Automated Assessment of Joint Synovitis Activity from Medical Ultrasound and Power Doppler Examinations Using Image Processing and Machine Learning Methods

MEDUSA

Prof. dr hab. Konrad Wojciechowski, Principal Investigator
Project partners

- Silesian University of Technology, Gliwice
- Polish Japanese Academy of Information Technology, Warsaw, Bytom
- Institute of Medical Technology and Equipment ITAM, Zabrze
- Helse Førde, Førde
- Høgskulen i Sogn og Fjordane, Førde

The ultrasound images for the MEDUSA project were created at the Section for Rheumatology; Central Hospital, Forde, Norway.
Outline: project MEDUSA

- Aim of the project
- Objectives of the MEDUSA Project
- Ultrasonographic pictures
- Grading of the synovitis
- Activity planning and challenges
- Idea
- Tasks
- Annotation editor
- Preprocessing
- Classifiers
- Results
- Medical verification
- Synovitis detection
- Results meaning
- Conclusions
Aim of the MEDUSA project

- Development of computer systems that can recognize and grade synovitis (joint inflammation) in USG
- Automatic detection of synovitis
- Automatic grading of hypertrophy
- Limited for MCP and PIP joints of fingers 2 to 5
- Photos taken in the medial line of the joint

Project implementation period 15th July 2013 - 30th October 2016
Amount of funding: 3 910 509,00 PLN
Ultrasonographic pictures
Grading of the synovitis

- Inflammation scale: 0 - 3
- Blood flow (in chosen area) scale: 0 - 3
- Very subjective diagnosis
Grading of the synovitis

0: no joint cavity widening (<0.5 mm)
1: minimal (0.5 - 1.9 mm)
2: middle (2 - 4 mm)
3: marked (> 4 mm)

0: no hypertrophy
1: minimal hypertrophy
2: moderate hypertrophy
3: extensive hypertrophy
Tasks

- Collecting pictures
- Preparation of annotation group and software
- Making annotation on ~3000 USG pictures
- Learning process
  - Filtration, preprocessing
  - Feature extraction
  - Classification
  - Evaluation
- Medical verification (38 pictures, 3 physicians)
Annotation editor

- Pictures taken by GE LOGIQ S8 Ultrasound machine and saved in database
- Marking different areas and set arthritis rating
Main method

• Joint detector + skin / bone detector = structural description
• Structural description $\rightarrow$ features for synovitis area extraction based on pixel classification.
• Structural description + synovitis area = features for inflammation level detection using classification.
Joint detector: Preprocessing

- Serie of image processing operation
- Enhance an image characteristic
- Prepare image for feature extraction suitable for specific classifier
Joint detector: Classifiers

- Nearest neighbor
- SVM
- Decision tree

Evaluation:
- ROC: true positive rate, false positive rate
- AUC analysis: trapezoid area sum
- Detector with greatest AUC value wins.
<table>
<thead>
<tr>
<th>mark</th>
<th>preparation</th>
<th>Feature extraction</th>
<th>classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM1</td>
<td>Gray scale, hist. equalization, Gaussian blur</td>
<td>SURF mixture</td>
<td>SVM</td>
</tr>
<tr>
<td>SVM2</td>
<td>Gray scale</td>
<td>SURF mixture</td>
<td>SVM</td>
</tr>
<tr>
<td>SVM3</td>
<td>Gray scale, hist. equalization, Gaussian blur</td>
<td>SURF</td>
<td>SVM</td>
</tr>
<tr>
<td>SVM4</td>
<td>Gray scale</td>
<td>SURF</td>
<td>SVM</td>
</tr>
<tr>
<td>SVM5</td>
<td>Gray scale, Gaussian blur</td>
<td>SURF</td>
<td>SVM</td>
</tr>
<tr>
<td>SVM6</td>
<td>Gray scale, Gaussian blur</td>
<td>SURF mixture</td>
<td>SVM</td>
</tr>
<tr>
<td>NN1</td>
<td>Gray scale, hist. equalization, Gaussian blur</td>
<td>SURF mixture</td>
<td>NN</td>
</tr>
<tr>
<td>NN2</td>
<td>Gray scale</td>
<td>SURF mixture</td>
<td>NN</td>
</tr>
<tr>
<td>NN3</td>
<td>Gray scale, hist. equalization, Gaussian blur</td>
<td>SURF</td>
<td>NN</td>
</tr>
<tr>
<td>NN4</td>
<td>Gray scale</td>
<td>SURF</td>
<td>NN</td>
</tr>
<tr>
<td>DT1</td>
<td>Gray scale, hist. equalization, Gaussian blur</td>
<td>SURF mixture</td>
<td>DT</td>
</tr>
<tr>
<td>DT2</td>
<td>Gray scale</td>
<td>SURF mixture</td>
<td>DT</td>
</tr>
<tr>
<td>DT3</td>
<td>Gray scale, hist. equalization, Gaussian blur</td>
<td>SURF</td>
<td>DT</td>
</tr>
<tr>
<td>DT4</td>
<td>Gray scale</td>
<td>SURF</td>
<td>DT</td>
</tr>
</tbody>
</table>
## Joint detector: Results – AUC

<table>
<thead>
<tr>
<th>SVM1</th>
<th>SVM2</th>
<th>SVM3</th>
<th>SVM4</th>
<th>SVM5</th>
<th>SVM6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,966</td>
<td>0,981</td>
<td>0,961</td>
<td>0,975</td>
<td>0,978</td>
<td>0,981</td>
</tr>
<tr>
<td>NN1</td>
<td>NN2</td>
<td>NN3</td>
<td>NN4</td>
<td>DT1</td>
<td>DT2</td>
</tr>
<tr>
<td>0,911</td>
<td>0,928</td>
<td>0,912</td>
<td>0,925</td>
<td>0,710</td>
<td>0,749</td>
</tr>
<tr>
<td>DT3</td>
<td>DT4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,889</td>
<td>0,877</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### SVM2
- Gray scale

