

A Consumer Photo Tagging Ontology

Concepts and Annotations



Stefanie Nowak

Semantic Audiovisual Systems

Fraunhofer IDMT

Corfu, 29.09.2009

stefanie.nowak@idmt.fraunhofer.de

slide 1



Outline

- Introduction
- Related Work
- Ontology for Photo Annotation
- Annotation Process
- Conclusion

Photo Tagging Ontology for ImageCLEF LS-VCDT 2009



Citylife
Outdoor
Night
Underexposed
Vehicle
No_Blur
No_Persons
No_Visual_Season

- ImageCLEF: Evaluation track of CLEF (Cross-language evaluation forum)
- LS-VCDT: Detect and annotate visual concepts in consumer photos.
- Dataset:
 - MIR Flickr 25.000 Image Dataset
 - Trainingset: 5.000 photos + EXIF data + ground truth annotations
 - Testset: 13.000 photos + EXIF data
- Photo Tagging Ontology to incorporate real-world knowledge!

slide 3

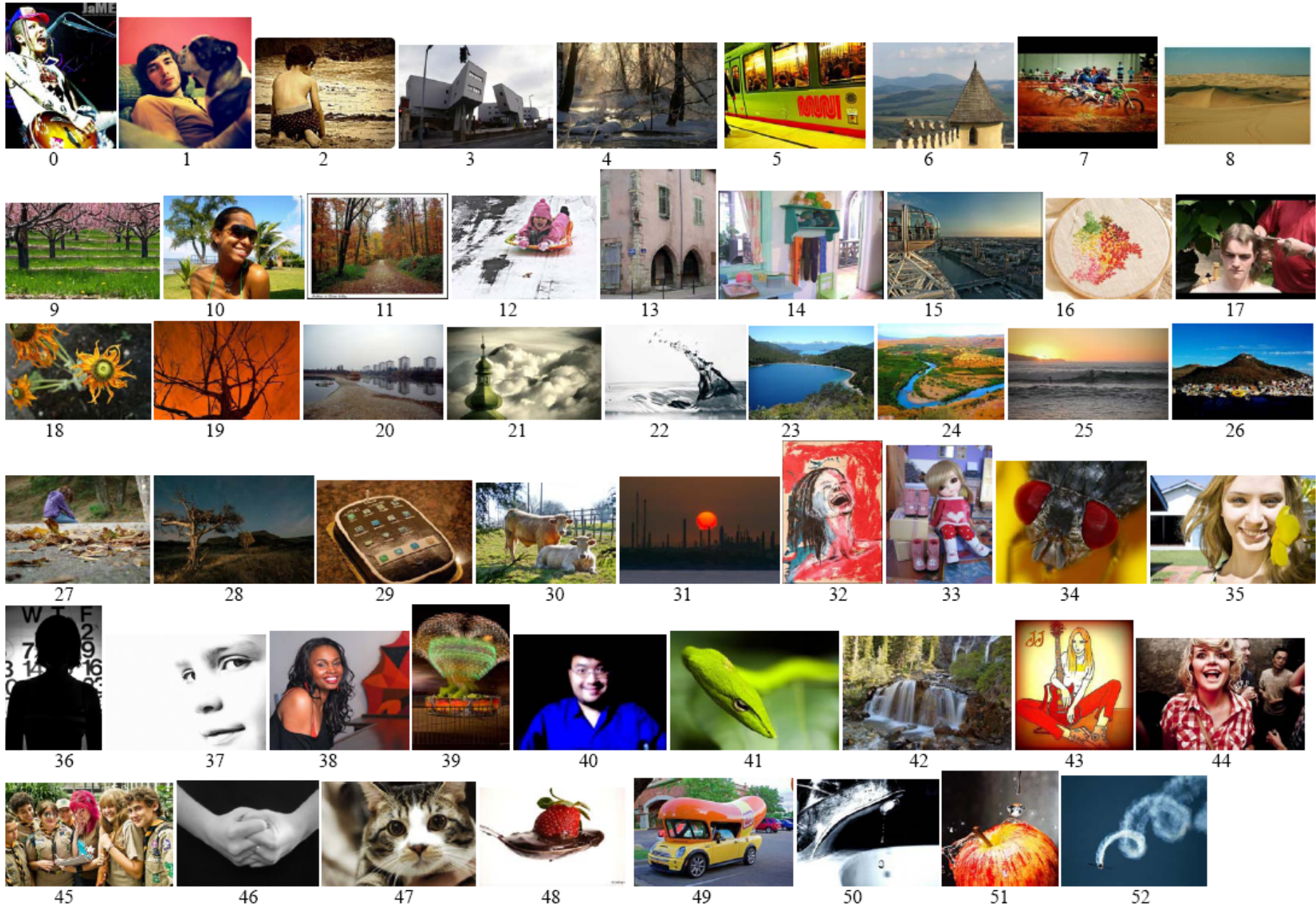
Related Work

