ImageCLEF Large Scale – Visual Concept Detection Task

Evaluation of multilabel image annotation incorporating domain knowledge and concept subjectivity

Stefanie Nowak

Semantic Audiovisual Systems

Fraunhofer IDMT

Brussels, 27.05.2009

stefanie.nowak@idmt.fraunhofer.de

Fraunhofer IDMT

THESEUS

- German research program (financed by BMWi)
- research on
  text recognition, ontologies, user interfaces, video and image analysis,
  evaluation strategies, visualization techniques, machine learning …

- CTC 8.4 Picture Analysis:
  Evaluation of photo and video analysis
  - Objective evaluation from third party
  - Measurement of improvement
  - Unknown datasets
  - International comparison

  → Organization of a task in international benchmark

THESEUS
New Technologies for the Internet of Services

Supported by
Federal Ministry of Economics and Technology

Fraunhofer IDMT

ImageCLEF

- Evaluation track of CLEF (Cross-language evaluation forum)
- History:
  - 2003: first image retrieval task, 4 participants
  - 2004: 17 participants for three tasks (~200 runs)
  - 2005: 24 participants for fours tasks (~300 runs)
  - 2006: 30 participants for four tasks (~300 runs)
  - 2007: 35 participants for four tasks
  - 2008: 45 participants submitted results (>1000 runs)

ImageCLEF 2008 / 2009

- Participation 2008:
- Total of 63 groups registered for five tasks
  - Photo Retrieval: 24 groups, 1042 runs
  - Medical Retrieval: 15 groups, 111 runs
  - WikipediaMM Retrieval: 12 groups, 77 runs
  - Visual Concept Detection: 11 groups, 53 runs
  - Medical Image Annotation: 6 groups, 24 runs

- ImageCLEF 2009
  - 6 tasks
  - New task: Large scale visual concept detection
  - 38 groups registered for LS-VCDT

  → 2 THESEUS partners

Large Scale - Visual Concept Detection Task 2009

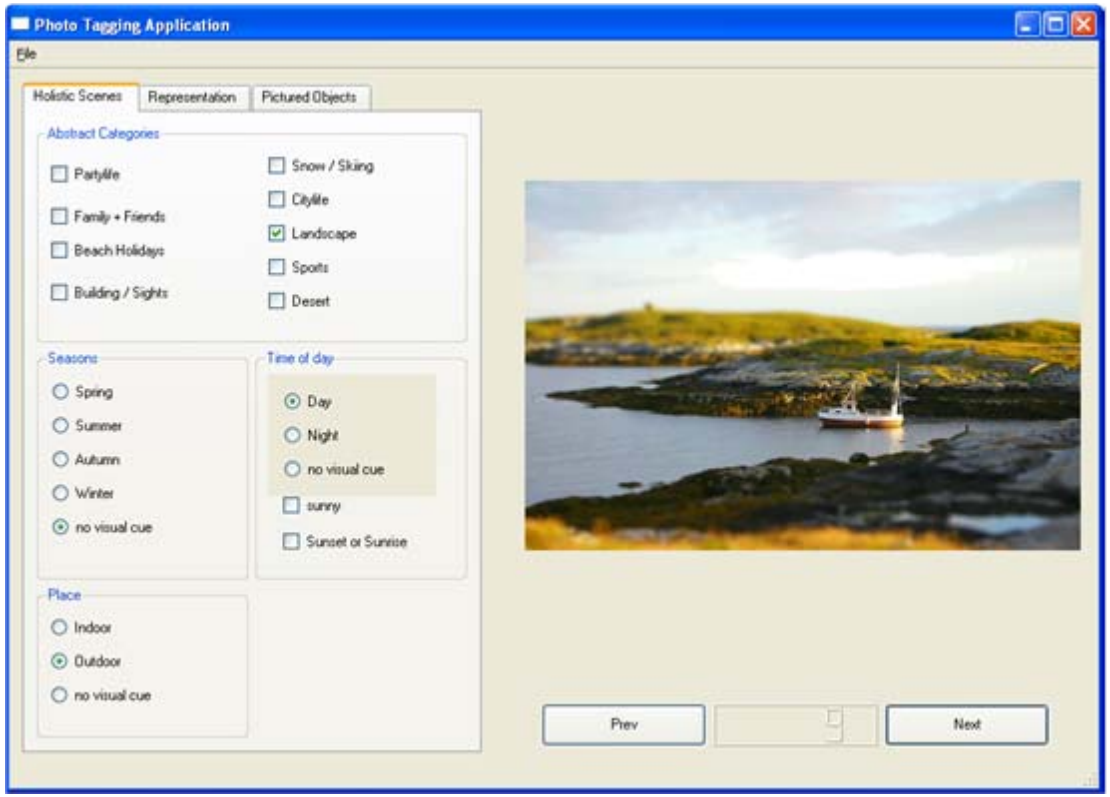
- Task:
  - Annotate the photos with all depicted visual concepts
  - Use provided real-world knowledge
- Main challenges:
  1) Can image classifiers scale to the large amount of concepts and data?
  2) Can an ontology (hierarchy and relations) help in large scale annotations?
- http://www.imageclef.org/2009/PhotoAnnotation

