Video Understanding
for Activity Recognition

Francois BREMOND

STARS project-team,

INRIA Sophia Antipolis, FRANCE

Francois.Bremond@inria.fr

http://www-sop.inria.fr/stars/

http://www-sop.inria.fr/members/Francois.Bremond/

Video Understanding

Objective: Designing systems for Real time recognition of human activities observed by video cameras.

Challenge: Bridging the gap between numerical sensors and semantic events.

Approach: Spatio-temporal reasoning and knowledge management.

Examples of human activities:

for individuals *(graffiti, vandalism, bank attack, cooking)*

for small groups *(fighting)*

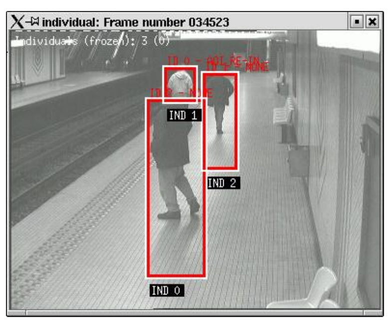
for crowd *(overcrowding)*

for interactions of people and vehicles *(aircraft refueling)*

Video Understanding Applications

- Strong impact for visual surveillance in transportation (metro station, trains, airports, aircraft, harbors)
- Control access, intrusion detection and Video surveillance in building
- Traffic monitoring (parking, vehicle counting, street monitoring, driver assistance)
- Bank agency monitoring
- Risk management (3D virtual realty simulation for crisis management)
- Video communication (Mediaspace)
- Sports monitoring (Tennis, Soccer, F1, Swimming pool monitoring)
- New application domains : Aware House, Health (HomeCare), Teaching, Biology, Animal Behaviors, …

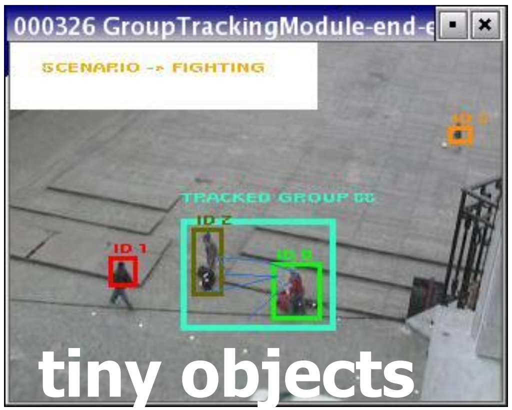
➢ Creation of a start-up Keeneo July 2005 (20 persons): http://www.keeneo.com/

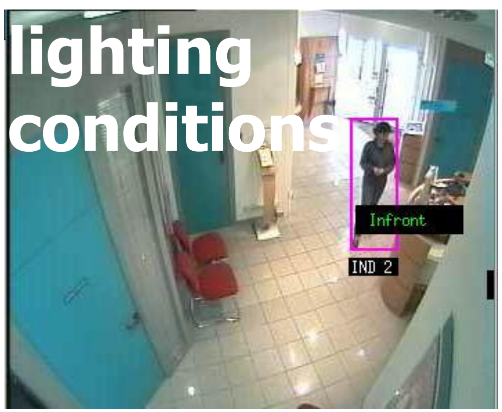


Video Understanding: Issues

Practical issues

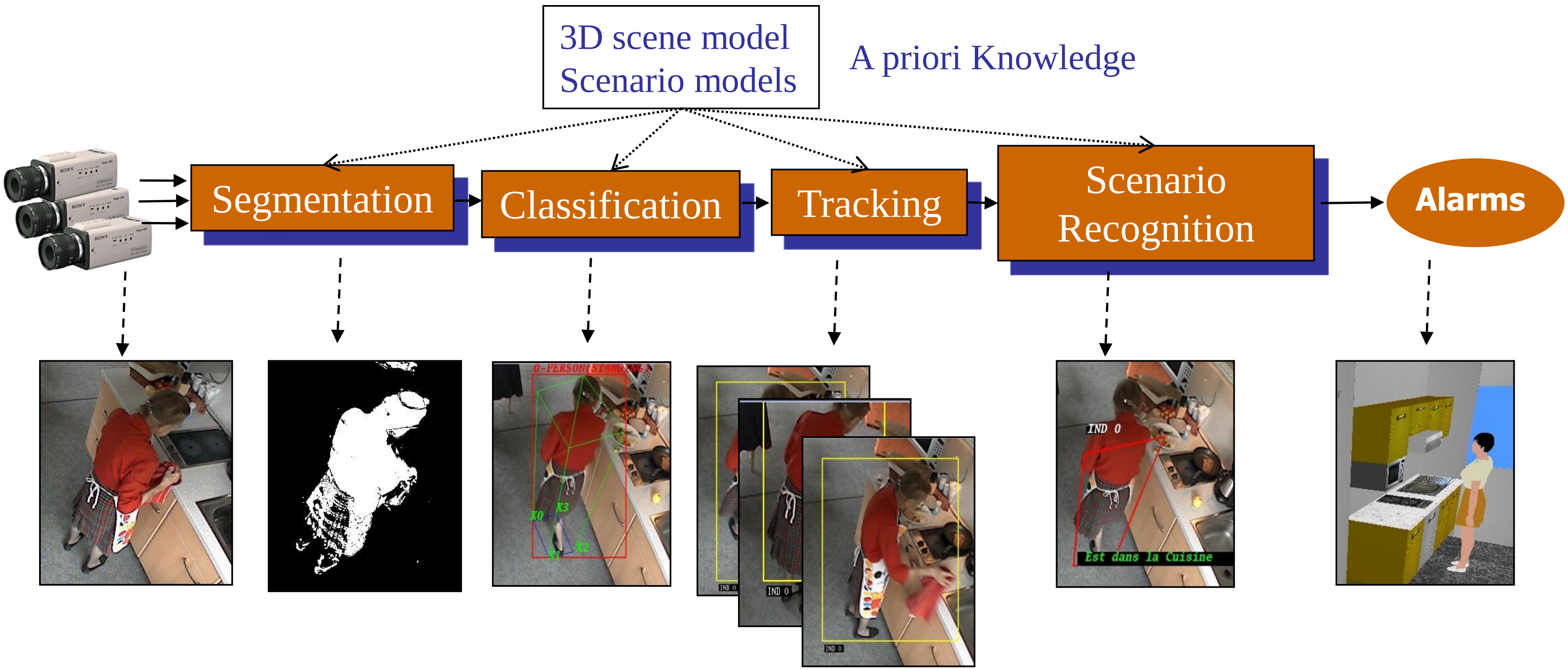
- Video Understanding systems have poor performances over time, can be hardly modified and do not provide semantics





