IMOTION

Heiko Schuldt, University of Basel, Switzerland
heiko.schuldt@unibas.ch

chist-era

University of Basel

KOÇ UNIVERSITY

UMONS Université de Mons

IMOTION at a Glance

- Project Title

  Intelligent Multimodal Augmented Video Motion Retrieval System (IMOTION)

- Project Start and Duration

  January 1st, 2014 – December 31st, 2016 (3 years)

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

Scientific Background

- **Large Multimedia Collections**
  - Private use, professional applications, education

- **Content-based Multimedia Retrieval**
  - Manual annotations not feasible
  - Use of inherent features (colors, shapes, objects, etc.)
  - Strong focus on image features

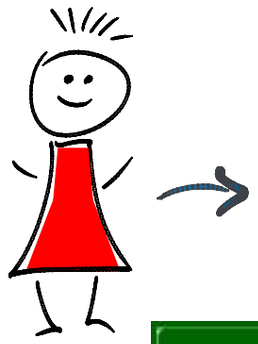
- **Video Retrieval**
  - ***Motion*** distinguishes video from still images
  - Efficient and effective content-based retrieval of (parts of) videos based on ***motion specification*** is lacking

↳ **IMOTION will develop and evaluate**
Innovative Multi-Modal User Interfaces for Interacting with Videos

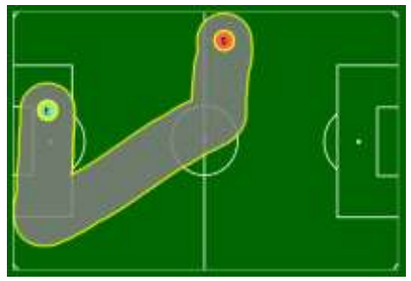
Key Challenges & Potential Impact

chist-era

Goal: Support novel types of **motion queries** in video collections



- **User Interaction for Query Specification**
  - Users can specify **motion paths of objects** in videos
  - via **sketches**, **gestures**, **natural language**, or combinations



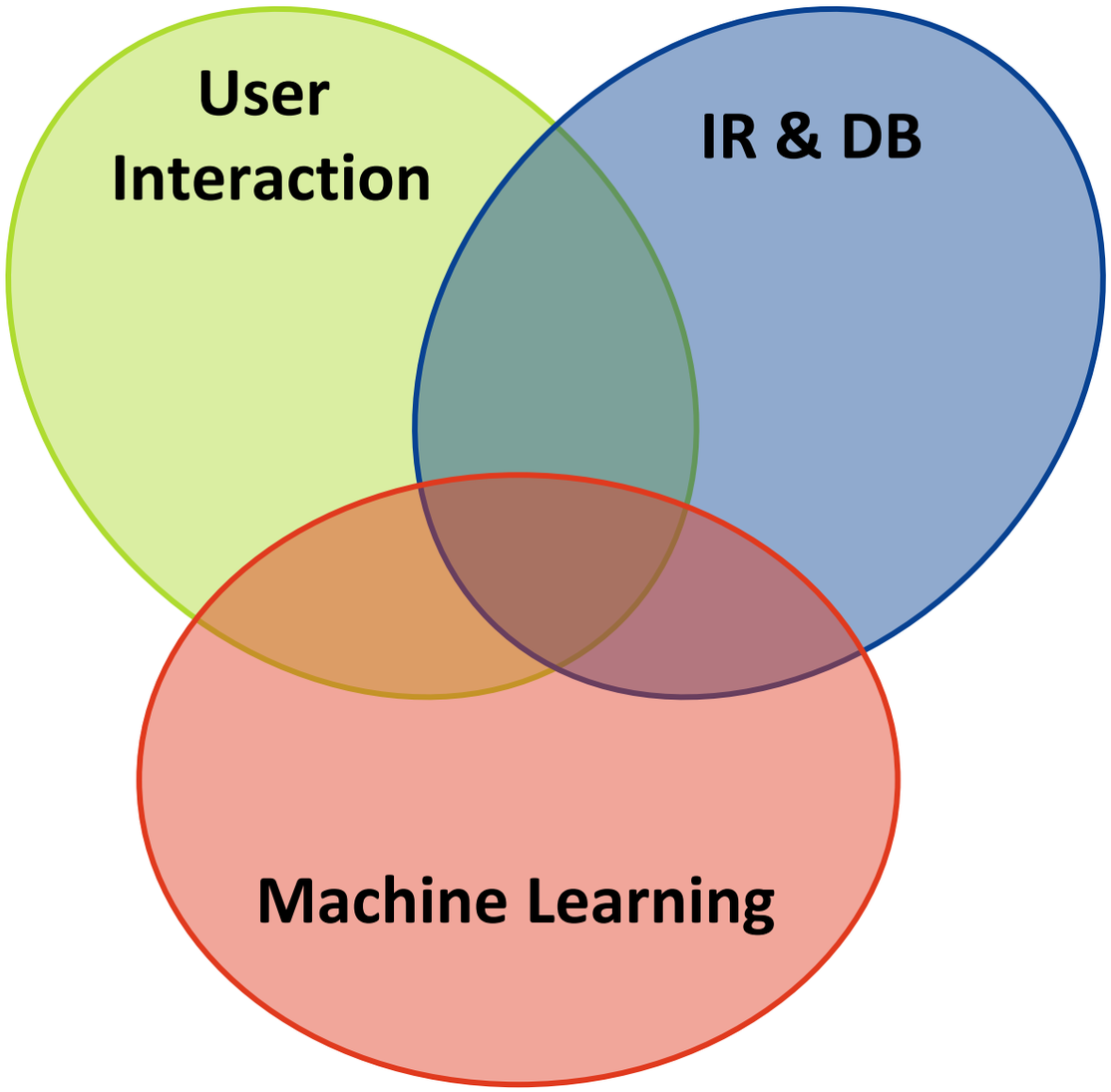
„Show me all shots on goal from the penalty area, with support from the left wing“

