Automated Object Retrieval from Large Video Surveillance Datasets

Cătălin Alexandru Mitrea[1,2]

[1] LAPI & Natural Computing Labs, University ”Politehnica” of Bucharest, 061071, Romania

[2] UTI GRUP, Bucharest, 020492, Romania

MMV Seminar - 04/03/2015

Outline

- Introduction (University & Company)
- Previous work
  - Problem statement
  - Proposed system
  - Surveillance datasets (Scouter & PEViD)
- Current work (MMV – QMUL)
  - DROP task
  - Main directions/ideas (initial results)
- Conclusions and future work
  - Acknowledge

Introduction (LAPI Lab)

- *Politehnica University of Bucharest*, Faculty of Electronics, Telecommunications and Information Technology
  - Image Processing and Analysis Lab (LAPI)

Research Areas

✓ Mathematical models (probabilistic / statistical, fuzzy, etc.);

✓ Color and multispectral image & video processing;

✓ Indexing and content-based retrieval algorithms for image and video databases.

✓ Parallel systems and fast algorithms for signal processing;

Introduction (Softrust Company)

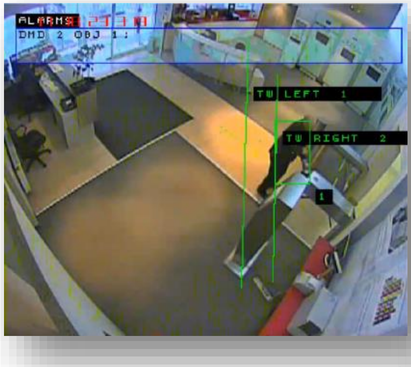
- *UTI Grup,* Softrust Vision Analytics Division
  - Young and specialized team in video analytics applications:

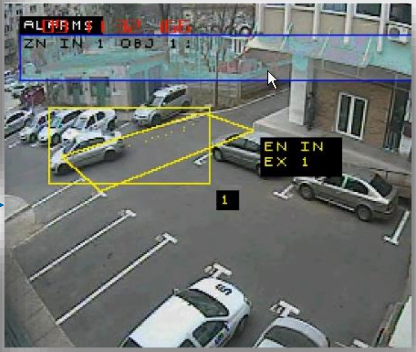
- Safety Zone
- Trip Wire
- Facial Recognition
- Auto Tracking

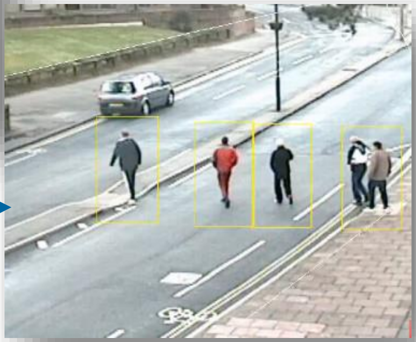
- Illegal Stopping / parking
- Detection of wrong direction

- Abandoned objects
- Missing objects

- Forensics







Previous work - premise

- High volume of video acquisition (~4mil CCTV cameras only in UK);
- Limited human resources.

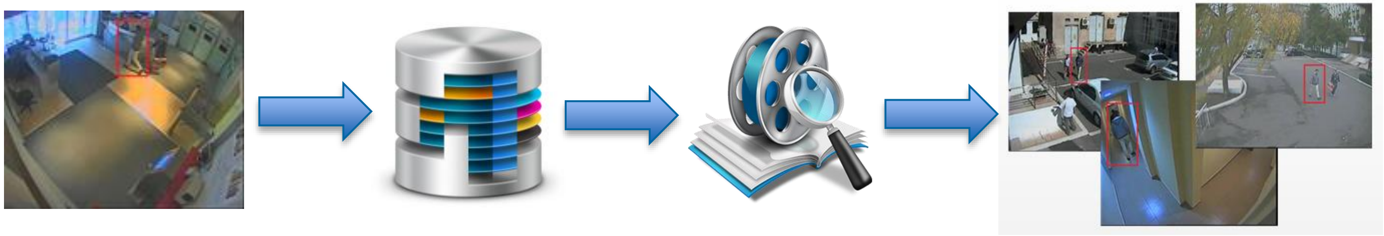
Solutions

- Intelligent video surveillance techniques:
  - Real-time identification and tracking of object of interest;
  - Behavior and incident detection;
  - Crowd analysis;
  - Content-based offline searching and indexing of objects (humans).

Previous work - Problem & Objectives

Problem statement

- Starting from a small sample (few frames) of the object to-be-found (human) => find (search) all relevant instances into a vast multisource video database.



