UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

ICCE2015, 15-18 March 2015 Kuwait City

Invited speech @International Conference on Computer Science and Engineering: Big data science for the social goods

Human Behavior understanding in video

Prof. Ing. Rita Cucchiara
Imagelab, Dipartimento di Ingegneria «Enzo Ferrari», Modena, Italy
Director of the Research Center in ICT Softech-ICT

http://imagelab.ing.unimore.it

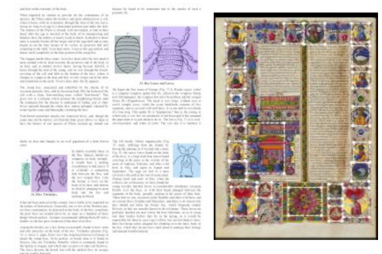
ING

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Imagelab : research

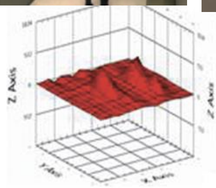
www.imagelab.unimore.it

- Pattern recognition and Image processing
- Medical Imaging (dermatology eu projects)
- Digitalized Document analysis (Encycl.Treccani)

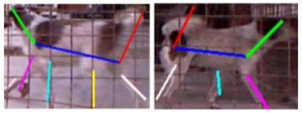


- Multimedia
- Multimedia big data annotation (RAI)

- 2D, 3D, wearable Computer vision
- <u>Augmented experiences in culture </u>and museums
- Experience with Wearable devices and IoT
- 2D and 3D augmented visits



- Computer vision for Behaviour analysis
- Children behaviour analysis
- <u>Surveillance in crowd</u>
- Animal behaviour



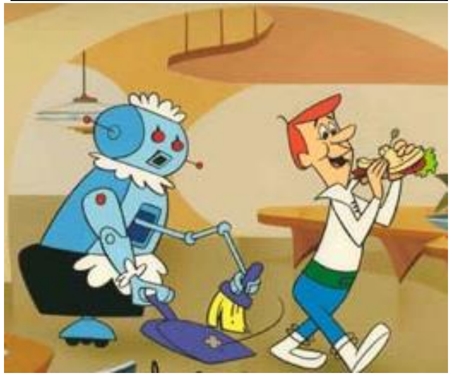
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Computer Vision and Human behaviour understanding

..with Computer Vision

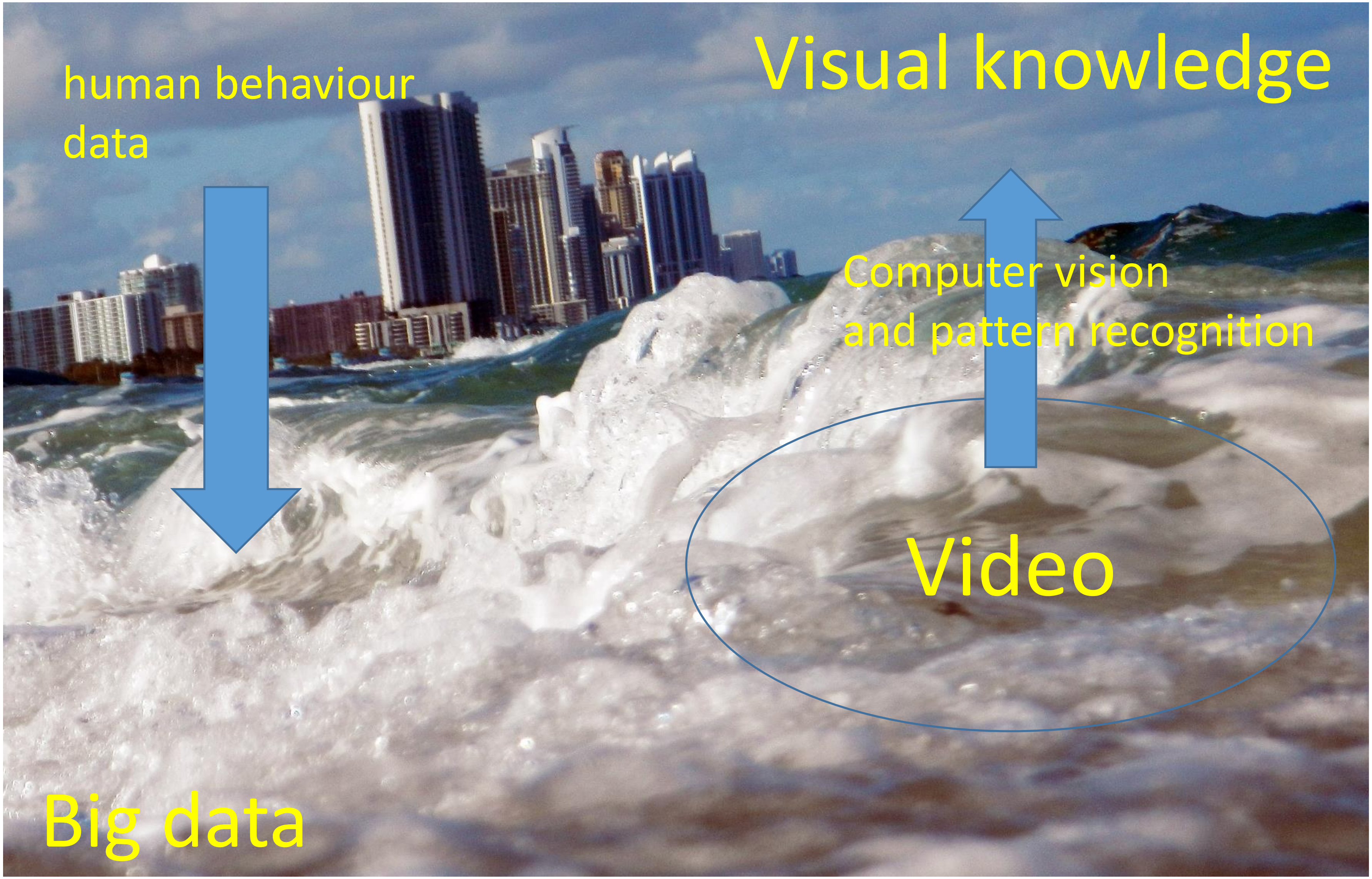
- **Computer Vision** is the scientific discipline studying **how to perceive and understand the world through visual data by computers.**

Can computer vision provide effectively Human behaviour understanding?



Can computer vision is useful Big data?

The visual data cycle



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Big data , the four Vs, and Human Behaviour Understanding

- Volume: video data are expected the 65% of volume of big data in 2015*; and human are the most important targets for human themselves*

- Velocity: Video streaming is an issue in many contexts, such as video surveillance: video understanding is necessary to filter only useful knowledge

- Variety: Video is only one powerful media where understand human behaviour that must be integrated with IoT data, textual and audio and social data…

- Veridicity: the main issue: extracting Human behaviour knowledge is complex and the data are noisy, erroneous and to be post-processed for assessing their reliability.

- T. Huang, "Surveillance Video: The Biggest Big Data," Computing Now, vol. 7, no. 2, Feb. 2014 - See more at: http://www.computer.org/web/computingnow/archive/february2014#sthash.b4UxnARn.dpuf

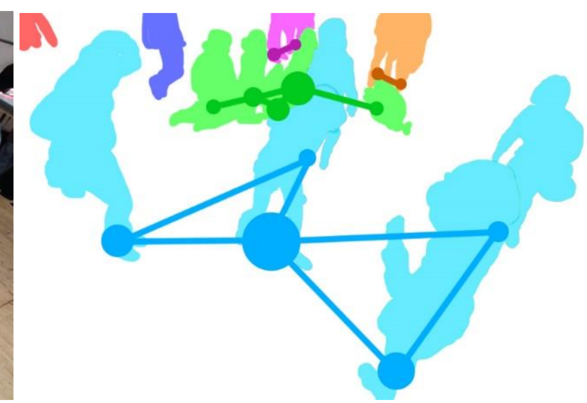
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Human behaviours

- What he is doing?
- What are they doing?

- Single and collective behaviours,
- Collaborative or not collaborative behaviours

Italian project
Surveillance in
cultural cities



Italian project “the educating city” cluster smart city 2014-2016

Understanding Human behaviour by video

- Why:

- To support sociologists' and psychologists' work

- To provide support for a huge number of applications, services and systems
- For on-line and off-line knowledge extraction by visual data

A long story…..

- 1997-2000 MIT Alex Pentland: PFINDER projects and understanding interactions
- 2006- datasets for action analysis (Weizmann ICCV2005)
- 5 workshops on HBU (from 2010: IAPR, AMI, IROS, ACM MM, ECCV)
  http://www.cmpe.boun.edu.tr/hbu/2014
- Chalearn workshops 2011- 2015; CVPR 2015 challenge "Looking at People"
- Now many datasets

978 R. Poppe / Image and Vision Computing 28 (2010) 976–990

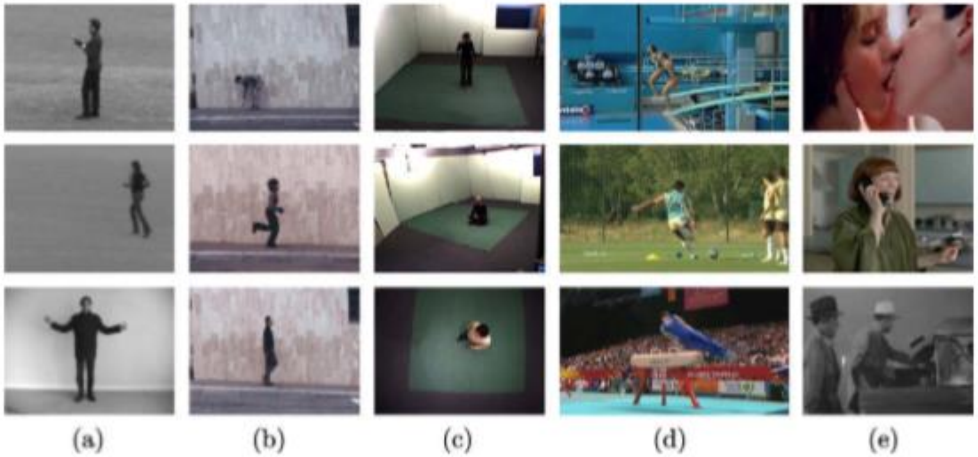
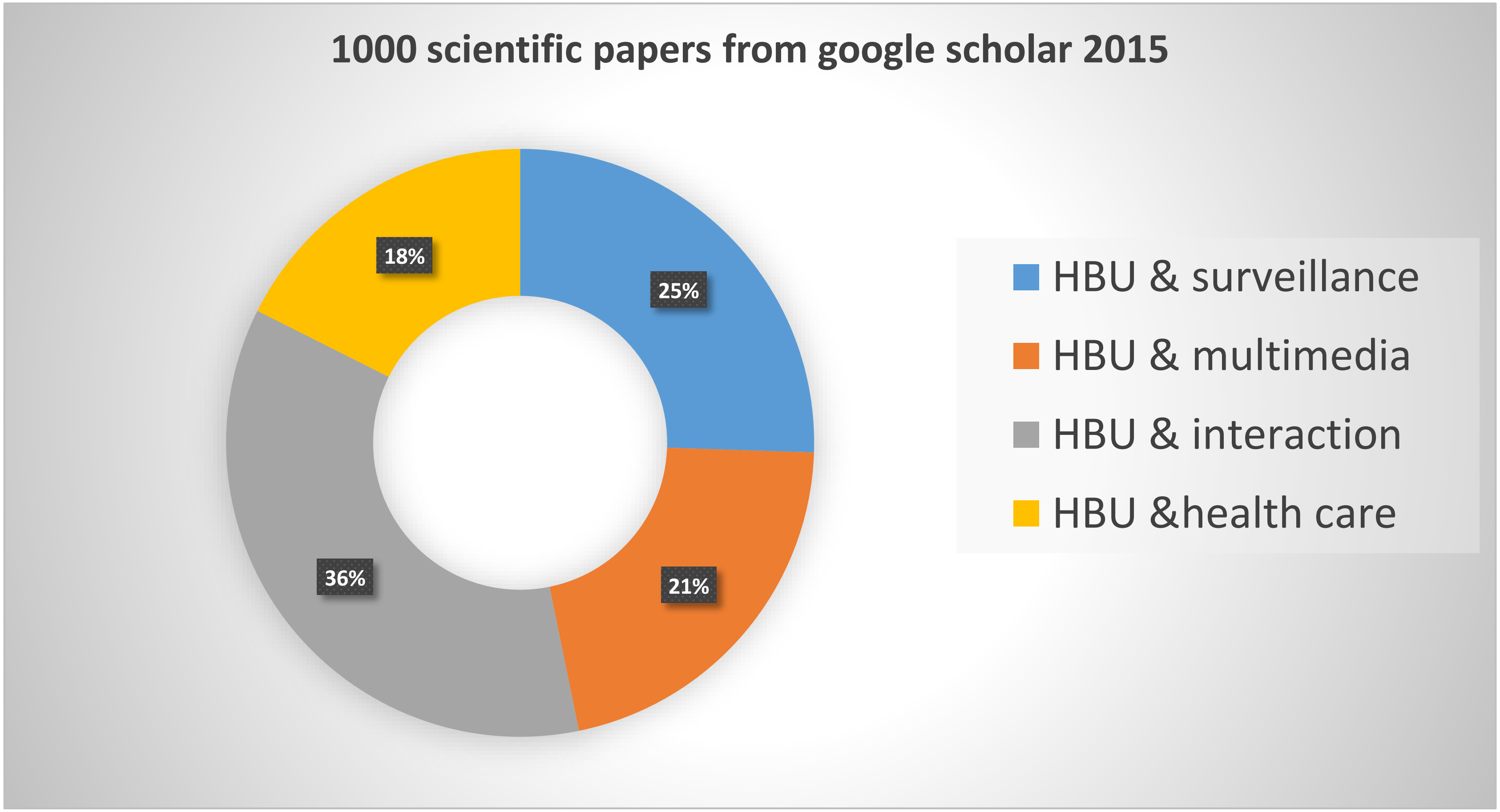


Fig. 1. Example frames of (a) KTH dataset, (b) Weizmann dataset, (c) Inria XMAS dataset, (d) UCF sports action dataset and (e) Hollywood human action dataset.

