AUTOMATIC LOCALIZED ANALYSIS OF
LONGITUDINAL CARTILAGE CHANGES

Liang Shan

A dissertation submitted to the faculty of the University of North Carolina at Chapel
Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in
the Department of Computer Science.

Chapel Hill
2014

Approved by:

Marc Niethammer

Stephen M. Pizer

Martin A. Styner

Hongtu Zhu

H. Cecil Charles

**ABSTRACT**

Liang Shan: AUTOMATIC LOCALIZED ANALYSIS OF
LONGITUDINAL CARTILAGE CHANGES
(Under the direction of Marc Niethammer)

Osteoarthritis (OA) is the most common form of arthritis; it is characterized by the loss of cartilage. Automatic quantitative methods are needed to screen large image databases to assess changes in cartilage morphology. This dissertation presents an automatic analysis method to quantitatively analyze longitudinal cartilage changes from knee magnetic resonance (MR) images.

A novel robust automatic cartilage segmentation method is proposed to overcome the limitations of existing cartilage segmentation methods. The dissertation presents a new and general convex three-label segmentation approach to ensure the separation of touching objects, i.e., femoral and tibial cartilage. Anisotropic spatial regularization is introduced to avoid over-regularization by isotropic regularization on thin objects. Temporal regularization is further incorporated to encourage temporally-consistent segmentations across time points for longitudinal data.

The state-of-the-art analysis of cartilage changes relies on the subdivision of cartilage, which is coarse and purely geometric whereas cartilage loss is a local thinning process and exhibits spatial non-uniformity. A novel statistical analysis method is proposed to study localized longitudinal cartilage thickness changes by establishing spatial correspondences across time and between subjects. The method is general and can be applied to other nonuniform morphological changes in other diseases.

**TABLE OF CONTENTS**

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1: INTRODUCTION

## 1.1 Motivation

Osteoarthritis (OA) is the most common form of arthritis and is a major cause of long-term disability in the US (Woolf and Pfleger, 2003). It is estimated that more than 16% of all adults 45 years or older suffer from symptomatic OA of the knee (CDC, 2008). The OA symptoms include swelling, pain, discomfort and problems in mobility and are caused by the progressive loss of joint cartilage (CDC, 2008). Figure 1.1 shows the anatomy of the human knee and illustrates the cartilage loss. While current treatment options help control pain, they are not able to reverse disease progression (Felson *et al.*, 2000b). Further drug research is essential to help OA patients.

Cartilage loss (Felson *et al.*, 2000a) is believed to be the dominating factor in OA. Studying cartilage morphological changes will help understand OA progression and drug effects. The only method currently accepted by regulators and clinicians for evaluating OA progression in hip or knee is through radiographs, which cannot measure cartilage directly but relies on joint-space width as a surrogate measure for cartilage thickness. Cartilage loss is usually thought of as nonuniform cartilage thinning and hence is difficult to quantify using x-ray projection images due to loss of three-dimensional spatial context. Magnetic resonance imaging (MRI) is a three-dimensional image technique able to directly measure cartilage volume and thickness and is thus more sensitive than radiographs in the detection of cartilage loss. Significant advances in MRI have resulted in the ability to quantify cartilage morphology (Eckstein *et al.*, 2006). Studies have shown that MRI allows reproducible quantification of cartilage morphology (Eckstein *et al.*, 1997, 1998; Raynauld, 2003; Raynauld *et al.*, 2004; Jones *et al.*, 2004; Cicuttini *et al.*, 2005;
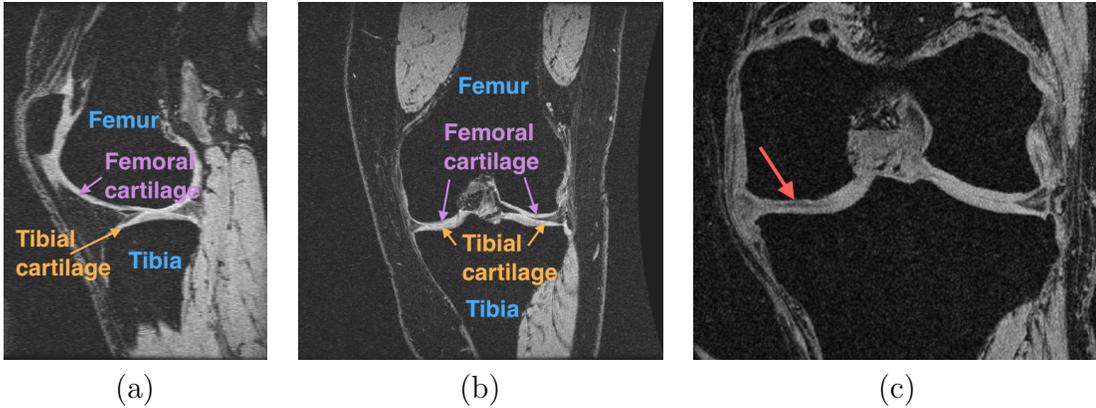
Figure 1.1: Example slice from T1 weighted MR images of human knee and illustration of cartilage loss. (a) A sagittal slice from a 3-dimensional T1 weighted MR image of a healthy knee. Bones are annotated in blue, femoral cartilage in purple and tibial cartilage in orange. (b) A coronal slice from the same MR image. (c) A coronal slice of an OA knee with cartilage loss indicated by the red arrow.

Eckstein *et al.*, 2006). Therefore, MRI is increasingly accepted as a primary method to evaluate progression of OA (Raynauld, 2003; Raynauld *et al.*, 2004; Cicuttini *et al.*, 2005).

Large image databases have been acquired for OA research. For instance, the database of the Osteoarthritis Initiative (Peterfy *et al.*, 2008) contains MR image data for 4,796 subjects. The Pfizer Longitudinal Study (PLS) (Eckstein *et al.*, 2008), a case-control study, consists of 97 normal control and 61 OA subjects. It would be of great value to fully analyze the MR images of these datasets in understanding the progression of OA. This would require robust fully-automatic computer-assisted image analysis methods because even if only a small amount of human intervention is needed for a single image, it would be extremely labor-intensive to analyze (tens of) thousands of images. Therefore a fully-automatic cartilage segmentation and longitudinal analysis from MR images is crucial to study OA.

In this thesis, **I present an end-to-end automatic method to study cartilage changes**, including cartilage segmentation, thickness computation, establishing spatial

correspondence and statistical analysis of longitudinal changes, among which the segmentation and the statistical analysis are the most challenging components. Cartilage registration is not a trivial task either because of the small size of the cartilage.

## 1.2 Cartilage segmentation

Automatic cartilage segmentation is a *challenging* problem in the field of medical image analysis for the following reasons.

1. Cartilage is a small and thin structure. There are many other tissues, e.g., muscles, in the MR image, which are irrelevant for the assessment of cartilage. This makes a *fully-automatic* segmentation challenging.

2. A segmentation method needs to be *robust* enough to be applied to large image datasets exhibiting variations in image appearance.

3. Femoral and tibial cartilage appear to be touching in an MR image, so conventional binary segmentation would merge them into a single object. Therefore, a new segmentation method is needed to produce separate segmentations for *touching objects*.

4. Isotropic spatial regularization as a means to deal with noisy image data is not an ideal choice for cartilage segmentation because it tends to shrink the segmentation boundary by cutting *thin* objects short.

5. Since subtle changes of cartilage might be indicative of early OA, it is desirable to study the longitudinal cartilage changes from a *temporally-consistent* segmentation which can mitigate image noise effects.

### 1.2.1 Existing methods

Automatic cartilage segmentation in the knee has been researched for decades. Many segmentation approaches have been proposed with varying levels of automation, e.g., region growing methods (Adams and Bischof, 1994), watershed methods (Beucher and Meyer, 1993), live wire (Mortensen and Barrett, 1995), active contours (Kass *et al.*, 1988; Caselles *et al.*, 1997a), model-based segmentation methods (Cootes *et al.*, 1995, 2001), graph-based methods (Felzenszwalb and Huttenlocher, 2004; Boykov and Kolmogorov, 2004), pattern recognition methods (Duda *et al.*, 2001), and atlas-based methods (Aljabar *et al.*, 2009). However, all the methods suffer from certain shortcomings, which are discussed below. A novel automatic segmentation method is proposed in this thesis to address the issues of the existing methods.

**Region growing methods**

The region growing algorithm (Adams and Bischof, 1994) iteratively examines neighboring pixels of "seed points" and determines whether the pixel neighbors should be added to the region. The method requires manual interaction to obtain the seed points and suffers from its sensitivity to image noise and parameter settings. Peterfy *et al.* (1994), Piplani *et al.* (1996), Eckstein *et al.* (1996, 1998) and Tamez-Pena *et al.* (1999) applied region growing methods to obtain *semi-automatic* cartilage segmentations.

**Watershed methods**

The watershed transform (Beucher and Meyer, 1993) is a segmentation method originating from the field of mathematical morphology. The image is considered as a topographic relief, where the height at each pixel is directly related to its gray level. If rain falls on the terrain, the watersheds correspond to the limits of the adjacent catchment

basins of the drops of water. Generally, the watershed transform is computed on the gradient magnitude image, so the boundaries are located at high gradient pixels. The main drawbacks of the watershed transform (Grau *et al.*, 2004) are over segmentation, sensitivity to noise, poor detection of areas with low contrast boundaries and thin structures. Ghosh *et al.* (2000) and Grau *et al.* (2004) used watershed method for *semi-automatic* cartilage segmentation.

**Live wire**

The live wire (also known as intelligent scissors) (Mortensen and Barrett, 1995) is an interactive segmentation method which allows the user to choose a contour by roughly tracing the object boundary. The *local* minimum cost path from the current cursor position to the last seed point is computed. Additional user interaction is needed if the computed path deviates from the desired one. The major limitation of this method is that heavy user interaction is necessary when many alternative minimal paths may exist. This method was applied by Steines *et al.* (2000) and Gougoutas *et al.* (2004) to compute *semi-automatic* cartilage segmentation in a slice-by-slice manner.

**Active contours**

An active contour model, also known as a snake, is an energy-minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges (Kass *et al.*, 1988). The active contour method had a tremendous impact in the segmentation community. The original formulation is not geometric, meaning there is not obvious relation between the parametrization of the contour and the geometry of the objects to be captured (Angenent *et al.*, 2006). Many extensions to and variations of the snakes have been proposed to address this issue. Caselles *et al.* (1997a) proposed a geometric model named geodesic active contours which minimizes a

weighted length of a closed curve. The curve that minimizes the weighted length will prefer to be on the image edges. Despite the popularity of these methods, most active contour methods suffer from a major drawback, high sensitivity to initialization, which is related to the non-convexity of the energy functional. The segmentation algorithm tends to get stuck in undesired local minima. Lynch *et al.* (2000), Kauffmann *et al.* (2003) and Brem *et al.* (2009) reported *semi-automatic* cartilage segmentation methods with active contour models.

The initial implementations of active contours used parameterizations which lead to non-convex optimization problems. More modern approaches (Appleton and Talbot, 2006) are more closely related to graph-cuts and are based on convex energy functions (see below).

**Graph-based methods**

Graph-based image segmentation techniques (Felzenszwalb and Huttenlocher, 2004) represent an image using a graph where each node corresponds to a pixel in the image and the edges connect certain pairs of neighboring pixels. A weight is associated with each edge based on some property of the pixels that it connects, such as image intensities. Then a segmentation corresponds to a cut of the graph. A global optimal solution may be obtained for graph cuts. Boykov and Kolmogorov (2004) presented an efficient way to compute the max-flow for computer vision related graph. The graph-based methods, partitioning the image with a graph cut, solve essentially the same optimization problem as active contours, which uses a curve to separate objects. However, the solution method is fundamentally different because one searches a labeling of graph nodes rather than the location of a contour as in the traditional ways of implementing active contours.

Shim *et al.* (2009) and Bae *et al.* (2009) utilized graph cuts to achieve *semi-automatic* cartilage segmentation guided by the placement of seed points by the user. Yin *et al.*

(2010b) proposed a method for simultaneous segmentation of multiple interacting objects, named LOGISMOS (layered optimal graph image segmentation of multiple objects and surfaces). The method incorporates multiple spatial inter-relationships in a single $n$-dimensional graph. Yin *et al.* (2010b) also presented an *automatic* method to segment cartilage using LOGISMOS.

Graph-based methods generally suffer from grid bias, also known as metrication errors. Large neighborhood systems are required to remove the bias, but they result in a large number of edges and thus increase the memory consumption.

## Model-based methods

In model-based segmentation, a statistical model is trained from a population of manual segmentations. The segmentation of a target image is computed iteratively. The model is fit to the target image and then refined based on the model and image features. Active shape models (ASM) (Cootes *et al.*, 1995) and active appearance models (ASM) (Cootes *et al.*, 2001) are the most frequently employed methods. Solloway *et al.* (1997) used ASM to segment femoral cartilage in a *semi-automatic* way. Vincent *et al.* (2010) applied AAM to compute *automatic* cartilage segmentation. Glocker *et al.* (2007) and Seim *et al.* (2010) proposed different statistical shape models for bone/cartilage segmentation.

Despite the wide application, these model-based methods are prone to local minima in the fitting process. In addition, these models are usually restricted to a given shape/appearance space and hence may not be able to capture pathologies well.

## Pattern recognition methods

In pattern recognition, the $k$ nearest neighbor ($k$NN) algorithm is a method for classifying objects based on closest training examples in the feature space. Folkesson *et al.*

(2007) proposed an *automatic* hierarchical voxel classification scheme using $k$NN. Dam and Loog (2008) accelerated the process through sparse classification. A support vector machine (SVM) (Cortes and Vapnik, 1995) constructs a hyperplane, which has the largest distance to the nearest training data point of any class, in a high-dimensional feature space for classification. Koo *et al.* (2009) proposed to use SVM to segment cartilage *automatically* from multi-contrast MR images. The spatial interactions between neighboring pixels are neglected in these methods. Nevertheless, it is straightforward to use these classifiers in combination with other methods. For example, one can use them to get likelihoods for foreground and background, which can then be used as unary terms in graph-cuts, and so on.

**Atlas-based methods**

An atlas (Aljabar *et al.*, 2009), 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 (Išgum *et al.*, 2009), namely single-atlas-based, average-shape atlas-based and multi-atlas-based methods. A statistical atlas was constructed by Glocker *et al.* (2007) for the patella cartilage from a set of pre-registered training images. The segmentation is solved through a single registration and thus is not robust to occasional registration failures.

In multi-atlas-based segmentation, multiple labeled images are registered to the query image independently. The propagated atlas labels are then fused into a single segmentation of the query image. This results in a method which is more robust to registration failures than single-atlas-based and average-shape-atlas-based segmentation methods, both of which need only one registration. The downside of multi-atlas-based segmentation is the high computation cost since it needs multiple registrations. Rohlfing *et al.* (2004) demonstrated that the multi-atlas-based segmentation is more accurate than the

other two types of atlas-based segmentation methods. In spite of the expensive computation, multi-atlas-based segmentation has been quite popular and successful in brain imaging (Aljabar *et al.*, 2009). Tamez-Peña *et al.* (2012) reported a multi-atlas-based cartilage segmentation approach.

### 1.2.2 Proposed method

**I propose a novel automatic cartilage segmentation method** which has the following advantages. First of all, the method is fully automatic and requires no user interaction (besides quality control). Secondly, the method is *robust* as it benefits from the multi-atlas-based strategies. Thirdly, both spatial and local appearance information are utilized in the segmentation. Local tissue classification is *probabilistic* (unlike Folkesson *et al.* (2007), Dam and Loog (2008) and Koo *et al.* (2009)), and it is combined with a spatial prior to generate the final segmentation within a segmentation model. Furthermore, the segmentation model is *convex* and thus allows for the computation of global optimal solutions, which cannot be guaranteed by active contour models (traditional implementations), ASM or AAM. The continuous formulation of the segmentation model makes the method free of grid bias that graph-based methods suffer from. The segmentation model also allows for the incorporation of spatial and temporal regularization.

It is difficult to segment the cartilage directly from the MR image due to its small size. Bones are relatively easy to segment. Their segmentations can help locate the cartilage given the spatial relationship between the bone and the cartilage. Given the advantages and success of multi-atlas-based methods, **I propose a multi-atlas-based bone and cartilage segmentation method**. I first compute the multi-atlas spatial priors for the bone. The bone priors are then combined with local likelihoods within a Bayesian model to generate the bone segmentation. The multi-atlas registration of the cartilage is then based on the corresponding bone segmentation. The cartilage segmentation is a joint
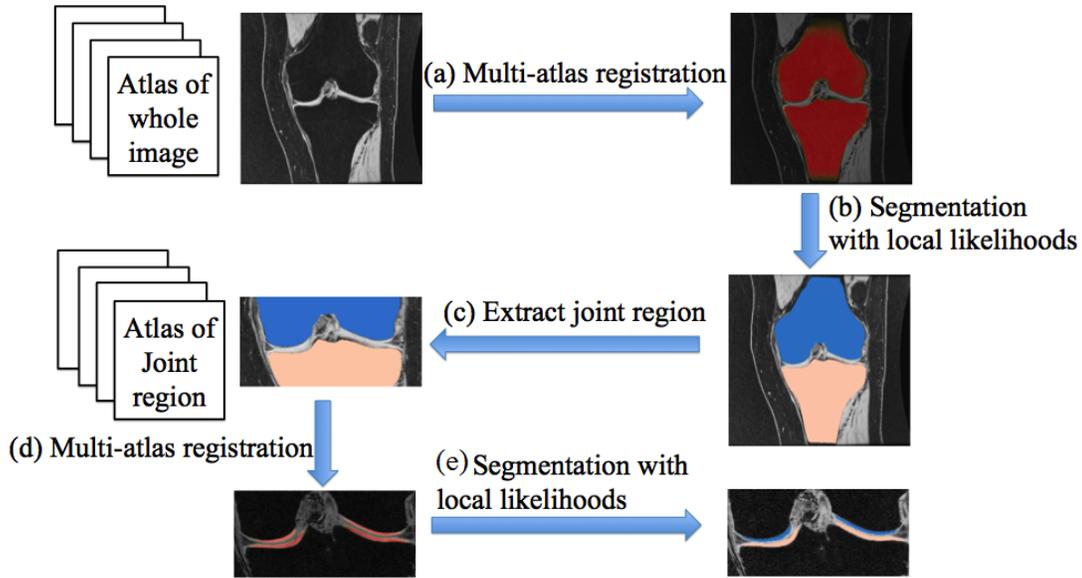
Figure 1.2: Cartilage segmentation pipeline. (a) Multi-atlas registration to compute the spatial prior for the bone. (b) Compute bone segmentation from the spatial prior and the local likelihood. (c) Extract the joint region. (d) Multi-atlas registration (based on bone segmentations) to compute the spatial prior for the cartilage. (e) Compute cartilage segmentation from the spatial prior and the local likelihood.

decision of the multi-atlas spatial priors and local likelihoods from a probabilistic tissue classification. Figure 1.2 shows the pipeline of multi-atlas cartilage segmentation.

The segmentation model for computing cartilage segmentation at step (e) in Fig. 1.2 is critical because the femoral and tibial cartilage appear to be touching in the MR image and it is desirable to separate them. Since binary segmentation methods tend to merge touching objects into a single one, **I propose a general three-label segmentation approach** which guarantees the separation of touching objects. The three-label segmentation is formulated as a convex optimization problem and therefore allows for the computation of global optimal solutions. The method is general, and it can be applied to other segmentation problems with two touching objects.

The commonly used isotropic spatial regularization is not an ideal choice for cartilage segmentation because it has a tendency to over-regularize at the tips of thin objects.

Therefore, **I propose an anisotropic spatial regularization for cartilage segmentation** and demonstrate it improves the segmentation accuracy in comparison to the isotropic one.

Temporally-consistent segmentation is desirable to study longitudinal change because subtle changes might be indicative of the disease. **A longitudinal three-label segmentation is proposed** to encourage temporal consistency across segmentations for longitudinal data. The longitudinal segmentation method is also general, and it can be applied to other segmentation problems with longitudinal data of two touching objects.

## 1.3   Cartilage thickness analysis

Statistical analysis of cartilage thickness changes is also a *challenging* problem for the following reasons.

1. Observational studies have shown that cartilage loss in OA may not be uniform throughout the cartilage but may affect certain regions (e.g., the center) more frequently and more strongly than others (Biswal *et al.*, 2002). Therefore, a localized analysis method is necessary to understand localized cartilage thinning. The localized analysis requires establishing spatial correspondence across time and subjects which is challenging, due to the small spatial amount of cartilage.

2. The other challenge is caused by the fact that cartilage thinning may happen to different locations for different subjects. Cartilage thinning at a particular location could happen to only a few subjects. Treating all OA subjects equally and using all of them for statistical analysis would be problematic.

### 1.3.1 Existing methods

Wirth and Eckstein (2008) studied average cartilage thickness change in defined anatomic subregions of the femorotibial joint. The tibial cartilage was divided into a central area of the total subchondral bone area and into anterior, posterior, internal, and external subregions surrounding it. In the weight-bearing femoral cartilage, central, internal, and external subregions were determined. However, this approach may be problematic because changes within a specific subregion could happen only to a few subjects whereas other subjects have strong progressions in different subregions.

Buck *et al.* (2009) proposed to use ordered values of subregional change in each subregion to focus on the thickness change alone regardless of in which subregion the change occurs in each subject. The ordered value approach was demonstrated to provide improved discrimination between healthy subjects and OA participants longitudinally. But the ordered value approach still relies on subdivisions of the cartilage and the average thickness within each subregion.

The subdivisions of the cartilage are purely geometric and necessarily coarse. Local changes (that happen to a smaller region than the size of a subregion) are weakened by averaging over a particular subregion and are impossible to recover. To fully understand the spatial pattern of OA progression, the analysis of *localized* cartilage thickness changes is necessary.

Williams *et al.* (2010) built spatial correspondence of cartilage thickness maps and studied local cartilage thickness. The method however treated all OA subjects equally without considering the spatial heterogeneity.

### 1.3.2 Proposed method

A particular reason for the lack of studies on localized cartilage changes is the challenging cartilage alignment problem because of its small volume. In my thesis, **I develop**

**a method to establish spatial correspondence of cartilage thickness maps be-
tween subjects and across time points.** The cartilage thickness is computed from the
obtained cartilage segmentation using a Laplace-equation approach (Yezzi and Prince,
2003) and then transformed to a common (the atlas) space based on the bone segmenta-
tion. The cartilage thickness maps are then comparable across time points and between
subjects.

Statistical analysis on localized cartilage changes is also challenging due to the fact
that cartilage thinning may happen at different locations for different subjects. The
heterogeneity of the longitudinal changes is not uncommon in the progression of other
diseases. **I propose a novel method to analyze the localized cartilage changes,
which can also be applied to study nonuniform local morphological changes
in other disease.** I first identify the thinning locations for each subject and then group
subjects into different clusters based on their thinning locations. The group difference
between normal control and OA subjects will be studied at each thinning location.

## 1.4    Thesis and contribution

Thesis: *Automatic, robust and accurate cartilage segmentations can be obtained through
multi-atlas-based registration and local tissue classification within a three-label segmenta-
tion framework allowing for spatial and temporal regularization. Spatially transforming
cartilage thickness maps into an atlas space enables statistical analysis on localized car-
tilage changes. Clustering of OA subjects improves statistical analysis due to the spatial
heterogeneity of cartilage loss.*

The scientific contributions are as follows:

1. I present a novel and general three-label segmentation method that ensures the sep-
   aration of touching objects and allows for spatial and temporal regularization. The
   three-label segmentation is formulated as a convex optimization problem, which

allows for the computation of global optimal solutions. The method can be applied to other segmentation problems with two touching objects.

2. I validate the automatic three-label cartilage segmentation on a sizable dataset consisting of more than 700 images. Specially, I study the impacts of different types of atlases (namely average-shape-atlas and multi-atlas) and different types of regularization (i.e., isotropic spatial regularization, anisotropic spatial regularization and temporal regularization) on cartilage segmentation accuracy.

3. I propose a novel method to analyze nonuniform localized cartilage changes, which can be applied to study nonuniform local morphological changes in other diseases.

4. I perform statistical analysis on a sizable longitudinal dataset containing about 150 subjects with 5 time points. The statistical analysis result of localized cartilage changes is presented and compared to that reported in literature using regional analysis.

The engineering contributions are the following:

1. I develop a new fully automatic three-label cartilage segmentation pipeline based on multi-atlas registration and local tissue classification.

2. I develop a method to establish spatial correspondences of knee cartilage across time points and between subjects, which allows for statistical analysis on localized cartilage thickness changes.

## 1.5   Overview of chapters

The remainder of this dissertation is organized as follows:

Chapter 2 presents a novel three-label segmentation approach which overcomes the limitation of the binary segmentation on touching objects.

Chapter 3 applies the proposed three-label segmentation to obtain automatic cartilage segmentation through multi-atlas-based registration and local tissue classification.

Chapter 4 presents the validation of the proposed cartilage segmentation method and compares the proposed method to existing methods.

Chapter 5 discusses the computation of cartilage thickness, transforming the cartilage thickness maps into an atlas space, proposes a new clustering-based statistical method, and presents the results of applying the new method on localized cartilage thickness changes.

Chapter 6 contains a discussion of the contributions and an outline of future work.

# CHAPTER 2: THREE-LABEL SEGMENTATION

## 2.1 Introduction

Image segmentation is a fundamental problem in the field of computer vision and medical image analysis. It is an essential component towards, for example, automated vision systems and medical applications. The aim is to find a partition of an image into a finite number of semantically important regions. Two classes of methods have gained tremendous popularity in the past decades. One class is active contour methods, including snakes (Kass *et al.*, 1988) and geodesic active contours and surfaces (Caselles *et al.*, 1997a,b). The other class takes a different approach by transforming the segmentation problem to a graph problem, e.g., graph cuts (Ford and Fulkerson, 1962).

Active contour methods evolve a curve or surface toward a structure of interest in an image. A major issue to these curve/surface evolution approaches methods is the sensitivity to the initialization. The quality of the segmentation result depends heavily on the choice of the initial contour, which means a bad initial condition can give an unsatisfactory result. The dependency on the initialization is due to the non-convexity of the objective energy functional. The global minimum of the energy corresponds to a contour of length zero, which has no practical sense for the image segmentation task. Therefore, a global optimal solution is not a desired result of these methods. It is difficult to pick the desired local minima from numerous local minima due to image noise and irrelevant objects. A number of heuristics have been proposed to avoid unwanted local minima, e.g., pressure forces designed to overcome shallow minima (Cohen, 1991), multi-resolution approaches to focus on objects that persist at high scales, and methods modifying the gradient descent to favor more significant contours (Xu and Prince,

1998). In spite of the advent of these heuristics, active contours typically require manual intervention which limits their application.

Graph-based image segmentation techniques (Felzenszwalb and Huttenlocher, 2004) represent an image using a graph where each node corresponds to a pixel in the image and the edges connect certain pairs of neighboring pixels. A weight is associated with each edge based on some property of the pixels that it connects, such as image intensities. Then a segmentation corresponds to a cut of the graph. A global optimal solution may be obtained for graph cuts. Boykov and Kolmogorov (2004) presented an efficient way to compute the max-flow. However, graph cuts suffer from discretization artifacts, which typically result in a preference for contours and surfaces to travel along the grid directions. Large neighborhood systems are required to remove the bias but results in a large number of edges and thus increase the memory consumption. Also, parallel implementations are not straightforward (Delong and Boykov, 2008).

## 2.2   Related work

Chan-Vese segmentation (Chan and Vese, 2001) aims at finding an optimal image partition into a uniform foreground and background region. It is classically formulated as a curve/surface-based optimization problem of the form

$$E(c_1, c_2, C) = Length(C) + \lambda_1 \int_{inside(C)} (c_1 - I(\mathbf{x}))^2 \, d\mathbf{x} + \lambda_2 \int_{outside(C)} (c_2 - I(\mathbf{x}))^2 \, d\mathbf{x}$$

(2.1)

where $I(\cdot)$ denotes image intensities, $\mathbf{x}$ is a pixel/voxel location, $c_1$ and $c_2$ are the intensity estimates for the interior and the exterior of the curve $C$ (their means). This is a non-convex optimization problem, which complicates the solution. In particular, the obtained solution may only be a local optimum of the optimization problem.

Appleton and Talbot (2006) recast active contours as a convex optimization problem by formulating the problem with respect to a labeling (indicator) function for foreground and background rather than with respect to a boundary curve/surface. Their approach can efficiently compute globally minimal curves and surfaces for image segmentation, stereo reconstruction, and other labeling problems. The method is isotropic (graph cuts are grid-biased) and optimal (active contours are suboptimal). The algorithm produces a globally maximal continuous flow at convergence. The energy functional Appleton and Talbot (2006) minimized is

$$E(u) = \int_\Omega g(\mathbf{x}) \, d\mathbf{x} \text{ with seed points}, u \in [0, 1] \tag{2.2}$$

where $\Omega$ is the spatial domain of the image, $\mathbf{x}$ is a pixel/voxel location in $\Omega$. Here, $u$ is a essentially binary labeling function for foreground ($u = 1$) and background ($u = 0$). And $g$ is an edge-indicator function, small at edges and large in uniform regions. The globally optimal solution $u$ can be computed efficiently. The globally minimal surface can be obtained trivially from the optimal $u$.

Bresson *et al.* (Bresson *et al.*, 2007) proposed the following convex functional for segmentation based on the Chan-Vese energy (Chan and Vese, 2001) to optimally partition the image into piecewise constant regions:

$$E(u) = \int_\Omega g\|\nabla u\| \, d\mathbf{x} + \lambda \int_\Omega \overbrace{\left((c_1 - I(\mathbf{x}))^2 - (c_2 - I(\mathbf{x}))^2\right)}^{r} u \, d\mathbf{x}, \quad u \in [0, 1]. \tag{2.3}$$

where $I(\cdot)$ denotes image intensities, $c_1$ and $c_2$ are the mean intensity estimates for the interior and the exterior of the segmentation respectively, $\Omega$ is the image domain, $u$ is an essentially binary labeling function, indicating foreground (1) and background (0), and $g$ is an edge-indicator function, small at edges and large in uniform regions. This model unifies the Chan-Vese model and the active contour model. The formulation (2.3)

is convex when the $c_1$ and $c_2$ are fixed. Chan-Vese method is not convex because $c_1$ and $c_2$ are optimized jointly with the segmentation.

In Shan *et al.* (2010), I proposed a novel binary segmentation extending the work by Bresson *et al.* The region-based term $r$ in (2.3) can be reformulated in a probabilistic way. One can replace $r$ by $\log \frac{P_i}{P_o}$ with $P_i$ and $P_0$ probabilities of a pixel belonging to foreground and background respectively (where equal Gaussian probability distributions result in the Chan-Vese energy).

To encourage segmentation stop at the desired boundary, a feature field $\mathbf{F}$ is added to favor boundaries with inward normal directions aligned with $\mathbf{F}$ itself. Thus, the overall optimization problem is to minimize the following convex energy functional

$$E(u) = \int_\Omega g\|\nabla u\| + ru + \mathbf{F} \cdot \nabla u \ d\Omega, \quad r = \log \frac{P_o}{P_i}. \tag{2.4}$$

The vector field $\mathbf{F}$ can be rewritten using the divergence theorem and integrated into the regional term $r$. The resulting segmentations are independent of initializations of function $u$ due to the convexity of energy functional (2.4).

All the methods discussed above result in binary segmentations, in which each pixel is labeled either foreground or background. Multi-label segmentation assigns distinct labels to different foreground objects and therefore allows for distinction among objects of interest. In the next section, I will discuss the advantage of multi-label segmentation through a synthetic example and present a three-label segmentation framework.