- **Machine Learning for Motion Features**
  - Extraction of **high-level motion descriptors**
  - based on **DNN** (ConvNets, dimensionality reduction)
- **Information Retrieval & Data Management for Efficient Search**
  - Index structures to jointly support **video features and motion metadata**
  - Distributed IR Engine: **Scalability** to very large collections
- **Quantitative and Qualitative Evaluation**

Consortium as a Whole …

Competences needed for IMOTION:

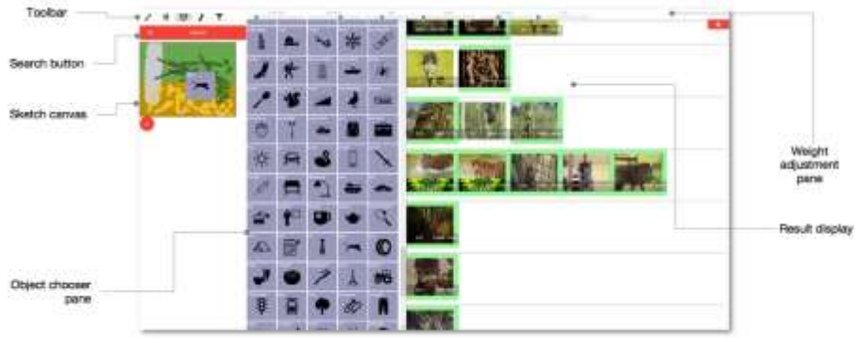
- **User Interaction**
  - Sketch-based Interfaces
  - Speech-based Interfaces
- **Information Retrieval & Databases**
  - Indexing, Retrieval Models
  - Distributed Data Management
- **Machine Learning**
  - Feature Extraction
  - Concept classification



University of Basel

Databases and Information Systems Group (Prof. Heiko Schuldt)

- **Multimedia Retrieval**, especially for **Big Data Collections**
  - *Video Similarity Search* [RGS 14, RGH+ 16]
  - *Query Types*: Seamless Combination of Boolean Retrieval and Similarity Search [GAS 14b]
  - *Very Large Collections*: Distribution in Map/Reduce Style [GAS 14a], progressive queries [GS 16]
  - *Dealing with ambiguity* [RTS 16]
  - *Video Feature Extraction* [SGS 14]
  - *Video Feature Combination* [RGS 14]