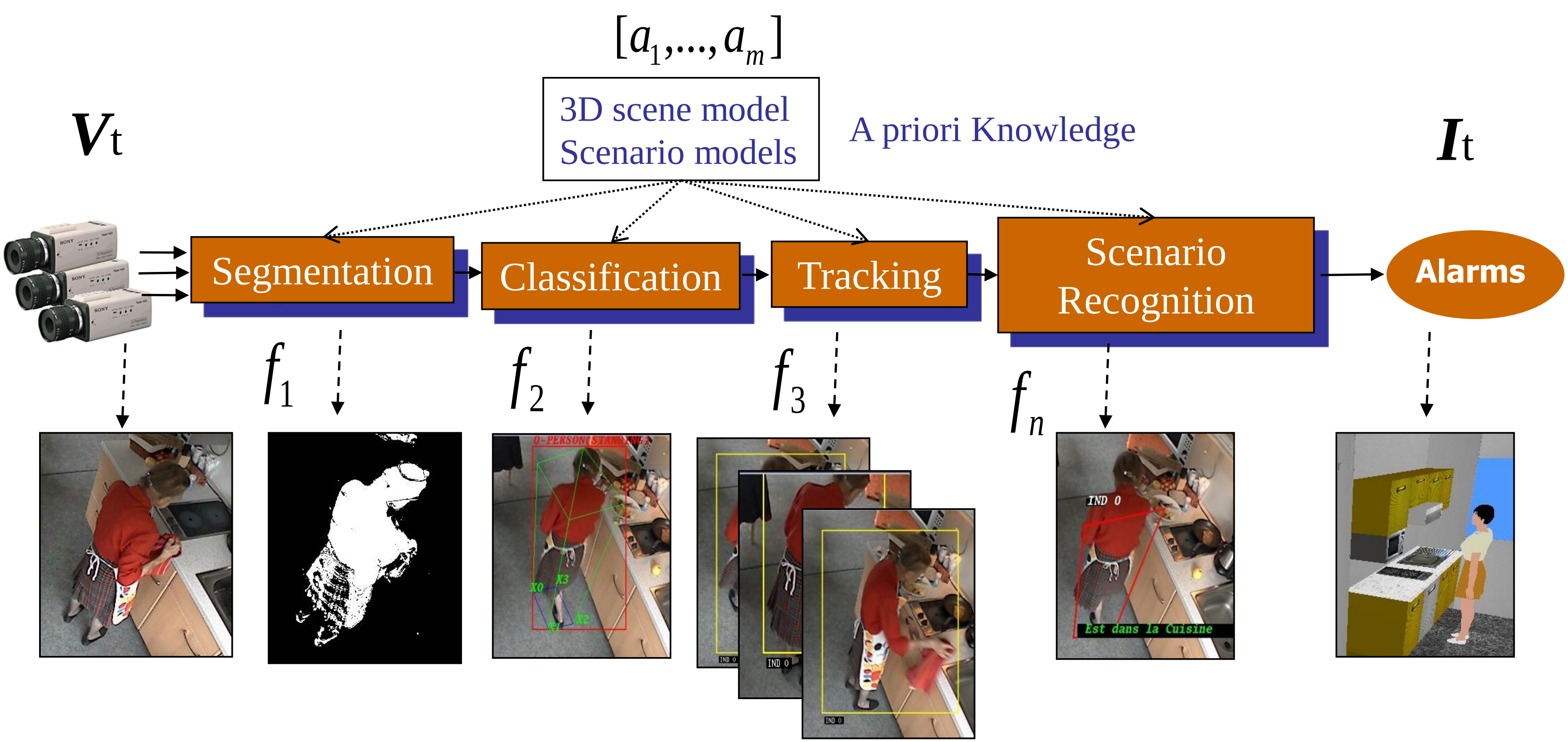
Video Understanding

Objective: *Real-time Interpretation of videos from pixels to events*



Video Understanding

Objective: *Real-time Interpretation of videos from pixels to events*



Video Understanding: Approach

Global framework for video understanding

Interpretation at time t :
moving objects, metadata, events, …

Video processing

Sensing data, signal

$$f_n \circ f_{1[a_1,\ldots,a_m]}(V_{[t-1,t]}, I_{t-1}) = I_t$$

Processing Parameters : thresholds, reference image,…
Contextual Information : sensor, static scene model,…
Knowledge : physical object models, scenario models,…

Video Understanding: Approach

5 Challenges in video understanding

5) Configuration, optimisation, **system generation** from specification

1) Robustness of Video Processing depending on data domains

2) spatio-temporal **reasoning,** uncertainty and semantics

Interpretation at time t : moving objects, metadata, events, …

Video processing

Sensing data, signal

$$f_n \circ f_{1[a_1,\ldots,a_m]}(V_{[t-1,t]}, I_{t-1}) = I_t$$

Processing Parameters : thresholds, reference image,…
Contextual Information : sensor, static scene model,…
Knowledge : physical object models, scenario models,…

3) Evaluation, Ground-truth, Metrics, videos

4) Knowledge **representation, learning**

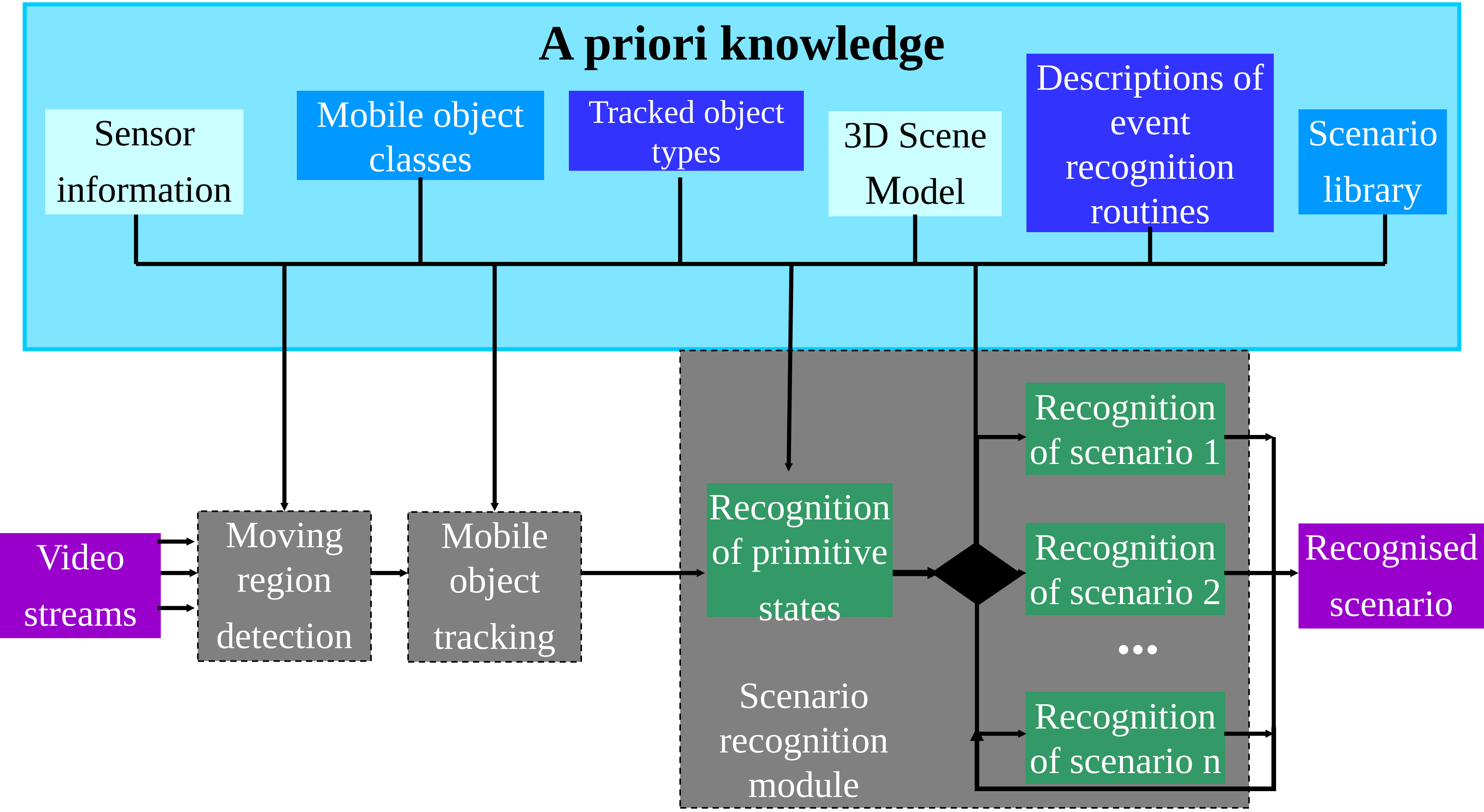
Video Understanding

Outline:

- Knowledge Representation : Scene Model
- Input of the Scenario Recognition process:
  - Object Detection, Object Tracking, Action Recognition
- Event/Scenario Representation
- Bag of Words
- Graphical models
- Temporal Scenario Recognition
  - Scenario representation
  - Recognition process
- Applications: recognition of several scenarios
- Learning Scenario Models

Knowledge Representation

Knowledge Representation



Knowledge Representation: 3D Scene Model - Context

Definition : a priori knowledge of the observed empty scene

- Cameras: 3D position of the sensor, calibration matrix, field of view,...
- 3D Geometry of physical objects (bench, trash, door, walls) and interesting zones (entrance zone) with position, shape and volume
- Semantic information : type (object, zone), characteristics (yellow, fragile) and its function (seat)

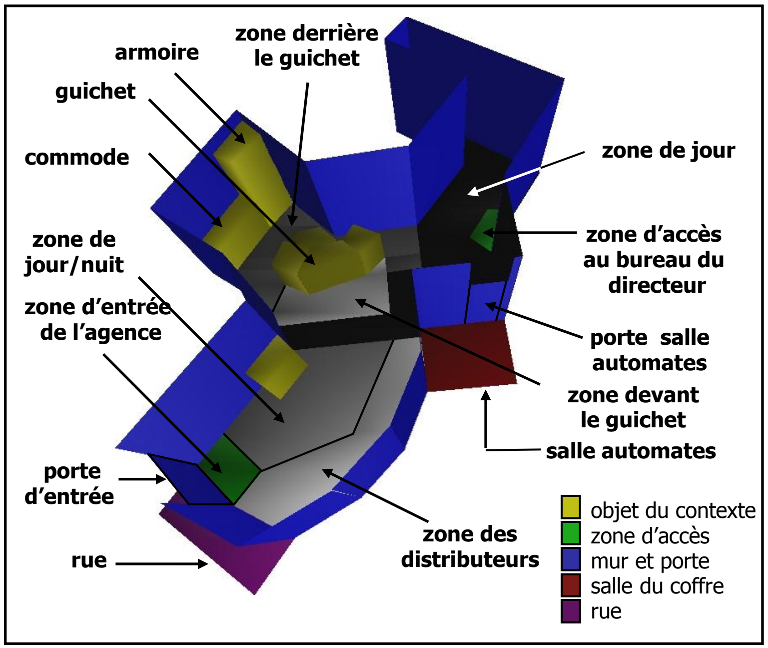
Role:

