A generic Framework for the Evaluation of content-based Image and Video Analysis Tasks in the Core Technology Cluster of THESEUS

Stefanie Nowak
Peter Dunker, Ronny Paduschek
Fraunhofer Institute for Digital Mediatechnology
Ilmenau, Germany

Aarhus, 16.09.2008
nwk@idmt.fraunhofer.de
Outline

» THESEUS

» Image and Video analysis in THESEUS

» Contests

» Concept

» Evaluation Framework

» First evaluations

» Image Segmentation

» Face Detection

» Conclusions
THESEUS

» German research program  
  (financed by Federal Ministry of Economy and Technology (BMWi))

» Focus: algorithms and web-based infrastructure for acquiring, processing and seeking of knowledge available from the web
  » Core Technology Cluster (CTC)
  » Use cases

» research on  
  text recognition, privacy, ontologies, user interfaces, video and image analysis, **evaluation strategies**, visualization techniques, machine learning

» Duration: 5 years
» Workpackage: Evaluation: Picture Analysis
» Lead: Fraunhofer Institute for Digital Mediatechnology (IDMT)

» Image and Video Analysis:
  » Image and video identification
  » Image and video similarity retrieval
  » Temporal shot, sub shot, scene change detection
  » Video genre detection
  » Video event detection
  » Image segmentation and spatio-temporal segmentation in videos
  » Image scene classification and annotation
  » Named entity retrieval
  » Face detection
  » ...

Goals for evaluation

» Measure improvement over time
» Comparison to other state-of-the-art algorithms
» Test with independent and unknown data

» Evaluation criteria:
  » Recognition rates
  » Processing times
  » Scalability
  » Robustness (e.g. regarding distortions)

» Functional range:
  » e.g. number of categories in categorization tasks
  » Supported formats, use of standards
<table>
<thead>
<tr>
<th>CONTEST</th>
<th>AREA</th>
<th>TASK</th>
</tr>
</thead>
<tbody>
<tr>
<td>PASCAL: Visual Object Class Challenge</td>
<td>Image</td>
<td>VOC 2008:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) Classification (presence / absence of objects)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) Object Detection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3) Pixel-wise object segmentation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4) Person Layout</td>
</tr>
<tr>
<td>Caltech Challenge</td>
<td>Image</td>
<td>Caltech 2007: Classification (1 out of 256)</td>
</tr>
<tr>
<td>ImageCLEF</td>
<td>Cross Language Image Retrieval</td>
<td>ImageCLEF 2008:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) Photographic Retrieval</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) Medical Retrieval</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3) Photographic Concept Detection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4) Automatic Medical Image Annotation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5) Image Retrieval from a Collection of Wikipedia Images</td>
</tr>
</tbody>
</table>
## Multimedia Analysis Contests

<table>
<thead>
<tr>
<th>CONTEST</th>
<th>AREA</th>
<th>TASK</th>
</tr>
</thead>
<tbody>
<tr>
<td>ImagEVAL</td>
<td>Image</td>
<td>ImagEVAL 2006:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) Recognition of transformed images</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) Text/Image mixed research</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3) Detection of text areas</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4) Detection of objects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5) Semantics Extraction</td>
</tr>
<tr>
<td>Berkely Segmentation</td>
<td>Image</td>
<td>Image segmentation and boundary detection</td>
</tr>
<tr>
<td>Benchmark</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREC Vid</td>
<td>Video</td>
<td>TREC Vid 2008:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) Surveillance event detection pilot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) High level feature extraction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3) Search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4) Rushes summarization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5) Content-based copy detection</td>
</tr>
<tr>
<td>VideOlympics</td>
<td>Video</td>
<td>Showcase: Video Retrieval</td>
</tr>
</tbody>
</table>
» Evaluation with international accepted databases and standards

» Independent evaluation with data that is not available for the algorithm developers

» Creation of testdata with standard distortions

» Comparison to state-of-the-art algorithms

» Development of a generic Evaluation Framework

» Eventually organizing tasks in benchmark related to research in Theseus
A generic evaluation framework will be developed to handle and measure all evaluations.

The key features of the framework are:

- easy extension to new formats and measures
- storing previous test results for comparison and measurement of improvements
- sophisticated visualizations for interactive reviewing
- generation of descriptive test results.
» 3 Test Cases:
  » 1) Retrieval
    » Input: Multimedia Document
    » Output: list of similar documents
    » search scenarios
  » 2) Keyword or Segment Indexing
    » Input: one media item
    » Output: description of item (holistic annotations, segment information, segment-based annotations)
    » face/object detection, classification
  » 3) Multimedia Enhancement
    » Input: multimedia document
    » Output: enhanced multimedia document
    » automatic distortions correction
Evaluation Framework

» Evaluation Manager
» Convert Input Data Module
» Load Data Module
» Visualization Module
» Evaluation Measures
» Significance Tests
Test Data

- Berkeley Segmentation Dataset (1)
  - 100 images for tests
  - 5-10 ground truth segmentations per image
- [Another test corpus with about 1000 images and 1 ground truth segmentation per image] → to be extended

Evaluation Measures

- 1) boundary-based measure from Berkeley Segmentation Benchmark (1)
- 2) region-based measure based on Normalized Hamming Distance (2)

Evaluation: Image Segmentation

» Why integrating an already existing benchmark?
  » Ground Truth
    » The ground truth was carefully collected
    » Subjectivity was minimized
    » Different granularity of segmentations are present
      → mirror different user expectations

» Comparison
  » Results are published on a website
  » Comparison from THESEUS to other segmentation results
  » Using same testing conditions
Normalized Hamming Distance

» Segmentation $S$, Ground Truth $G$

\[ S = \{ R_1^1, R_1^2, \ldots, R_1^m \} \quad G = \{ R_2^1, R_2^2, \ldots, R_2^n \} \]

» Directional Hamming Distance:

\[ D_H(S \Rightarrow G) = \sum_{R_2^i \in G} \sum_{R_1^k \neq R_1^j, R_1^k \cap R_2^i \neq \emptyset} |R_2^i \cap R_1^k| \]

» Normalized Hamming Distance:

\[ p = 1 - \frac{D_H(S \Rightarrow G) + D_H(G \Rightarrow S)}{2 \times |S|} \]

» Missing Rate:

\[ E^m_R = \frac{D_H(S \Rightarrow G)}{|S|} \]

» False Alarm Rate:

\[ E^f_R = \frac{D_H(G \Rightarrow S)}{|S|} \]
Evaluation: Image Segmentation

Graphical User Interface

[Image of a graphical user interface with a window labeled "imag_view" showing a horse on the left and a scatter plot on the right.]
Evaluation: Face Detection

» Test Data
  » >350 images
  » 1000 faces
    » Manually annotated
    » Set will be increased in the future
  » Ground Truth: Bounding Box

» How to relate bounding box for detected face and ground truth?
  » Position of BB: Euclidean Distance \( \Delta dist_{xy} < \xi w_1 \), with \( \xi = 0.5 \)
  » Relation of sizes of BB:

\[
 w' = \frac{|w_2 - w_1|}{w_1} \quad w' < \psi, \quad \text{with} \quad \psi = 0.4
\]
Evaluation: Face Detection

Result visualization:
Future Developments

» Generic evaluation framework for evaluation of image and video analysis algorithms
  » Complemented throughout THESEUS

» No duplication of existing contests
  » Submit algorithms to contests
    » E.g. image classification to VOC 2008
    » Video identification to TRECVid (?)
  » Organize task in Benchmark
Thank you

http://theseus-programm.de/
http://www.idmt.fraunhofer.de