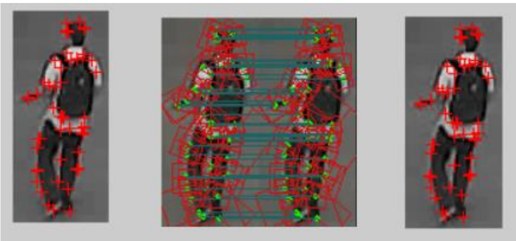
Objectives:

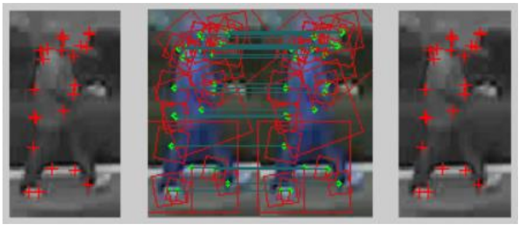
- Develop a system for providing content-based search capabilities within multi-source video surveillance footage.
- Introduce an indexed dataset containing surveillance videos recorded in a real public institution (Scouter DB).

Previous work - Trends in literature

Main methods and directions

- Large video databases processing techniques [Snoek,IEEE 2010];
- Content descriptors extraction (color, texture, shape, temporal and motion, audio [Ionescu, LNCS 2011]);
- Feature points (SIFT, SURF [Stottinger, IEEE 2010]);
- Fusion (BoW, Boosting, fisher kernel representations [Mironica, ACM 2013]).
- Intuitive interfaces for video query and data mining [Shah, IEEE MultiMedia 2007];





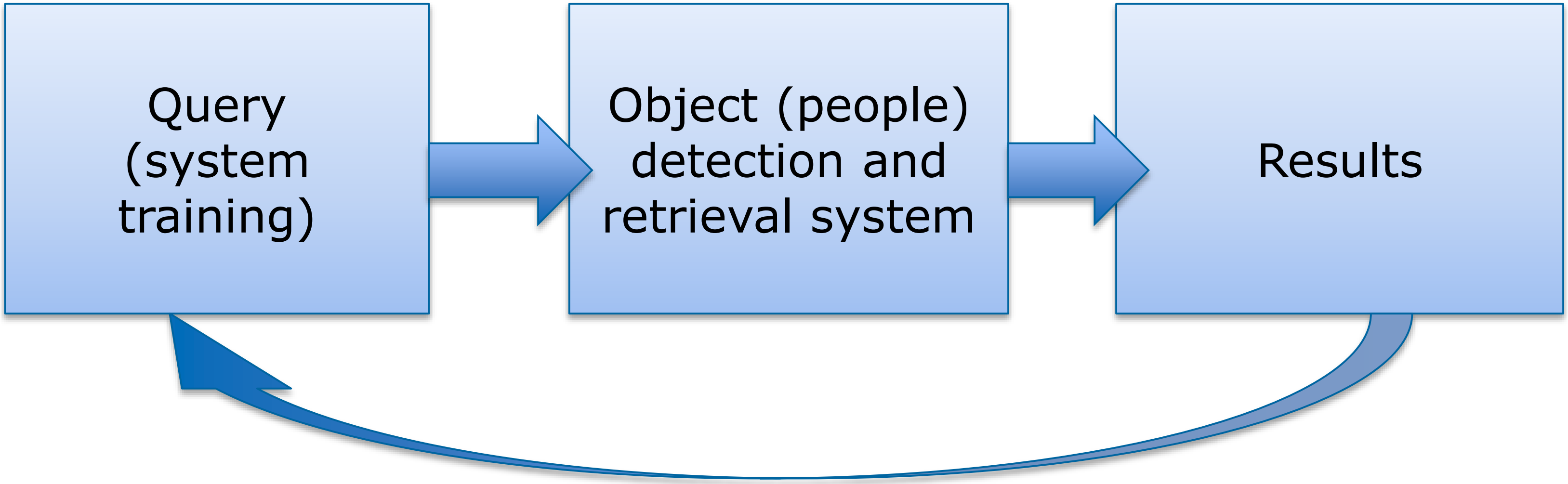
Drawbacks

- Computation complexity
- Difficult to implement for "real field" systems
- Not all methods are suitable for video surveillance datasets

perspectives - e.g., multiple source cameras, different weather conditions, different setups - e.g., indoor vs. outdoor, appearances, etc.