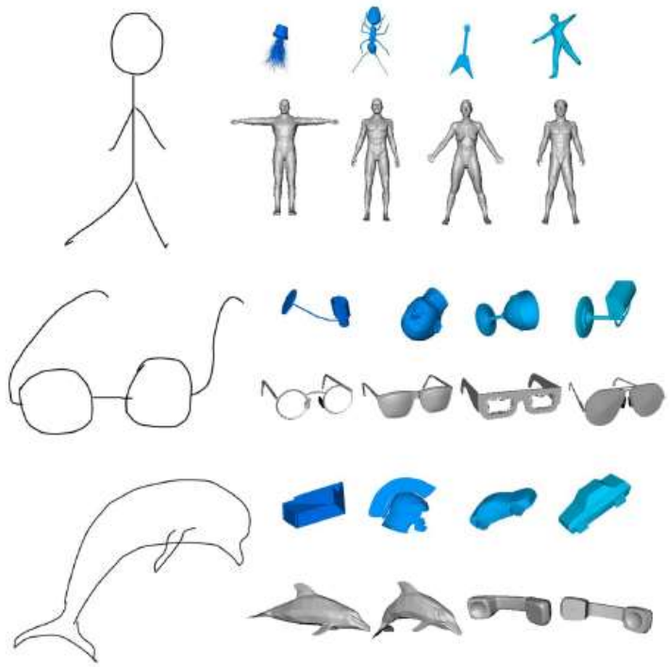


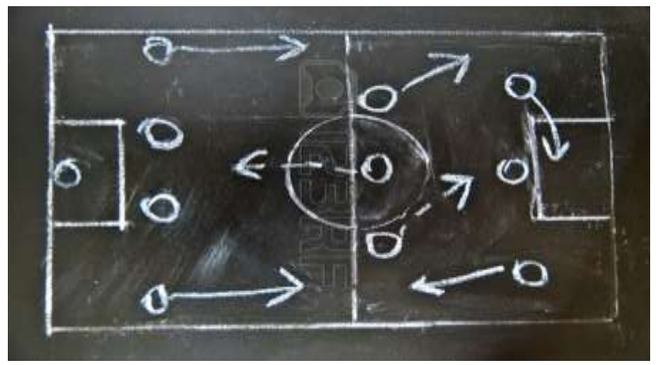
- **User Interfaces for Multimedia Queries**
  - Sketches: Interactive Paper, Tablets, Mobile Devices
  - Gestures: IR-based gesture recognition

Koç University

Intelligent User Interfaces Laboratory (Prof. T. Metin Sezgin)

- **User Interfaces for Sketch-based Interaction**
  - Sketch Recognition
  - Early recognition through auto-completion
- **Multimodal Motion Retrieval**
  - *Sketch Interfaces:* Retrieval of 3D articulated shapes through 2D sketch-based input.
  - *Speech Interfaces:* Description of motion queries
  - *Multimodal fusion:* Co-reference resolution.
  - *Use case:* Soccer videos

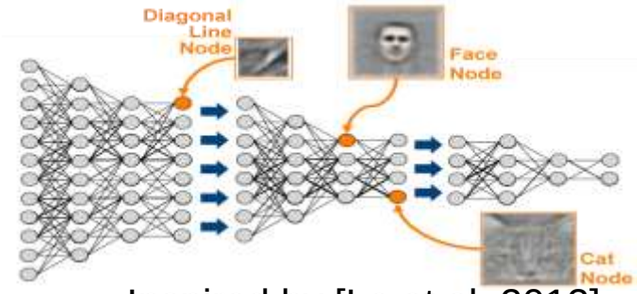




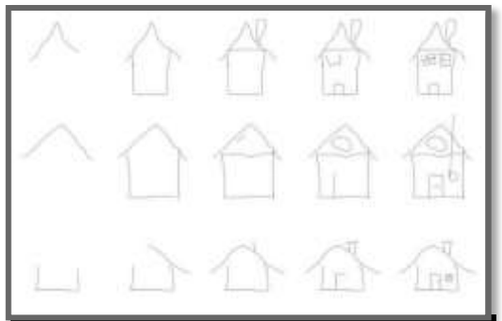
University of Mons

Research Center for Creative Technologies (Dr. Stéphane Dupont)

- **Deep Neural Networks (DNNs) for Multimedia Retrieval**
  - Feature extraction (often unsupervised learning): learning representations of sketch/image/video content
  - Classification (supervised learning): recognizing concepts and actions
  - Regression (supervised learning): mapping sketch queries to real content features