Large Scale - Visual Concept Detection Task 2009

```
Citylife
Outdoor
Night
Underexposed
Vehicle
No_Blur
No_Persons
No_Visual_Season
```

- Annotations
  - Multiple annotations
  - 53 visual concepts
  - Most: holistic visual concepts
  - Objective vs. subjective impression of annotators
  - Organization in a small ontology
  - Format
    - Plain text format
    - Rdf-xml
- Trainingset: 5.000 photos + ground truth annotations
- Testset: 13.000 photos

Ground Truth Acquisition



- MIR Flickr 25.000 image dataset
- C++ Tagging Tool
- Guideline for annotation
  - 1 of n concepts
  - Optional concepts
- Validation step (2 persons)

→ 18.000 photos annotated

→ 43 persons
(min 30 photos, max 2500 photos)

Fraunhofer IDMT

Ground Truth Acquisition – Validation Step*

- Well-annotated concepts
  - Top 5:
    - Outdoor
    - No visual season
    - Small Group, No Persons
    - Clouds
    - Sunny

- Difficult concepts
  - statistical:
    - Overexposed
    - Autumn
    - Lake
    - Winter
    - Out of focus
  - Number of changed annotations:
    - Partly blurred (378)
    - Landscape (266)
    - Macro (198)
    - Day (187)
    - Still Life (116)
    - Trees (93)

* The numbers refer to the validation of the trainingsset (5000 photos)

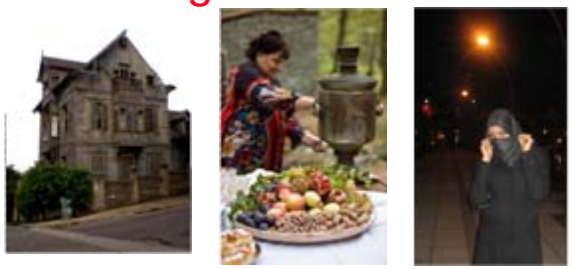
Fraunhofer IDMT

Validation – Problems in annotation

- Misunderstanding of photographic terms
  - Overexposed:

    correct:

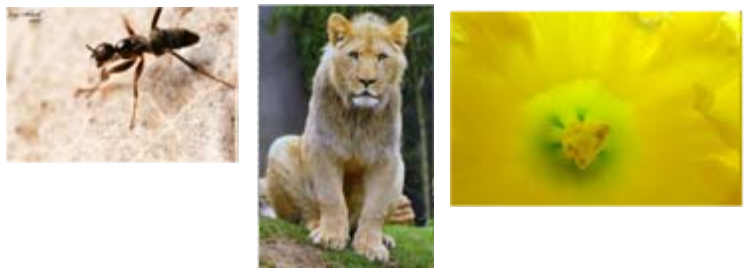
    wrong:



- Bad concept descriptions
  - Landscape / Nature

    should be:

    not:



- Semantic associations
  - E.g. Christmas tree in living room → winter
- What is really visible in the photos?