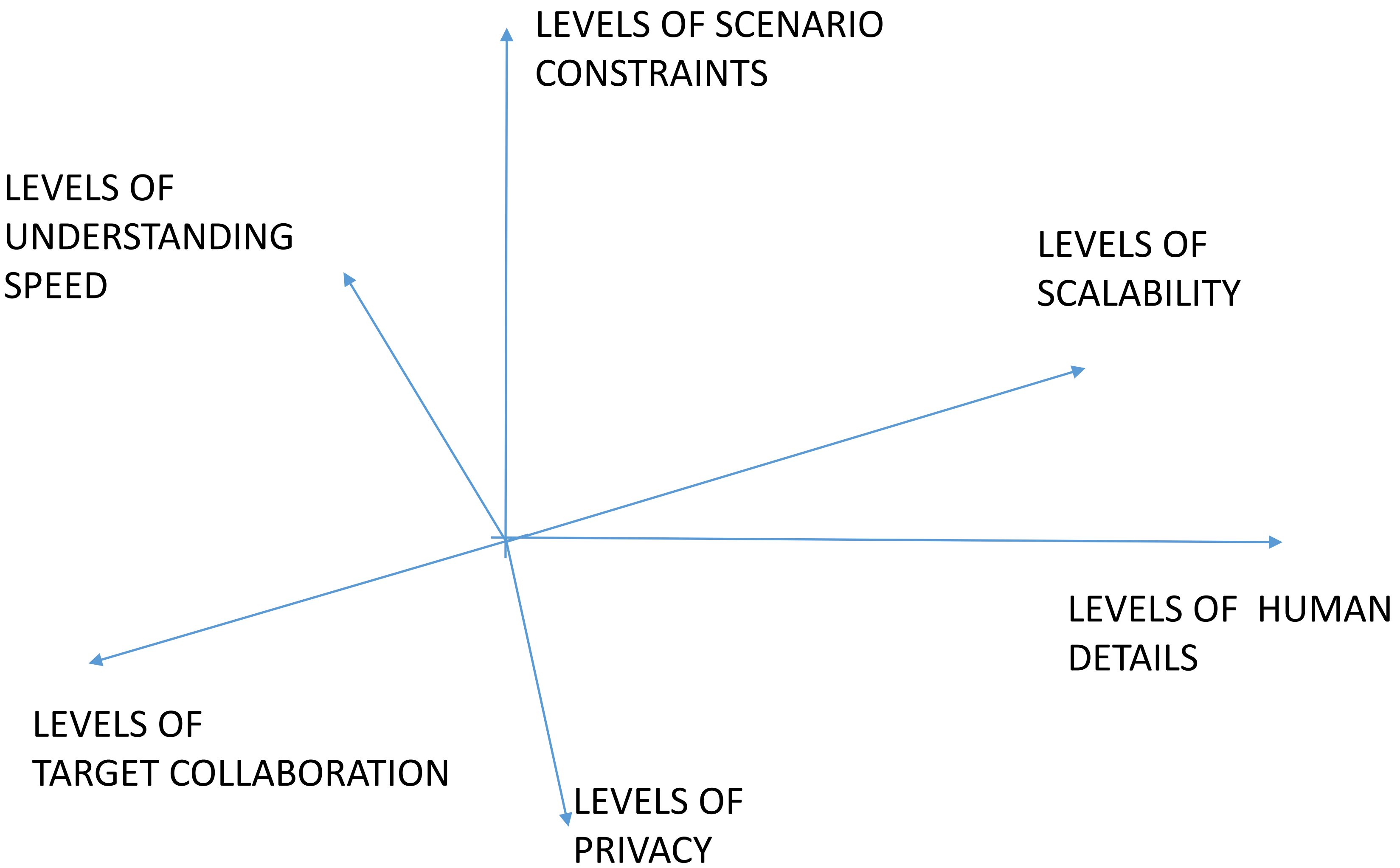
- From R. Poppe "A survey on vision based action recognition" Image and vision computing 2010

Research in HBU

- HBU by vision:
- More than 1000 papers from 2010 to 2015



Different dimensions..



A DIMENSION: the levels of details

- Levels of details

Human
parts

Human
Full body

Human(s)
In the environemnt

Humans
In crowd

Lev 1: Expressions and gestures

Understanding humans for New natural Human Computer interaction systems *

What are they doing?

Deaf Sign language

Picture from Benjamin Lewis UCLA

Maja Pantic, Alex Pentland, Anton Nijholt and Thomas Huang uman Computing and Machine Understanding of Human Behavior: A Survey ICMI 2006

Lev2: body actions

Fig. 7. Sample input frame of the Weizmann dataset

Roberto Vezzani, Davide Baltieri, and Rita Cucchiara, HMM Based Action Recognition with Projection Histogram Features ICPRW2010 supported by EU THIS Project

Lev. 3 people in the environment

- What are they doing?

Real-time surveillance

(big) Data analysis
for digital forensics

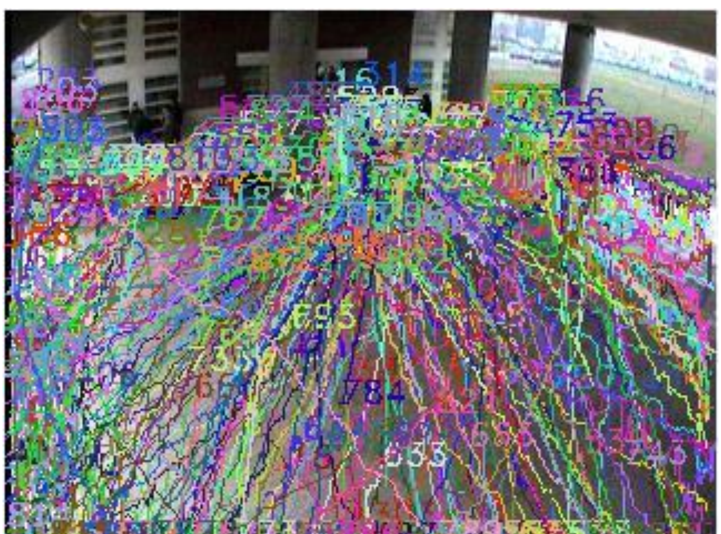
ENVI-VISION

EGO-VISION

Lev 4: social activities and behavior

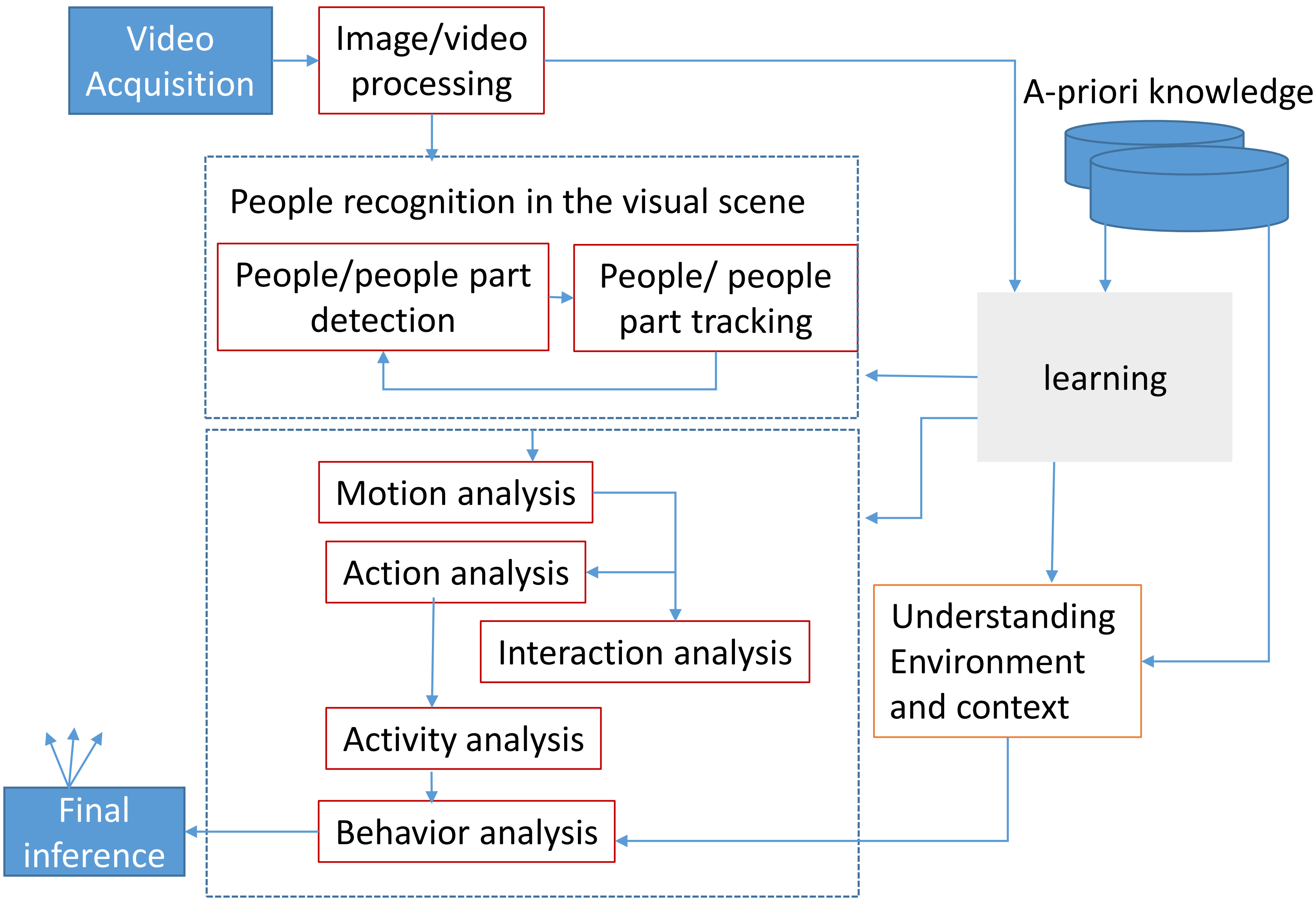
- Detecting, tracking and understanding groups in crowd

What are they doing?

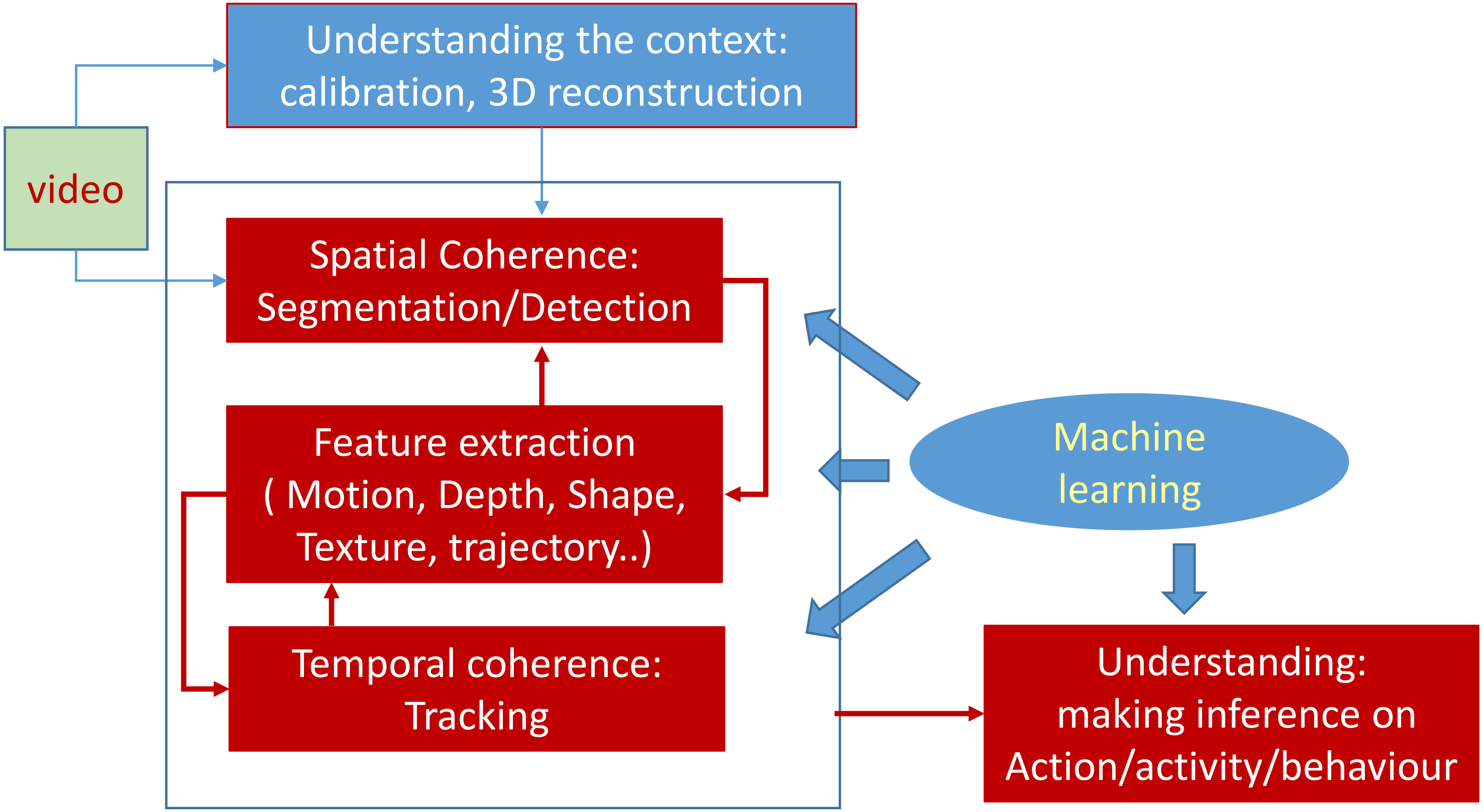


Alex Prager Crowd #8 (city Hall) 2013

Common framework



The big challanges



A big role of machine learning

2015 30.000$ in prizes from Microsoft in Looking at People competition
Special issue on PAMI 2015

A second dimension: a constrained scenario

- The (1-..) constraints in the scenario

Health ,
Sport training

Human
Computer
Interaction

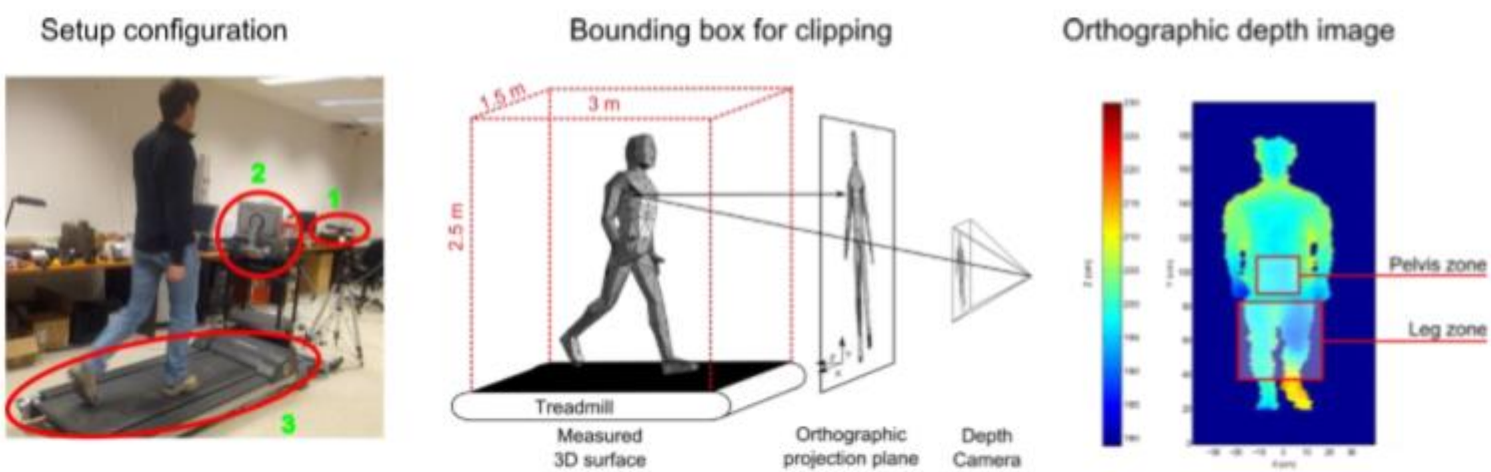
Surveillance

Humans
In crowd

Multimedia

1 constrained applications: eg health

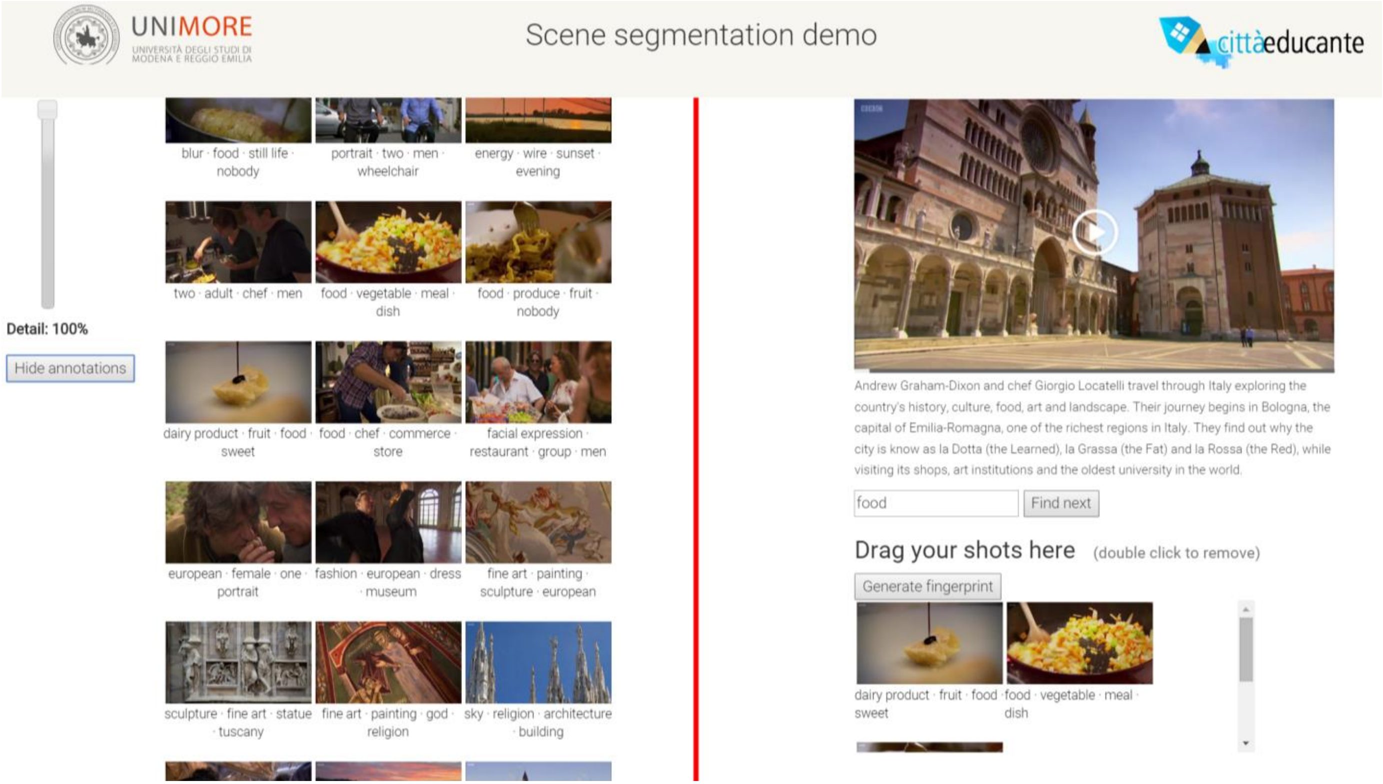
- Acquisition in a very constrained environment
- From marker-based to markerless 3D clouds
- Specific setup, Reconstruction and measures
- learning as a support of final dyagnosis