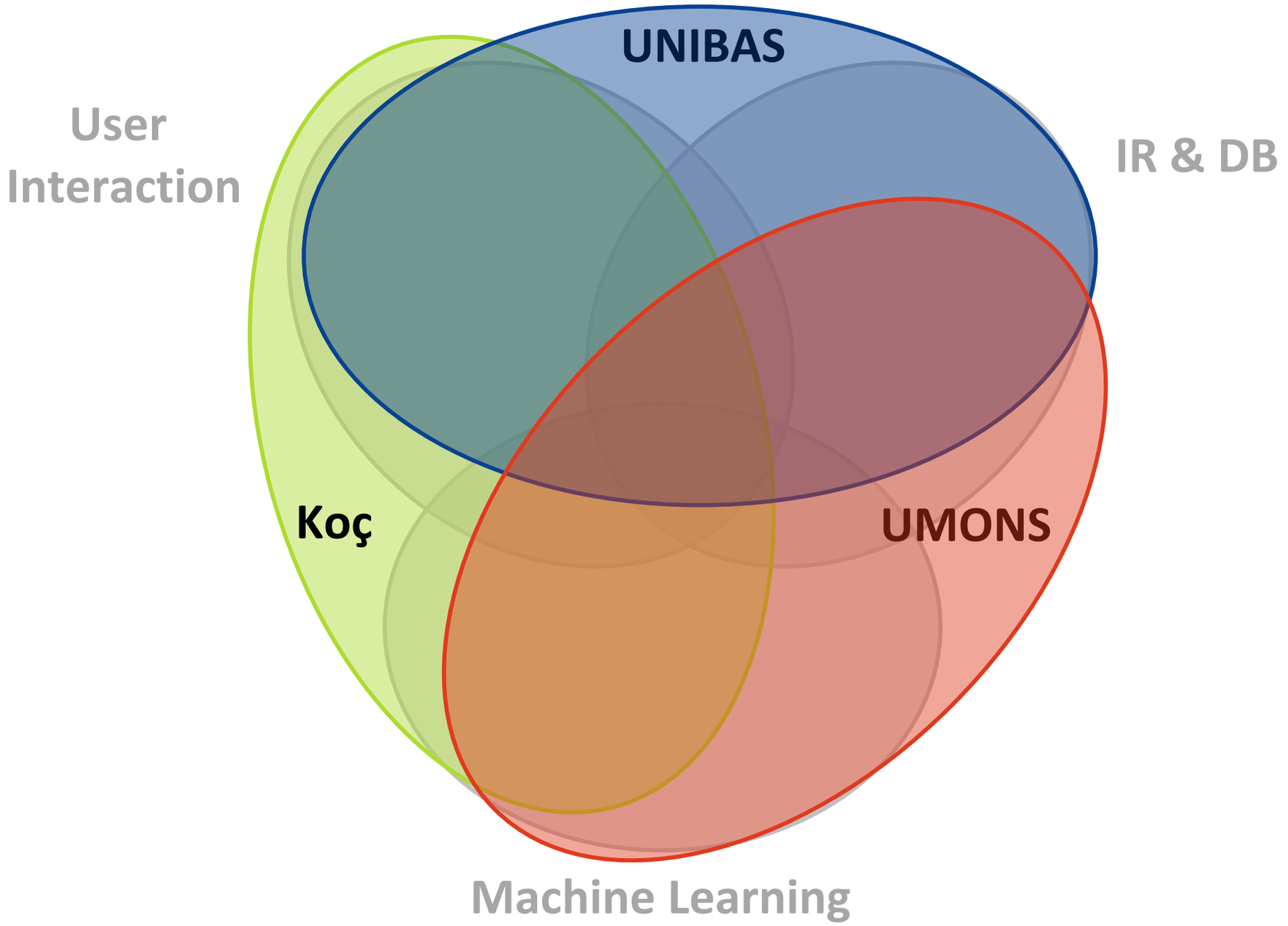
Inspired by [Le et al. 2012]



- **Latest Developments**
  - *Improved sketch recognition [SDM 15, SDM 16b]*
    - using temporal nature of sketching process
    - comparative study of fusion schemes
    - sketch-to-image search using common ConvNet features
  - *Improved action recognition [SKP+ 15, SDM 16a]*
    - comparative study of optical flow algorithms
    - fusion of spatial and temporal streams
  - *Beyond single concept: multi-label multi-instance*
    - new approach & benchmarking under development

