Semantic Context Forests for Learning-Based Knee Cartilage Segmentation in 3D MR Images

MICCAI 2013: Workshop on Medical Computer Vision

Authors:

Quan Wang, Dijia Wu, Le Lu, Meizhu Liu,
Kim L. Boyer, and Shaohua Kevin Zhou

Rensselaer

SIEMENS

Background

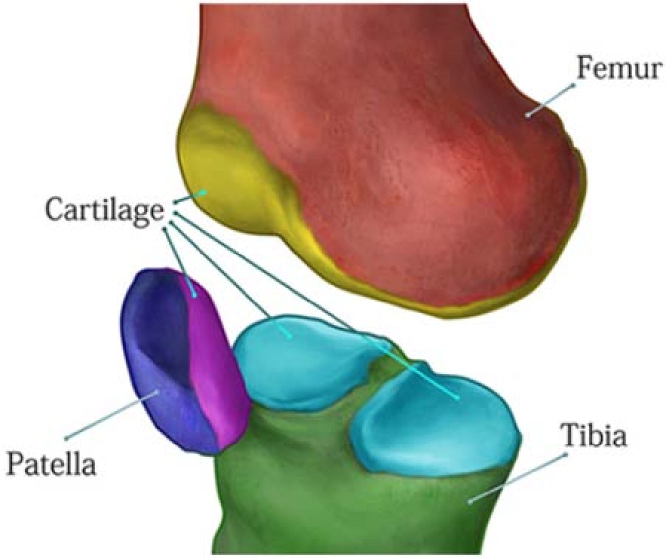
- Knee cartilage analysis is important:
  - Needed for study of cartilage morphology and physiology
  - Required for surgical planning of knee osteoarthritis (OA)

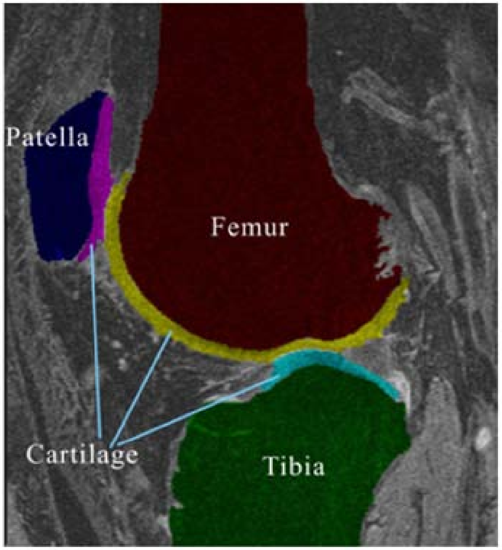
- Lots of research in knee cartilage segmentation:
  - SKI10 – MICCAI 2010 Grand Challenge
    - http://www.ski10.org/
  - Publications on TMI, CVIU, MRI, *etc*.

Knee Joint Anatomy

- Three knee bones:
  - Femur
  - Tibia
  - Patella

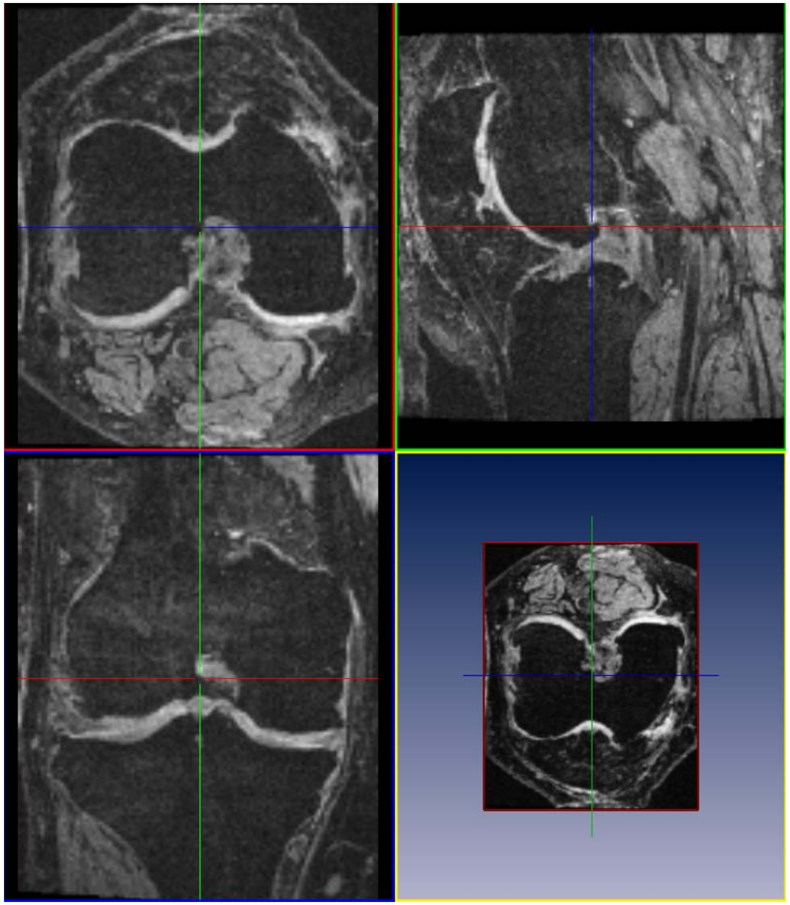
- Three knee cartilages:
  - Femoral cartilage
  - Tibial cartilage (2 pieces)
  - Patellar cartilage





Our Dataset

- The Osteoarthritis Initiative (OAI) dataset
  - 176 volumes
- "iMorphics" annotations
  - Cartilage ground truth
- Modality
  - 3D MR images
- Resolution
  - 0.365mm × 0.365mm × 0.7mm
- Volume size
  - 384 × 384 × 160
- Cohort
  - Progression: all subjects show symptoms of OA



Challenges

- Large appearance variations
  - Inhomogeneous intensities and textures

Challenges

Naïve voxel classification would fail

- Large appearance variations
  - Inhomogeneous intensities and textures

Challenges

Naïve voxel classification would fail

- Large appearance variations
  - Inhomogeneous intensities and textures
- Diffuse boundaries

Challenges

- Large appearance variations
  - Inhomogeneous intensities and textures
- Diffuse boundaries

Naïve voxel classification would fail

Direct graph cuts or random walks would fail

Challenges

Naïve voxel classification would fail

- Large appearance variations
  - Inhomogeneous intensities and textures

Direct graph cuts or random walks would fail

- Diffuse boundaries
- Large shape variations
  - Shape of cartilage varies tremendously due to bone shape variations and severity of disease

Challenges

Naïve voxel classification would fail

- Large appearance variations
  - Inhomogeneous intensities and textures

Direct graph cuts or random walks would fail

- Diffuse boundaries

Shape models are not reliable

- Large shape variations
  - Shape of cartilage varies tremendously due to bone shape variations and severity of disease

Challenges

- Large appearance variations
  - Inhomogeneous intensities and textures
- Diffuse boundaries
- Large shape variations
  - Shape of cartilage varies tremendously due to bone shape variations and severity of disease
- Multiple cartilages
  - Need to avoid overlapping

Naïve voxel classification would fail

Direct graph cuts or random walks would fail

Shape models are not reliable

Challenges

- Large appearance variations
  - Inhomogeneous intensities and textures

Naïve voxel classification would fail

- Diffuse boundaries

Direct graph cuts or random walks would fail

- Large shape variations
  - Shape of cartilage varies tremendously due to bone shape variations and severity of disease

Shape models are not reliable

- Multiple cartilages
  - Need to avoid overlapping

Better not to segment different cartilages separately

Intuitions

- Each cartilage only grows on certain regions of its corresponding bone surface
- Bone segmentation is much easier than cartilage segmentation
  - Larger size
  - More regular shape
  - More discriminative intensity distribution

Existing Methods

- Folkesson: voxel classification
  - Only intensity/texture features
  - No bone segmentation
- Shan: atlas-based

Existing Methods

Pixel: picture element
Voxel: volume element

- Folkesson: voxel classification
  - Only intensity/texture features
  - No bone segmentation
- Shan: atlas-based

Existing Methods

Pixel: picture element
Voxel: volume element

- Folkesson: voxel classification
  - Only intensity/texture features
  - No bone segmentation
- Shan: atlas-based

Poor performance

Existing Methods

Pixel: picture element
Voxel: volume element

- Folkesson: voxel classification
  - Only intensity/texture features
  - No bone segmentation
- Shan: atlas-based

Poor performance

- Vincent: active appearance models
  - Build models for (1) bones + cartilages, and (2) each cartilage separately

Pixel: picture element
Voxel: volume element

Existing Methods

- Folkesson: voxel classification
  - Only intensity/texture features
  - No bone segmentation
- Shan: atlas-based

Poor performance

- Vincent: active appearance models
  - Build models for (1) bones + cartilages, and (2) each cartilage separately
- Bone-cartilage interface (BCI) based methods
  1. Yin: BCI + multi-column graph cuts
  2. Fripp: BCI + 1D normal search
  3. Lee: BCI + graph cuts

Existing Methods

Pixel: picture element
Voxel: volume element

- Folkesson: voxel classification
  - Only intensity/texture features
  - No bone segmentation
- Shan: atlas-based

Poor performance

- Vincent: active appearance models
  - Build models for (1) bones + cartilages, and (2) each cartilage separately

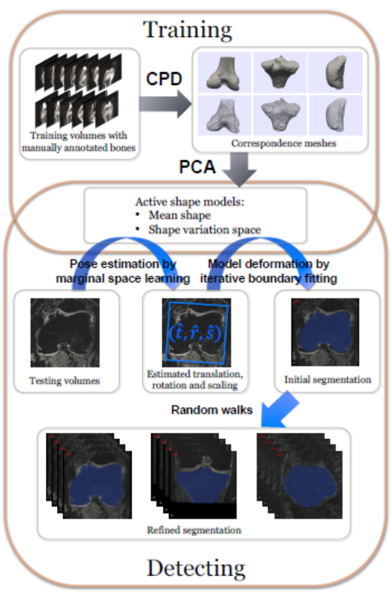
- Bone-cartilage interface (BCI) based methods
  1. Yin: BCI + multi-column graph cuts
  2. Fripp: BCI + 1D normal search
  3. Lee: BCI + graph cuts

