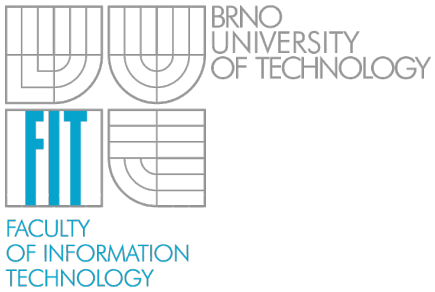
SUNAR

Surveillance Network Augmented by Retrieval

Petr Chmelař

Brno University of Technology, Faculty of Information Technology
Bozetechova 2, 612 00 Brno, CZ
www.fit.vutbr.cz/~chmelarp



AVSS 2009, Genova

Outline

0. Abstract
1. Introduction
   - Motivation
   - History
   - Goals
2. SUNAR system
   - Data source
   - Vision module CVM
   - Retrieval module VRM
     - cleaning
     - persistence
     - training
     - integration
   - Human interface HMI
3. Evaluation experiments
4. No conclusions, just future work

Abstract

The work dealt with an information retrieval based wide area surveillance system Sunar (Surveillance Network Augmented by Queries) being developed as an open source software at FIT BUT. It includes many experimental techniques to be evaluated by NIST at the Multiple Camera Person Tracking Challenge as a part of the AVSS 2009 Conference.

In brief, we used OpenCV Library for tracking in Computer Vision Modules processing the surveillance video. We have improved some methods and added feature extraction... to the library. Information about objects and the area under surveillance is cleaned, integrated, indexed and stored in Video Retrieval Modules. It is based on the PostgreSQL database that is extended to be capable of similarity and spatio-temporal information retrieval and some data analysis and mining in global context.

Motivation

Evening Standard (2007) shows statistics of crime-fighting CCTV cameras in Great Britain: The country's more than 4.2 million CCTV cameras catch each British resident as many as 300 times each day.

BBC News (in 2007) informed that half a million pounds a year is spent on cameras helping to pick up litter. Yet 80% of crime is unsolved.

We think that high quality crime investigation is the best prevention.

History

Faculty of IT, Brno University of Technology has started to develop an IR-based multi-camera tracking system in 2006 to be at the top of the state of the art and to be useful in any accident investigation.

During the time the team had about 10 people (as I remember, unsorted, without short-term students' projects):

Martin Heckel, Igor Potucek, Stana Sumec, Vita Beran, Pavel Zak, Ivo Reznicek, Ales Lanik, Josef Mlich, Honza Navratil, Vasek Simek, Jaroslav Zendulka, Karel Chytka, Petr Castek, ...

Content Analysis and Retrieval Technologies to Apply Knowledge Extraction to massive Recording

We have solved EU FP7 CARETAKER (Rome and Torino metro data), NIST Event detection pilot (Gatwick data) and we have some private data (Brno metropolitan police, faculty surveillance system, ...)

