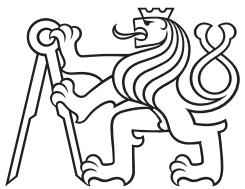
Knowledge Extraction from Multimedia Content

extended version

Petr Pulc

Department of Applied Mathematics
Faculty of Information Technology
Czech Technical University in Prague

October 13, 2022



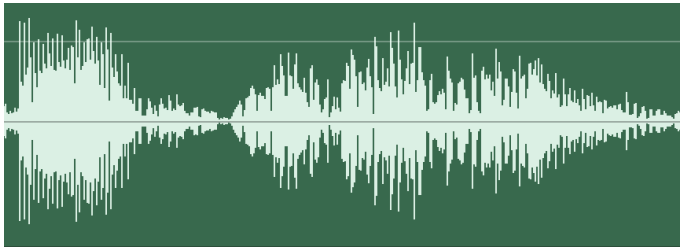
Outline

Introduction – Problem statement

- Extensive archive of multimedia content
- Various types of recorded material

- Need of a way how to navigate the material
- Significantly limited computational resources

Introduction – Multimedia content basics

- “Hello everyone!”
-  → hɛˈləʊ ˈ ɛvrɪwʌn ↑
-  → ?

Introduction – Vision

Create a **framework** that **describes or labels** text, sound, or video documents **only with the tools most suitable for the job** at a given time **on minimal subset** of the original material.

Introduction – Content duplicity

Introduction – Content duplicity

Introduction – State of the Art

- Generic methods with restricted semantics
- — We aim to bridge this gap —
- Single-use methods expensively constructed for solving specific problems
- Methods utilising universal approximators and black-box methods

Introduction – State of the Art

- Generic methods with restricted semantics
- — **We aim to bridge this gap** —
- Single-use methods expensively constructed for solving specific problems
- Methods utilising universal approximators and black-box methods

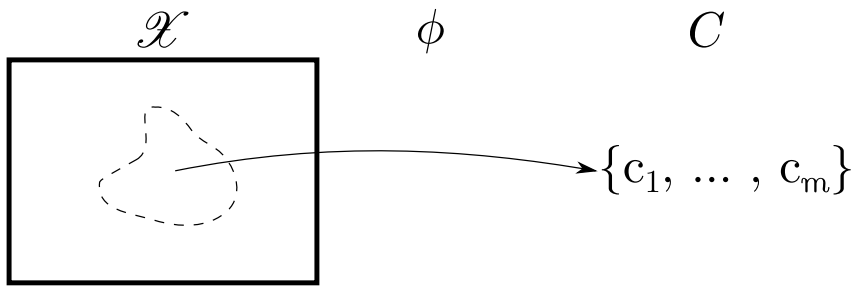
Contributions of the Thesis

1. Theoretical approach to knowledge extraction from various multimedia content
   (hierarchical classification framework)
2. Extraction of semantic features from data aggregates
   (classification methods based on low-level processing)
3. Leveraging information redundancy in the video signal
   (reducing video signal to simpler representations)
4. Inclusion of unsupervised knowledge extraction methods
   (to traditionally strictly supervised approaches)

Contribution 1

Knowledge extraction framework

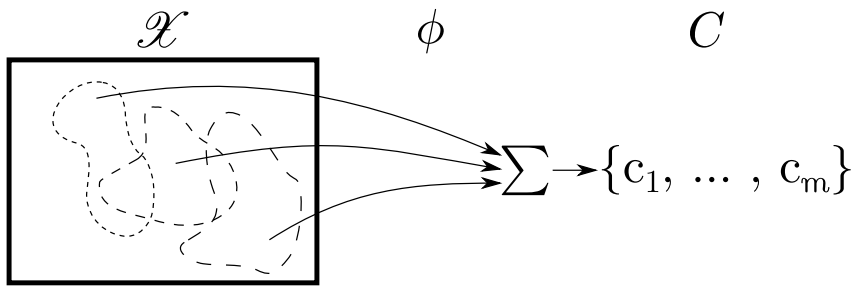
Instead of designing a monolithic classifier or combining the output of multiple classifiers, we propose utilising a meta-learning approach, where we use a subset of the original content for assessing the suitability of particular more complex processing methods.



Contribution 1

Knowledge extraction framework

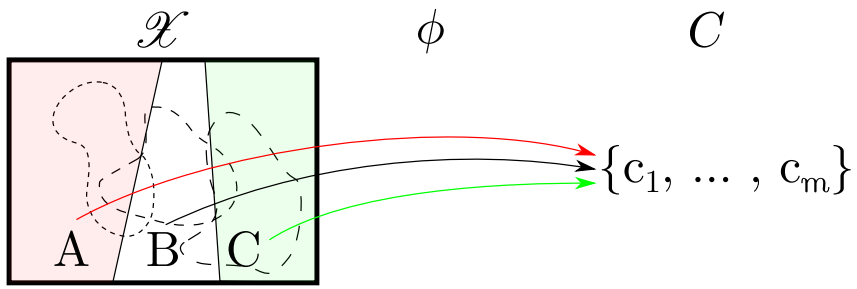
Instead of designing a monolithic classifier or combining the output of multiple classifiers, we propose utilising a meta-learning approach, where we use a subset of the original content for assessing the suitability of particular more complex processing methods.