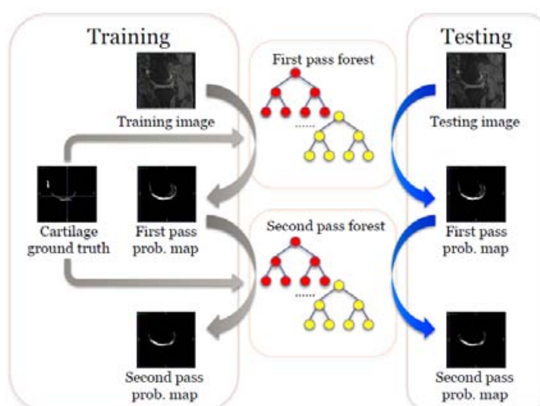
Very complicated

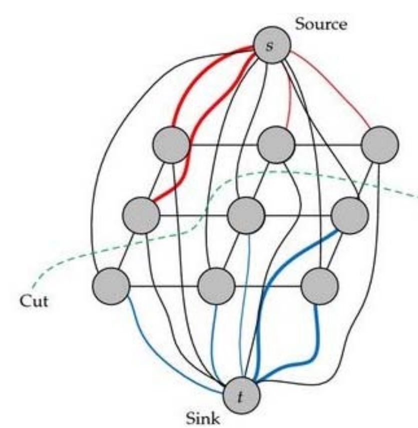
Overview of Our Method

- Diagram:

Bone segmentation by marginal space learning → Voxel classification by random forests → Graph cuts refinement

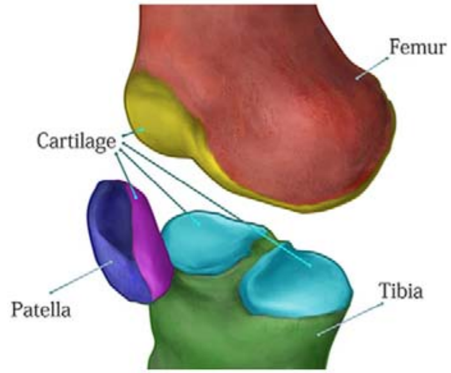


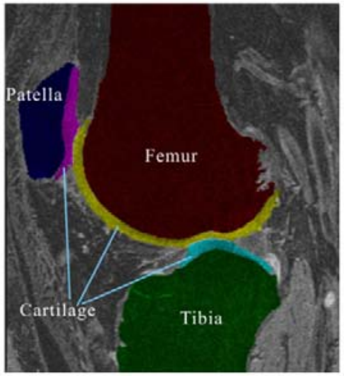




Bone Segmentation

- Bone segmentation is needed to construct distance-based features
- Bone segmentation is much easier than cartilage segmentation
- We segment the 3 knee bones:
    - Femur
    - Tibia
    - Patella





Bone Segmentation Pipeline

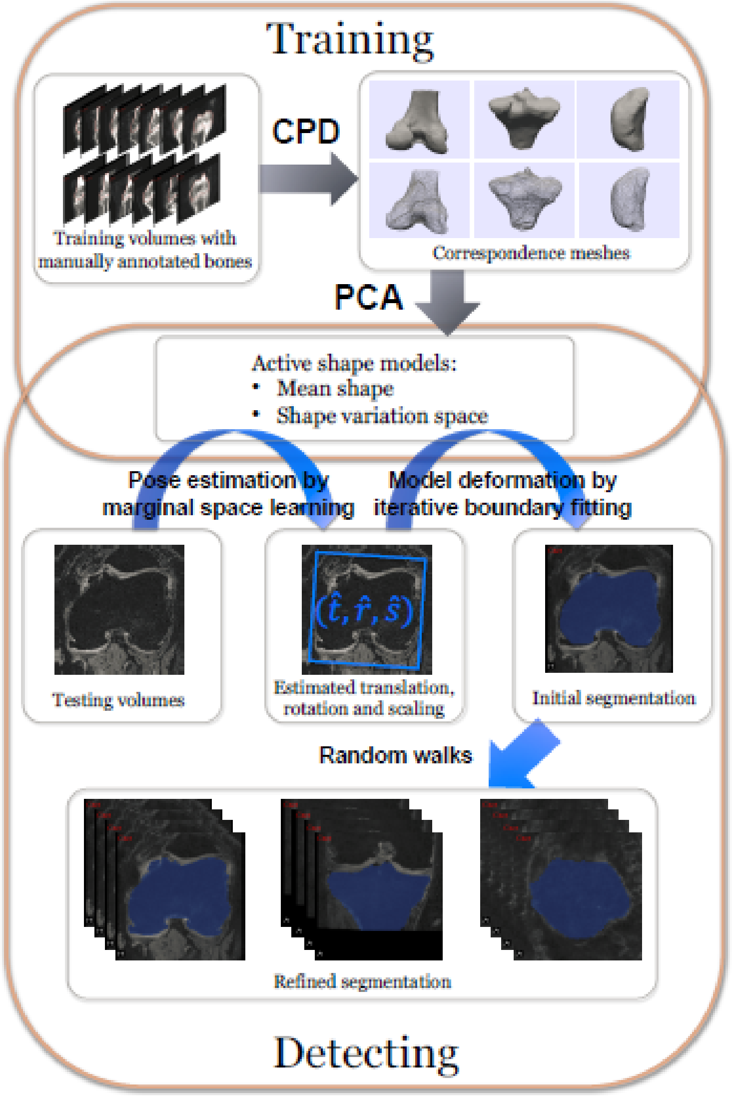
- Step 1: Construct correspondence meshes using Coherent Point Drift [1]
- Step 2: Train PCA models for each bone [2]
- Step 3: Detect bones in images using PCA models
- Step 4: Use random walks to refine segmentation [3]

- [1] A. Myronenko and X. Song. Point set registration: Coherent point drift. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(12):2262–2275, Dec. 2010.

- [2] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models–their training and application. Computer Vision and Image Understanding, 61(1):38–59, 1995.

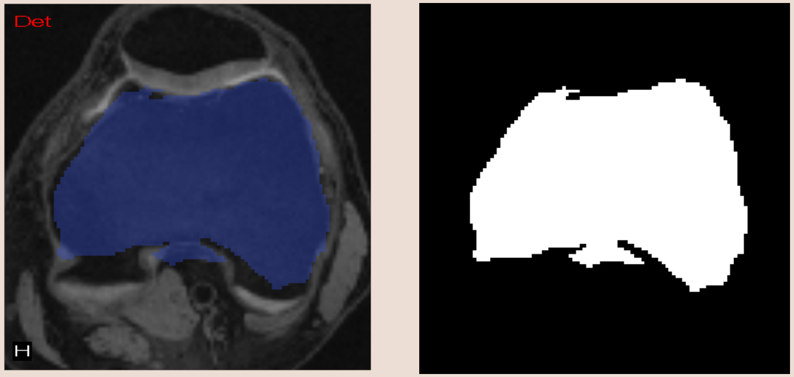
- [3] L. Grady. Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(11):1768–1783, Nov. 2006.

Bone Segmentation Pipeline



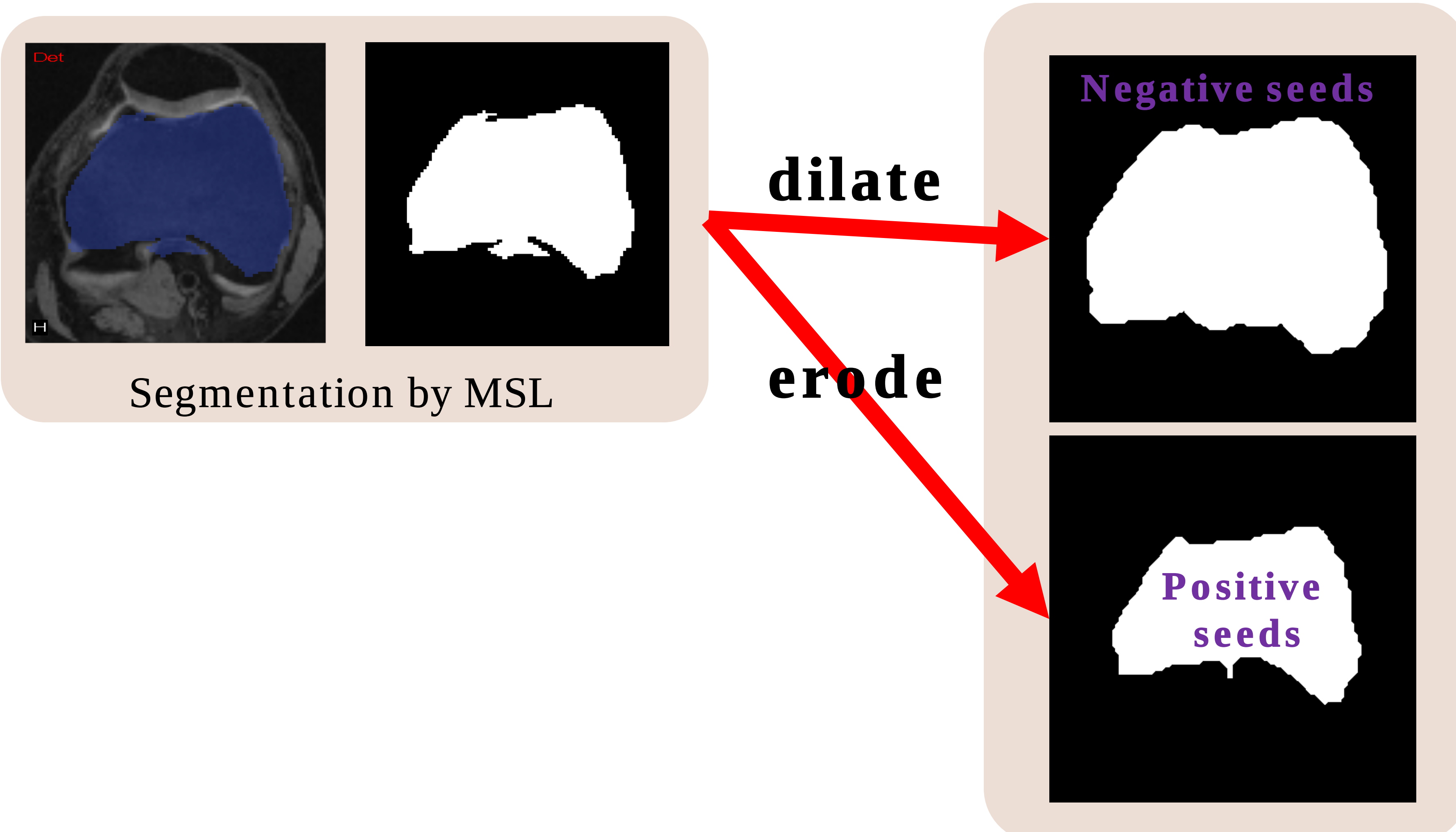
- Training:
  - Train shape models
- Detecting
  1. Bounding box by marginal space learning (MSL)
  2. Model deformation by boundary fitting
  3. Refine with random walks