Ground Truth - Ambiguities

- How many persons are depicted?
  - Single?
  - Small group (2-5)?
  - Big group (> 5)?
  - No persons?

- Which photo is a portrait photo?

- Annotation Rules:
  - Parts of persons are no persons
  - Drawn persons are only persons in a canvas
  - Portrait is defined to depict persons or animals

Ontology

- 81 classes
  - 53 visual concepts + structural classes
  - E.g.
    - Scene Description
    - Representation
    - Illumination
    - Content Elements
    - …
- 19 object properties
- Expressivity of *ALCHIQ(D)*

Fraunhofer IDMT

Evaluation Measures

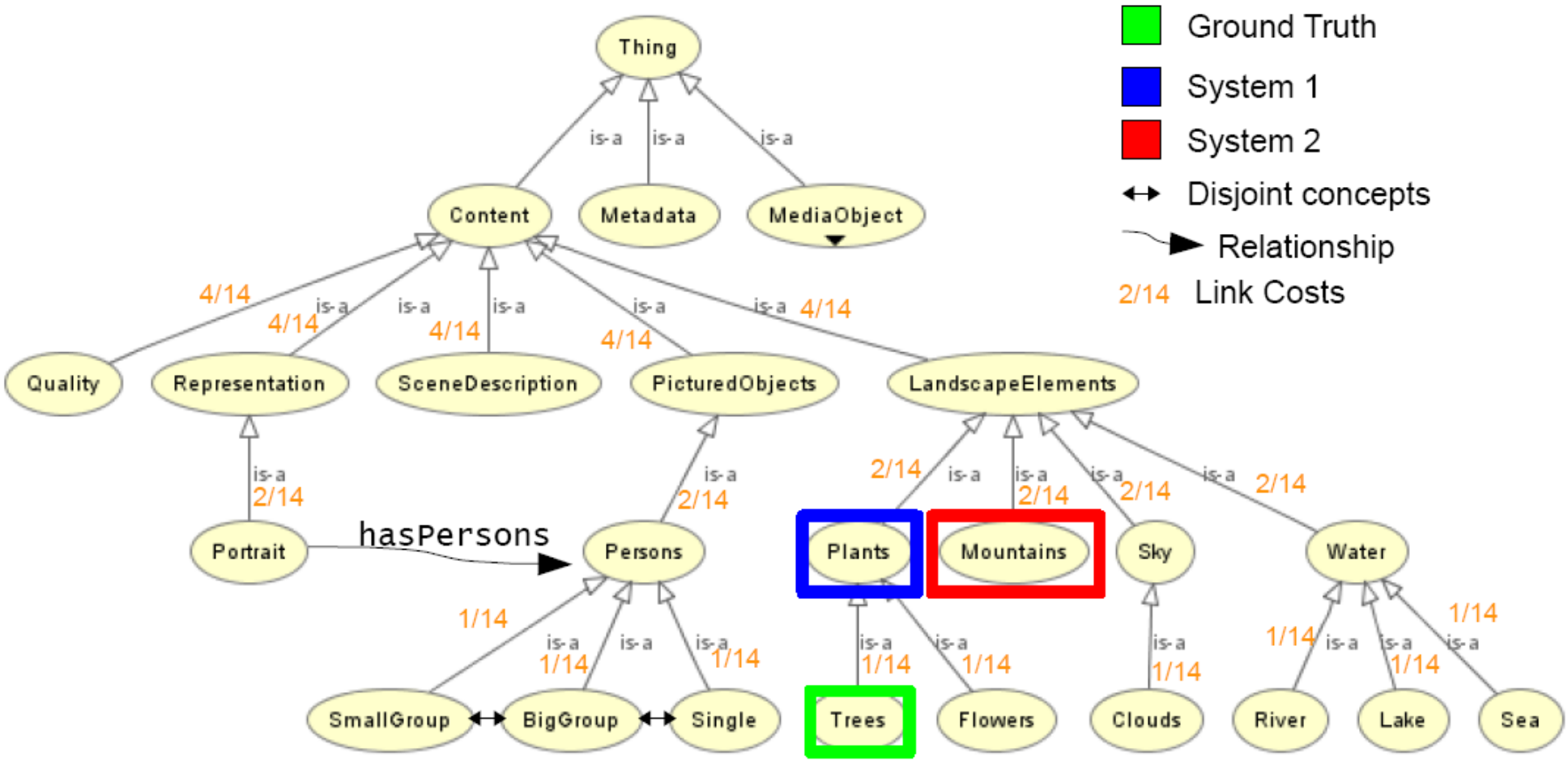
- 1) Evaluation per concept
  - EER and AUC
  - Same measure as in last years