- to keep the interpretation independent from the sensors and the sites : many sensors, one 3D referential
- to provide additional knowledge for behavior recognition

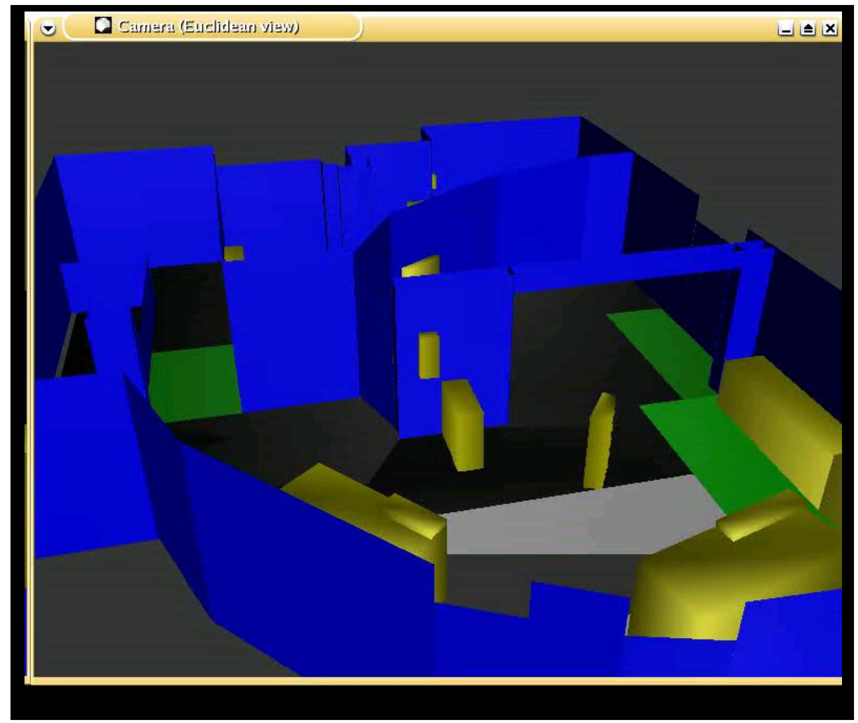
Knowledge Representation : 3D Scene Model

3D Model of 2 bank agencies

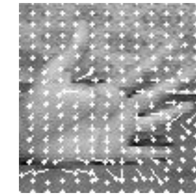
Les Hauts de Lagny



Villeparisis



Object (People) detection

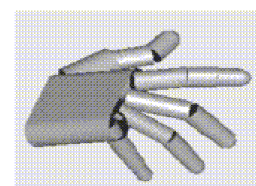


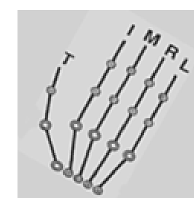
Estimation of Motion

- Need of textured objects
    - Optical Flow: Estimation of apparent motion (pixel intensity between 2 frames)
    - Local descriptors (patches, tracklets, gradients (SURF, HOG), color histograms, moments over a neighborhood)

Object model

- Need of mobile object model
    - 2D appearance model (shape, color, pixel template)
    - 3D articulate model





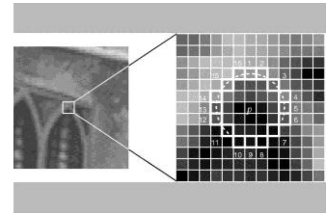
Reference image subtraction

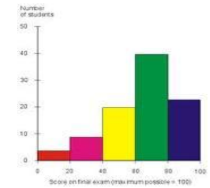
- Need of static cameras
    - Most robust approach (model of background image)
    - Most common approach even in case of PTZ, mobile cameras

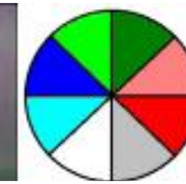
Complex Scenes: People detection

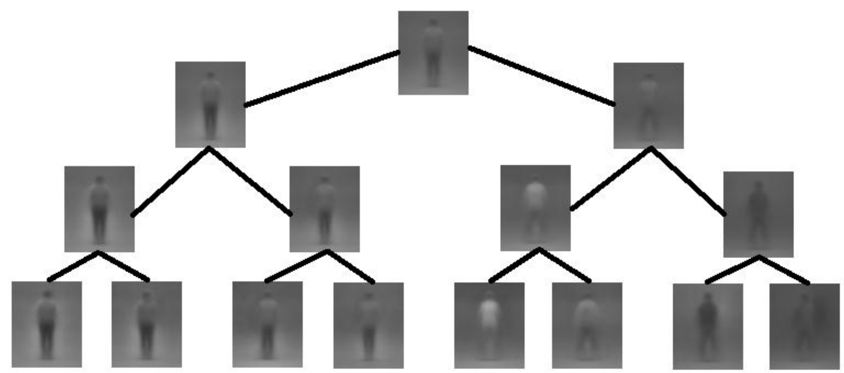
Issues in Local Descriptor People Detection:

- **Features:**
  - HOG, LBP, Covariance Matrix, Haar, SIFT, Granules
- **Learning paradigm:**
  - Adaboost, Hierarchical trees, SVM
- **Training / testing databases:**
  - Camera view point, distortion, resolution,
  - Occlusion, pose,
  - Background samples
- **Processing time:**
  - Training (best feature selection)
  - Detection (scanning window sampling rate, multi-resolution)
- **Filtering:**
  - Overlapping scanning window, candidate selection
  - 3D constraint, motion segmentation,
- **Body parts:**
  - Global detection
  - Model based association
  - E.g. head, torso, legs ...





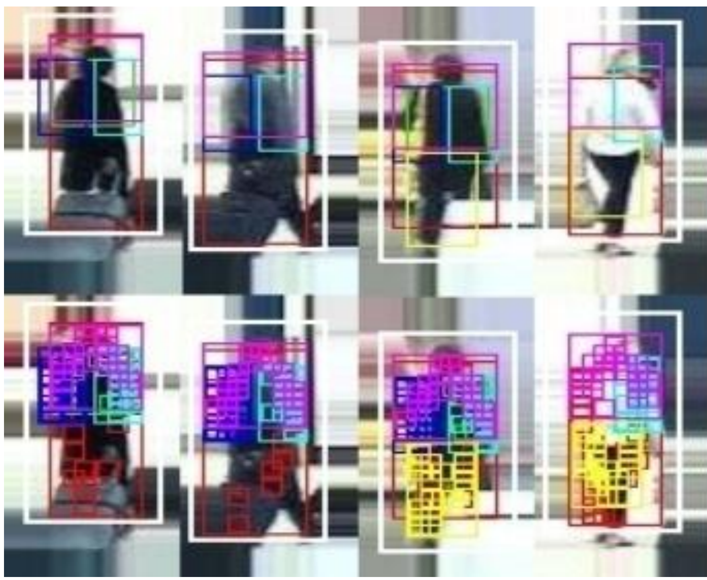




Complex Scenes: People detection

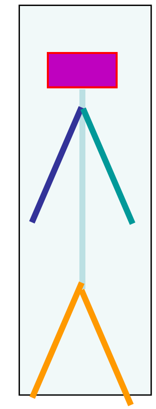
Body part combination

- Body parts combination:
  - Detected body parts (HOG detector trained on manually selected areas of the person)
  - Example below in TrecVid camera 1