Contribution 1

Knowledge extraction framework

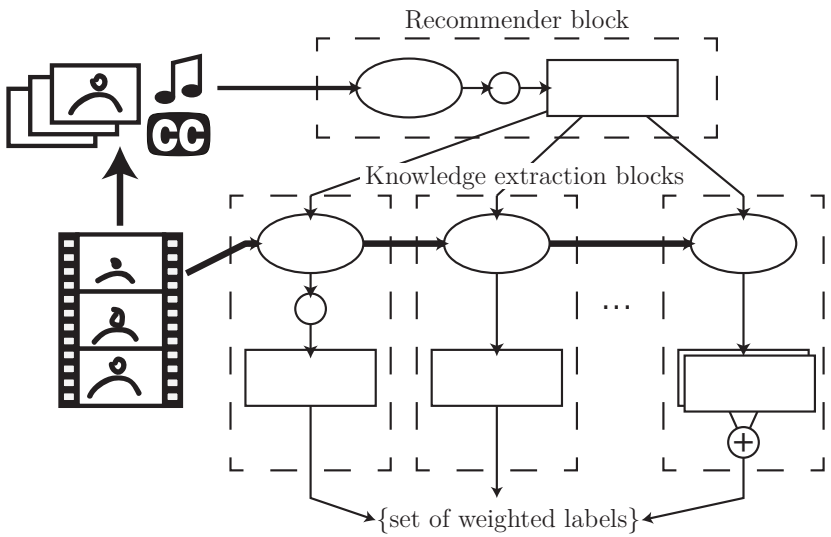
Instead of designing a monolithic classifier or combining the output of multiple classifiers, we propose utilising a meta-learning approach, where we use a subset of the original content for assessing the suitability of particular more complex processing methods.



Contribution 1

Knowledge extraction framework

Instead of designing a monolithic classifier or combining the output of multiple classifiers, we propose utilising a meta-learning approach, where we use a subset of the original content for assessing the suitability of particular more complex processing methods.