- Kodak benchmark on the detection of semantic concepts in consumer videos
 - 100 semantic multimodal concepts (user studies) → lexicon
 - Domain: consumer videos
 - 7 categories: **subject activity**, orientation, **location**, traditional subject matter, occasion, audio and camera motion
→ also audio analysis and temporal information needed
- Large-scale concept ontology for multimedia (LSCOM)
 - >1000 concepts
 - Domain: News videos
 - 6 categories: objects, **activities/events**, **scenes/locations**, people, graphics and program categories
→ also audio analysis and temporal information needed

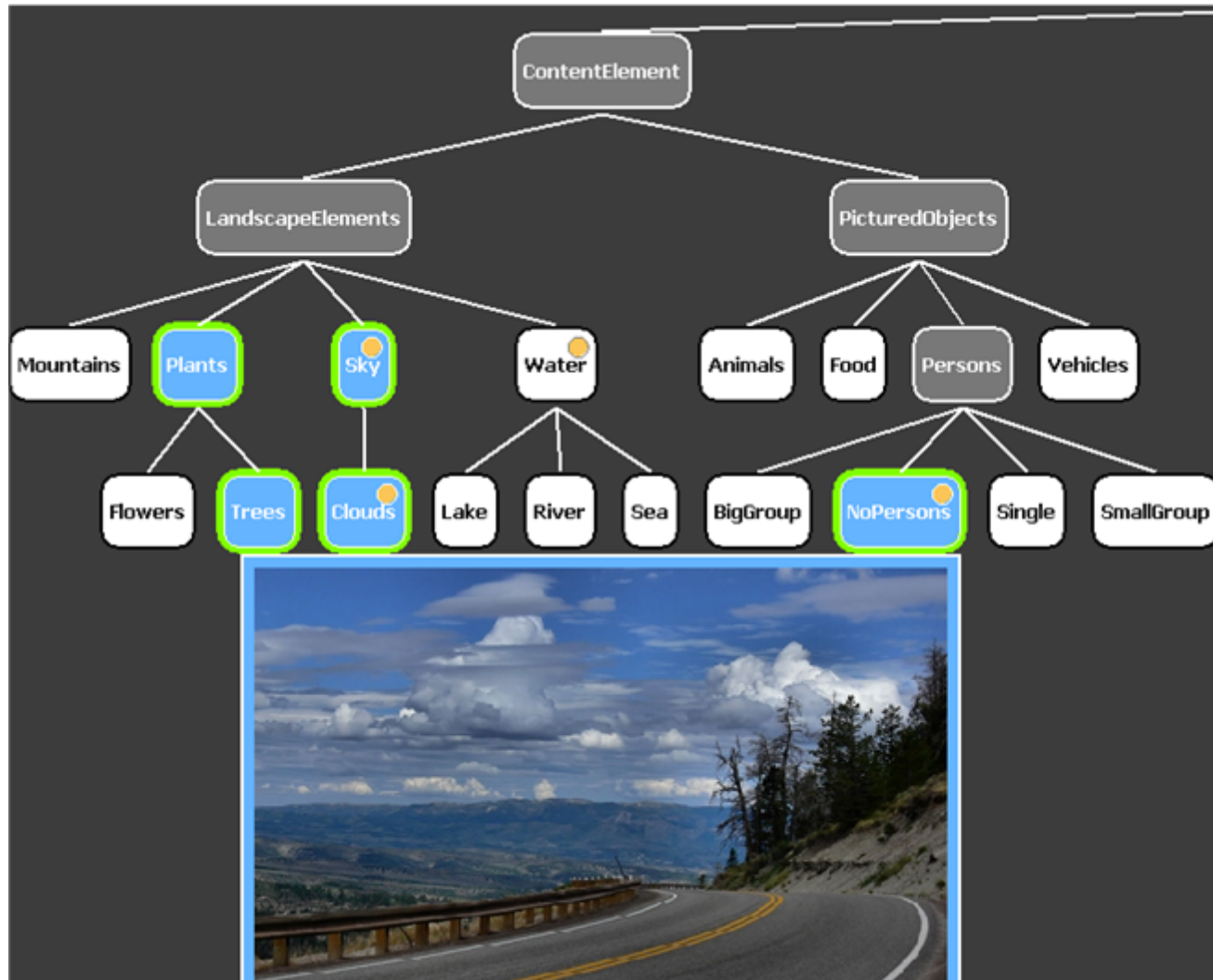
Ontology for Photo Annotation

- Requirements from THESEUS program
- Concepts should be holistically present
→ no object detection
- 4 top-level categories:
 - Content Element
 - Scene Description
 - Representation
 - Quality
 - → 53 visual concepts

