# Segmentation of Synovitis

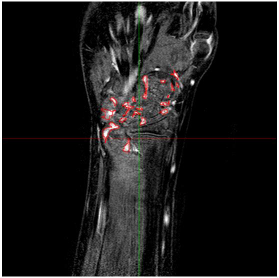
C. Basso A. Verri

Dipartimento di Informatica e Scienze dell'Informazione
Università di Genova

11/07/2007

# Goals

- The volume of the inflamed synovia might be an early predictor for the severity of the disease (it is in the case of Osteoarthritis and Rheumatoid arthritis).
- Manual annotation of the inflamed synovia is extremely time-consuming and prone to variability.



## Goal

provide a semi-automatic quantification method to ***speed up*** the process and ***reduce variability***.

# Data

- T1W-FFE3D FAT, coronal MRI of wrists/hips, acquired at  1.5 min after injection of a contrast bolus.
- The contrast enhances the inflamed synovia, which can be segmented.
- Other structures are enhanced (blood vessels, sinew).

# Approach: Segmentation as Classification

the segmentation problem is seen as a problem of classifying voxels

- manually annotated exams are used for training-validating-testing a set of classifiers, combined by bagging;
- the features are the intensities of the voxels and of their 1-neighbors (in 3D);
- each classifier is a linear combination of Gaussian kernels centered on the examples;
- the width of the kernels is chosen as the median distance between training examples;
- the training is performed by $\nu$-method, an iterative method originally devised for solving inverse problems.

# Annotations

we need annotations for the whole training-validation-testing pipeline

- We provide a software tool for performing annotations.
- Freehand drawing was preferred to polylines.
- Current status:
  - 4 wrists by two clinicians (8 annotations)
  - 1 wrist annotated together

# Annotation Tool

Features:

- loading and visualization of DICOM data;
- support for drawing, saving and loading freehand and elliptic annotations;
- works on different platforms: Win, Mac and Linux;
- controls on image zoom and contrast;
- misc: multiplanar view, maximum intensity projections, thresholding, ...

# Analysis of Manual Annotations

## Inter-observer analysis

- Correlation, per slice: $r = 0.92$, $p < 0.001$
- Absolute difference per slice, avg = 30.5%, median = 14.4%
- Absolute difference per exam, avg = 15.2%, median = 12.1%
- Classification overlap = 63.0%

Computed over 4 exams, with 2 different annotations

## Sensitivity to Morphological Operations

- 1px Dilation (avg), per exam: 33.3%
- 1px Erosion (avg), per exam: -31.4%

Computed over 9 annotations.

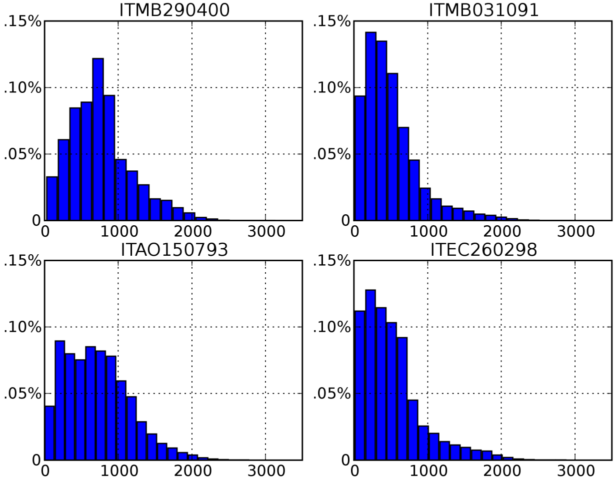
# Analysis of Training Examples

Percentage of overlap between histograms of the intensities for positive and negative examples.
The lower the overlap, the easier the classification task.

| | Obs. 1 | Obs. 2 |
|---|---|---|
| ITAO | **46.9**% | **55.3**% |
| ITEC | 29.5% | 38.3% |
| ITMB03 | 35.9% | 39.0% |
| ITMB29 | 35.8% | **43.1**% |