Edouard Auvinet, Franck Multon and Jean Meunier; New Lower-Limb Gait Asymmetry Indices Based on a Depth Camera, Sensor 2015

2 multimedia video annotation

- Unconstrained scenario
- Detecting people behviour in (educational) video
- From Italian national broadcast RAI
- Deep learning for concept detection



2.HCI and collaborative applications

- **Human computer interaction**
  - ***(general assumptions… often too general)***

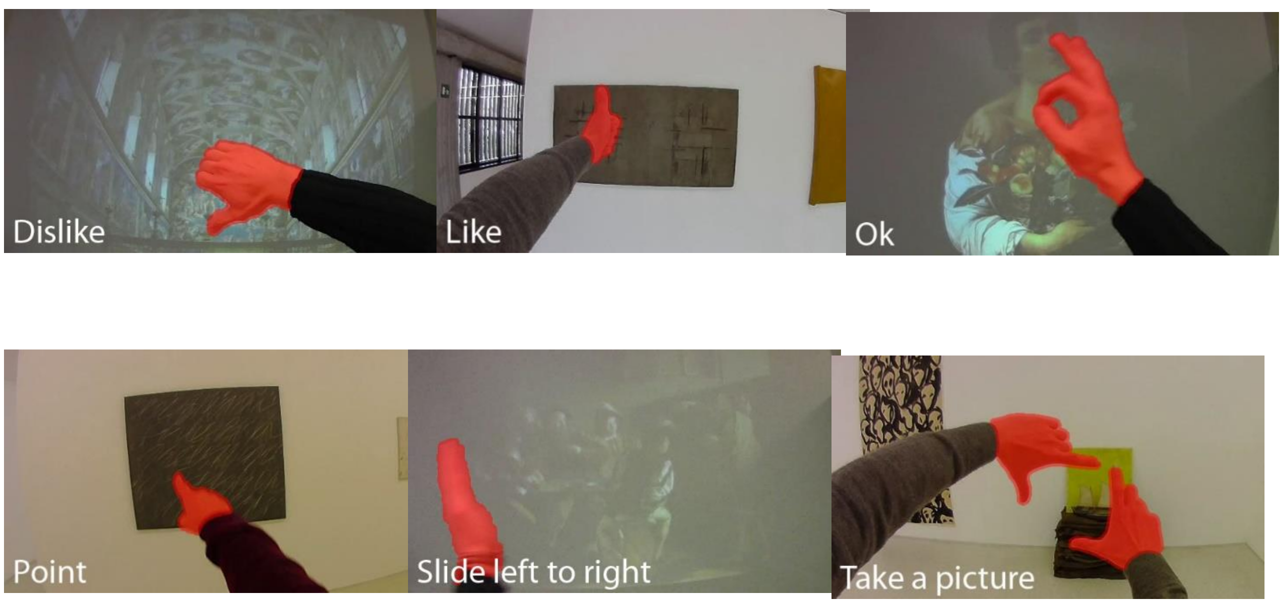
- Less constrained environment but with strong assumptions (e.g. the person knows what has to be done)
- Real-time processing (often with embedded solutions) but with an acceptance temporal window
- Collaborative environment: goal high precision; if recall is not enough, human-in-the-loop can handle it
- General purpose but a-priori defined features (eg hand colors, point trajectories..)
- Learning by very few examples (often by a single person only) but the Learning space is well defined (eg with few ambiguities)

→ all Kinect applications in HCI!

the gesture recognition market is estimated to grow at a CAGR of 29.2 % year from 2013-2018 for Gaming and Entertainment, Healthcare, Automotive applications, Educations and serious gaming (By USA Markets and Markets 2014)

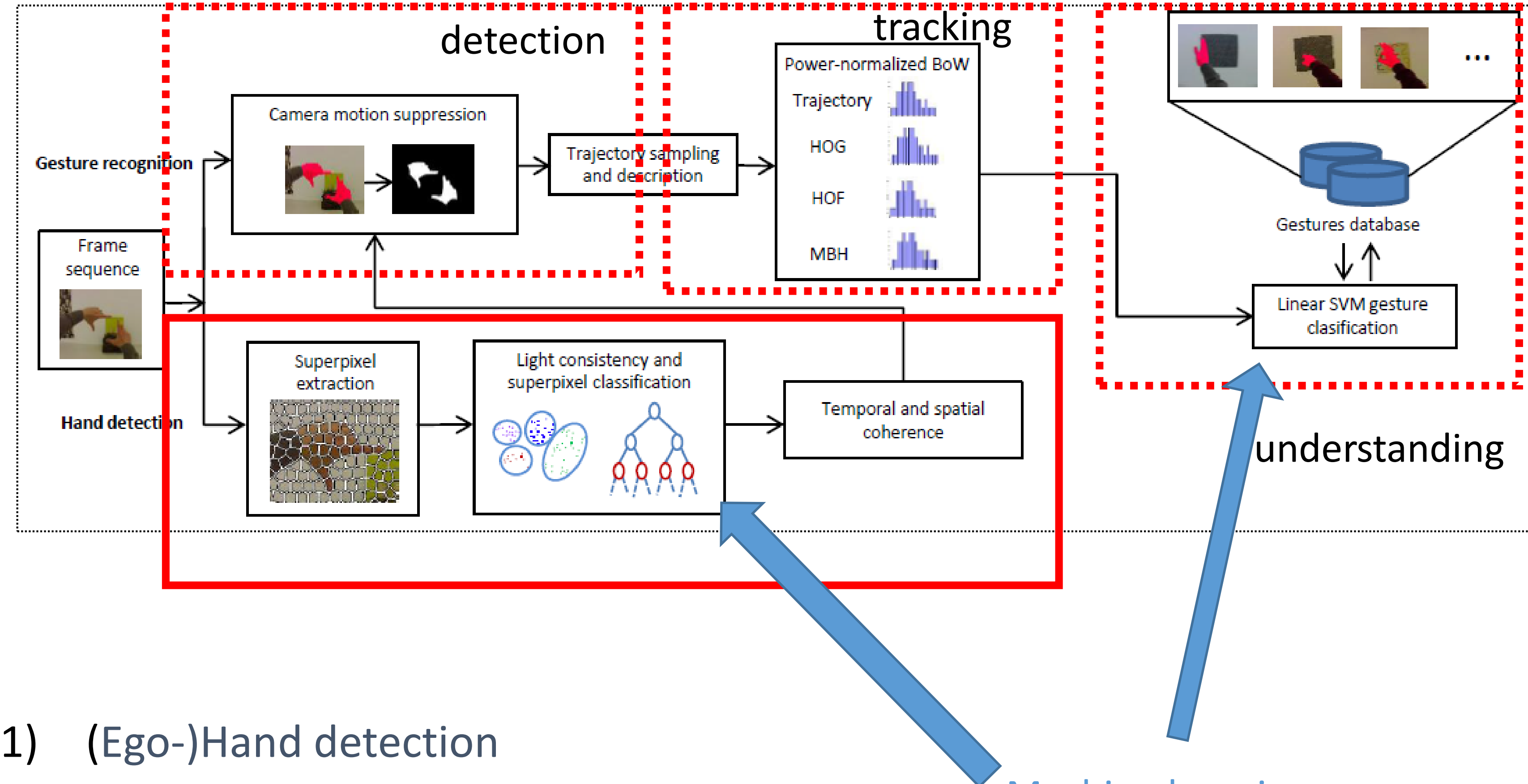
An example:

- Understanding human behaviour in interacting with artistic objects for augmented experiences
- HBI: Gesture analysis, associated with context



L. Baraldi, F. Paci, G. Serra, L. Benini, R. Cucchiara «Gesture Recognition using Wearable Vision Sensors to Enhance Visitors' Museum Experiences" IEEE Sensors 2015

Ego-Gesture recognition



Machine learning

1) (Ego-)Hand detection
2) (Ego-)Camera motion suppression
3) Feature extraction
4) Classification and gesture recognition

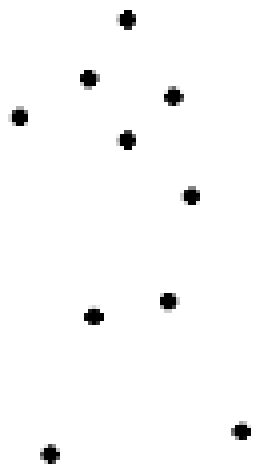
3. Surveillance

People Surveillance

one of the biggest topic in computer vision

- On-line, fast, <u>not collaborative at all</u>.
- Partially constrained (eg. calibration)

Humans recognize motion and recognize by motion



- Segmentation/ detection
- Simple features (motion, appearance)
- Tracking: the big challenge

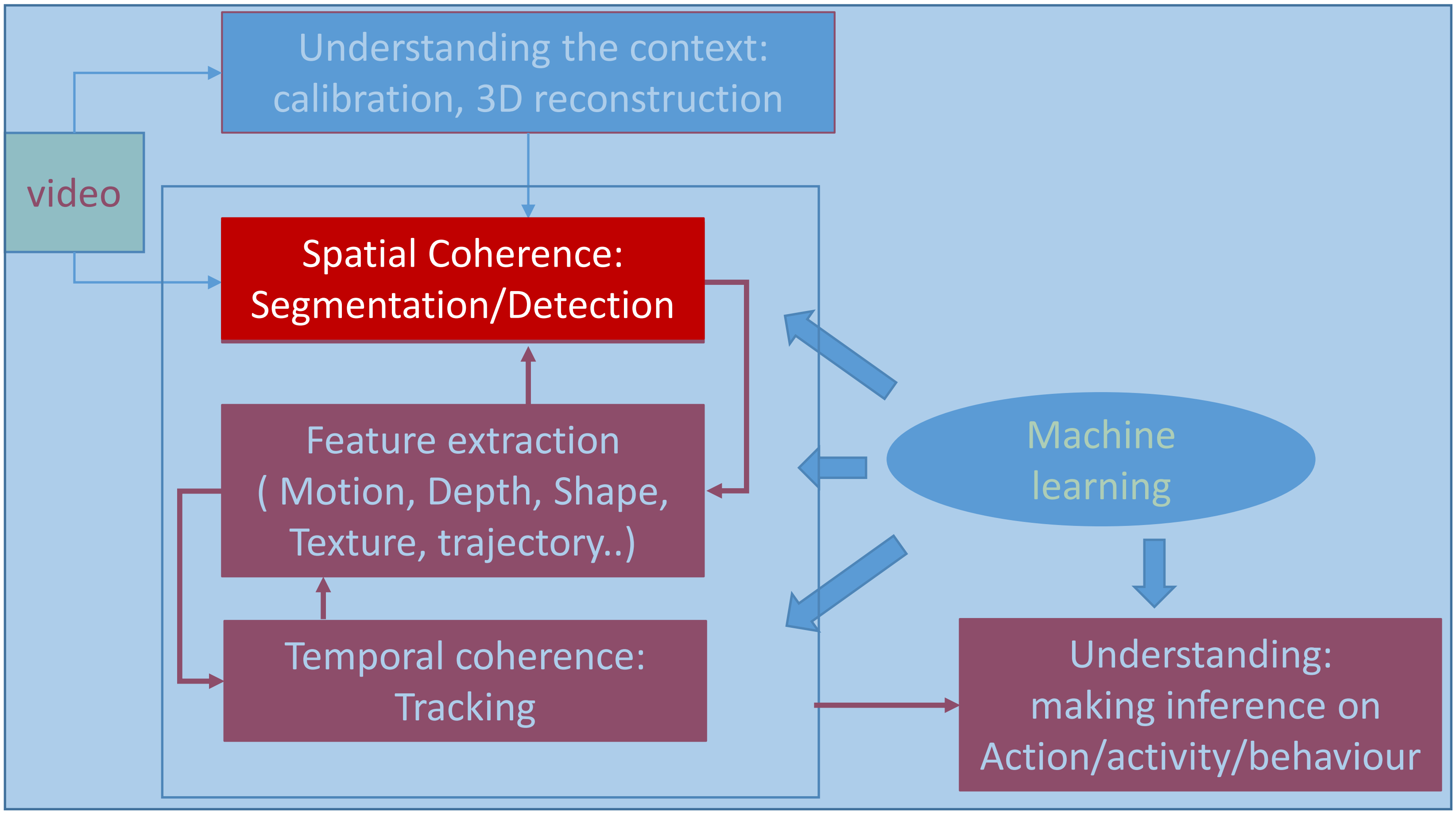
- Learning and high reasoning: with many examples

HBU in Surveillance

For single people and crowd:

Why HBU for surveillance?

- Monitoring dangerous/forbidden zone
- Access control
- Single people iterative activity recognition
- Extracting common behaviour in crowd
- Understanding anomalous behaviours (eg unfrequent trajectories)
- Recognizing specific behaviour (e.g. suspicious behaviour, for terrorism, social engagement for children..)



Understanding the context:
calibration, 3D reconstruction
video
Spatial Coherence:
Segmentation/Detection
Feature extraction
(Motion, Depth, Shape,
Texture, trajectory..)
Temporal coherence:
Tracking
Machine
learning
Understanding:
making inference on
Action/activity/behaviour

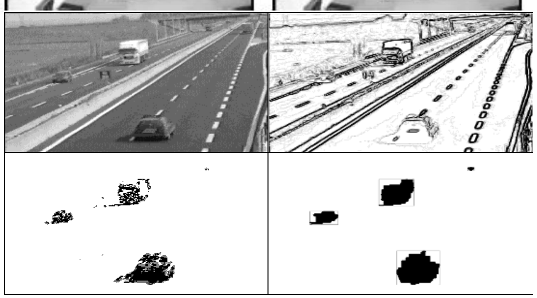
Segmentation/detection in surveillance

?