Example of detected with corresponding HOG cells

Detection examples



person

omega

left arm

right arm

torso

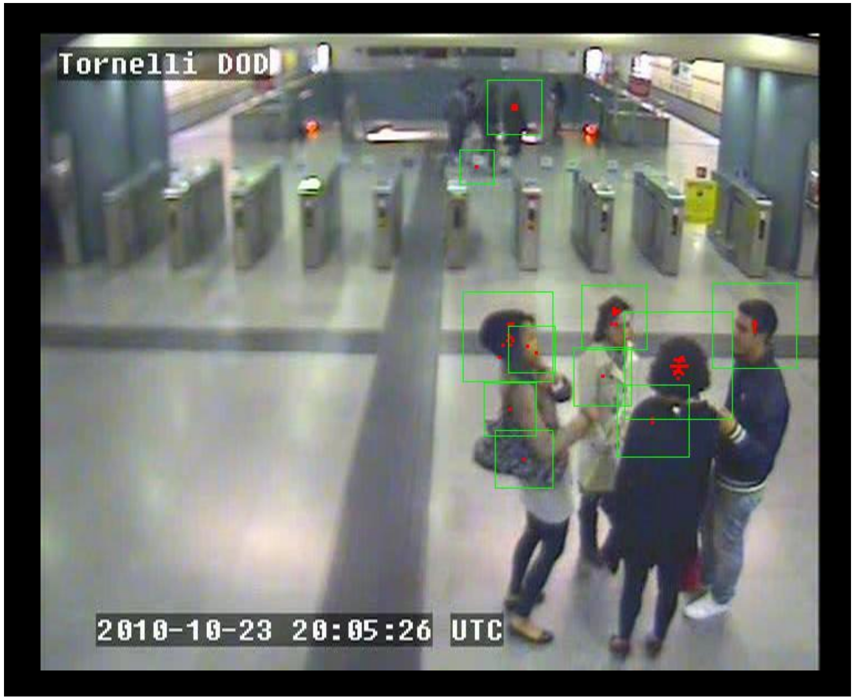
legs

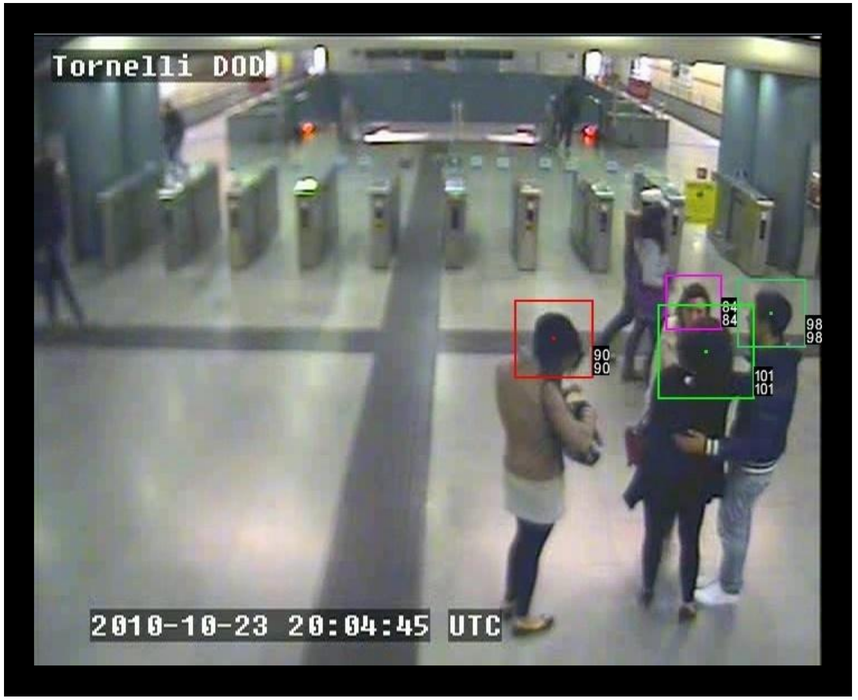
Head detection and tracking results

Training head database: selection of 32x32 head images from publicly available MIT, INRIA and NLDR datasets. A total of 3710 images were used

Training background dataset: selection of 20 background images of TrecVid and 5 background images of Torino ‘Biglietattrice.

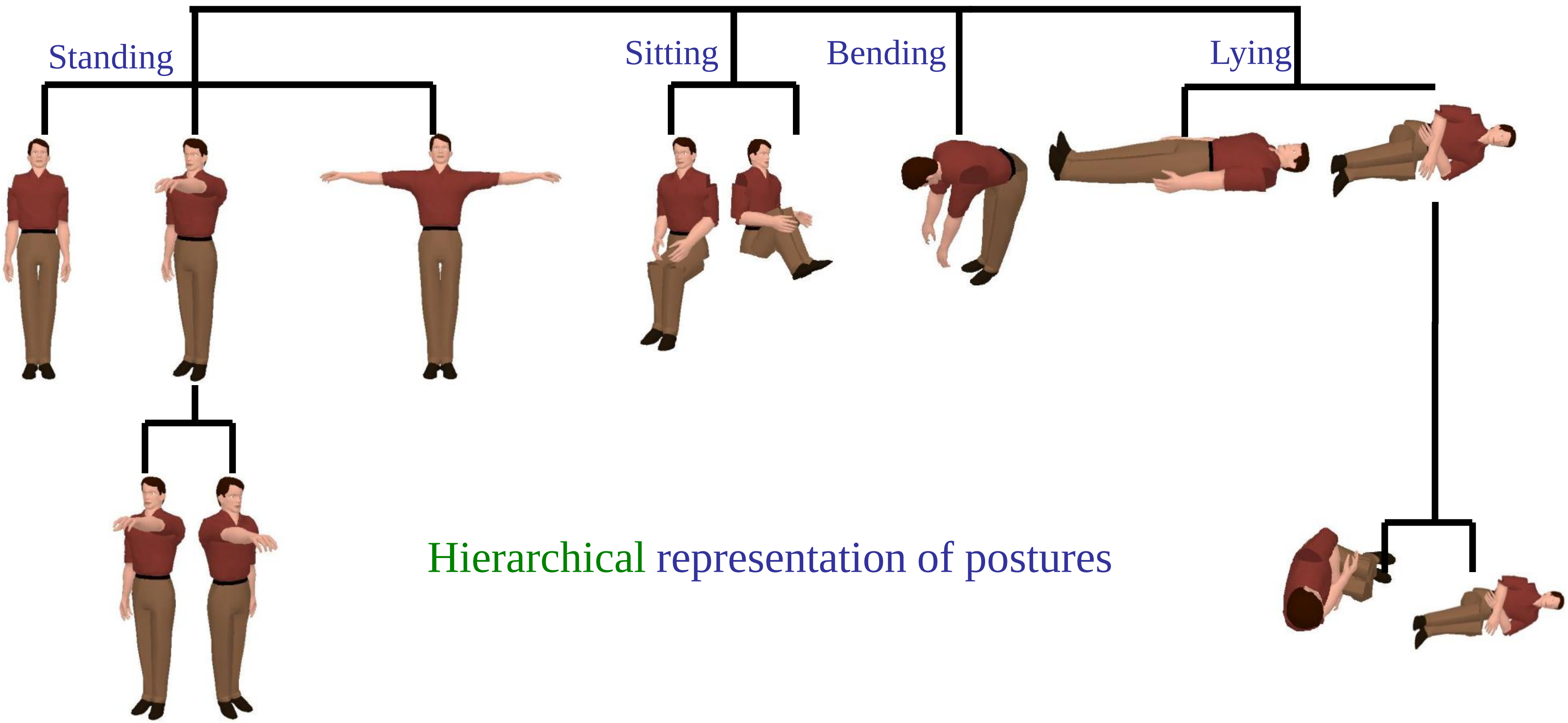
Speed: Once integral images are computed, the algorithm reaches ~ **1fps** for 640x480 pixels





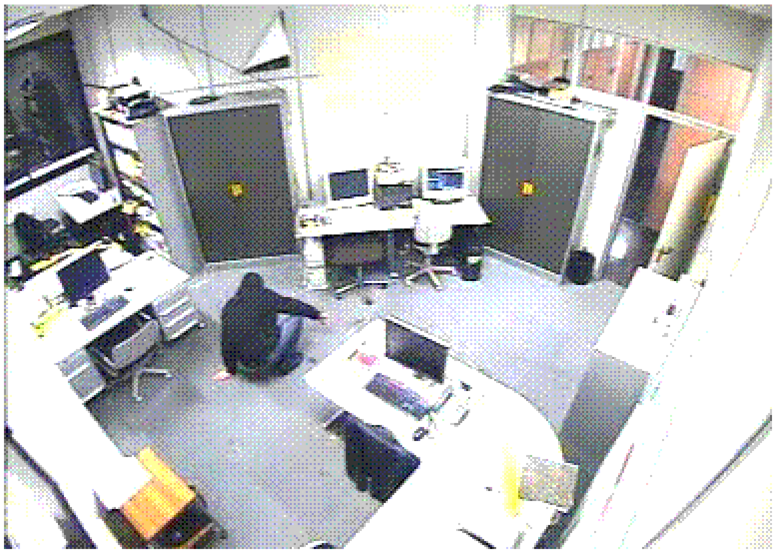
Left: head detection examples and right: tracking examples in Torino underground