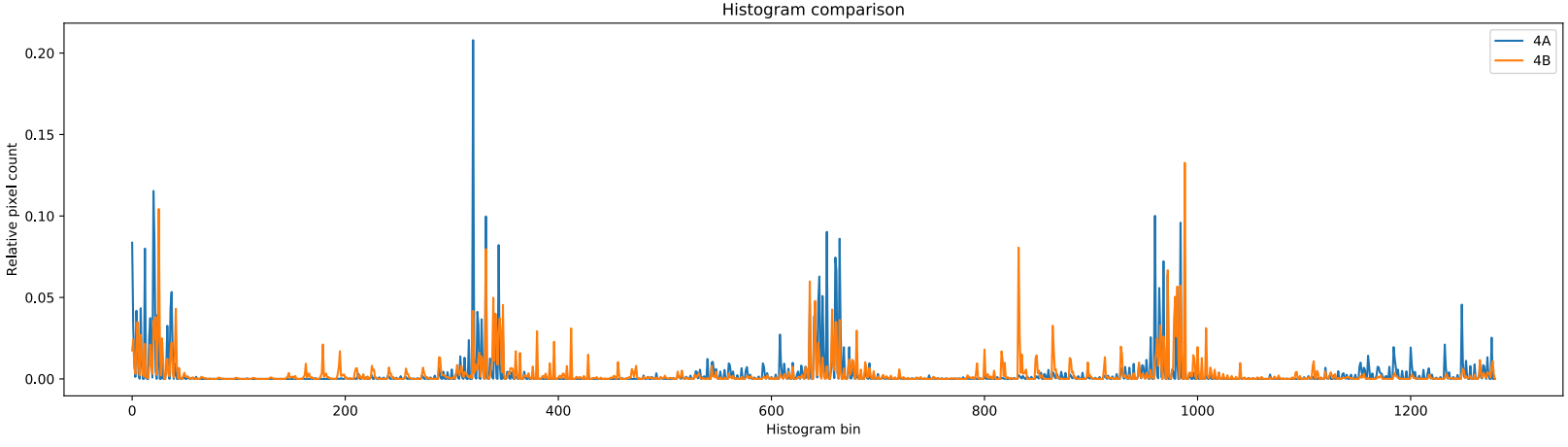


Contribution 2

Classification from low-level descriptors

(a) Room 4A

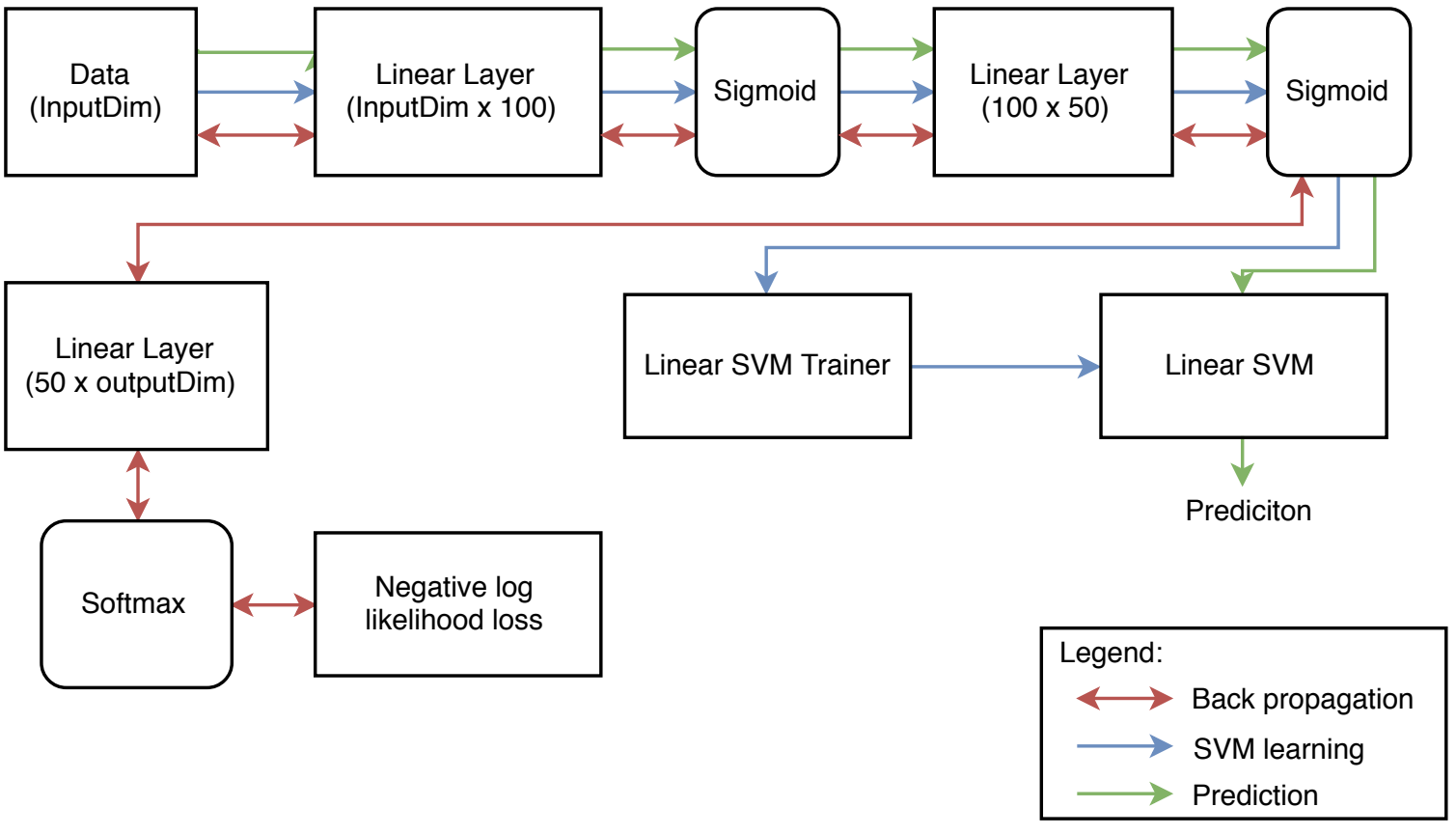
(b) Room 4B



(c) Comparison of HSV histogram on 2x2 grid

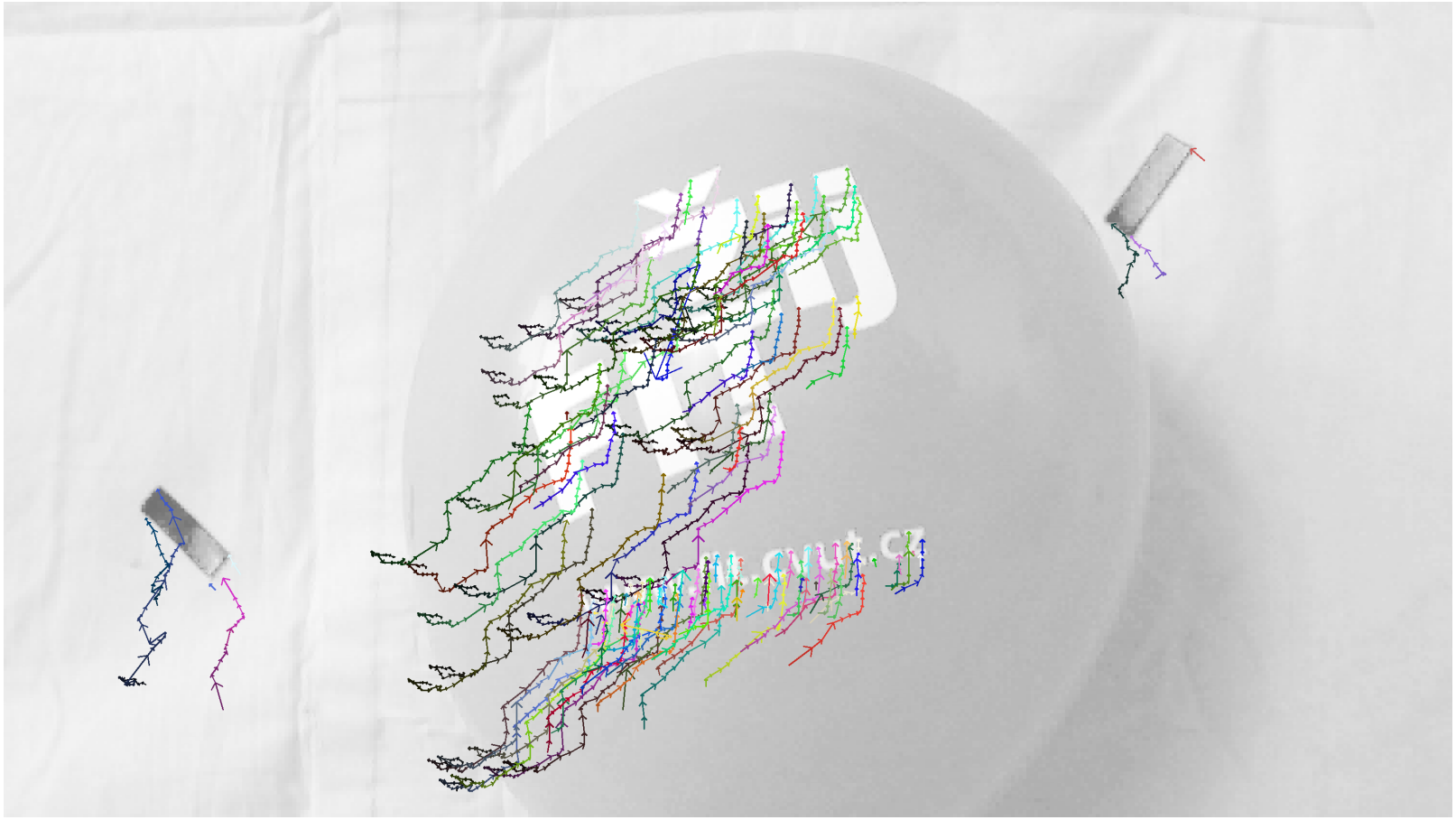
Contribution 2

Classification from low-level descriptors



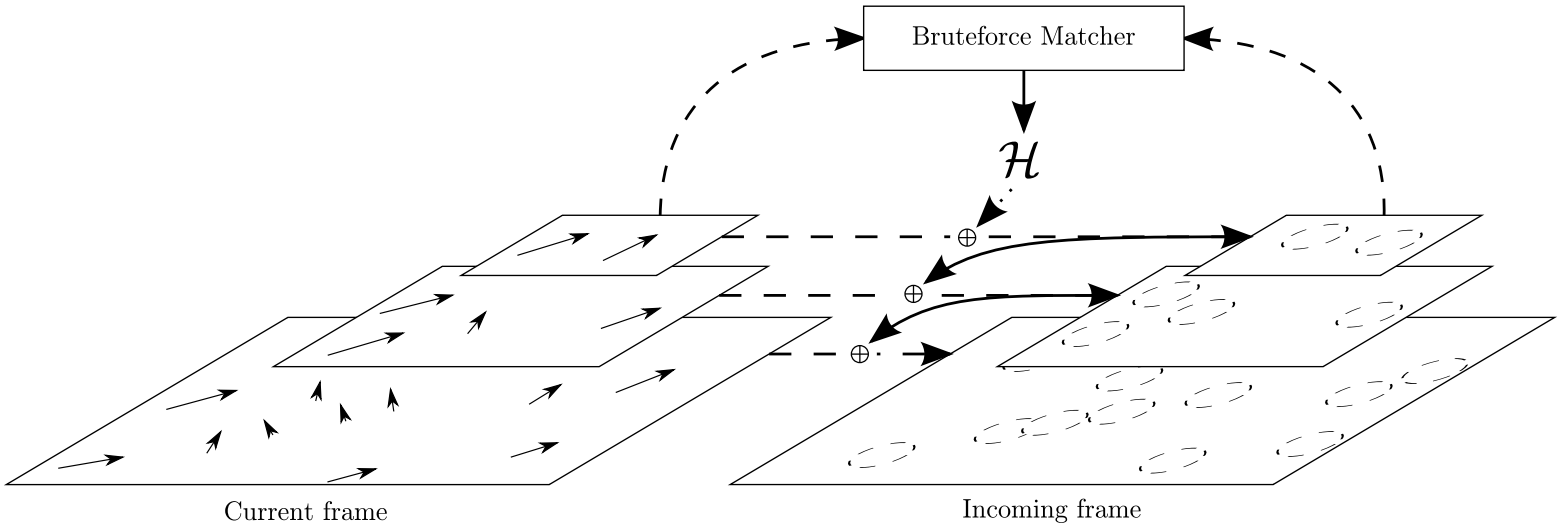
Contribution 3

Leveraging signal redundancy



Contribution 3

Leveraging signal redundancy



Contribution 3

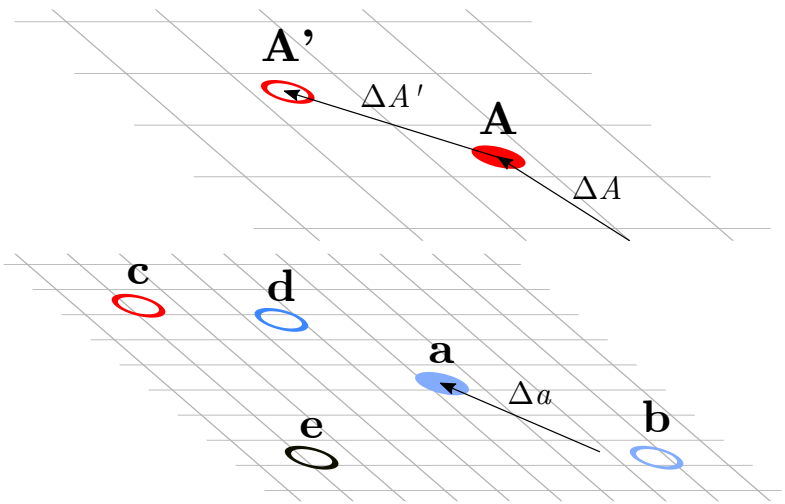
Leveraging signal redundancy

Animation key:

- • Current frame
- ◦ Incoming frame

color ≈ description