## 2.3 Three-label segmentation

In this section, I will discuss the limitations of binary segmentations and propose a novel convex three-label segmentation framework, which ensures the separation of the two foreground objects, e.g., femoral and tibial cartilage.
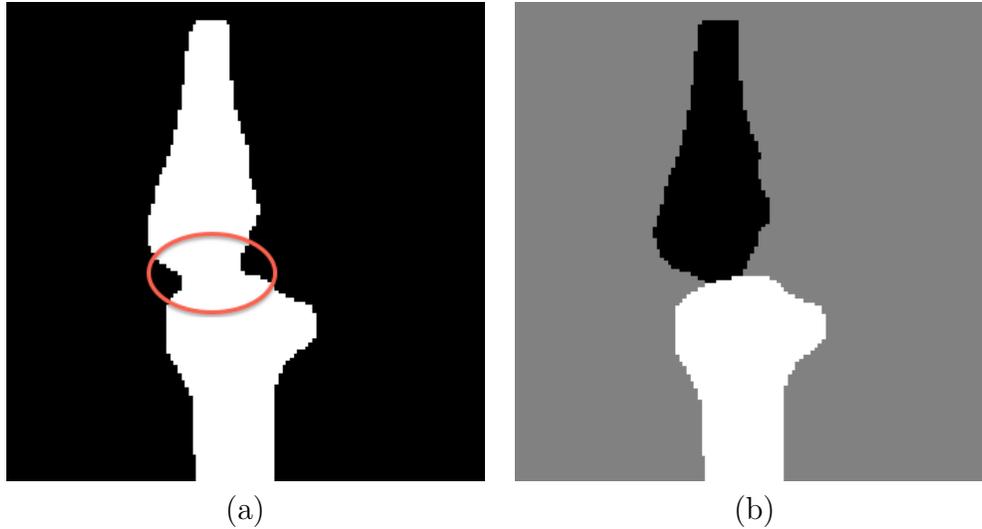
Figure 2.1: Synthetic example comparing binary and three-label segmentations. (a) Binary segmentation result. Femur and tibia are segmented as one object and the boundary in the joint region is not captured well due to regularization effects. (b) Proposed three-label segmentation. The boundaries between bones and background are preserved.

### 2.3.1    Limitation of binary segmentations

Only one object is segmented from binary segmentation, which uses two labels, i.e., foreground and background. In many cases, there is more than one object which needs to be segmented. If the objects of interest are spatially separate, a binary segmentation may still be sufficient (one can decide after the fact which one is which). But if objects touch, binary segmentation is not enough. Also because of spatial regularization, the joint boundary of touching objects will be smoothed out. Multi-label segmentation can keep objects separated and is therefore particularly suited to segment touching objects. Figure 2.1 demonstrates the limitations of a binary versus a three-label segmentation method for a synthetic bone case. In the binary segmentation (Fig. 4.3(a)), the two bones are regarded as one object and the joint boundary (where the two bones touch) is smoothed out due to spatial regularization. The three-label segmentation (Fig. 4.3(b)) distinguishes the two bones with different labels. The joint boundary is preserved because of the distinct labels.

### 2.3.2 A general three-label segmentation method

The three-label case is a specialization of the multi-label segmentation method by Zach *et al.* (2009). Using three labels allows for a symmetric formulation with respect to the background segmentation class.

A multi-label segmentation is a mapping from an image domain $\Omega$ 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 $\Omega$ to label $\Lambda(\mathbf{x})$ in label space. The goal is to find a labeling function that minimizes an energy functional of the form:

$$E(\Lambda) = \int_\Omega c(\mathbf{x}, \Lambda(\mathbf{x})) + V(\nabla\Lambda, \nabla^2\Lambda, ...) \, d\mathbf{x}, \tag{2.5}$$

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. This optimization problem is generally difficult to solve as the data costs are typically highly non-convex. By embedding the labeling assignment into higher dimensions and appropriately restating the energy (2.5), a convex formulation can be obtained.

The different labelings can be encoded through a level function $u$ defined in higher dimensions

$$u(\mathbf{x}, l) = \begin{cases} 1 & \text{if } \Lambda(\mathbf{x}) < l, \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

which maps the Cartesian product of the image domain $\Omega$ 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 2.2 illustrates the relation between $u$ and $\Lambda$ for the three-label case.

With the level function $u$, the label cost term can be rewritten as

$$\int_\Omega c(\mathbf{x}, \Lambda(\mathbf{x})) = \int_\mathcal{D} c(\mathbf{x}, l) \, |\nabla_l u(\mathbf{x}, l)| \, d\mathbf{x}dl, \ \mathcal{D} = \Omega \times \mathcal{L}, \tag{2.7}$$

using the same technique as Pock *et al.* (2008) proposed. The regularization energy in 2.5) can be written in terms of the gradient of the level function as

$$\int_{\mathcal{D}} \psi_{\mathbf{x},l}(\nabla u) d\mathbf{x} dl, \ \mathcal{D} = \Omega \times \mathcal{L}, \tag{2.8}$$

where $\psi_{\mathbf{x},l}$ defines the regularization term and is convex and 1-positively homogeneous. By combining the two terms, one can rewrite (2.5) purely in terms of $u$,

$$E(u) = \int_{\mathcal{D}} \psi_{\mathbf{x},l}(\nabla u) + c(\mathbf{x},l) |\nabla_l u(\mathbf{x},l)| \ d\mathbf{x} dl, \ \mathcal{D} = \Omega \times \mathcal{L}. \tag{2.9}$$

As $u$ can only take two integer values, i.e., 0 and 1, the functional (2.9) is not convex. The non-convex constraint is usually relaxed to a convex one on $u \in [0,1]$. The relaxation transforms the original hard optimization problem into a related problem that is convex and therefore easier to solve. Zach *et al.* (2009) has proven that the resulting solution $u$ to the relaxed problem is essentially binary and thresholding of an essentially binary optimal solution of energy (2.9) yields to an equally globally optimal solution to the original discrete problem.

Model (2.5) is in general asymmetric with respect to the labels, since the design of the level function (2.6) 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 that treats the two objects symmetrically.
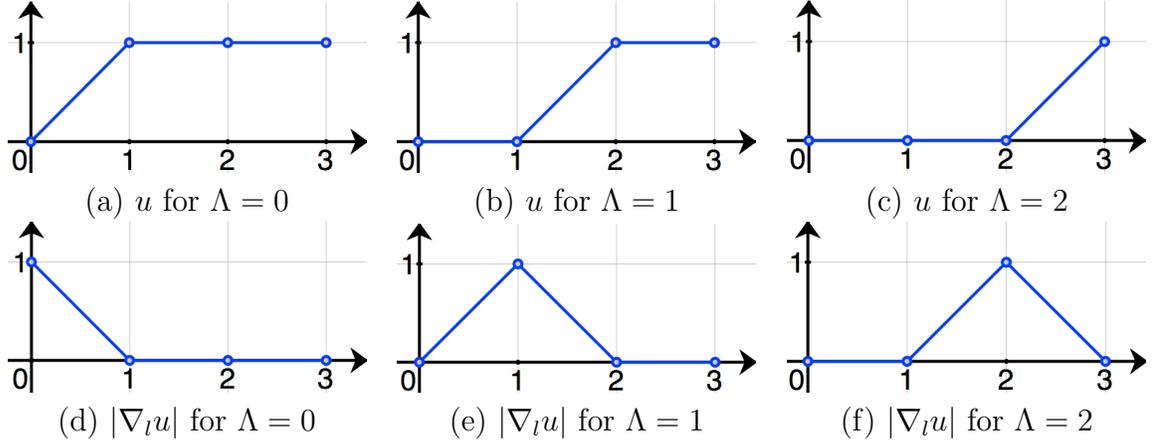
Figure 2.2: Values of $u$ and $|\nabla_l u|$ for different label assignments in a three-label segmentation (abscissa $l$). Assuming a discretization with forward differences. $|\nabla_l u|$ determines the label assignment.

### 2.3.3 Three-label segmentation with isotropic regularization

The three-label segmentation energy functional is formulated as (the explicit dependence of $\mathbf{x}$ and $l$ is omitted)

$$E(u) = \int_{\mathcal{D}} g \, \|\nabla_{\mathbf{x}} u\| + c \, |\nabla_l u| \ d\mathbf{x} dl, \ \mathcal{D} = \Omega \times \mathcal{L}$$

$$\text{subject to } u \in [0,1], \ u(\mathbf{x}, 0) = 0, \ u(\mathbf{x}, 3) = 1.$$

(2.10)

This is relaxed optimization problem with a convex constraint $u \in [0,1]$ as opposed to the original problem with the non-convex constraint $u \in \{0,1\}$.

Minimizing energy (2.10) 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 $u$, $\nabla_{\mathbf{x}} u = (\partial u/\partial x, \partial u/\partial y, \partial u/\partial y)^T$ 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. Parameter $g$ is set to be a non-negative constant. The formulation is convex, so a global optimum can be computed. One can apply an iterative gradient descent/ascent scheme for the optimization. See section 2.3.5 for details. The three-label

23
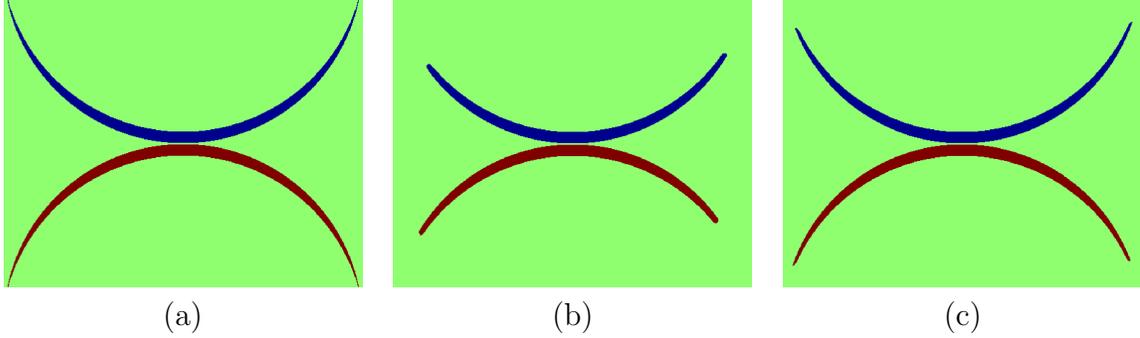
Figure 2.3: Synthetic example comparing isotropic and anisotropic regularization. (a) original image to be segmented; (b) and (c) three-label segmentation results with isotropic and anisotropic regularization respectively. Anisotropic regularization avoids over-regularization at the tips of the synthetic shape.

segmentation can then be computed from the discontinuity set of $u$.

### 2.3.4   Three-label segmentation with anisotropic regularization

The isotropic regularization in model (2.10) treats all directions equally, which is not an ideal choice for long and thin objects like the cartilage. To customize the segmentation model (2.10) for cartilage segmentation, I replace the isotropic regularization term, $g$, by an anisotropic one

$$E(u) = \int_{\mathcal{D}} \|\mathbf{G}\nabla_{\mathbf{x}} u\| + c|\nabla_l u| \ d\mathbf{x}dl, \ \mathcal{D} = \Omega \times \mathcal{L},$$

$$u \in [0,1], \ u(\mathbf{x},0) = 0, \ u(\mathbf{x},3) = 1, \tag{2.11}$$

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 2.3 illustrates the problem with isotropic regularization which tends to shrink the segmentation boundary by cutting thin objects short and the benefit from anisotropic regularization. I choose $\mathbf{G}$ as

$$\mathbf{G} = g\left[\mathbf{I} + (\alpha - 1)\mathbf{n}\mathbf{n}^T\right], \alpha \in [0,1], \tag{2.12}$$

24
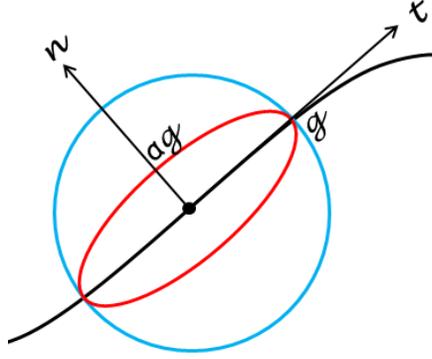
Figure 2.4: Difference between isotropic and anisotropic regularization. The black curve is an edge in an image. The regularization is illustrated at a pixel (the dot). The blue circle indicates the isotropic case where regularization is enforced equally in every direction. The red ellipse shows the anisotropic situation where less regularization is applied in the normal direction and more in the tangent direction.

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. 2.4 for an illustration of isotropic versus anisotropic regularization. Since the normal direction to the cartilage surface is not known a-priori, I approximate it by the normal direction to the bone-cartilage interface which can be determined from the segmentations of femur and tibia. The energy functional (2.11) is also convex and therefore a global optimum can also be computed. Again, an iterative gradient descent/ascent scheme is applied for the optimization. See section 2.3.5 for details.

### 2.3.5 Numerical solution

This section discusses an iterative scheme to optimize (2.11). Solving (2.10) is a special case with $\mathbf{G} = g\mathbf{I}$ ($\mathbf{I}$ is the identity matrix). I introduce two dual variables $\mathbf{p}$ (a vector field) and $q$ (a scalar field) and rewrite (2.11) as

$$E(u, \mathbf{p}, q) = \int_{\mathcal{D}} \langle \mathbf{p}, \mathbf{G}\nabla_{\mathbf{x}}u \rangle + q\nabla_l u \ d\mathbf{x}dl,$$

$$\text{subject to } \|\mathbf{p}\| \leq 1, \ |q| \leq c,$$

(2.13)

in which $\langle \cdot, \cdot \rangle$ represents inner products. Minimizing (2.11) with respect to $u$ is equivalent to minimizing (2.13) with respect to $u$ and maximizing it with respect to $\mathbf{p}$ and $q$.

Taking the variation yields

$$\delta E(u, \mathbf{p}, q; \delta u, \delta \mathbf{p}, \delta q) \tag{2.14}$$

$$= \frac{\partial}{\partial \epsilon} \int_{\Omega} \langle \mathbf{p} + \epsilon \delta \mathbf{p}, \mathbf{G} \nabla_{\mathbf{x}} u + \epsilon \mathbf{G} \nabla_{\mathbf{x}} \delta u \rangle + (q + \epsilon \delta q)(\nabla_l u + \epsilon \nabla_l \delta u) \ d\mathbf{x} dl \tag{2.15}$$

$$= \int_{\Omega} \langle \mathbf{p}, \mathbf{G} \nabla_{\mathbf{x}} \delta u \rangle + \langle \delta \mathbf{p}, \mathbf{G} \nabla_{\mathbf{x}} u \rangle + q \nabla_l \delta u + \delta q \nabla_l u \ d\mathbf{x} dl \tag{2.16}$$

$$= \int_{\Omega} (- \operatorname{div}(\mathbf{G} \mathbf{p}) - \nabla_l q) \delta u + \langle \delta \mathbf{p}, \mathbf{G} \nabla_{\mathbf{x}} u \rangle + \delta q \nabla_l u \ d\mathbf{x} dl \tag{2.17}$$

The gradient descent/ascent update scheme of (2.13) is

$$\mathbf{p}_t = -\mathbf{G} \nabla_x u, \ \|\mathbf{p}\| \le 1 \tag{2.18}$$

$$q_t = -\nabla_l u, \ |q| \le c \tag{2.19}$$

$$u_t = -\operatorname{div}_{\mathbf{x}}(\mathbf{G} \mathbf{p}) - \nabla_l q \tag{2.20}$$

The iterative scheme will lead to a global optimum upon convergence (Appleton and Talbot, 2006) because of the convexity of (2.11). 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 $\overset{\circ}{\mathcal{D}} = \mathcal{D} \setminus (S \cup T) = \Omega \times \{1, 2\}$. Zach *et al.* (2009) proposed to terminate the iterations when the gap between primal energy and dual energy is sufficiently small. The dual energy is computed as

$$\begin{aligned} E^*(u) &= \int_S \operatorname{div}(\mathbf{G} \mathbf{p}) + \nabla_l q \ d\mathbf{x} dl \\ &+ \int_{\overset{\circ}{\mathcal{D}}} \min(0, \operatorname{div}(\mathbf{G} \mathbf{p}) + \nabla_l q) \ d\mathbf{x} dl \end{aligned} \tag{2.21}$$

Zach *et al.* (2009) proved that the solution $u$ is essentially binary and monotonically

increasing after convergence and thresholding of the essentially binary solution $u$ yields an equally globally optimal solution to the original discrete problem.

## 2.4 Multi-label segmentation

The three-label segmentation (2.10 and 2.11) is a special case of multi-label segmentation. The background label is placed between the two foreground labels so that the same amount (if desired, can also be intentionally different) of spatial regularization is applied to each object with respect to the background. Such symmetric labeling is impossible when the number of labels is greater than three. This section therefore presents a multi-label segmentation method which requires no such symmetric setting and is capable of handling any number of objects.

### 2.4.1 Limitation of three-label segmentation

Figure 2.5 shows the regularization issues when applying (2.10) to segment four objects from the background. A symmetric labeling is not possible with more than two objects for model (2.10). Objects with labels closer to the background label get less regularized. In Fig. 2.5, the background is labeled zero so the object with largest label gets the most regularization. As the parameter $g$ in energy (2.10) increases, the objects disappear in the descending order of their labels.

### 2.4.2 Multi-label segmentation

The issue comes from the spatial regularization term $g \left\| \nabla_{\mathbf{x}} u \right\|$ in the energy (2.10) because the penalty depends on label ordering. Figure 2.9 (a) provides an illustration of the dependence. Each pixel is associated with a vector $u$ with six elements (first one is always zero, and last one is always one according to definition). Changing the label from the background ($u = [0, 1, 1, 1, 1, 1]$) to object 1 whose $u = [0, 0, 1, 1, 1, 1]$ gets penalty

27
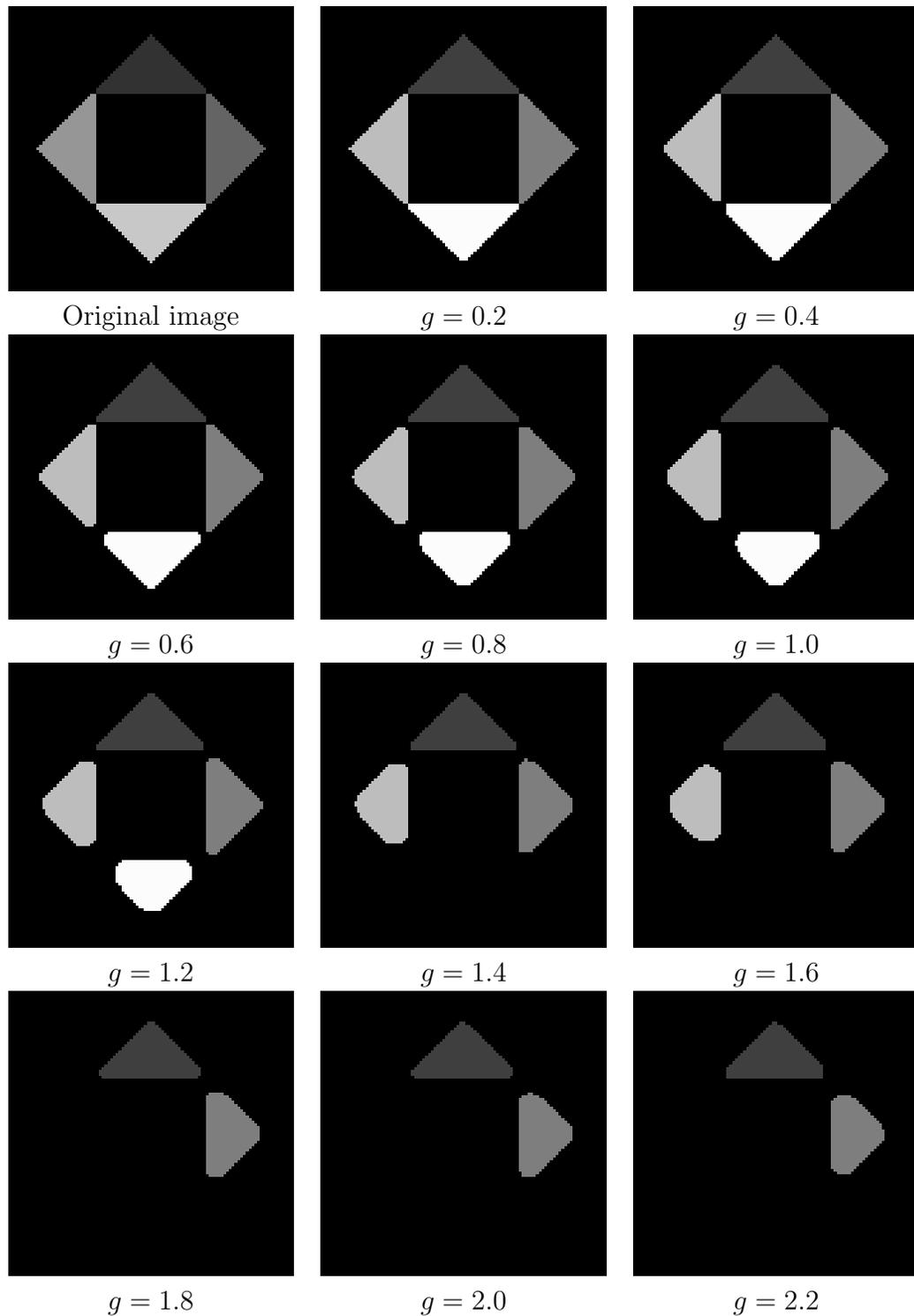
Figure 2.5: Synthetic example demonstrating the regularization bias of isotropic three-label segmentation applied to multiple objects. Four objects are to be segmented from the background. The labels are assigned as follows. Background: 0, top object: 1, right object: 2, left object: 3, bottom object: 4.
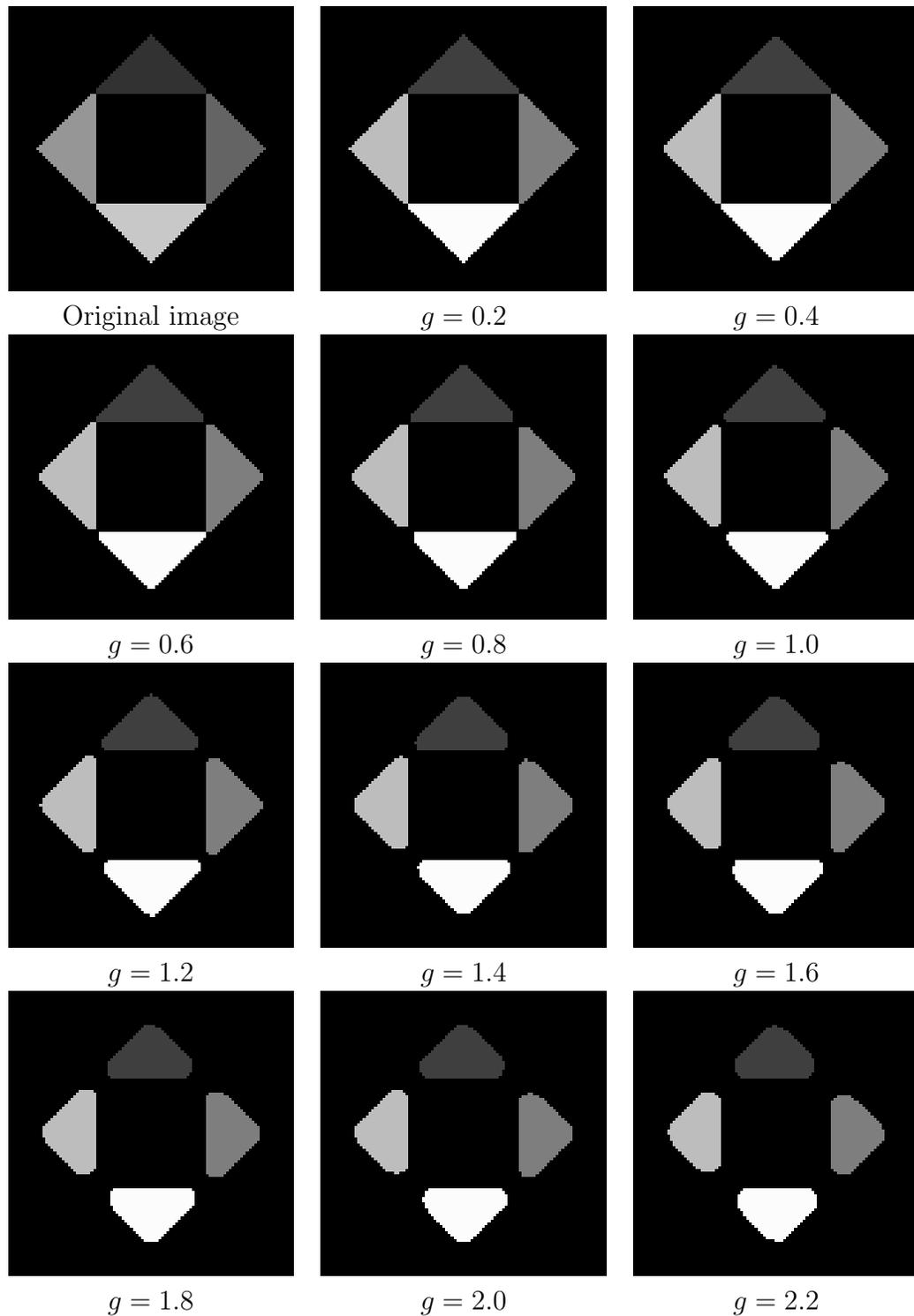
Figure 2.6: Synthetic example demonstrating no regularization bias of isotropic multi-label segmentation. Four objects are to be segmented from the background. The labels are assigned as follows. Background: 0, top object: 1, right object: 2, left object: 3, bottom object: 4.
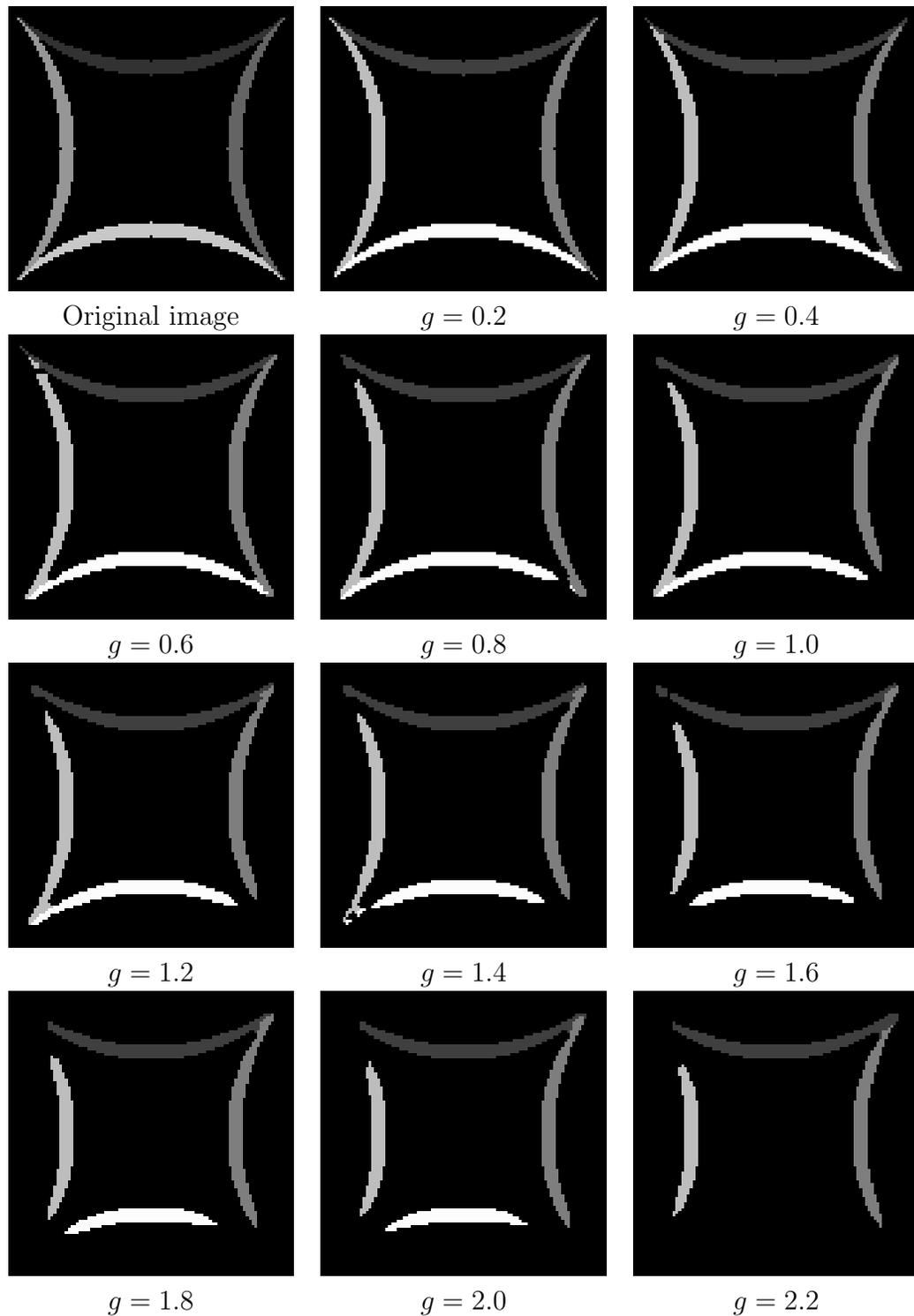
Figure 2.7: Synthetic example demonstrating the regularization bias of anisotropic three-label segmentation applied to multiple objects. Four objects are to be segmented from the background. The anisotropy parameter $\alpha$ is set to be 0.2. The labels are assigned as follows. Background: 0, top object: 1, right object: 2, left object: 3, bottom object: 4.
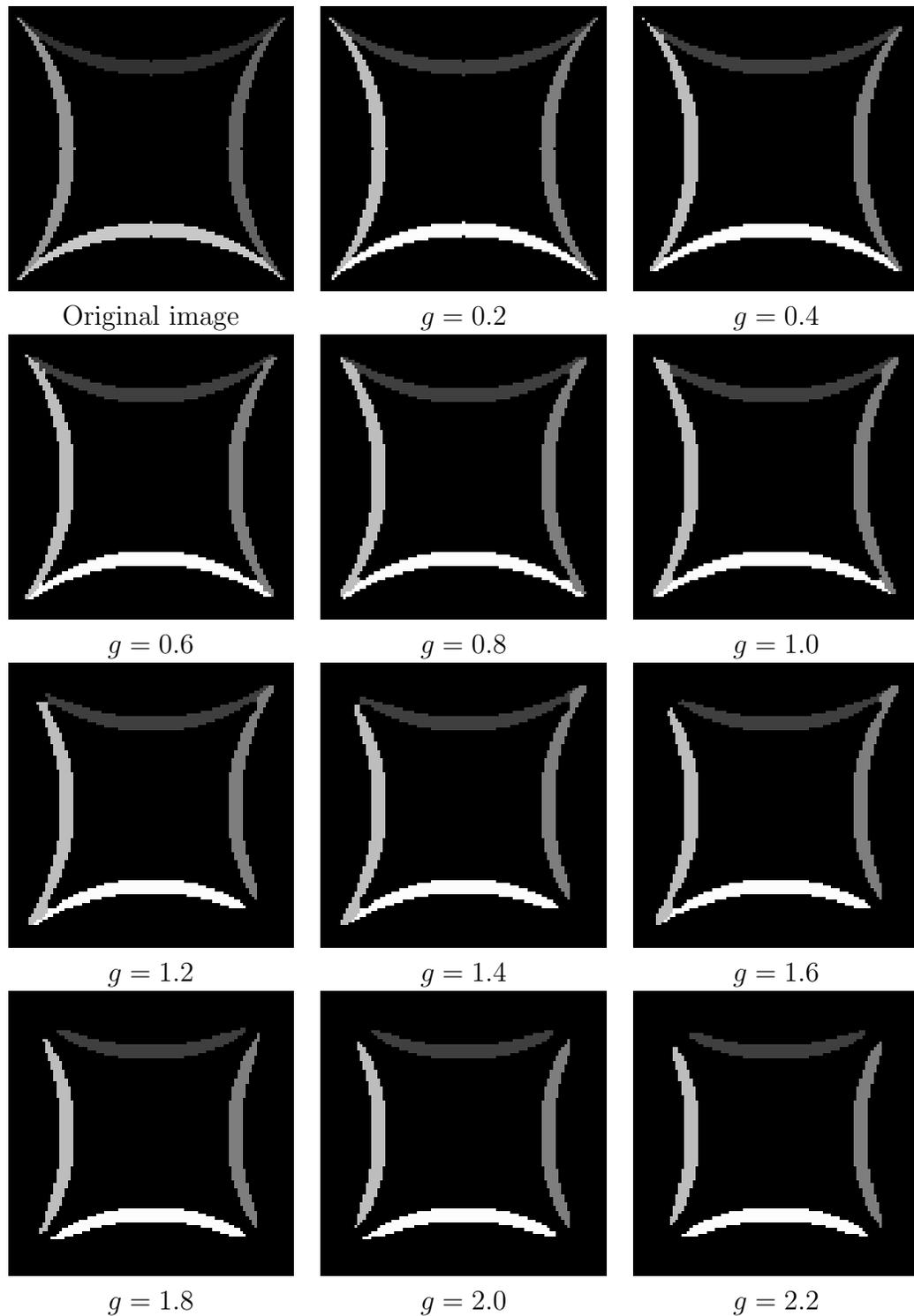
Figure 2.8: Synthetic example demonstrating no regularization bias of anisotropic multi-label segmentation. Four objects are to be segmented from the background. The anisotropy parameter $\alpha$ is set to be 0.2. The labels are assigned as follows. Background: 0, top object: 1, right object: 2, left object: 3, bottom object: 4.