Posture Recognition : Set of Specific Postures



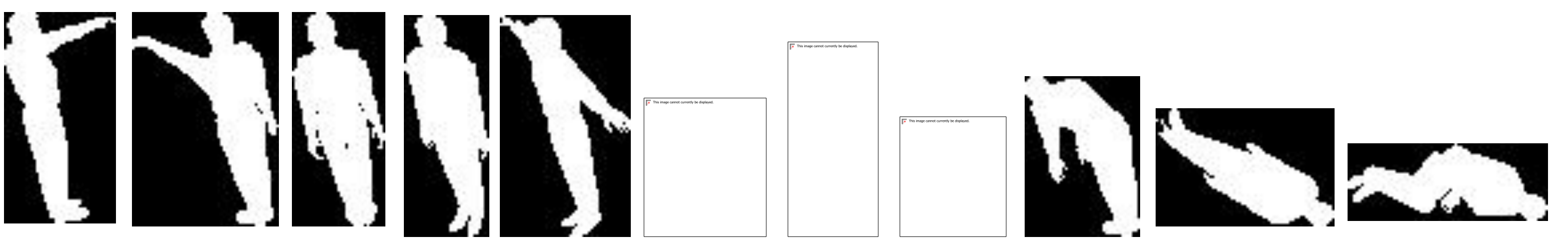
Hierarchical representation of postures

Posture Recognition : silhouette comparison



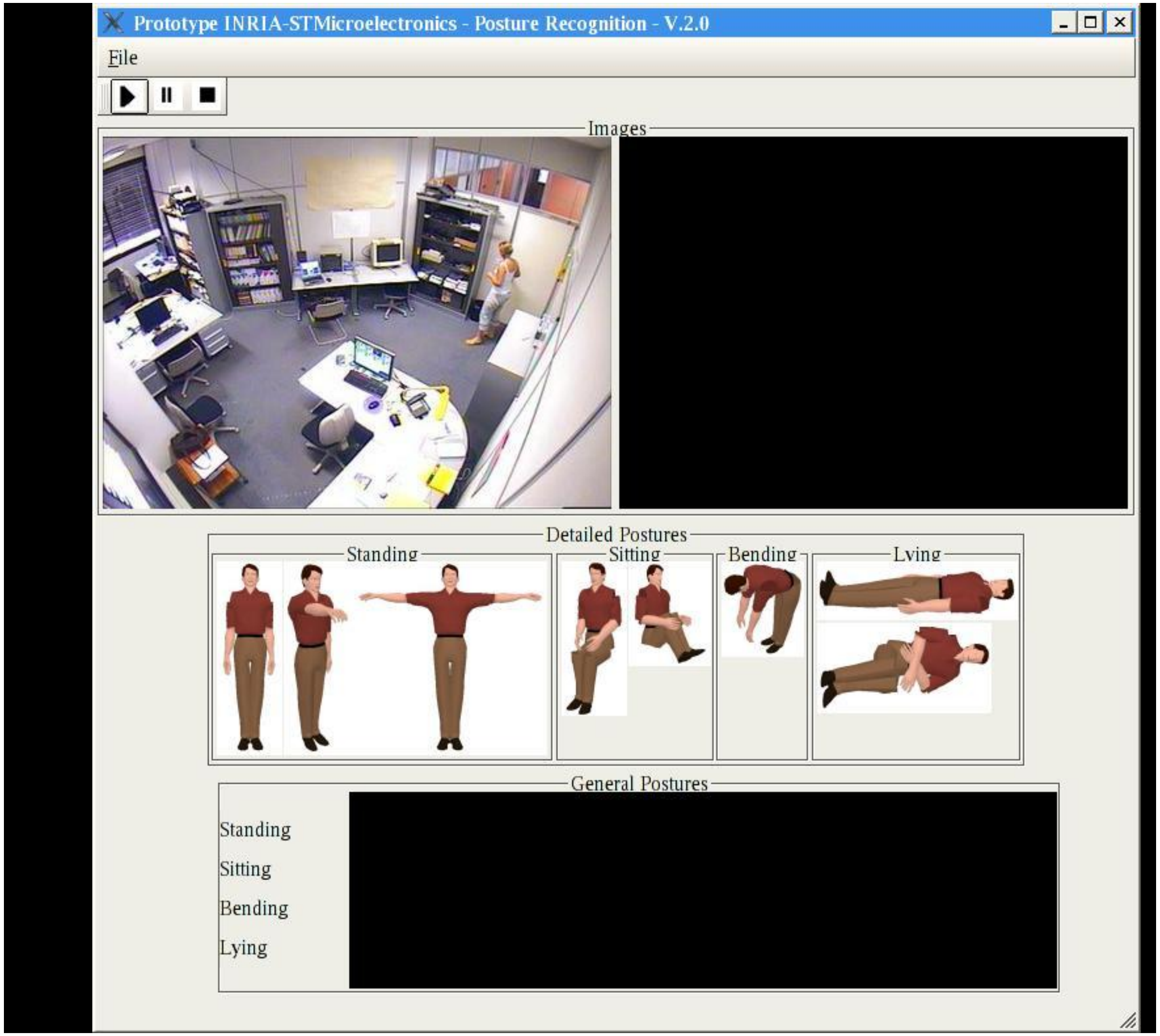
Real world

Virtual world



Generated silhouettes

Posture Recognition : results



Event/Scenario Recognition

Outline:

- Event/Scenario Representation
- Bag of Words
- Graphical models
- Temporal Scenario Recognition
  - Scenario representation
  - recognition process
- Applications: recognition of several scenarios
- Learning Scenario Models

Event Representation: Video Event Ontology

Definition:

- Video Event Ontology: a set of concepts and relations is used as a reference between all the actors of the domain to describe knowledge

Properties:

- Enable experts to describe video events of interest (e.g. composite event) and to structure the knowledge: ontology of the application domain.
- Share knowledge between developers: ontology of visual concepts (e.g. a stopped mobile object)
- Ease communication between developers and end users and enable performance evaluation: ontology of the video understanding process (what should be detected: mobile object (a parked car), object of interest (a door), visible object (occluded person))
- Architecture interoperability: separation between specification and knowledge description

Event Representation

Several entities are involved in the scene understanding process:

- **Moving region**: any intensity change between images.

- **Context object**: predefined static object of the scene environment (entrance zone, wall, equipment, door...).

- **Physical object** : any moving region which has been tracked and classified (person, group of persons, vehicle, … etc).

- **Physical object of interest**: meaningful object, but depending on applications (person/ door, parked vehicle, … etc).

Event Representation

Actions, States, Streams, Threads, Events, Situations, chronicles, behaviours, activities and scenarios… : a large variety

- more or less composed of sub-events (running/fighting).
- involving few/many actors (football game).
- general (standing)/sensor and application/view (sit down, stop) dependent.
- spatial granularity: the view observed by one camera/the whole site.
- temporal granularity: instantaneous/long term with complex relationships (synchronize).

➢ 3 levels of complexity depending on the complexity of temporal relations and on the number of physical objects :

- non-temporal constraint relative to one physical object (sitting). Intuitive combination of feature probabilities to get better precision.
- temporal sequence of sub-scenarios relative to one physical object (open the door, go toward the chair then sit down). Filtering noisy input, versus meaningful changes.
- complex temporal constraints relative to several physical objects (A meets B at the coffee machine then C gets up and leaves). Need of logic reasoning (declarative, expressive) but sensitive to vision errors.

Event Representation

Video events: real world notions corresponding to short actions (coherent unit of motion) up to activities.

- *Primitive State:* a spatio-temporal property linked to vision routines involving one or several actors, valid at a given time point or stable on a *time interval*

    Ex : « close», « walking», « sitting»

- *Composite State: a combination of primitive states*
- *Primitive Event:* significant change of states

    Ex : « enters», « stands up», « leaves »

- *Composite Event:* a combination of states and events. Corresponds to a long term (symbolic, application dependent) activity.

    Ex : « fighting», « vandalism»

Event Recognition

Several formalisms can be used:

- **Event representation**:
  - n-ary tree, frame, aggregate (structure).
  - finite state automaton, sequence (evolution).
  - graph, set of constraints.

- **Event recognition**:
  - Feature based routine.
  - Classification, Bayesian, neural network, SVM, clustering, BoW.
  - DBN, HMM, Petri net.
  - Stochastic grammar, Prolog.
  - Constraint propagation, verification of temporal constraints.