1. Compute a salient point motion correction vector $\Delta a - \Delta A$
2. Add result to motion prediction from the layer above $\widehat{\Delta a'}$ and consider only candidates within some spatial distance



Contribution 3

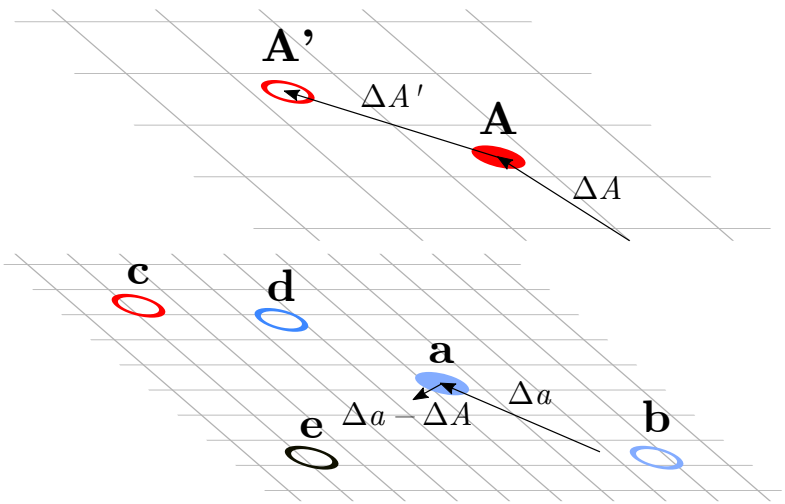
Leveraging signal redundancy

Animation key:

- ● Current frame
- ○ Incoming frame

color ≈ description

1. Compute a salient point motion correction vector $\Delta a - \Delta A$
2. Add result to motion prediction from the layer above $\widehat{\Delta a'}$ and consider only candidates within some spatial distance



Contribution 3

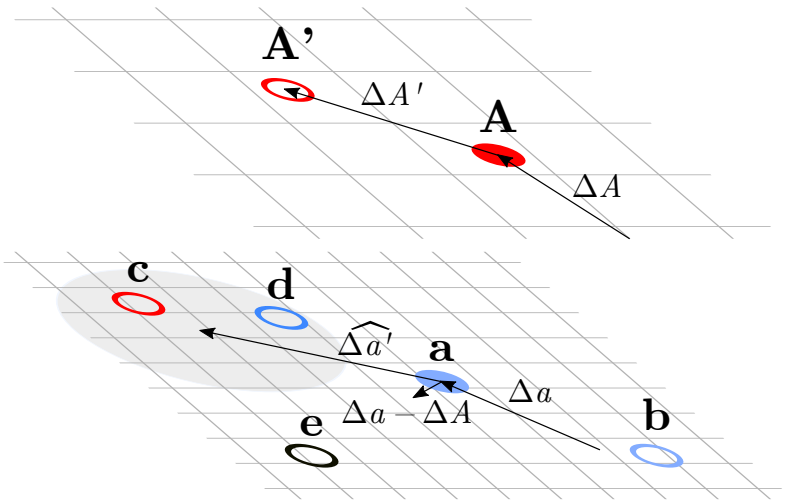
Leveraging signal redundancy

Animation key:

- • Current frame
- ◦ Incoming frame

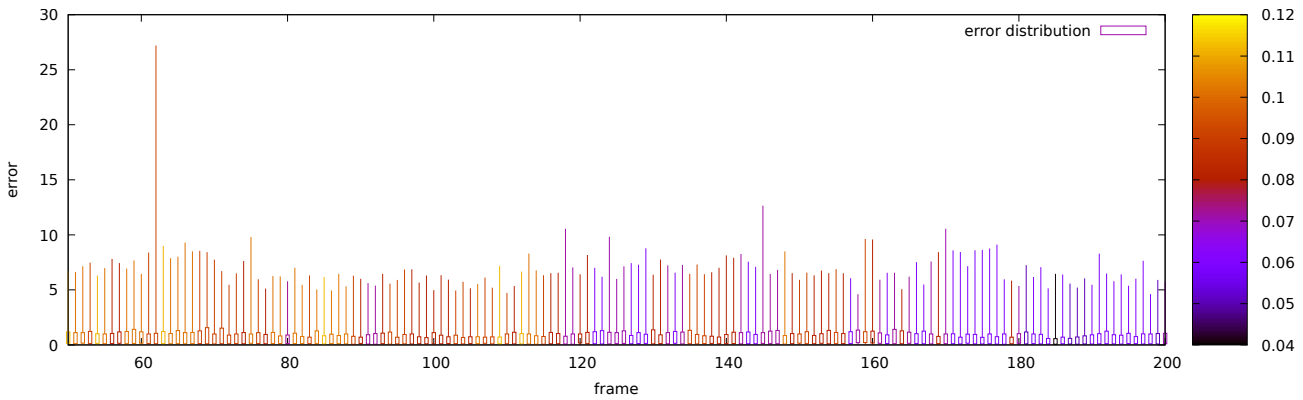
color $\approx$ description

1. Compute a salient point motion correction vector $\Delta a - \Delta A$
2. Add result to motion prediction from the layer above $\widehat{\Delta a'}$ and consider only candidates within some spatial distance



Contribution 3

Leveraging signal redundancy




chair 0.829
handbag 0.772
person 0.843
person 0.999
person 0.967
person 0.828
bench 0.962
person 0.997
person 0.999
umbrella 0.895
person 0.703


vodafone
4
1
5
7
2
3
0
6

Contribution 4

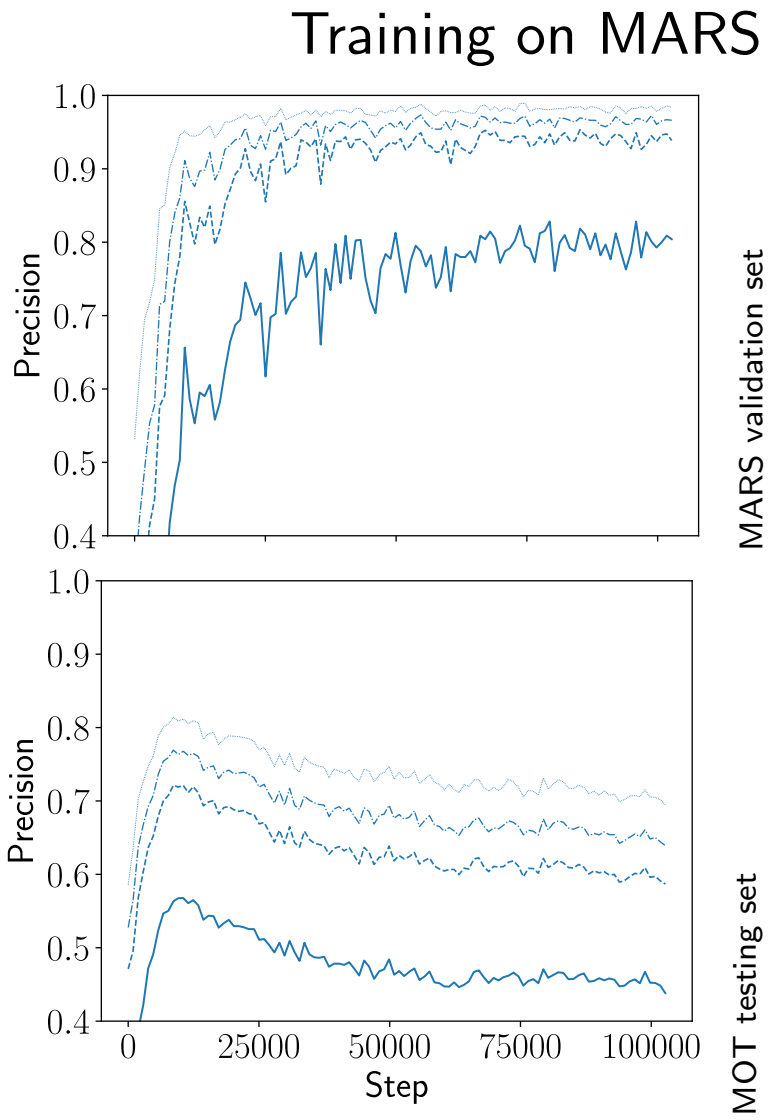
Enhancing supervised methods with unsupervised tracking