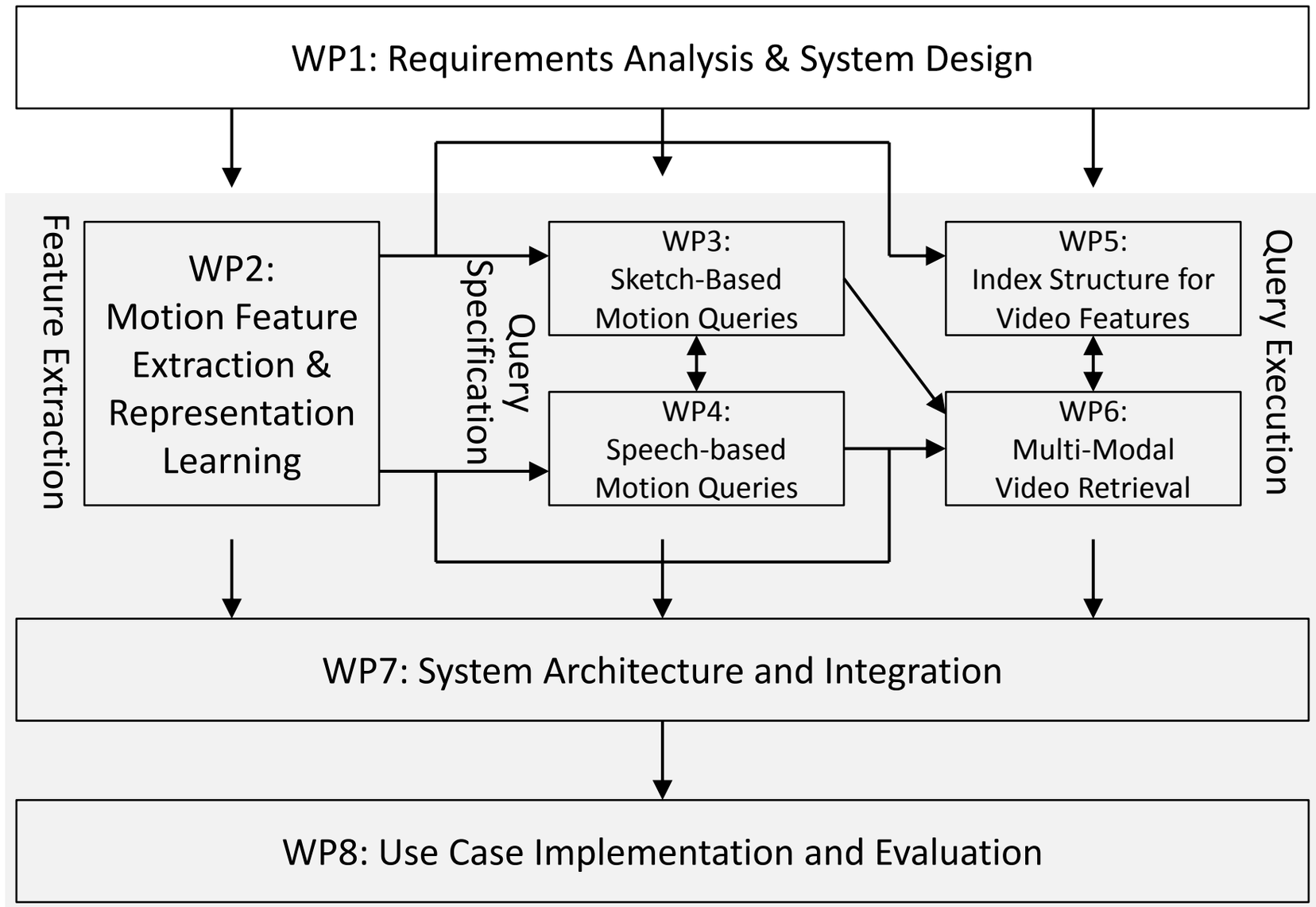
… Consortium as a Whole

- Small, but highly complementary consortium



IMOTION Workplan

- WP 2 (UMONS)
  - Feature Extraction
- WPs 3 & 4 (Koç)
  - UI for Query Specification
- WPs 5 & 6 (UNIBAS)
  - Query Execution
- WP 7 (all)
  - Build and continuously update integrated system
- WP 8 (all)
  - Evaluate integrated system