Refinement by Random Walks

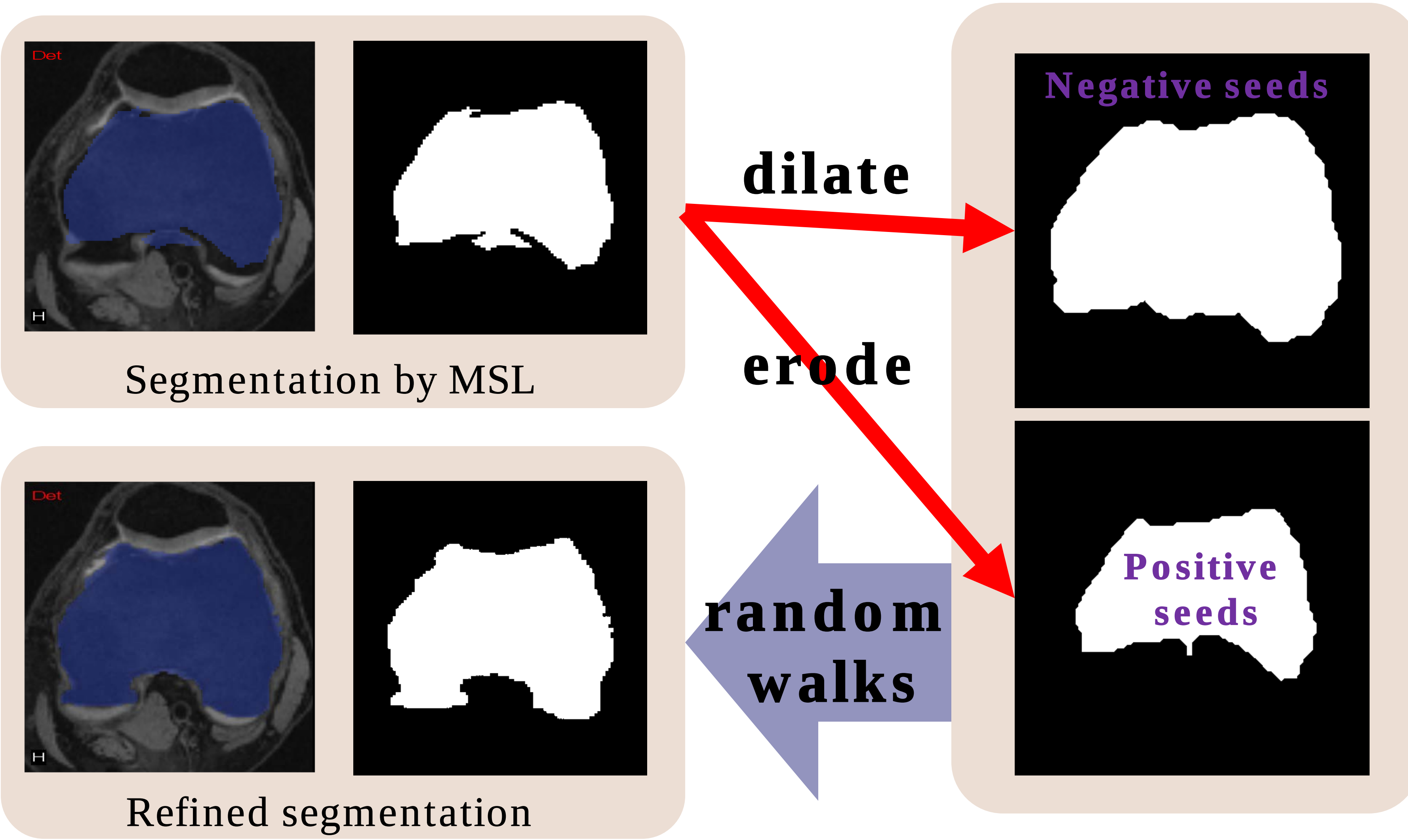


Segmentation by MSL

Refinement by Random Walks

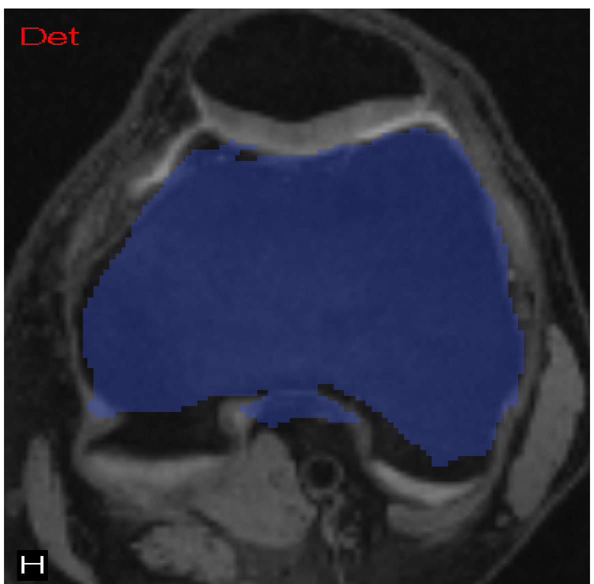


Refinement by Random Walks

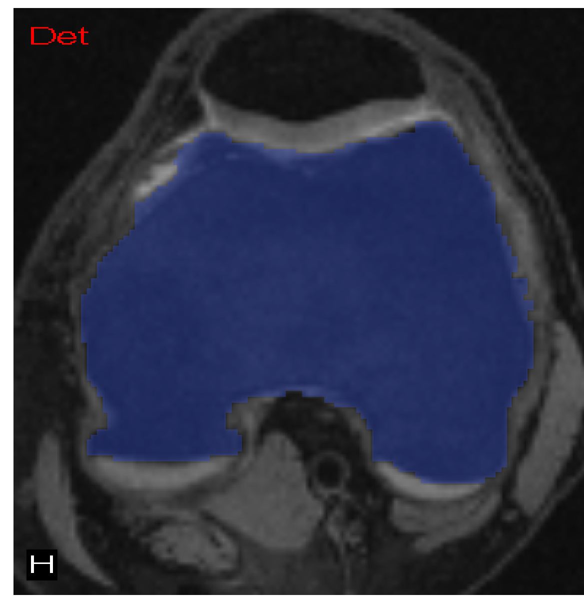


Bone Segmentation Performance

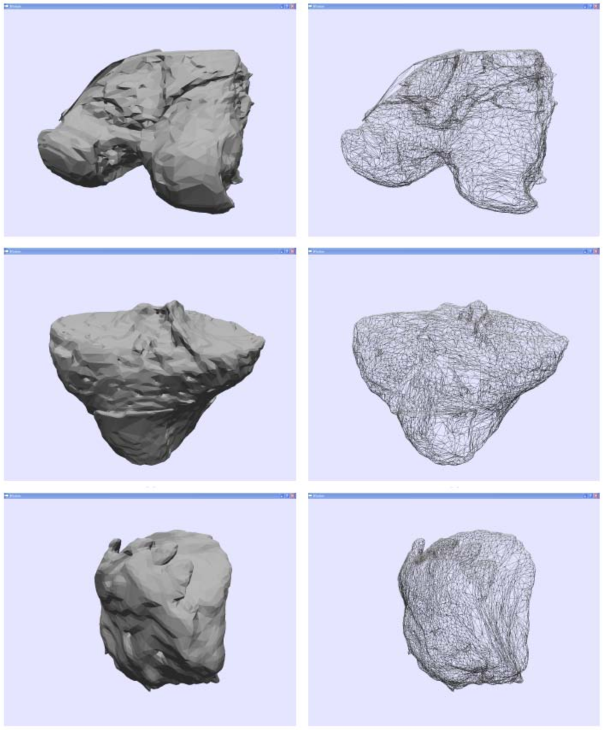
	Femur DSC	Tibia DSC	Patella DSC
Before random walks	92.37%±1.58%	94.64%±1.18%	92.07%±1.47%
After random walks	**94.86**%±1.85%	**95.96**%±1.64%	**94.31**%±2.15%

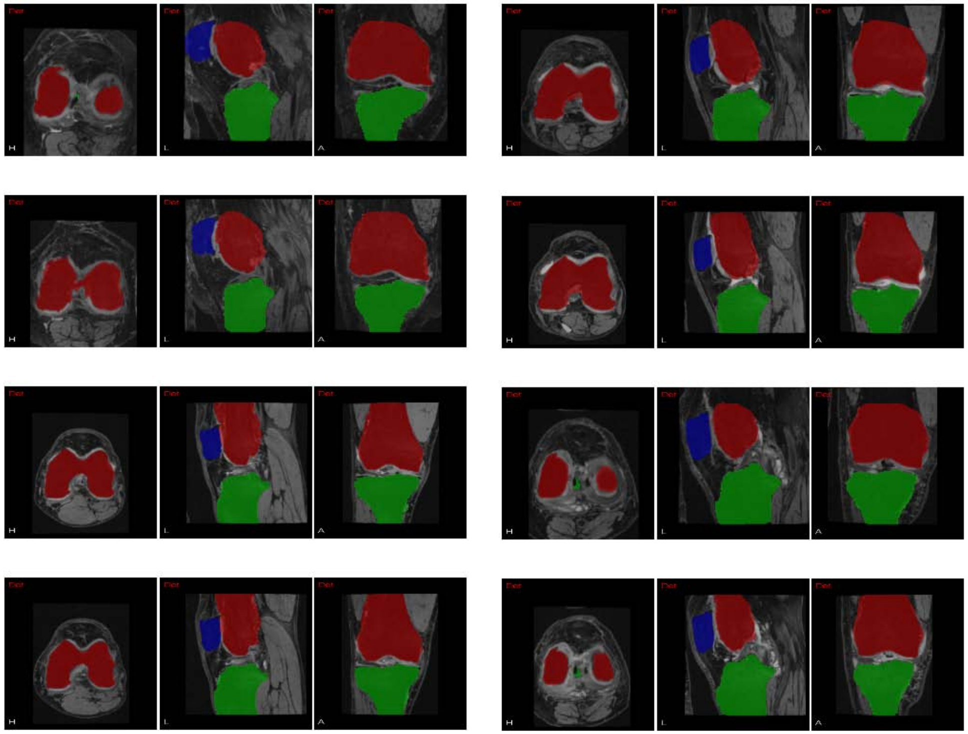


Random walks refinement



Bone Segmentation Examples





Resulting meshes

Resulting masks

Red: femur

Green: tibia

Blue: patella

Overview of Cartilage Segmentation

- 4-class voxel classification for cartilages:
  - Background
  - Femoral cartilage
  - Tibial cartilage
  - Patellar cartilage
- Feature for classification
  - Intensity-based features
  - Distance-based features
  - Semantic context features (RSID&RSPD)
- Classifier
  - Multi-pass random forests (auto-context)
  - Only classify those voxels close to the bone surface (20mm)