Ontology for Photo Annotation: Concepts



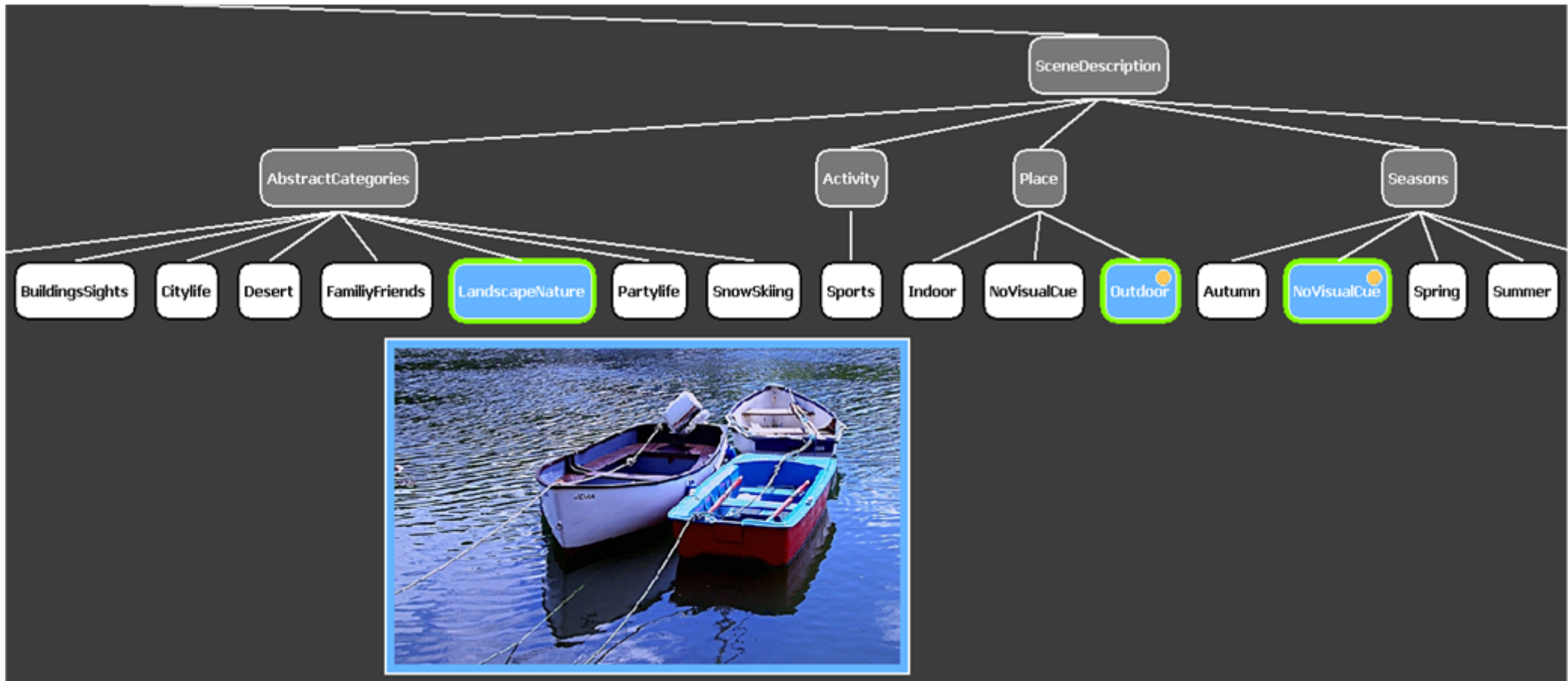
Ontology for Photo Annotation: Content Element