1. The penalty between background and object 4 (whose $u = [0, 0, 0, 0, 0, 1]$) will result in penalty 4 as four elements are different. Therefore the spatial regularization depends on the label ordering, meaning a different permutation of labels will result in a different segmentation result. Labels that are farther away from the background will be more smoothed out.

To alleviate the label ordering issue, I propose a novel multi-label segmentation energy functional

$$E(u) = \int_{\mathcal{D}} g \, \|\nabla_{\mathbf{x}} \nabla_l u\| + c \, |\nabla_l u| \, d\mathbf{x} dl, \ \mathcal{D} = \Omega \times \mathcal{L}$$

$$\text{subject to } u \in [0, 1], \ u(\mathbf{x}, 0) = 0, \ u(\mathbf{x}, L) = 1. \tag{2.22}$$

Comparing to (2.10), the regularization term is defined by the spatial gradients of $\nabla_l u$, rather than $u$. The reason for such a change can be explained by Fig. 2.1(b) where $\nabla_l u$ contains a single "1" element for each object. Any two objects differ by 2 when comparing $\nabla_l u$. Figure 2.6 demonstrates the advantage of the proposed multi-label segmentation which has no regularization bias therefore is capable of handling any number of objects. The resulting segmentation is independent of label ordering.

The multi-label segmentation problem with anisotropic regularization is formulated as

$$E(u) = \int_{\mathcal{D}} \|\mathbf{G} \nabla_{\mathbf{x}} \nabla_l u\| + c \, |\nabla_l u| \, d\mathbf{x} dl, \ \mathcal{D} = \Omega \times \mathcal{L}$$

$$\text{subject to } u \in [0, 1], \ u(\mathbf{x}, 0) = 0, \ u(\mathbf{x}, L) = 1. \tag{2.23}$$

Again, the different comparing to (2.11) is the regularization is computed through the spatial gradient of $\nabla_l u$, rather than $u$. The matrix $\mathbf{G}$ is defined in the same way as in (2.11).
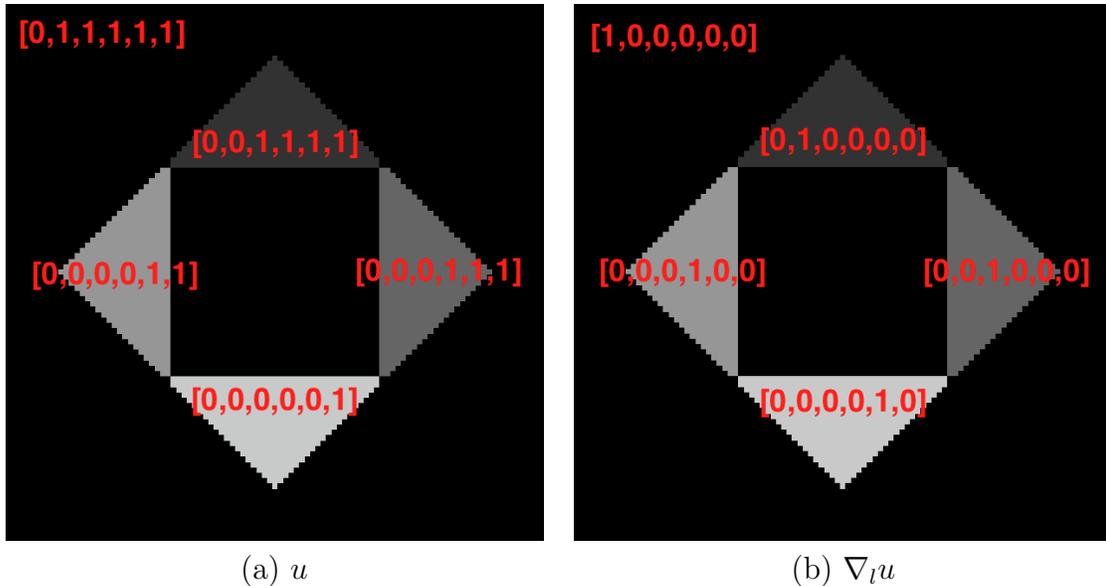
Figure 2.9: Function $u$ and $\nabla_l u$ for each object. The labels are assigned as follows. Background: 0, top object: 1, right object: 2, left object: 3, bottom object: 4.

Figure 2.5 shows the problematic regularization of (2.11) when applied to more than three objects. Energy (2.23) overcomes the issue and is therefore capable of dealing with any number of objects. The segmentation results are robust to label ordering.

The three-label energy (2.11) gives "essentially binary" solutions. It can also be proved that the thresholding of a "essentially binary" solution yields an equally globally optimal solution (Zach *et al.*, 2009). However, this property does not hold for the multi-label energy (2.23). there is no guarantee the final solution is "essentially binary" property. Energy (2.22) corresponds to the linear programming relaxation of Markov random fields with a Potts discontinuity prior. Discrete Potts models are known to be NP-hard. Therefore the optimization of (2.22) yields a globally optimal solution to the relaxed problem, rather than the original discrete problem. However, the solutions are usually quite binary in practice and can be regarded as close approximations to the true optimal solutions.

### 2.4.3   Numerical solution

This section discusses an iterative scheme to optimize (2.23). Solving (2.22) is a special case with $\mathbf{G} = g\mathbf{I}$ ($\mathbf{I}$ is the identity matrix). I introduce two dual variables $\mathbf{p}$ (vector field) and $q$ (scalar field) and rewrite (2.11) as

$$E(u, \mathbf{p}, q) = \int_{\mathcal{D}} \langle \mathbf{p}, \mathbf{G}\nabla_{\mathbf{x}}\nabla_l u \rangle + q\nabla_l u \; d\mathbf{x}dl, \tag{2.24}$$

$$\text{subject to } \|\mathbf{p}\| \leq 1, \; |q| \leq c,$$

in which $\langle \cdot, \cdot \rangle$ represents inner products. Minimizing (2.23) with respect to $u$ is equivalent to minimizing (2.24) with respect to $u$ and maximizing it with respect to $\mathbf{p}$ and $q$.

Taking the variation, we have

$$\delta E(u, \mathbf{p}, q; \delta u, \delta \mathbf{p}, \delta q) \tag{2.25}$$

$$= \frac{\partial}{\partial \epsilon} \int_{\Omega} \langle \mathbf{p} + \epsilon\delta\mathbf{p}, \mathbf{G}\nabla_{\mathbf{x}}\nabla_l u + \epsilon\mathbf{G}\nabla_{\mathbf{x}}\nabla_l \delta u \rangle + (q + \epsilon\delta q)(\nabla_l u + \epsilon\nabla_l \delta u) \; d\mathbf{x}dl \tag{2.26}$$

$$= \int_{\Omega} \langle \mathbf{p}, \mathbf{G}\nabla_{\mathbf{x}}\nabla_l \delta u \rangle + \langle \delta\mathbf{p}, \mathbf{G}\nabla_{\mathbf{x}}\nabla_l u \rangle + q\nabla_l \delta u + \delta q\nabla_l u \; d\mathbf{x}dl \tag{2.27}$$

$$= \int_{\Omega} (\nabla_l (\text{div}(\mathbf{G}\mathbf{p})) - \nabla_l q) \, \delta u + \langle \delta\mathbf{p}, \mathbf{G}\nabla_{\mathbf{x}}\nabla_l u \rangle + \delta q\nabla_l u \; d\mathbf{x}dl \tag{2.28}$$

The gradient descent/ascent update scheme of (2.24) is

$$\mathbf{p}_t = -\mathbf{G}\nabla_x\nabla_l u, \; \|\mathbf{p}\| \leq 1 \tag{2.29}$$

$$q_t = -\nabla_l u, \; |q| \leq c \tag{2.30}$$

$$u_t = \nabla_l (\text{div}(\mathbf{G}\mathbf{p})) - \nabla_l q \tag{2.31}$$

The iterative scheme will lead to a global optimum upon convergence (Appleton and Talbot, 2006) because of the convexity of (2.23). Let $S$ and $T$ denote the source and sink sets: $S = \Omega \times \{0\}$, $T = \Omega \times \{L\}$. The region without sources or sinks is denoted as

$\overset{\circ}{\mathcal{D}} = \mathcal{D} \setminus (S \cup T) = \Omega \times \{1, 2, \cdots, L-1\}$. Iterations are terminated when the gap between the primal energy (2.11) and the dual energy is sufficiently small. The dual energy is computed as

$$
\begin{aligned}
E^*(u) = \int_S \nabla_l q - \nabla_l \left( \text{div} \left( \mathbf{G} \mathbf{p} \right) \right) \ d\mathbf{x} dl \\
+ \int_{\overset{\circ}{\mathcal{D}}} \min(0, \nabla_l q - \nabla_l \left( \text{div} \left( \mathbf{G} \mathbf{p} \right) \right)) \ d\mathbf{x} dl
\end{aligned}
\tag{2.32}
$$

The multi-label segmentation can be easily recovered from the discontinuity set of $u$. Unlike the three-label segmentation, there is no guarantee that solution $u$ is essentially binary after convergence. Thresholding of $u$ does not necessarily yield a globally optimal solution to the original discrete problem.

## 2.5  Conclusion

The main contribution of this chapter is the novel three-label and multi-label segmentation framework.

The proposed three-label segmentation framework is general and guarantees separation of the touching objects, e.g., touching bones in Fig. 4.3. By placing the background label in the middle, one can obtain a symmetric labeling with respect to the background. The anisotropic regularization has general applicability for the segmentation of thin objects, e.g., knee cartilage. An important advantage of the segmentation framework is its convexity, which guarantees that a globally optimal solution to the relaxed problem can be computed, which then yields a globally optimal solution to the original problem after thresholding due to its essentially binary property.

The multi-label segmentation framework is proposed to overcome the regularization bias of the three-label segmentation when applied to more than three labels (two objects, one label is for the background). Again, the relaxed formulation is convex, and one can

compute the globally optimal solution. Unfortunately, the "essentially binary" property is not guaranteed in this case. Therefore, the thresholded solution may not be a globally optimal solution to the discrete formulation, which is a Potts model and known to be NP-hard. However, the solution can be a close approximation to the true optimal solution as it converges to solutions which are close to binary in practice.

In the next chapter, I will apply the proposed three-label segmentation to segment human knee cartilage, i.e., femoral and tibial cartilage. The background label is placed between the labels for femoral and tibial cartilage for symmetry.

# CHAPTER 3: AUTOMATIC MULTI-ATLAS-BASED THREE-LABEL CARTILAGE SEGMENTATION

## 3.1 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 (Woolf and Pfleger, 2003). 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 (Eckstein *et al.*, 1998, 2006) 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-progression modifying 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 (Eckstein *et al.*, 2008) 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 chapter, I 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.* (2007) proposed a voxel-based hierarchical classification scheme for cartilage segmentation. Fripp *et al.* (2010) 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.* (2010b). Vincent *et al.* (2010) applied multi-start and hierarchical active appearance modeling to segment cartilage. Texture analysis (Dodin *et al.*, 2010) has also been employed in cartilage segmentation. Seim *et al.* (2010) utilized prior knowledge on the variation of cartilage thickness. Voxel-based classification approaches have been investigated for segmenting multi-contrast MR data by Koo *et al.* (2009) and Zhang and Lu (2011).

To allow for localized analysis and the suppression of unlikely 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 (Rohlfing *et al.*, 2004) 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.* (2007), which used a statistical shape atlas from a set of pre-aligned knee images, and the work by Tamez-Peña *et al.* (2012) using a multi-atlas-based method, are two exceptions. The proposed segmentation method is most closely related to Tamez-Peña *et al.* (2012) as both methods make use of multi-atlas segmentation strategies. However, I significantly extend the prior work by Tamez-Peña *et al.* (2012). In particular:

1) I propose a convex three-label segmentation method which allows for anisotropic spatial regularization (described in detail in chapter 2). 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. The method incorporates spatial priors via atlas information (see item 2) and local segmentation label likelihoods through appearance classification comparing both $k$

nearest neighbors ($k$NN) classification and classification by a support vector machine (SVM).

2) I 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) In chapter 4, I 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 (Heimann *et al.*, 2010) 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, i.e. femoral/tibial cartilage and background. Furthermore, I demonstrate that anisotropic regularization within this segmentation model is less sensitive to parameter settings than isotropic regularization and yields more accurate cartilage segmentations.

2) I 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) The 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 the proposed segmentation method to automatically achieve accurate cartilage segmentations for large imaging studies. The required robustness of the segmentation method is achieved 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.
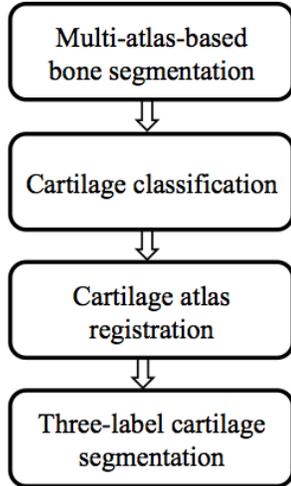
Figure 3.1: Cartilage segmentation pipeline.

Figure 3.1 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 remainder of this chapter is organized as follows: Section 3.2 clarifies the atlas terminology and briefly discuss atlas-based segmentation methods. Section 3.3 describes the multi-atlas-based bone segmentation method. The probabilistic cartilage classification is explained in section 3.4. Sections 3.5 and 3.6 discuss the average-shape-atlas-based and multi-atlas-based cartilage segmentation, respectively. This chapter closes with conclusions and future work. Chapter 4 shows the experimental results on the PLS dataset and compares the proposed method to other methods by making use of the SKI10 dataset.

## 3.2 Atlas-based segmentation

An atlas (Aljabar *et al.*, 2009), 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 (Išgum *et al.*, 2009), namely single-atlas-based, average-shape atlas-based, and multi-atlas-based methods. The work by Glocker *et al.* (2007) falls into the second group. The work by Tamez-Peña *et al.* (2012) 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.* (2004) demonstrated that multi-atlas-based segmentation is more accurate than the other two atlas-based segmentation methods. I will therefore follow a multi-atlas strategy in what follows.

## 3.3 Multi-atlas-based bone segmentation

The labeling cost $c$ in (2.10) 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 $\mathbf{x}$:

$$c(\mathbf{x}, l) = -log(P(l|I(\mathbf{x}))) = -log\left(\frac{p(I(\mathbf{x})|l) \cdot P(l)}{p(I(\mathbf{x}))}\right). \tag{3.1}$$

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, I assume a simple model (3.2) to estimate bone likelihoods,

$$p(I(\mathbf{x})|\text{FB}) = p(I(\mathbf{x})|\text{TB}) = exp(-\beta I(\mathbf{x})), \tag{3.2}$$

where $\beta$ 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 (3.1), I employ a multi-atlas registration approach followed by label fusion. Suppose there are $N$ atlases $A_i$ and their bone segmentations are $S_i^{\text{FB}}$ and $S_i^{\text{TB}}$ ($i = 1, 2, ..., N$). Registration from an atlas $A_i$ to a query image $I$ is an affine registration $T_i^{\text{affine}}$ followed by a B-Spline registration $T_i^{\text{bspline}}$. Averaging all $N$ propagated atlas labels yields a spatial prior of femur and tibia for the query image:

$$\begin{aligned} p(\text{FB}) &= \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{bspline}} \circ T_i^{\text{affine}} \circ S_i^{\text{FB}}\right), \\ p(\text{TB}) &= \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{bspline}} \circ T_i^{\text{affine}} \circ S_i^{\text{TB}}\right). \end{aligned} \tag{3.3}$$

Now that I have computed the spatial priors and the local likelihoods, I integrate them into (3.1) and solve (2.10) to obtain the three-label bone segmentation. The bone segmentation will help locate the cartilage in atlas-based cartilage segmentation.

## 3.4   Probabilistic classification

I 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:

$$c(\mathbf{x}, l) = -log(P(l|\mathbf{f}(\mathbf{x}))) = -log\left(\frac{p(\mathbf{f}(\mathbf{x})|l) \cdot p(l)}{p(\mathbf{f}(\mathbf{x}))}\right), \tag{3.4}$$

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.

I compute the spatial prior $p(l)$ in two different ways: using an average-shape-atlas registration and a multi-atlas registration (see sections 3.5 and 3.6). I compare the performance of both approaches in chapter 4. The local likelihood term $p(\mathbf{f}(\mathbf{x})|l)$ is obtained from a probabilistic classification based on local image appearance. I investigate classification based on a probabilistic $k$ nearest neighbors ($k$NN) (Duda *et al.*, 2001) as well as by a support vector machine (SVM) (Cortes and Vapnik, 1995). For classification I use a reduced set of features compared to Folkesson *et al.* (2007): intensities at three scales, first-order derivatives in three directions at three scales, and second-order derivatives in the axial direction at 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 Folkesson *et al.* (2007) and Koo *et al.* (2009) is the probabilistic nature of the classification, which allows an easy incorporation of the classification result into the Bayesian framework. Further, the final segmentations are generated by a segmentation method with anisotropic regularization, whereas no regularization was used in Folkesson *et al.* (2007) nor Koo *et al.* (2009). I demonstrate in chapter 4 that spatial regularization helps improve the segmentation accuracy and anisotropic regularization yields better accuracy than isotropic regularization.

### 3.4.1 Classification using $k$NN

I estimate the data likelihoods for femoral and tibial cartilage, $p(\mathbf{f}(\mathbf{x})|l)$, of (3.4) by probabilistic $k$NN classification (Duda *et al.*, 2001). I use a one-versus-other classification strategy and the expert segmentations of femoral and tibial cartilage to build the $k$NN 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 $k$NN classifier given a query voxel $\mathbf{x}$ with its feature vector $\mathbf{f}(\mathbf{x})$ are

$$
\begin{aligned}
p(\mathbf{f}(\mathbf{x})|\text{FC}) &= \frac{n^{\text{FC}}(\mathbf{f}(\mathbf{x}))}{k}, \\
p(\mathbf{f}(\mathbf{x})|\text{BG}) &= \frac{n^{\text{BG}}(\mathbf{f}(\mathbf{x}))}{k}, \\
p(\mathbf{f}(\mathbf{x})|\text{TC}) &= \frac{n^{\text{TC}}(\mathbf{f}(\mathbf{x}))}{k}.
\end{aligned}
\tag{3.5}
$$

Here $n^{\text{FC}}$, $n^{\text{TC}}$, $n^{\text{BG}}$ 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 $k$NN is sensitive to the number of training samples, I scale the outputs according to the

training class sizes to balance the three classes.

### 3.4.2 Classification using an SVM

An alternative approach to compute the local likelihoods is to use a support vector machine (SVM) (Cortes and Vapnik, 1995), which constructs a hyperplane maximally separating classes given a labeled training set. Koo *et al.* (2009) used two-class SVM to segment cartilage automatically from multi-contrast MR images. I apply LIBSVM (Chang and Lin, 2011) 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 $p(\mathbf{f}(\mathbf{x})|\text{TC})$. I compare the SVM and the $k$NN probabilistic classification methods in chapter 4.

### 3.5 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 there are $N$ images with expert segmentations. One can pick the segmentation of one image as the reference to bring all the segmentations to the same position. Specifically, I register the femur segmentation $S_i^{\text{FB}}$ ($i = 1, 2, ..., N$) and the tibial segmentation $S_i^{\text{TB}}$ ($i = 1, 2, ..., N$) to the reference femur and tibia segmentations with affine transforms $T_i^{\text{FB}}$ and $T_i^{\text{TB}}$ respectively. The femoral and tibial cartilage segmentations $S_i^{\text{FC}}$ and $S_i^{\text{FC}}$ are propagated accordingly. The average bone and cartilage atlas $A_{avg}$

(including $A_{avg}^{\mathrm{FB}}$, $A_{avg}^{\mathrm{TB}}$, $A_{avg}^{\mathrm{FC}}$ and $A_{avg}^{\mathrm{TC}}$) is computed by

$$
\begin{aligned}
A_{avg}^{\mathrm{FB}} &= \frac{1}{N} \sum_{i=1}^{N} \left( T_i^{\mathrm{FB}} \circ S_i^{\mathrm{FB}} \right), \\
A_{avg}^{\mathrm{TB}} &= \frac{1}{N} \sum_{i=1}^{N} \left( T_i^{\mathrm{TB}} \circ S_i^{\mathrm{TB}} \right), \\
A_{avg}^{\mathrm{FC}} &= \frac{1}{N} \sum_{i=1}^{N} \left( T_i^{\mathrm{FB}} \circ S_i^{\mathrm{FC}} \right), \\
A_{avg}^{\mathrm{TC}} &= \frac{1}{N} \sum_{i=1}^{N} \left( T_i^{\mathrm{TB}} \circ S_i^{\mathrm{TC}} \right).
\end{aligned}
\tag{3.6}
$$

Given a query image $I$, I have computed the bone segmentation $S^{\mathrm{FB}}$ and $S^{\mathrm{TB}}$ from section 3.3. The atlas femur $A^{\mathrm{FB}}$ and tibia $A^{\mathrm{FB}}$ are registered to the segmentation of femur $S^{\mathrm{FB}}$ and tibia $S^{\mathrm{TB}}$ with affine transforms $T^{\mathrm{FB}}$ and $T^{\mathrm{TB}}$. The spatial prior for each cartilage is then computed by propagating each cartilage atlas with the corresponding transform,

$$
\begin{aligned}
p(\mathrm{FC}) &= T^{\mathrm{FB}} \circ A_{avg}^{\mathrm{FC}}, \\
p(\mathrm{TC}) &= T^{\mathrm{TB}} \circ A_{avg}^{\mathrm{TC}}, \\
p(\mathrm{BG}) &= 1 - p(\mathrm{FC}) - p(\mathrm{TC}).
\end{aligned}
\tag{3.7}
$$

These spatial priors and the local likelihoods from section 3.4 are integrated into (3.4) and the cartilage segmentation is obtained by optimizing the three-label segmentation energy with anisotropic regularization (2.11).

## 3.6 Multi-atlas-based cartilage segmentation

This section presents an alternative approach to computing the spatial prior for cartilage. I make use of multi-atlas registration, rather than average-shape-atlas registration

described in section 3.5. 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.

Suppose there are $N$ atlases $A_i$ including their femur segmentations $S_i^{\mathrm{FB}}$, tibia segmentations $S_i^{\mathrm{TB}}$, femoral cartilage segmentations $S_i^{\mathrm{FC}}$ and tibial cartilage segmentations $S_i^{\mathrm{TC}}$ ($i = 1, 2, ..., N$). For a query image $I$, we have the bone segmentation $S^{\mathrm{FB}}$ and $S^{\mathrm{TB}}$ from section 3.3.

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

One can simply take the average of the registered cartilage atlas segmentations to compute the spatial priors, which is majority voting Rohlfing *et al.* (2004) label fusion:

$$
\begin{aligned}
p(\mathrm{FC}) &= \frac{1}{N} \sum_{i=1}^{N} \left( T_i^{\mathrm{FB}} \circ S_i^{\mathrm{FC}} \right), \\
p(\mathrm{TC}) &= \frac{1}{N} \sum_{i=1}^{N} \left( T_i^{\mathrm{TB}} \circ S_i^{\mathrm{TC}} \right).
\end{aligned}
\tag{3.8}
$$

One can also apply a locally-weighted label fusion strategy (Išgum *et al.*, 2009), which was shown to yield a better segmentation accuracy than a majority voting strategy. In this case, I choose to favor the atlases which locally agree better with the cartilage likelihoods $p(\mathbf{f}(\mathbf{x})|\mathrm{FC})$ and $p(\mathbf{f}(\mathbf{x})|\mathrm{TC})$ from the probabilistic classification in section 3.4. The spatially varying weighting functions $\lambda_i^{\mathrm{FC}}$ for the femoral cartilage and $\lambda_i^{\mathrm{TC}}$ for the tibial cartilage are calculated as

$$
\begin{aligned}
\lambda_i^{\mathrm{FC}}(\mathbf{x}) &= \frac{1}{\alpha \left| T_i^{\mathrm{FB}} \circ S_i^{\mathrm{FC}} - p(\mathbf{f}(\mathbf{x})|\mathrm{FC}) \right| + \epsilon}, \\
\lambda_i^{\mathrm{TC}}(\mathbf{x}) &= \frac{1}{\alpha \left| T_i^{\mathrm{TB}} \circ S_i^{\mathrm{TC}} - p(\mathbf{f}(\mathbf{x})|\mathrm{TC}) \right| + \epsilon},
\end{aligned}
\tag{3.9}
$$

followed by a small amount of diffusion smoothing. I choose $\alpha = 0.2$ and $\epsilon = 0.001$ in

my implementation. The spatial prior for each cartilage is then the weighted average of the propagated atlas cartilage segmentations

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

$$p(\text{TC}) = \sum_{i=1}^{N} \frac{\lambda_i^{\text{TC}}}{\sum_{i=1}^{N} \lambda_i^{\text{TC}}} \left( T_i^{\text{TB}} \circ S_i^{\text{TC}} \right). \tag{3.10}$$

Recently, non-local patch-based label fusion techniques have been proposed by Coupé *et al.* (2011) and Rousseau *et al.* (2011). 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 registration 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 $\mathbf{x}$. I 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 $N$, $S_i^{\text{FC}}$ refers to the femoral cartilage segmentation of the $i$-th atlas, and $I_i$ is the $i$-th atlas appearance. For the femoral cartilage, the probability is computed as

$$p^{\text{FC}}(\mathbf{x}) = \frac{\sum\limits_{i=1}^{N} \sum\limits_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w^{\text{FC}}(\mathbf{x},\mathbf{y}) \tilde{S}_i^{\text{FC}}(\mathbf{y})}{\sum\limits_{i=1}^{N} \sum\limits_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w^{\text{FC}}(\mathbf{x},\mathbf{y})}, \tag{3.11}$$

$$w^{\text{FC}}(\mathbf{x},\mathbf{y}) = \exp\left( \frac{\sum\limits_{\mathbf{x}' \in \mathcal{P}(\mathbf{x})} \sum\limits_{\mathbf{y}' \in \mathcal{P}(\mathbf{y})} \left( I(\mathbf{x}') - \tilde{I}_i^{\text{FC}}(\mathbf{y}') \right)^2}{h^{\text{FC}}(\mathbf{x})} \right), \tag{3.12}$$

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

$$h^{\text{FC}}(\mathbf{x}) = \min_{\substack{1 \le i \le N \\ \mathbf{y} \in \mathcal{N}(\mathbf{x})}} \sum_{\substack{\mathbf{x}' \in \mathcal{P}(\mathbf{x}) \\ \mathbf{y}' \in \mathcal{P}(\mathbf{y})}} \left( I(\mathbf{x}') - \tilde{I}_i^{\text{FC}}(\mathbf{y}') \right)^2 + \epsilon. \tag{3.13}$$

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 chapter 4. 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 3.4 are integrated into (3.4) and the cartilage segmentation is obtained by optimizing the three-label segmentation energy with anisotropic regularization (2.11).

## 3.7 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, I first correct the MRI bias field (Sled *et al.*, 1998), scale image intensities to a common range, and then perform edge-preserving smoothing using curvature flow (Sethian, 1999).

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. I 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 (3.2). The priors and the likelihoods are combined in (3.1) and then integrated in the three-label segmentation (2.10), the global optimal solution of which produces the bone segmentation.

Once I have the bone segmentation, I perform the probabilistic classification ($k$NN 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 (2.12) is computed by taking the gradient of the diffusion smoothed three-label bone segmentation result within the joint area. Finally, the local likelihoods and the spatial priors are integrated into the three-label segmentation to generate the cartilage segmentation.

## 3.8  Conclusion

The major contribution of this chapter is the fully-automatic cartilage segmentation method that incorporates local classification (from image appearance) and shape information (from atlas registration) into the three-label segmentation framework discussed in chapter 2.

I used a multi-atlas-based bone segmentation to guide the registration of a cartilage atlas. I obtained cartilage segmentation using an average shape atlas or multiple atlases with various label fusion techniques to obtain spatial cartilage priors within the three-label segmentation framework, which incorporates anisotropic regularization to improve segmentation performance (shown in chapter 4) for the thin femoral and tibial cartilage layers.

The overall pipeline is fully-automatic (besides quality control), which enables the method to be applied to large image databases. The robustness due to multi-atlas-based strategies also makes the proposed method appropriate for large datasets.

The next chapter presents the validation result on two different large datasets, i.e.,

the PLS dataset (Eckstein *et al.*, 2008) and SKI10 dataset (Heimann *et al.*, 2010), on which I compare the proposed method to existing ones quantitatively.

The major drawback of the proposed 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 (Aljabar *et al.*, 2009). Such a selection strategy can be integrated into the segmentation pipeline and is expected to further improve segmentation performance. I will explore atlas selection for cartilage segmentation in future work.

# CHAPTER 4: VALIDATION OF CARTILAGE SEGMENTATION

## 4.1  Introduction

This chapter presents the validations results of the proposed cartilage segmentation method in chapter 3 on two datasets, PLS (Eckstein *et al.*, 2008) and SKI10 (Heimann *et al.*, 2010).

The main validation is performed on the PLS dataset. I compare different atlas choices for cartilage segmentation, namely average-shape-atlas and multi-atlas with various label fusion strategies (majority voting, locally-weighted voting and patch-based label fusion). I also compare the two probabilistic tissue classification, i.e., $k$NN and SVM. The impact of isotropic and anisotropic regularization is also studied.

To compare to other existing methods quantitatively, I test the proposed method on a publicly available dataset, SKI10.

## 4.2  Data description

The main dataset is the PLS dataset, containing 706 T1-weighted (3D oblique coronal spoiled gradient recalled) 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) (Kellgren and Lawrence, 1957) 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 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, one should expect partial femoral cartilage

Table 4.1: 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     |

segmentations and full tibial cartilage segmentations. Expert bone segmentations are available for the baselines of 18 subjects.

## 4.3 Bone validation

I validate the 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 with respect to the expert segmentations using the Dice similarity coefficient (DSC) (Dice, 1945) defined as

$$\text{DSC} = \frac{2\,|S \cap R|}{|S| + |R|}, \tag{4.1}$$

where $S$ and $R$ represent two segmentations. Table 4.1 and Fig. 4.1 show the validation results of the bone segmentation with and without regularization (corresponding to $g > 0$ and $g = 0$ in model (2.10) 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.

As can be seen from Fig. 4.2, 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. I use the bone segmentation with spatial regularization $g = 0.5$ to compute the cartilage segmentations for the remaining experiments in this section.

(a) Femur　　　　　　　　　(b) Tibia
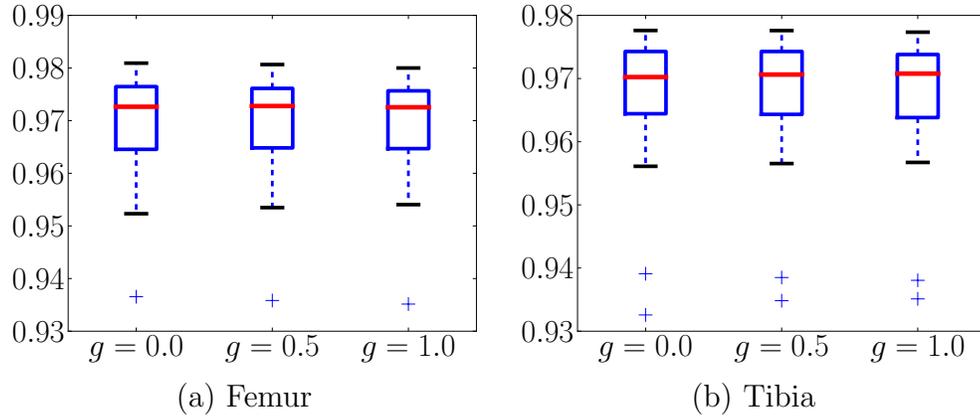
Figure 4.1: Box plots of DSC for femur and tibia with different amount of regularization on 18 test images. The center red line is the median and the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.