### SVM6
- Gray scale, Gaussian blur

### SURF mixture
- SVM
Bone and Skin: filtering

10 filters used: Gaussian smoothing, first and second derivative, Laplacian as well as different threshold filters on them.
Bone and Skin: samples selection
Bone and Skin: features generation
Bone and Skin: visual results

Exemplary detections using Random Forest classifier
## Bone and Skin: evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>Time (Learn)</th>
<th>Time (Predict)</th>
<th>Precision</th>
<th>Recall</th>
<th>Fall-out (Err)</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-NN</td>
<td>6s</td>
<td>2.5min</td>
<td>0.87</td>
<td>0.83</td>
<td>0.11</td>
</tr>
<tr>
<td>Decision Trees</td>
<td><strong>14s</strong></td>
<td><strong>1s</strong></td>
<td><strong>0.95</strong></td>
<td><strong>0.97</strong></td>
<td><strong>0.05</strong></td>
</tr>
<tr>
<td>Naive Bayesian</td>
<td>0.5s</td>
<td>3.5s</td>
<td>0.78</td>
<td>0.99</td>
<td>0.26</td>
</tr>
<tr>
<td>SVM</td>
<td>7.5min</td>
<td>16s</td>
<td>0.91</td>
<td>0.78</td>
<td>0.07</td>
</tr>
<tr>
<td>Random Forest</td>
<td><strong>8s</strong></td>
<td><strong>6s</strong></td>
<td><strong>0.98</strong></td>
<td><strong>0.84</strong></td>
<td><strong>0.02</strong></td>
</tr>
<tr>
<td>Ada Boost</td>
<td>3.5min</td>
<td>1min</td>
<td>0.78</td>
<td>0.94</td>
<td>0.24</td>
</tr>
</tbody>
</table>
Synovitis region: features

- Pixel intensity values
- Histograms
- Distance to bone, skin, joint
- Gaussian blur
Synovitis region: training

• Random Forest classifier
  Selected as the best compromise between accuracy and training/detection speed among others: k-Nearest Neighbours, Decision Trees, Naive Bayesan, SVM, Adaboost

• Training and evaluation
  237 annotated pictures used for training
  94 annotated pictures used for testing (evaluation)

False Discovery Rate:  \( FDR = \frac{FalsePositives}{FalsePositives+TruePositives} \)
Synovitis region: post-processing

- Synovitis pixels detection
- Aggregation filters
- Merging regions
- Final result
Inflammation Level detection

• Features for synovitis area recognized in previous step:
  • Geometrical moments:
    • area,
    • hull area,
    • ellipse axis,
  • Structural moments:
    • Sum and average distance to bone
    • Sum and average distance to skin
    • Sum and average distance to joint
Inflammation Level detection

- Feature vector is extracted for each image:
  - 121 pictures in training set
  - 204 pictures in evaluation set.
- Optimization of parameters and training set with cross-validation
- Tests on evaluation set.
IL detection results

Measures:
- **accuracy** = true positive count / all count
- **misclassification error** is average difference between recognised and real inflammation level;

Confusion matrix:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>28</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>42</td>
</tr>
<tr>
<td>5</td>
<td>88</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

accuracy = 0.77451;
misclassification error = 0.254902;
Medical verification

• Qualified rheumatologist using our GE LOGIQ S8 Ultrasound machine took images.
• We have collected 403 pictures from 20 subjects. All pictures were sent to developed dedicated database.
• 38 pictures was chosen to verification
• One of the pictures got grading respectively level 0, 3, and 1 from three experienced physicians, and 1 from software – what does it mean?
• Gold Standard
Inaccurate synovitis detection - marked area under bone lines
Inaccurate synovitis detection – synovitis area underestimated
Correct detection
Results meaning

- **Weighted Kappa Cohen**

<table>
<thead>
<tr>
<th>Observer Comparison</th>
<th>Kappa Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>observer 1 - observer 2</td>
<td>0.588</td>
</tr>
<tr>
<td>observer 1 - observer 3</td>
<td>0.513</td>
</tr>
<tr>
<td>observer 2 - observer 3</td>
<td>0.727</td>
</tr>
<tr>
<td>observer 1 - software</td>
<td>0.370</td>
</tr>
<tr>
<td>observer 2 - software</td>
<td>0.361</td>
</tr>
<tr>
<td>observer 3 - software</td>
<td>0.326</td>
</tr>
<tr>
<td>observer 1 – gold standard</td>
<td>0.682</td>
</tr>
<tr>
<td>observer 2 – gold standard</td>
<td>0.895</td>
</tr>
<tr>
<td>observer 3 – gold standard</td>
<td>0.832</td>
</tr>
<tr>
<td>software - gold standard</td>
<td>0.380</td>
</tr>
</tbody>
</table>
Results meaning

- Observer 2 - Observer 3
- Observer 1 - Observer 3
- Observer 1 - Observer 2

Categories: more, ok, less
Results meaning
Conclusions

• Automated system gives acceptable results.
• Correlations is not very good but it is definitely much better than random results.
• For every possible weights for kappa Cohen coefficient physicians have ‘better’ results.
• Preliminary software makes mistakes like:
  • - detection under bones
  • - fragmented inflammation area
  • - inflammation area is away from joint
• All the above problems are not physically possible and can be easy(?) eliminated in next version.
Thank you for your attention