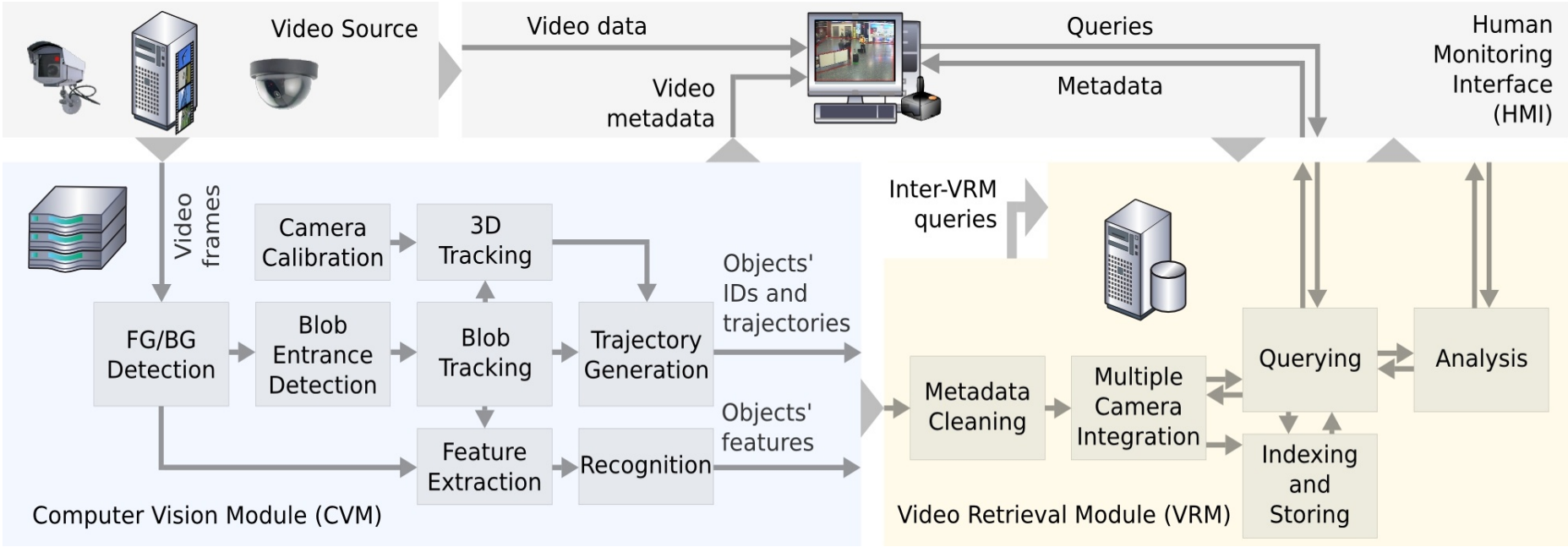
Goal

The goal is simply to perform real-time tracking, object and event detection which is producing metadata;
to clean, integrate, index and store the metadata to be able of querying and analyzing it.

The information requests are of two types:

- On-line for identity preservation, event detection (delayed real-time)
- Off-line to query all the metadata from processed camera records in the wide area when an accident, crime, a natural or human disaster occurs (no way to real-time)

System design



SUNAR architecture

There is no need for:

- calibration
- central module or database
- special hardware

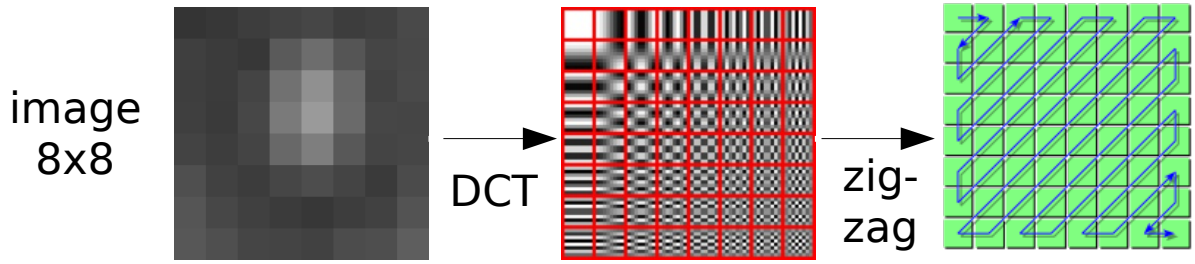
System - vision

The video source is based on anything digital (own FFMPEG util)...

The output is *metadata of objects* and the environment - *local identification* of objects, *spatio-temporal* (position, dimension, speed) and *visual description* (color layout, ...)

The SUNAR's CVM module is based on OpenCV, except:

- BG/FG detection (based on RGB cone beginning at #000 [])
- feature extraction while trajectory generation (MPEG-7)



Ready, but not used - not helpful (... everybody wears black):

- other descriptors (texture, face, SIFT, MSER, SURF)
- object (and event) recognition
- network video stream capture/server (FFMPEG)
- 3D calibration

[] Carmona E.J., Martinez-Cantos J. and Mira J. 2008. A new video segmentation method of moving objects based on blob-level knowledge. Pattern Recognition Letters. Volume 29. p 272-285.