- Object-based concepts
- 2 subcategories:
 - 1) Landscape Elements
 - Hierarchical Structure
 - Optional concepts
 - 2) Pictured Objects
 - Optional concepts
 - Subcategory: Persons
 - Disjoint concepts

slide 7

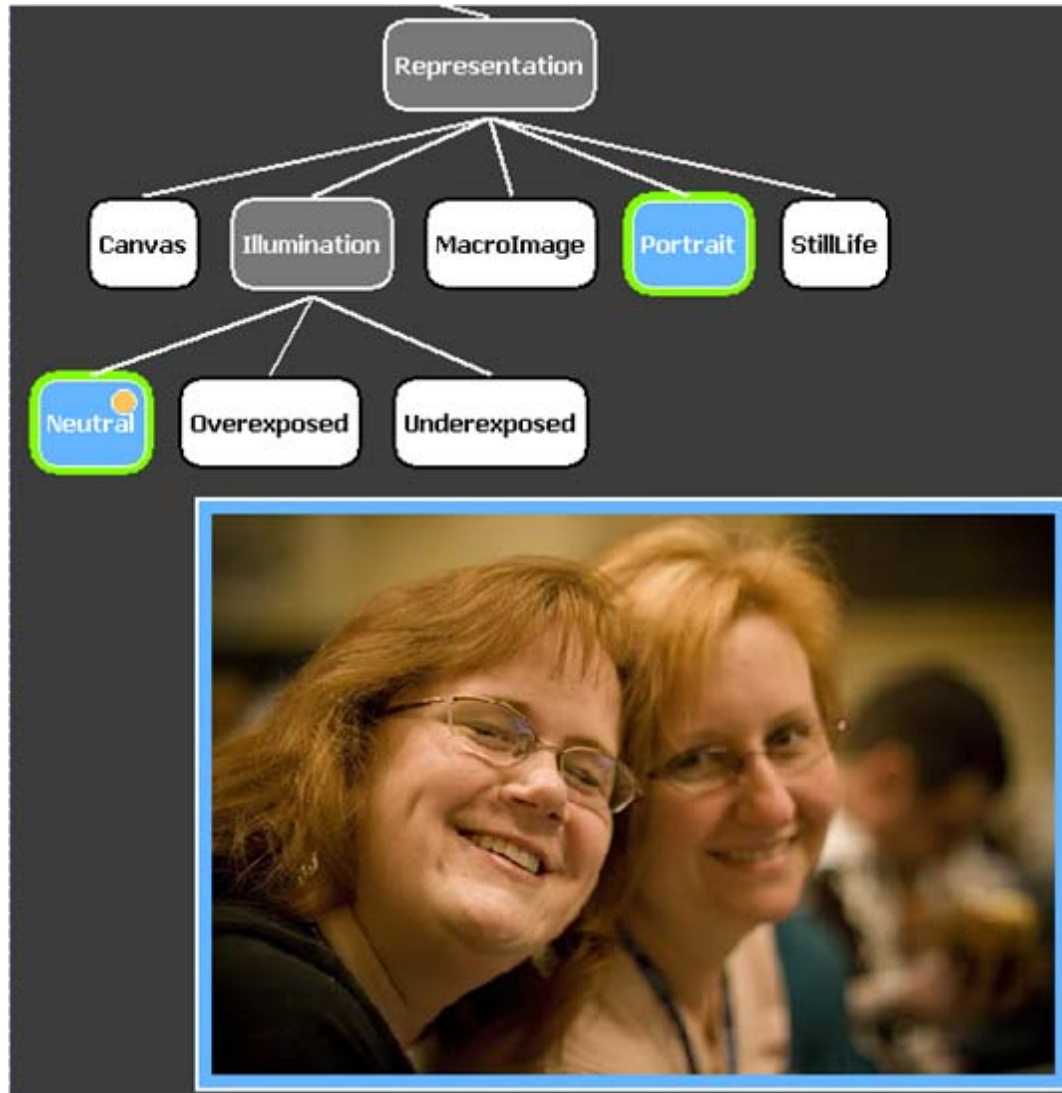
Ontology for Photo Annotation: Scene Description



- 5 subcategories: Abstract Categories, Activity, Place, Seasons and Time of Day
 - Abstract Categories + Activity: optionally modeled
 - Place + Seasons + Time: (mostly) disjoint

