IMOTION (and vitrivr)

Heiko Schuldt

Databases & Information Systems Research Group (DBIS)

University of Basel, Switzerland

heiko.schuldt@unibas.ch
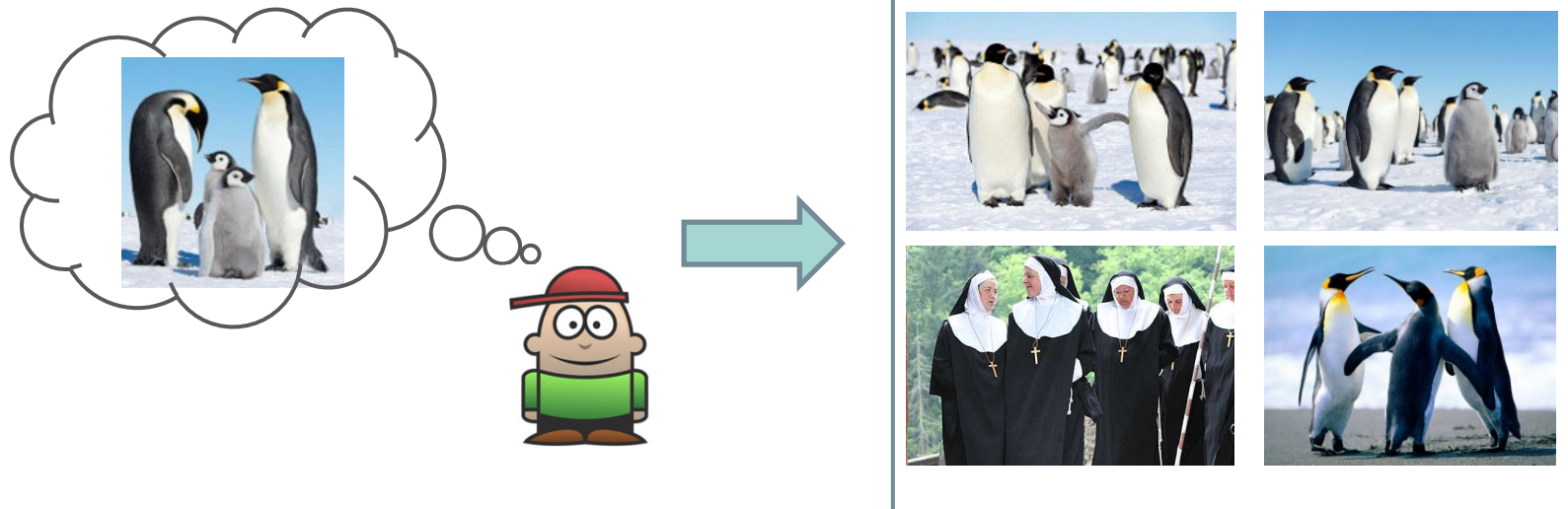
https://dbis.dmi.unibas.ch

The IMOTION Project

- IMOTION: Intelligent Multimodal Augmented Video Motion Retrieval System
  - CHIST-ERA IUI Call (Intelligent User Interfaces)
  - Duration: 2014 – 2016 (plus one year extension)

- Project Partners
  - University of Basel (UNIBAS), Switzerland (Coordinator)
  - Koç University (Koç), Istanbul, Turkey
  - University of Mons (UMONS), Belgium

- vitrivr: open source multimedia search system
  - vitrivr.org and github.com/vitrivr/
  - Databases: $ADAM_{pro}$, Cottontail-DB
  - Retrieval engine: Cineast
  - Front-end: Web-based, virtual reality (VR), augmented reality (AR)
  - Participation at Google Summer of Code 2016 & 2018