Overview of Cartilage Segmentation

- 4-class voxel classification for cartilages:
  - Background
  - Femoral cartilage
  - Tibial cartilage
  - Patellar cartilage
- Feature for classification
  - Intensity-based features
  - Distance-based features
  - Semantic context features (RSID&RSPD)
- Classifier
  - Multi-pass random forests (auto-context)
  - Only classify those voxels close to the bone surface (20mm)

Largely reduces computational cost

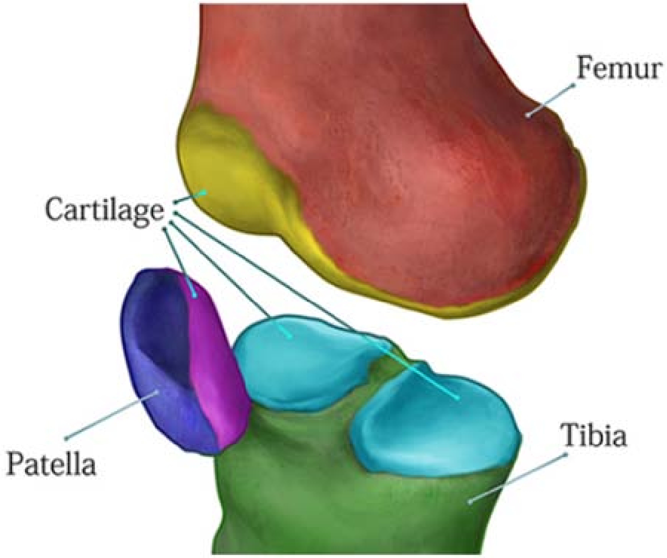
Intensity-Based Features

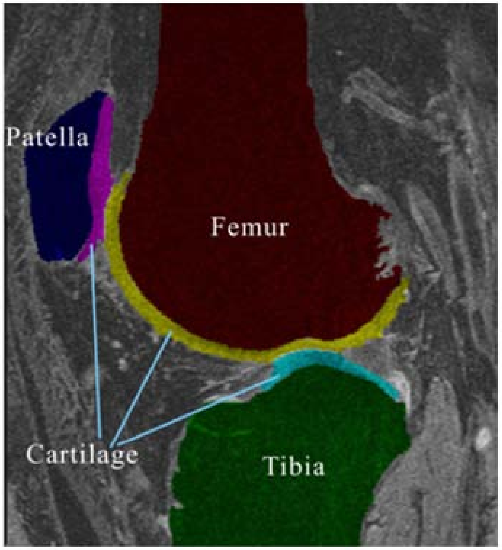
- Intensity:

$$I(\mathbf{x})$$

- Gradient magnitude:

$$\|\nabla I(\mathbf{x})\|$$





Distance-Based Features (1)

- Signed distances to bones
  - We perform signed distance transform to each segmented bone
  - The signed distances at each voxel, and their linear combinations comprise our features:

F: femur	$d_F(\mathbf{x})$
T: tibia	$d_T(\mathbf{x})$
P: patella	$d_P(\mathbf{x})$

Distance-Based Features (1)

- Signed distances to bones
  - We perform signed distance transform to each segmented bone
  - The signed distances at each voxel, and their linear combinations comprise our features:

F: femur	$d_F(\mathbf{x})$	$d_F(\mathbf{x}) + d_T(\mathbf{x})$
T: tibia	$d_T(\mathbf{x})$	$d_F(\mathbf{x}) + d_P(\mathbf{x})$
P: patella	$d_P(\mathbf{x})$	

Sum:
Whether voxel is between 2 bones?

Distance-Based Features (1)

- Signed distances to bones
  - We perform signed distance transform to each segmented bone
  - The signed distances at each voxel, and their linear combinations comprise our features:

F: femur	$d_F(\mathbf{x})$	$d_F(\mathbf{x}) + d_T(\mathbf{x})$
T: tibia	$d_T(\mathbf{x})$	$d_F(\mathbf{x}) + d_P(\mathbf{x})$
P: patella	$d_P(\mathbf{x})$	$d_F(\mathbf{x}) - d_T(\mathbf{x})$
		$d_F(\mathbf{x}) - d_P(\mathbf{x})$

Sum:
Whether voxel is between 2 bones?

Difference:
Which bone is closer?

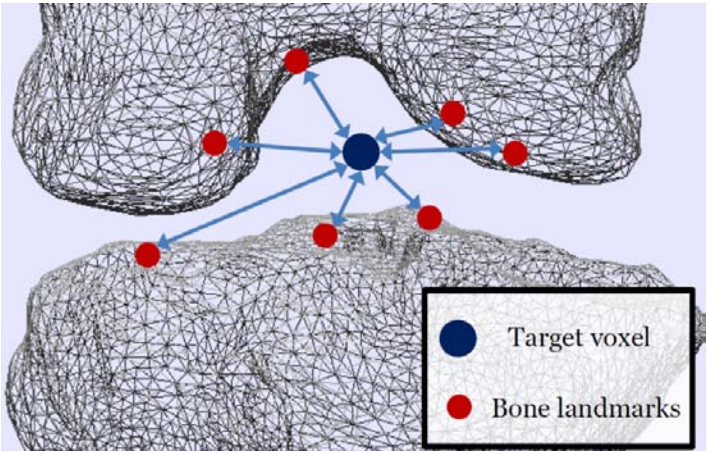
Distance-Based Features (2)

- Distances to densely registered bone landmarks
  - We measure the distance from a voxel to each landmark on the joint bone mesh

$$f_{11}(\mathbf{x}, \zeta) = \left\| \mathbf{x} - \mathbf{z}_{\zeta} \right\|$$

  - $\mathbf{z}_{\zeta}$ is the spatial coordinates of the ζth landmark on the bone mesh (ζ: index of landmark)

This feature group replaces the estimation of Bone-Cartilage Interface (BCI)



Semantic Context Features (1)

- Random shift intensity difference (RSID)

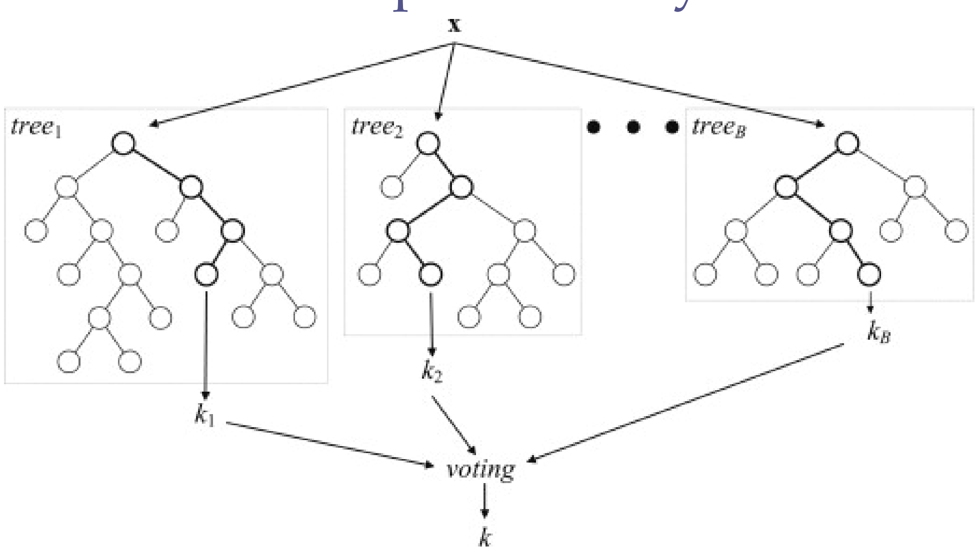
$$f_{10}(\mathbf{x}, \mathbf{u}) = I(\mathbf{x} + \mathbf{u}) - I(\mathbf{x})$$

  - The spatial shift **u** is randomly generated in training

- Distances to landmarks (f_{11}) and RSID (f_{10}) involve random parameters (ζ and **u**), thus they are both "feature groups"

Classifier: Random Forests

- We use multi-class random forests as our classifier
  - Reasons for our choice:
    1. Although training is slow, decision is very fast
    2. Classification results are probabilities, which can be used to construct new features (discussed later)
    3. Very easy to implement
    4. Forest size and depth are easy to customize



Classifier: Random Forests

- Training:
  - Use maximal entropy reduction principle
  - Tree depth: 18
  - At each non-leaf node, generate 1000 (feature, threshold) pairs
  - At each leaf node, compute the probability of being:
    - Background, femoral cartilage, tibial cartilage, patellar cartilage
  - Number of trees in a forest: 60

Classifier: Random Forests

Best separates different classes

- Training:
  - Use maximal entropy reduction principle
  - Tree depth: 18
  - At each non-leaf node, generate 1000 (feature, threshold) pairs
  - At each leaf node, compute the probability of being:
    - Background, femoral cartilage, tibial cartilage, patellar cartilage
  - Number of trees in a forest: 60

Classifier: Random Forests

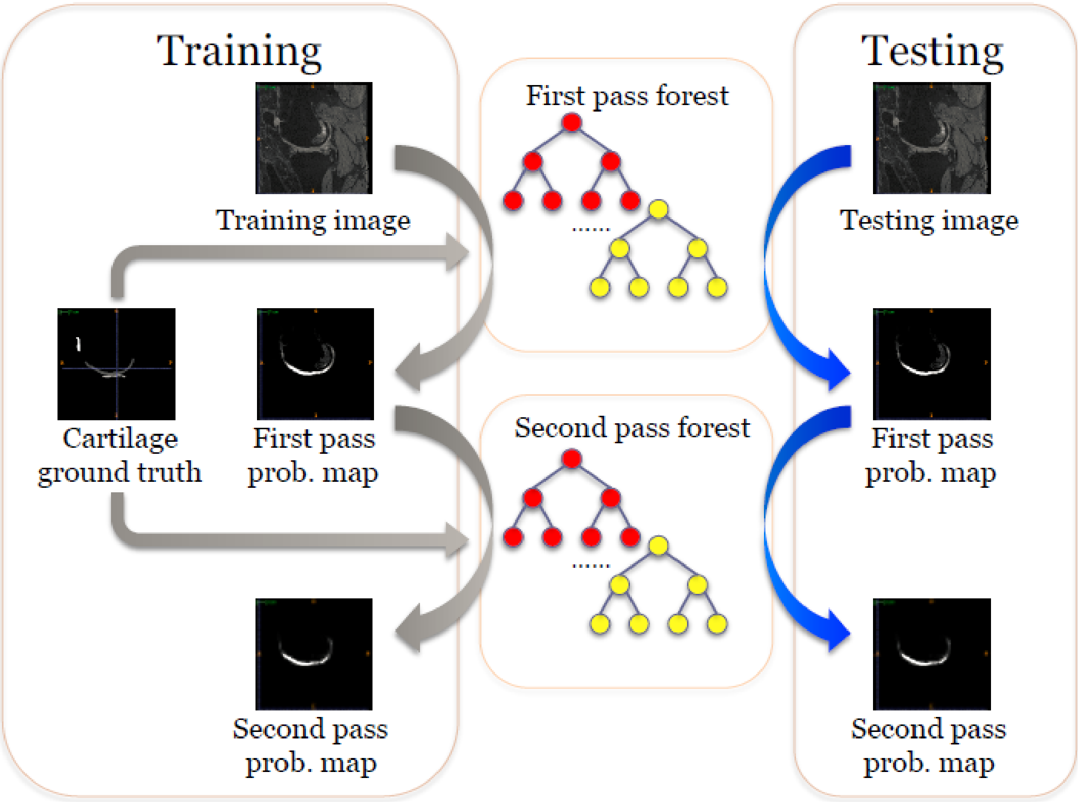
Best separates different classes

- Training:
  - Use maximal entropy reduction principle
  - Tree depth: 18
  - At each non-leaf node, generate 1000 (feature, threshold) pairs
  - At each leaf node, compute the probability of being:
    - Background, femoral cartilage, tibial cartilage, patellar cartilage
  - Number of trees in a forest: 60