slide 8

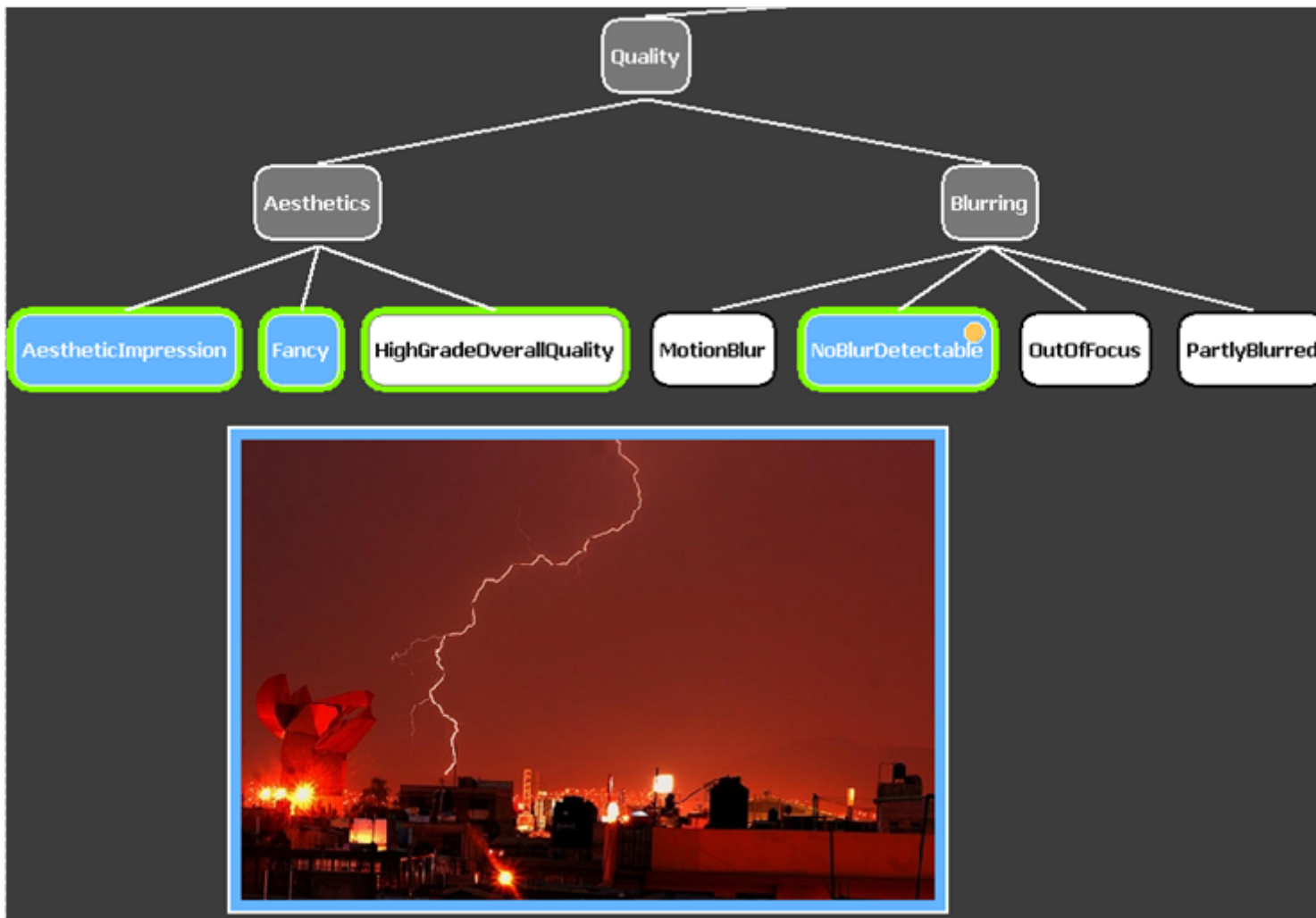
Ontology for Photo Annotation: Representation



- Concepts do not refer to content but on its representation
- Optionally modeled
- 1 subcategory: Illumination
 - Disjoint