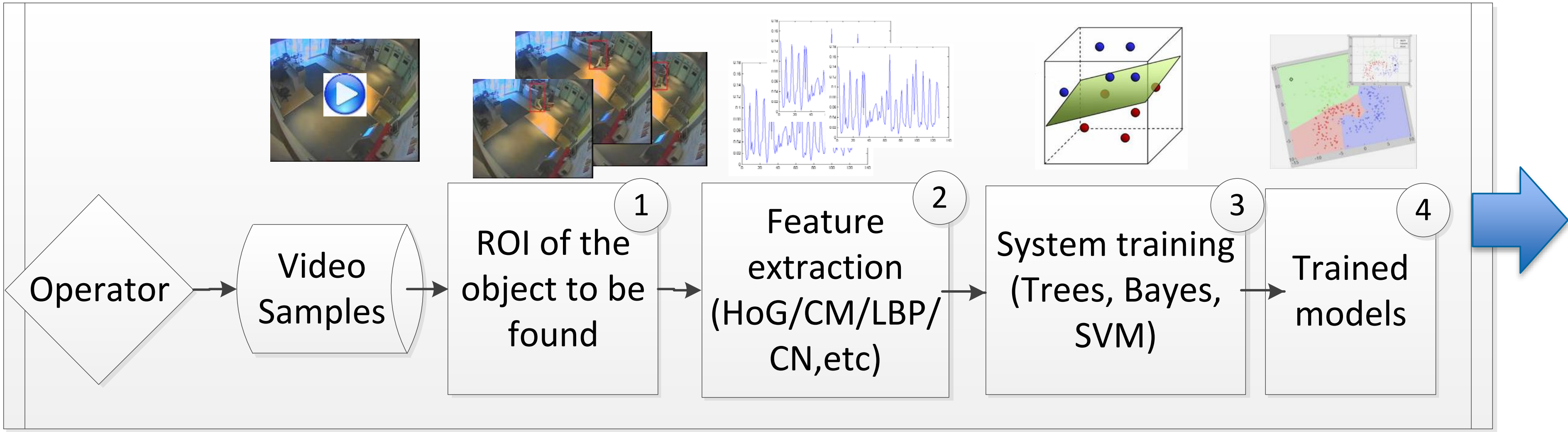
Previous work - Proposed system

- Block diagram



Previous work - Proposed system (2)

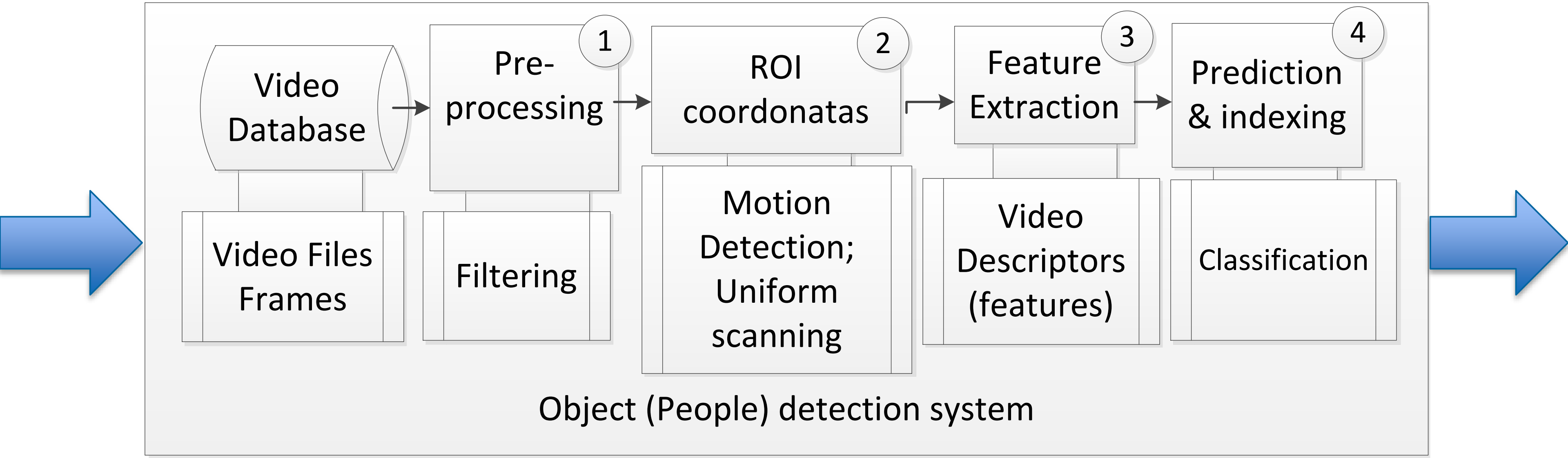
- Query (system training)



Operator → Video Samples → 1 ROI of the object to be found → 2 Feature extraction (HoG/CM/LBP/CN,etc) → 3 System training (Trees, Bayes, SVM) → 4 Trained models

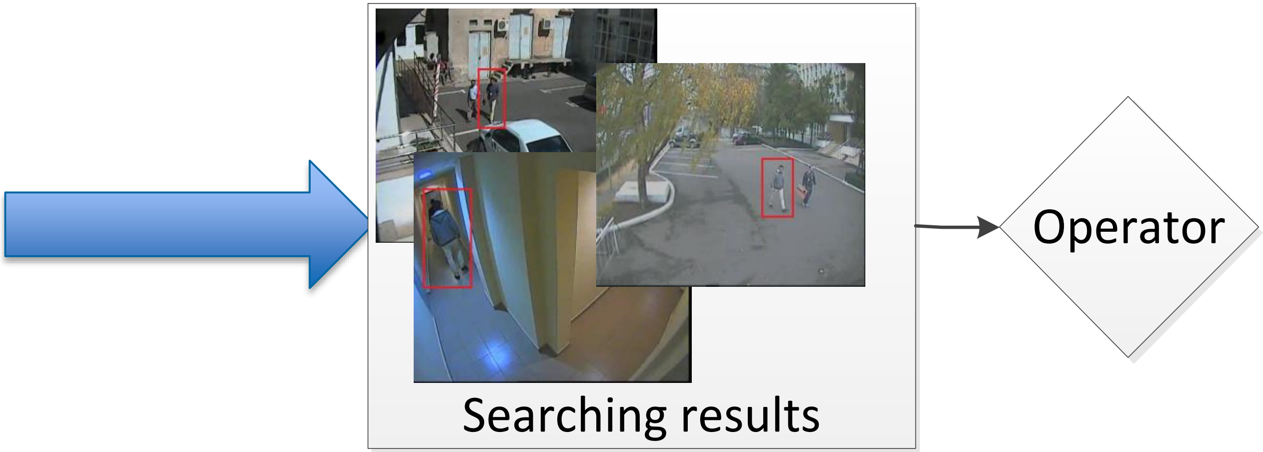
Previous work - Proposed system (3)