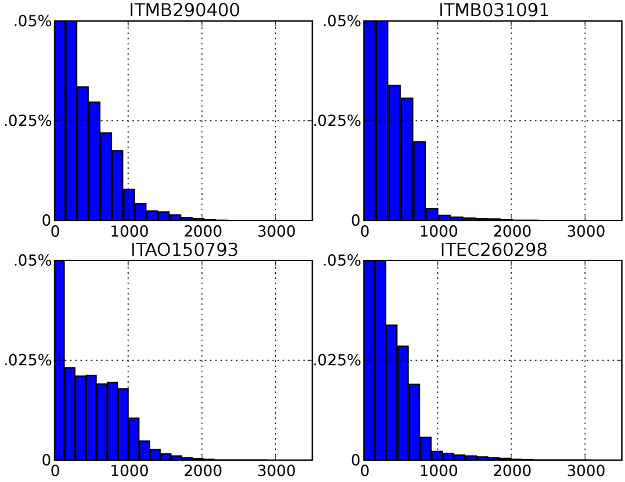
# Analysis of Training Examples

Intensity histograms of the exemplar voxels (pos+neg).



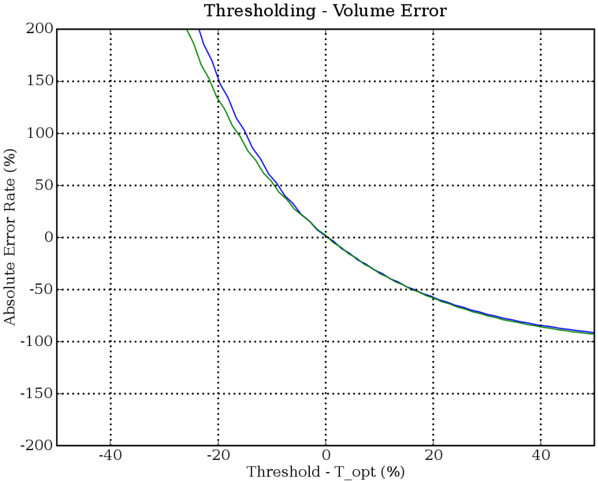
# Analysis of Training Examples

Intensity histograms of all voxels.