slide 9

Ontology for Photo Annotation: Quality

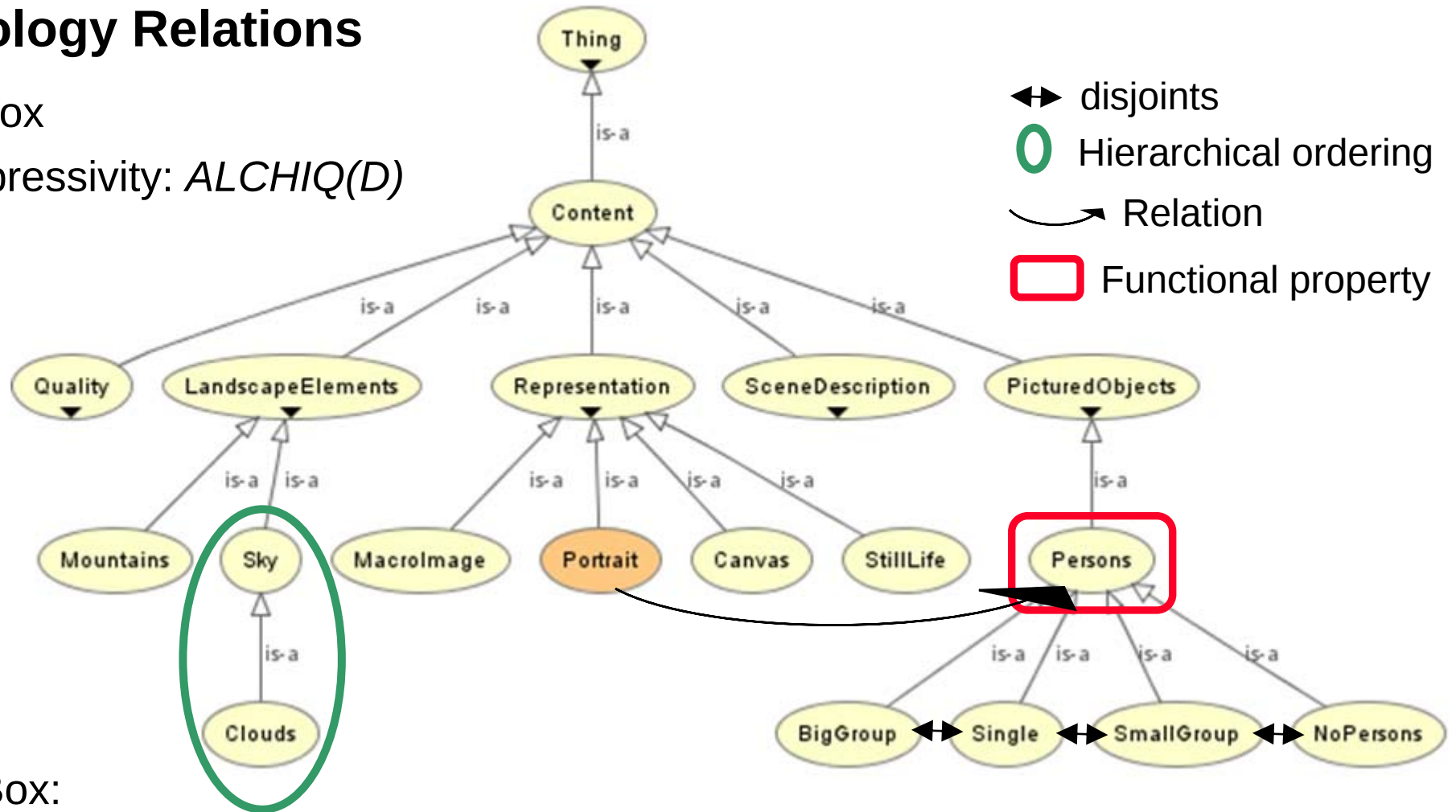


- Grade of Quality
- 2 subcategories:
 - 1) Aesthetics
 - modeled optional
 - very subjective
 - 2) Blurring
 - Modeled disjoint