Trade-off between computational cost and performance

Multi-Pass Random Forests

- After 1-pass random forest, we use the resulting probabilities to train a second pass
- Similar idea to cascaded classifiers, auto-context, *etc.*



Semantic Context Features (2)

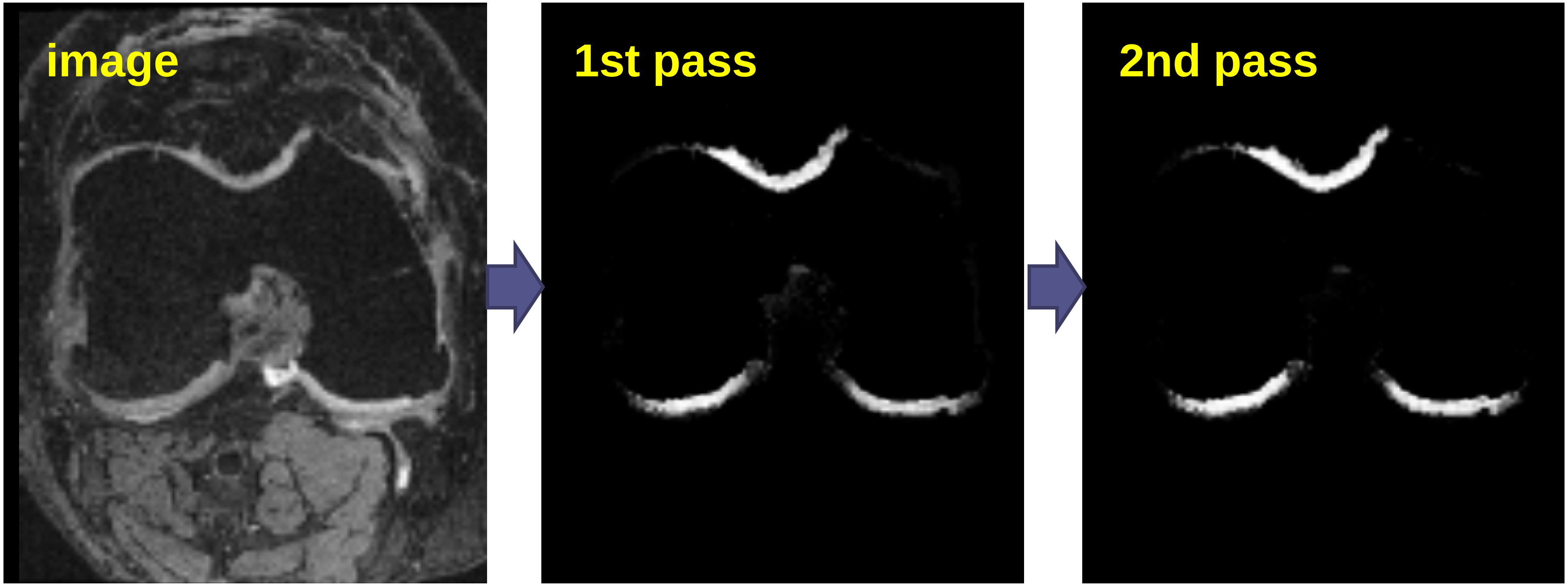
- In the second pass, we construct probability features and random shift probability difference (RSPD) features

F: femur	$P_F(\mathbf{x})$	$P_F(\mathbf{x}+\mathbf{u}) - P_F(\mathbf{x})$
T: tibia	$P_T(\mathbf{x})$	$P_T(\mathbf{x}+\mathbf{u}) - P_T(\mathbf{x})$
P: patella	$P_P(\mathbf{x})$	$P_P(\mathbf{x}+\mathbf{u}) - P_P(\mathbf{x})$

  - The shift **u** is randomly generated in training
  - Similar to random shift intensity difference features

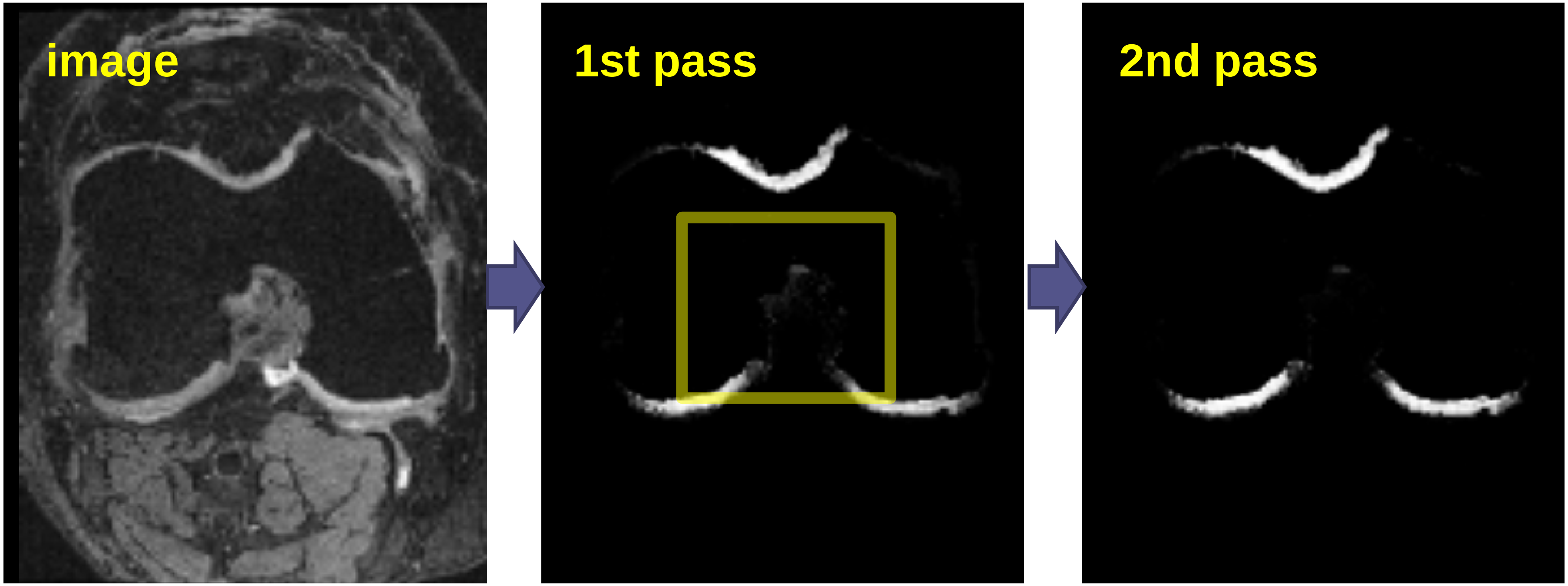
Probability Maps from Multi-Pass

- Image, 1st pass and 2nd pass probability map of femoral cartilage
- We can see, in each new pass we get cleaner results



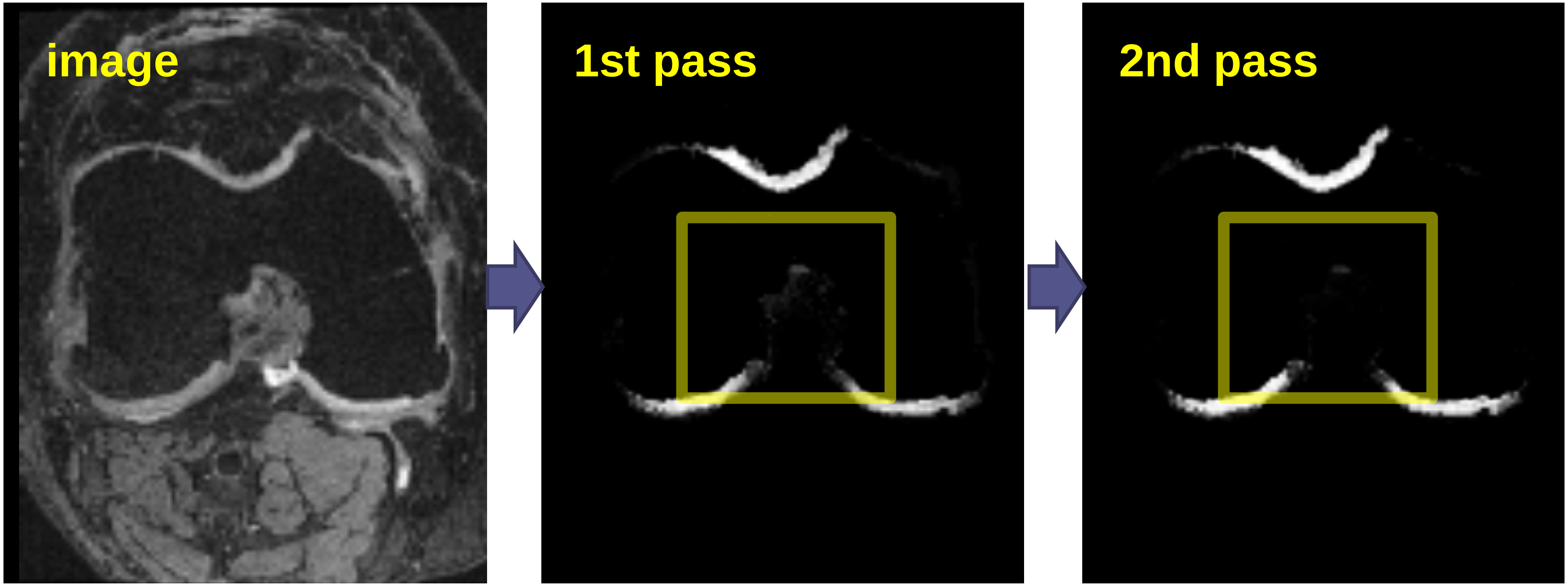
Probability Maps from Multi-Pass

- Image, 1st pass and 2nd pass probability map of femoral cartilage
- We can see, in each new pass we get cleaner results



Probability Maps from Multi-Pass

- Image, 1st pass and 2nd pass probability map of femoral cartilage
- We can see, in each new pass we get cleaner results



Graph Cuts Refinement

- The multi-label graph cuts algorithm
  - 4 labels:
    - Background
    - Femoral cartilage
    - Tibial cartilage
    - Patellar cartilage
  - Algorithm [4]
    - α-expansion
    - α-β-swap

Using probabilities from multi-pass forests

[4] Yuri Boykov, Olga Veksler, Ramin Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *TPAMI*, 2001.