Event Recognition : Issues

Performance: robustness of real-time (vision) algorithms

Bridging the gaps at different abstraction levels:

- From sensors to image processing
- From image processing to 4D (3D + time) analysis
- From 4D analysis to semantics

Uncertainty management:

- uncertainty management of noisy data (imprecise, incomplete, missing, corrupted)
- formalization of the expertise (fuzzy, subjective, incoherent, implicit knowledge)

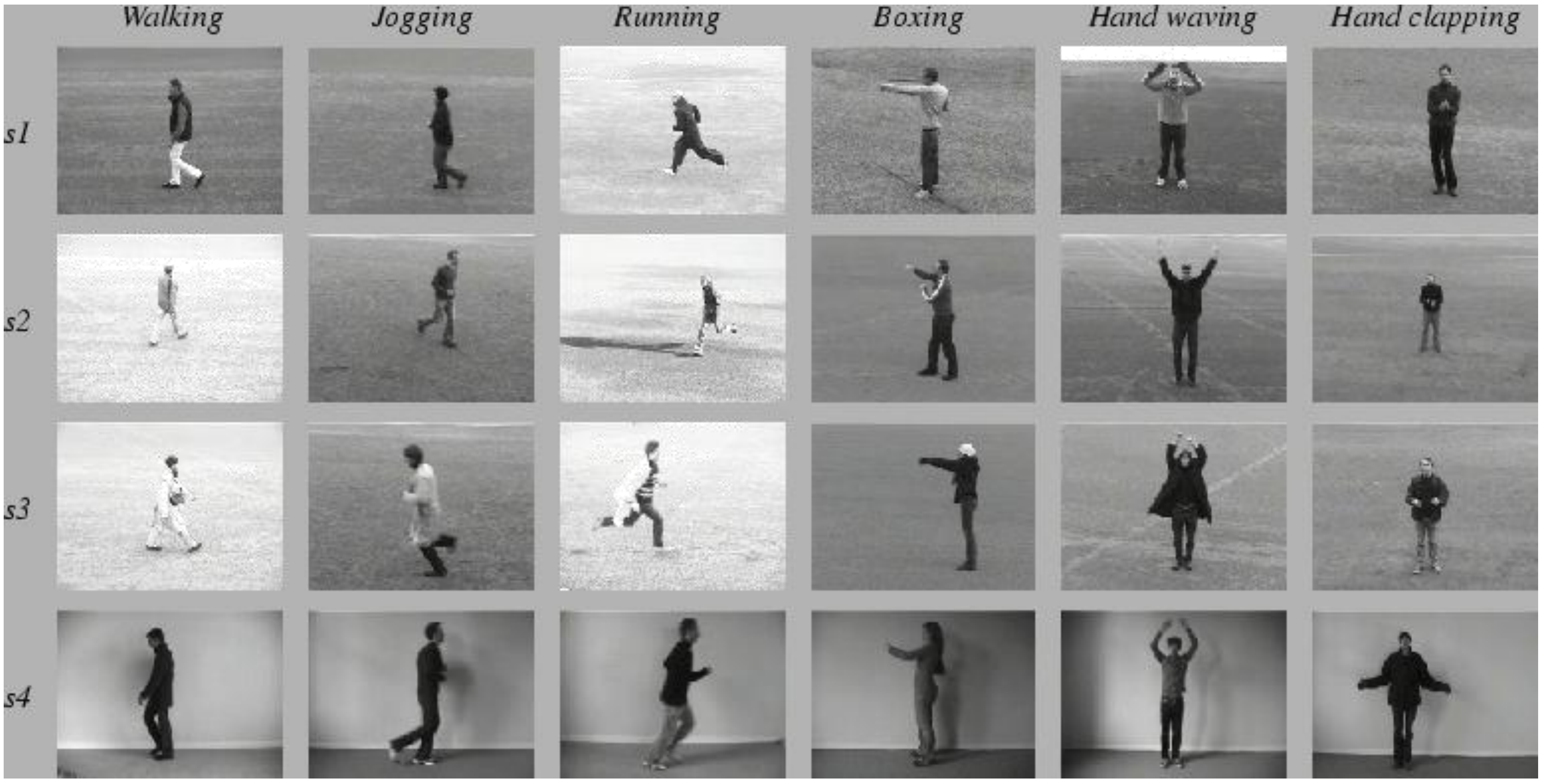
Independence of the models/methods versus:

- Sensors (position, type), scenes, low level processing and target applications
- several spatio-temporal scales

Knowledge management :

- Bottom-up versus top-down, focus of attention
- Regularities, invariants, generic models and context awareness
- Knowledge acquisition versus ((none, semi)-supervised, incremental) learning techniques
- Formalization, modeling, ontology, standardization

Action Recognition (MB. Kaaniche, P. Bilinski)

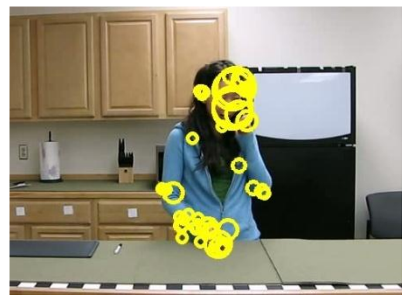


Type of gestures and actions to recognize

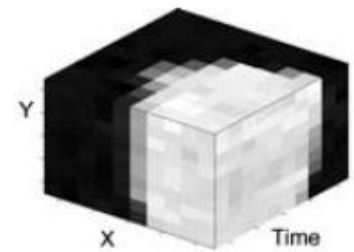
Action Recognition Algorithms

Videos

Point detector



Point descriptor



All feature vectors

BOW model

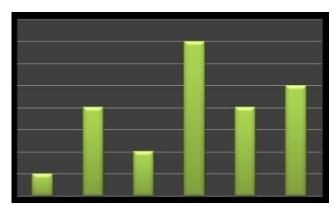
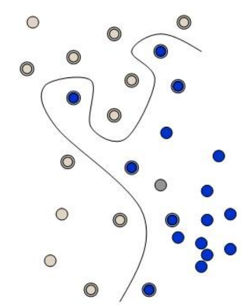
Codebook generation

Bag-of-words model

Offline Learning:

All training feature vectors → Codebook generation **(different sizes)** → Database

Online recognition:

All testing feature vectors → Assignment to the closest codeword →  → Non-linear SVM 

Histogram of codewords

ADL Dataset

ADL - Results

Codebook size / Descriptor	HOG [72-bins]	HOF [90-bins]	HOG-HOF [162-bins]	HOG3D [300-bins]
Size 1000	85.33%	**90.00%**	**94.67%**	**92.00%**
Size 2000	**88.67%**	90.00%	92.67%	91.33%
Size 3000	83.33%	89.33%	94.00%	90.67%
Size 4000	86.67%	89.33%	94.00%	85.00%
Best	**88.67% (4)**	**90.00% (3)**	**94.67% (1)**	**92.00% (2)** (7% diff)

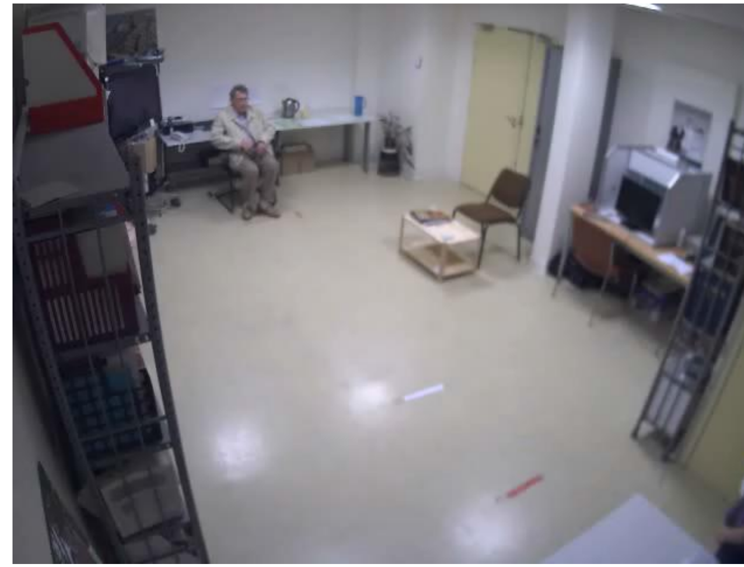
SOA: 96% Wang [CVPR11]

Issues in Action Recognition

- Different detectors (Hessian, Dense sampling, STIP...)
- Different parameters of descriptors (grid size, ...)
- Different classifiers (k-NN, linear-SVM, ...)
- Different clustering algorithms (Bossa Nova, Fisher Kernels,…)
- Different resolutions of videos
- Generic to other datasets (IXMAS, UCF Sports , Hollywood, Hollywood2, YouTube, ...)
- Finer actions, more discriminative, without context...