slide 10

Ontology Relations

- T-Box
- Expressivity: *ALCHIQ(D)*



- A-Box:
 - Individual per concept
 - Individual per photo → photoID

Annotation Process

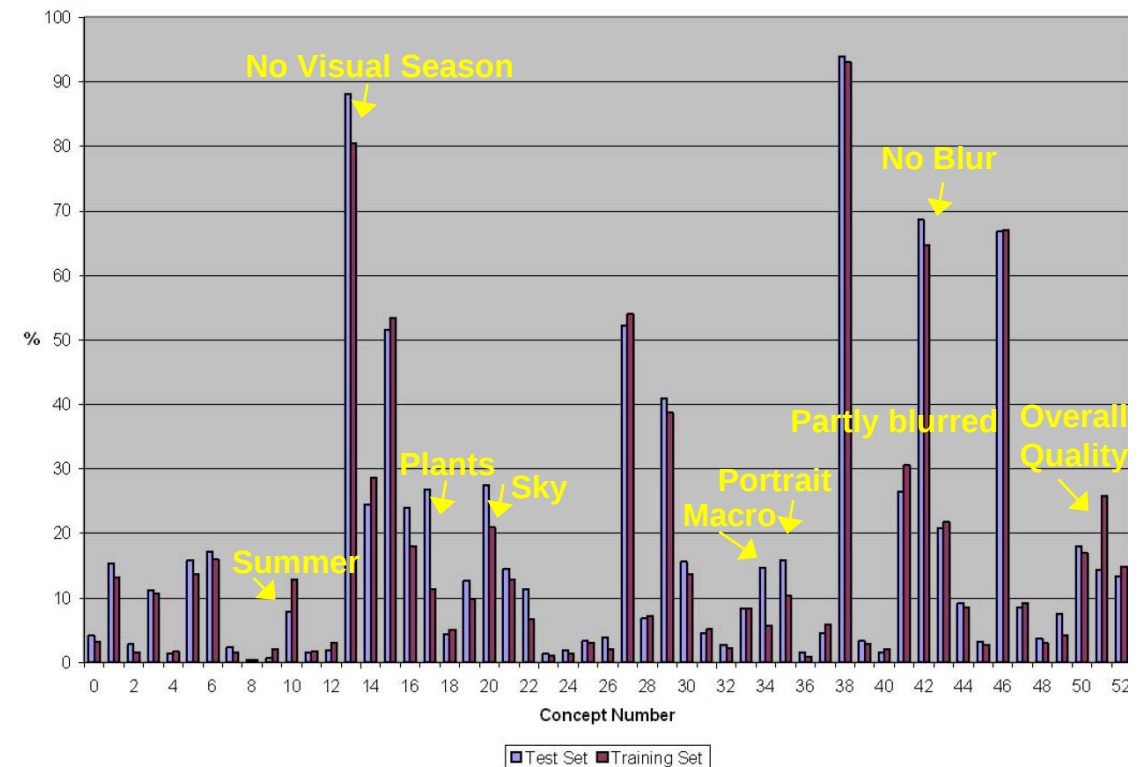
1) Annotation Step

- 18.000 photos annotated
- 43 persons (min 30 photos, max 2500 photos)
- Guideline for annotation

2) Validation Step

- 3 persons
- Screening of photos
 - a) annotated with X
 - b) not annotated with X

3) Annotator Agreements



2) Validation Step*

- Well-annotated concepts

- Top 5:

- Outdoor
 - No visual season
 - Small Group, No Persons
 - Clouds
 - Sunny

- Deleted concepts:

- Post-Processed
 - HDR Image

- 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)

slide 13

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

2) Validation – Problems in annotation

- Misunderstanding of photographic terms

- Overexposed:

correct:



wrong:



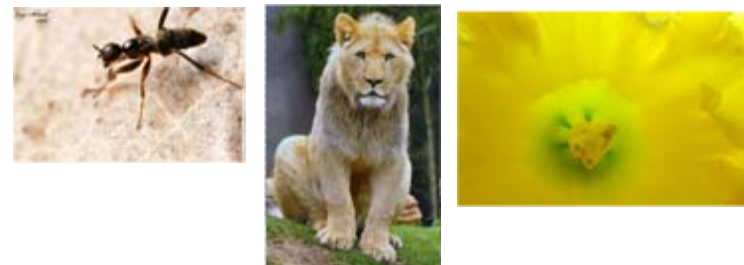
- Bad concept descriptions

- Landscape / Nature

should be:



not:



- Semantic associations

- E.g. Christmas tree in living room → winter

- What is really visible in the photos?

slide 14

1) Annotation / 2) Validation - 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

1) Annotation / 2) Validation: Aesthetic Concepts

- Not validated:
 - Fancy
 - Overall quality
 - Aesthetic impression
- Problems:
 - Explanation of aesthetic concepts not sufficient
 - Opinion changes during time

➡ Personalized Aesethetics [Datta et al.]

➡ Simplicity, realism and utilized basic techniques [Ke et al.] as guideline for annotation?

