



THESEUS

Forschungsprogramm für eine
neue internetbasierte Wissensinfrastruktur

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





- » THESEUS
- » Image and Video analysis in THESEUS
- » Contests
- » Concept
- » Evaluation Framework
- » First evaluations
 - » Image Segmentation
 - » Face Detection
- » Conclusions



- » German research program
(financed by Federal Ministry of Economy and Technology (BMWii))
- » 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
 - » ...



- » 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



CONTEST	AREA	TASK
PASCAL: Visual Object Class Challenge	Image	VOC 2008: 1) Classification (presence / absence of objects) 2) Object Detection 3) Pixel-wise object segmentation 4) Person Layout
Caltech Challenge	Image	Caltech 2007: Classification (1 out of 256)
ImageCLEF	Cross Language Image Retrieval	ImageCLEF 2008: 1) Photographic Retrieval 2) Medical Retrieval 3) Photographic Concept Detection 4) Automatic Medical Image Annotation 5) Image Retrieval from a Collection of Wikipedia Images



CONTEST	AREA	TASK
ImagEVAL	Image	ImagEVAL 2006: <ol style="list-style-type: none">1) Recognition of transformed images2) Text/Image mixed research3) Detection of text areas4) Detection of objects5) Semantics Extraction
Berkely Segmentation Benchmark	Image	Image segmentation and boundary detection
TRECVID	Video	TRECVID 2008: <ol style="list-style-type: none">1) Surveillance event detection pilot2) High level feature extraction3) Search4) Rushes summarization5) Content-based copy detection
VideOlympics	Video	Showcase: Video Retrieval



- » 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

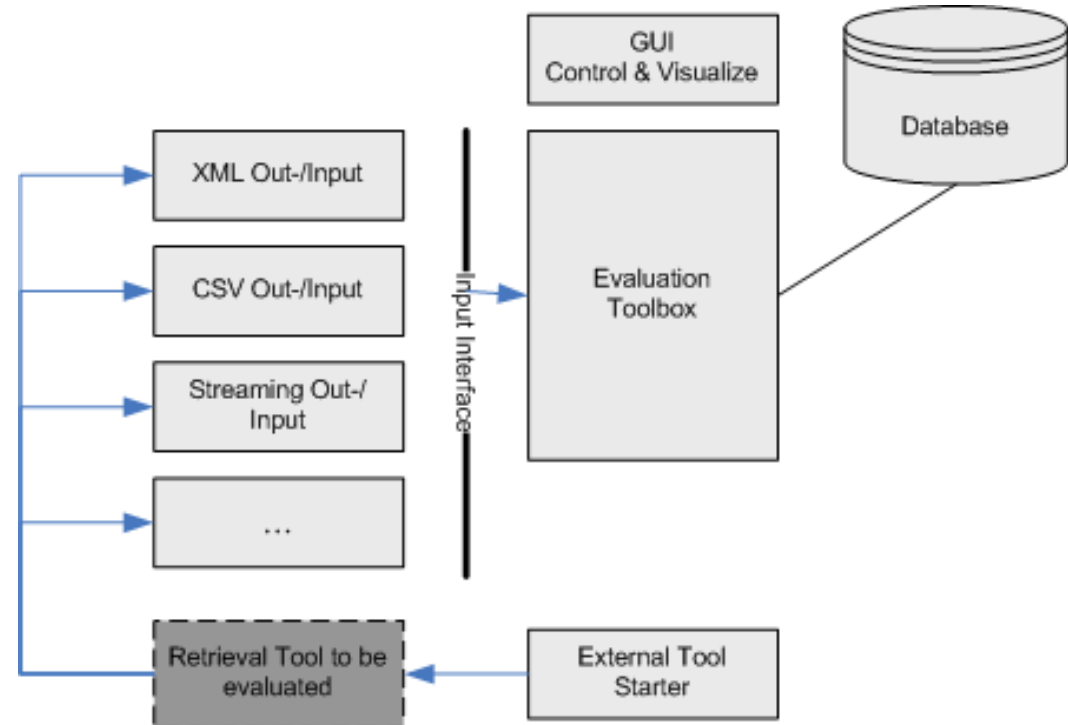
- » The key features of the framework are:
 - » easy extension to new formats and measures
 - » storing previous test results for comparison and measurement of improvements
 - » sophisticated visualizations for interactive reviewing
 - » generation of descriptive test results.

