De Universiteit van Amsterdam ⬢ UvA

Easy Categorization of Large
Collections by Automatic Analysis
and Information Visualization

Marcel Worring

Informatics Institute, UvA

De Universiteit van Amsterdam ⬢ UvA

Visual information

Sound and Vision Archive

Yearly 15.000 hours
video added

In biodiversity

Humanities Image Database Project

De Universiteit van Amsterdam ⬢ UvA

What we have

De Universiteit van Amsterdam ⬢ UvA

What we want?

Insight

De Universiteit van Amsterdam ⬢ UvA
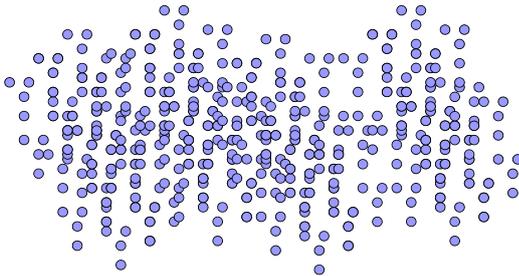
From image collection to insight

"The purpose of computing is insight, not numbers"

Richard Hamming (1915-1998)

"The purpose of information visualization is insight"

Card et.al. 1999

De Universiteit van Amsterdam ⬢ UvA

Characteristics of insight

Complex
Insight is complex, involving all or large amounts of the given data in a synergistic way, not simply individual data values.

Deep
Insight builds up over time, accumulating and building on itself to create depth often generating further questions and, hence, further insight.

Insight

Relevant
Insight is deeply embedded in the data domain, connecting the data to existing domain knowledge and giving it relevant meaning going beyond dry data analysis, to relevant domain impact.
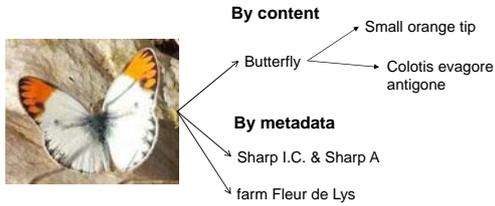
Qualitative
Insight is not exact, can be uncertain and subjective, and can have multiple levels of resolution.

Unexpected
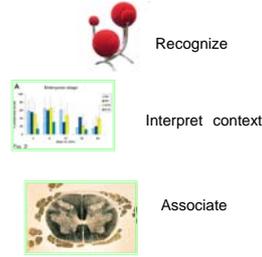Insight is often unpredictable, serendipitous, and creative.

Categorization

De Universiteit van Amsterdam UvA

Images get their meaning through iterative labeling / categorization

By content

Butterfly → Small orange tip
→ Colotis evagore antigone

By metadata

Sharp I.C. & Sharp A

farm Fleur de Lys

Category membership and statistics are core in getting insight

---

Man versus machine

De Universiteit van Amsterdam UvA

What humans are good at

Recognize

Interpret context

Associate

What machines are good at
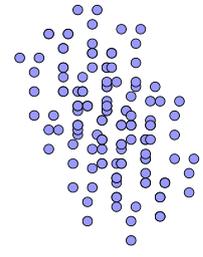
Bulk processing

Similarity can be abstract

De Universiteit van Amsterdam UvA

- Example: democracy

THE MAP OF FREEDOM

DEMOCRACY

Semantically Similar appearance

De Universiteit van Amsterdam UvA

- Example: pills

EUVISION Technologies
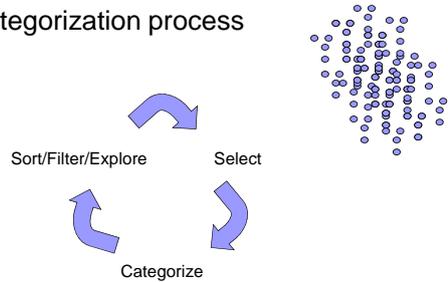

Consequences

De Universiteit van Amsterdam UvA

- Abstract
  - Only human, focus on visualizing collection and its metadata
- Semantically similar
  - Cooperation possible

The categorization process

De Universiteit van Amsterdam UvA

Sort/Filter/Explore → Select

Categorize

Slide 1:

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Ref: Nguyen2008

GalaxyBrowser

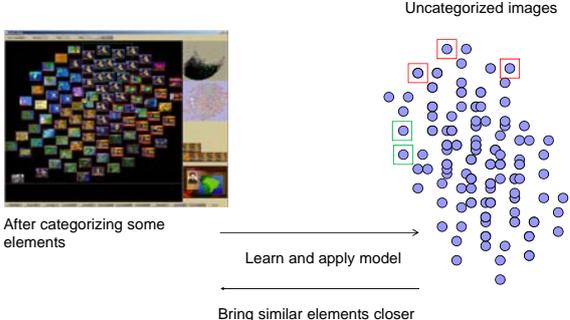**Visualize in a way that similar images are close together**

Shows the collection

Similar images can be categorized together

induced by dissimilarity

Slide 2:

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Ref: Nguyen2008

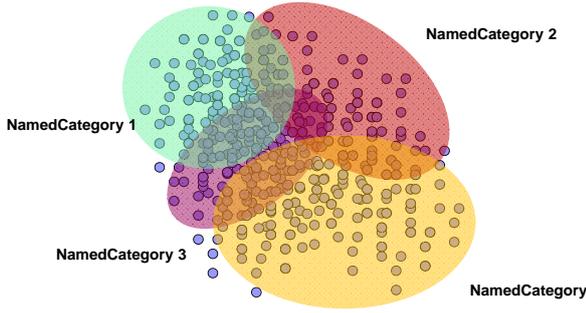## GalaxyBrowser

Uncategorized images

After categorizing some elements

Learn and apply model

Bring similar elements closer

Slide 3:

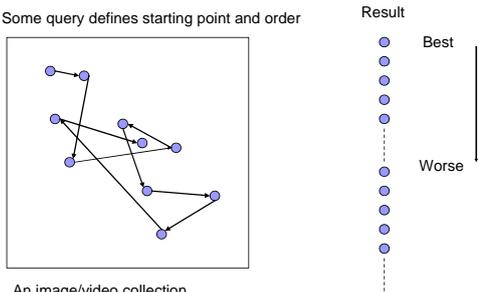DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Categorizing information

NamedCategory 2

NamedCategory 1

NamedCategory 3

NamedCategory 4

Slide 4:

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Ranking of data

Some query defines starting point and order

Result

Best
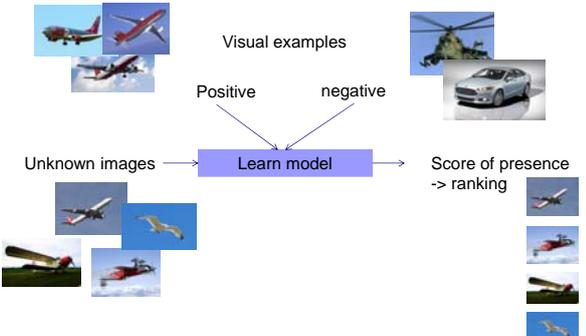

Worse

An image/video collection

Slide 5:

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Finding rankings: Concept detection

Visual examples
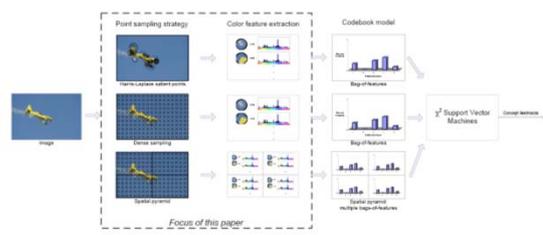
Positive   negative

Unknown images → Learn model → Score of presence -> ranking

Slide 6:

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

van de Sande, CIVR 2008, Snoek Trecvid 2008

Concept detection

One of the best detection pipelines

50 - 500 concepts
Wildlife, Face,
People marching, ……

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Ref: Snoek
IEEE Computer 2010

Quality is steadily improving

We do a good job for simple concepts
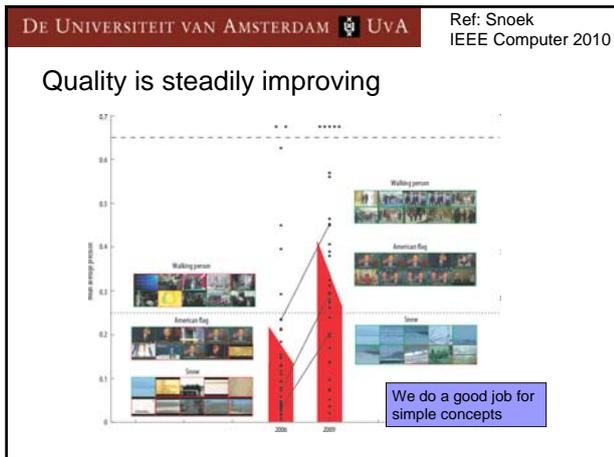
DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Multimedia Analysis