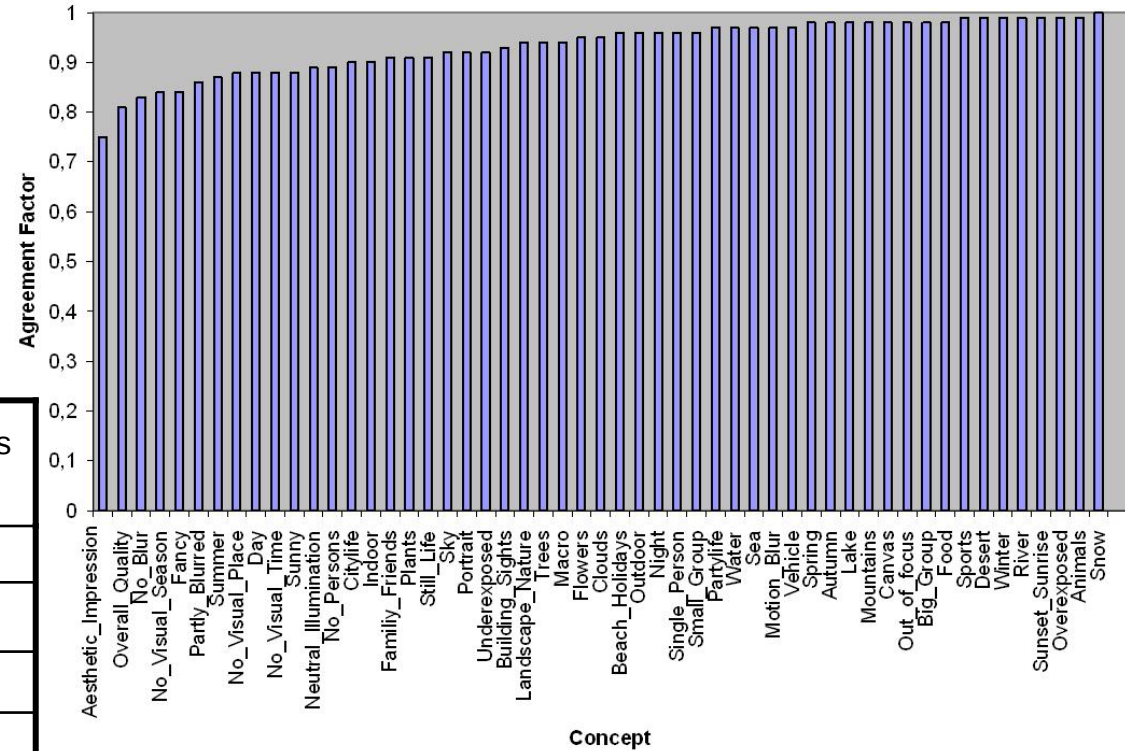
3) Annotator Agreements

- 100 photos were annotated by 11 persons
- 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
- Agreement:
 - Mean over optional concepts: 93,84%
 - [Mean over optional concepts (annotated at least 1 time): 77,85%]
 - Mean over 1 of n concepts: 92,47%

slide 17

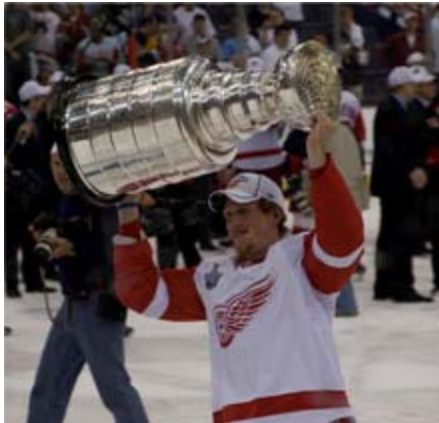
3) Annotator Agreements – Concept View

Optional Concept	Ø (min 1 time annotated)	Ø over all photos	No. 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



3) Annotator Agreements – Photo View

■ High Agreement on photo



■ Low Agreement on photo



	Sports	Sunny	Sky	Portrait	Ø 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)

	Beach	Landscape	Sky	Snow	Ø 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)

3) Annotator Agreements

- LS-VCDT:
 - Annotator Agreements per concept
 - Average over all concepts (Tagging presence and absence)
- No major differences between scores with (HSA) / without agreements (HS)
- Ranking remains
- Mean Difference: 0,028
 - Profit of agreements: +
 - Suffer from agreements: -

Team	HSA	HS	Tendency
XRCE	0,830	0,810	-
CVIUI2R	0,828	0,808	-
FIRST	0,815	0,794	-
Kameyama Lab	0,809	0,787	-
LEAR	0,792	0,770	-
Wroclaw Uni	0,790	0,765	-
ISIS	0,783	0,760	-
apexlab	0,780	0,759	-
INAOE TIA	0,759	0,732	-
CEA LIST	0,752	0,726	-
MRIM	0,741	0,711	+
UAIC	0,724	0,691	+
bpacad	0,708	0,678	+
MMIS	0,618	0,576	+
LSIS	0,549	0,498	+
AVEIR	0,516	0,479	+
LIP6	0,445	0,415	+
IAM Southampton	0,419	0,396	-
Telecom Paristech	0,390	0,361	+
random	0,384	0,351	+

Conclusion & Future Work

- Photo Tagging Ontology for consumer photos
 - Concept Structure + Definitions
- Annotation Process of LS-VCDT

- Overlaps to Kodak video concepts:
➡ additional concepts for representation, quality

- Large number of concepts for evaluation initiative

- Future Issues:
 - Incorporate Metadata (EXIF Data, Flickr Tags)
 - User Studies
 - Comparison to MIRFlickr Tags

slide 21

Thank you very much.



Stefanie Nowak
Semantic Audiovisual Systems
Fraunhofer IDMT
www.idmt.fraunhofer.de



Supported by



Federal Ministry of
Economics
and Technology

slide 22