Multi-label Graph Cuts

- Target:
  - Minimize

$$E(f) = E_{smooth}(f) + E_{data}(f)$$

$$E(f) = \sum_{\{p,q\} \in N} V_{p,q}(f_p, f_q) + \sum_{p \in P} D_p(f_p)$$

  - f: label configuration
  - P: the set of all voxels
  - N: neighborhood system
  - D_p: regional energy
  - $V_{p,q}$: boundary energy

Graph Configuration

- Regional energy:

$$D_p(f_p) = \min\{K, -\lambda \ln P(f_p)\}$$

- Boundary energy:

$$V_{p,q}(f_p, f_q) = u_{\{p,q\}} \cdot \delta(f_p \neq f_q)$$

$$u_{\{p,q\}} = \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{\mathrm{dist}(p,q)}$$

Graph Configuration

- Regional energy:

$$D_p(f_p) = \min\{K, -\lambda \ln P(f_p)\}$$

Probability from multi-pass forests

- Boundary energy:

$$V_{p,q}(f_p, f_q) = u_{\{p,q\}} \cdot \delta(f_p \neq f_q)$$

$$u_{\{p,q\}} = \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{\mathrm{dist}(p,q)}$$

Graph Configuration

- Regional energy:

$$D_p(f_p) = \min\{K, -\lambda \ln P(f_p)\}$$

Probability from multi-pass forests

- Boundary energy:

$$V_{p,q}(f_p, f_q) = u_{\{p,q\}} \cdot \delta(f_p \neq f_q)$$

$$u_{\{p,q\}} = \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{\mathrm{dist}(p,q)}$$

Parameters: K, λ, σ

Experiments

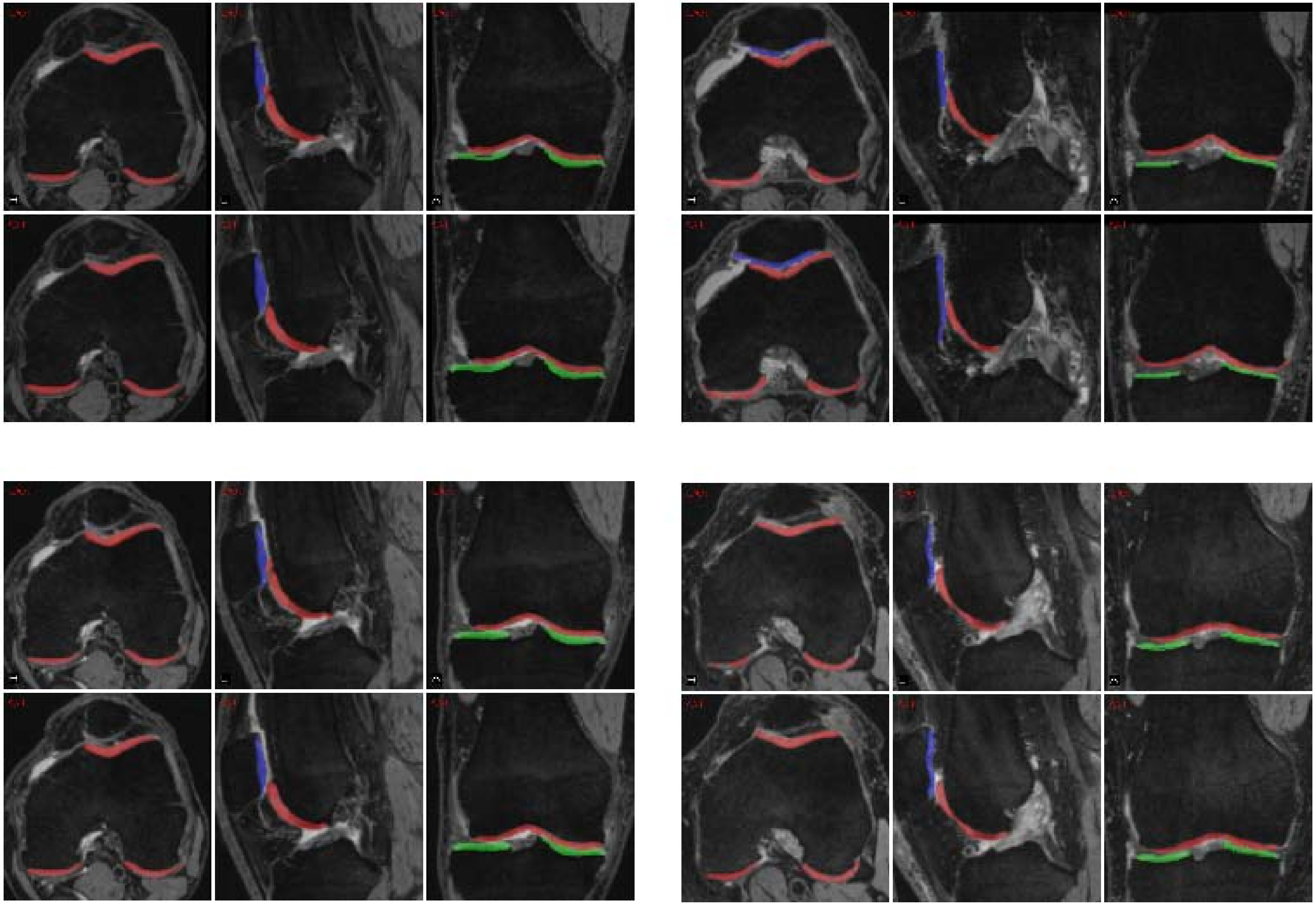
- Dataset:
  - As mentioned before, we use 176 volumes from OAI
- Evaluation protocol:
  - We perform a three-fold cross validation
- Measurement:
  - We report the Dice similarity coefficient (DSC) of three cartilages

Experimental Results

Author	Dataset	Femoral Cartilage DSC Mean	Femoral Cartilage DSC Std.	Tibial Cartilage DSC Mean	Tibial Cartilage DSC Std.	Patellar Cartilage DSC Mean	Patellar Cartilage DSC Std.
Shan et al. (2012)	18 SPGR images	78.2%	5.2%	82.6%	3.8%	–	–
Folkesson et al. (2007)	139 Esaote C-Span images	77%	8.0%	81%	6.0%	–	–
Fripp et al. (2010)	20 FS SPGR images	84.8%	7.6%	82.6%	8.3%	83.3%	13.5%
Lee et al. (2011)	10 images in OAI	82.5%	–	80.8%	–	82.1%	–
Yin et al. (2010)	60 images in OAI	84%	4%	80%	4%	80%	4%
Proposed method	OAI, D_1 subset (58 images)	85.47%	3.10%	84.96%	3.82%	78.56%	9.38%
	OAI, D_2 subset (58 images)	85.20%	3.65%	83.52%	4.08%	80.79%	7.40%
	OAI, D_3 subset (60 images)	84.22%	3.05%	82.74%	3.84%	78.12%	9.63%
	OAI, overall (176 images)	84.96%	3.30%	83.74%	4.00%	79.16%	8.88%

- Dataset:
  - We are using the largest dataset (176 volumes)
  - D_1, D_2 and D_3 are 3 subsets for cross validation
- Remarks:
  - Our method has competitive DSC performance, but since people use different datasets, these numbers are not directly comparable in the strict sense

Example Segmentation



Red:
Femoral cart.

Green:
Tibial cart.

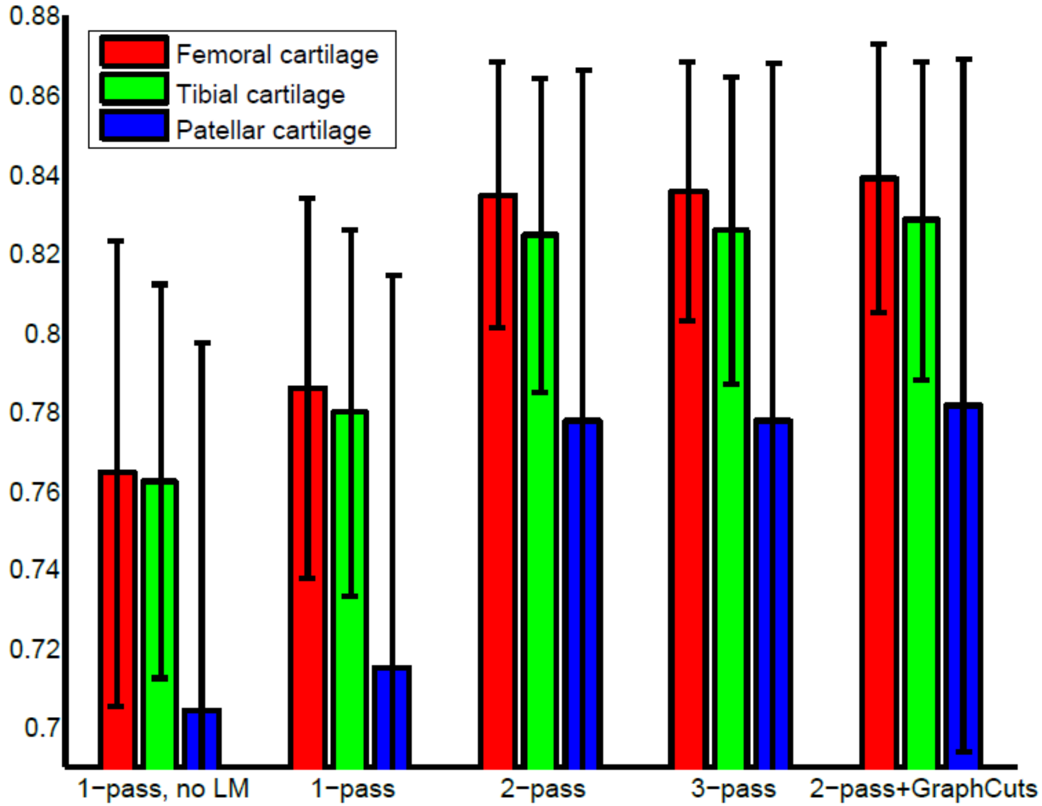
Blue:
Patellar cart.