Contribution 4

Enhancing supervised methods with unsupervised tracking

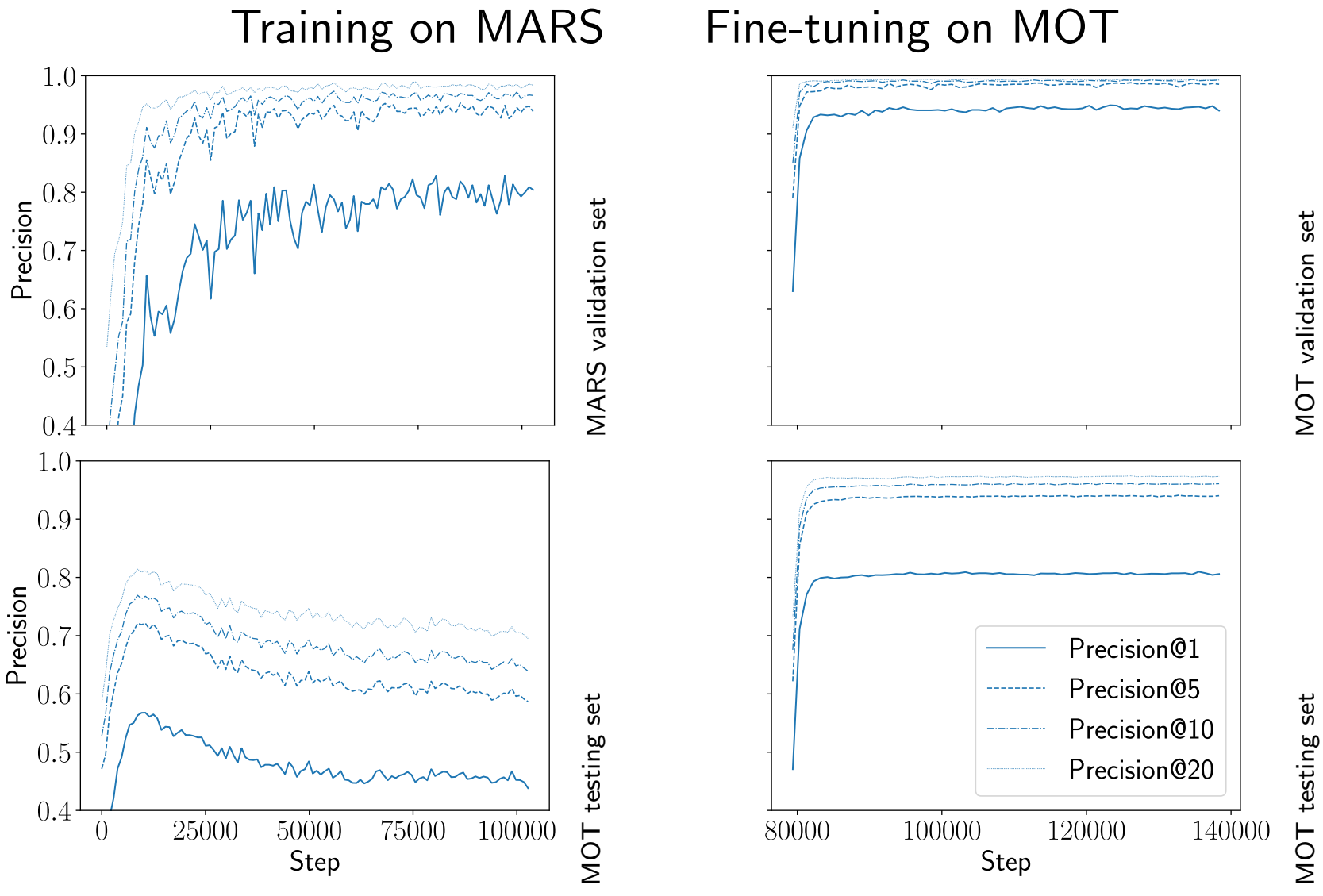
Contribution 4

Enhancing supervised methods with unsupervised tracking



Contribution 4

Enhancing supervised methods with unsupervised tracking



Contributions of the Thesis

1. Theoretical approach to knowledge extraction from various multimedia content
   (hierarchical classification framework)
2. Extraction of semantic features from data aggregates
   (classification methods based on low-level processing)
3. Leveraging information redundancy in the video signal
   (reducing video signal to simpler representations)
4. Inclusion of unsupervised knowledge extraction methods
   (to traditionally strictly supervised approaches)

Thank you for your attention!

Petr Pulc
pulcpetr@fit.cvut.cz

Acknowledgements: The research reported in this thesis has been supported by the Czech Science Foundation (GAČR) grants 13-17187S, 17-01251 and 18-18080S, Grant Agency of the Czech Technical University in Prague SGS17/210/OHK3/3T/18, and by the Institutional Support for Long-term Conceptual Development of Research Organization programme, provided by Ministry of Education, Youth and Sports, Czech Republic.

Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140).

If dimensionality reduction is an optimisation problem, what criteria should be met?

The lowest possible dimensionality that retains the data properties needed for given task.

Cont.

How can one consider histogram representation as one of the dimensionality reduction techniques?

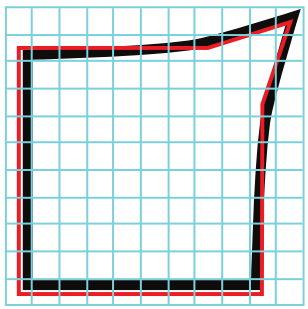
A feature selection technique, which subsamples or completely omits some of the dimensions, retaining only the number of occurrences.

Is "cosine distance" an example of a distance (metric)?

No – does not hold triangle inequality.

However, this term is commonly defined as $1 - S_{cos}(A, B)$ for convenience.

Comment on "Polygonal approximation" in Table 4.2



[0,1], [7,1], [10,0], [9,3], [9,10], [0,10]

How does your hierarchical multimedia processing framework fit in with the meta-learning characteristic that refers to learning algorithms that learn from other learning algorithms?

Meta-learning Selection of a model for the same classifier based on statistical properties of the data.

Our approach Selection of a feature extractor, classifier and its model, based on some properties of the data.

How is the recommendation block trained?