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

(c) Segmentation result　　　　　(d) Expert segmentation
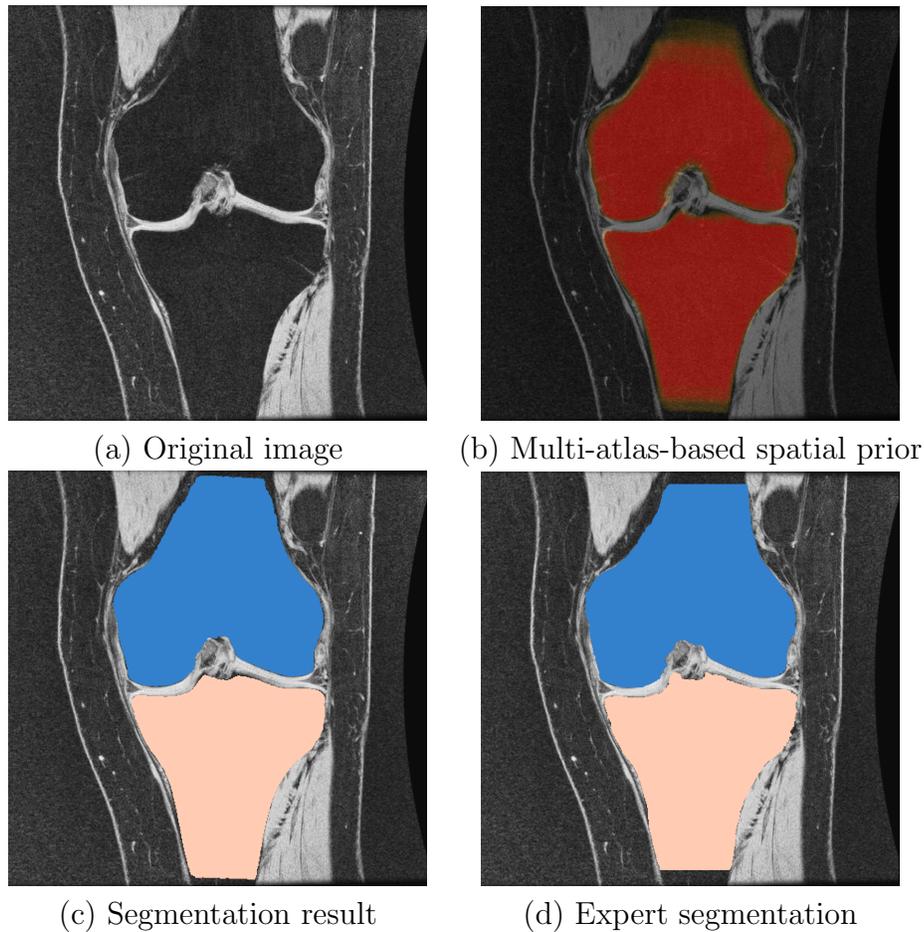
Figure 4.2: Bone segmentation of one example slice in coronal view.

(a) Binary segmentation    (b) Three-label segmentation    (c) Expert segmentation
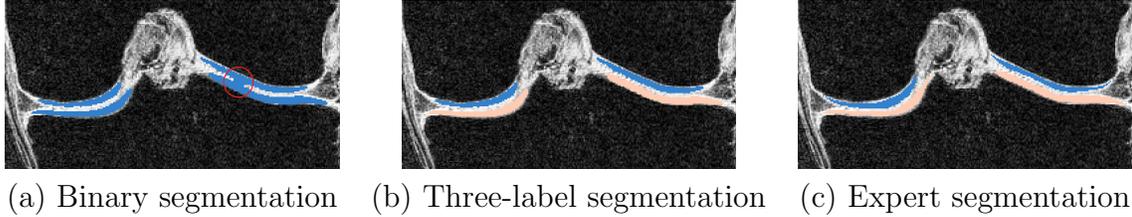
Figure 4.3: Example comparing binary and three-label segmentation methods. (a) is the binary segmentation result. (b) is the three-label segmentation result in which femoral and tibial cartilage have distinct labels. (c) is the expert segmentation. In (a), as the red circle indicates, the lateral (right) 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.

## 4.4 Cartilage validation

Figure 4.3 illustrates the beneficial behavior of the 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.

I build an average shape atlas of bone and cartilage from the expert bone and cartilage segmentations of the 18 images. Figure 4.4 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.

In the average-shape-atlas-based cartilage segmentation, I use the atlas built from 18 images (each from a different subject) to segment cartilage of the remaining 137 subjects. Within the 18 subjects, I 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: I 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 $k$NN

(a)                   (b)
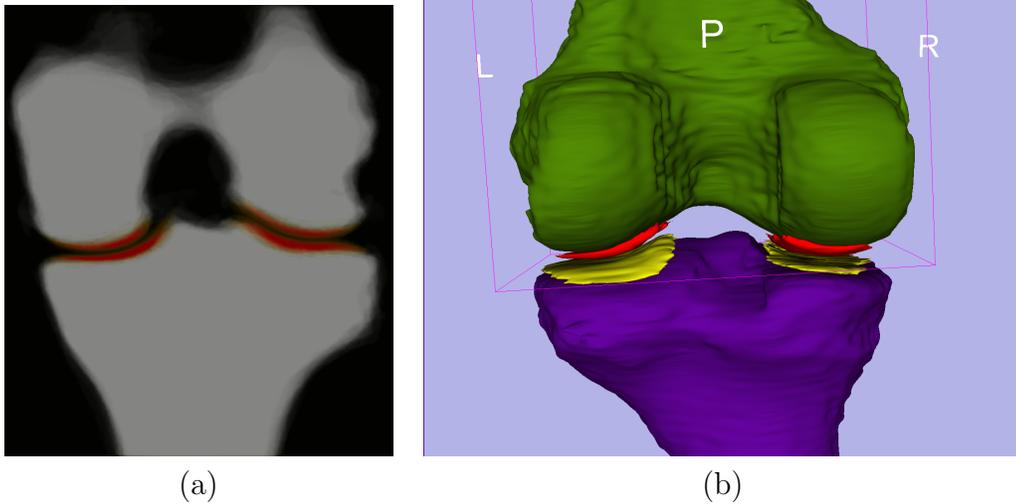
Figure 4.4: Average atlas built from 18 images. (a) is a slice of the probabilistic atlas of femoral and tibial bone and cartilage (red) overlaid on the bone in coronal view. Saturated red denotes high probability. (b) is a 3-dimensional rendering of the thresholded atlas of femur (green), tibia (purple), femoral (red) and tibial cartilage (yellow).

and SVM are chosen in the same way.

In the non-local patch-based label fusion, I upsample the images to approximately isotropic resolution and search for similar $5\times5\times5$ patches within a $9\times9\times9$ neighborhood.

Figures 4.6 and 4.7 compare the two local classification methods, i.e., $k$NN 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, $k$NN and SVM generate similar mean DSC. The SVM improves the mean DSC by a considerable amount over the $k$NN 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 $k$NN. SVM performs better than $k$NN for tibial cartilage which is segmented in its entirety.

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

(c) Segmentation result      (d) Expert segmentation
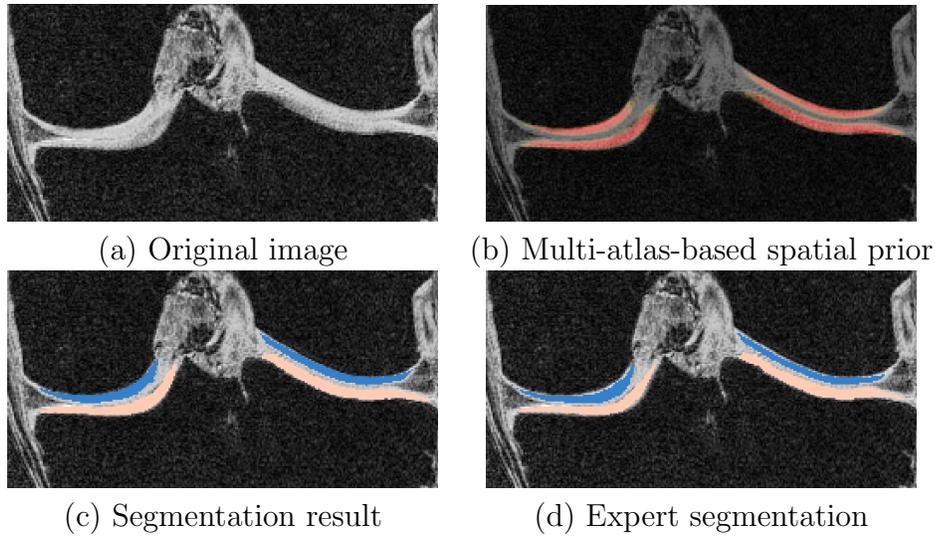
Figure 4.5: Cartilage segmentation of one example slice in coronal view. Only joint region is shown.



(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
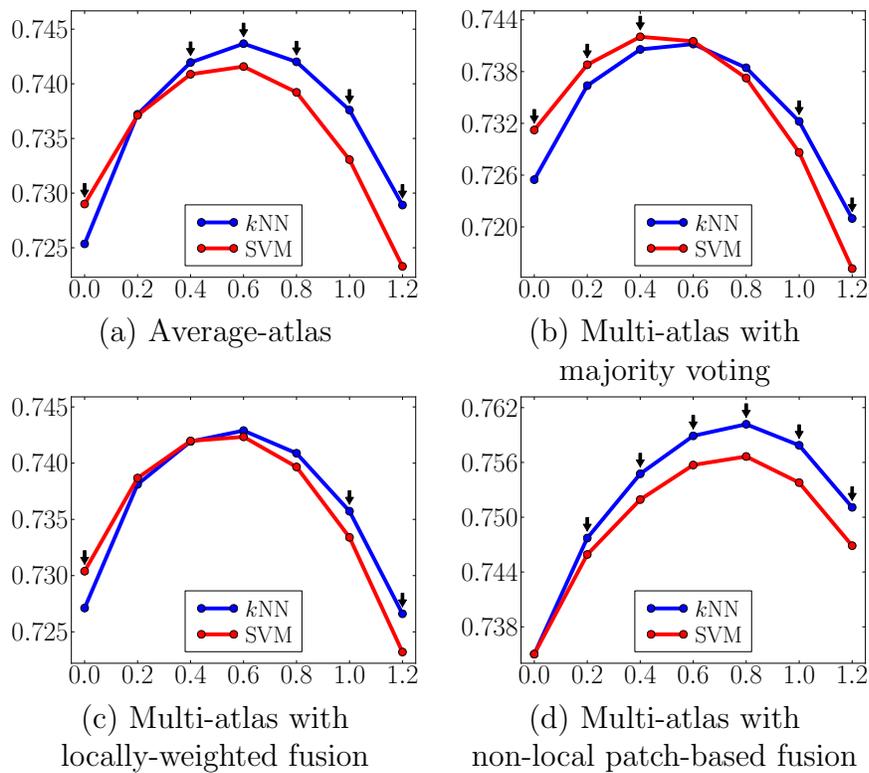
Figure 4.6: Comparison of $k$NN and SVM based on the mean DSC (ordinate) with varying amount of isotropic regularization (abscissa $g$) under different atlas choices for the femoral cartilage. The black downarrows ($\Downarrow$) indicate statistically significant differences between the two methods at 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
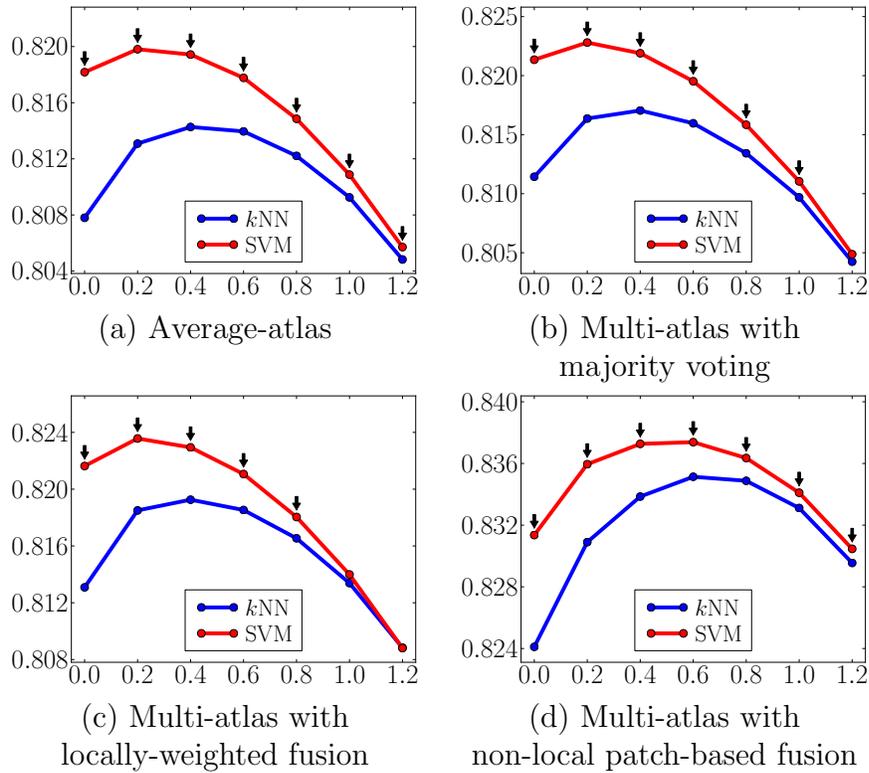
Figure 4.7: Comparison of $k$NN and SVM based on the mean DSC (ordinate) with varying amount of isotropic regularization (abscissa $g$) under different atlas choices for the tibial cartilage. The black downarrows ($\Downarrow$) indicate statistically significant superiority of SVM to $k$NN at corresponding spatial regularization settings via paired t-tests at a significance level of 0.05.

(a) Femoral cartilage $k$NN  (b) Femoral cartilage SVM

(a) Tibial cartilage $k$NN  (b) Tibial cartilage SVM
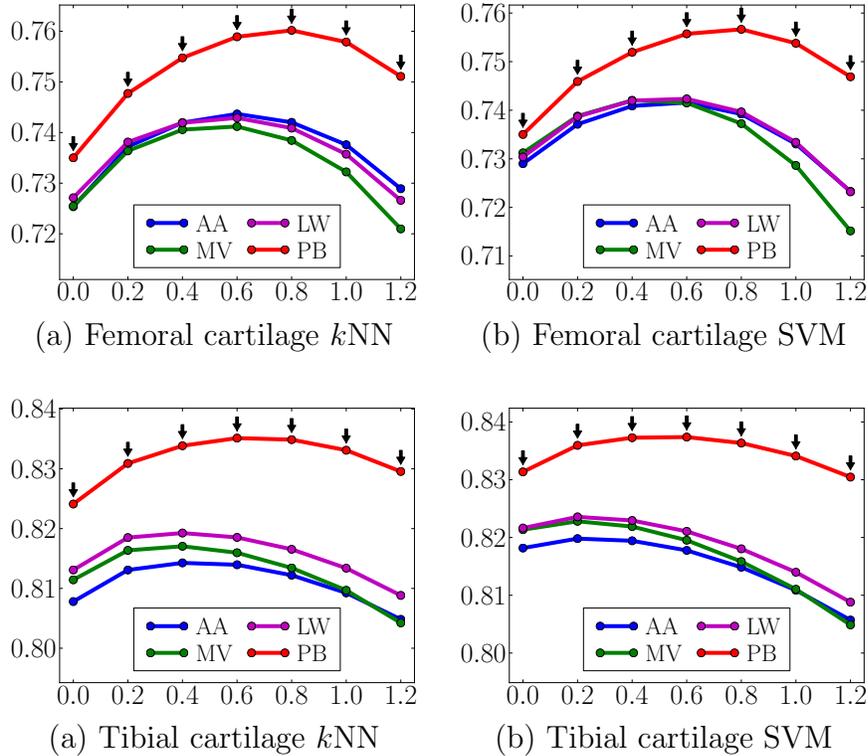
Figure 4.8: Comparisons of mean DSC (ordinate) from different atlas choices for different amount of **isotropic** regularization (abscissa $g$). AA average-shape-atlas. MV multi-atlas with majority voting. LW multi-atlas with locally-weighted fusion. PB multi-atlas with non-local patch-based fusion. 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.

Figure 4.8 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 4.9 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 4.10 shows the advantage of anisotropic regularization. The isotropic regularization has a tendency to cut long and thin objects short as shown in Fig. 4.10 (a) at

(a) Femoral cartilage $k$NN     (b) Femoral cartilage SVM

(a) Tibial cartilage $k$NN     (b) Tibial cartilage SVM
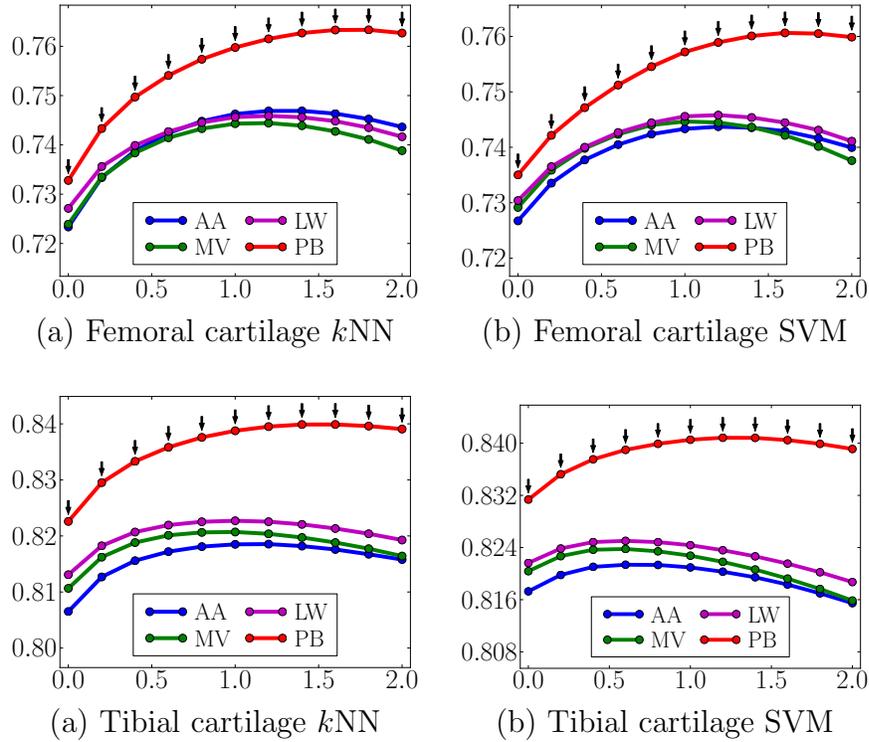
Figure 4.9: Comparisons of mean DSC (ordinate) from different atlas choices for different amount of **anisotropic** regularization (abscissa $g$). The parameter $\alpha$ controlling the anisotropy is set to be 0.2. AA average-shape-atlas. MV multi-atlas with majority voting. LW multi-atlas with locally-weighted 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) Isotropic        (b) Anisotropic        (c) Expert segmentation
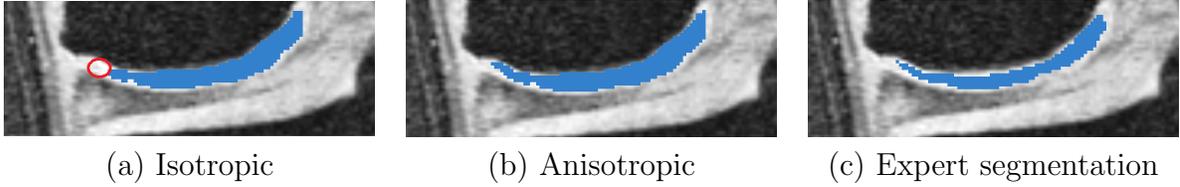
Figure 4.10: Improvement by anisotropic regularization. (a) uses isotropic regularization and misses circled region. (b) uses anisotropic regularization and captures the missing region in (a). (c) is the expert segmentation.



(a) Femoral cartilage        (b) Tibial cartilage
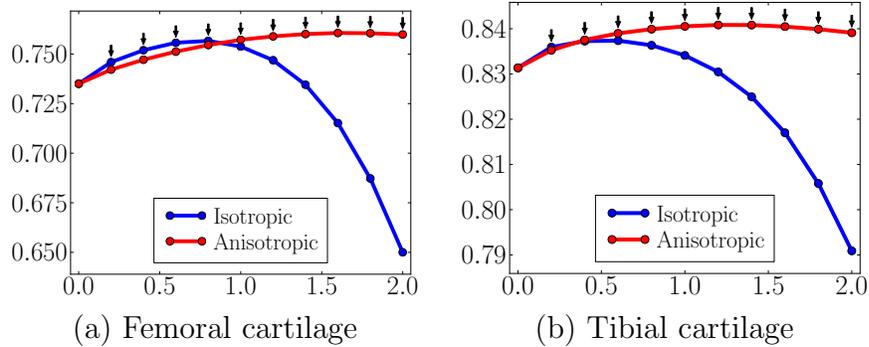
Figure 4.11: Change of mean DSC for femoral and tibial cartilage with isotropic and anisotropic regularization over the amount of regularization $g$ (abscissa). The parameter $\alpha$ is set to be 0.2 for all anisotropic tests. All tests use SVM and non-local patch-based label fusion. The black downarrows ($\Downarrow$) indicate statistically significant differences between the two methods at corresponding spatial regularization settings via paired t-tests at a significance level of 0.05.

the medial femoral cartilage. Anisotropic regularization, on the other hand, avoids this problem (see Fig. 4.10 (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 illustrated in Fig. 4.11 (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. 2.4). The faster drop-off in the isotropic case indicates a stronger dependency on the parameter settings for isotropic regularization.

To further illustrate segmentation behavior, I show the box plots of the DSC for
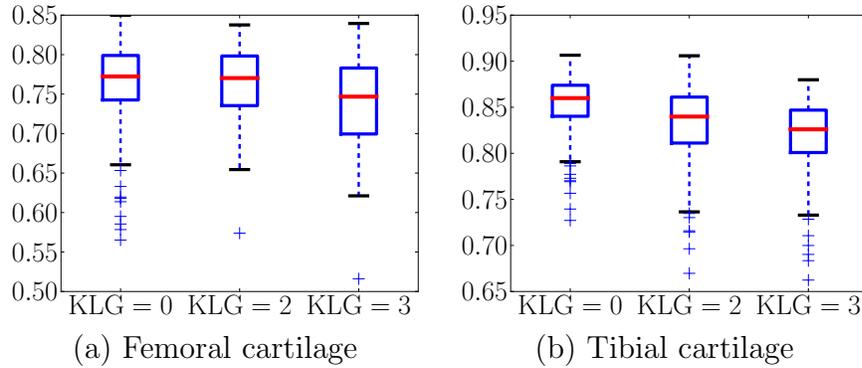
Figure 4.12: Boxplots of DSC for different KLG. We choose the best strategy combination, SVM and non-local patch-based label fusion with an anisotropic regularization with $g = 1.4$ and $\alpha = 0.2$.

different progression levels, i.e., KLG for femoral and tibial cartilage in Fig. 4.12. As expected, a slight deterioration is observed in segmentation accuracy for larger KLG as it is more challenging to segment pathological knee cartilage.

Figure 4.13 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 (Yezzi and Prince, 2003). I compute the correlation coefficient of local thickness maps from the expert and the proposed segmentations for each image. Figure 4.14 shows box plots of Pearson's correlation coefficients for different KLG. Thicknesses of the automatic and the expert segmentations are strongly correlated. 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.

Table 4.2: 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. FC: femoral cartilage, TC: tibial cartilage.

| | DSC | | | Sensitivity | | | Specificity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | STD | Mean | Median | STD | Mean | Median | STD |
| FC | 76.0% | 76.8% | 4.8% | 85.3% | 86.7% | 7.2% | 99.8% | 99.8% | 0.05% |
| TC | 84.1% | 84.7% | 3.7% | 89.4% | 89.9% | 4.1% | 99.8% | 99.8% | 0.07% |



(a) Femoral cartilage  (b) Tibial cartilage
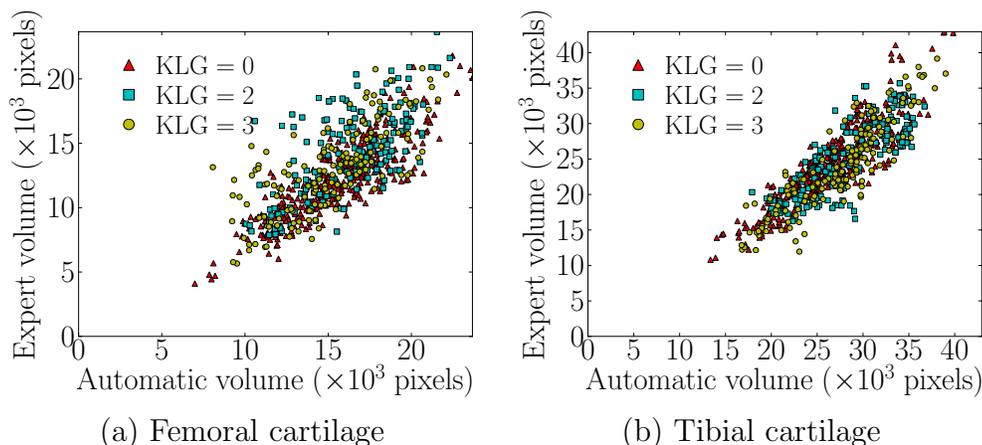
Figure 4.13: Scatter plots of segmentation volumes (number of pixels). The best strategy combination is used, SVM and non-local patch-based label fusion with an anisotropic regularization with $g = 1.4$ and $\alpha = 0.2$.



(a) Femoral cartilage  (b) Tibial cartilage
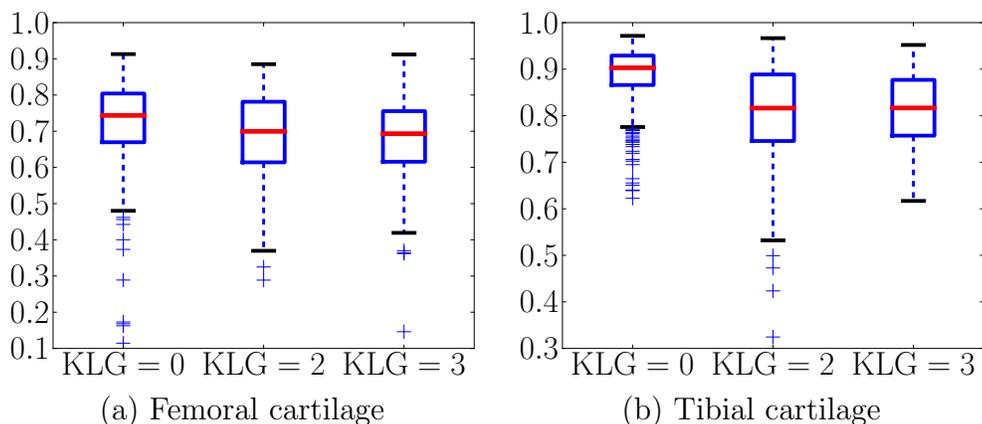
Figure 4.14: Boxplots of Pearson's correlation coefficients of local cartilage thickness for different KLG. The best strategy combination is used, SVM and non-local patch-based label fusion with an anisotropic regularization with $g = 1.4$ and $\alpha = 0.2$.

## 4.5 Comparison to other methods

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

### 4.5.1 Comparisons based on the SKI10 dataset

To compare to other algorithms I use the data from the cartilage segmentation challenge SKI10 (Heimann *et al.*, 2010). I 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. I restrict myself to comparisons between the top 8 methods. The proposed method ranks 5/16 overall. However, as I will discuss below the proposed method performs as well as the top method on volume overlap error for cartilage segmentation (or equivalently Dice similarity coefficient) which as I argue is the most important of the performance measures. For simplicity I denote the methods as Rank 1 to Rank 8 to simplify readability. Tables 4.3 and 4.4 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 4.15 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 $R$.

$$\text{VOE} = 100 \left( 1 - \frac{|S \cap R|}{|S \cup R|} \right), \tag{4.2}$$

$$\text{VD} = 100 \frac{|S| - |R|}{|R|}. \tag{4.3}$$

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

$$\text{DSC} = \frac{200 - 2 \times \text{VOE}}{200 - \text{VOE}}. \tag{4.4}$$

The proposed method achieves excellent performance on VOE and DSC. The VD is best at zero: the proposed 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 I regard them as more informative than VD for the assessment of cartilage segmentation differences.

Table 4.3 compares the proposed method to other methods based on the different SKI10 validation measures. Specifically, I test if scores of competing methods are significantly better than for our method.