System – retrieval

The input is CVM metadata and *queries* (other VRMs and HMIs) - it doesn't process any video (and thus may be programmed in Java)

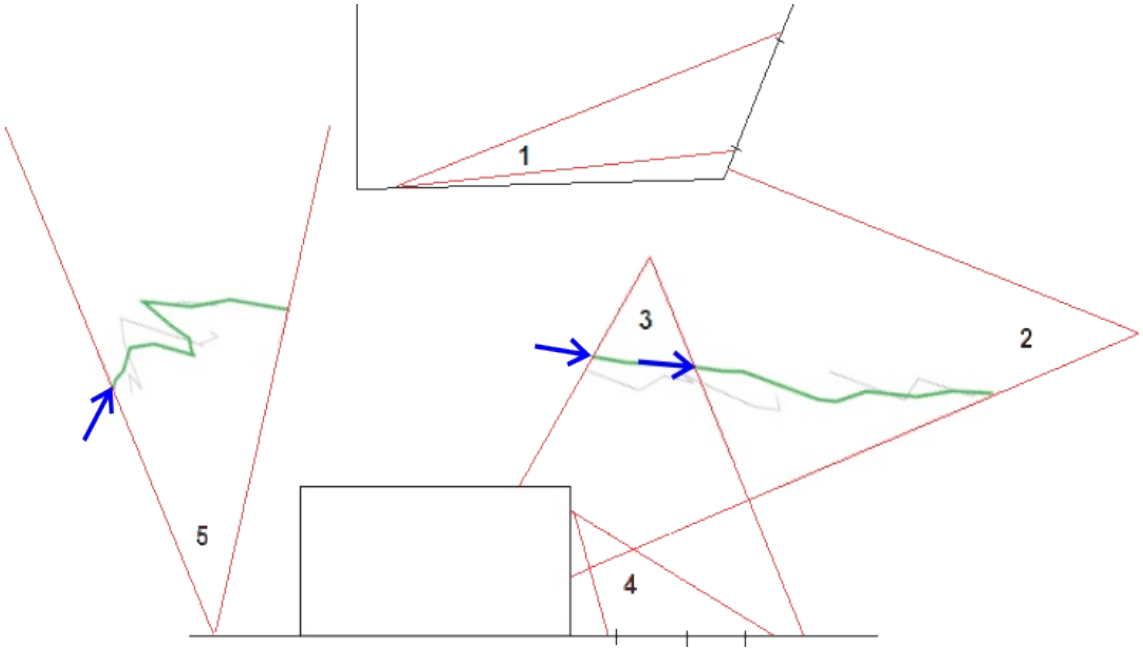
The output should be *clean*, *integrated* and *relevant* metadata...

VRMs are distributed “hearts” of the surveillance network:

- Clean metadata (filter, summarize, remove noise)
- Normalize it in time and space (color, lighting, 3D bias)
- Train (and test) the system (thanks to AVSS, NIST)
- Semi-supervised training (using IR capabilities)
- Persistently store the information (PostgreSQL)
- Integrate identifiers (IDs) of objects in the wide area
  - based on the previous occurrence of an object and its appearance
  - querying neighboring VRMs
- Information Retrieval capability
  - Querying (nontrivial! - spatio-temporal, similar)
  - Analysis (spatio-temporal OLAP functionality, off-line)
  - Data mining (classification, clustering of objects and trajectories)

System – cleaning

Kalman filtering of evolving properties (location, size) ... using classic and inverted direction (Kalman state)