Several approaches:

- Reusing existing classification method for initial labelling
  - and hand-crafting class–method association.
  - and rule-mining class–method association wrt. precision of the recommended classifier.
- Constructing a custom classifier that directly outputs the class of recommended processing.

What is the single novel idea in the doctoral thesis, no one ever came up with before?

Contributions sorted by the amount of novelty at time of publication:

- Leveraging video signal redundancy with a novel hierarchical approach to salient point matching.
- Enhancing supervised method of object "tracking by detection" with datasets created by unsup. object tracking.
- Application of a meta-learning principle to disjoint modalities and processing methods.
- Experiments with scene classification from low-dimensional still-image features and low-level descriptors.

Have you compared your methods of **video scene recognition** with any state-of-the-art competitors?

In [Šabata, T., Pulc, P., Holeňa, M.: Semi-supervised and Active Learning in Video Scene Classification from Statistical Features. IAL 2018 — ECML-PKDD 2018, 1613-0073.] we compared our approach with state-of-the-art inception-style neural network [Wang, L., Guo, S., Huang, W., Xiong, Y., Qiao, Y.: Knowledge guided disambiguation for large-scale scene classification with multi-resolution CNNs. IEEE Transactions on Image Processing, 2055-2068.] and concluded that the LSUN challenge winner is unfit for classification of specific scenes, although capable for scene type classification.

Have you compared your methods of **multiple object tracking** with any state-of-the-art competitors?

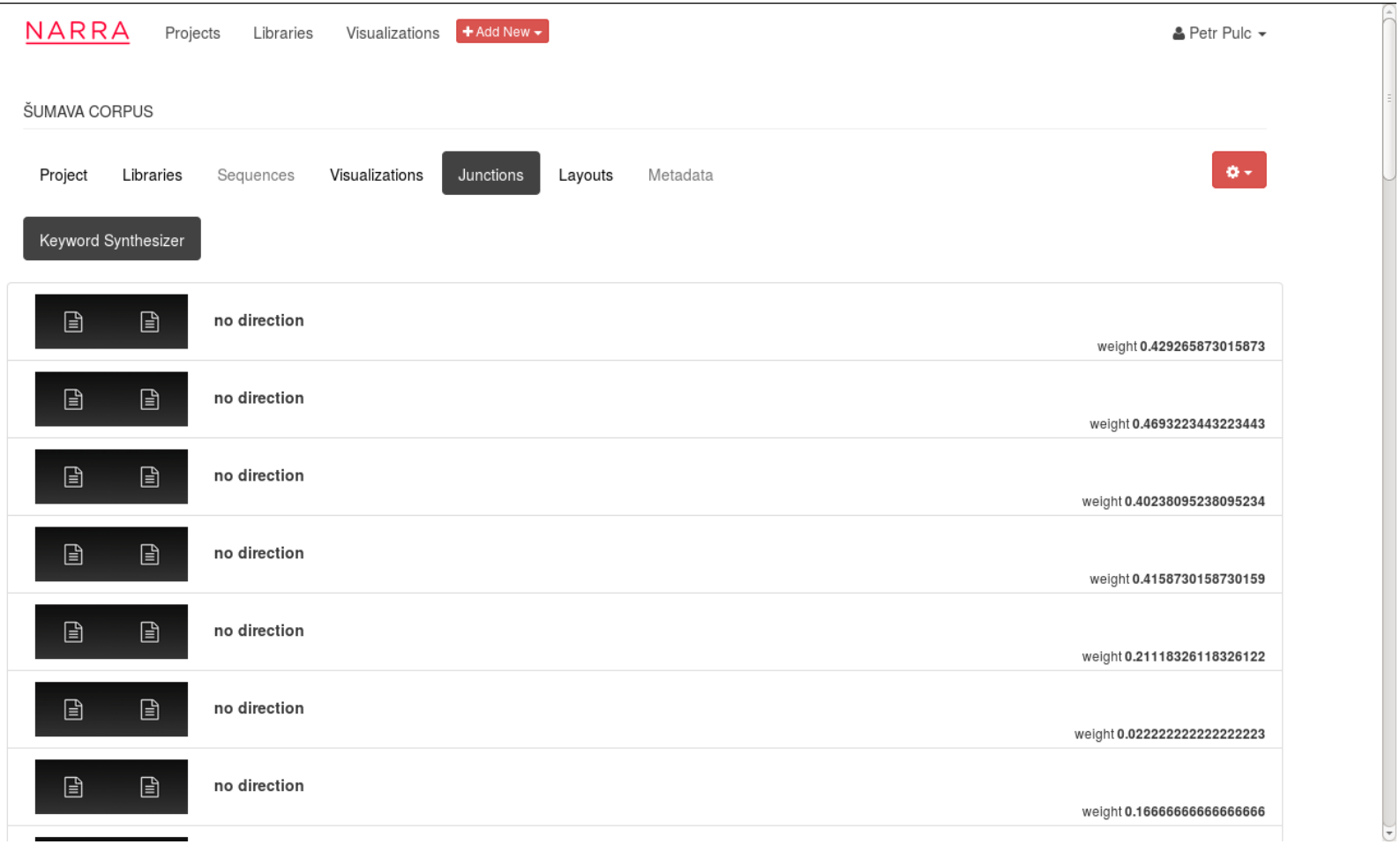
In [Pulc, P., Holeňa, M.: Towards Real-time Motion Estimation in High-Definition Video Based on Points of Interest. FedCSIS 2017, 2300-5963.] and [Pulc, P., Holeňa, M.: Hierarchical Motion Tracking Using Matching of Sparse Features. SITIS 2018, 978-1-5386- 9385-8.] we compared the internal principles and performance with a baseline [Lucas, B. D., Kanade, T.: An iterative image registration technique with an application to stereo vision. IJCAI'81] and concluded that the results are comparable, but we can obtain them much faster on higher image resolution.