The proposed method achieves statistically significantly better accuracy than most of the other methods regarding VOE and DSC before and after multiple comparison correction. Table 4.4 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

(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
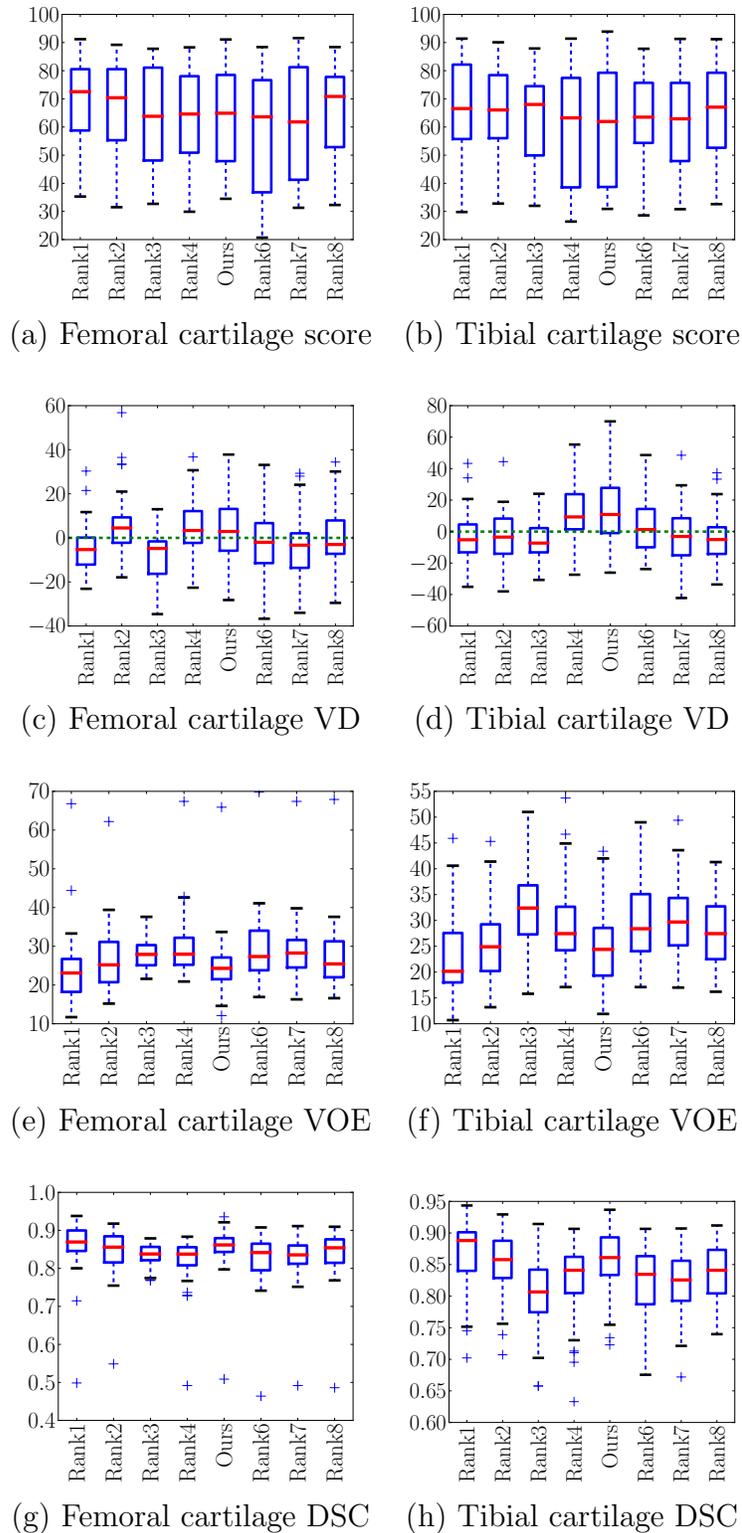
Figure 4.15: Box plots of segmentation measures for femoral and tibial cartilage from top 8 ranking methods on SKI10 challenge. The center red line is the median and the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

Table 4.3: Results of statistical tests (paired $t$ tests for score, VOE and DSC, Wilcoxon signed-rank tests for VD) between different methods. The proposed 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 author but using a slightly different combination, i.e., probabilistic $k$NN and locally-weighted label fusion.

| Rank | Team | Femoral cartilage | | | |
| --- | --- | --- | --- | --- | --- |
| | | Score | VD | VOE | DSC |
| 1 | Imorphics (Vincent *et al.*, 2010) | NS/NS | +/+ | NS/NS | NS/NS |
| 2 | ZIB (Seim *et al.*, 2010) | NS/NS | +/NS | +/NS | +/NS |
| 3 | UPMC_IBML | NS/NS | +/+ | +/NS | +/NS |
| 4 | SNU_SPL | NS/NS | NS/NS | +/+ | +/+ |
| 6 | UIiibiKnee (Yin *et al.*, 2010a) | NS/NS | −/− | +/+ | +/+ |
| 7 | shan_unc | NS/NS | +/+ | +/+ | +/+ |
| 8 | BioMedIA (Wang *et al.*, 2013) | NS/NS | +/+ | +/+ | +/+ |
| Rank | Team | Tibial cartilage | | | |
| | | Score | VD | VOE | DSC |
| 1 | Imorphics (Vincent *et al.*, 2010) | NS/NS | −/− | NS/NS | NS/NS |
| 2 | ZIB (Seim *et al.*, 2010) | NS/NS | −/− | NS/NS | NS/NS |
| 3 | UPMC_IBML | NS/NS | −/− | +/+ | +/+ |
| 4 | SNU_SPL | NS/NS | NS/NS | +/+ | +/+ |
| 6 | UIiibiKnee (Yin *et al.*, 2010a) | NS/NS | −/− | +/+ | +/+ |
| 7 | shan_unc | NS/NS | −/− | +/+ | +/+ |
| 8 | BioMedIA (Wang *et al.*, 2013) | NS/NS | −/− | +/+ | +/+ |

Table 4.4: Statistics summary (mean, median and standard deviation) of DSC from the top ranking methods. Rank 7 was also submitted by the author but using a different combination, i.e., probabilistic $k$NN and locally-weighted label fusion.

| Rank | Team | Femoral cartilage | | |
|---|---|---|---|---|
| | | Mean | Median | STD |
| 1 | Imorphics (Vincent *et al.*, 2010) | 0.861 | 0.869 | 0.065 |
| 2 | ZIB (Seim *et al.*, 2010) | 0.845 | 0.856 | 0.058 |
| 3 | UPMC_IBML | 0.836 | 0.838 | 0.028 |
| 4 | SNU_SPL | 0.821 | 0.838 | 0.059 |
| **5** | **shan_unc (proposed)** | **0.856** | **0.862** | **0.057** |
| 6 | UIiibiKnee (Yin *et al.*, 2010a) | 0.824 | 0.842 | 0.067 |
| 7 | shan_unc | 0.828 | 0.836 | 0.060 |
| 8 | BioMedIA (Wang *et al.*, 2013) | 0.840 | 0.854 | 0.062 |
| Rank | Team | Tibial cartilage | | |
| | | Mean | Median | STD |
| 1 | Imorphics (Vincent *et al.*, 2010) | 0.865 | 0.888 | 0.054 |
| 2 | ZIB (Seim *et al.*, 2010) | 0.850 | 0.858 | 0.049 |
| 3 | UPMC_IBML | 0.805 | 0.807 | 0.057 |
| 4 | SNU_SPL | 0.824 | 0.841 | 0.058 |
| **5** | **shan_unc (proposed)** | **0.859** | **0.861** | **0.047** |
| 6 | UIiibiKnee (Yin *et al.*, 2010a) | 0.825 | 0.834 | 0.056 |
| 7 | shan_unc | 0.820 | 0.826 | 0.051 |
| 8 | BioMedIA (Wang *et al.*, 2013) | 0.836 | 0.841 | 0.048 |

ranked method. In particular, one does not observe statistically significant performance differences 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 the proposed method for femoral cartilage segmentation with respect to the second ranked method by Seim *et al.* (2010). This suggests that the proposed 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 (Vincent *et al.*, 2010). However, 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. The proposed 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 the overall segmentation method, the proposed method remains flexible enough to also capture cartilage variations not strictly contained in the atlas set.

Note that the SKI10 (Heimann *et al.*, 2010) 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 over-segmentation at pathological regions; e.g., segmenting across regions of total cartilage loss or segmenting osteophytes. Figure 4.16 shows an example illustrating total cartilage loss and the challenge to define a reliable gold standard segmentation.
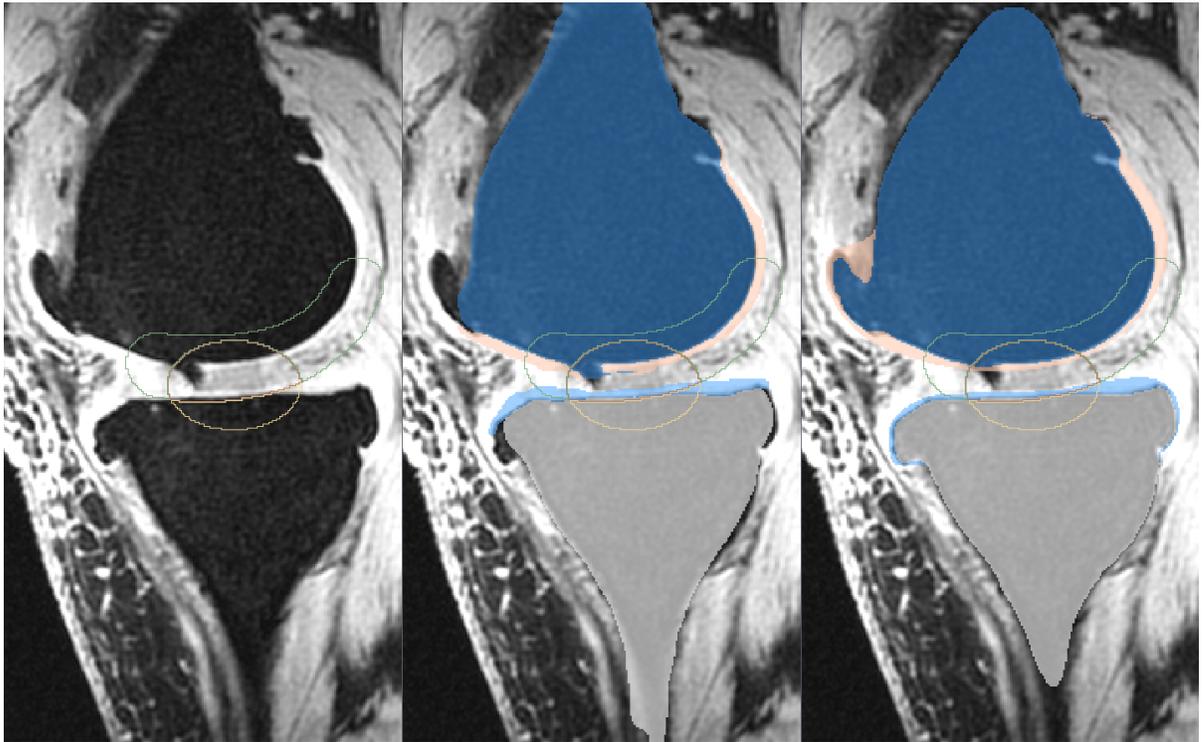
(a) Original image      (b) Automatic segmentation      (c) Expert segmentation

Figure 4.16: An example slice from SKI10 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. The proposed segmentation successfully delineates it, but the expert segmentation fails to do so.

### 4.5.2   Qualitative comparison to other methods

The methods that have not been tested on SKI10 (Heimann *et al.*, 2010) dataset are not directly comparable to our method because of different datasets. Note that the proposed method compares favorably to other methods, however, none of the competing methods were validated on datasets as large as the dataset I test on (with more than 700 images for the PLS data alone). For example, Folkesson *et al.* (2007) tested on 139 images, (Fripp *et al.*, 2010) 20 images, Tamez-Peña *et al.* (2012) used 12 images and Yin *et al.* (2010b) 60 images. Hence, my validation dataset is an order of magnitude larger than for most other existing studies.

### 4.6   Conclusion

The method proposed in chapter 3 is robust, because multi-atlas-based methods can overcome occasional registration failures. This is a critical aspect when applying the approach to larger datasets, such as the OAI dataset.

The contribution of this chapter is the extensive validation of the proposed cartilage segmentation method on the PLS and the SKI10 dataset. The PLS dataset contains 706 images from 155 subjects. It is the largest dataset (an order of magnitude larger than most exciting studies) that has ever been validated on in literature.

It can be concluded that best segmentation strategy (from the choices mentioned in this chapter) is 1) multi-atlas-based segmentation with patch-based label fusion for spatial prior, 2) SVM for local likelihood and 3) three-label segmentation with anisotropic regularization.

I also demonstrated that the proposed segmentation strategy performs as well as the top-ranking methods on the SKI10 dataset.

# CHAPTER 5: LONGITUDINAL THREE-LABEL CARTILAGE SEGMENTATION

## 5.1 Introduction

Since subtle changes of cartilage might be indicative of early OA, it is desirable to study longitudinal cartilage changes using a temporally-consistent segmentation which can mitigate image noise effects. However, the methods presented in chapter 3 treat each image volume separately without exploring the temporal consistency within the same subject. For brain segmentation, Xue *et al.* (2006) proposed a longitudinal segmentation method by adding a temporal consistency constraint term to a fuzzy clustering segmentation.

In this chapter, I propose a novel general longitudinal three-label segmentation approach and apply it to the cartilage segmentation problem. The longitudinal three-label segmentation is an extension of segmentation methods (2.10 and 2.11) presented in chapter 2 and is also formulated as a convex optimization problem. A temporal consistency term is added to the existing three-label segmentation (2.10 and 2.11). I make use of temporally-independent bone and cartilage segmentations to transform the longitudinal image data from native image space to a common longitudinal image space for each subject.

The contributions of this chapter include a novel general longitudinal three-label segmentation method, the application of the proposed method to achieve a fully automatic longitudinal cartilage segmentation and the evaluation of longitudinal segmentation on a sizable longitudinal dataset consisting of more than 700 images against temporally-independent segmentation.

## 5.2    Longitudinal three-label segmentation

Rather than treating each image separately, one can make use of the temporal consistency of longitudinal image data from the same subject to improve the segmentation. I propose a novel longitudinal three-label segmentation method to mitigate possible noise effect and encourage segmentation consistency across time points. The method is general and capable of solving other segmentation problems with two objects.

Assuming the longitudinal image data for a given subject has been registered into a common space, I formulate the longitudinal three-label segmentation energy as

$$E(u) = \int_{\mathcal{D}} \|\mathbf{G}\nabla_{\mathbf{x}}u\| + c\,|\nabla_l u| + h\,|\nabla_t u|\ d\mathbf{x}dldt,\ \mathcal{D} = \Omega \times \mathcal{L} \times \mathbf{T}$$

$$\text{subject to } u \in [0,1],\ u(\mathbf{x},0,t) = 0,\ u(\mathbf{x},3,t) = 1, \tag{5.1}$$

where $\mathbf{T}$ is the time domain; $\nabla_t u$ is the gradient in time dimension, $\nabla_t u = \partial u/\partial t$; $h > 0$ controls the temporal regularization. Matrix $\mathbf{G}$ determines the spatial regularization as defined in (2.12). This is a convex formulation and yields a global optimal solution. As the cartilage is a thin structure this energy (defined over a 5-dimensional space) can still be efficiently optimized. Figure 5.1 demonstrates the benefit of the longitudinal segmentation which is more resistant to image noise than the temporally-independent segmentation.

### 5.2.1    Numerical solution

This section discusses an iterative scheme to optimize (5.1). I introduce three dual variables $\mathbf{p}$ (vector field), $q$ (scalar field) and $s$ (scalar field) and rewrite (5.1) as

$$E(u, \mathbf{p}, q, s) = \int_{\mathcal{D}} \langle \mathbf{p}, \mathbf{G}\nabla_{\mathbf{x}}u \rangle + q\nabla_l u + s\nabla_t u\ d\mathbf{x}dldt,$$

$$\text{subject to } \|\mathbf{p}\| \leq 1,\ |q| \leq c,\ |s| \leq h, \tag{5.2}$$

(a) $t = 0$         (b) $t = 1$         (c) $t = 2$

(d) $t = 0$         (e) $t = 1$         (f) $t = 2$
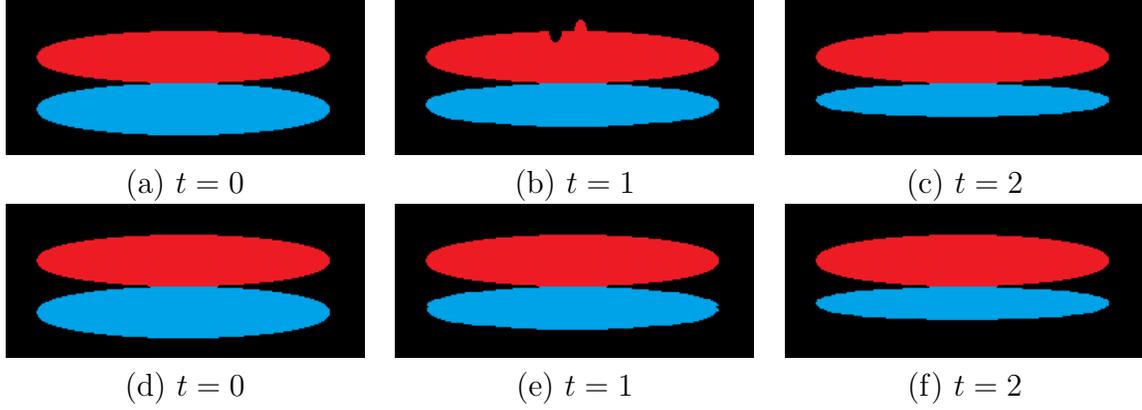
Figure 5.1: Synthetic example demonstrating the benefit of longitudinal segmentation. (a), (b) and (c) are longitudinal images and (b) has image noise. (d), (e) and (f) are the corresponding segmentation from the longitudinal segmentation model (5.1). The segmentation result is resistant to image noise (red object) and still captures the atrophy of the blue object.

in which $\langle \cdot, \cdot \rangle$ represents inner products. Minimizing (5.1) with respect to $u$ is equivalent to minimizing (5.2) with respect to $u$ and maximizing it with respect to $\mathbf{p}$, $q$ and $s$.

Taking the variation yields

$$\delta E(u, \mathbf{p}, q; \delta u, \delta \mathbf{p}, \delta q, \delta s) \tag{5.3}$$

$$= \frac{\partial}{\partial \epsilon} \int_{\Omega} \langle \mathbf{p} + \epsilon \delta \mathbf{p}, \mathbf{G} \nabla_{\mathbf{x}} u + \epsilon \mathbf{G} \nabla_{\mathbf{x}} \delta u \rangle + (q + \epsilon \delta q)(\nabla_l u + \epsilon \nabla_l \delta u)$$

$$+ (s + \epsilon \delta s)(\nabla_t u + \epsilon \nabla_t \delta u) \; d\mathbf{x} dl dt \tag{5.4}$$

$$= \int_{\Omega} \langle \mathbf{p}, \mathbf{G} \nabla_{\mathbf{x}} \delta u \rangle + \langle \delta \mathbf{p}, \mathbf{G} \nabla_{\mathbf{x}} u \rangle + q \nabla_l \delta u + \delta q \nabla_l u + s \nabla_t \delta u + \delta s \nabla_t u \; d\mathbf{x} dl dt \tag{5.5}$$

$$= \int_{\Omega} (-\operatorname{div}(\mathbf{G}\mathbf{p}) - \nabla_l q - \nabla_t s)\delta u + \langle \delta \mathbf{p}, \mathbf{G} \nabla_{\mathbf{x}} u \rangle + \delta q \nabla_l u + \delta s \nabla_t u \; d\mathbf{x} dl dt \tag{5.6}$$

The gradient descent/ascent update scheme of (5.2) is

$$\mathbf{p}_t = -\mathbf{G}\nabla_x u, \ \|\mathbf{p}\| \leq 1 \tag{5.7}$$

$$q_t = -\nabla_l u, \ |q| \leq c \tag{5.8}$$

$$s_t = -\nabla_t u, \ |s| \leq h \tag{5.9}$$

$$u_t = -\operatorname{div}_\mathbf{x}(\mathbf{Gp}) - \nabla_l q - \nabla_t s \tag{5.10}$$

The iterative scheme will lead to a global optimum upon convergence (Appleton and Talbot, 2006) because of the convexity of (5.1). Let $S$ and $T$ denote the source and sink sets: $S = \Omega \times \{0\} \times \mathbf{T}, \ T = \Omega \times \{3\} \times \mathbf{T}$. The region without sources or sinks is denoted as $\overset{\circ}{\mathcal{D}} = \mathcal{D} \setminus (S \cup T) = \Omega \times \{1, 2\} \times \mathbf{T}$. The dual energy is

$$
\begin{aligned}
E^*(u) = & \int_S \operatorname{div}(\mathbf{Gp}) + \nabla_l q + \nabla_t s \ d\mathbf{x}dldt \\
& + \int_{\overset{\circ}{\mathcal{D}}} \min(0, \operatorname{div}(\mathbf{Gp}) + \nabla_l q + \nabla_t s) \ d\mathbf{x}dldt.
\end{aligned}
\tag{5.11}
$$

The iterations are terminated when the duality gap between the primal energy (5.1) and the dual energy (5.11) 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 $u$. It has been shown that thresholding of the essentially binary solution $u$ of (5.1) yields to a equally globally optimal solution to the original discrete problem(Zach *et al.*, 2009).

## 5.3 Registration of longitudinal images

The longitudinal segmentation model (5.1) requires registered longitudinal data. However the images at different time points are not aligned. Therefore one needs to register

the longitudinal images of a given subject into a common space (longitudinal space) before performing longitudinal segmentation. Here I use the temporally-independent bone and cartilage segmentation discussed in chapter 3.

Specifically I use independent rigid transformations based on the femur and tibia segmentations to initialize rigid transformations between the temporally-independent femoral and tibial cartilage segmentations (for improved registration robustness). The labeling cost for (5.1) can then be computed as

$$c(\mathbf{x}, l, t) = -log(p_t(l|\mathbf{f}(\mathbf{x}))) = -log\left(\frac{p_t(\mathbf{f}(\mathbf{x})|l) \cdot p_t(l)}{p_t(\mathbf{f}(\mathbf{x}))}\right), \tag{5.12}$$

where the subscript $t$ represents the likelihoods at the $t$-th time point which are propagated from the native image space:

$$
\begin{aligned}
p_t(\mathbf{f}(\mathbf{x})|\text{FC}) &= R_t^{\text{FC}} \circ R_t^{\text{FB}} \circ p(\mathbf{f}(\mathbf{x})|\text{FC}), \\
p_t(\mathbf{f}(\mathbf{x})|\text{TC}) &= R_t^{\text{TC}} \circ R_t^{\text{TB}} \circ p(\mathbf{f}(\mathbf{x})|\text{TC}), \\
p_t(\text{FC}) &= R_t^{\text{FC}} \circ R_t^{\text{FB}} \circ p(\text{FC}), \\
p_t(\text{TC}) &= R_t^{\text{TC}} \circ R_t^{\text{TB}} \circ p(\text{TC}).
\end{aligned}
\tag{5.13}
$$

Here, $R_t$ denotes a rigid transform for the $t$-th time point. Superscripts specify the registration ("FC": femoral cartilage, "TC": tibial cartilage, "FB": femoral bone, "TB": tibial bone). Local classification likelihoods $p(\mathbf{f}(\mathbf{x})|\text{FC})$ and $p(\mathbf{f}(\mathbf{x})|\text{TC})$ are computed from the trained SVM classifier as in chapter 3. Spatial priors $p(\text{FC})$ and $p(\text{FC})$ are obtained from patch-based label fusion as discussed in chapter 3.

The choice of the common space is important. To avoid overlap of femoral and tibial cartilage after registration, I choose the first time-point (baseline image) as the common space because it is expected to have the thickest cartilage. The longitudinal segmentation of each cartilage can then be obtained by optimizing (5.1) with labeling cost defined by

(5.12).

## 5.4   Experimental results

This section compares the proposed longitudinal segmentation against the temporally-independent segmentation presented in chapter 3 on the PLS dataset, which is described in section 4.2. The SKI10 dataset is not a longitudinal dataset and can therefore not be used for validation.

Expert cartilage segmentations are available for all images in the native image space. The femoral cartilage segmentation is drawn only on the weight-bearing part while the tibial cartilage segmentation covers the entire region.

As in chapter 3, I use images from the baseline images from 18 subjects as training atlases (for multi-atlas registration and local classification) to segment the remaining 137 subjects. Within these 18 subjects, each subject is tested using the other 17 images for training.

Figure 5.2 compares the mean Dice similarity coefficient (DSC) of the longitudinal segmentation of each cartilage with increasing amount of temporal regularization. The DSC is computed in both the longitudinal image space and the native image space. The mean DSC increases with an appropriate amount of temporal regularization, which demonstrates the advantage of the longitudinal segmentation model. The difference of the mean DSC in the two spaces is due to the rigid transformations and resamplings.

The resampling of probabilities ($p(\mathbf{f}(\mathbf{x})|\text{FC})$, $p(\mathbf{f}(\mathbf{x})|\text{TC})$, $p(\text{FC})$, and $p(\text{FC})$) diminishes sharp features and thus has a smoothing effect. Less spatial regularization is needed for the longitudinal segmentation to generate satisfactory results than the temporally-independent segmentation. It was concluded in chapter 4 that $g = 1.4$ and $\alpha = 0.2$ yields best segmentation accuracy. For the longitudinal segmentation, using $g = 0.2$ and $\alpha = 0.2$ results in the best segmentation accuracy.

(a) Femoral cariltage, longitudinal space

(b) Femoral cartilage, native space

(c) Tibial cartilage, longitudinal space

(d) Tibial cartilage, native space
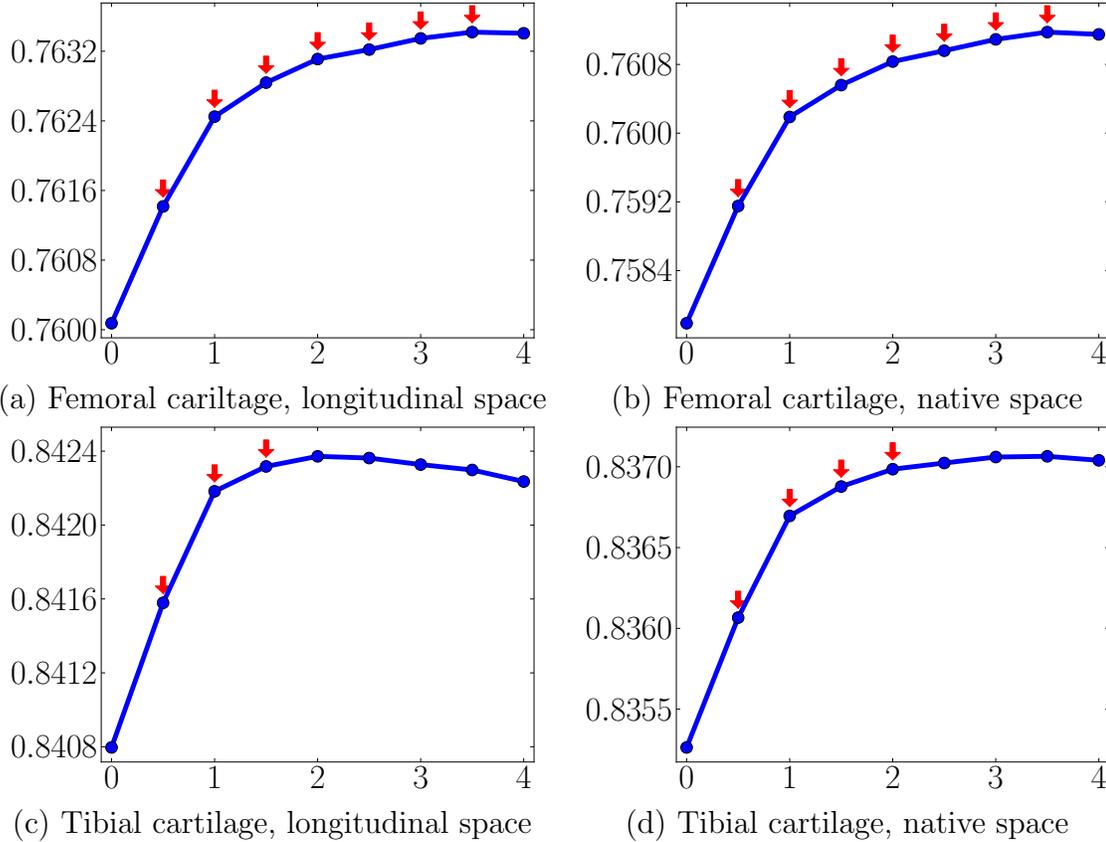
Figure 5.2: Change of mean DSC for femoral and tibial cartilage over the amount of temporal regularization $h$ (abscissa). The spatial regularization control $g$ is set to be 0.2 and the anisotropy parameter $\alpha$ is set to be 0.2. The computation of DSC in the longitudinal image space is done by transforming the expert segmentations to the longitudinal image space. The DSC is also computed in the native image space by transforming the longitudinal segmentations back into the native image space. Improvement of the mean DSC by including temporal regularization is observed in both spaces. The red downarrows ($\Downarrow$) indicate statistically significant differences between the current and the previous temporal regularization settings via paired t-tests at a significance level of 0.05.

Table 5.1 compares the validation statistics in the native image space from the longitudinal segmentation and temporally-independent segmentation under their own best parameter settings. Since there is only one expert segmentation which is drawn timepoint-by-timepoint (and therefore not expected to be longitudinally consistent) one cannot directly assess improvements in segmentation accuracy of the longitudinal method with respect to the individual segmentations. Table 5.1 indicates that the temporal smoothing inherent in the longitudinal model does not greatly affect the result with respect to *temporally-independent* manual segmentations. However, more temporary consistency is more biologically plausible, which is achieved by the longitudinal segmentation.

To evaluate the improvement of temporal consistency, I use the weighted sum of number of label changes over time as the temporal consistency measure (TCM) defined by

$$\text{TCM} = \sum_{i=1}^{n-1} \sum_{\mathbf{x} \in \Omega} \frac{|S_{i+1}(\mathbf{x}) - S_i(\mathbf{x})|}{t_{i+1} - t_i}, \tag{5.14}$$

where $n$ is number of time points $(n > 1)$, $\mathbf{x}$ is a voxel in the image domain $\Omega$, $S_i$ is the segmentation and $t_i$ is the time at $i$-th time point. The weighting is based on the assumption that label changes are more likely to occur over a longer time period. For those patients who have only one scan available $(n = 1)$, the measure is set to be 0. I compare the TCM of temporally-independent, longitudinal and expert segmentations of the femoral cartilage for all the subjects in Fig. 5.3. The longitudinal segmentation achieves the best temporal consistency. Note that the expert segmentations are drawn independently in each native image space which explains the high TCM. The same comparison result is also observed for the tibial cartilage.

## 5.5 Conclusion

In this chapter, a novel longitudinal three-label segmentation approach is proposed to encourage the temporal consistency of the segmentation of longitudinal data. The

Table 5.1: Compare longitudinal and temporally-independent segmentation in terms of mean, median and standard deviation of DSC, sensitivity and specificity computed in the native image space.FC: Femoral cartilage. TC: Tibial cartilage. Lon: Longitudinal segmentation with $g = 0.2$, $\alpha = 0.2$ and $h = 3.0$. Ind: Temporally-independent segmentation with $g = 1.4$ and $\alpha = 0.2$.

|  |  | DSC | | | Sensitivity | | | Specificity | | |
|---|---|---|---|---|---|---|---|---|---|---|
|  |  | Mean | Median | STD | Mean | Median | STD | Mean | Median | STD |
| FC | Lon | 76.1% | 76.7% | 4.6% | 84.9% | 86.1% | 6.7% | 99.8% | 99.8% | 0.05% |
|  | Ind | 76.0% | 76.8% | 4.8% | 85.3% | 86.7% | 7.2% | 99.8% | 99.8% | 0.05% |
| TC | Lon | 83.7% | 84.5% | 3.7% | 89.1% | 89.4% | 3.9% | 99.8% | 99.8% | 0.07% |
|  | Ind | 84.1% | 84.7% | 3.7% | 89.4% | 89.9% | 4.1% | 99.8% | 99.8% | 0.07% |



(a) Femoral cariltage

(b) Tibial cartilage
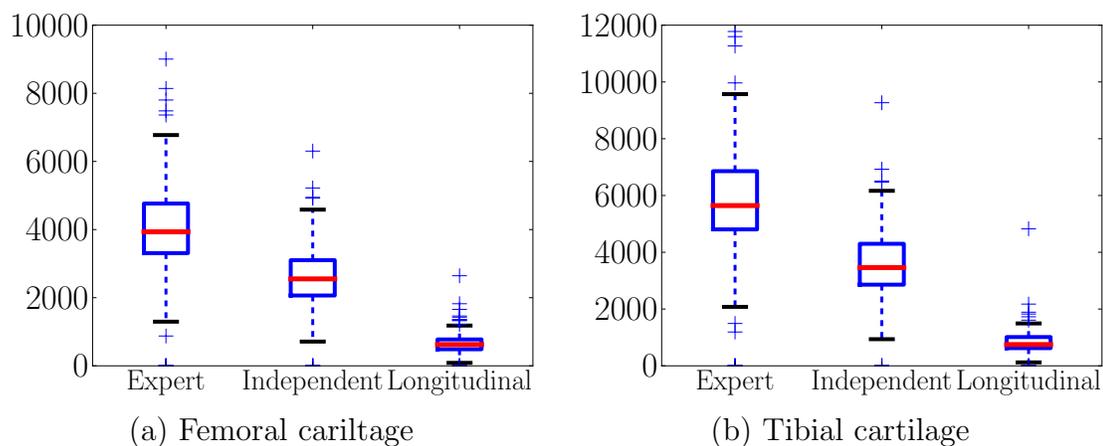
Figure 5.3: Box plots of temporary consistency measure (TCM) of each cartilage from expert, temporally-independent ($g = 1.4$, $\alpha = 0.2$) and longitudinal ($g = 0.2$, $\alpha = 0.2$, $h = 3.0$) segmentations.

approach is general and can be applied to other longitudinal segmentation problems with two objects. I also propose a pipeline to register longitudinal knee images from native image space to the common (longitudinal) space to establish correspondence across time for each individual subject. Experimental results on cartilage segmentation demonstrate the improvement in temporal consistency using the proposed longitudinal segmentation against the temporally-independent segmentation.

In the next chapter, I will discuss statistical analysis of cartilage thickness given the segmentations obtained from chapter 3 and 5.

# CHAPTER 6: LOCALIZED ANALYSIS OF LONGITUDINAL CARTILAGE THICKNESS

## 6.1 Introduction

Statistical analysis of cartilage thickness changes faces two major challenges. The first one is the lack of spatial correspondence across subjects and/or time. As cartilage loss in OA may not be uniform throughout the cartilage (Biswal *et al.*, 2002), advanced statistical analysis methods are necessary to understand localized cartilage thinning. However, the localized analysis requires establishing spatial correspondence across time and subjects, which is challenging due to the small volume of cartilage in relation to the rest of the knee. The second challenge is the spatial heterogeneity of cartilage progression across subjects. This is caused by the fact that cartilage thinning may happen at different locations in different subjects. Cartilage thinning at a particular location may only be consistent for a subset of the full subject population. Thus, treating all OA subjects equally and using all of them for standard statistical methods would be problematic.

Existing longitudinal OA studies often rely on subdivisions of the (manually segmented) cartilage into corresponding regions (Wirth and Eckstein, 2008; Buck *et al.*, 2009) to avoid global cartilage measures without requiring registration. However, such subdivisions are purely geometric and necessarily coarse. Local changes (that happen to a smaller region than the size of a subregion) are weakened by averaging over a particular subregion and are difficult to capture. To fully understand the spatial pattern of OA progression, the analysis of *localized* cartilage thickness changes is necessary.