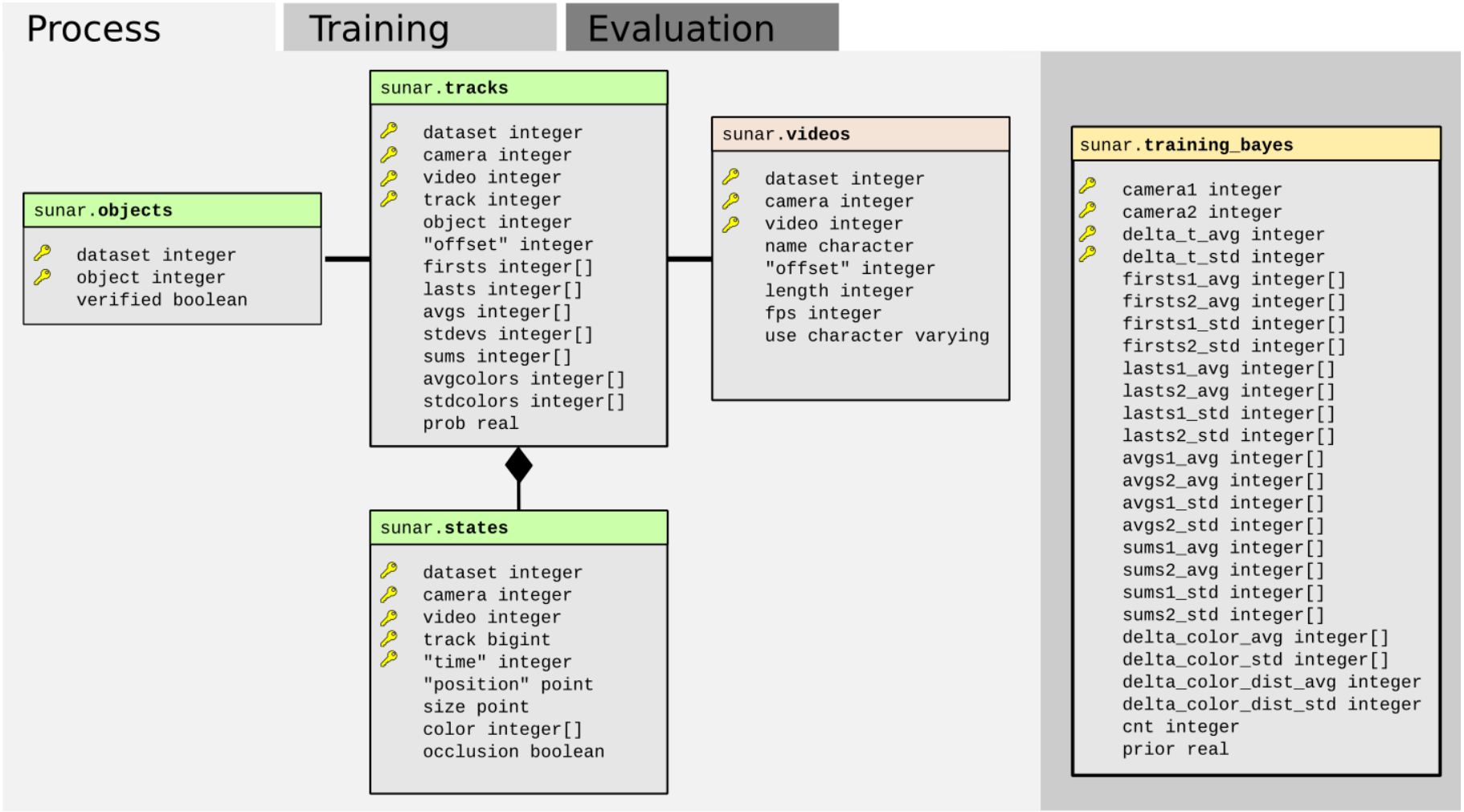
Summarization of (hopefully) constant attributes (color features) (implemented as GMM/EM modeling but not necessarily)

On-line noise removal (short, non-moving trajectories, small objects) according to the previous experiences (eg. t < 1s)

System – persistence

PostgreSQL extended by:

- Similarity search (Eucleidean, Chebyshev, cosine ... distance)
- Temporal extension (to the spatial, extended by overlap **%**)