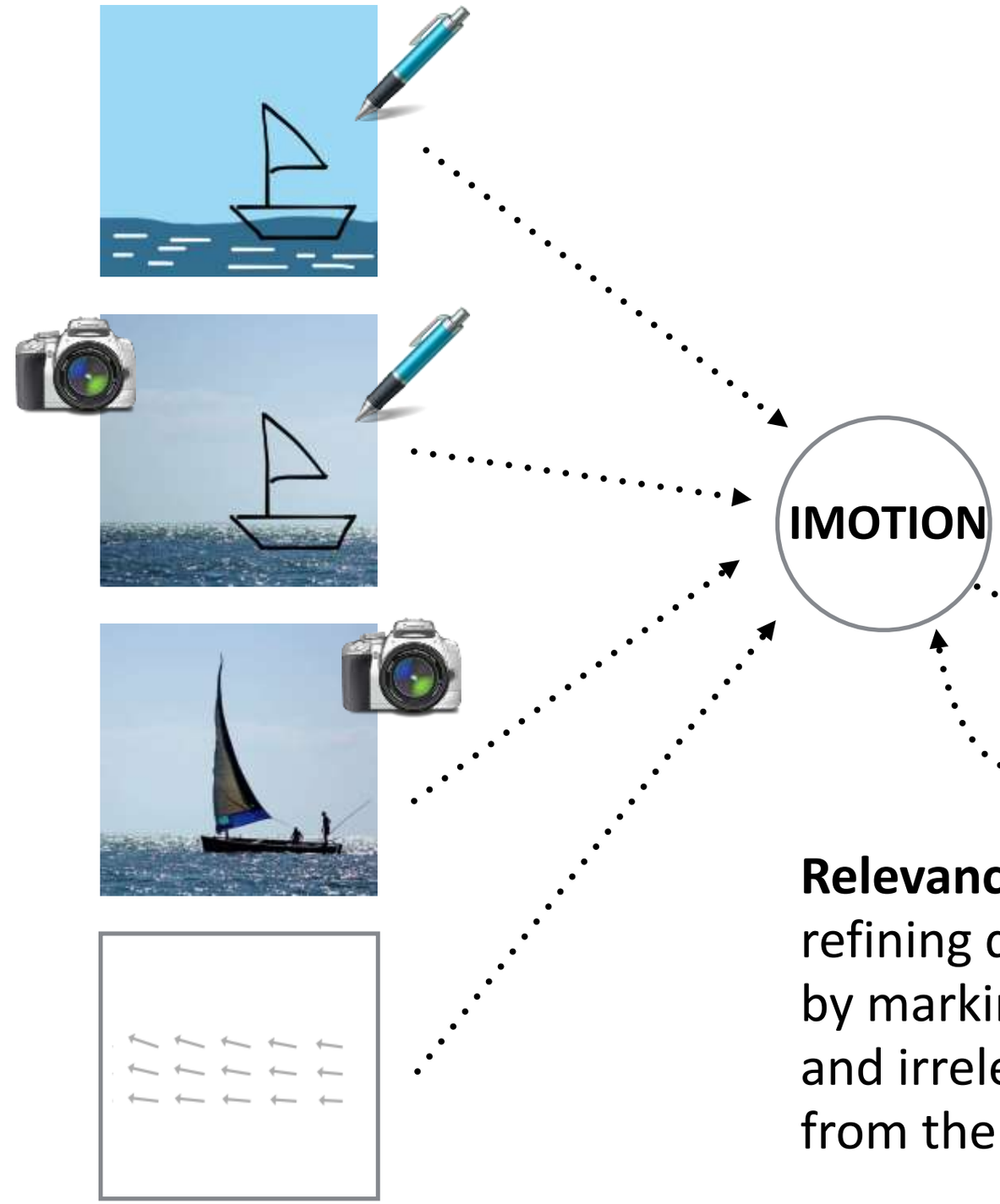
IMOTION System: Query Types

- **Query-by-Sketch**
  using a rough
  hand-drawn sketch

- **Query-by-Sketch** with
  Query-by-Example
  by adding/removing
  parts from sample image

- **Query-by-Example**
  based on query image
  or a result object

- **Query-by-Motion**
  motion of objects across
  frames via flow fields

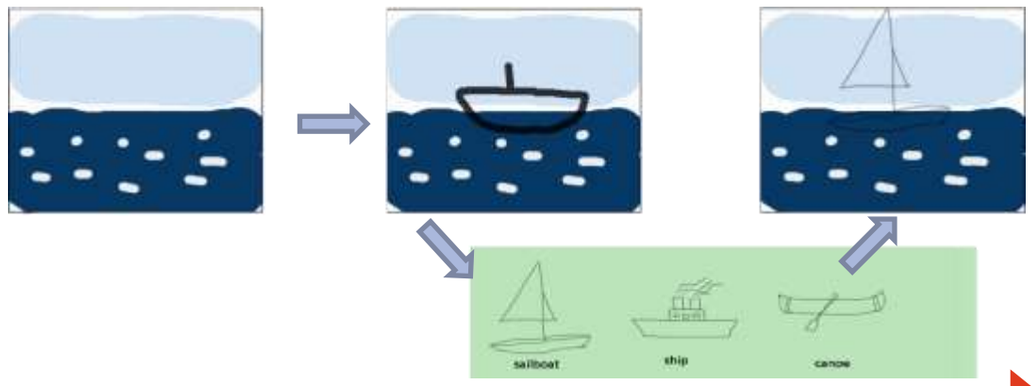


Relevance Feedback
refining query results
by marking relevant
and irrelevant elements
from the result list

IMOTION System: Overview

chist-era

Online



Query Sketch / Image

Ranked Result List

Retrieval Runtime

Query Generation

Score Fusion and Re-ranking

Feature Module

Feature Module

...