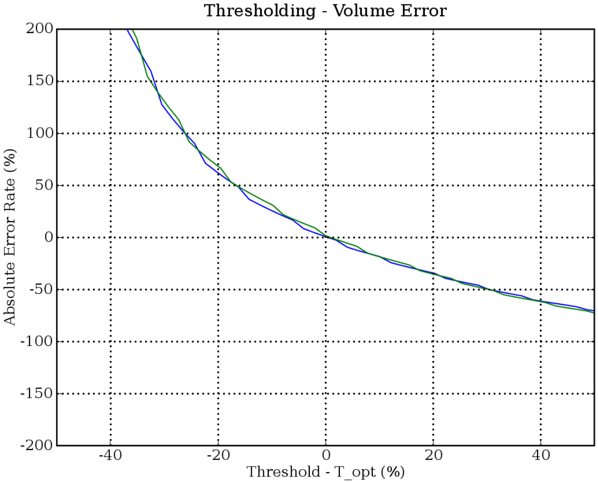
# Analysis of Hard Thresholding

ITAO



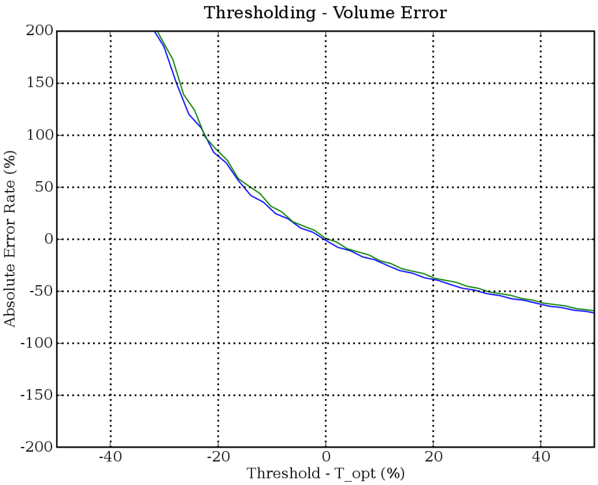
# Analysis of Hard Thresholding

ITEC



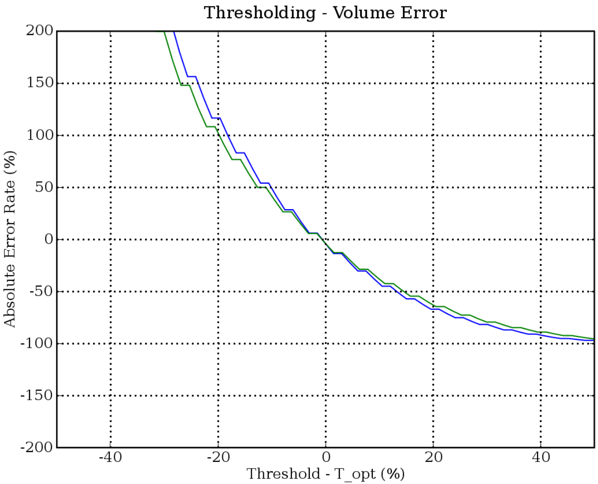
# Analysis of Hard Thresholding

ITMB03



# Analysis of Hard Thresholding

ITMB29

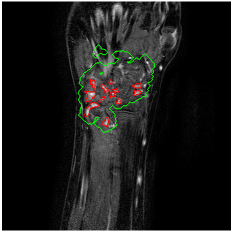


# Experimental Setup

for each exam

- use the other exams for training and validation;
- train 5 classifiers on 1.000 voxels;
- validate each classifier on 5 different samples of 1.000 voxels;
- combine the classifiers by bagging (averaging);
- test the classifiers on 20.000 voxels;

the exemplar voxels are drawn from the annotated synovia (red contour) and surrounding regions (green contour).


# Normalization by Ranking

examples from different exams have different intensity histograms

- one exam is randomly chosen as reference;
- the intensities of its voxels are sorted in an ordered list $I_r = (i_0, \ldots, i_N)$, where $N$ is the number of voxels and $i_j \leq i_{j+1}$ for $j = 0, \ldots, N$;
- given another exam, the intensity of each voxel is set to the intensity of its equal-rank voxel in the reference exam; that is, if a voxel has rank $k$ in the sorted list of intensities, then its new intensity will be $i_k$

# Results

Results for classifiers trained and tested on one observer. ***Normalized by ranking***. Tuning of the offset.

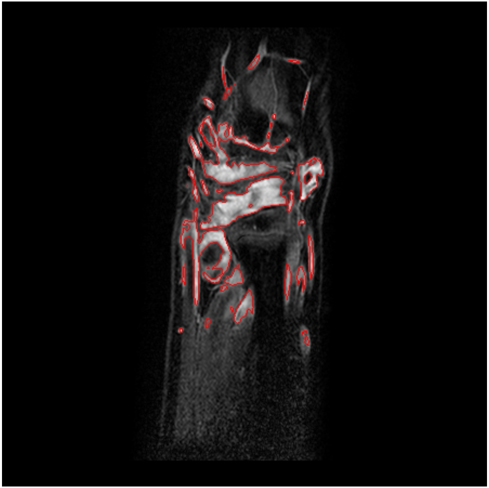
| | $b_0$ | $b_0 + 0.1$ | $b_0 + 0.2$ | $b_0 + 0.3$ | $b_0 + 0.4$ |
|---|---|---|---|---|---|
| | Classification Error | | | | |
| ITAO | 8.7% | 8.9% | 9.1% | 9.6% | 10.4% |
| ITEC | 8.2% | 7.8% | 7.6% | 7.4% | 7.6% |
| ITMB03 | 7.7% | 7.3% | 7.6% | 8.0% | 8.4% |
| ITMB29 | 8.7% | 8.9% | 8.8% | 9.4% | 9.6% |
| mean | 8.3% | 8.2% | 8.3% | 8.6% | 9.0% |
| $\sigma$ | 0.48% | 0.81% | 0.79% | 1.07% | 1.24% |
| | Volume Error | | | | |
| ITAO | −38.4% | −21.1% | −12.3% | 6.4% | 22.6% |
| ITEC | −49.3% | −42.7% | −36.0% | −24.8% | −11.3% |
| ITMB03 | −27.8% | −16.2% | −2.1% | 11.2% | 26.4% |
| ITMB29 | −34.8% | −20.8% | −6.5% | 5.1% | 25.1% |
| mean | −37.6% | −25.2% | −14.2% | −0.5% | 15.7% |
| $\sigma$ | 9.0% | 11.9% | 15.1% | 16.4% | 18.1% |