- Object (people) detection and retrieval system



Previous work - Proposed system (4)

- ➢ Results



- ▪ Different instances of the object-to-be-found returned to user from the entire database

Previous work - Content descriptors (know-how)

- **CN** features (Color Naming histogram – color descriptor) [Van De Weijer, CVPR 1994];

  11 colors distribution: "black", "blue", "brown", "gray", "green", "orange", "pink", "purple", "red", "white" and "yellow".

- **CM** (Color Moments –color descriptor) [Stricker, SPIE 1995];

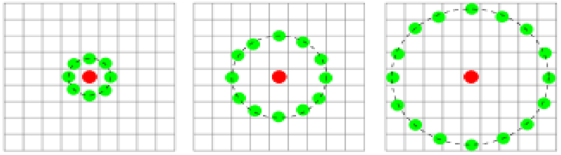
  Color similarity. Three central moments of an image’s color distribution: mean, standard deviation and skewness.

- **CSD** (Color Structured Descriptor –color descriptor) [Ojala, ICPR 2002];

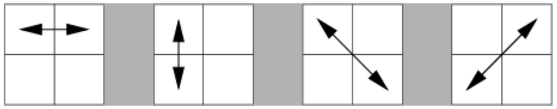
  Color accumulation and local spatial distribution of colors.

Previous work - Content descriptors (know-how)

- **LBP** (Local Binary Pattern – texture descriptor) [Ojala, ICPR 1994];

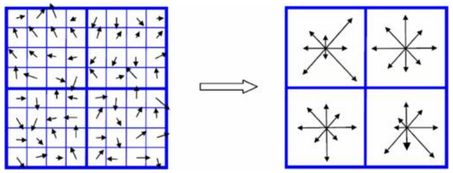


- **Haralick** (texture descriptor) [Haralick, TSMC 1973];



->co-occurrence matrices generated using each of these directions

- **HoG** features (Histogram of Oriented Graphs – shape-based descriptor) - [Dalal, CVPR 2005]

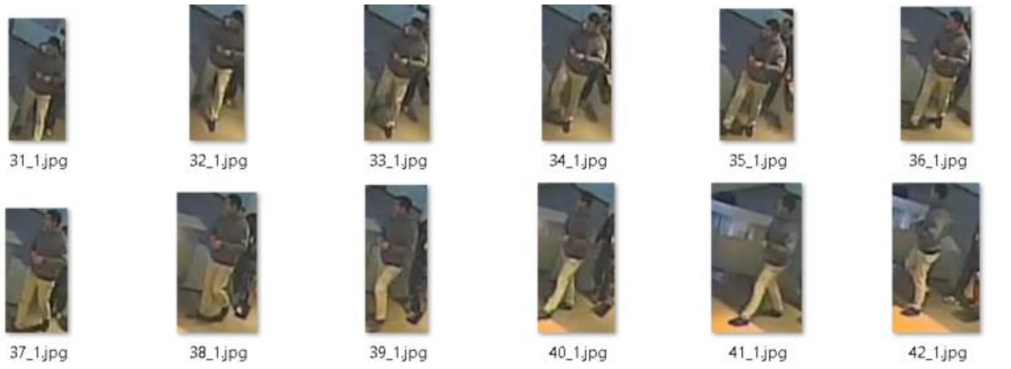


- **SIFT | SURF** descriptors (Scale-invariant feature transform | Speeded Up Robust Features) [Lowe, ICCV 1999 | Herbert, ECCV 2006];

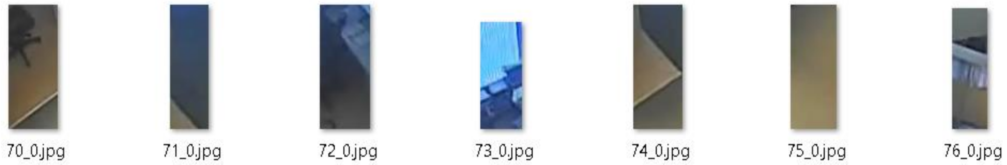
Previous work – Classifiers (know-how)

- 6 classifiers used (5 standard and 1 proposed)