New Approaches for Searching in Multimedia …

- Keyword Search: Search based on (manually added) textual descriptions

- Query-by-Example: Similarity to query object

Sources: • http://moviespictures.org
• Penguins by kyuubidemon98 Michel CC BY-SA 3.0 via http://kyuubidemon98.deviantart.com/art/penguins-156283137
• Emperor Penguins by Christopher Michel CC BY-SA 2.0 via flickr -- https://www.flickr.com/photos/cmichel67/11240231654,
https://www.flickr.com/photos/cmichel67/11240225716, https://www.flickr.com/photos/cmichel67/11240219084
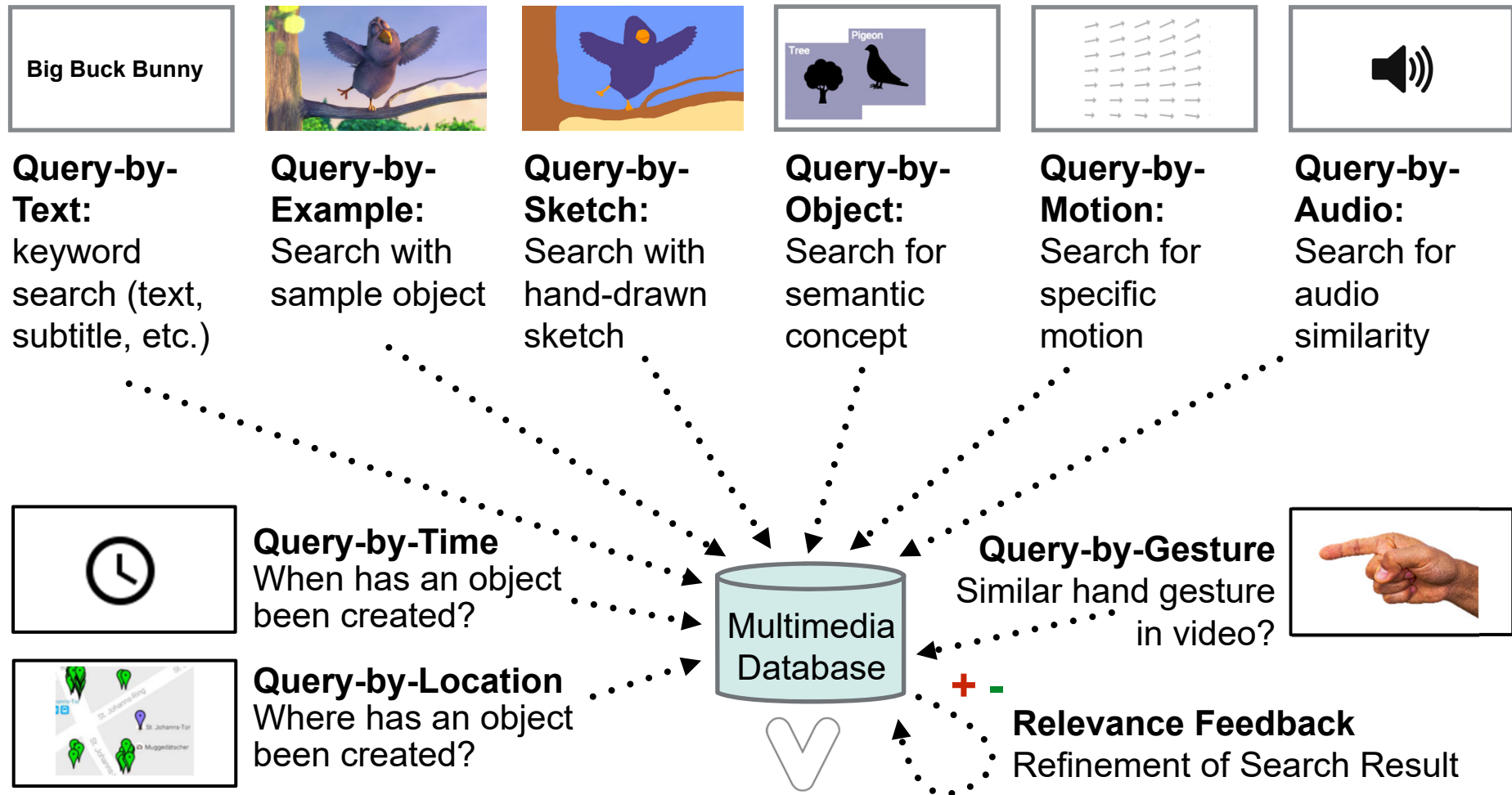
... New Approaches for Searching in Multimedia