Feature Module

SQL Query

SQL Query

SQL Query

Results

Results

Ranked NN List

Ranked NN List

Feature Vector Database

Offline

- Semantic Image Features
- Motion Features
- Color / Edge Features
- …

Video

Input Decoder

Video Segmenter

Extraction Runtime

Features

Main Results & Achievements

- 15 scientific publications
- Participation at the 4th Video Search Showcase (VSS) competition [RGS+15], co-located with the Multimedia Modeling Conference (MMM'15)
  - IMOTION finished in second place overall and has been winner of the visual search task
- Participation at the 2016 Video Browser Showdown (VBS), co-located with MMM'16
  - IMOTION has participated in two modes: manually operated [RGH+ 16] and fully automated [RGT+ 16]
- Best demo award at MMM'16 [RGH+ 16]
- Demo at IUI conference on sketch auto-completion [TGR+ 16]
- Open source version of search engine and database backend (vitrivr) supported by Google Summer of Code 2016
- Plan: evaluation in TRECVID 2016, VBS 2017

Project Management …

- **Project has started on January 1st, 2014**
- **Staffing**
  - UNIBAS
    - 1 PostDoc for 2 years (C. Tănase)
    - 2 PhD students for 3 years (I. Giangreco, L. Rossetto)
  - Koç
    - 1 PostDoc for 2 years (Y. Sahillioğlu)
    - 1 MSc student (O. Altıok)
  - UMONS
    - 1 PhD student for 3 years (O. Seddati)
- **Internal project meetings (twice a year)**
  - February 2014 kick-off in Basel
  - 09/2014 in Mons, 01/2015 in Istanbul, 08/2015 in Mons, 04/2016 in Basel
  - plus bilateral technical meetings

… Project Management

- **Student/scientists exchanges**
  - eNTERFACES 2015 in Mons (August 2015): four week on-site collaboration
    www.enterface.net/enterface15/wp-content/uploads/2015/02/VideoSketcher.pdf
- **Financial reporting**
  - UNIBAS and UMONS: in plan
  - Koç: underspent
- **Dissemination, Infrastructure, Collaboration**
  - Project Website: **`imotion-project.eu`**
  - Repository, mailing list, etc. has been set up
- **Sustainability**
  - Project deliverables published as scholarly papers
  - IMOTION software in open source license
  - Collections and evaluation metrics released to the research community
    OSVC: Open Short Video Collection

IMOTION Publications (1/4)

[TGR+ 16] Claudiu Tănase, Ivan Giangreco, Luca Rossetto, Heiko Schuldt, Omar Seddati, Stéphane Dupont, Ozan Can Altıok, Metin Sezgin: *Semantic Sketch-Based Video Retrieval with Autocompletion*. In: Proceedings of the 21st ACM International Conference on Intelligent User Interfaces (IUI'16), Sonoma, CA, USA, March 2016.

[GS 16] Ivan Giangreco, Heiko Schuldt: *ADAM$_{pro}$: Database Support for Big Multimedia Retrieval*. Datenbank-Spektrum, Special Issue on Big Data & Information Retrieval, Springer, March 2016.

[RGH+ 16] Luca Rossetto, Ivan Giangreco, Silvan Heller, Claudiu Tănase, Heiko Schuldt, Stéphane Dupont, Omar Seddati, Metin Sezgin, Ozan Can Altıok, Yusuf Sahillioğlu: *IMOTION – Searching for Video Sequences using Multi-Shot Sketch Queries*. In: Proceedings of the 22nd International Conference on Multimedia Modeling (MMM 2016) - Video Browser Showdown Track, Miami, FL, USA, January 2016. Springer LNCS, Vol. 9517, pp. 377-382.

[RGT+ 16] Luca Rossetto, Ivan Giangreco, Claudiu Tănase, Heiko Schuldt, Stéphane Dupont, Omar Seddati, Metin Sezgin, Yusuf Sahillioğlu: *iAutoMotion – an Autonomous Content-based Video Retrieval Engine*. In: Proceedings of the 22nd International Conference on Multimedia Modeling (MMM 2016) - Video Browser Showdown Track, Miami, FL, USA, January 2016. Springer LNCS, Vol. 9517, pp. 383-387.

IMOTION Publications (2/4)

[RGH+ 16] Luca Rossetto, Ivan Giangreco, Silvan Heller, Claudiu Tănase, Heiko Schuldt: *Searching in Video Collections using Sketches and Sample Images – The Cineast System*. In: Proceedings of the 22nd International Conference on Multimedia Modeling (MMM 2016), Miami, FL, USA, January 2016. Springer LNCS, Vol. 9517, pp. 336-341. **Best Demo Award.**

[RTS 16] Luca Rossetto, Claudiu Tănase, Heiko Schuldt: *Dealing with ambiguous Queries in Multimodal Video Retrieval*. In: Proceedings of the 22nd Int'l Conf. on Multimedia Modeling (MMM 2016), Miami, FL, USA, January 2016. Springer LNCS, Vol. 9517, pp. 898-909.