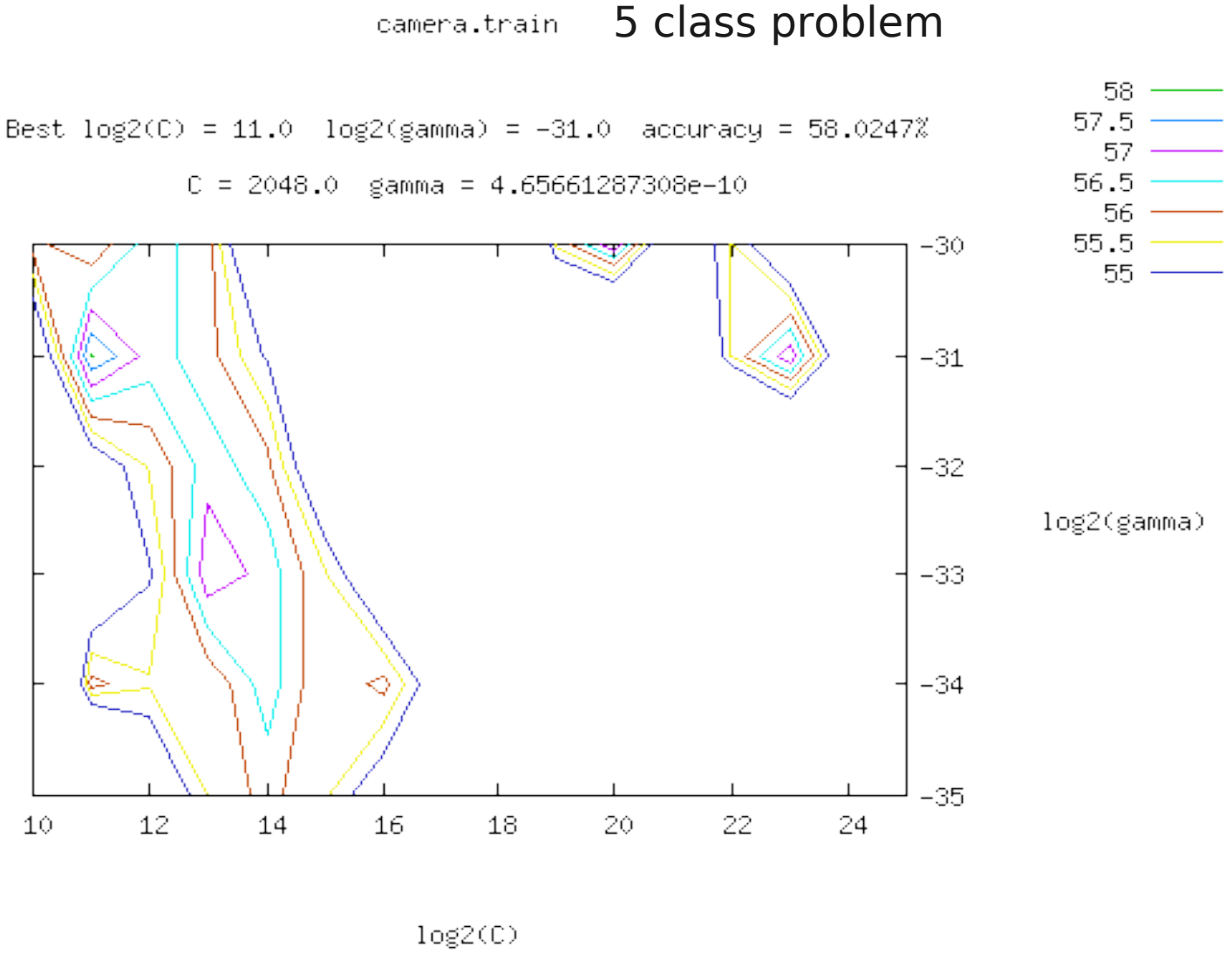
Extracted trajectories are mapped to that NIST gave us
(optimized, but usually there are more extracted for one annotated)

The trajectories are sorted (time_starts ASC) and normalized:

- appearance bias removed (lighting, color bias and 3D parameters),
- the time delta is computed (negative if cams are overlapping :)

from	to	hovers	avg(dt)	min(dt)	max(dt)	avg(dc)	min(dc)	max(dc)
1	2	2	71,0	62	80	95,18	83,44	106,91
2	2	27	101,3	14	358	69,89	33,82	118,38
2	3	44	-86,0	-650	182	73,08	33,06	126,84
3	2	40	-94,1	-568	240	76,4	29,26	129,19
3	3	31	59,4	2	234	62,6	29,02	108,14
3	4	2	2,0	-26	30	51,32	31,78	70,85
3	5	4	533,0	324	680	62	50,39	82,86
4	2	1	96,0	96	96	93,13	93,13	93,13
4	3	3	-107,3	-120	-96	44,89	25,12	77,76
5	3	2	597,0	584	610	77,62	67,99	87,25
5	5	5	29,6	2	60	59,6	49,22	76,48

Other cleaned features are used (Kalman states, summarized attribs)



System – integration

The primary function of the VRM is to *identify objects*
VRM integrate identifiers (k) of objects in the wide area based on:

- the previous occurrence of an object (o)
- and its state (appearance, s)

$$\kappa*(o,s) = \operatorname{argmax}_k P(k|o,s) \approx P(o|k)\,P(s|k)$$

Using the trained model, the task is as follows:

- Find all possible previous trajectories (from all possible cams/VRMs)
- Count the probability of their appearance $P(o|k)$
- Multiply by their normalized appearance Chebyshev distance $P(s|k)$
- Order by the product … submmit to NIST

Evaluation problems:

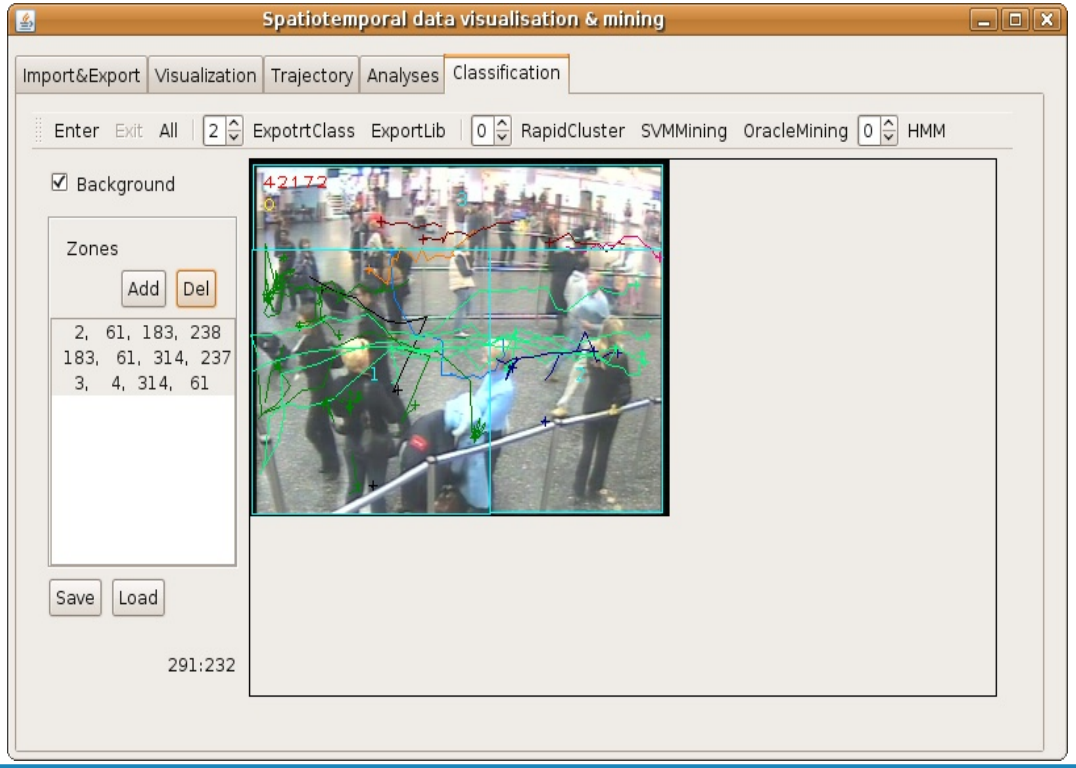
- Starter (1 s) not found by the tracker ?!
  ⇉ extend linearly or by the Kalman filter
- Which tracks to retrieve (precision x recal) ??
  ⇉ recursion: 1_starter_track – 4 – 3 – 2 – 1 – 1 – 1 (7, acc. to dry run)
- Submission validation ⇉ work more precisely :)

System – Human Interface

The human interface is ugly (command line) and video … next slide

Our previous (STVM) project is nicer and it does the IR:

- Trajectory queries and operations (begin, visit, meet, merge, …)
- Spatio-temporal OLAP functionality (trajectory aggregation)
- Data mining (classification, clustering of objects and trajectories)



Evaluation experiments

MCSPT

Cam ID	#GT	MOTA	CorDet	Precision	Recall
1	229	-2,11	157	0,197	0,686
2	4093	-2,13	283	0,031	0,069
3	4238	-0,66	220	0,068	0,052
4	311	0,00	0	NaN	0,000
5	3525	-0,76	110	0,038	0,031
	12396	-1,18	770	0,047	0,062

CPSPT

Cam ID	#GT	MOTA	CorDet	Precision	Recall
1	102	0,18	84	0,560	0,824
2	2547	-1,54	210	0,048	0,082
3	2280	-0,62	246	0,129	0,108
4	134	-0,52	7	0,083	0,052
5	256	-1,44	17	0,042	0,066
	5319	-1,08	564	0,082	0,106

VIDEO

Future work

The SUNAR system works … low-crowded scenes

The SUNAR system works poorly … crowded scenes, wearing black

To be done:

- tracker, tracker, tracker, … working in crowded scenes
- improved annotations assignment to improve the training
- better features, object classification
- 3D calibration
- a nice interface
- a lot of work...

Wanted:

- time
- people
- money
- projects
- cooperation

Thank you

... for your questions

chmelarp@fit.vutbr.cz