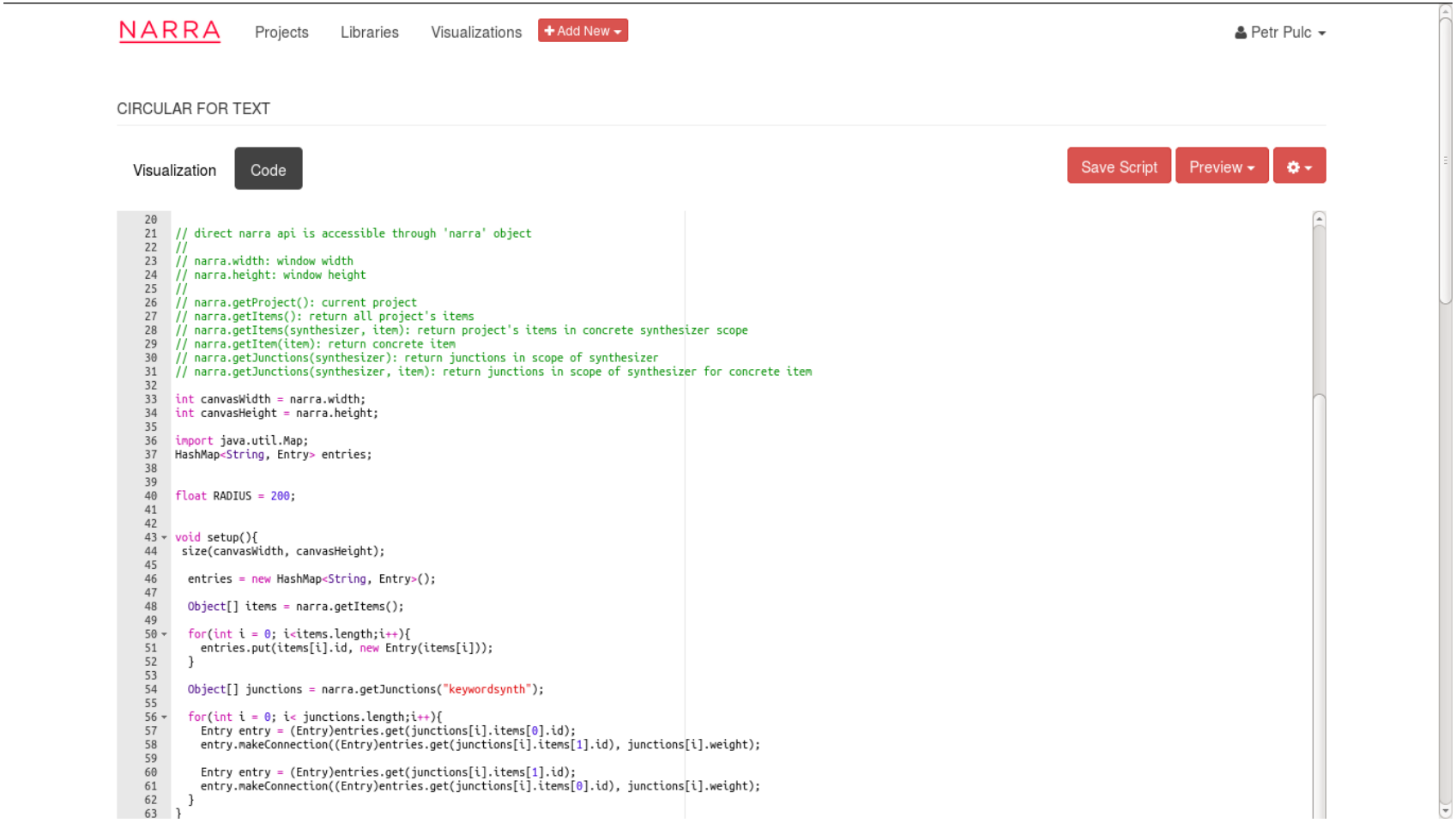
Can you discuss relation to Content-Based Video Retrieval problems?

The ultimate goal of our research is to provide the indexing approaches for video retrieval systems with:

- lower computational requirements, and
- increased efficacy.

To meet this goal we proposed the utilisation of a meta-learning approach, where some of the classification power will be dedicated towards selection of the processing methods with highest information gain. Cont.


NARRA
Projects
Libraries
Visualizations
+ Add New
Petr Pulc
ŠUMAVA CORPUS
Project
Libraries
Sequences
Visualizations
Junctions
Layouts
Metadata
Keyword Synthesizer
no direction
weight 0.429265873015873
no direction
weight 0.4693223443223443
no direction
weight 0.40238095238095234
no direction
weight 0.4158730158730159
no direction
weight 0.21118326118326122
no direction
weight 0.022222222222222223
no direction
weight 0.16666666666666666


NARRA
Projects
Libraries
Visualizations
+ Add New
Petr Pulc
CIRCULAR FOR TEXT
Visualization
Code
Save Script
Preview
// direct narra api is accessible through 'narra' object
//
// narra.width: window width
// narra.height: window height
//
// narra.getProject(): current project
// narra.getItems(): return all project's items
// narra.getItems(synthesizer, item): return project's items in concrete synthesizer scope
// narra.getItem(item): return concrete item
// narra.getJunctions(synthesizer): return junctions in scope of synthesizer
// narra.getJunctions(synthesizer, item): return junctions in scope of synthesizer for concrete item
int canvasWidth = narra.width;
int canvasHeight = narra.height;
import java.util.Map;
HashMap<String, Entry> entries;
float RADIUS = 200;
void setup(){
size(canvasWidth, canvasHeight);
entries = new HashMap<String, Entry>();
Object[] items = narra.getItems();
for(int i = 0; i<items.length;i++){
entries.put(items[i].id, new Entry(items[i]));
}
Object[] junctions = narra.getJunctions("keywordsynth");
for(int i = 0; i< junctions.length;i++){
Entry entry = (Entry)entries.get(junctions[i].items[0].id);
entry.makeConnection((Entry)entries.get(junctions[i].items[1].id), junctions[i].weight);
Entry entry = (Entry)entries.get(junctions[i].items[1].id);
entry.makeConnection((Entry)entries.get(junctions[i].items[0].id), junctions[i].weight);
}
}

NARRA

ŠUMAVA CORPUS