1. Naive Bayes
2. Nearest Neighbor
3. Decision Trees
4. Random Forests
5. Support Vector Machines
6. FSVC – Fast Support Vector Classifier



E.g. of 12 training samples of true class

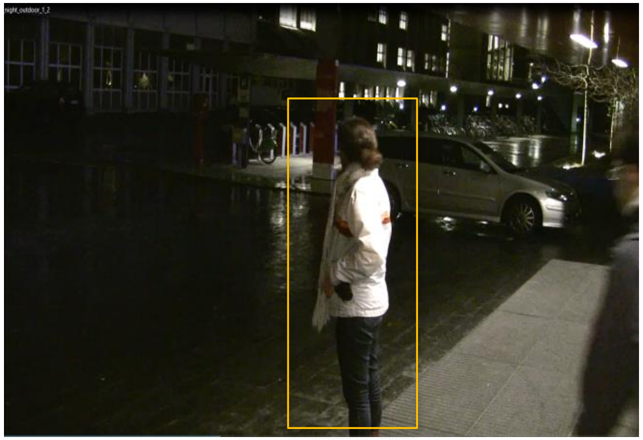


E.g. 7 training samples of false class

Previous work - Datasets

- Manually indexed SCOUTER Database
  - 30 video files (3 different days x 10 cameras @ 6 to 10 FPS
  - 704 x 576 resolution
  - ~36,000 annotated frames (2 people–scenarios!)
  - humans varies from 50 x 50 pixels to ~250 x 350 pixels.

- Manually indexed PEViD-HD Database
  - 21 video files
  - recorded at ~ 25 FPS
  - FullHD resolution
  - ~10,000 annotated frames (14 people–scenarios!)
  - humans varies from ~200 x 300 pixels to ~650 x 850 pixels.



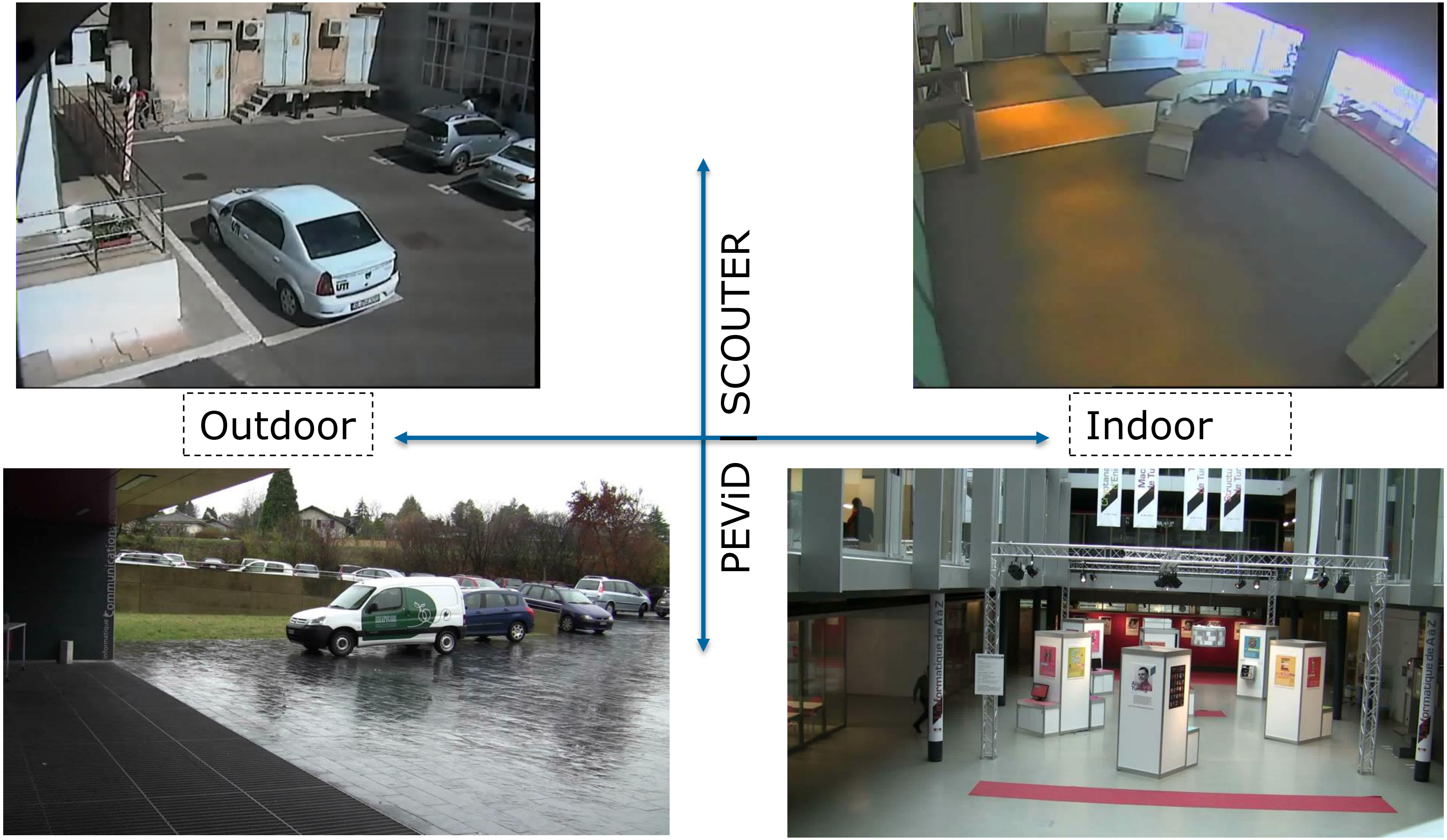
Scouter free download link: http://uti.eu.com/pncd-scouter/rezultate.html
PEViD free download link: http://mmspg.epfl.ch/pevid-hd