Issues in Action Recognition

- Finer actions, more discriminative



Event Recognition: Specific Routines
Advisor project: F. Cupillard, A. Avanzi,…

Event Recognition: Specific Routines
Results in metro station

Scenario Recognition :
Running

Scenario:Running
-> ALARM

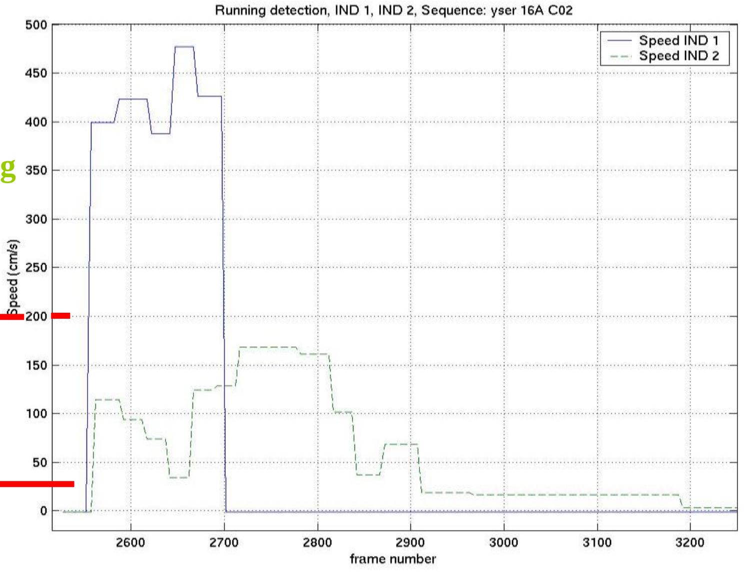
State: walking

State: stopped

running

walking

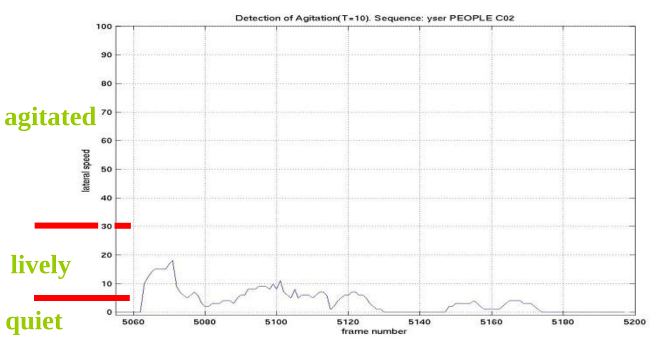
stopped

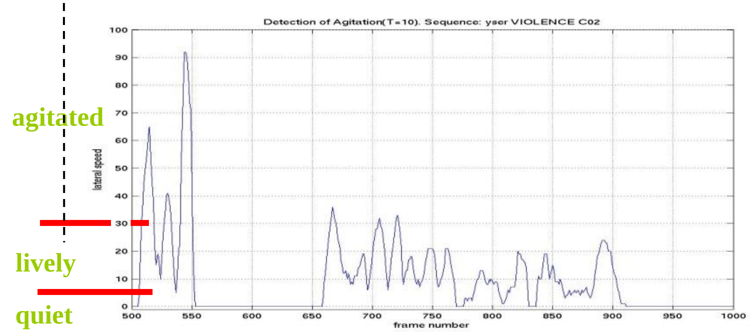


Event Recognition: Specific Routines
Results in metro station

Scenario:Agitated Behaviour
-> ALARM

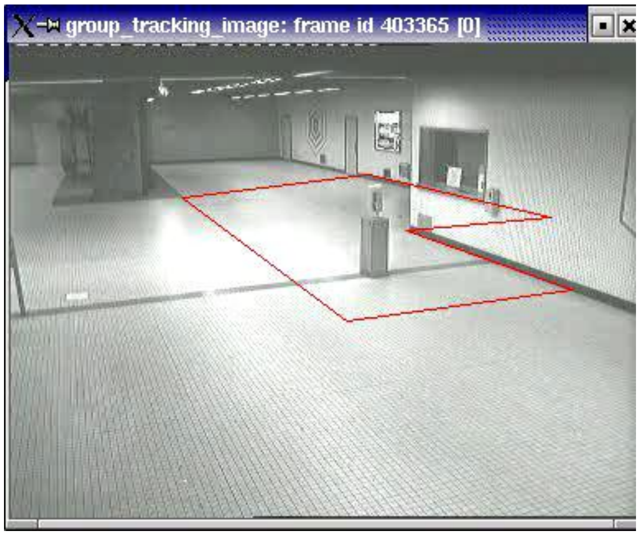
State: Lively





Event Recognition: automaton

The scenario "A Group of people blocks an Exit" is based on a Finite state automaton

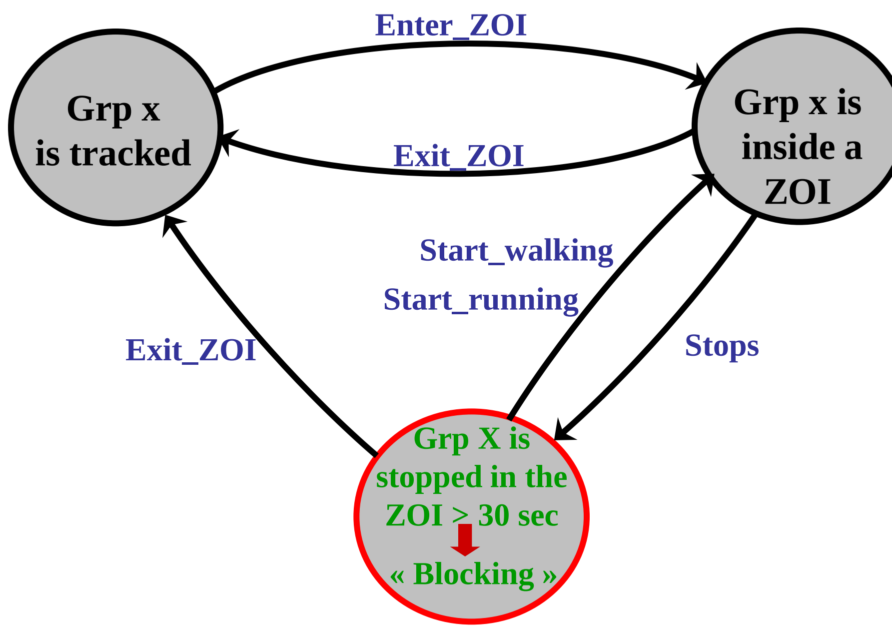


Mobile objects Detection

Group Tracking

Blocking

Recognition of the behaviour
« a Group of people
blocks an Exit»



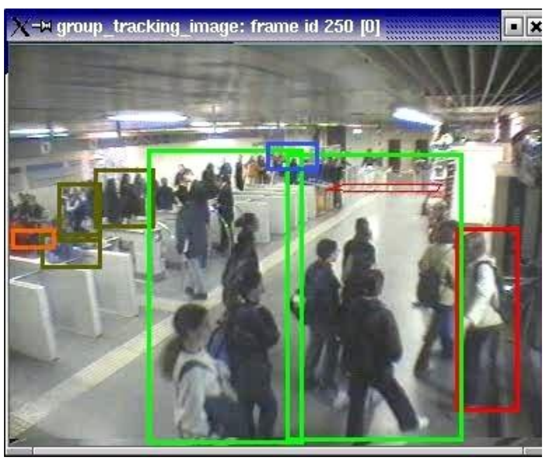
Event Recognition: Brussels and Barcelona Metros

Group behavior

Blocking

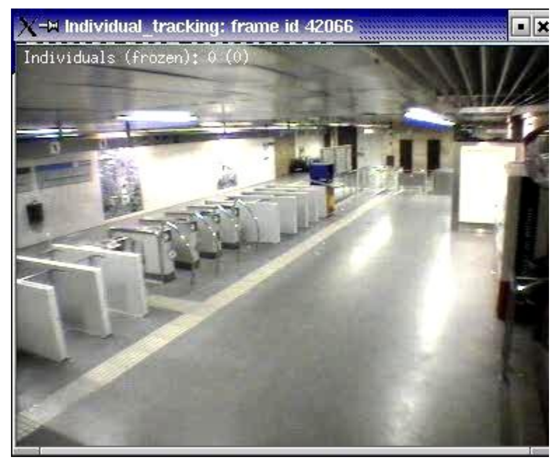
Group behavior

Fighting



Crowd behavior

Overcrowding



Individual behavior

Jumping over barrier

Event Recognition using Posture

Event Recognition : automaton

- Recognition of five behaviors "Blocking", "Fighting", "Jumping over barrier", "Vandalism" and "Overcrowding" in 2003 (FP6 Advisor).