[YSC+ 15] Kemal Tugrul Yesilbek, Cansu Sen, Serike Cakmak, Tevfik Metin Sezgin*: SVM-Based Sketch Recognition: Which Hyperparameter Interval to Try*, In: Proceedings of the workshop on Sketch-Based Interfaces and Modeling (SBIM '15), pages 117-121, June 2015.

[AYS16] Ozan Can Altıok, Kemal Tugrul Yesilbek, Tevfik Metin Sezgin: *What Auto-Completion Tells Us About Sketch Recognition*. In: Proceedings of the Sketch-Based Interfaces and Modeling Symposium, Lisbon, May 2016.

[SKP+ 15] Omar Seddati, Emre Kulah, Gueorgui Pironkov, Stéphane Dupont, Saïd Mahmoudi, Thierry Dutoit: *UMONS at MediaEval 2015 Affective Impact of Movies Task including Violent Scenes Detection*, In: Working Notes Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, 14-15 September 2015.

IMOTION Publications (3/4)

[RGS+15] Luca Rossetto, Ivan Giangreco, Heiko Schuldt, Stéphane Dupont, Omar Seddati, Metin Sezgin, Yusuf Sahillioğlu: *IMOTION – a Content-based Video Retrieval Engine.* In: Proceedings of the 21st MultiMedia Modelling Conference (MMM2015) - Video Search Showcase Track, Sydney, Australia, pp. 255-260, January 2015, Springer LNCS, Vol. 8936.

[SDM 15] Omar Seddati, Stéphane Dupont, Said Mahmoudi: *Deepsketch: deep convolutional neural networks for sketch recognition and similarity search*, in Proceedings of 13th International Workshop on Content-Based Multimedia Indexing (CBMI), Prague, Czech Republic, 10-12 June 2015.

[GAS 14a] Ivan Giangreco, Ihab Al Kabary, Heiko Schuldt: *ADAM — A Database and Information Retrieval System for Big Multimedia Collections*. In: Proceedings of the 3rd International Congress on Big Data, Anchorage, USA, pp. 406 – 413, June 2014.

[GAS 14b] Ivan Giangreco, Ihab Al Kabary, Heiko Schuldt. *ADAM — A System for Jointly Providing IR and Database Queries in Large-Scale Multimedia Retrieval.* In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14), Gold Coast, Australia, July 2014.

[RGS 14] Luca Rossetto, Ivan Giangreco, Heiko Schuldt: *Cineast: A Multi-Feature Sketch-Based Video Retrieval Engine.* In: Proceedings of the 16th IEEE International Symposium on Multimedia (ISM2014), Taichung, Taiwan, December 2014. IEEE.

IMOTION Publications (4/4)

[SGS 14] Fabio Sulser , Ivan Giangreco, Heiko Schuldt: *Crowd-based Semantic Event Detection and Video Annotation for Sports Videos*. In: Proceedings of the 3rd International ACM Workshop on Crowdsourcing for Multimedia, Orlando, USA, pp. 63-68, November 2014.

[SDM 16a] Omar Seddati, Stéphane Dupont, Saïd Mahmoudi: *Réseaux de neurones convolutionnels profonds pour la reconnaissance d'action dans les vidéos*. In: Proceedings of 18ème édition CORESA 2016 (COmpression et REprésentation des Signaux Audiovisuels), Nancy, France, May 2016.

[SDM 16b] Omar Seddati, Stéphane Dupont, Saïd Mahmoudi: *DeepSketch 2: Deep Convolutional Neural Networks for Partial Sketch Recognition*. In: Proceedings of the 14th International Workshop on Content-Based Multimedia Indexing (CBMI), Bucharest, Romania, June 2016.

[DAB+ 15] Stéphane Dupont, Ozan Can Altıok, Aysegül Bumin, Ceren Dikmen, Ivan Giangreco, Silvan Heller, Emre Külah, Gueorgui Pironkov, Luca Rossetto, Yusuf Sahillioğlu, Heiko Schuldt, Omar Seddati, Yusuf Setinkaya, Metin Sezgin, Claudiu Tănase, Emre Toyan, Sean Wood, Doguhan Yeke: *VideoSketcher: Innovative Query Modes for Searching Videos through Sketches, Motion and Sound*. In: Proceeding the 11th International one-month Summer Workshop on Multimodal Interfaces (eNTERFACE'2015), Mons, Belgium (in press).

chist-era

http://www.shutterstock.com