Often it works

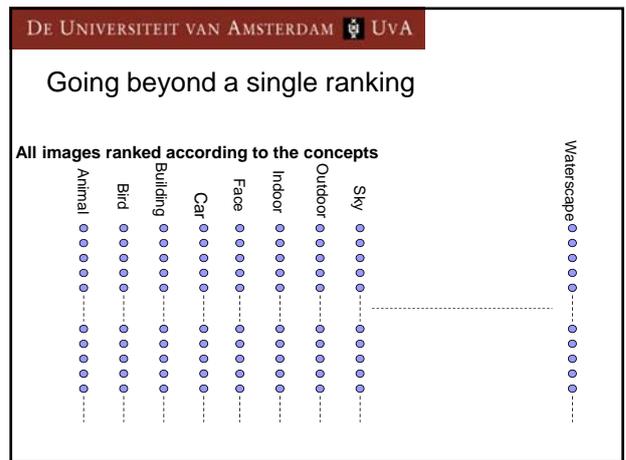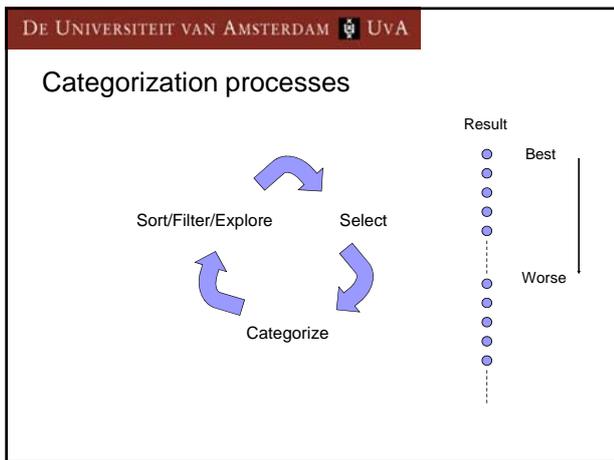**Rover Mini Cooper**

Concepts with score higher than 0.2

Car 0.716
Ground_Vehicles 0.555
Vehicle 0.767
- Tags
- Views
- Model
- Date and Time
- ownername
☑ Concepts

Often it doesn't

il decollo

Concepts with score higher than 0.2

Adult 0.263
Building 0.356
Indoor 0.311
Male_Person 0.223
- Tags
- Views
- Model
- Date and Time
- ownername
☑ Concepts

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Categorization processes

Sort/Filter/Explore   Select

Categorize

Result

Best

Worse

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

Going beyond a single ranking

All images ranked according to the concepts

Animal   Bird   Building   Car   Face   Indoor   Outdoor   Sky   Waterscape

---

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

De Rooij, Worring, vWijk,
2010

An advanced tool

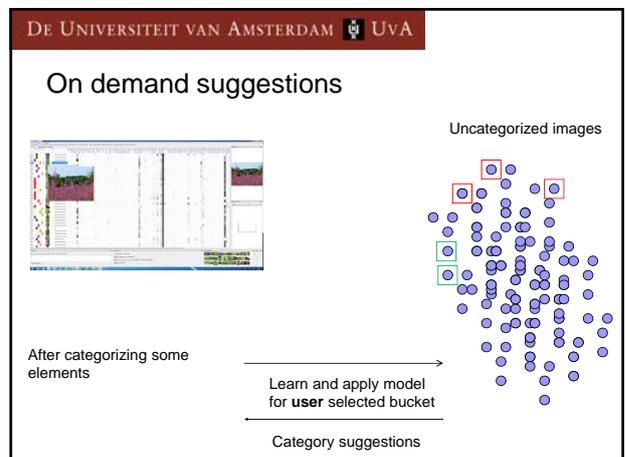All rankings visualized at the same time sorted according to one column

Learn new concepts on the fly

Bulk categorization

Search for specific elements

Interactive Exploration of Heterogeneous Information Sources

DE UNIVERSITEIT VAN AMSTERDAM 🛡 UvA

On demand suggestions

Uncategorized images

After categorizing some elements

Learn and apply model for **user** selected bucket

Category suggestions

4

Slide 1: De Universiteit van Amsterdam · UvA — **Unobtrusive assistance**
- Uncategorized images
- Continuously observe what happens
- Learn and apply model for **system** selected bucket
- Category suggestions

Slide 2: De Universiteit van Amsterdam · UvA — **An advanced tool**
- Explore different dimensions in image scatterplots

Slide 3: De Universiteit van Amsterdam · UvA — **Multimedia PivotTables**
- Tags as row variable
- Statistics over the buckets/categories
- The images corresponding to this tag, sorted according to the concept
- Filters

Slide 4: De Universiteit van Amsterdam · UvA — **Towards insight: biodiversity**

Slide 5: De Universiteit van Amsterdam · UvA — **Towards insight: showing statistics**

Slide 6: De Universiteit van Amsterdam · UvA — **Easy categorization**
- Automatic processing
- User control
- Interactive categorization
- System suggestions

Slide 1

DE UNIVERSITEIT VAN AMSTERDAM UvA

Towards insight

Complex

Multimodal categorization
Multimedia pivot tables

Deep

Incremental categorization
Use as filters and variables

Insight Impact

Relevant

User in control
Work with domain experts

Qualitative

Categories with user annotations

Unexpected

Find patterns / outliers
Incremental categorization

Slide 2

DE UNIVERSITEIT VAN AMSTERDAM UvA

A final quote

Albert Einstein wrote:

"*Computers are incredibly fast, accurate, and …….. stupid.*

Humans are incredibly slow, inaccurate and ….. brilliant.

Together they are powerful beyond imagination."

Slide 3

DE UNIVERSITEIT VAN AMSTERDAM UvA

Thanks for your attention

m.worring@uva.nl

http://www.science.uva.nl/~worring

D S
R C Data Science
Research Center