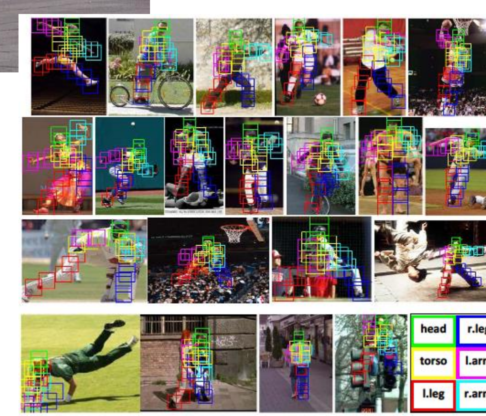
Segmentation by motion

Detection with learning and classifiers

Differential Motion



HOG, part based.. People detectors



Background suppression

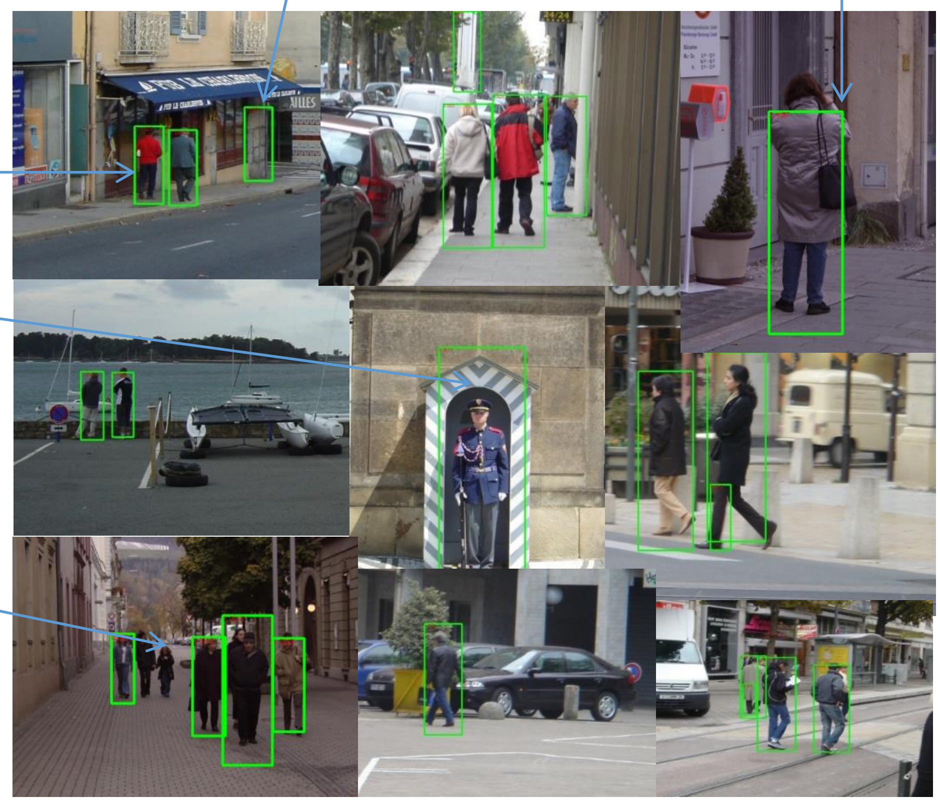
R. Cucchiara, C. Grana, M. Piccardi, A. Prati "Detecting Moving Objects, Ghosts and Shadows in Video Streams", IEEE *Trans on PAMI,* 2003

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

People detection

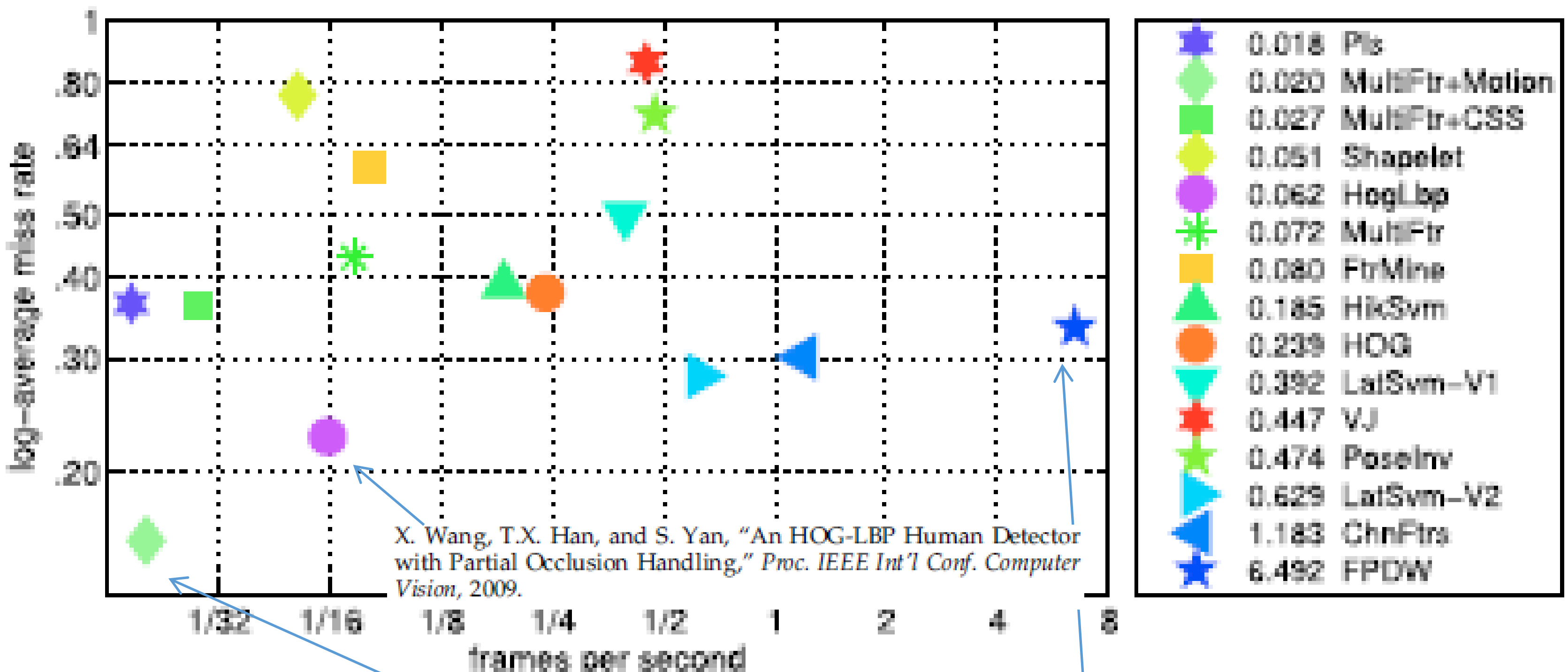
- ☺ very general for pedestrian
- Without any constraints: from mobile, moving cameras, wearable…)

- ☹ someErrors….



Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona, “**Pedestrian Detection: An Evaluation of the State of the Art**” IEEE Trans. PAMI 2012

Speed and accuracy



X. Wang, T.X. Han, and S. Yan, "An HOG-LBP Human Detector with Partial Occlusion Handling," *Proc. IEEE Int'l Conf. Computer Vision*, 2009.

(a) accuracy versus runtime for pedestrians over 100 pixels

P. Dollár, S. Belongie, and P. Perona, "The Fastest Pedestrian Detector in the West," *Proc. British Machine Vision Conf.*, 2010.

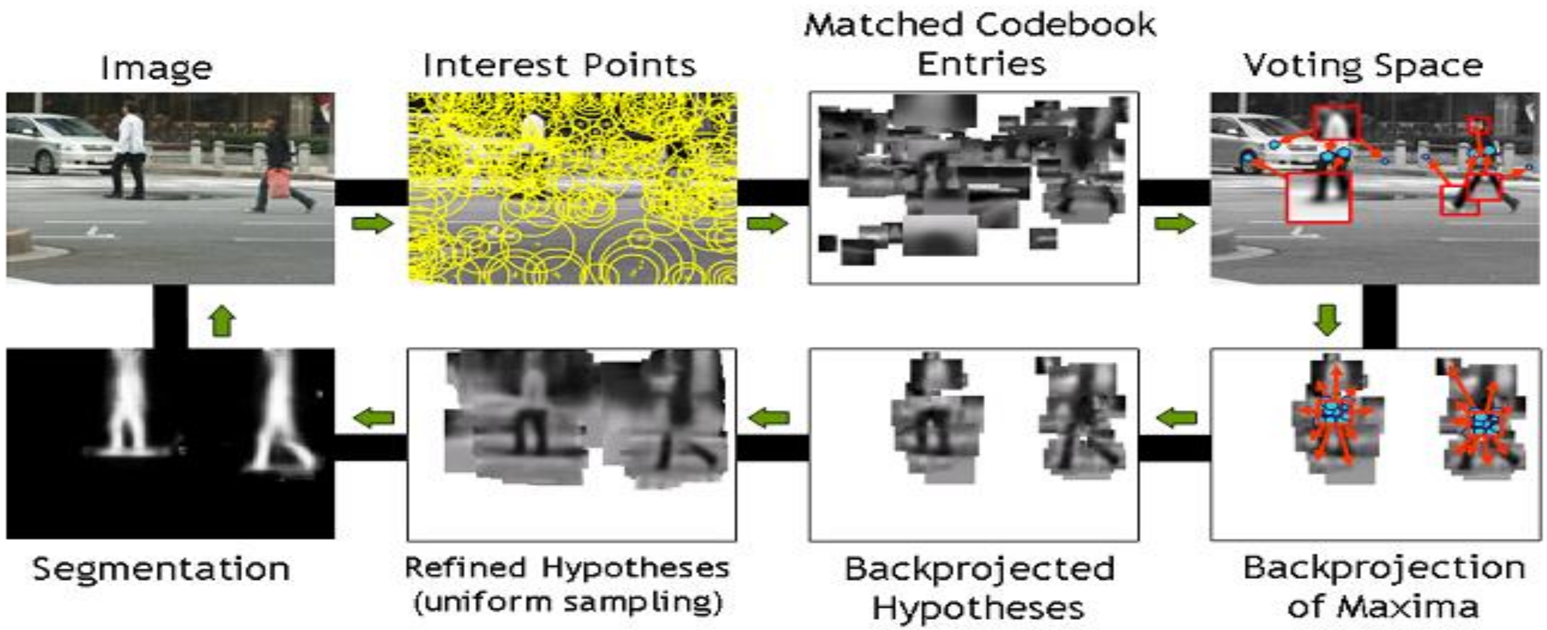
(a) Caltech [3] (b) Caltech-Japan [3]

S. Walk, N. Majer, K. Schindler, and B. Schiele, "New Features an Insights for Pedestrian Detection," *Proc. IEEE Conf. Compute Vision and Pattern Recognition*, 2010.

Detecting people &target

1) The initial attempts: Generate hypotheses (local detector)

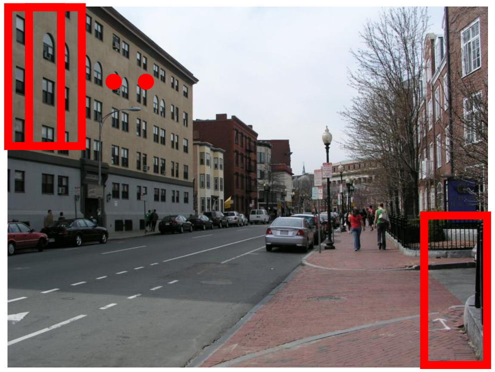
- SIFT, HOG etc



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

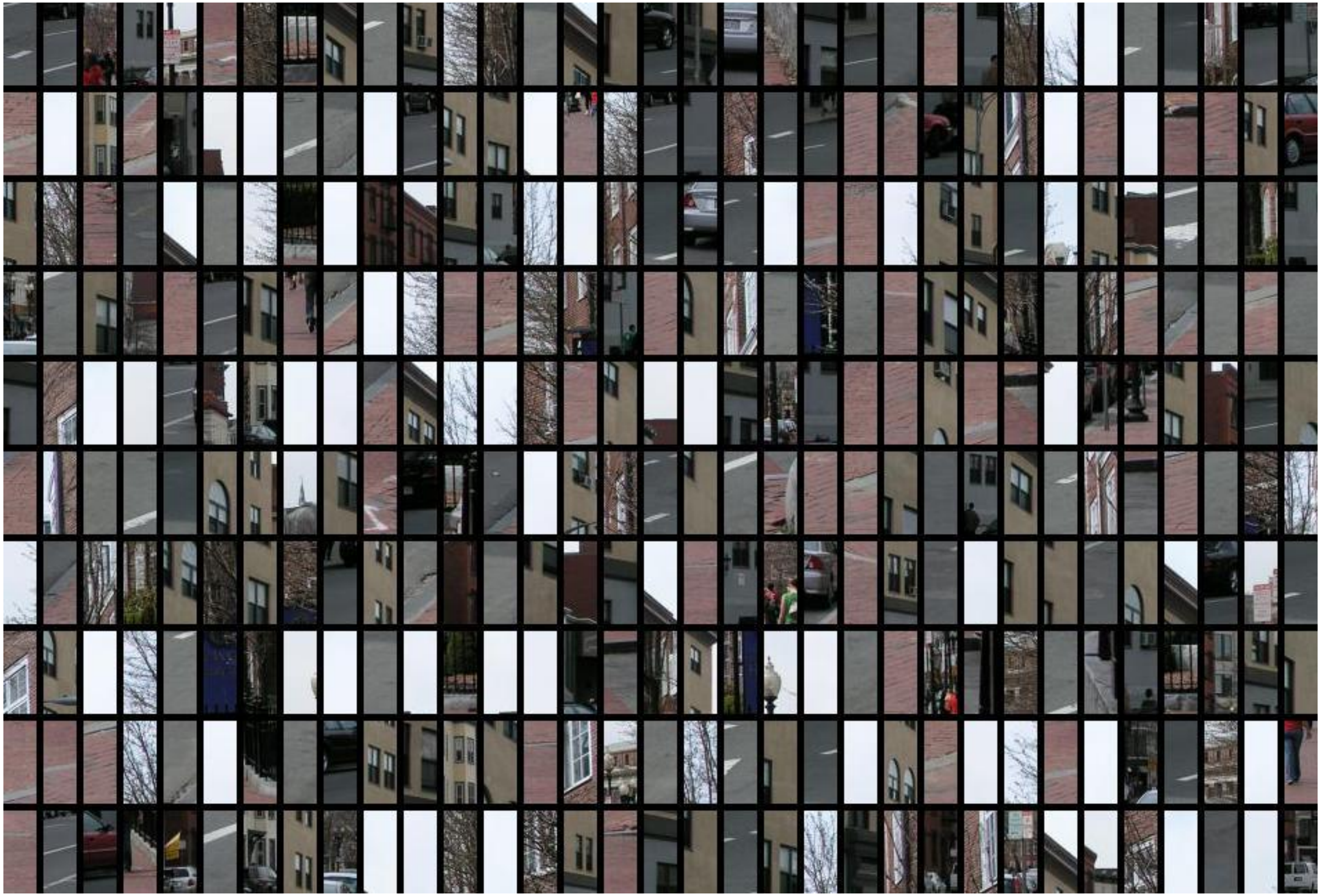
Detecting people &target

- 2) Most of the approaches: better to see everywhere: **Sliding windows**



Thanks to Derek Hoiem Illinois univ.