- Query-by-Sketch: Searching using a low-level representation

Edinburgh-Castle by Christian Bickel -
https://commons.wikimedia.org/wiki/File:Edinburg004.jpg

- Motion Query: Specify motion (in videos)

- Query-by-Concept: Searching using real-world concepts

king penguin (4.96037)
penguin (4.88750)
spenisciform seabird (4.81464)
seabird (4.00825)
aquatic bird (2.31044)

Sun
Mountain
Grass

- Query-by-Semantic Sketch: Assign concepts to regions in an object

- Query-by-Time/Space: Similarity to location or time of capture of query object

- … and a combination of all these modes

Query Modes in vitrivr

Big Buck Bunny

Query-by-Text: keyword search (text, subtitle, etc.)

Query-by-Example: Search with sample object

Query-by-Sketch: Search with hand-drawn sketch

Query-by-Object: Search for semantic concept

Tree · Pigeon

Query-by-Motion: Search for specific motion

Query-by-Audio: Search for audio similarity

Query-by-Time When has an object been created?

Query-by-Location Where has an object been created?

Multimedia Database

Query-by-Gesture Similar hand gesture in video?

+ −

Relevance Feedback Refinement of Search Result

vitrivr: From the Query to the Result

vitrivr

User interface

Retrieval engine CINEAST

Storage: Cottontail-DB

Query by

– Query interpretation and execution
– Result merging

– Data storage and retrieval

– Sketch
– Example
– Semantic concept
– Keyword
– Motion
– Audio

Feature module | Feature module | Feature module | Feature module

query

result

Online phase: search

Offline phase: feature extraction

– Feature extraction

Images
Videos
Audio
3D

collection

upload

$$\begin{bmatrix} 12 \\ 3 \\ 23 \\ 0 \\ 2 \\ 0 \\ .... \\ 255 \end{bmatrix} \begin{bmatrix} 36 \\ 11 \\ 0 \\ 4 \\ 112 \\ 6 \\ .... \\ 55 \end{bmatrix} \begin{bmatrix} 91 \\ 2 \\ 243 \\ 198 \\ 17 \\ 201 \\ .... \\ 188 \end{bmatrix} \cdots \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \\ 2 \\ 255 \\ .... \\ 1 \end{bmatrix}$$

store

How to Evaluate Multimedia Retrieval?

- Multimedia Retrieval is largely interactive
  - Human in the loop (result browsing, query refinement, etc.)

Existing evaluation campaigns:

- TRECVID:
  - Very comprehensive, offline, but: not interactive
  - Ad-hoc Video Search, Activity Detection, Instance Search, Video to Text

- VBS (Video Browser Showdown) @ MMM Conference
  - Interactive, but: only on-site, small collection, online
  - Visual Known Item Search (search a video sequence that has been shown)
  - Textual Known Item Search (search for a video described textually)
  - Ad-hoc Video Search

- LSC (Lifelog Search Challenge) @ ICMR Conference
  - Interactive, but: very small collection, only on-site, online
  - Textual Known Item Search (search for images described textually)

Our Contribution: Video Dataset V3C*

V3C: Vimeo Creative Commons Collection (available at NIST)