# Normalization Comparisons

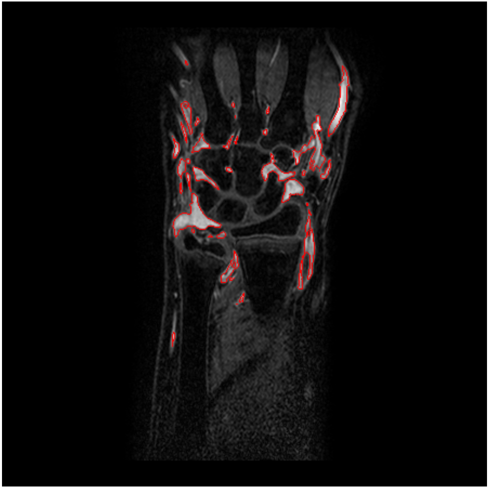
Results for classifiers trained and tested on one observer. Different normalizations.

| | - | Scaling | Ranking |
|---|---|---|---|
| | Classification Error | | |
| ITAO | 9.8% | 8.5% | 8.7% |
| ITEC | 9.3% | 7.7% | 8.2% |
| ITMB03 | 7.8% | 7.0% | 7.7% |
| ITMB29 | 8.7% | 8.0% | 8.7% |
| mean | 8.9% | 7.8% | 8.3% |
| $\sigma$ | 0.86% | 0.63% | 0.48% |
| | Volume Error | | |
| ITAO | 1.0% | $-53.6\%$ | $-38.4\%$ |
| ITEC | $-64.6\%$ | $-9.7\%$ | $-49.3\%$ |
| ITMB03 | $-71.4\%$ | $-47.2\%$ | $-27.8\%$ |
| ITMB29 | $-6.4\%$ | $-43.1\%$ | $-34.8\%$ |
| mean | $-35.4\%$ | $-38.4\%$ | $-37.6\%$ |
| $\sigma$ | 37.9% | 19.6% | 9.0% |