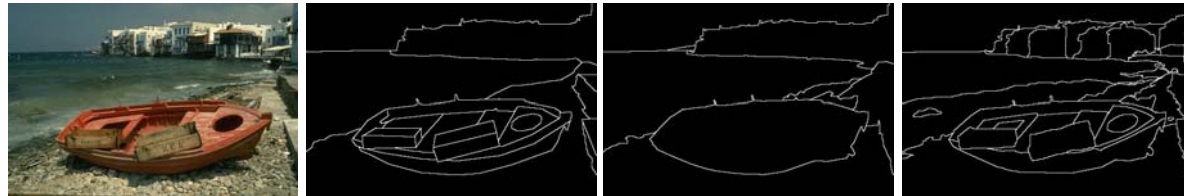


- » 3 Test Cases:
 - » 1) Retrieval
 - » Input: Multimedia Document
 - » Output: list of similar documents
 - » → search scenarios
 - » 2) Keyword or Segment Indexing
 - » Input: one media item
 - » Output: description of item (holistic annotations, segment information, segment-based annotations)
 - » → face/object detection, classification
 - » 3) Multimedia Enhancement
 - » Input: multimedia document
 - » Output: enhanced multimedia document
 - » → automatic distortions correction



- » Evaluation 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)

- » (1) D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *Proc. Eighth Int'l Conf. Computer Vision*, vol. 2, pp. 416–423, 2001.
- » (2) Q. Huang and B. Dom, "Quantitative methods of evaluating image segmentation," *IEEE International Conference on Image Processing*, vol. 3, pp. 53–56, 1995.



- » 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



- » Segmentation S , Ground Truth G

$$S = \{R_1^1, R_1^2, \dots, R_1^m\} \quad G = \{R_2^1, R_2^2, \dots, R_2^n\}$$

- » 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:

$$p = 1 - \frac{D_H(S \Rightarrow G) + D_H(G \Rightarrow S)}{2 * |S|}$$

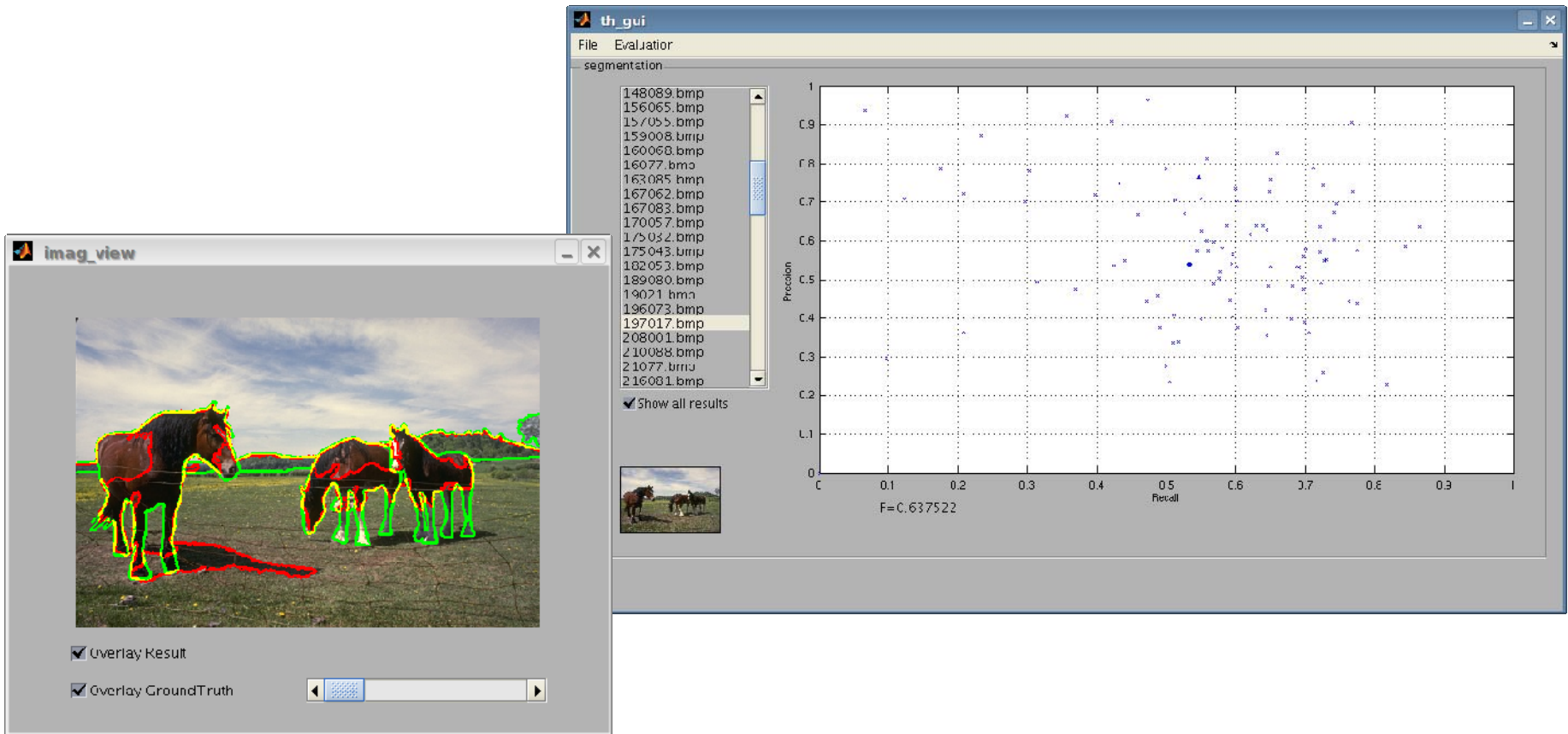
- » Missing Rate:

$$E_R^m = \frac{D_H(S \Rightarrow G)}{|S|}$$

- False Alarm Rate:


$$E_R^f = \frac{D_H(G \Rightarrow S)}{|S|}$$


Graphical User Interface

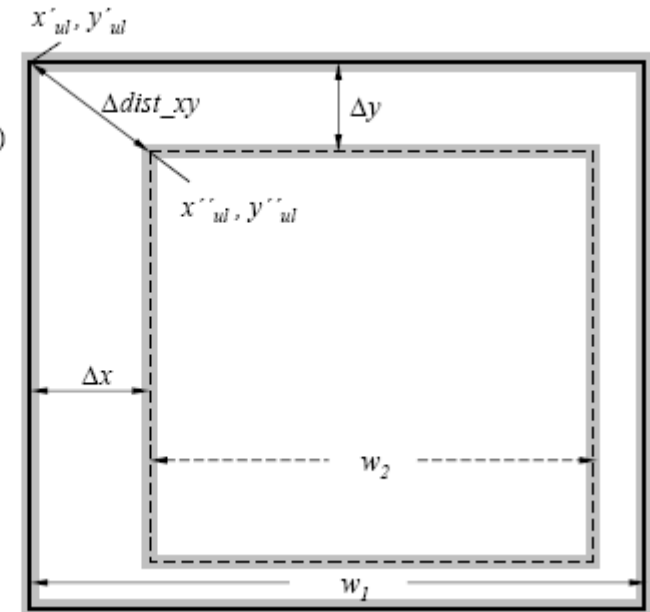


» Test Data

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

 bounding box (ground truth)

 bounding box (detected)



» 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:

$$w' = \frac{|w_2 - w_1|}{w_1} \quad w' < \psi, \quad \text{with } \psi = 0.4$$



Result visualization:

The screenshot shows the 'th_gui' application window. The main area displays an image of two men, with two green bounding boxes around their faces. Below the image are navigation buttons: 'Last Image', '33', 'Go', and 'Next Image'. There are also checkboxes for 'Draw Ground Truth' and 'Draw Detected Faces', and a small table showing counts for Ground Truth, True Positives, False Positives, and False Negatives.

	TPos	FPos	FNeg	MultiTP
1	2	0	0	0
2	2	0	0	0
3	2	0	0	0
4	3	0	0	0
5	2	0	0	0
6	3	0	0	0
7	9	0	0	0
8	1	0	0	0
9	2	0	0	0
10	2	0	1	0
11	3	0	0	0
12	2	0	0	0
13	2	0	0	0
14	1	0	0	0
15	1	0	0	0
16	8	1	0	0
17	1	0	1	0
18	1	0	0	0
19	3	0	1	0
20	2	0	1	0
21	1	0	0	0
22	2	0	0	0
23	1	0	1	0
24	3	1	0	0
25	1	0	0	0
26	1	0	0	0
27	2	0	0	0
28	1	0	0	0
29	1	0	0	0
30	0	1	1	0

The right-hand panel contains a 'Results Summary' section with the following values:

- GroundTruth: 943
- (Multi)TruePos: 705
- FalsePos: 92
- FalseNeg: 238
- Precision: 0.88457
- Recall: 0.74761
- DetectTime: 229.558

Below the summary are buttons for 'Choose Result Directory', 'Save Results', 'Save Drawed Images', and 'Choose Image Directory'.

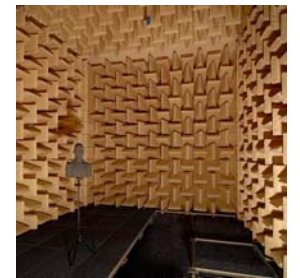


- » Generic evaluation framework for evaluation of image and video analysis algorithms
 - » Complemented throughout THESEUS

- » 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>