Previous work - Datasets (Comparison)

- With other standard video datasets

	KTH	Weizmann	Scouter	PEViD -HD
Max. Resolution (W x H)	160 x120	180 x 144	704 x 576	1920 x 1080
Human Height in Pixels	80 – 100	60 - 70	50 - 350	200 - 850
Human to video height ratio	65 to 85%	42 to 50%	10 to 60%	15 to 65 %
Scenes Viewpoint Type	Side	Side	Varying	Varying
Natural Background Clutter	No	No	Yes	Yes
Incidental Objects/Activities	No	No	Yes	Yes
Multiple annotations on movers	No	No	Yes	Yes

Video dataset examples



Current work (MMV – QMUL)

➢ DROP - *Distinctive regions of patterns task;*

✓ Briefly the task consist of accurately identifying the same regions or patterns of interest in a large set of images from videos, starting from just a few (one) samples as an example.

✓ Patterns of interest include color regions, tattoos, logos and any *other distinctive feature* that appear in a given anchor image.

Current work (Scotland Yard DB)

- Main issues
  - PTZ cameras (high image shifting/zoom from frame to frame – unsuitable or difficult for motion based detection);
  - Low image quality and high noise (need for adaptable & robust feature extraction algorithms);
  - Different perspectives (need for scale invariant feature extraction algorithms);
  - Very few (or one) sample used as reference for searching (or training).

DROP Task

❑ Samples

- ➢ Stealing in supermarket scenario;
- ➢ Find DROP (e.g. “hat”):