What the detector sees

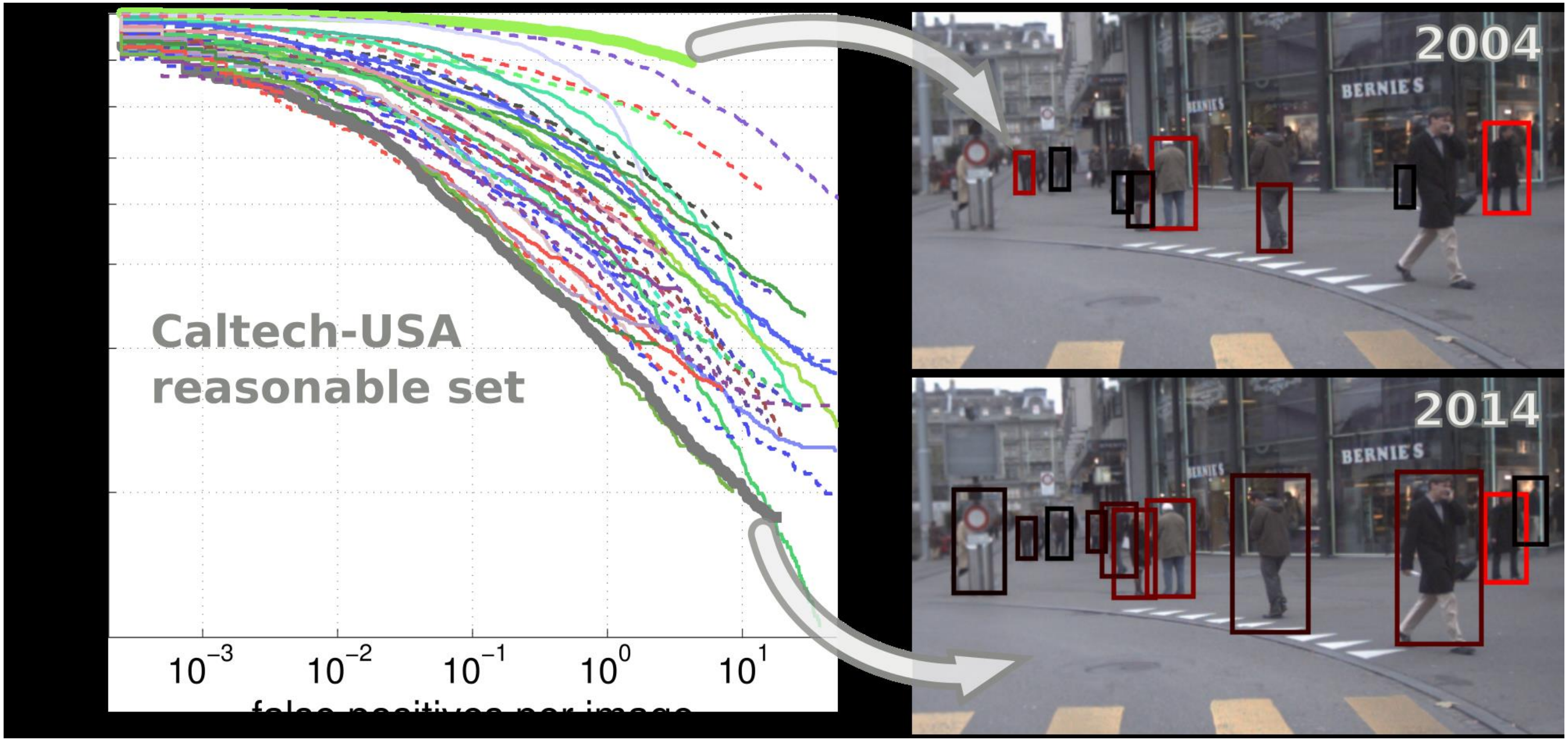


- And repeated at each possible scale... and than, learning

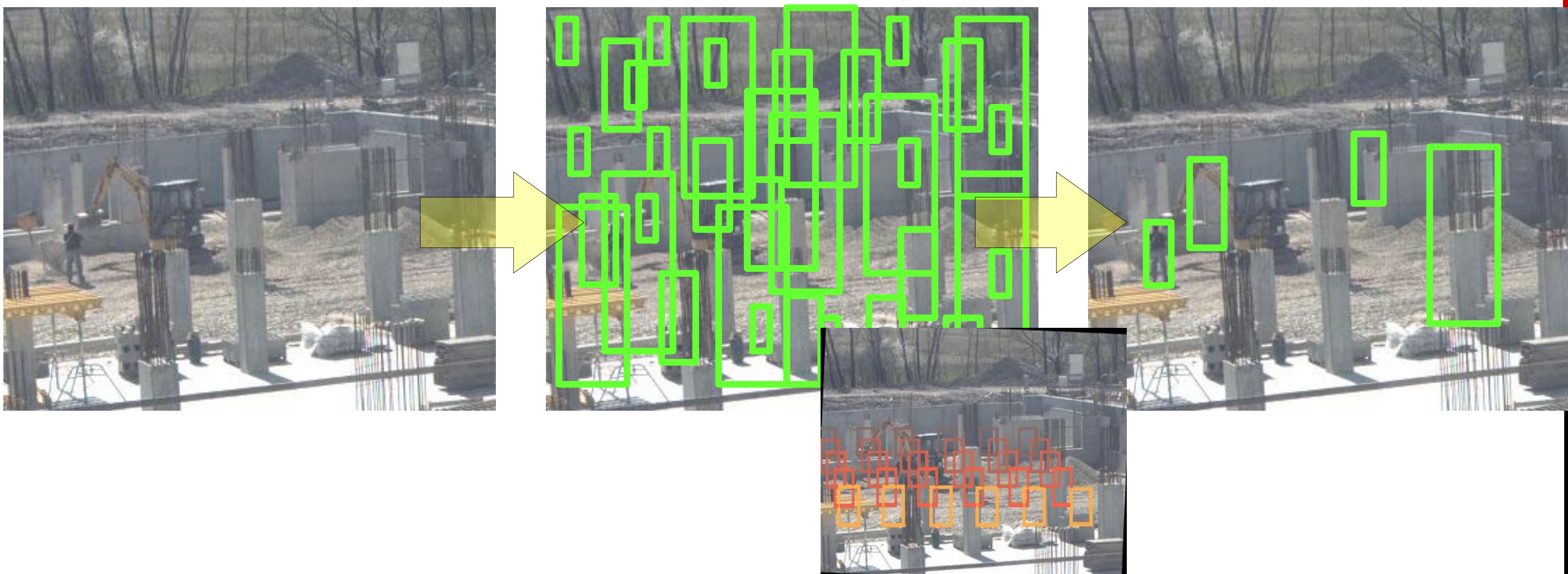
Thanks to Derek Hoiem Illinois univ.

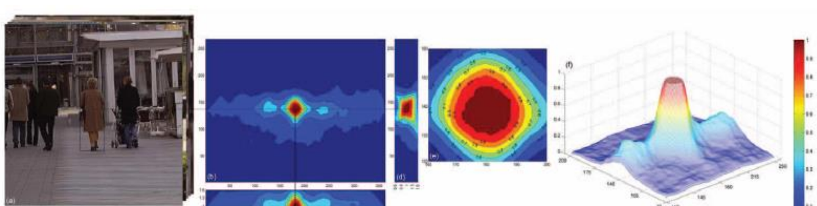
A conclusive study

- **Benenson.. Schiele. Ten years of pedestrian detection, what have we learned ? ECCV2014**

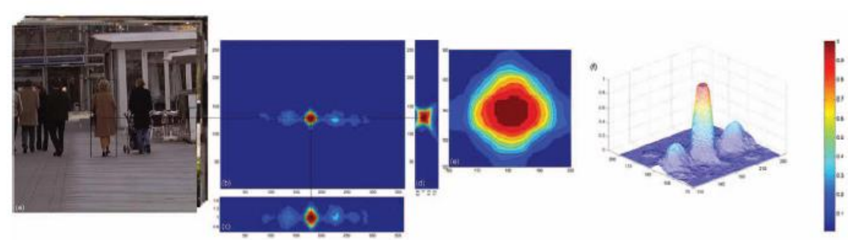


Speed and accuracy

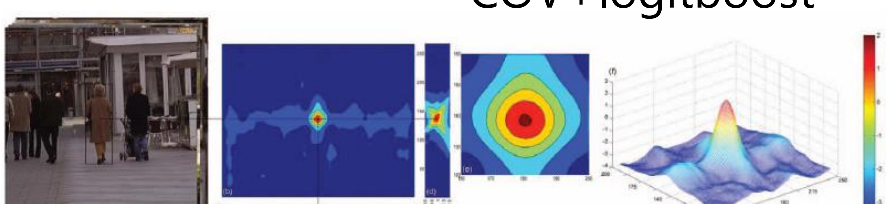




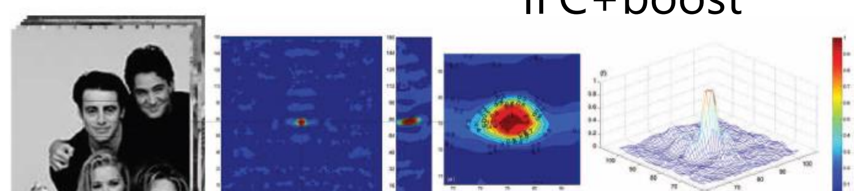
COV+logitboost



IFC+boost



HOG+SVM



VJ+boost

Particle- windows for speeding-up

A probabilistic bayesian paradigm for object detection:
"estimate obj. detection as a pdf"

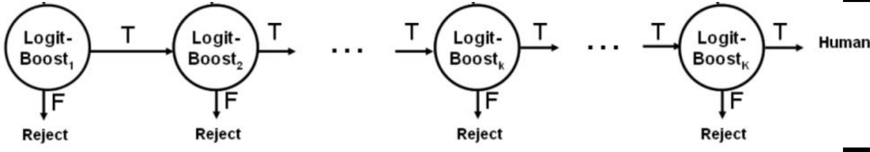
Set $q_0(\boldsymbol{X}) = U(\boldsymbol{X})$
for i=1..m **do**
Draw N_i samples from $q_{i-1}(\boldsymbol{X})$
Assign a Gaussian kernel to each sample
Compute the measurement on each sample $s_i^{(j)}$
Obtain the measurement density function at step

$$p_i(\boldsymbol{Z}|\boldsymbol{X}) = \sum \pi_i^{(j)} \cdot \mathcal{N}\left(s_i^{(j)}, \Sigma_i^{(j)}\right)$$

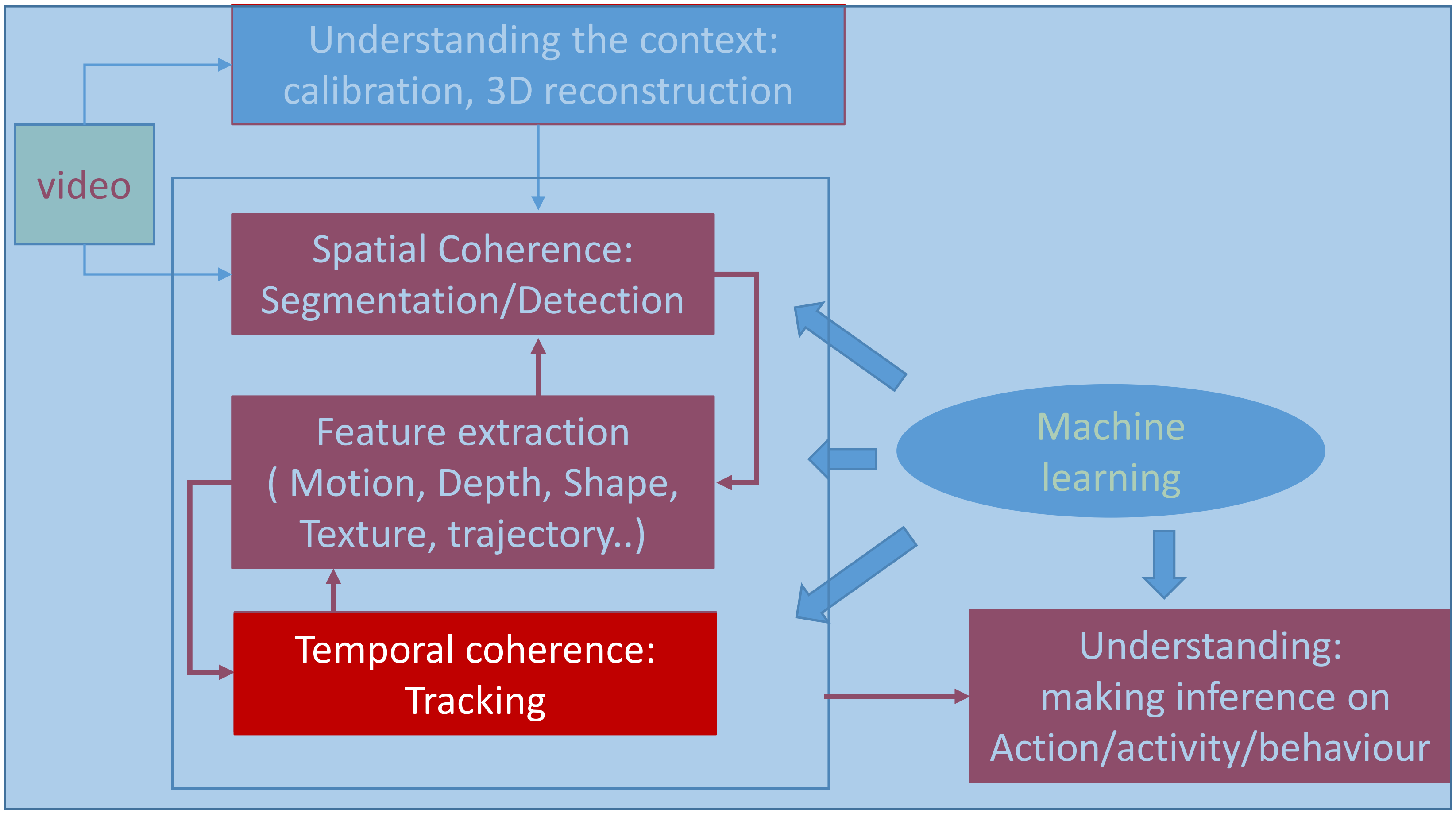
Compute the new proposal distribution:

$$q_i(\boldsymbol{X}) = (1-\alpha_i)\, q_{i-1}(\boldsymbol{X}) + \alpha_i \frac{p_i(\boldsymbol{Z}|\boldsymbol{X})}{\int p_i(\boldsymbol{Z}|\boldsymbol{X}) d\boldsymbol{X}}$$

end for



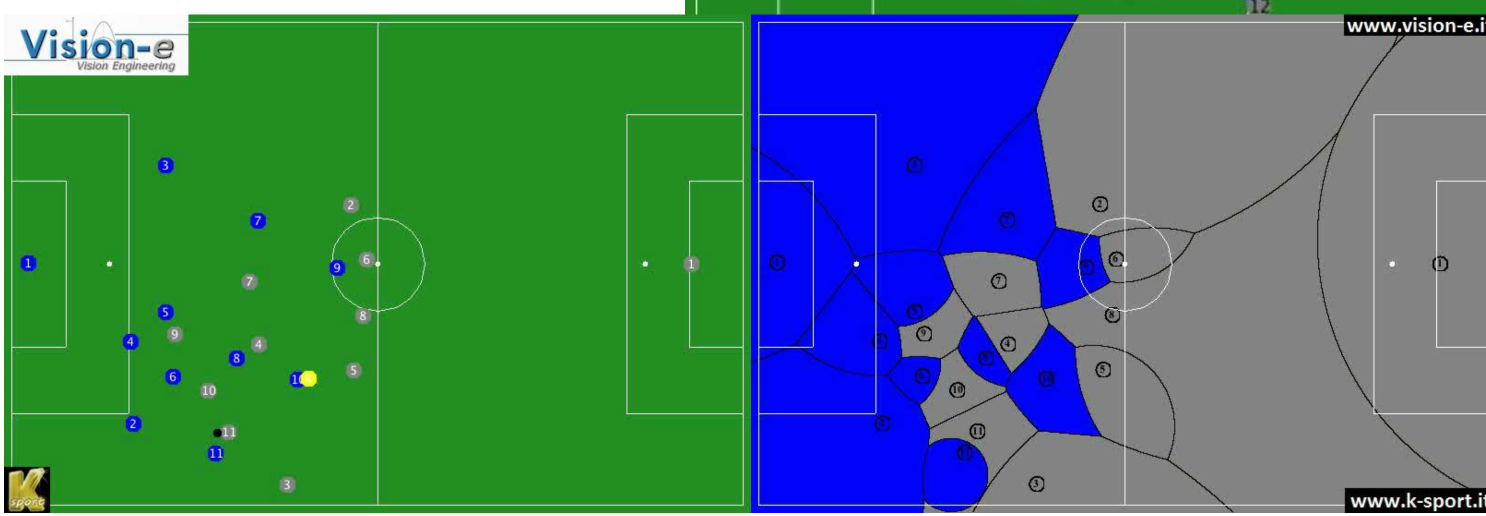
G.Gualdi, A.Prati, R.Cucchiara Multi-Stage Particle Windows for Fast and Accurate Object Detection
IEEE Transactions on PAMI Aug. 2012


Understanding the context:
calibration, 3D reconstruction
video
Spatial Coherence:
Segmentation/Detection
Feature extraction
(Motion, Depth, Shape,
Texture, trajectory..)
Temporal coherence:
Tracking
Machine
learning
Understanding:
making inference on
Action/activity/behaviour

Detection and tracking

Tracking by detection: using people detection for initialize ROI-based tracking (eg particle filter)

In semi-constrained world
Tarcking is possible



tracking

- Tracking is the hardest problem.