To avoid the shortcomings of existing cartilage analysis methods, I propose (i) an automatic analysis method establishing a common (atlas) spatial coordinate system for

cartilage analysis of the knee and (ii) a clustering-based statistical method to approach the spatial heterogeneity across OA subjects. The proposed method is applied to the PLS dataset.

## 6.2   Cross-sectional vs. longitudinal analysis[1]

Before presenting the proposed method for longitudinal cartilage analysis, let me first discuss why longitudinal analysis is necessary in the first place for longitudinal datasets.

Cross-sectional studies collect observations at one time point or over a short time period in a given population. In a longitudinal study, a single measurement is collected repeatedly over time on each subject in the study. Because of the repeated observation at the individual level, longitudinal studies have more power than cross-sectional studies. Suppose a cross-sectional study involves people of different ages. One can only tell the difference among differently aged subjects. With a longitudinal study, one will be able to address the effects of aging on a single individual's response. Suppose there are samples of two groups, say men and women, in a cross-sectional study. One can describe the difference in level between the two groups at the current time. With a longitudinal study, analysts will additionally be able to describe the separate trends over time of each group and a single individual's trend.

The defining feature of longitudinal data is that multiple observations within subject can be ordered across time. The temporal ordering of the measurement is important because measurements closer in time within a subject are likely to be more similar than observations further apart in time. A positive correlation between longitudinal measurements is often expected. Ignoring the correlation can lead to biased inference of the standard errors of the coefficients.

---

[1]Most of the analysis in this section are from `http://ehs.sph.berkeley.edu/hubbard/longdata/webfiles/Chapter_3_Feb_2006.pdf`

### 6.2.1 Cross-sectional analysis of longitudinal data

Applying cross-sectional data analysis methods, e.g., linear regression, results in incorrect estimation of variability. Suppose there are $n$ subjects and subject $i$ has $m_i$ observations. The random variable $Y_{i,j}$ denotes the response observed from subject $i$ $(1 \leq i \leq n)$ at time $t_j$ $(1 \leq j \leq m_i)$. Let

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i,1} \\ \vdots \\ Y_{i,m_i} \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} 1 & t_{i,1} \\ \vdots & \vdots \\ 1 & t_{i,m_i} \end{pmatrix}, \text{and } \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}. \tag{6.1}$$

The ordinary linear regression finds the line that "best" fits the data by minimizing the total amount of deviation

$$E(\boldsymbol{\beta}) = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2, \tag{6.2}$$

with $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ and $\beta_0, \beta_1$ are, respectively, the intercept and the slope of the fitted line. Note that the residual sums of squares treats each observation equally, making no distinction between longitudinal observations on the same individual from those on different individuals. A closed-form solution (often referred as ordinary least square estimator) of $\boldsymbol{\beta}$ can be calculated by minimizing (6.2)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{6.3}$$

Let us take a look at the expectation of the random vector $\boldsymbol{\beta}$ to see how it is affected

by the correlation of the $Y_{i,j}$'s.

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}\right) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E\left(\mathbf{Y}\right) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}.
\end{aligned}
\tag{6.4}
$$

One can see from the derivation above that the expectation of parameter $\boldsymbol{\beta}$ does not depend on the covariance of the $Y_{i,j}$'s. However, model (6.2) makes no distinction between changes in variable $\mathbf{X}$ that occur over longitudinal observations within the same individual and changes of $\mathbf{X}$ across different individuals. To take advantage of the longitudinal structure of the data, one must properly parameterize the regression model to separate out the longitudinal effects from those due to selection bias.

Let us also take a look at the variability of the least square estimator $\hat{\boldsymbol{\beta}}$.

$$
\begin{aligned}
\text{var}(\hat{\boldsymbol{\beta}}) &= \text{var}\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}\right) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\,\text{var}(\mathbf{Y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}.
\end{aligned}
\tag{6.5}
$$

The covariance matrix of $\hat{\boldsymbol{\beta}}$ depends on the covariance matrix of $\mathbf{Y}$. As (6.2) assumes all observations are independent, $\text{var}(\mathbf{Y})$ is a diagonal matrix. Taking $\text{var}(\mathbf{Y}) = \sigma^2\mathbf{I}$ results in

$$
\begin{aligned}
\text{var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.
\end{aligned}
\tag{6.6}
$$

If the assumption of independence is invalid (which is the case for longitudinal data), the derivation above introduces bias into the estimation of the variability of $\hat{\boldsymbol{\beta}}$, translating into incorrect inference.

85

### 6.2.2   Linear mixed effects model for longitudinal data

The general form of linear mixed effects model can be expressed as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \text{ for the } i\text{th subject, } i = 1, \ldots, n. \qquad (6.7)$$

Matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ are design matrices for fixed effects and random effects, respectively. Suppose $\mathbf{X}_i$ and $\mathbf{Z}_i$ are the same.

$$\mathbf{X}_i = \mathbf{Z}_i = \begin{pmatrix} 1 & t_{i,1} \\ \vdots & \vdots \\ 1 & t_{i,m_i} \end{pmatrix}. \qquad (6.8)$$

Parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ characterizes the fixed effects, i.e., the systematic part of the response. Parameter $\mathbf{b}_i = (b_{0i}, b_{1i})^T$ describes the random effects due to among-subject variation. $\mathbf{b}_i \sim \mathcal{N}_k(\mathbf{0}, \mathbf{D})$ and $\mathbf{D}$ is a covariance matrix characterizing variation due to among-subject sources, assumed the same for all subjects. The dimension of $\mathbf{D}$ depends on the number of among-subject random effects in the model. Within-subject deviations are characterized by vector $\mathbf{e}_i \sim \mathcal{N}_{m_i}(\mathbf{0}, \mathbf{R}_i)$. Here, $\mathbf{R}_i$ is a $m_i \times m_i$ matrix that characterizes variance and correlation due to within-subject sources.

With these assumptions, we have

$$E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}$$

$$var(\mathbf{Y}_i) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \mathbf{R}_i = \boldsymbol{\Sigma}_i \qquad (6.9)$$

$$\mathbf{Y}_i \sim \mathcal{N}_{n_i}\left(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i\right)$$

The maximum-likelihood estimator for $\boldsymbol{\beta}$ can be written as

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^{n} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Y}_i \tag{6.10}$$

Note that this is not a closed-form solution since $\hat{\boldsymbol{\Sigma}}_i^{-1}$ is also unknown. This estimator is often called the generalized least squares estimator of $\boldsymbol{\beta}$. When the sample size $n$ grows infinitely large, we have

$$E(\hat{\boldsymbol{\beta}}) = E\left[ \left( \sum_{i=1}^{n} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Y}_i \right] \approx \boldsymbol{\beta}, \tag{6.11}$$

and

$$var(\hat{\boldsymbol{\beta}}) \approx \left( \sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right)^{-1} \tag{6.12}$$

$$= \left( \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^{-1}, \tag{6.13}$$

with

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma}_n \end{pmatrix}.$$

The generalized least square estimator is *not* numerically equivalent to the ordinary least square estimator described in section 6.2.1. Therefore the cross-sectional analysis produces wrong estimation of the covariance, which can be avoided by the linear mixed effects model.

## 6.3 Spatial correspondence of cartilage thickness

Existing methods partition the cartilage into several subregions, which is coarse and arbitrary, to establish spatial correspondence across subjects and time. This section presents an automatic method to establish a common (atlas) spatial coordinate system for cartilage analysis of the knee.

Given an cartilage segmentation, local thickness is computed using a Laplace equation approach (Yezzi and Prince, 2003). Thickness measurements are performed in native image space for expert manual segmentations and for the temporally-independent segmentations. For the longitudinal segmentations, thickness is computed in the longitudinal space, i.e., the space of the baseline image of a given subject.

The thicknesses are then mapped to a common atlas space. I use an affine followed by a B- Spline transform based on the bone segmentations for each cartilage thickness volume. I then project each thickness volume down to an axial slice by taking the median thickness value along the axial direction (where thicknesses are approximately constant). The thickness maps are then comparable across time-points and subjects.

Figure 6.1 illustrates the pipeline which takes the original 3D MR image and converts to the 2D thickness map for each cartilage. Examples of thickness maps are shown in Fig. 6.2.

## 6.4 Longitudinal analysis of localized cartilage thickness changes

The questions of interest in this chapter include i) if there is a significant difference of baseline thickness between OA and normal control (NC) subjects, and ii) if there is a significant difference of longitudinal thickness change between OA and NC subjects.

To answer these two questions, simply applying statistical models to all the subjects would be problematic because that would treat all OA subjects equally and cartilage thinning is a spatially heterogeneous process. Because of the fact that cartilage thinning
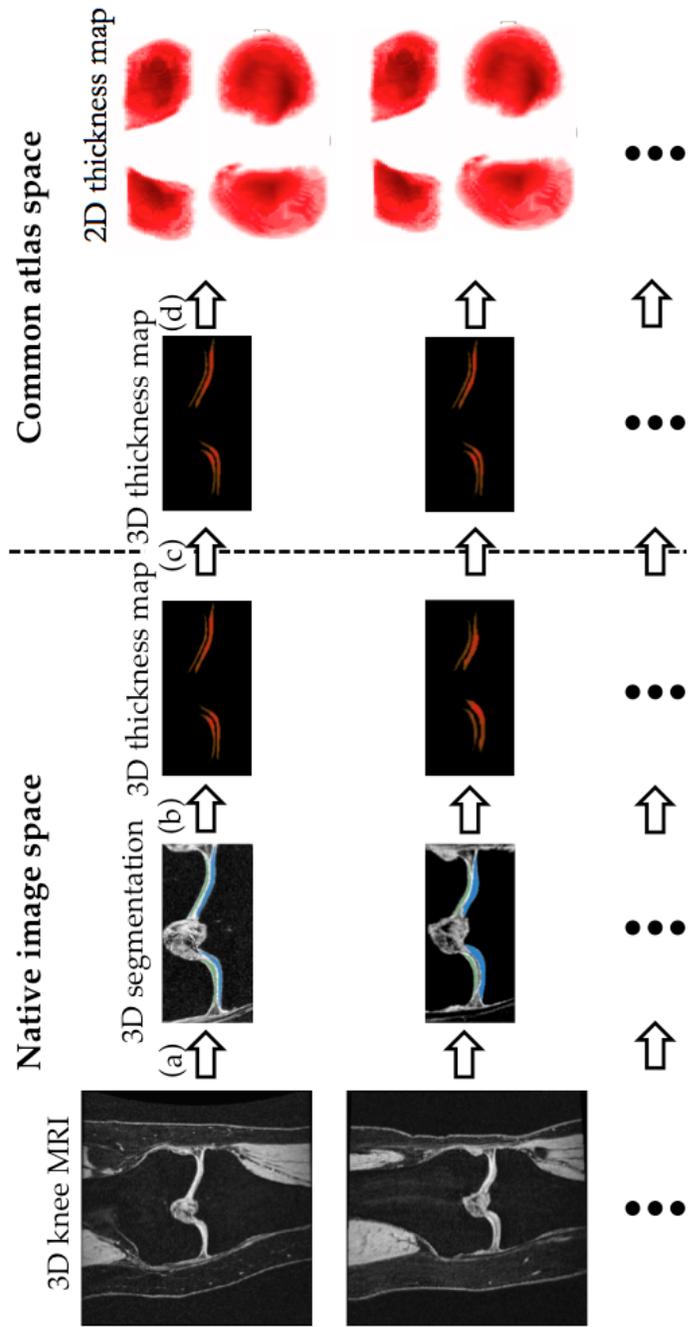
Figure 6.1: Thickness computation pipeline. (a) Extract the joint region and compute cartilage segmentations. (b) Compute cartilage thickness. (c) Register all thickness maps to a common atlas space. (d) Project 3D thickness onto a 2D plane.

(a) Femoral cartilage  (b) Tibial cartilage
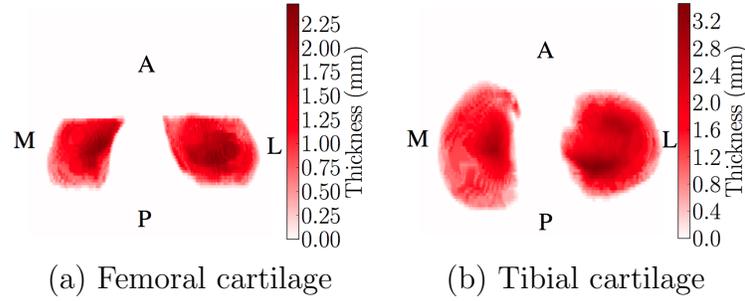
Figure 6.2: Examples of 2D cartilage thickness maps. M: medial; L: lateral; A: anterior; P: posterior.

may happen to different locations for different subjects, some subjects exhibit similar thinning patterns than other subjects. It is therefore sensible to cluster the OA subjects into a set of groups. The subjects within the same group have similar thinning patterns. In this section, I propose a clustering-based statistical method to approach the spatial heterogeneity of cartilage loss.

Suppose there are $n$ subjects and subject $i$ has $m_i$ time points at $\{t_1, \ldots, t_{m_i}\}$. Let $y_{i,j}(\mathbf{x})$ denote the thickness of subject $i$ $(1 \leq i \leq n)$ at time point $t_j$ $(1 \leq j \leq m_i)$ and pixel location $\mathbf{x}$. Let us further define

$$
\mathbf{Y}_i(\mathbf{x}) = \begin{pmatrix} y_{i,1}(\mathbf{x}) \\ \vdots \\ y_{i,m_i}(\mathbf{x}) \end{pmatrix}, \text{ and } \mathbf{X}_i = \begin{pmatrix} 1 & t_{i,1} \\ \vdots & \vdots \\ 1 & t_{i,m_i} \end{pmatrix}.
\tag{6.14}
$$

The delta function $\delta_i$ for subject $i$ is defined as

$$
\delta_i = \begin{cases} 1 & \text{if subject } i \text{ is NC, i.e., KLG} = 0, \\ 0 & \text{if subject } i \text{ has OA, i.e., KLG} > 0. \end{cases}
\tag{6.15}
$$

The following linear mixed model is applied at each pixel location $\mathbf{x}$. Parameters $\beta_0$ and $\beta_1$ are the intercept and the slope, respectively, of *fixed* effects for NC subjects, and $\beta_{2i}$ and $\beta_{3i}$ are the intercept and the slope of *random* effects for an NC subject $i$.

Parameters $\beta_4$ and $\beta_5$, respectively, model the *difference* of slope and intercept of *fixed* effects between NC and OA subjects, and $\beta_{6i}$ and $\beta_{7i}$ describe the slope and the intercept of random effects for OA subject $i$.

$$\mathbf{Y}_i(\mathbf{x}) = \mathbf{X}_i \begin{pmatrix} \beta_0(\mathbf{x}) \\ \beta_1(\mathbf{x}) \end{pmatrix} + (1 - \delta_i)\, \mathbf{X}_i \begin{pmatrix} \beta_{2i}(\mathbf{x}) \\ \beta_{3i}(\mathbf{x}) \end{pmatrix}$$
$$+ \delta_i \mathbf{X}_i \begin{pmatrix} \beta_4(\mathbf{x}) \\ \beta_5(\mathbf{x}) \end{pmatrix} + \delta_i \mathbf{X}_i \begin{pmatrix} \beta_{6i}(\mathbf{x}) \\ \beta_{7i}(\mathbf{x}) \end{pmatrix} + \mathbf{e}_i(\mathbf{x}). \qquad (6.16)$$

The simplified version of model (6.16) is

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}^{\mathrm{f,nc}} + (1 - \delta_i)\mathbf{X}_i \boldsymbol{\beta}_i^{\mathrm{r,nc}} + \delta_i \mathbf{X}_i \boldsymbol{\beta}^{\mathrm{f,oa}} + \delta_i \mathbf{X}_i \boldsymbol{\beta}_i^{\mathrm{r,oa}} + \mathbf{e}_i, \qquad (6.17)$$

where $\boldsymbol{\beta}^{\mathrm{f,nc}} = (\beta_0, \beta_1)^T$, $\boldsymbol{\beta}^{\mathrm{r,nc}} = (\beta_{2i}, \beta_{3i})^T \sim \mathcal{N}(\mathbf{0}, \mathbf{D}^{\mathrm{nc}})$, $\boldsymbol{\beta}^{\mathrm{f,oa}} = (\beta_4, \beta_5)^T$, and $\boldsymbol{\beta}^{\mathrm{r,oa}} = (\beta_{6i}, \beta_{7i})^T \sim \mathcal{N}(\mathbf{0}, \mathbf{D}^{\mathrm{oa}})$. Matrices $\mathbf{D}^{\mathrm{nc}}$ and $\mathbf{D}^{\mathrm{oa}}$ are $2 \times 2$ covariance matrices characterizing variation due to among-subject sources for NC and OA group respectively. Within-subject deviations are characterized by vector $\mathbf{e}_i \sim \mathcal{N}_{m_i}(\mathbf{0}, \mathbf{R}_i)$. Here, $\mathbf{R}_i$ is a $m_i \times m_i$ matrix that characterizes variance and correlation due to within-subject sources. The parameters in (6.16) or (6.17) are estimated by maximum likelihood estimation.

The parameter of interest is $\boldsymbol{\beta}^{\mathrm{f,oa}}$, containing $\beta_4$, the difference of baseline thickness between OA and NC subjects, and $\beta_5$, the difference of thickness change rate between OA and NC subjects. As of so far, $\boldsymbol{\beta}^{\mathrm{f,oa}}$ is estimated using all OA subjects and the inconsistency of thinning locations across subjects would undermine the difference between NC and OA subjects. I therefore propose to cluster OA subjects to explore the subgroup information within them so that the subjects within a group have relative consistency in cartilage thinning.

Once the parameters are estimated, I perform clustering on the OA subjects based on their random effects, i.e, $\boldsymbol{\beta}_i^{\mathrm{r,oa}}$. As model (6.17) is solved at every pixel location (of interest), each OA subject is then associated with two random effects maps, i.e., $\beta_{6i}$ and $\beta_{7i}$. The two maps are vectorized and concatenated to compose a long vector, on which the following clustering is based. To account for different scales of the two maps, a normalization step is performed prior to the clustering, by dividing the estimated standard error in a pixel-wise way. I then apply the $k$-means algorithm (MacQueen, 1967) to cluster OA subjects into a number of groups. In the experiments, I choose 4 groups for each cartilage. Femoral and tibial cartilage are treated separately.

Each OA cluster is then studied separately versus NC subjects. Assume the OA clusters are indexed by $c$ ($1 \le c \le k$). Let $c = 0$ represent NC subjects. The delta function $\delta_i(c)$ for cluster $c$ is defined by

$$
\delta_i(c) = \begin{cases} 1 & \text{if subject } i \text{ belongs to cluster } c, \\ 0 & \text{otherwise.} \end{cases} \tag{6.18}
$$

The model for clustering-based analysis is

$$
\begin{aligned}
\mathbf{Y}_i(\mathbf{x}) = \mathbf{X}_i \begin{pmatrix} \beta_0(\mathbf{x}) \\ \beta_1(\mathbf{x}) \end{pmatrix} + \delta_i(0)\mathbf{X}_i \begin{pmatrix} \beta_{2i}(\mathbf{x}) \\ \beta_{3i}(\mathbf{x}) \end{pmatrix} \\
+ \sum_{c=1}^{k} \delta_i(c)\mathbf{X}_i \begin{pmatrix} \beta_{4,c}(\mathbf{x}) \\ \beta_{5,c}(\mathbf{x}) \end{pmatrix} + \sum_{c=1}^{k} \delta_i(c)\mathbf{X}_i \begin{pmatrix} \beta_{6i,c}(\mathbf{x}) \\ \beta_{7i,c}(\mathbf{x}) \end{pmatrix} + \mathbf{e}_i(\mathbf{x}).
\end{aligned} \tag{6.19}
$$

The difference of (6.19) from (6.16) is the OA subjects are divided into $k$ clusters and each cluster is studied separately against NC subjects. Parameters $\beta_{4,c}$ and $\beta_{5,c}$ characterize the difference of fixed effects (intercept and slope, respectively) of cluster $c$ comparing to NC group. Parameters $\beta_{6i,c}$ and $\beta_{7i,c}$ characterize the difference of random effects

(intercept and slope, respectively) of cluster $c$ comparing to NC group.

The simplified version of model (6.19) is

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}^{\text{f,nc}} + \delta_i(0)\mathbf{X}_i \boldsymbol{\beta}_i^{\text{r,nc}} + \sum_{c=1}^{k} \delta_i(c)\mathbf{X}_i \boldsymbol{\beta}_c^{\text{f,oa}} + \sum_{c=1}^{k} \delta_i(c)\mathbf{X}_i \boldsymbol{\beta}_{i,c}^{\text{r,oa}} + \mathbf{e}_i, \qquad (6.20)$$

where $\boldsymbol{\beta}_c^{\text{f,oa}} = (\beta_{4,c}, \beta_{5,c})^T$, and $\boldsymbol{\beta}_c^{\text{r,oa}} = (\beta_{6i,c}, \beta_{7i,c})^T \sim \mathcal{N}(\mathbf{0}, \mathbf{D}^{\text{oa}})$. Matrices $\mathbf{D}^{\text{nc}}$ and $\mathbf{D}^{\text{oa}}$ are $2 \times 2$ covariance matrices characterizing variation due to among-subject sources for NC and OA group resectively. Again, the parameters in (6.19) or (6.20) are estimated by maximum likelihood estimation.

Since the OA subjects in the same cluster exhibit similar cartilage thinning patterns, it is expected to observe more significant difference after clustering ($\beta_{4,c}$ and $\beta_{5,c}$) than before clustering ($\beta_4$ and $\beta_5$).

## 6.5 Experimental results

This section discusses the results of applying the proposed clustering-based thickness analysis. Again, I use the PLS dataset, which is described in section 4.2. The 2D thickness maps are from 3 different sources, expert segmentations, automatic temporally-independent (chapter 3) and longitudinal (chapter 5) segmentations. The analysis is performed only in masked regions. The masks (one for each cartilage) contain pixels at which $\geq 680$ out of 706 images have non-zero thickness values.

As femoral and tibial cartilage are studied independently, I will discuss each cartilage separately and start with the femoral cartilage.

### 6.5.1 Results of femoral cartilage

Let us consider the estimates of $\beta_0$, $\beta_1$, $\beta_4$ and $\beta_5$, and their $p$-values. Note for all $p$-values presented in this section, the false discover rate (FDR) adjustment method (Yekutieli and Benjamini, 1999) is employed to adjust $p$-values for multiple comparisons.

First of all, let us look at the parameter estimation before clustering. Figure 6.3 shows parameter estimation of (6.16) before clustering for femoral cartilage using thickness maps obtained from expert segmentations. Figures 6.4 and 6.5 show the results from temporally-independent and longitudinal segmentations, respectively. The estimates of $\beta_0$ from different segmentation methods are similar and the significance levels are also similar. No significant longitudinal change rate for the NC subjects, $\beta_1$, is observed in Figures 6.3 or 6.4. But significant thinning of NC subjects is observed from Fig. 6.5 which uses longitudinal segmentations. All of the three figures (6.3, 6.4 and 6.5) show significant difference of baseline thickness between NC and OA subjects. No significant difference of change rates between NC and OA subjects, $\beta_5$, is observed in any one of the three figures.

Now let us consider the parameter estimation after clustering. Figure 6.6 shows estimates of parameter $\beta_{4,c}$ of (6.19) after clustering for femoral cartilage using thickness maps obtained from expert segmentations. Figures 6.7 and 6.8 show the results from temporally-independent and longitudinal segmentations, respectively. From these three figures (6.6, 6.7 and 6.8), the first two clusters ($\beta_{4,1}$ and $\beta_{4,2}$) show significant difference (thinning and thickening) of baseline thickness of OA subjects comparing to NC subjects. One of the reasons for the thickening may be due to the swelling that OA patients suffer from.

The other two clusters ($\beta_{4,3}$ and $\beta_{4,4}$) exhibit distinct patterns. Cluster 3 shows cartilage loss on the medial side of the femoral cartilage and cluster 4 on the lateral side. This observation is consistent across different segmentation methods. The significance

levels are much higher than those before clustering. This suggests that cartilage loss may happen at different locations across OA subjects and the clustering helps to divide patients with similar thinning patterns into the same group.

The longitudinal change rates are not significant in any of the clusters for expert or temporally-independent segmentation. Only one cluster using longitudinal segmentation demonstrates significant difference of changing rate compared to NC subjects (see Fig. 6.9(a) and (b)).

### 6.5.2 Results of tibial cartilage

Figures 6.10, 6.11 and 6.12 present parameter estimation results before clustering. The estimates of $\beta_0$ and the $p$-values are in agreement across segmentations. Statistically significant thinning of NC subjects ($\beta_1$) is detected from expert and longitudinal segmentations, but not the temporally-independent segmentation. Comparing to NC subjects, OA subjects have statistically significant smaller baseline thickness ($\beta_4$), which is observed from all three segmentations. No significant difference of changing rate between NC and OA groups is seen from any of the three segmentations.

After clustering, the estimated difference of baseline thickness ($\beta_{4,c}$) for each OA cluster against NC are shown in Fig. 6.13, 6.14 and 6.15 from expert, temporally-independent and longitudinal segmentations, respectively. The results are comparable across segmentation. Cluster 1 show significant thickening on the medial side and cluster 2 is significantly thicker on the lateral side. Cluster 3 shows significant thinning on the lateral side and cluster 4 has significant thinning on both sides. Clustering is beneficial for detecting thinning (thickening) regions as more regions are identified with significance which were not discovered before clustering. The higher significance levels suggest the diversity of thickness changes as well as the hidden subgroup structure among OA subjects.

(a) Estimates of $\beta_0$

(b) $p$-values of $\beta_0$

(c) Estimates of $\beta_1$

(d) $p$-values of $\beta_1$

(e) Estimates of $\beta_4$

(f) $p$-values of $\beta_4$

(g) Estimates of $\beta_5$
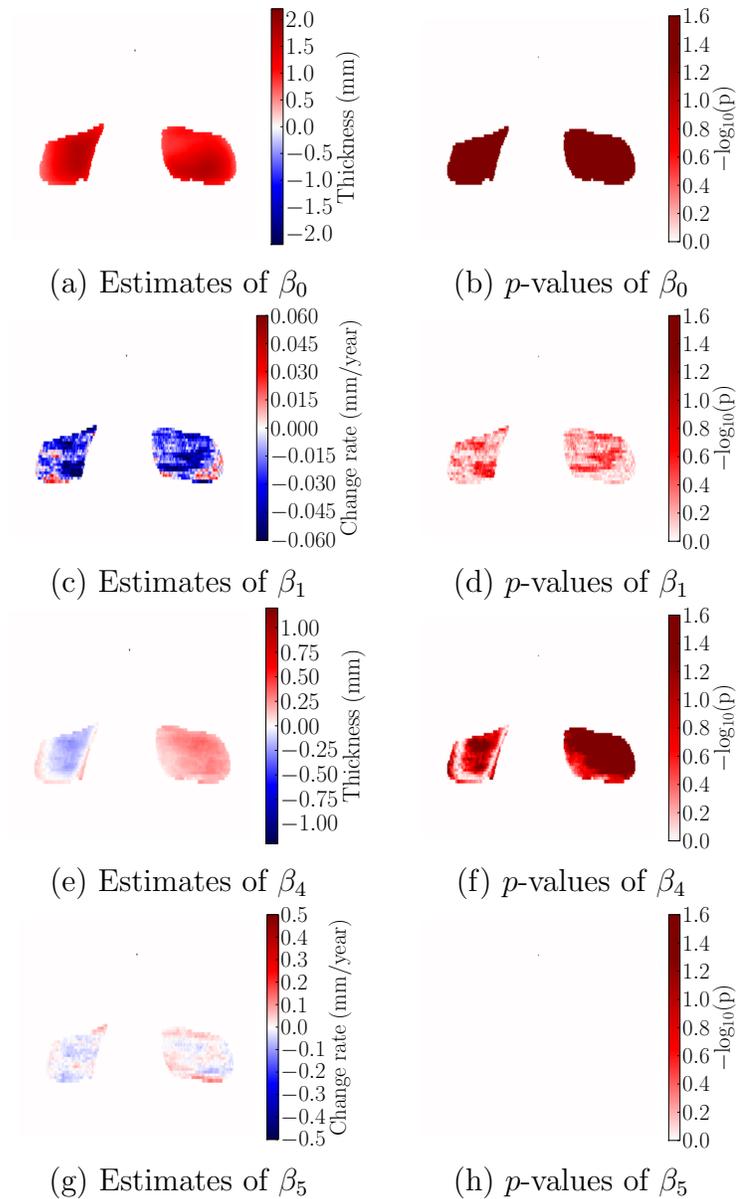
(h) $p$-values of $\beta_5$

Figure 6.3: Thickness analysis result (before clustering) of femoral cartilage comparing OA against NC subjects. Thickness maps are from expert segmentations. Left column: estimates of parameters. Right column: FDR adjusted $p$-values for testing null hypothesis that the corresponding parameter is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_0$      (b) $p$-values of $\beta_0$

(c) Estimates of $\beta_1$      (d) $p$-values of $\beta_1$

(e) Estimates of $\beta_4$      (f) $p$-values of $\beta_4$

(g) Estimates of $\beta_5$      (h) $p$-value of $\beta_5$
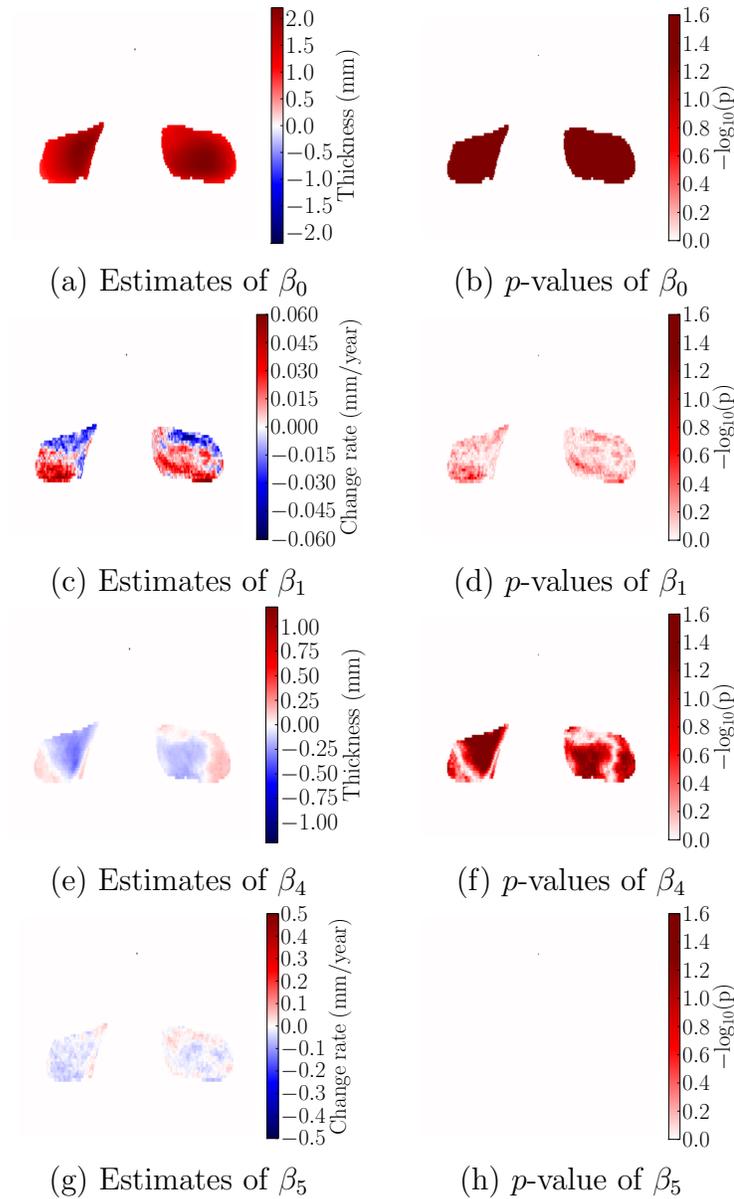
Figure 6.4: Thickness analysis result (before clustering) of femoral cartilage comparing OA against NC subjects. Thickness maps are from automatic temporally-independent segmentations. Left column: estimates of parameters. Right column: FDR adjusted $p$-values for testing null hypothesis that the corresponding parameter is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_0$      (b) $p$-values of $\beta_0$