Upper row:
Our result

Lower row:
Ground truth

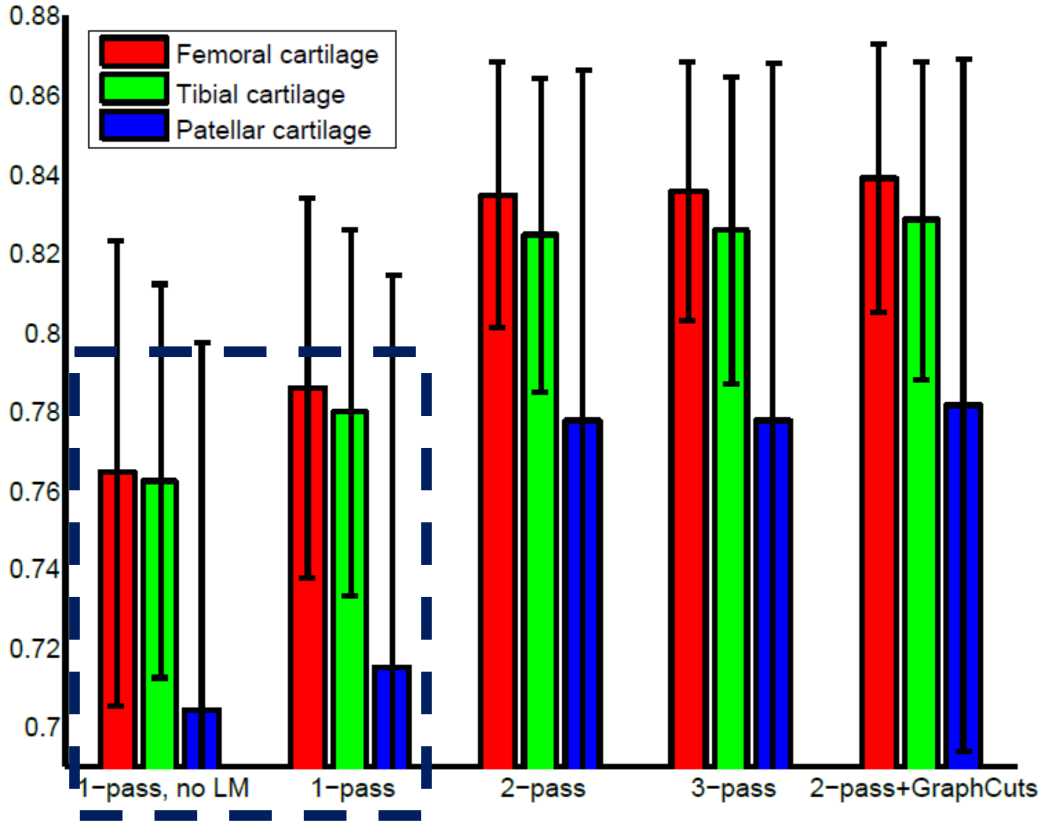
Comparative Study

- How does each component contribute to final performance:



Comparative Study

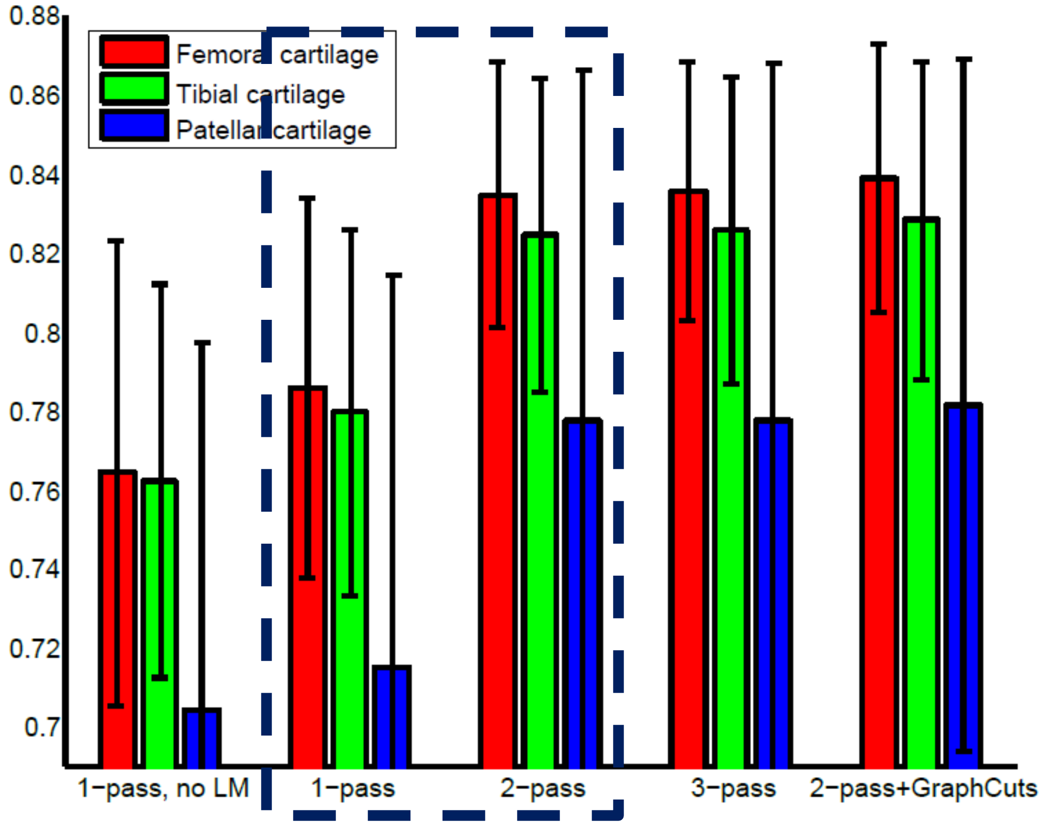
- How does each component contribute to final performance:



Distances to landmarks make a big difference

Comparative Study

- How does each component contribute to final performance:

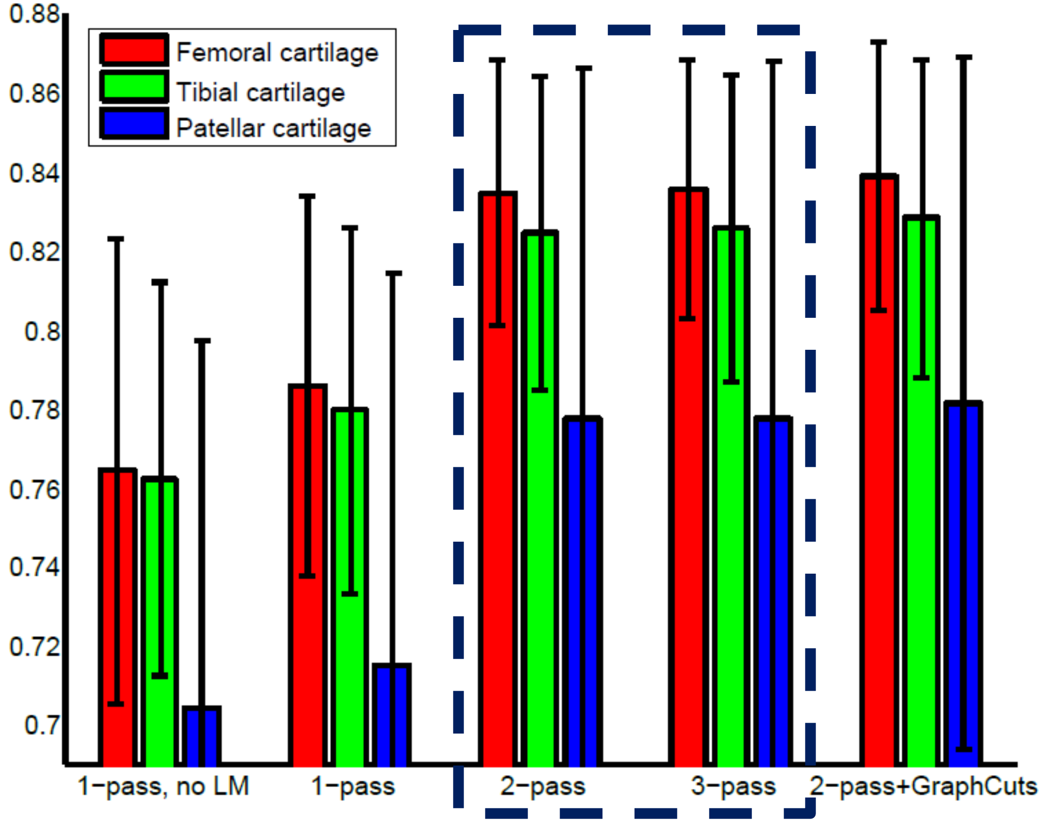


The 2nd pass forest largely improves performance

Comparative Study

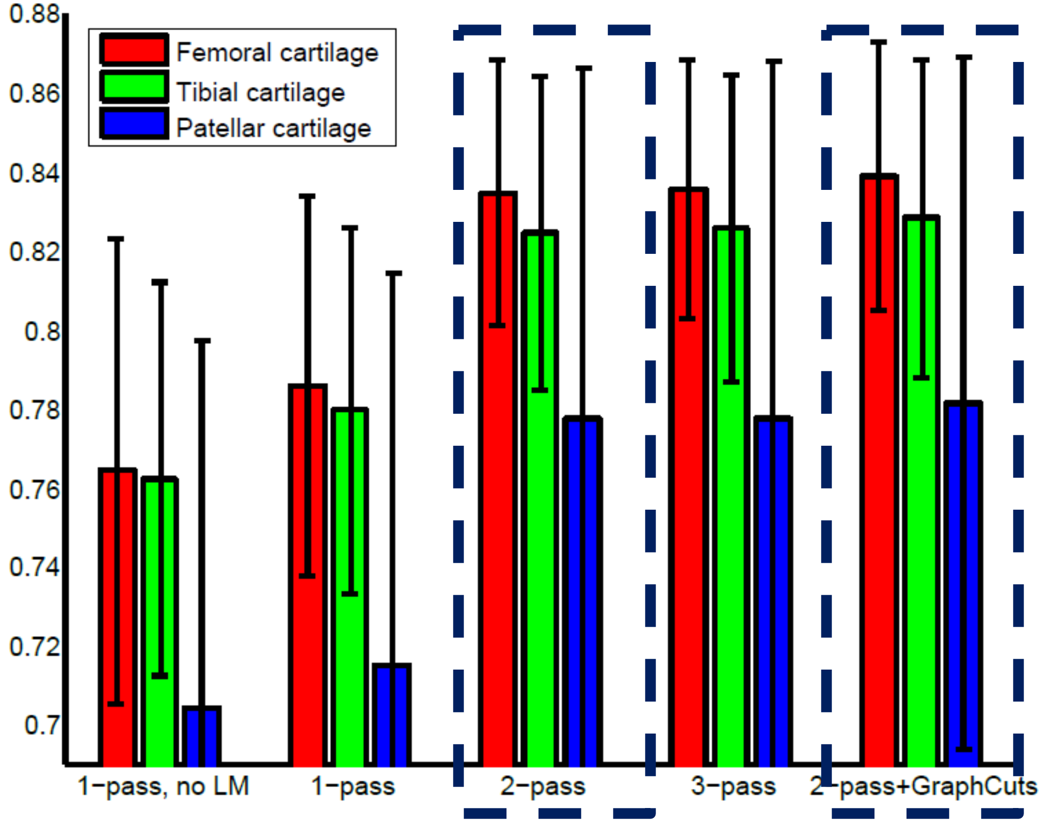
- How does each component contribute to final performance:

The 3rd pass forest doesn't bring much improvement



Comparative Study

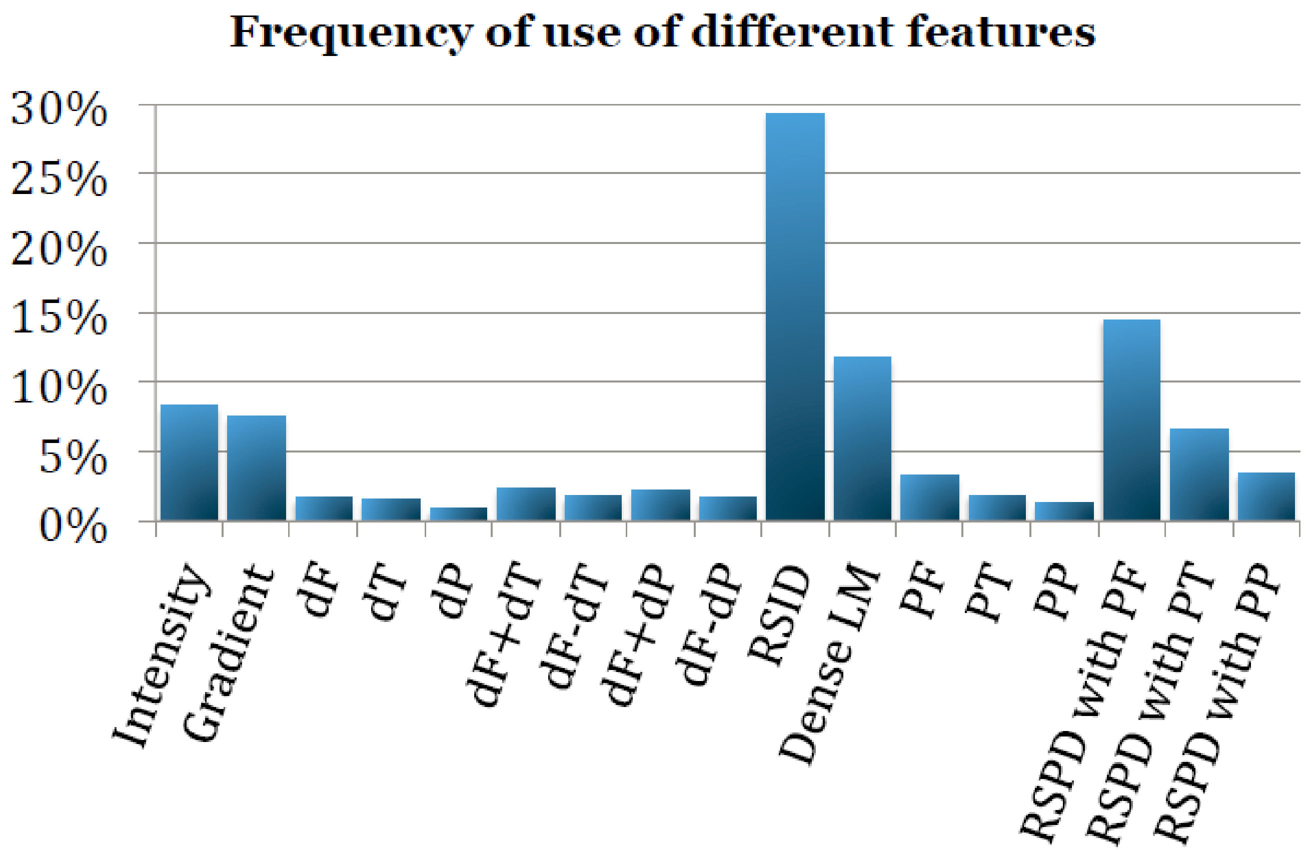
- How does each component contribute to final performance:



Graph cuts slightly improves performance

Comparative Study

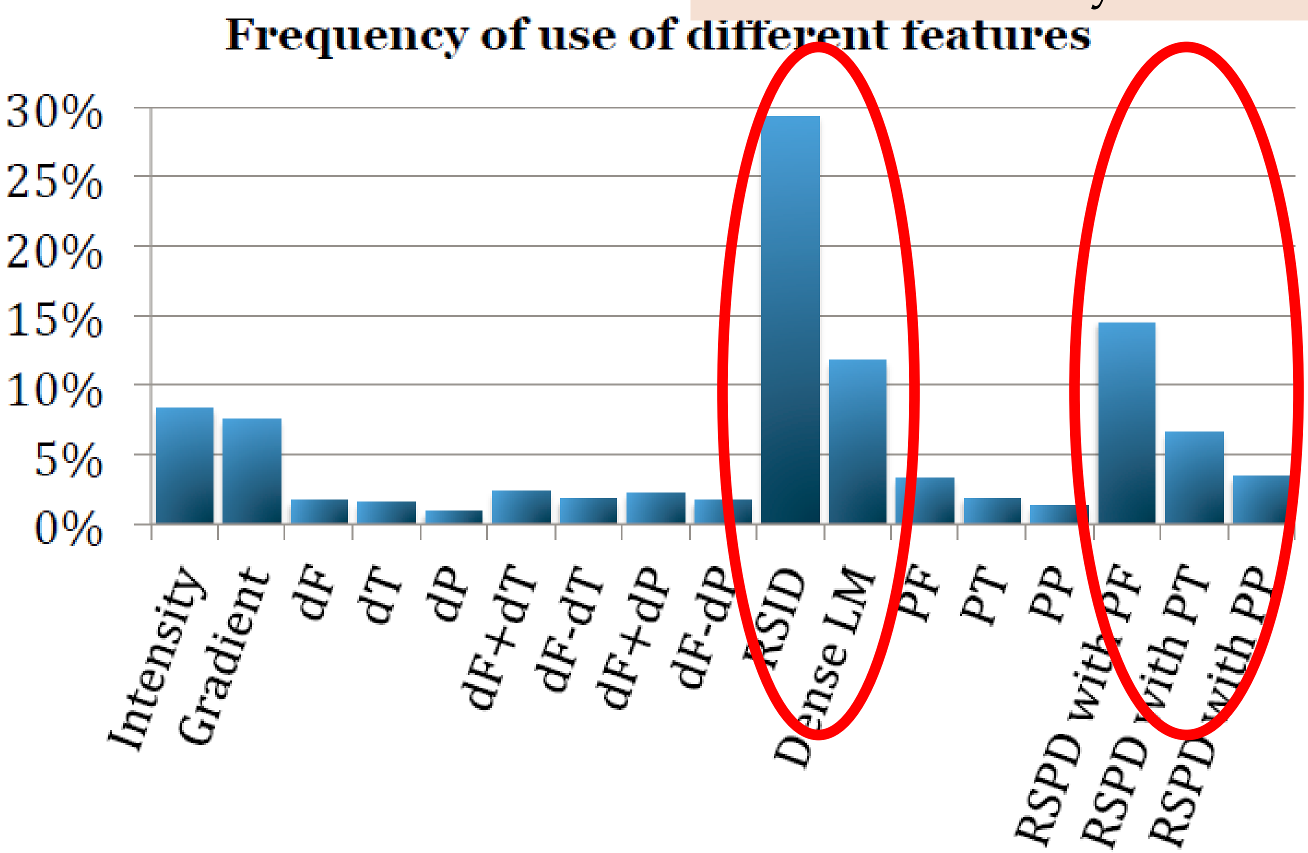
- How often is each feature used in resulting forests:



Comparative Study

- How often is each feature used in the forests:

Distances to landmarks and semantic context features are very useful!



Conclusions

1. We have built a complete system for 3D MR segmentation of knee bones and cartilages.

Conclusions

1. We have built a complete system for 3D MR segmentation of knee bones and cartilages.

   Segmentation of one volume including 3 bones and 3 cartilages takes about 2 minutes on our machine.

Conclusions

2. Our method and system produce highly accurate segmentation results. The reported DSC is close to or higher than those reported in literature.

Conclusions

2. Our method and system produce highly accurate segmentation results. The reported DSC is close to or higher than those reported in literature.

   However, the DSC numbers are not directly comparable in the strict sense, since people use different dataset.

   Our dataset is the largest one compared with others' work.

Conclusions

3. The distance to densely registered landmarks is a very effective feature. It replaces the estimation of bone-cartilage interface (BCI).

Conclusions

3. The distance to densely registered landmarks is a very effective feature. It replaces the estimation of bone-cartilage interface (BCI).

   Moreover, it is a wise way to **combine shape models and learning-based methods**. It encodes the spatial constraints between bones and cartilages into the random forests.

Conclusions

3. The distance to densely registered landmarks is a very effective feature. It replaces the estimation of bone-cartilage interface (BCI).

   Moreover, it is a wise way to **combine shape models and learning-based methods**. It encodes the spatial constraints between bones and cartilages into the random forests.

   We expect good performance of this method in the segmentation of other objects (*e.g*. organs) and other modalities (*e.g*. CT, ultrasound).

Semantic Context Forests for Learning-Based Knee Cartilage Segmentation in 3D MR Images

MICCAI 2013: Workshop on Medical Computer Vision

Authors:

Quan Wang, Dijia Wu, Le Lu, Meizhu Liu,
Kim L. Boyer, and Shaohua Kevin Zhou