- Finding the visual invariance among frames
- Appearance?
- Motion?
- Space continuity?

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

..and tracking single (people) target

- **Is tracking a solved problem?**
- We tried to answer this questions in an "**experimental evaluation**"
- Even in case of single target tracking*
- - a very large dataset
- of 14 categories of challenges

315 video ALOV++
http://www.alov300.org
http://imagelab.ing.unimo.it/dsm

- - a large set of performance measures

MOTA; OTA; Deviaton….
F-Measure
SURVIVAL CURVES..

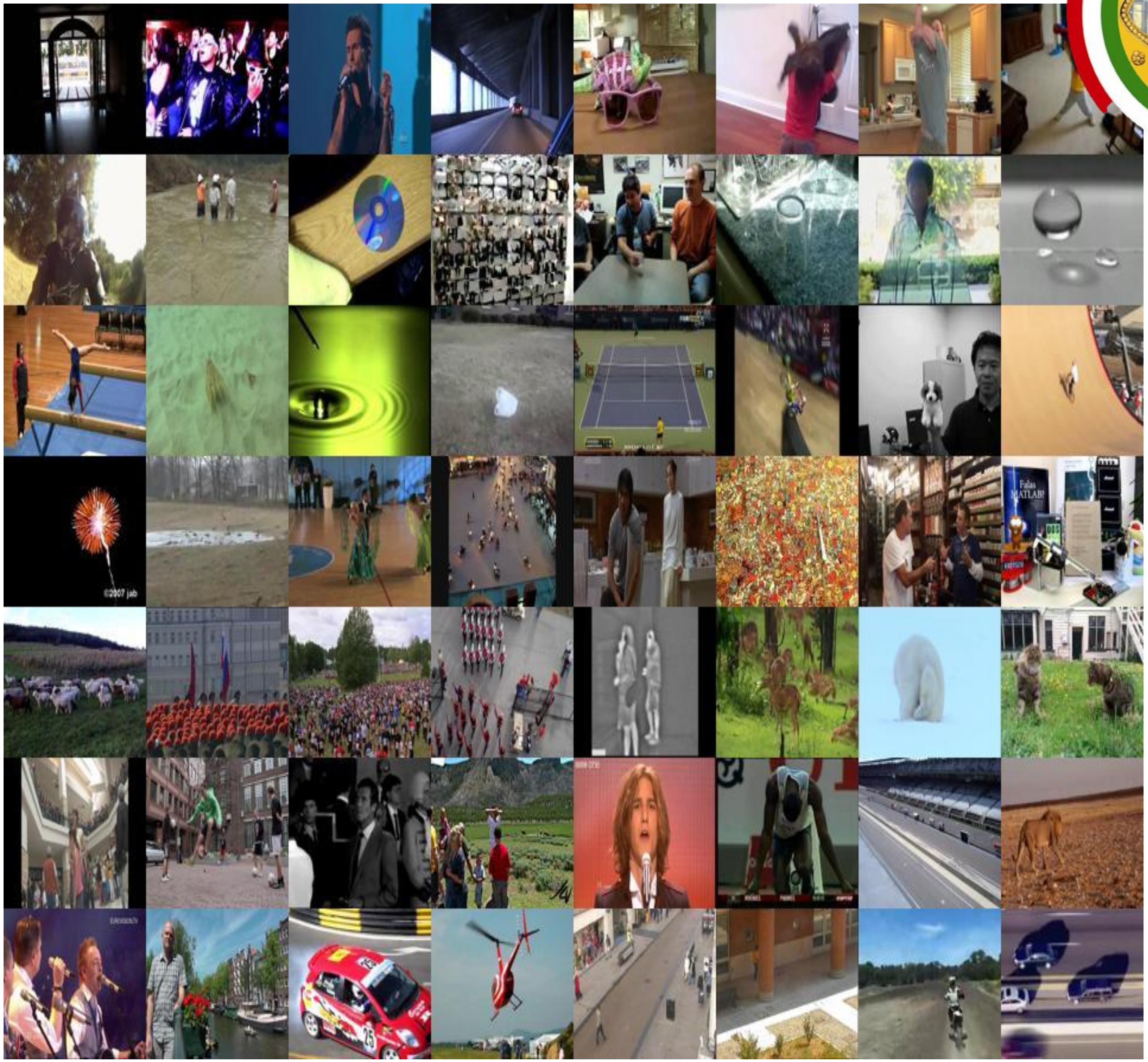
- - a large experimentation
- (with code available over 3 clusters in 3 labs)

19 trackers
BASELINES
STATE OF THE ART

* *D.Chu, A.Smeulders, S.Calderara, R.Cucchiara, A. Dehghan, M.Shah* **Visual Tracking: an Experimental Survey [TPAMI 2013]**

14 tracking challenges in 313 videos

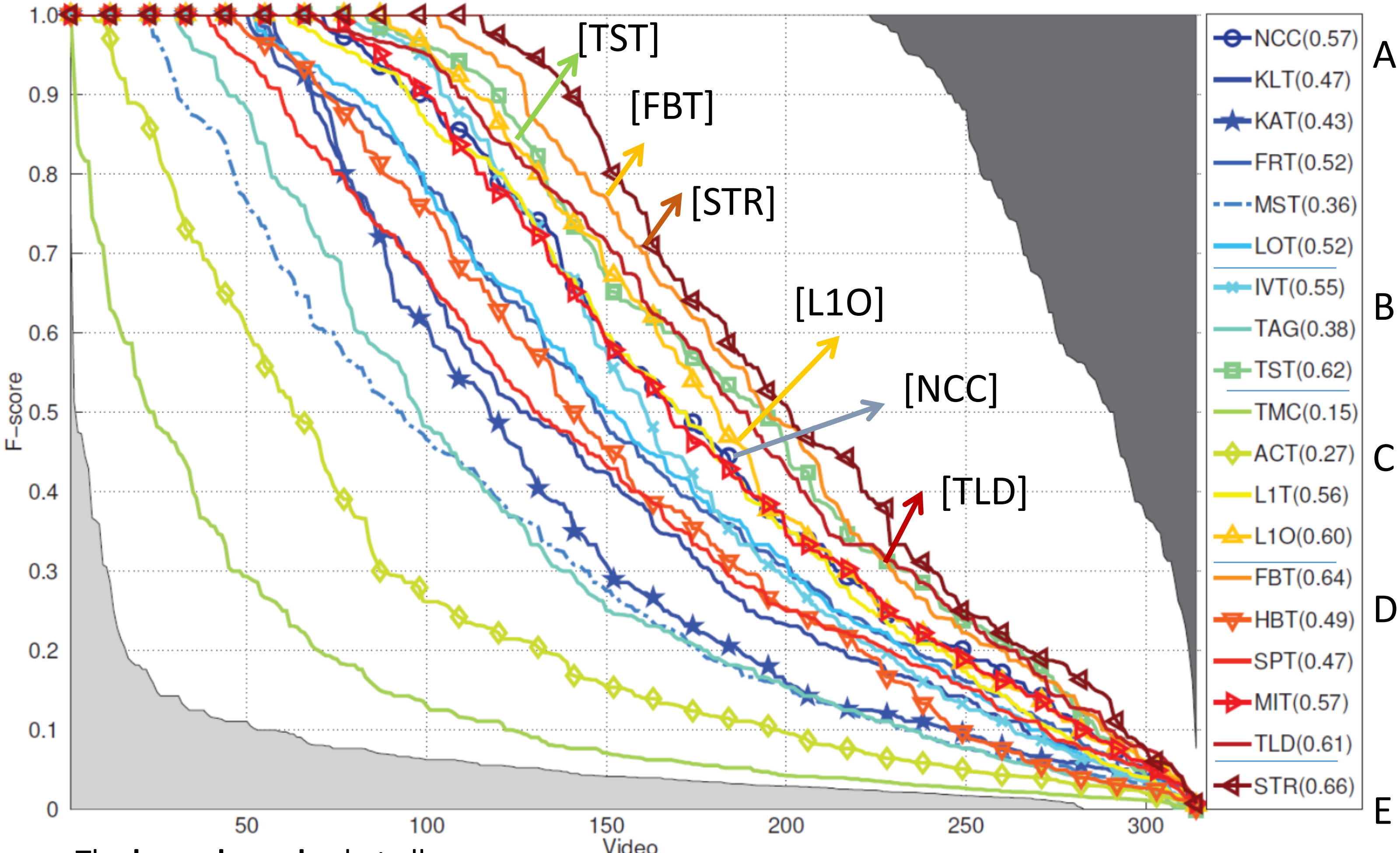
- 01-LIGHT
- 02-SURFACECOVER
- 03-SPECULARITY
- 04-TRANSPARENCY
- 05-SHAPE
- 06-MOTIONSMOOTHNESS
- 07-MOTIONCOHERENCE
- 08-CLUTTER
- 09-CONFUSION
- 10-LOWCONTRAST
- 11-OCCLUSION
- 12-MOVINGCAMERA
- 13-ZOOMINGCAMERA
- 14-LONGDURATION



A comprehensive view Survival curve

About the 30%, correctly tracked only

The **upper bound,** taking the best of all trackers at each frame 10%

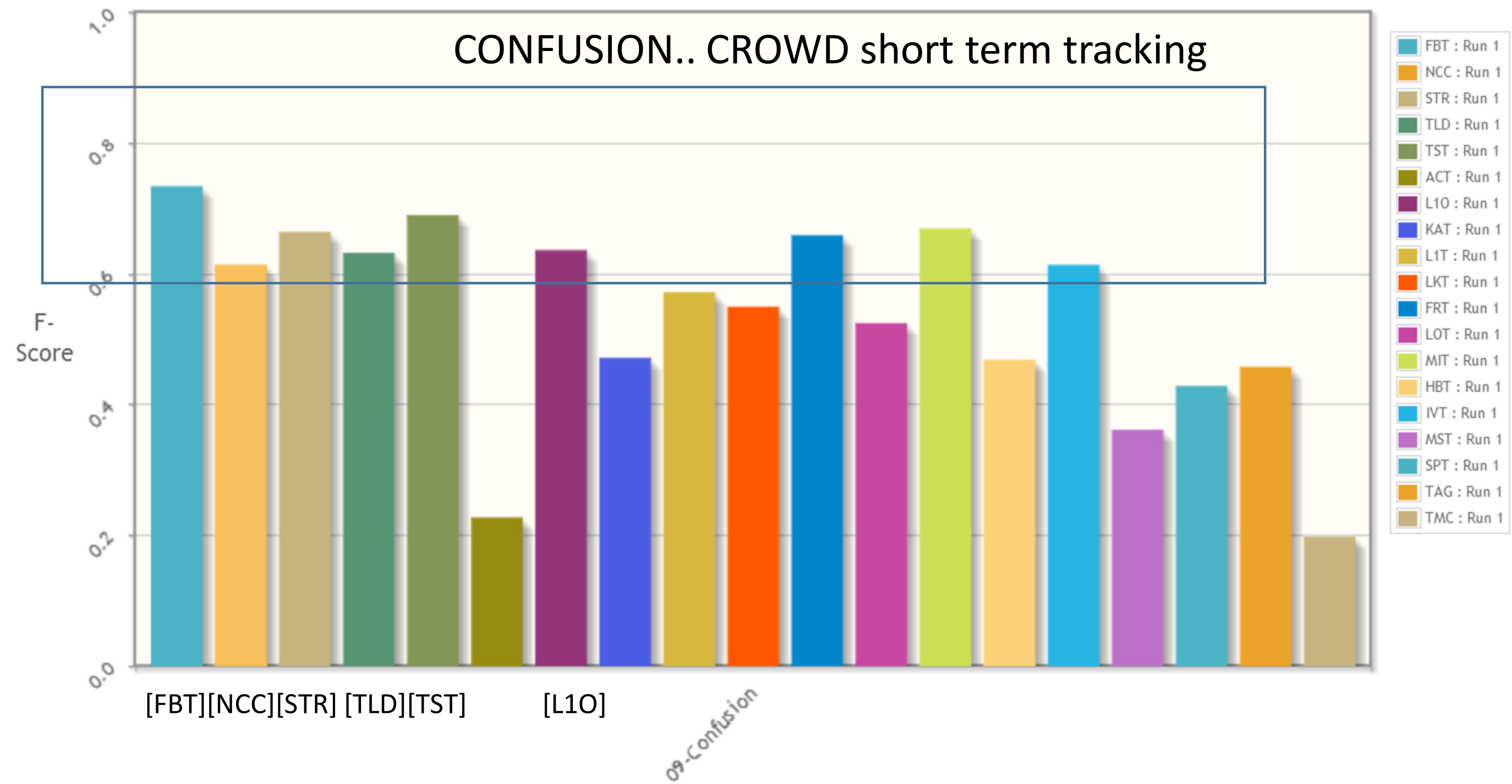


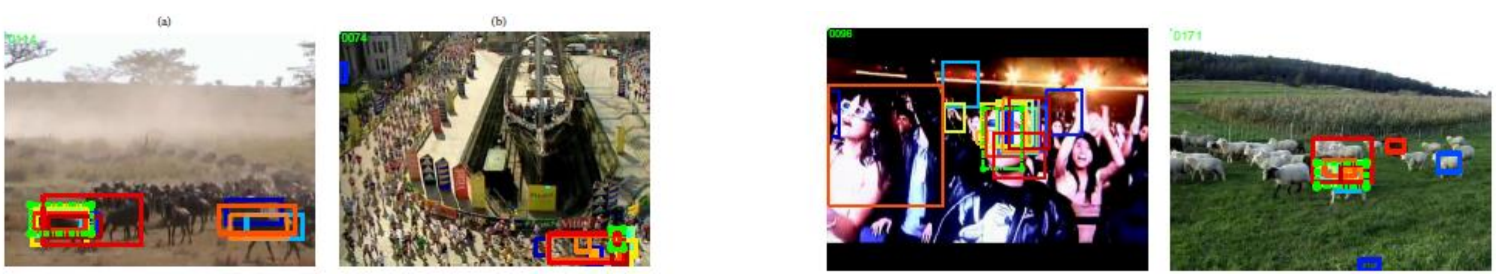
The **lower bound,** what all trackers can do 7%