(c) Estimates of $\beta_1$      (d) $p$-values of $\beta_1$

(e) Estimates of $\beta_4$      (f) $p$-values of $\beta_4$

(g) Estimates of $\beta_5$      (h) $p$-values of $\beta_5$
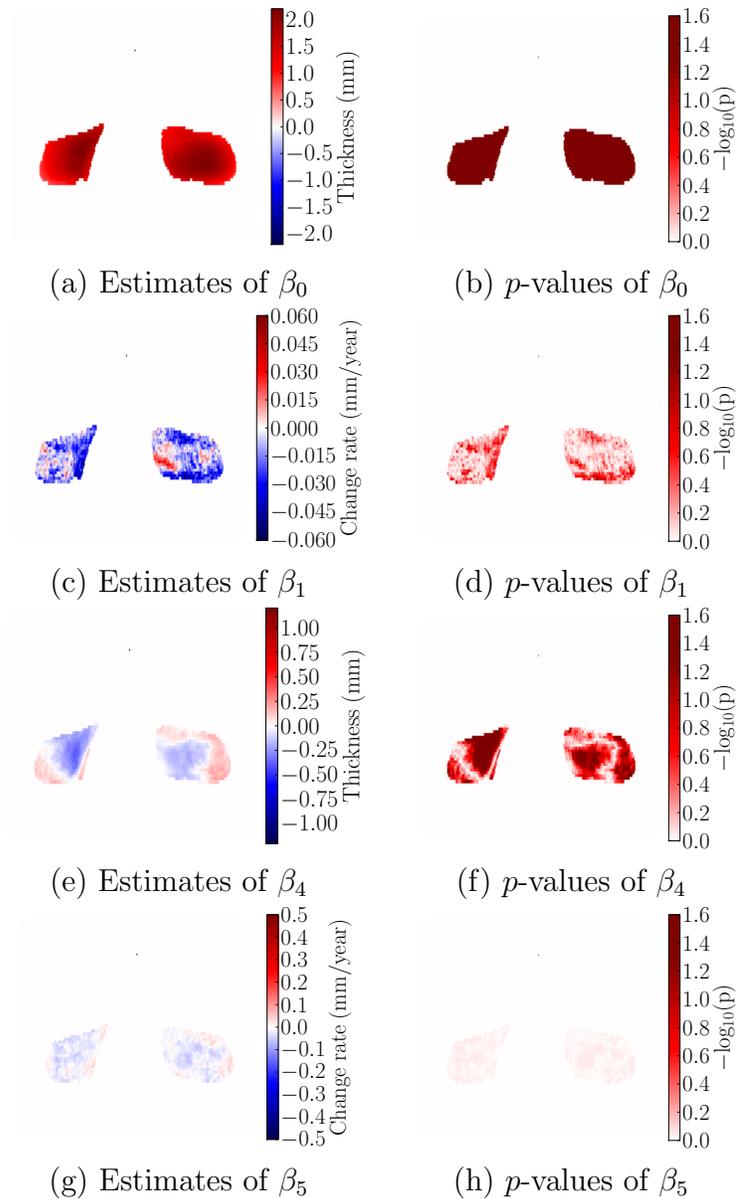
Figure 6.5: Thickness analysis result (before clustering) of femoral cartilage comparing OA against NC subjects. Thickness maps are from automatic longitudinal segmentations. Left column: estimates of parameters. Right column: FDR adjusted $p$-values for testing null hypothesis that the corresponding parameter is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_{4,1}$       (b) $p$-values of $\beta_{4,1}$

(c) Estimates of $\beta_{4,2}$       (d) $p$-values of $\beta_{4,2}$

(e) Estimates of $\beta_{4,3}$       (f) $p$-values of $\beta_{4,3}$

(g) Estimates of $\beta_{4,4}$       (h) $p$-values of $\beta_{4,4}$
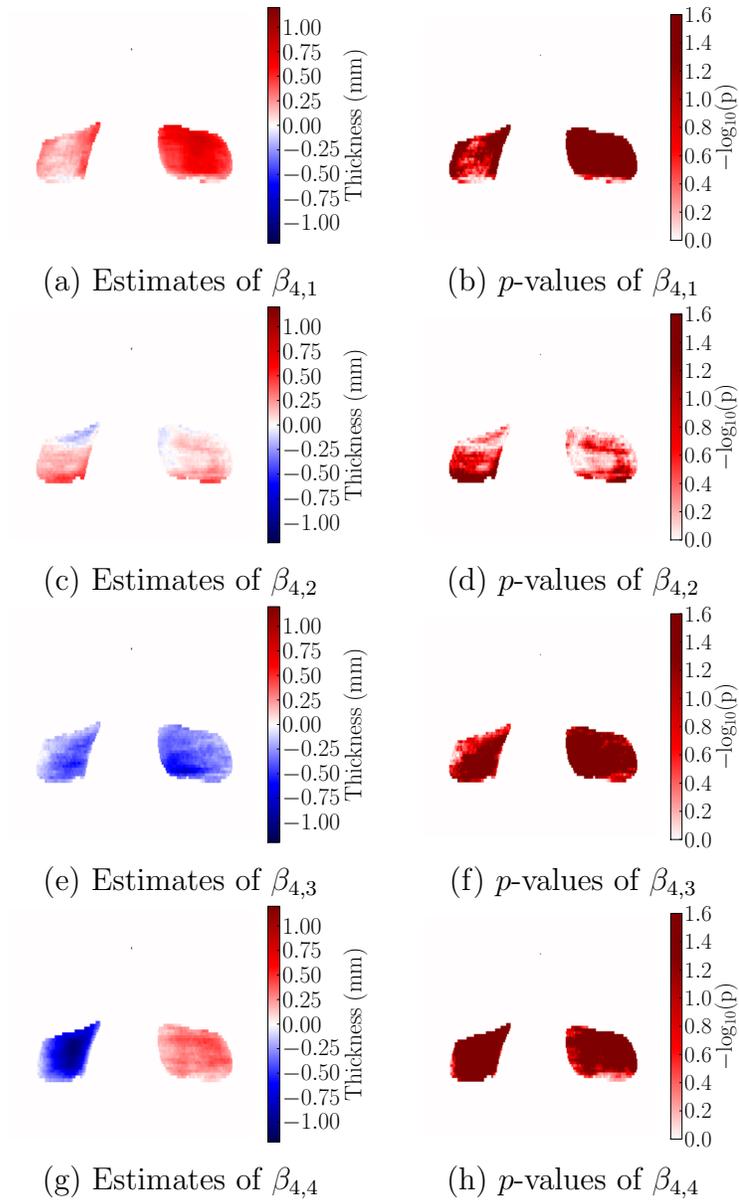
Figure 6.6: Thickness analysis result (after clustering), $\beta_{4,c}$, of femoral cartilage comparing OA against NC subjects. Thickness maps are from expert segmentations. Left column: estimates of parameter $\beta_{4,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{4,c}$ is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_{4,1}$       (b) $p$-values of $\beta_{4,1}$

(c) Estimates of $\beta_{4,2}$       (d) $p$-values of $\beta_{4,2}$

(e) Estimates of $\beta_{4,3}$       (f) $p$-values of $\beta_{4,3}$

(g) Estimates of $\beta_{4,4}$       (h) $p$-values of $\beta_{4,4}$
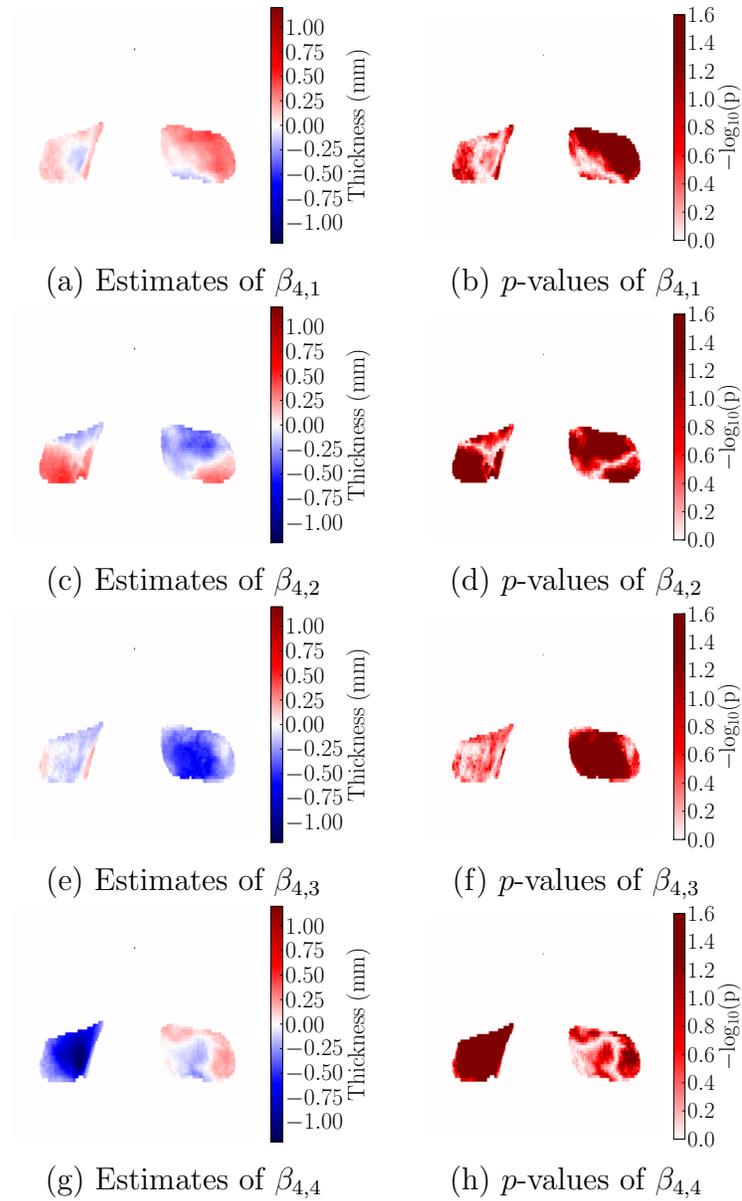
Figure 6.7: Thickness analysis result (after clustering), $\beta_{4,c}$, of femoral cartilage comparing OA against NC subjects. Thickness maps are from automatic temporally-independent segmentations. Left column: estimates of parameter $\beta_{4,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{4,c}$ is zero. $-log_{10}0.05 \approx 1.3$.
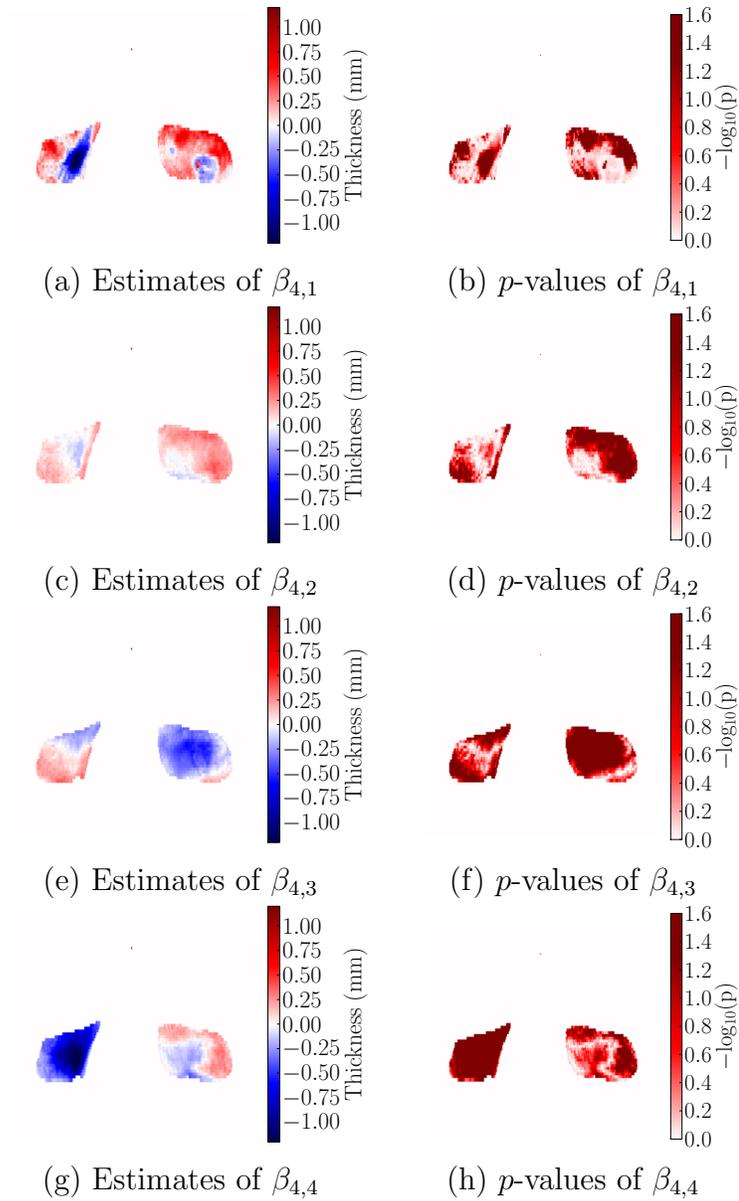
Figure 6.8: Thickness analysis result (after clustering), $\beta_{4,c}$, of femoral cartilage comparing OA against NC subjects. Thickness maps are from automatic longitudinal segmentations. Left column: estimates of parameter $\beta_{4,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{4,c}$ is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_{5,1}$      (b) $p$-values of $\beta_{5,1}$
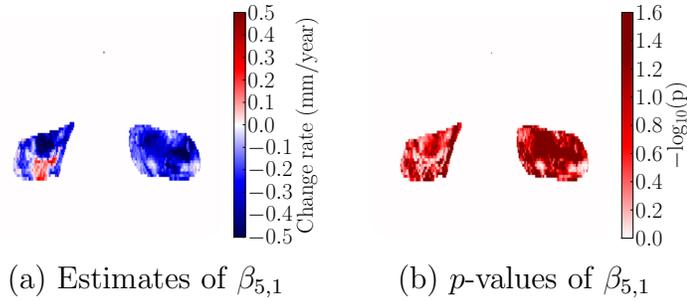
Figure 6.9: Thickness analysis result (after clustering), $\beta_{5,c}$, of femoral cartilage comparing OA against NC subjects. Thickness maps are from automatic longitudinal segmentations. Left column: estimates of parameter $\beta_{5,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{5,c}$ is zero. $-log_{10}0.05 \approx 1.3$.

Only a few small regions exhibit significant change rate in cluster 3 of expert segmentations and cluster 2 of longitudinal segmentations. See Fig. 6.16 and 6.17 for illustration.

### 6.5.3 Comparison to regional analysis

The results show significant thickening at each cartilage for certain clusters. The thickening is consistent with cartilage swelling or hypertrophy observed as a sign of early OA (Watson *et al.*, 1996; Calvo *et al.*, 2001, 2004; Eckstein *et al.*, 2011). Buck *et al.* (2010) suggested that OA may not be a one-way-road of cartilage loss and that particularly in early OA, cartilage changes may occur in both directions simultaneously, i.e., cartilage thinning and cartilage thickening.

Regional analysis was able to detect significant change rate in the central region and some peripheral regions as well (Wirth *et al.*, 2009; Buck *et al.*, 2009; Eckstein *et al.*, 2011). However, the majority of locations show insignificant change rate with the proposed method. One possible explanation is the difference in datasets. Wirth *et al.* (2009); Eckstein *et al.* (2011) used a subset from the OAI data. I used the PLS dataset instead.

Another possible reason is the difference in number of time points used. Existing regional analysis results were obtained from only two time points. For example, Buck

(a) Estimates of $\beta_0$        (b) $p$-values of $\beta_0$

(c) Estimates of $\beta_1$        (d) $p$-values of $\beta_1$

(e) Estimates of $\beta_4$        (f) $p$-values of $\beta_4$

(g) Estimates of $\beta_5$        (h) $p$-values of $\beta_5$
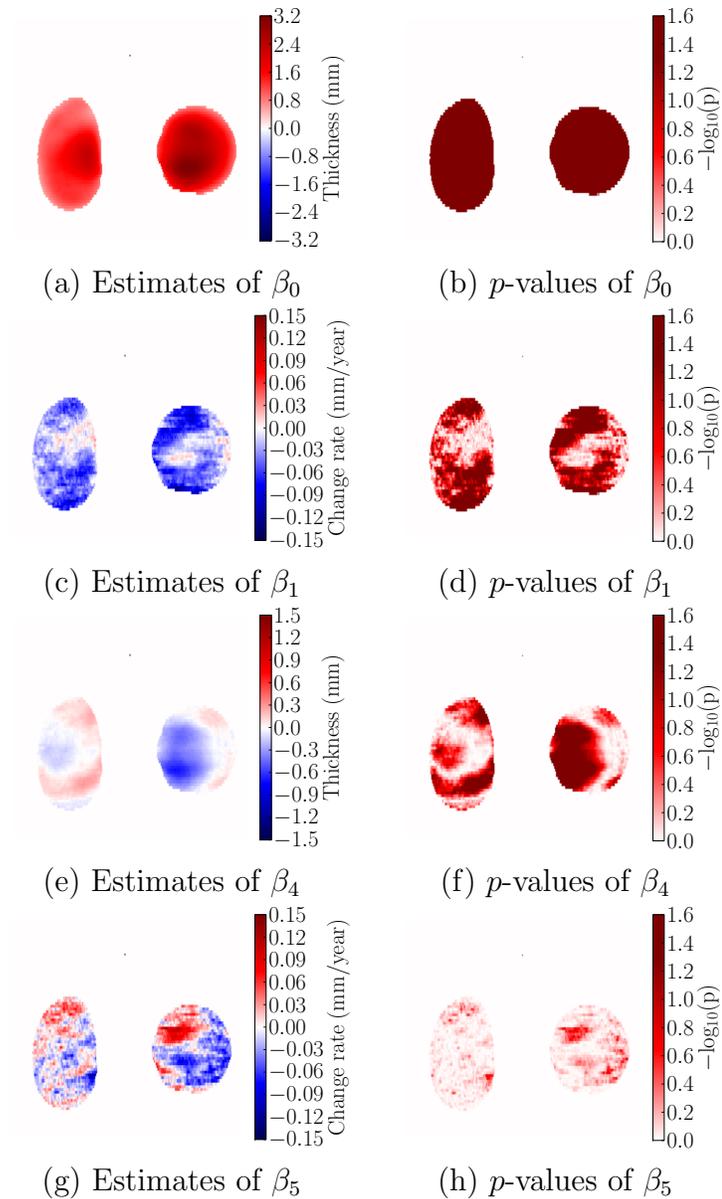
Figure 6.10: Thickness analysis result (before clustering) of tibial cartilage comparing OA against NC subjects. Thickness maps are from expert segmentations. Left column: estimates of parameters. Right column: FDR adjusted $p$-values for testing null hypothesis that the corresponding parameter is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_0$      (b) $p$-values of $\beta_0$

(c) Estimates of $\beta_1$      (d) $p$-values of $\beta_1$

(e) Estimates of $\beta_4$      (f) $p$-values of $\beta_4$

(g) Estimates of $\beta_5$      (h) $p$-values of $\beta_5$
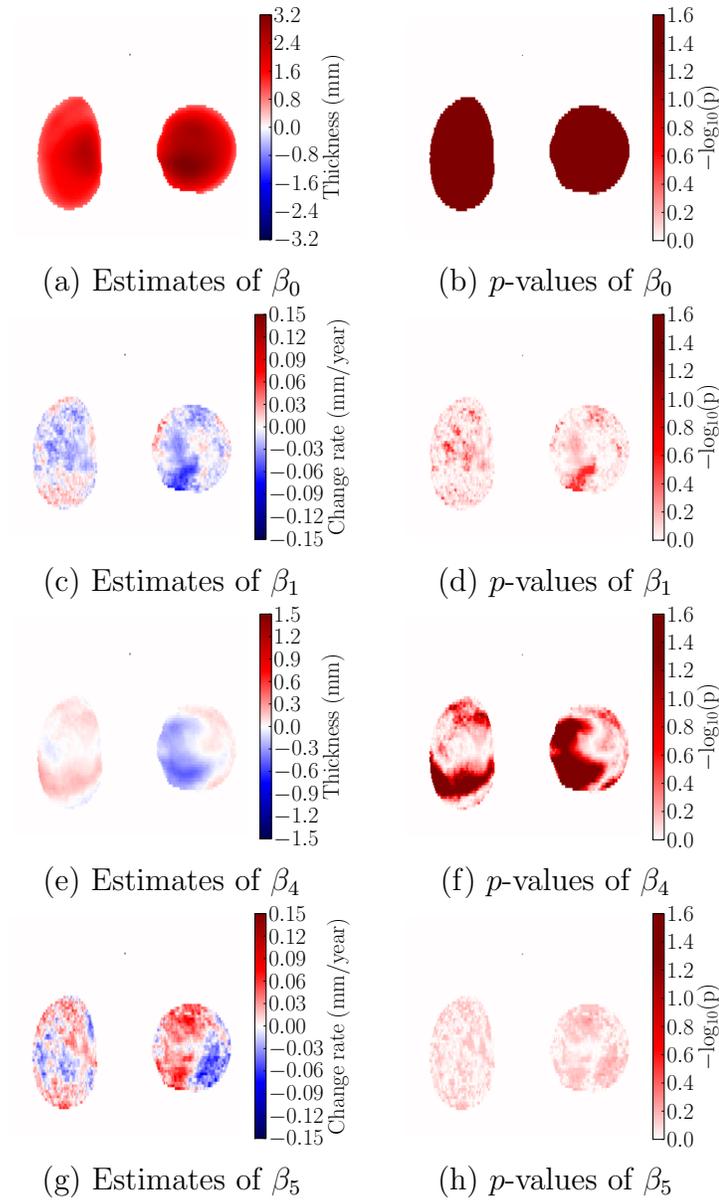
Figure 6.11: Thickness analysis result (before clustering) of tibial cartilage comparing OA against NC subjects. Thickness maps are from automatic temporally-independent segmentations. Left column: estimates of parameters. Right column: FDR adjusted $p$-values for testing null hypothesis that the corresponding parameter is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_0$      (b) $p$-values of $\beta_0$

(c) Estimates of $\beta_1$      (d) $p$-values of $\beta_1$

(e) Estimates of $\beta_4$      (f) $p$-values of $\beta_4$

(g) Estimates of $\beta_5$      (h) $p$-values of $\beta_5$
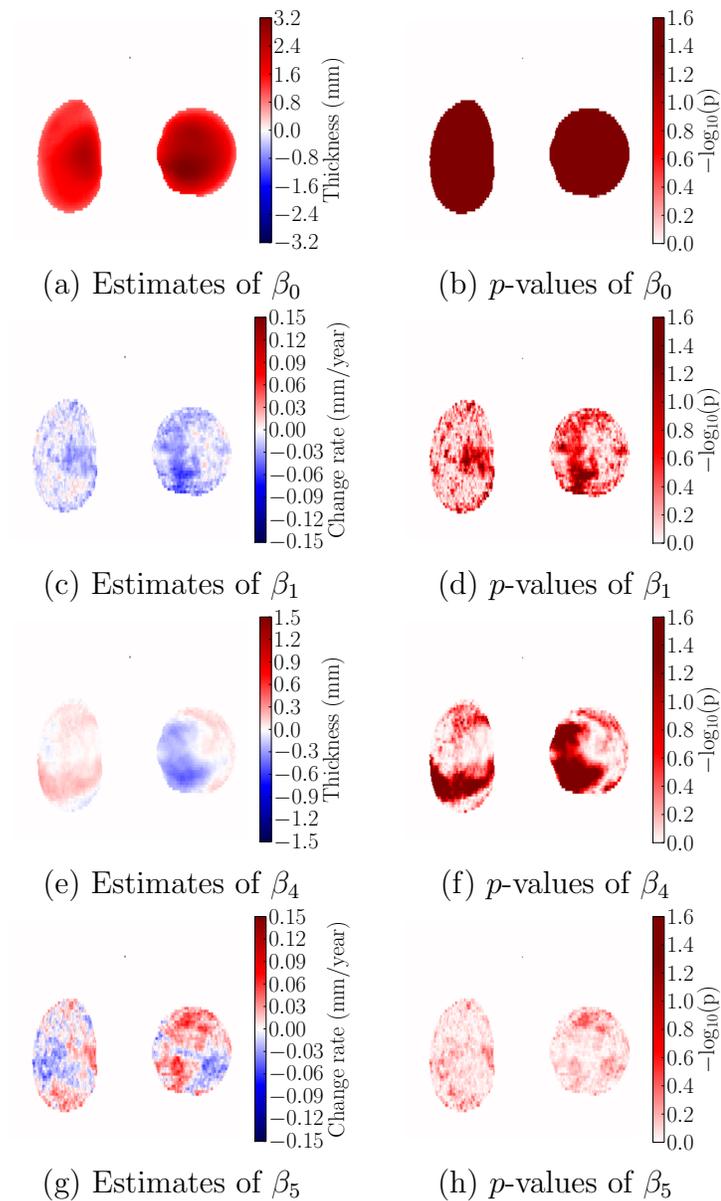
Figure 6.12: **Thickness analysis result (before clustering) of tibial cartilage comparing OA against NC subjects. Thickness maps are from automatic longitudinal segmentations.** Left column: estimates of parameters. Right column: FDR adjusted $p$-values for testing null hypothesis that the corresponding parameter is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_{4,1}$      (b) $p$-values of $\beta_{4,1}$

(c) Estimates of $\beta_{4,2}$      (d) $p$-values of $\beta_{4,2}$

(e) Estimates of $\beta_{4,3}$      (f) $p$-values of $\beta_{4,3}$

(g) Estimates of $\beta_{4,4}$      (h) $p$-values of $\beta_{4,4}$
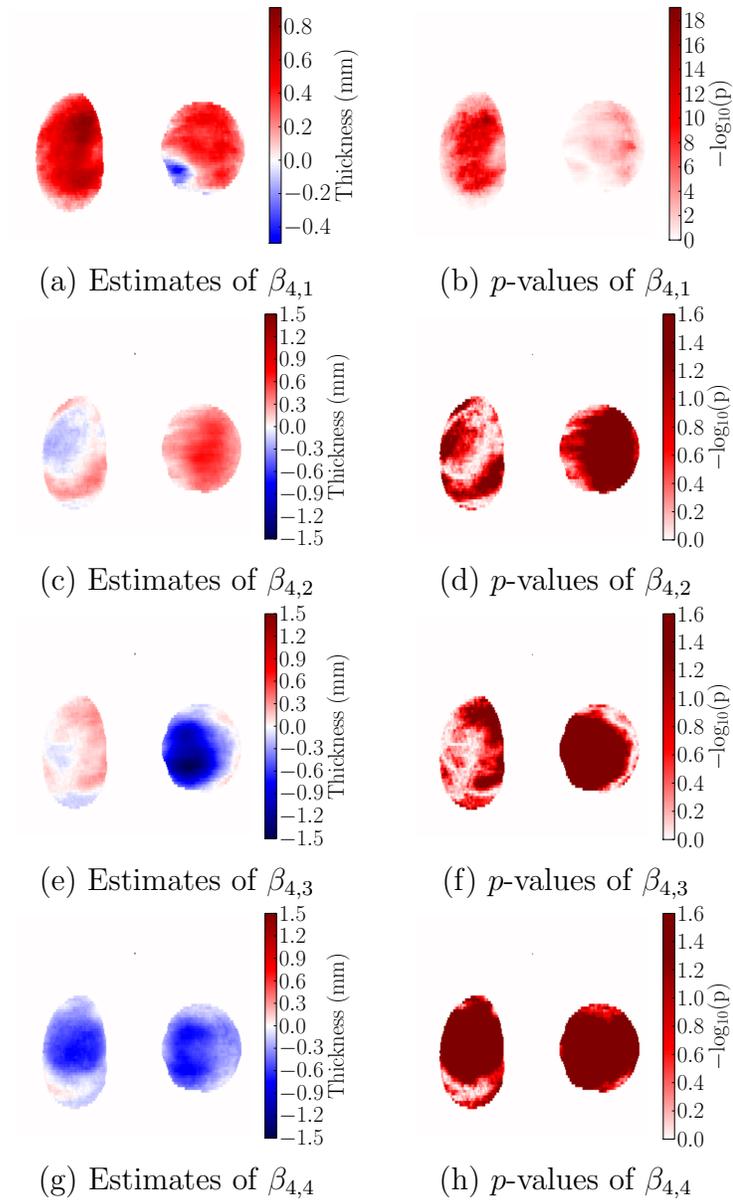
Figure 6.13: Thickness analysis result (after clustering), $\beta_{4,c}$, of tibial cartilage comparing OA against NC subjects. Thickness maps are from expert segmentations. Left column: estimates of parameter $\beta_{4,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{4,c}$ is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_{4,1}$      (b) $p$-values of $\beta_{4,1}$

(c) Estimates of $\beta_{4,2}$      (d) $p$-values of $\beta_{4,2}$

(e) Estimates of $\beta_{4,3}$      (f) $p$-values of $\beta_{4,3}$

(g) Estimates of $\beta_{4,4}$      (h) $p$-values of $\beta_{4,4}$
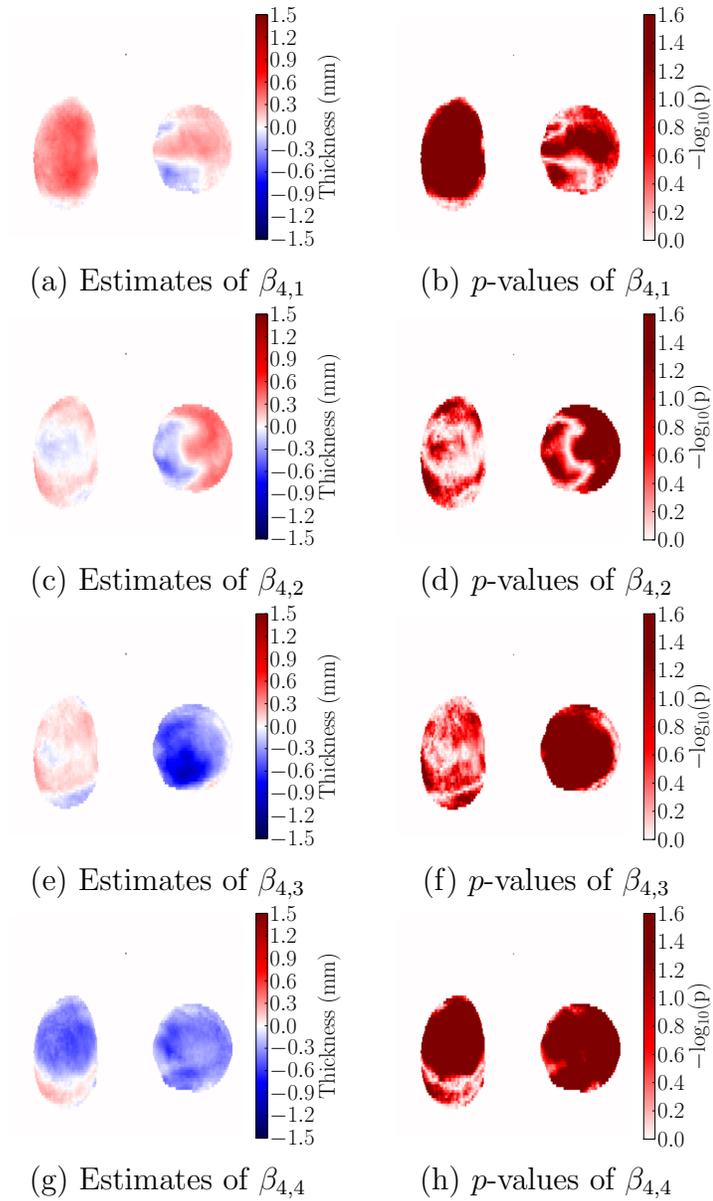
Figure 6.14: Thickness analysis result (after clustering), $\beta_{4,c}$, of tibial cartilage comparing OA against NC subjects. Thickness maps are from automatic temporally-independent segmentations. Left column: estimates of parameter $\beta_{4,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{4,c}$ is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_{4,1}$

(b) $p$-values of $\beta_{4,1}$

(c) Estimates of $\beta_{4,2}$

(d) $p$-values of $\beta_{4,2}$

(e) Estimates of $\beta_{4,3}$

(f) $p$-values of $\beta_{4,3}$

(g) Estimates of $\beta_{4,4}$

(h) $p$-values of $\beta_{4,4}$
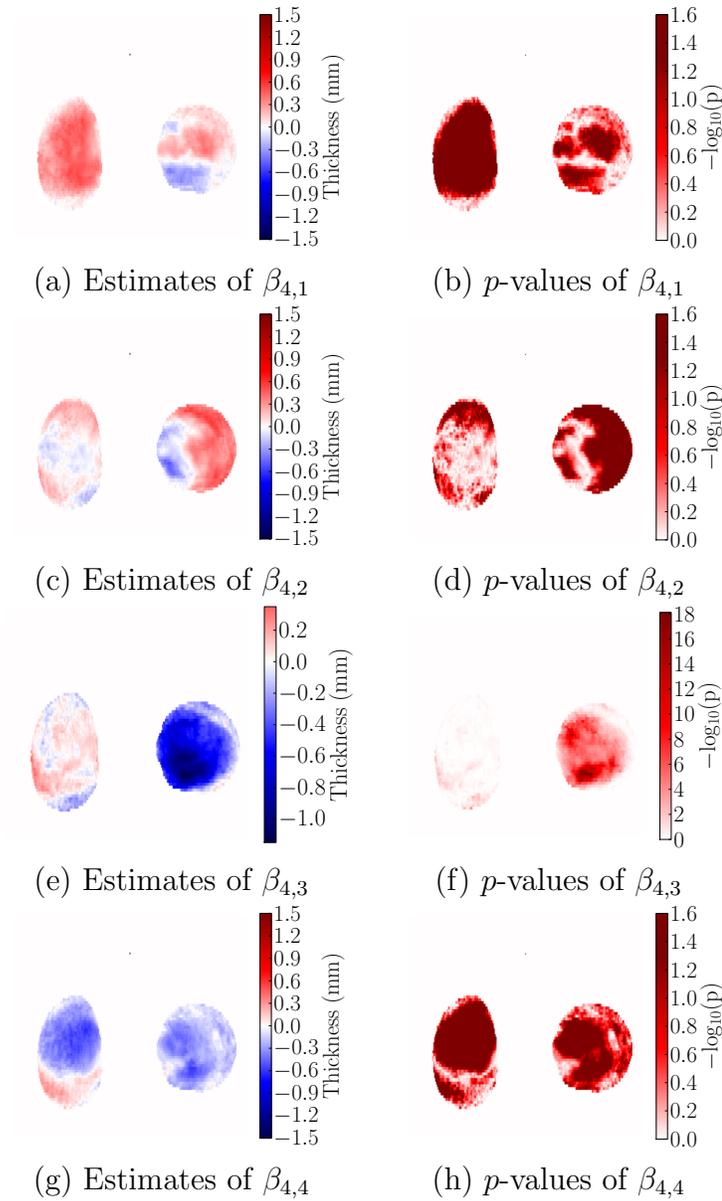
Figure 6.15: Thickness analysis result (after clustering), $\beta_{4,c}$, of tibial cartilage comparing OA against NC subjects. Thickness maps are from automatic longitudinal segmentations. Left column: estimates of parameter $\beta_{4,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{4,c}$ is zero. $-log_{10}0.05 \approx 1.3$.