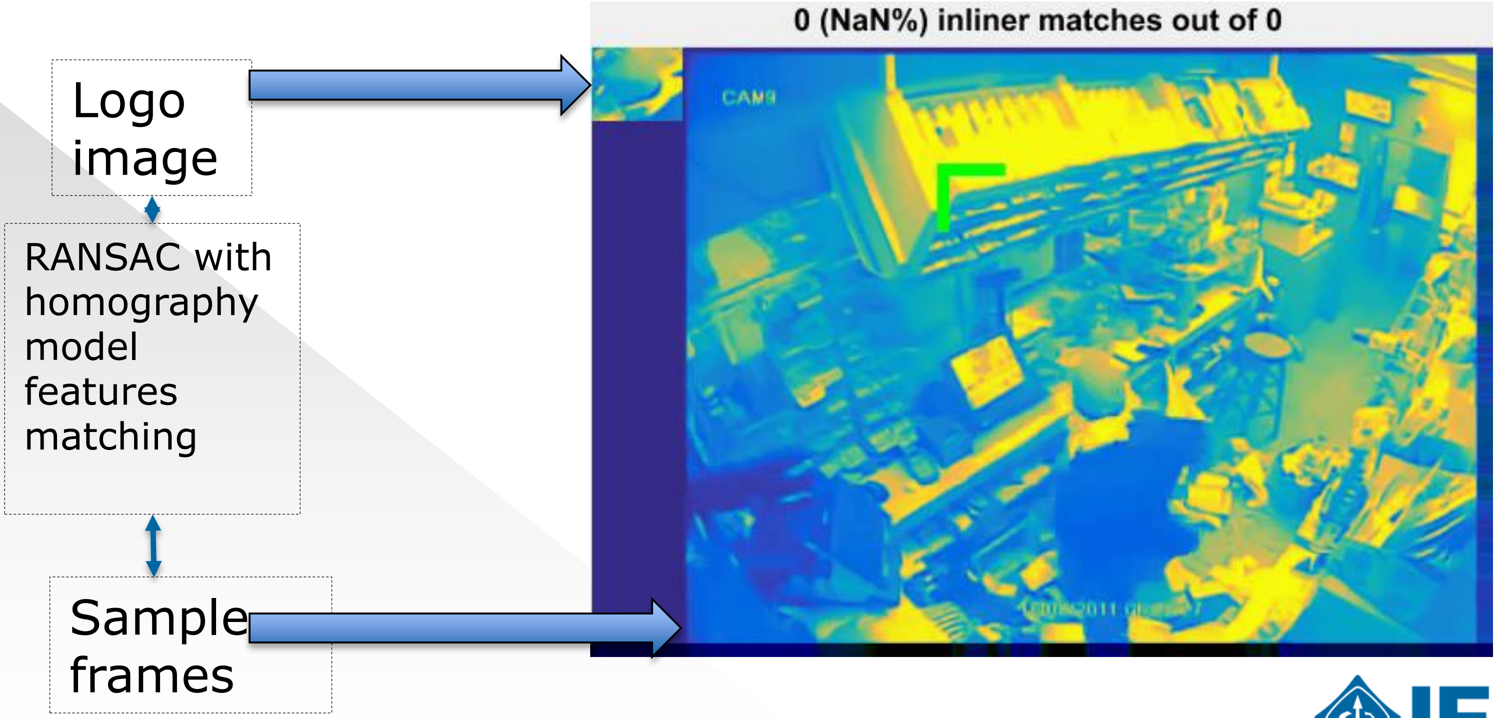
~ 1000 frames containing thief with

“hat” drop.

Source Youtube: https://www.youtube.com/watch?v=wYRiSO_VyF8)

DROP Task (directions – 1)

□ Feature Points based (dense SIFT, moSIFT, SURF, ORB, etc.)



Logo image

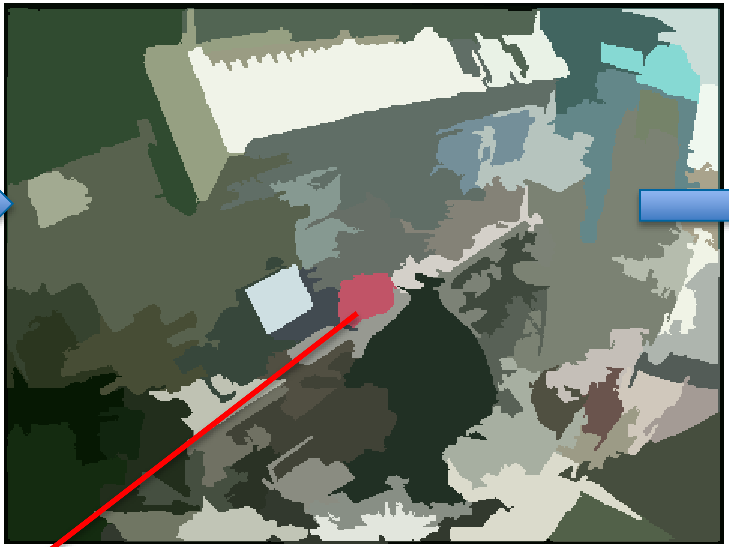
RANSAC with homography model features matching

Sample frames

DROP Task (directions – 2)