UNIMORE

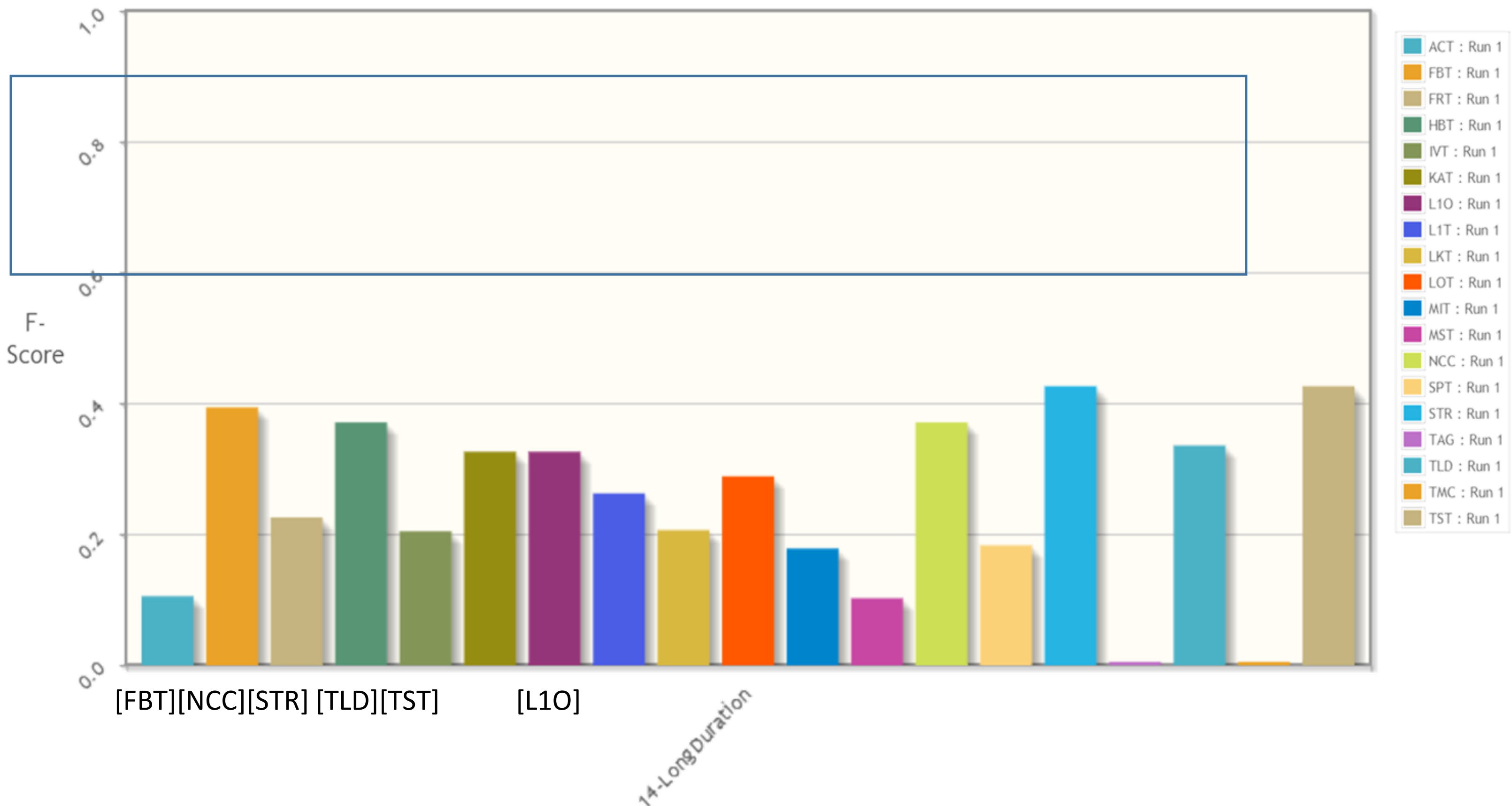
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

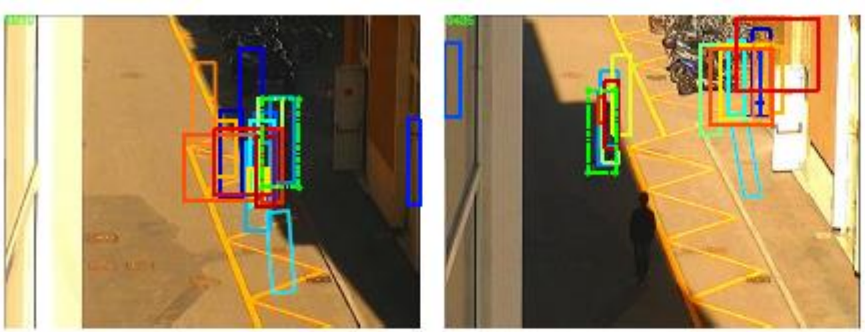
hallenge: trackers comparison





challenge: trackers comparison

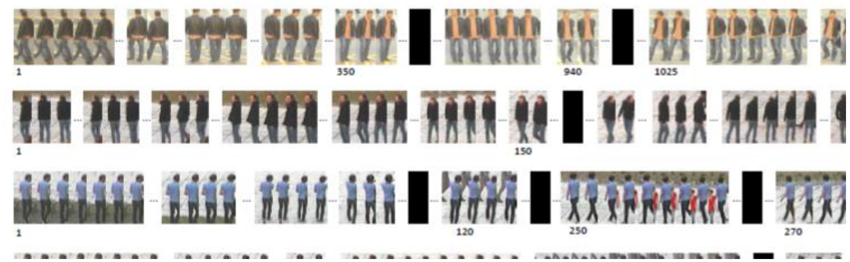


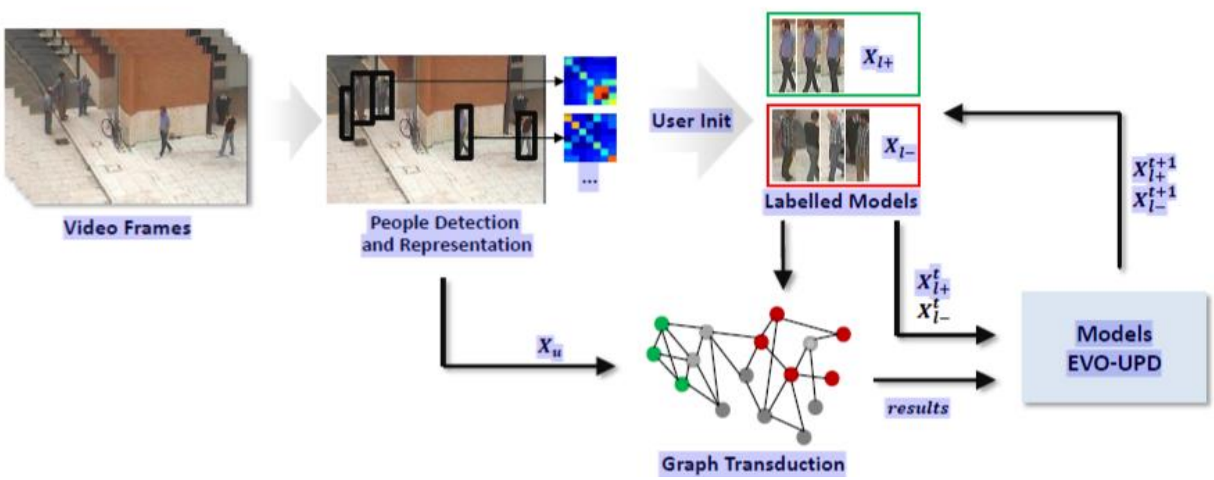


We need more effort

tracking

1. Tracking by background suppression: appearance based tarcking
2. <u>Tracking by detection</u>



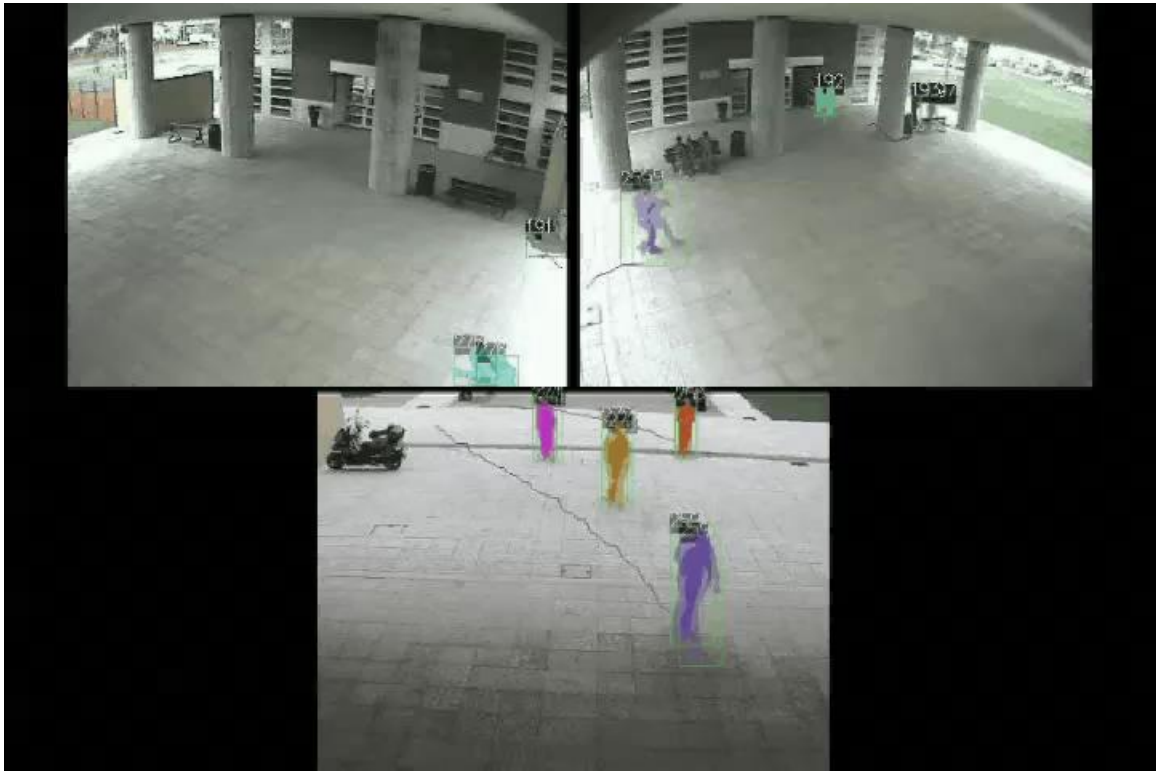


D. Coppi, S.Calderara, R. Cucchiara “Transductive People Tracking in Unconstrained Surveillance” Transactions on CSVT 2015


Stazione_Ferroviaria_76 - 256 sottopasso 2
28 Apr 2010 - 09:34:08/08 ora
Stazione_Ferroviaria_76 - 256 sottopasso 2

Tracking (few) people

- Tracking few people in a constrained environment: «solved problem»

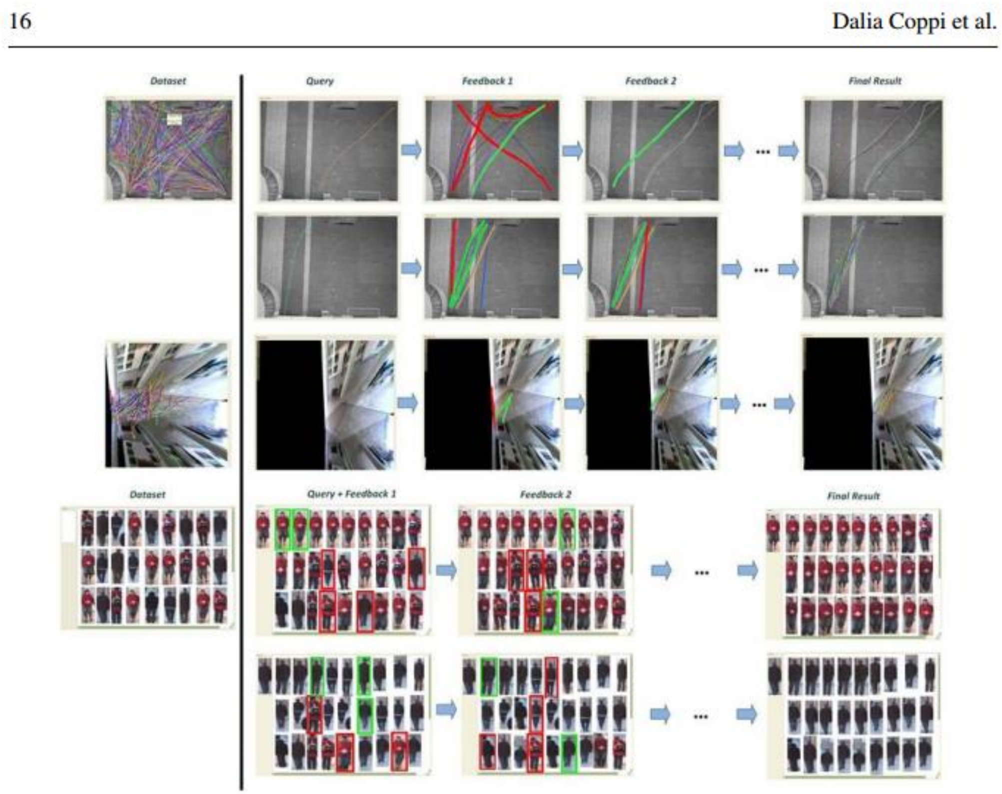


S. Calderara, R. Cucchiara, A. Prati, "Bayesian-competitive Consistent Labeling for People Surveillance"in IEEE Trans. On PAMI , 2008

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Behavior understanding

- Searching from anomalous behaviour
- query systems for forensics applications



Understanding anomalous behavior

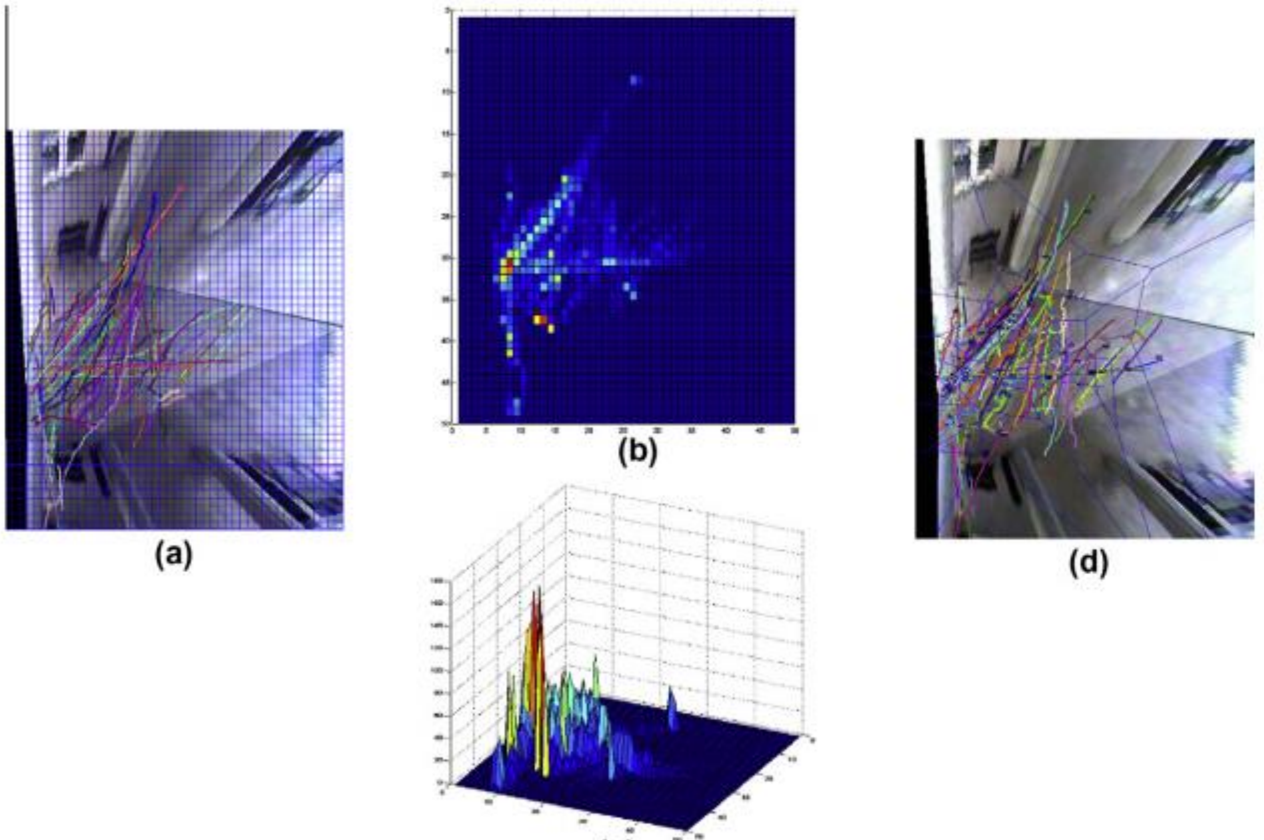


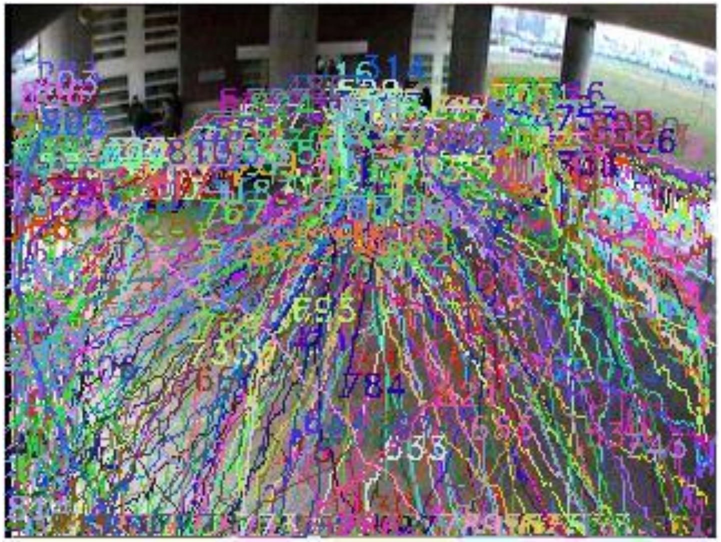
Fig. 2. Irregular partitioning of the image area through Voronoi diagrams: (a) Reports the first regular division of the image ($50 \times 50 = 2500$ cells in this example); (b) shows the top view of the 2D histogram, while (c) shows a side view; and (d) shows the resulting Voronoi diagram with 50 cells.