(a) Estimates of $\beta_{5,3}$       (b) $p$-values of $\beta_{5,3}$
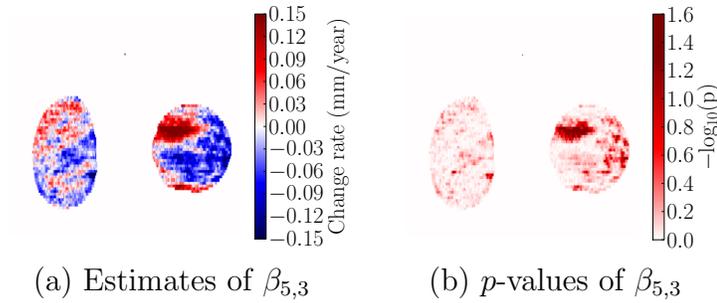
Figure 6.16: Thickness analysis result (after clustering), $\beta_{5,c}$, of tibial cartilage comparing OA against NC subjects. Thickness maps are from expert segmentations. Left column: estimates of parameter $\beta_{5,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{5,c}$ is zero. $-log_{10}0.05 \approx 1.3$.



(a) Estimates of $\beta_{5,2}$       (b) $p$-values of $\beta_{5,2}$
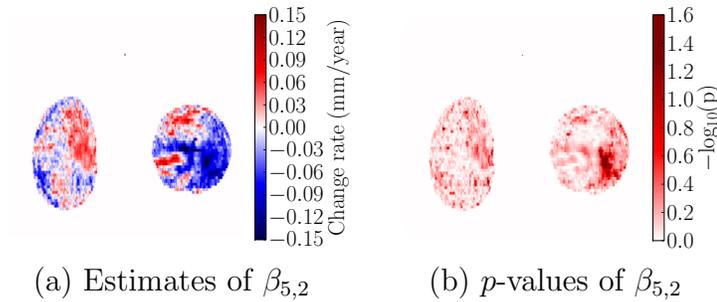
Figure 6.17: Thickness analysis result (after clustering), $\beta_{5,c}$, of tibial cartilage comparing OA against NC subjects. Thickness maps are from automatic longitudinal segmentations. Left column: estimates of parameter $\beta_{5,c}$. Right column: FDR adjusted $p$-values for testing null hypothesis that $\beta_{5,c}$ is zero. $-log_{10}0.05 \approx 1.3$.

*et al.* (2009) used baseline and 2-year followup, and Wirth *et al.* (2009) used baseline and 1-year followup. I used 5 time points at baseline, 3, 6, 12 and 24 months. The linear model may not be the best model to describe how thickness change over time. It is possible that cartilage thinning progresses slowly at first and rapidly later, in which case, including 3-month and 6-month data would cause underestimation of the change rate. Using baseline and 2-year followup only might result in a more statistically significant change rate.

It is also reasonable that the insignificant change rates are due to the limited baseline thickness for degeneration. The significant change rates are only found at locations with large baseline thickness. See Fig. 6.9, 6.16 and 6.17.

Even though fewer regions are found with significant change rate, the proposed method is able to uncover change rates at large magnitudes. The change rates reported in Buck *et al.* (2009) and Wirth *et al.* (2009) are far smaller than image resolution. For example, Buck *et al.* (2009) estimated the difference of change rate between KLG0 and KLG3 to be 0.061 $mm$/year with $p$-value 0.001 from baseline and 2-year followup. This change rate means the difference is 0.12 $mm$ in 2 years which is far smaller compared to image resolution ($0.3 \times 0.3 \times 1.0\ mm$). The proposed approach detects significant change rate at $\geq 0.15\ mm$/year, which translates to 1 pixel thinner over 2 years. This is a more reasonable rate for popular MRI resolutions in knee imaging. The increased magnitudes of change rates suggest that cartilage loss is a local process and regional average analysis is too coarse for subtle changes.

## 6.6   Conclusion

This chapter presents a method to automatically analyze localized cartilage thickness changes given cartilage segmentations. The method first establishes spatial correspondence across subjects and time via registration. The following analysis is pixel-wise

therefore avoids subdivision of the cartilage region. Additionally, due to the heterogeneous nature of cartilage thinning, the method clusters OA subjects into several groups according to their thinning patterns. Each cluster is then studied separately against NC subjects. The analysis results show that clustering improves the significance levels of the difference between healthy and diseased group and also helps to detect more thinning locations that would not be detected without.

# CHAPTER 7: DISCUSSION

## 7.1 Summary of contributions

In this section the thesis statement and contributions presented in chapter 1 are revisited. Following each claim is a discussion that relates to the claim and summarizes how that claim is addressed in this dissertation.

The contributions are as follows:

1. *I present a novel and general three-label segmentation method that ensures the separation of touching objects and allows for spatial and temporal regularization. The three-label segmentation is formulated as a convex optimization problem, which allows for the computation of global optimal solutions. The method can be applied to other segmentation problems with two touching objects.*

I proposed a three-label segmentation framework in section 2.3. The method is general and guarantees separation of the touching objects. By placing the background label in the middle, one can obtain a symmetric labeling with respect to the background. The anisotropic regularization has general applicability for the segmentation of thin objects, e.g., knee cartilage. An important advantage of the segmentation framework is its convexity, which guarantees that a globally optimal solution to the relaxed can be computed, which then yields a globally optimal solution to the original problem after thresholding due to its essentially binary property.

Later in section 2.4, I proposed a multi-label segmentation framework to overcome the regularization bias of the three-label segmentation when applied to more than three labels (two objects, one label is for the background). Again, the relaxed formulation is

convex and one can compute the globally optimal solution. Unfortunately, the "essentially binary" property is not guaranteed in this case. Therefore the thresholded solution may not be a globally optimal solution to the discrete formulation, which is a Potts model and known to be NP-hard. However, the solution can be a close approximation to the real optimal solution as the indicator function is quite binary after convergence in practice.

A novel longitudinal three-label segmentation approach was proposed in section 5.2 to encourage the temporal consistency of the segmentation of longitudinal data. The approach is also general and can be applied to other longitudinal segmentation problems with two objects.

2. *I develop a new fully automatic three-label cartilage segmentation pipeline based on multi-atlas registration and local tissue classification.*

The fully automatic cartilage segmentation pipeline was proposed in chapter 3. I used a multi-atlas-based bone segmentation (described in section 3.3) to guide the registration of a cartilage atlas. I obtained cartilage segmentation using an average shape atlas (section 3.5) or multiple atlases with various label fusion techniques (section 3.6) to obtain spatial cartilage priors within the three-label segmentation framework which incorporates anisotropic regularization to improve segmentation performance (shown in chapter 4) for the thin femoral and tibial cartilage layers.

For longitudinal knee image datasets, e.g., the PLS dataset, I proposed in section 5.3 a pipeline to register longitudinal images from native image space to the common (longitudinal) space to establish correspondence across time for each individual subject, so that the longitudinal segmentation can be applied to encourage temporary consistency.

The overall pipeline is fully-automatic (besides quality control), which enables the method to be applied to large image databases. The robustness due to multi-atlas-based strategies also make the proposed method appropriate for large datasets.

*3. I validate the automatic three-label cartilage segmentation on a sizable dataset consisting of more than 700 images. Specially, I study the impacts of different types of atlases (namely average-shape-atlas and multi-atlas) and different types of regularization (i.e., isotropic spatial regularization, anisotropic spatial regularization and temporal regularization) on cartilage segmentation accuracy.*

Extensive validation of the proposed approach was presented in sections 4.3 and 4.4. It can be concluded that best segmentation strategy is 1) multi-atlas-based segmentation with patch-based label fusion for spatial prior, 2) SVM for local likelihood and 3) three-label segmentation with anisotropic regularization. I also demonstrated that the proposed segmentation strategy performs as well as the top-ranking methods on the SKI10 dataset in section 4.5.

I demonstrated the improvement in temporal consistency using the longitudinal segmentation over the temporally-independent segmentation in section 5.4.

*4. I develop a method to establish spatial correspondences of knee cartilage across time points and between subjects, which allows for statistical analysis on localized cartilage thickness changes.*

I proposed in section 6.3 an automatic analysis method establishing a common (atlas) spatial coordinate system for cartilage analysis of the knee. The spatial correspondence enables localized analysis of thickness changes.

*5. I propose a novel method to analyze nonuniform localized cartilage changes, which can be applied to study nonuniform local morphological changes in other diseases.*

Due to the fact that cartilage thinning may happen to different locations for different subjects, some subjects exhibit similar thinning patterns than other subjects. I proposed a new clustering-based statistical method to analyze cartilage thickness changes in section 6.4. The OA subjects were clustered into a few groups and those within the same group have similar thinning patterns. Each cluster is then studied separately comparing

to NC subjects.

6. *I perform statistical analysis on the PLS dataset. The statistical analysis result of localized cartilage changes is presented and compared to that reported in literature using regional analysis.*

The analysis results shown in section 6.5 demonstrate that clustering improves the significance levels of the difference between healthy and diseased group and also helps to detect more thinning locations that would not be detected without. The proposed method also resulted in detection of change rates at larger magnitudes than conventional subregion-based analysis.

Having fulfilled the contributions above, I present again the thesis statement.

**Thesis**: *Automatic, robust and accurate cartilage segmentations can be obtained through multi-atlas-based registration and local tissue classification within a three-label segmentation allowing for spatial and temporal regularization. Spatially transforming cartilage thickness maps into an atlas space enables statistical analysis on localized cartilage changes. Clustering of OA subjects improves statistical analysis due to the spatial heterogeneity of cartilage loss.*

## 7.2 Future work

There are a number of questions and directions for future research related to this dissertation. In this section a collection of these topics is briefly reviewed.

**Application to the OAI dataset**

The objective of this dissertation is to develop an automatic and robust method to process large image datasets. While the PLS is a sizable dataset, the OAI dataset is far larger, containing longitudinal images from 4,796 subjects. There are 88 knees who have been manually segmented at baseline and 12-month visits by iMorphics that are publicly

available. The proposed method could first be validated on the delineated images. As expert segmentations are not available for most of the images in OAI dataset, it would be of tremendous value to apply a fully automatic and well validated method (on the PLS and partially OAI datasets) to obtain cartilage segmentations and perform statistical analysis on longitudinal cartilage thickness changes. Without an automatic method, analysis of the shear amount of data would be difficult.

**Atlas selection**

The major drawback of the proposed cartilage segmentation 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 (Aljabar *et al.*, 2009). Such a selection strategy can be integrated into the segmentation pipeline and is expected to further improve segmentation performance. I will explore atlas selection for cartilage segmentation in future work.

**Local likelihoods for the bones**

Currently, the bone likelihoods are estimated from a simple model (3.2) based on the fact that bones appear dark in T1 weighted MR images, which might not be the case in other imaging modalities, e.g., T2 weighted where the cortical bone appears dark and bone marrow appears bright. A more general method for the estimation of bone likelihoods would be similar to that for cartilage likelihoods, namely, probabilistic classification. Local features could be computed and then used to train a classier to distinguish bone and non-bone voxels in a probabilistic manner. This extension would make the entire segmentation pipeline applicable to other image modalities.

## Cartilage overlap avoidance in longitudinal space

The baseline image space is chosen as the common space to register longitudinal images of an individual because cartilage was thought to be thickest at the baseline visit. However, the following statistical analysis suggests that cartilage thickening could happen due to swelling and hypertrophy, which was also observed in previous literature (Watson *et al.*, 1996; Calvo *et al.*, 2001, 2004; Buck *et al.*, 2010). So the baseline image is no longer an ideal choice for the common longitudinal space as thicker cartilage of followup visits could possibly overlap in this space and it would result in under-segmentation of the followup visits. One potential and straightforward fix would be pulling femur and tibia apart in the baseline space to make room for the later possible thicker cartilage.

## Neighborhood dependence for statistical analysis

The statistical analysis of cartilage thickness was performed in a pixel-by-pixel manner. Neighboring pixels are treated as independent of each other. However, it is reasonable to assume dependency among neighboring pixels. A pixel surrounded by pixels most of which exhibit cartilage loss is more likely to have reduced thickness than a pixel whose neighboring pixels demonstrate no changes over time. Markov random fields could be used to incorporate neighborhood dependence.

## Correlation between femoral and tibial cartilage

OA subjects are clustered into subgroups based on estimated random effects of femoral cartilage thickness, and into a different set of subgroups based on estimated random effects of femoral cartilage thickness. The correlation (if there is any) between femoral and tibial cartilage is broken by different sets of clusters. In future work, I will experiment with clustering using both cartilages jointly so that possible correlations between them could be revealed.

**Including more patient information**

OA is a complex disease whose etiology is partially unclear. There is growing evidence for the role of systemic factors (such as genetics, dietary intake, estrogen use, and bone density) and of local biomechanical factors (such as muscle weakness, obesity, and joint laxity) (Felson *et al.*, 2000a). To understand the disease, statistical analysis could be conducted including aforementioned risk factors into statistical models.

# BIBLOGRAPHY

Adams, R. and Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern analysis and machine intelligence*, **16**(6), 641–647.

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

Angenent, S., Pichon, E., and Tannenbaum, A. (2006). Mathematical methods in medical image processing. *Bulletin of the American Mathematical Society*, **43**, 365–396.

Appleton, B. and Talbot, H. (2006). Globally mimimal surfaces by continuous maximal flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**(1), 106–118.

Bae, K. T., Shim, H., Tao, C., Chang, S., Wang, J. H., Boudreau, R., and Kwoh, C. K. (2009). Intra- and inter-observer reproducibility of volume measurement of knee cartilage segmented from the OAI MR image set using a novel semi-automated segmentation method. *Osteoarthritis Cartilage*, **17**, 1589–1597.

Beucher, S. and Meyer, F. (1993). The morphological approach to segmentation: The watershed transform. In *Mathematical Morphology in Image Processing*, volume 12, pages 433–481. New York: Marcel Dekker, E. R. Dougherty edition.

Biswal, S., Hastie, T., Andriacchi, T. P., Bergman, G. A., Dillingham, M. F., and Lang, P. (2002). Risk factors and progressive cartilage loss in the knee: A longitudinal magnetic resonance imaging study in forty-three patients. *Arthritis & Rheumatism*, **46**, 2884–2892.

Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern analysis and machine intelligence*, **26**(9), 1124–1137.

Brem, M. H., Brem, M. H., Lang, P. K., Neumann, G., Schlechtweg, P. M., Schneider, E., Jackson, R., Yu, J., Eaton, C. B., Hennig, F. F., Yoshioka, H., Pappas, G., and Duryea, J. (2009). Magnetic resonance image segmentation using semi-automated software for quantification of knee articular cartilage - initial evaluation of a technique for paired scans. *Skeletal Radiology*, **28**(5), 505–511.

Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.-P., and Osher, S. (2007). Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, **28**, 151–167.

Buck, R. J., Wyman, B. T., Graverrand, M.-P. L., Hudelmaier, M., Wirth, W., and Ecksten, F. (2009). Does the use of ordered values of subregional change in cartilage thickness improve the detection of disease progression in longitudinal studies of osteoarthritis? *Arthritis and Rheumatism*, **61**(7), 917–924.

Buck, R. J., Wyman, B. T., Graverand, M. P. L., Hudelmaier, M., Wirth, W., Eckstein, F., and 9001140-A investigators (2010). Osteoarthritis may not be a one-way-road of cartilage loss–comparison of spatial patterns of cartilage change between osteoarthritic and healthy knees. *Osteoarthritis Cartilage*, **18**(3), 329–35.

Calvo, E., Palacios, I., Delgado, E., Ruiz-Cabello, J., Hernndez, P., Snchez-Pernaute, O., Egido, J., and Herrero-Beaumont, G. (2001). High-resolution MRI detects cartilage swelling at the early stages of experimental osteoarthritis. *Osteoarthritis Cartilage*, **9**(5), 463–72.

Calvo, E., Palacios, I., Delgado, E., Snchez-Pernaute, O., Largo, R., Egido, J., and Herrero-Beaumont, G. (2004). Histopathological correlation of cartilage swelling detected by magnetic resonance imaging in early experimental osteoarthritis. *Osteoarthritis Cartilage*, **12**(11), 878–86.

Caselles, V., Kimmel, R., and Sapiro, G. (1997a). Geodesic active contours. *International Journal of Computer Vision*, **22**(11), 61–79.

Caselles, V., Kimmel, R., Sapiro, G., and Sbert, C. (1997b). Minimal surfaces based object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, 394–398.

CDC (2008). Arthritis related statistics, Center of Disease Control (CDC). `http://www.cdc.gov/arthritis/data_statistics/arthritis_related_statistics.htm`.

Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, **10**, 266–277.

Chang, C.-C. and 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`.

Cicuttini, F., Hankin, J., Jones, G., and Wluka, A. (2005). Comparison of conventional standing knee radiographs and magnetic resonance imaging in assessing progression of tibiofemoral joint osteoarthritis. *Oesteoarthritis Cartilage*, **13**, 722–727.

Cohen, L. D. (1991). On active contour models and balloons. *Computer Vision, Graphics, and Image Processing: Image Understanding*, **53**(2), 211–218.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape model - their training and application. *Computer Vision and Image Understanding*, **61**(1), 38–59.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(6), 681–685.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, **20**(3), 273–197.

Coupé, P., Manjȵn, J., Fonov, V., Prusessner, J., Robles, M., and Collins, D. (2011). Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage*, **59**(4), 3736–3747.

Dam, E. B. and Loog, M. (2008). Efficient segmentation by sparse pixel classification. *IEEE Transactions on Medical Imaging*, **27**(10), 1525–1534.

Delong, A. and Boykov, Y. (2008). A scalable graph-cut algorithm for N-D grids. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Dice, L. (1945). Measures of the amount of ecologic association between species. *Ecology*, pages 297–302.

Dodin, P., Pelletier, J. P., Martel-Pelletier, J., and Abram, F. (2010). Automatic human knee cartilage segmentation from 3-d magnetic resonance images. *IEEE Transactions on Biomedical Engineering*, **57**(11), 2699–2711.

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

Eckstein, F., Gavazzeni, A., Sittek, H., Haubner, M., Lösch, A., Milz, S., Englmeier, K. H., Schulte, E., Putz, R., and Reiser, M. (1996). Determination of knee joint cartilage thickness using three-dimensional magnetic resonance chondro-crassometry (3D MR-CCM). *Magnetic Resonance in Medicine*, **36**(2), 256–265.

Eckstein, F., Adam, C., Sittek, H., Becker, C., Milz, S., Schulte, E., and et al. (1997). Non-invasive determination of cartilage thickness throughout joint surfaces using magnetic resonance imaging. *Journal of Biomechanics*, **30**, 285–289.

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

Eckstein, F., Cicuttini, F., Raynauld, J.-P., Waterton, J. C., and Peterfy, C. (2006). Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): morphological assessment. *Osteoarthritis and Cartilage*, **14**, A46–A75.

Eckstein, F., Buck, R. J., Burstein, D., Charles, H. C., Crim, J., Hudelmaier, M., Hunter, D. J., Hutchins, G., Jackson, C., Kraus, V. B., Lane, N. E., Link, T. M., Majumdar, L. S., Mazzuca, S., Prasad, P. V., Schnitzer, T. J., Taljanovic, M. S., Vaz, A., Wyman, B., Graverand, M.-P. H. L., and on behalf of the A9001140 Study Group (2008). Precision of 3.0 tesla quantitative magnetic resonance imaging of cartilage morphology in a multicentre clinical trial. *Annals of the Rheumatic Diseases*, **67**, 1683–1688.

Eckstein, F., Hudelmaier, M., and Wirth, W. (2011). Quantitative MR imaging of cartilage morphology in osteoarthritis. *Cartilage Imaging: Significance, Techniques, and New Developments*.

Felson, D. T., Lawrence, R. C., Dieppe, P. A., Hirsch, R., Helmick, C. G., Jordan, J. M., Kington, R. S., Lane, N. E., Nevitt, M. C., Zhang, Y., Sowers, M., McAlindon, T., Spector, T. D., Poole, A. R., Yanovski, S. Z., Ateshian, G., Sharma, L., Buckwalter, J. A., Brandt, K. D., and Fries, J. F. (2000a). Osteoarthritis: new insights, part 1: the disease and risk factors. *Annals of Internal Medicine*, **133**(8), 635–646.

Felson, D. T., Lawrence, R. C., Hochberg, M. C., McAlindon, T., Dieppe, P. A., Minor, M. A., Blair, S. N., Berman, B. M., Fries, J. F., Weinberger, M., Lorig, K. R., Jacobs, J. J., and Goldberg, V. (2000b). Osteoarthritis: new insights, part 2: treatment approaches. *Annals of Internal Medicine*, **133**(9), 726–737.

Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, **59**(2), 167–181.

Folkesson, J., Dam, E. B., Olsen, O. F., Pettersen, P. C., and Christiansen, C. (2007). Segmenting articular cartilage automatically using a voxel classification approach. *IEEE Transactions on Medical Imaging*, **26**(1), 106–115.

Ford, J. L. R. and Fulkerson, D. R. (1962). *Flows in networks*. Princeton University Press.

Fripp, J., Crozier, S., Warfield, S. K., and Ourselin, S. (2010). Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee. *IEEE Transactions on Medical Imaging*, **29**(1), 21–27.

Ghosh, S., Beuf, O., Ries, M., Lane, N. E., Steinbach, L. S., Link, T. M., and Majumdar, S. (2000). Watershed segmentation of high resolution magnetic resonance images of articular cartilage of the knee. In *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE*, volume 4, pages 3174–3176.

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

Gougoutas, A. J., Wheaton, A. J., Borthakur, A., Shapior, E. M., Kneeland, J. B., Udupa, J. K., and Reddy, R. (2004). Cartilage volume quantification via live wire segmentation. *Academic Radiology*, **11**(12), 1189–1195.

Grau, V., Mewes, A. U. J., Alcaniz, M., Kikinis, R., and Warfield, S. (2004). Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on Medical Imaging*, **23**(4), 447–458.

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

Išgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M. A., and 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**(7), 1000–1010.

Jones, G., Ding, C., Scott, F., Glisson, M., and Cicuttini, F. (2004). Early radiographic osteoarthritis is associated with substantial changes in cartilage volume and tibial bone surface area in both males and females. *Osteoarthritis Cartilage*, **12**(2), 169–174.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active conour models. *International Journal of Computer Vision*, pages 321–331.

Kauffmann, C., Gravel, P., Godbout, B., Gravel, A., Beaudoin, G., Raynauld, J. P., Materl-Pelletier, J., Pelletier, J. P., and de Guise, J. A. (2003). Computer-aided method for quantification of cartilage thickness and volume changes using mri: validation study using a synthetic model. *IEEE Transactions on Biomedical Engineering*, **50**(8), 978–988.

Kellgren, J. and Lawrence, J. (1957). Radiological assessment of osteoarthritis. *Annals of Rheumatic Diseases*, **16**(4), 494–502.

Koo, S., Hargreaves, B. A., and Gold, G. E. (2009). Automatic segmentation of articular cartilage from MRI. *Patent, US 2009/0306496*.

Lynch, J. A., Zaim, S., Zhao, J., Stork, A., Peterfy, C. G., and Genant, H. K. (2000). Cartilage segmentation of 3D MRI scans of the osteoarthritic knee combining user knowledge and active contours. In *Proceedings of SPIE. Interna- tional Society for Optical Engineering 2000*, volume 3979, pages 925–935.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif. University of California Press.

Mortensen, E. N. and Barrett, W. A. (1995). Intelligent scissors for image composition. In *SIGGRAPH '95 Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 191–198.

Peterfy, C. G., van Dijke, C. F., Janzen, D. L., Glüer, C. C., Namba, R., Jajumdar, S., Lang, P., and Genant, H. K. (1994). Quantification of articular cartilage in the knee with pulsed saturation transfer subtraction and fat-suppressed MR imaging: optimization and validation. *Radiology*, **192**(2), 485–491.

Peterfy, C. G., Schneider, E., and Nevitt, M. (2008). The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis and Cartilage*, **16**, 1433–1441.

Piplani, M. A., Disler, D. G., McCauley, T. R., Holmes, T. J., and Cousins, J. P. (1996). Articular cartilage volume in the knee: semiautomated determination from three-dimensional reformations of MR images. *Radiology*, **198**(3), 855–859.

Pock, T., Schoenemann, T., Graber, G., Bischof, H., and Cremers, D. (2008). A convex formulation of continuous multi-label problems. In *In ECCV, pages III: 792805*.

Raynauld, J. P. (2003). Quantitative magnetic resonance imaging of articular cartilage in knee osteoarthritis. *Current Opinions in Rheumatology*, **15**(5), 647–650.

Raynauld, J. P., Martel-Pelletier, J., Berthiaume, M. J., and et al. (2004). Quantitative magnetic resonance imaging evaluation of knee osteoarthritis progression over two years and correlation with clinical symptoms and radiologic changes. *Arthritis Rheumatology*, **50**(2), 476–487.

Rohlfing, T., Brandt, R., Menzel, R., and 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.

Rousseau, F., Habas, P., and Studholme, C. (2011). A supervised patch-based approach for human brain labeling. *IEEE Transactions on Medical Imaging*, **30**(10), 1852–1862.

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

Sethian, J. A. (1999). *Level set methods and fast marching methods - evolving interfaces in computational geometry, fluid mechanics, computer vision and materials science*. Cambridge Press.

Shan, L., Zach, C., Styner, M., Charles, C., and Niethammer, M. (2010). Automatic bone segmentation and alignment from MR knee images. *SPIE Medical Imaging*.

Shim, H., Chang, S., Tao, C., Wang, J.-H., Kwoh, C. K., and Bae, K. T. (2009). Knee cartilage: efficient and reproducible segmentation on high-spatial-resolution MR images with the semiautomated graph-cut method. *Radiology*, **251**(2), 548–556.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A non-parametric method for automatic correction of intensity non-uniformity in MRI data. *IEEE Transactions on Medical Imaging*, **17**, 87–97.

Solloway, S., Hutchinson, C. E., Waterton, J. C., and Taylor, C. J. (1997). The use of active shape models for making thickness measurements of articular cartilage from MR images. *Magnetic Resonance in Medicine*, **37**(6), 943–952.

Steines, D., Cheng, C., Wong, A., Berger, F., Napel, S., and Lang, P. (2000). Segmentation of osteoarthritic femoral cartilage from MR images. In *Computer Assisted Radiology and Surgery: Proceedings of the 14th International Congress and Exhibition*, pages 303–308.

Tamez-Peña, J., Farber, J., González, P. C., Schrever, E., Schneider, E., and Totterman, S. (2012). Unsupervised segmentation and quantification of anatomical knee features: data from the osteoarthritis initiative. *IEEE Transactions on Biomedical Engineering*, **59**(4), 1177–86.

Tamez-Pena, J. G., Totterman, S., and Parker, K. J. (1999). Unsupervised statistical segmentation of multispectral volumetric MR images. In *Proceedings of the SPIE Conference on Image Processing*, volume 3661, pages 300–311.

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

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

Watson, P. J., Carpenter, T. A., Hall, L. D., and Tyler, J. A. (1996). Cartilage swelling and loss in a spontaneous model of osteoarthritis visualized by magnetic resonance imaging. *Osteoarthritis Cartilage*, **4**(3), 197–207.

Williams, T. G., Holmes, A. P., Waterton, J. C., Maciewicz, R. A., Hutchinson, C. E., Moots, R. J., Nash, A. F. P., and Taylor, C. J. (2010). Anatomically correspondd regional analysis of cartilage in asymptomatic and osteoarthritic knees by statistical shape modelling of the bone. *IEEE Transactions of Medical Imaging*, **29**(8), 1541–1559.

Wirth, W. and Eckstein, F. (2008). A technique for regional analysis of femorotibial cartilage thickness based on quantitative magenetic resonance imaging. *IEEE Transactions on Medical Imaging*, **27**(6), 737–744.

Wirth, W., Graverand, M.-P. H. L., Wyman, B. T., Maschek, S., Hudelmaier, M., Hitzl, W., Nevitt, M., Eckstein, F., and the OAI Investigator Group (2009). Regional analysis of femorrotibial cartilage loss in a subsample from the osteoarthritis initiative processing subcohort. *Osteoarthritis Cartilage*, **17**(3), 291–297.

Woolf, A. D. and Pfleger, B. (2003). Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization*, **81**(0), 646–656.

Xu, C. and Prince, J. L. (1998). Snakes, shapes and gradient vector flow. *IEEE Transactions on Image Processing*, **7**(3), 359–369.

Xue, Z., Shen, D., and Davatzikos, C. (2006). CLASSIC: Consistent longitudinal alignment and segmentation for serial image computing. *NeuroImage*, **30**, 388–399.

Yekutieli, D. and Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, **82**(12), 171–196.

Yezzi, A. J. and Prince, J. L. (2003). An Eulerian PDE approach for computing tissue thickness. *IEEE Transactions on Medical Imaging*, **22**(10), 1332–1339.

Yin, Y., Williams, R., Anderson, D. D., and 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*, pages 241–250.

Yin, Y., Zhang, X., Williams, R., Wu, X., Anderson, D. D., and Sonka, M. (2010b). LOGISMOS - layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint. *IEEE Transactions on Medical Imaging*, **29**(12), 2023–2037.

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

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