- – Size: 28'450 videos, 4.8TB
- – Overall length: ~ 3'800 hours, duration between 3 and 60 minutes each
- – Raw content and metadata (shot segmentation, etc.)
- – Used since 2019 at TRECVID and VBS

| Partition	V3C1	V3C2	V3C3	Total
File Size	1.3 TB	1.6 TB	1.8 TB	4.8 TB
No. of videos	7'475	9'760	11'215	28'450
Overall video duration	1'000 hrs 23 min	1'300 hrs 52 min	1'500 hrs 8 min	3'801 hrs 25 min
Mean video duration	8 min 2 sec	7 min 59 sec	8 min 1 sec	8 min 1 sec
No. of segments	1'082'659	1'425'454	1'635'580	4'143'693

Luca Rossetto, Heiko Schuldt, George Awad, Asad A. Butt: *V3C – A Research Video Collection*. In: Proceedings of the 25th International Conference on Multimedia Modeling (MMM 2019), Thessaloniki, Greece, January 2019.

* Joint work with Luca Rossetto, University of Zürich.

Our Contribution: Evaluation Server DRES*

DRES: Distributed Retrieval Evaluation Server

- – Open source evaluation system
- – Supports on-site and remote interactive retrieval evaluations
- – Provides detailed statistics and analyzes
- – Used since 2020 at LSC and from 2021 onwards at VBS
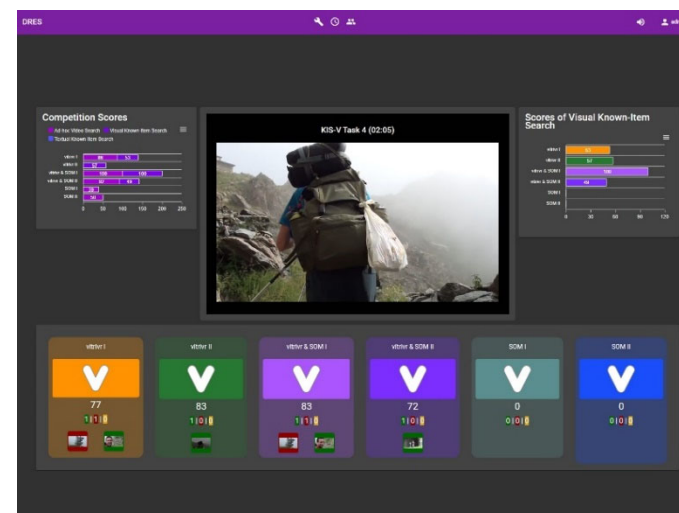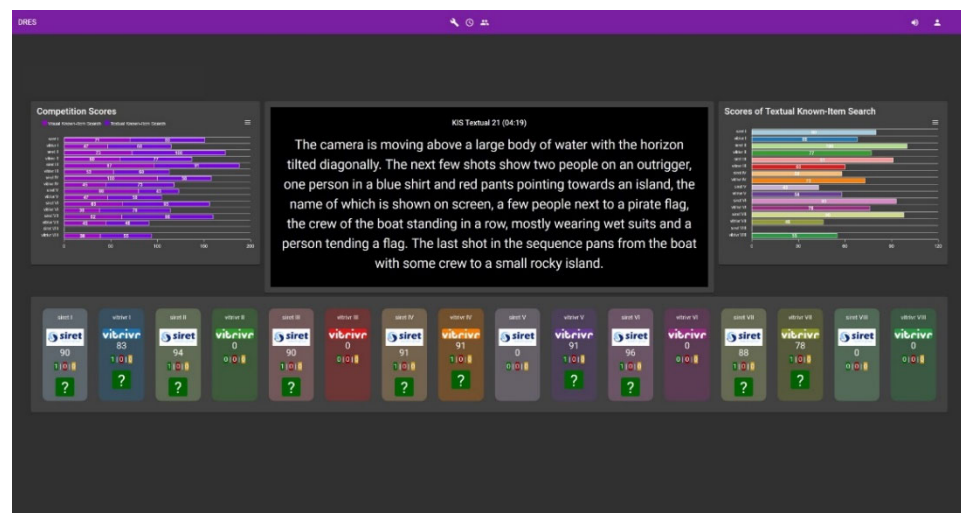
Luca Rossetto, Ralph Gasser, Loris Sauter, Abraham Bernstein, Heiko Schuldt: *A System for Interactive Multimedia Retrieval Evaluations*. To appear in: Proceedings of the 27th International Conference on Multimedia Modeling (MMM 2021), Prague, Czech Republic, June 2021.