- Segmentation -> Classification

Original Image

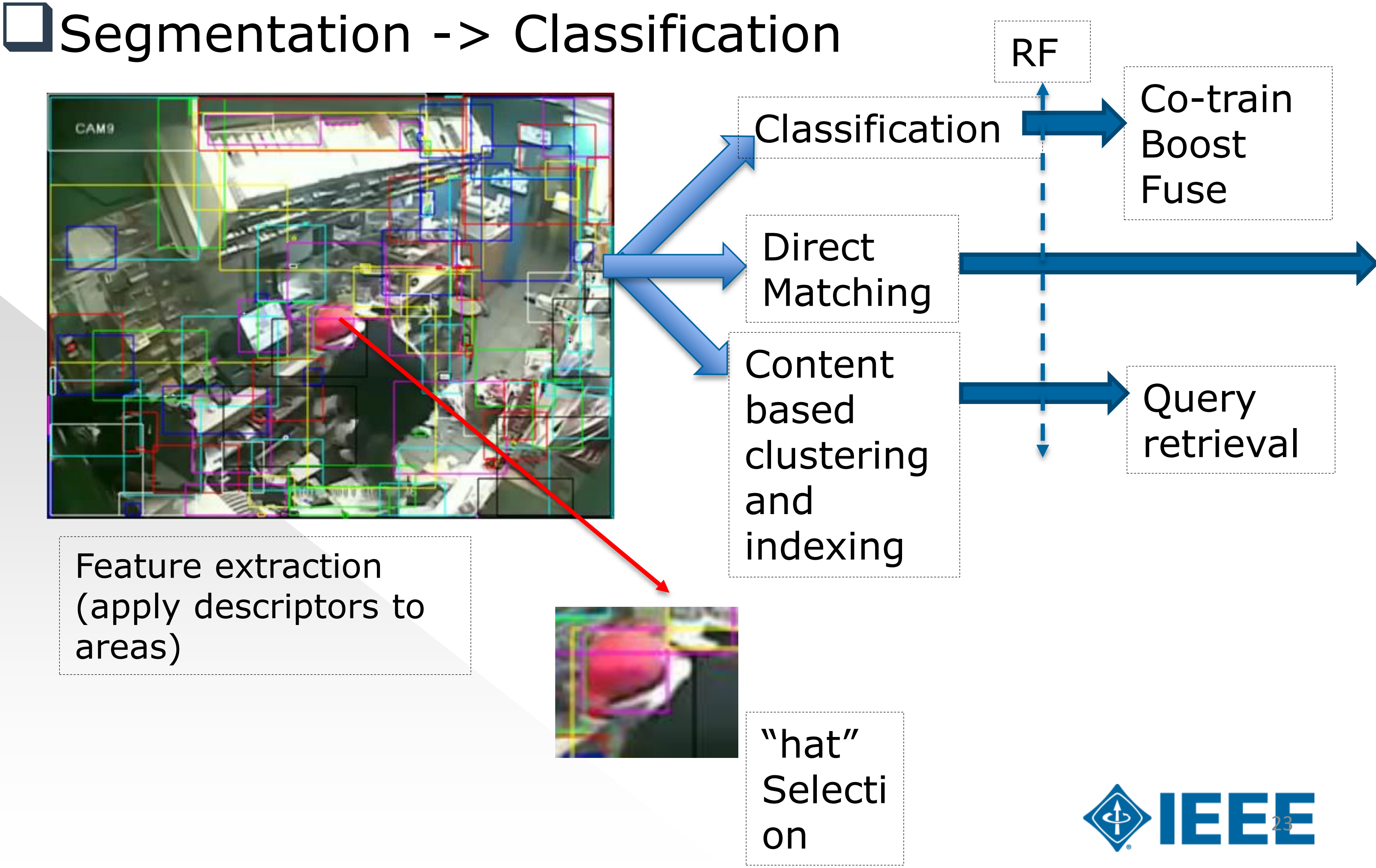


“Hat” Segment

Segmentation Mean Shift based

DROP Task (directions – 2)

- Segmentation -> Classification



RF

Classification

Co-train
Boost
Fuse

Direct Matching

Content based clustering and indexing

Query retrieval

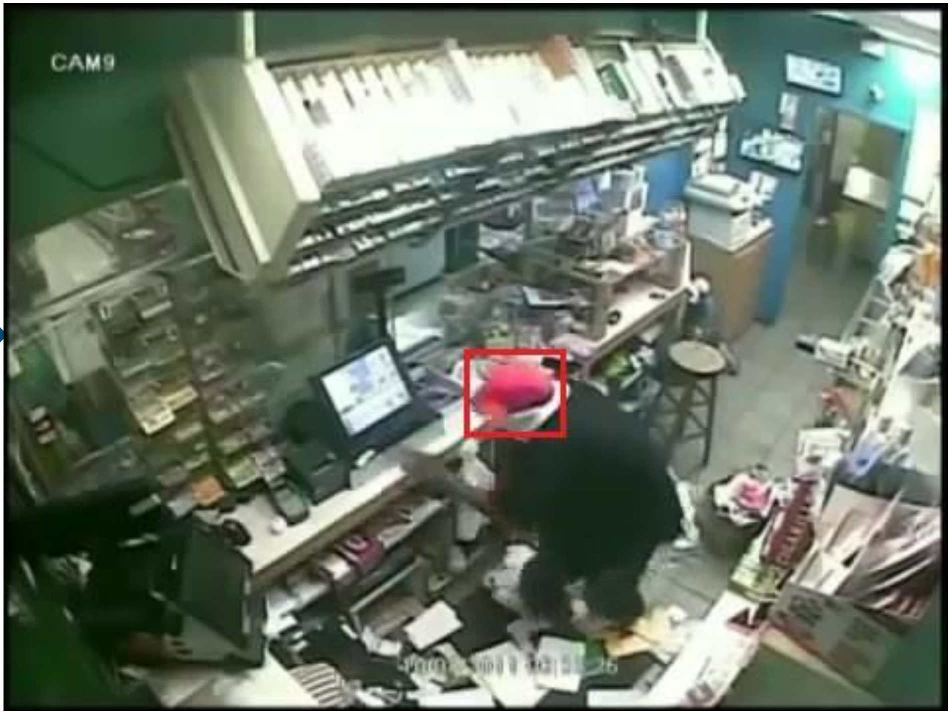
Feature extraction (apply descriptors to areas)

“hat” Selection

DROP Task

- ❑Example

Two descriptors agreement (chi-square distance based matching)



DROP Task (directions – 2)

❑Challenges & Initial Conclusions

- **For real-world datasets (Scotland DB) “academic” state-of-the art algorithms needs to be adapted to new challenges:**
  - How to deal with low quality/noisy data sets (image enhancement for some alg. can decrease performance!);
  - How to “learn” starting from few (or one) samples (fusion, boosting or co-training techniques might be suitable for task);
- **The classification-based approach seems a suitable perspective to solve multi-instances object retrieval (search) if there are enough samples to train the “decisioners”.**
  - Artificial sampling algorithms are investigating (Steerable Pyramids, Dual-Tree Complex Wavelet Transform, etc.);
- **For performance assessment, proper and relevant ground truth needs to be developed.**

Acknowledge

- This work is supported by the SCOUTER project (PN-II-IN-DPST-28DPST/30.08.2013)

Thank you!