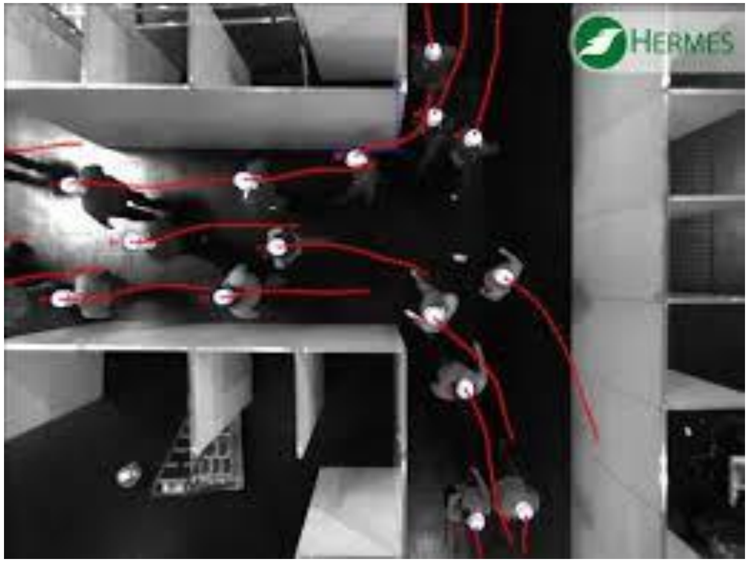
S. Calderara, U. Heinemann, A. Prati, R. Cucchiara, N. Tishby, **["Detecting Anomalies in People's Trajectories using Spectral Graph Analysis"]()**
Computer Vision and Image Understanding, vol. 115, n. 8, pp. 1099-1111, 2011

Trajectory analyisis

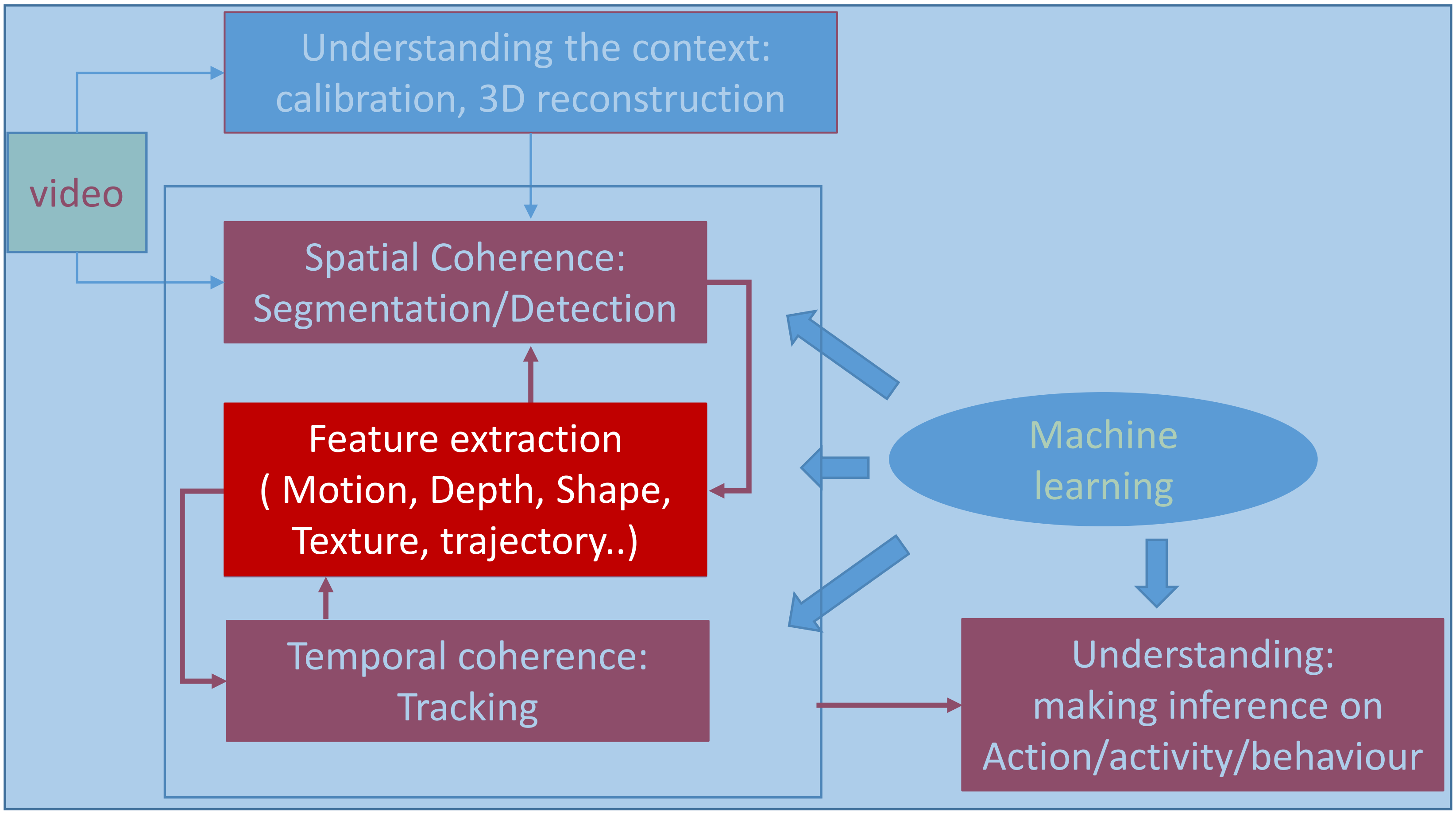
- Detection , tracking , trajectory acquisition
- With few people in the scene
- with controlled experiments

(helmets) PEtrack

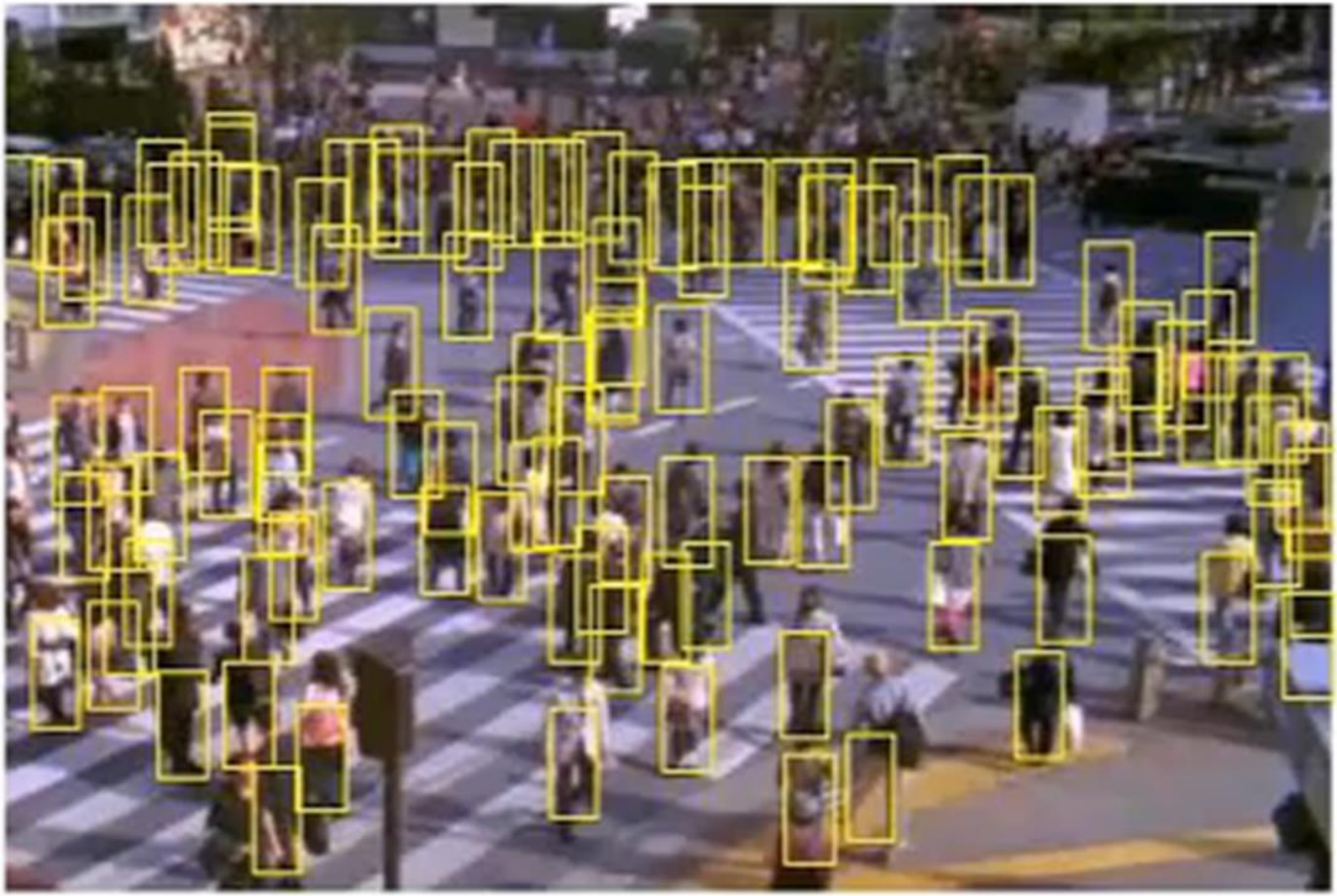




But in real crowd the problem is far to be solved..


Understanding the context:
calibration, 3D reconstruction
video
Spatial Coherence:
Segmentation/Detection
Feature extraction
(Motion, Depth, Shape,
Texture, trajectory..)
Temporal coherence:
Tracking
Machine
learning
Understanding:
making inference on
Action/activity/behaviour

Detection and tracklets



Groups of People

- If tracking were solved…

If the trajectories of every pedestrian in the scene (more or less) were available..

would we be able to discern the behaviour of groups?

Detecting social groups in crowds

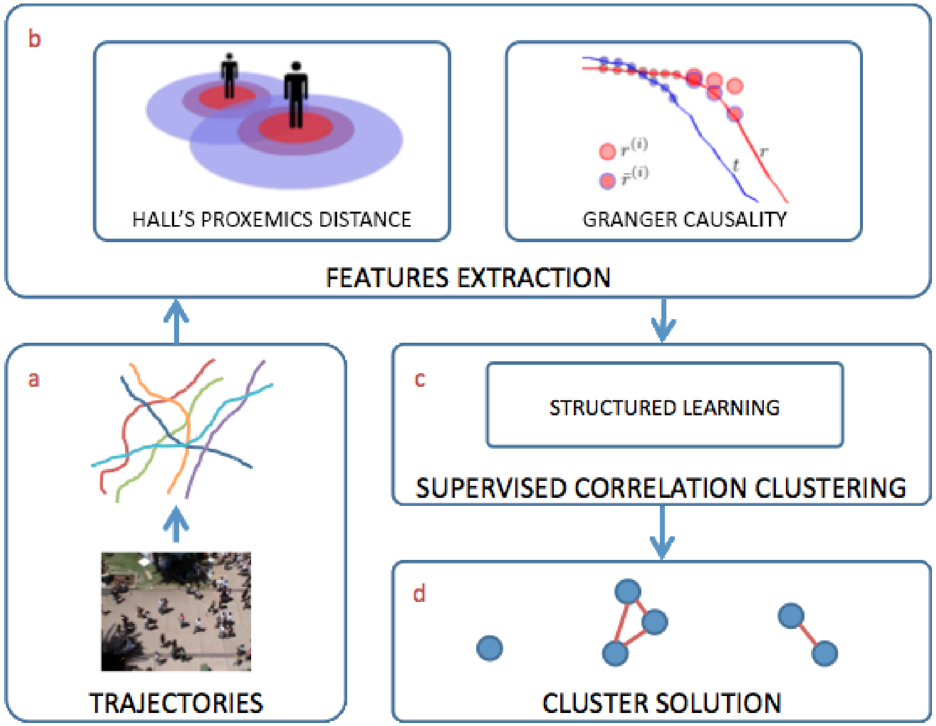
- Group detection: **learn to partition** into groups the pedestrians being part of a crowd observing pairwise relations and transitivities.*
- Integrating two cues:

1. HALL'S PROXEMICS

- Hall's proxemics theory [1] defines ***reaction bubbles*** around every individual and
- the **interaction between pairs** of individuals can be classified according to a quantization of their **mutual distance**

2. GRANGER CAUSALITY

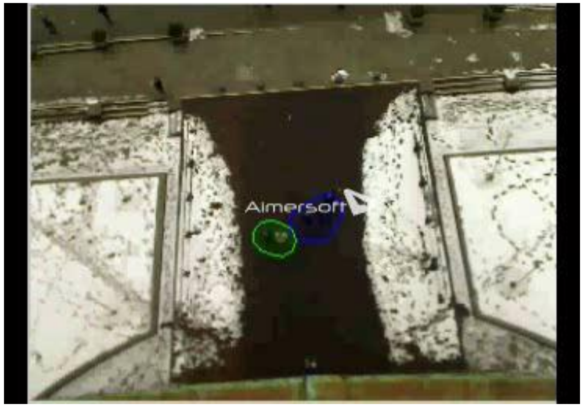
- Intuition: two pedestrian belonging to the same group will probably influence each other position and direction![2]
- The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another



* Structured learning for detection of social groups in crowd Solera, Calderara, Cucchiara, AVSS 2013

Results

Features: Proxemics and Granger causality
Structure function: pair-wise correlation clustering
Group detection: Structured SVM [groups]



F.Solera, S.Calderara, R.Cucchiara"Socially Constrained Structural Learning for Groups Detection in Crowd" II rev PAMI

Conclusions and Open Problems

- Human behavior understanding in video: **a big challenge for big data**
- The problem is not in reasoning, or in data mining

  but in *extracting the correct spatial and temporal visual knowledge*
- Enormous improvements in very few years
- Some challenges "have been solved"
- Most of them have been not solved yet:
  - Human interaction with the environment in an unconstrained and not collaborative way
  - Understanding intention before behavior
  - Working on crowd and in cluttered scenario
  - Working in big footage from the web without text annotation
  - Working on streaming and scalable data….
  - …….

The way ahead is still long…

QUESTIONS??

http://imagelab.ing.unimo.it

Interdipartimental Research Center in ICT
Tecnopolo di Modena
Emila Romagna High Technology Network

PEOPLE

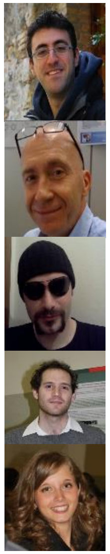
Rita Cucchiara

Costantino Grana

Roberto Vezzani

Simone Calderara

Augusto Pieracci



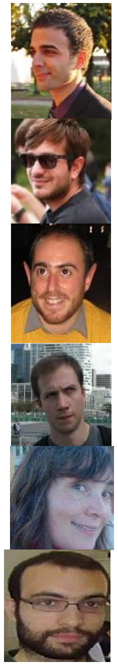
Giuseppe Serra

Paolo Santinelli

Martino Lombardi

Michele Fornaciari

Dalia Coppi



Marco Manfredi

Francesco Solera

Simone Pistocchi

Fabio Battilani

Patrizia Varini

Stefano Alletto,