  → No „real“ multi-label scenario evaluation

- 2) Evaluation per photo
  - Correlation between ground truth and annotated label set for each photo
    - Hierarchy of concepts
    - Domain knowledge
    - Annotator agreements

- 3) Processing Times

Evaluation of Multi-Label Annotations



[Stefanie Nowak,Hanna Lukashevich, Multilabel Classification Evaluation using Ontology Information, IRMLES Workshop 2009]

Fraunhofer IDMT

Evaluation of Multi-Label Annotations

- Predicted set of labels: P

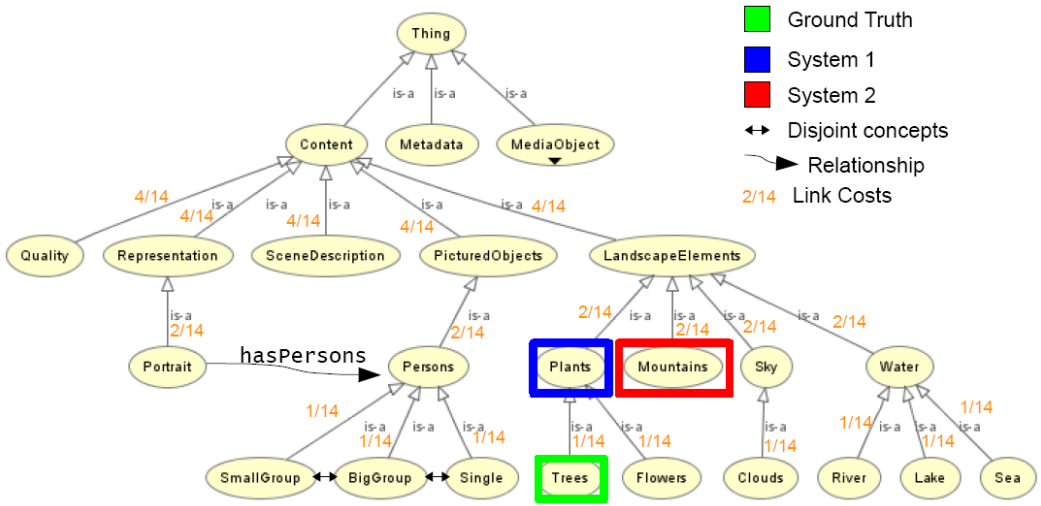
$$P' = P \setminus (P \cap G)$$

- Ground Truth set of labels: G

$$G' = G \setminus (P \cap G)$$

- Costs c_i for each link:

$$c_i = \frac{2^{(i-1)}}{2 \cdot \sum_{i=1}^{N} 2^{(i-1)}}$$



$$match(P, G) = \sum_{l_i \in P'} \left((\min_{l_j \in G} cost(l_i, l_j)) \cdot a(l_j{}^*) \right) + \sum_{l_j \in G'} \left((\min_{l_i \in P} cost(l_i, l_j)) \cdot a(l_j) \right)$$

$$\text{with } l_j{}^* = \text{argmin}_{l_j \in G}(cost(l_i, l_j))$$

$$score(X) = \left(1 - \frac{match(P, G)}{|P \cup G|} \right)^{\alpha}$$

Fraunhofer IDMT

Annotator Agreements

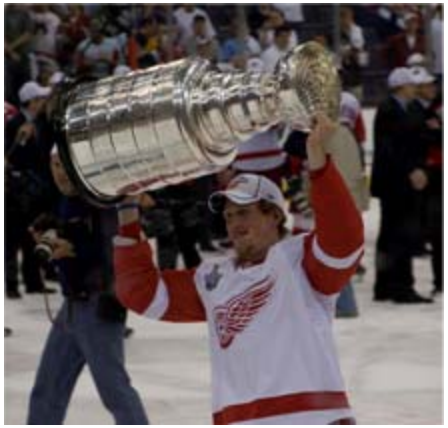
- How to interpret a decision of an annotator?
  - Optional concepts:
    - Tagging presence of concepts?
    - Tagging presence and absence of concepts?
  - 1 of n concepts
    - Annotator is forced to annotate one of the n concepts

- 100 photos were annotated by 11 persons
  - Mean over tagged optional concepts: 77,85%
  - Mean over optional concepts: 93,84%
  - Mean over 1 of n concepts: 92,47%

- Deleted concepts:
  - Post-Processed, HDR Image

Annotator Agreements – Photo View

- High Agreement on photo



	Sports	Sunny	Sky	Portrait	**Mean Agreement**
Tagged by	11 / 11	1 / 11	0/11	1/11	-
Tagging decision performed	100%	90% (as not sunny)	- (0%)	90% (as no portrait)	**92%** (for all min 1 time tagged optional concepts)
Overall percentage	100%	90%	100%	90%	**99%** (for all optional concepts)

- Low Agreement on photo

	Beach	Landscape	Sky	Snow	**Mean Agreement**
Tagged by	4 / 11	2 / 11	7/11	0/11	-
Tagging decision performed	63% (no beach)	81% (no landscape)	63%	- (0%)	**68%** (for all min 1 time tagged optional concepts)
Overall percentage	63%	81%	63%	100%	**86%** (for all optional concepts)

Fraunhofer IDMT

Annotator Agreements – Concept View

Optional Concept	Mean over all photos (min 1 time annotated)	Mean over all photos	Number of photos annotated
Snow	**0%**	**100%**	**0/100**
Buildings / Sights	70%	93%	24/100
Aesthetic	70%	75%	84/100
Family / Friends	74%	91%	35/100
…	…	…	…
Landscape	85%	94%	37/100
Animals	89%	99%	9/100
Desert	**90%**	**99%**	**1/100**

Announcement: THESEUS/ImageCLEF Pre-workshop

- Workshop on Visual Information Retrieval Evaluation
- Topics:
  - Evaluation of Visual Information Retrieval and Annotation Methods
  - Image Retrieval / Image Annotation with Application on Photos, Medical data and Robotic vision
  - Multilabel Image Annotation supported by Knowledge Structures (Ontologies)
- Important Dates:
  - July 15, 2009 - Paper Deadline (Extended Abstract, 1-2 pages)
  - August 15, 2009 - Authors Notification
  - September 01, 2009 - Final Paper Submission (Camera Ready, 6-8 pages)
  - September 29, 2009 - Theseus/ImageCLEF Workshop in Corfu, Greek

Thank you very much.

Stefanie Nowak
Semantic Audio-Visual Systems
Fraunhofer IDMT
www.idmt.fraunhofer.de

Fraunhofer IDMT