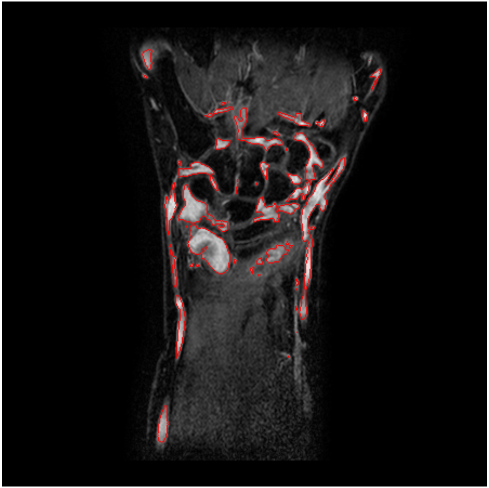
# Results on Unlabeled Data



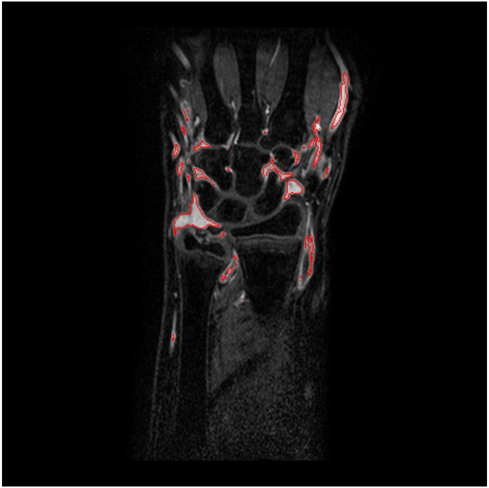
# Results on Unlabeled Data



# Results on Unlabeled Data

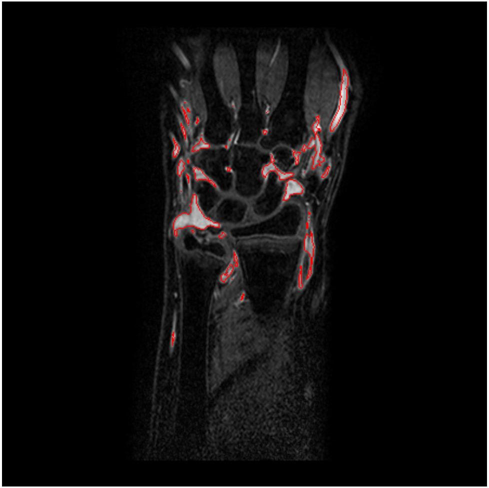


# Offset Tuning



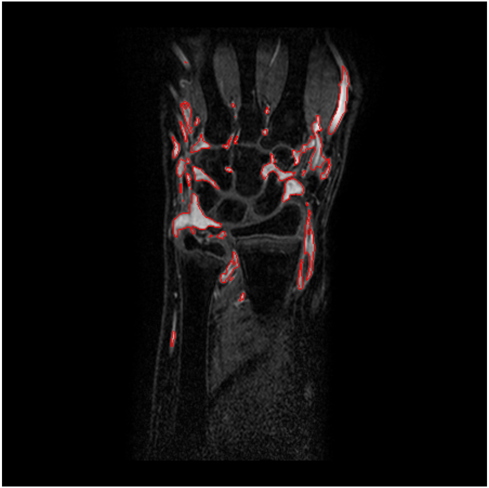
$$b = b_0 - 0.3$$

# Offset Tuning



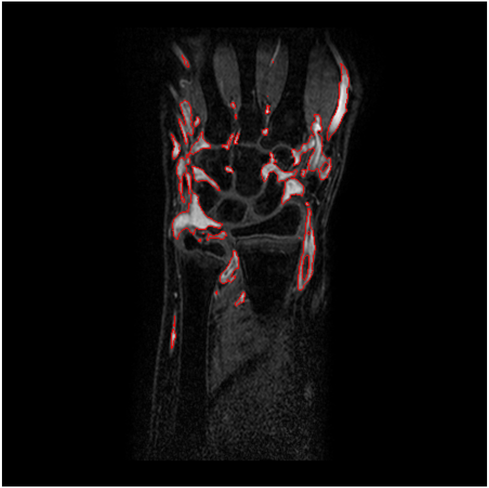
$$b = b_0$$

# Offset Tuning



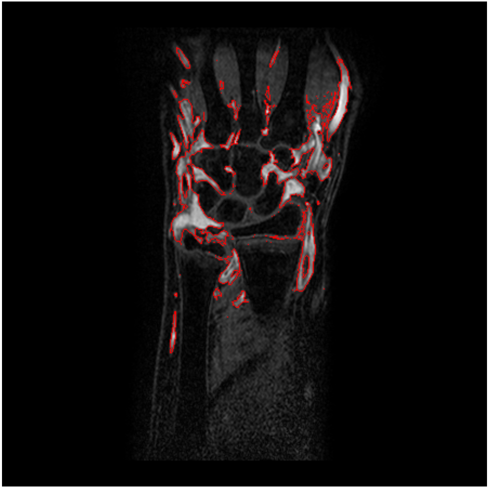
$$b = b_0 + 0.3$$

# Offset Tuning



$$b = b_0 + 0.6$$

# Offset Tuning



$$b = b_0 + 0.9$$

# Next Steps

Other enhanced structures (blood vessels, sinew) are selected.
They have to be removed

Mixed approach:

- some structures (e.g. vessels) might be detected based on their shape, and can be removed automatically;
- however, there might be structures more difficult to detect, which will have to be removed manually.