* Joint work with Luca Rossetto, University of Zürich.

Remote Interactive Multimedia Retrieval

- So far: only on-site, once per year, attached to a conference
- With V3C / DRES:
  - Remote evaluations (test run with siret, Charles University Prague in 2020)
  - More often, more detailed, more insightful
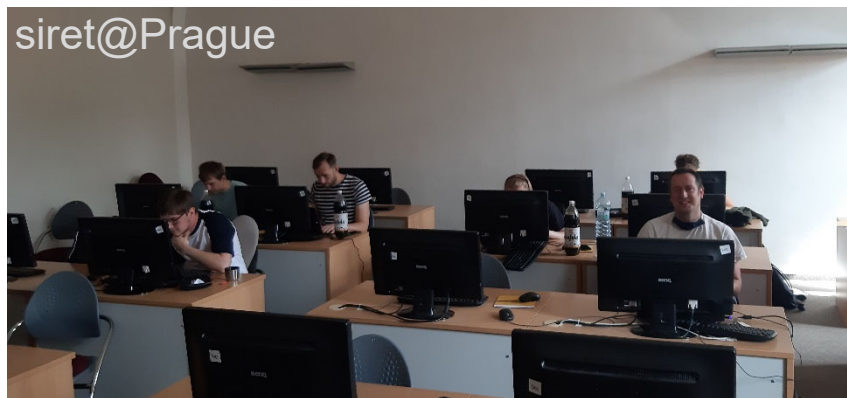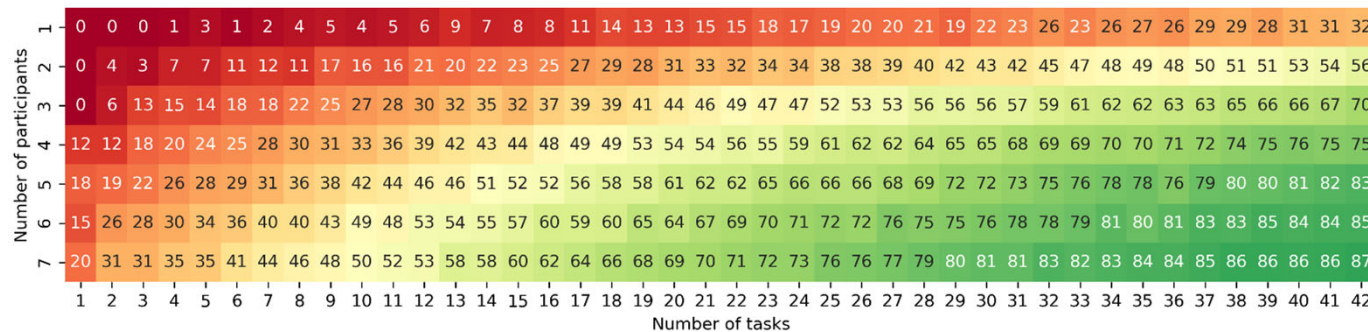
The Future of Interactive Multimedia Retrieval

- **Synchronous and asynchronous evaluations**
  - No travel involved
  - Higher frequency, longer duration

- **Standardized evaluation metrics**
  - Possibility to compare results across retrieval campaigns
  - Higher statistical significance
  - Standardized logs → more thorough analysis

- **Open science, reproducible results** (despite of user involvement)
  - Systems, configuration parameters, and data open source
  - Objective judgements

Thank you for your attention!

contact: heiko.schuldt@unibas.ch