- Tested on 50 metro sequences (10 hours) and one week live recognition

- True positive per sequence: 70% ("Fighting") to 95% ("Blocking")

- False positive per sequence: 5% ("Fighting", "Jumping over barrier") to 0% (others)

However :

- Sensitive to noise
- Difficulties to tune to get best performance

Scenario Recognition: Temporal Constraints

Work done in collaboration with T. Vu

Event Representation

Representation Language to describe Temporal Events of interest.

A video event is mainly constituted of five parts:

- Physical objects: all real world objects present in the scene observed by the cameras

  Mobile objects, contextual objects, zones of interest

- Components: list of states and sub-events involved in the event
- Forbidden Components: list of states and sub-events that must not be detected in the event
- Constraints: symbolic, logical, spatio-temporal relations between components or physical objects
- Action: a set of tasks to be performed when the event is recognized

Event Representation

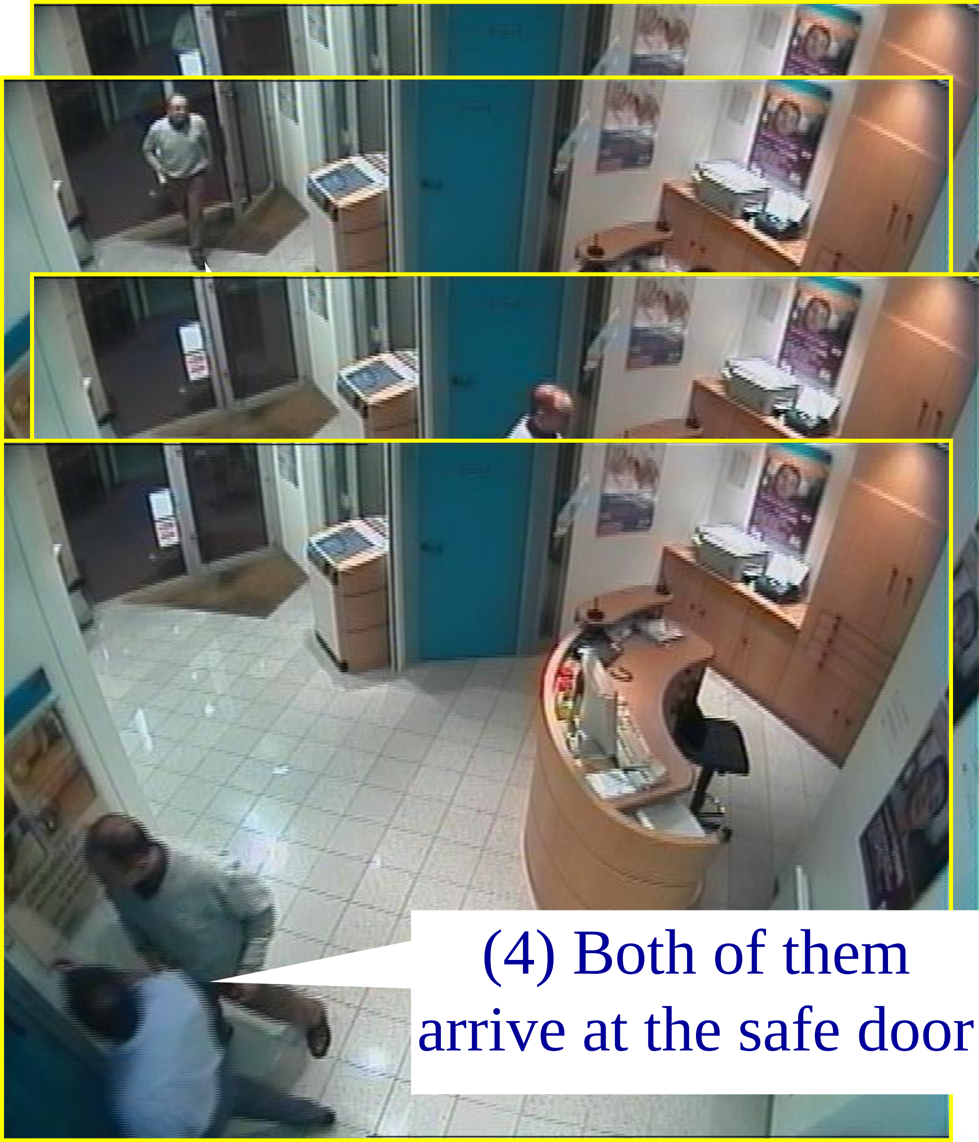
Representation Language to describe Temporal Events of interest.

Example: a “Bank_Attack” scenario model

composite-event (***Bank_attack***,
physical-objects ((employee : **Person**), (robber : **Person**))
components(
(e1 : primitive-state ***inside_zone*** (employee, "Back"))
(e2 : primitive-event ***changes_zone*** (robber, "Entrance", "Infront"))
(e3 : primitive-state ***inside_zone*** (employee, "Safe"))
(e4 : primitive-state ***inside_zone*** (robber, "Safe")))
constraints ((e2 ***during*** e1)
(e2 ***before*** e3)
(e1 ***before*** e3)
(e2 ***before*** e4)
(e4 ***during*** e3))
action (“Bank attack!!!”))

Scenario Representation

A “Bank attack” scenario instance



Scenario Recognition: Temporal Constraints

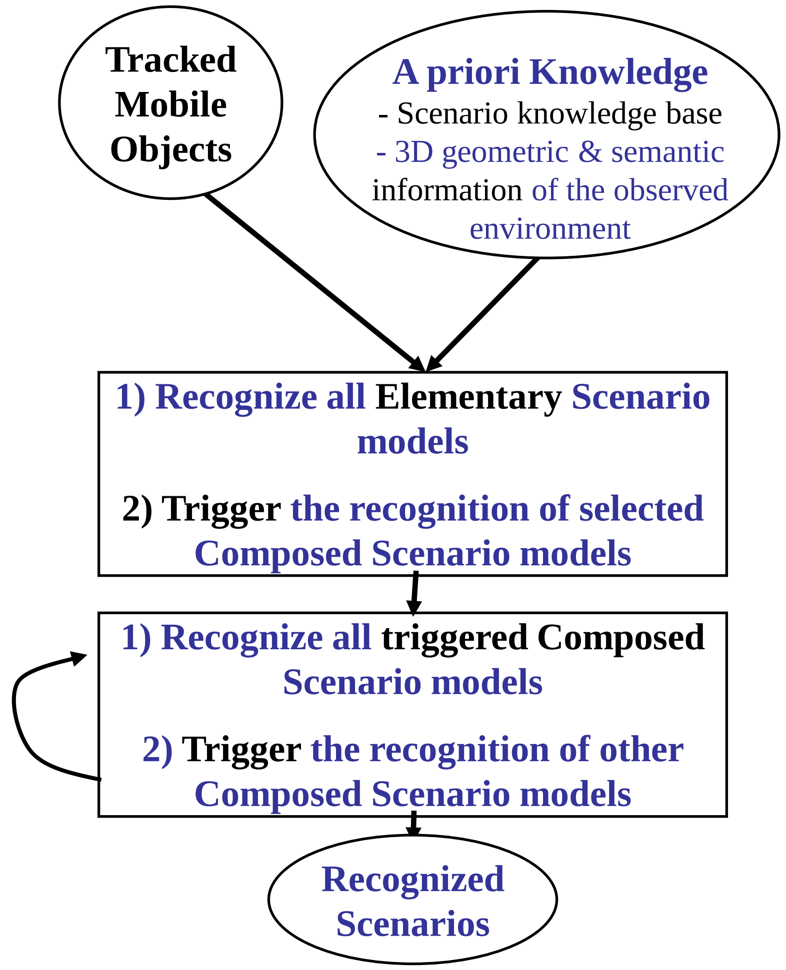
- Overview of the recognition process
- Recognition of elementary scenarios
- Scenario compilation
- Recognition of composed scenarios
- Prediction and uncertainty
- Example of the recognition of a “Bank attack” scenario and more…

Scenario Recognition: Temporal Constraints (T. Vu)

- **Scenario** (algorithmic notion): any type of video events

- Two types of scenarios:
    - **elementary** (primitive states)
    - **composed** (composite states and events).

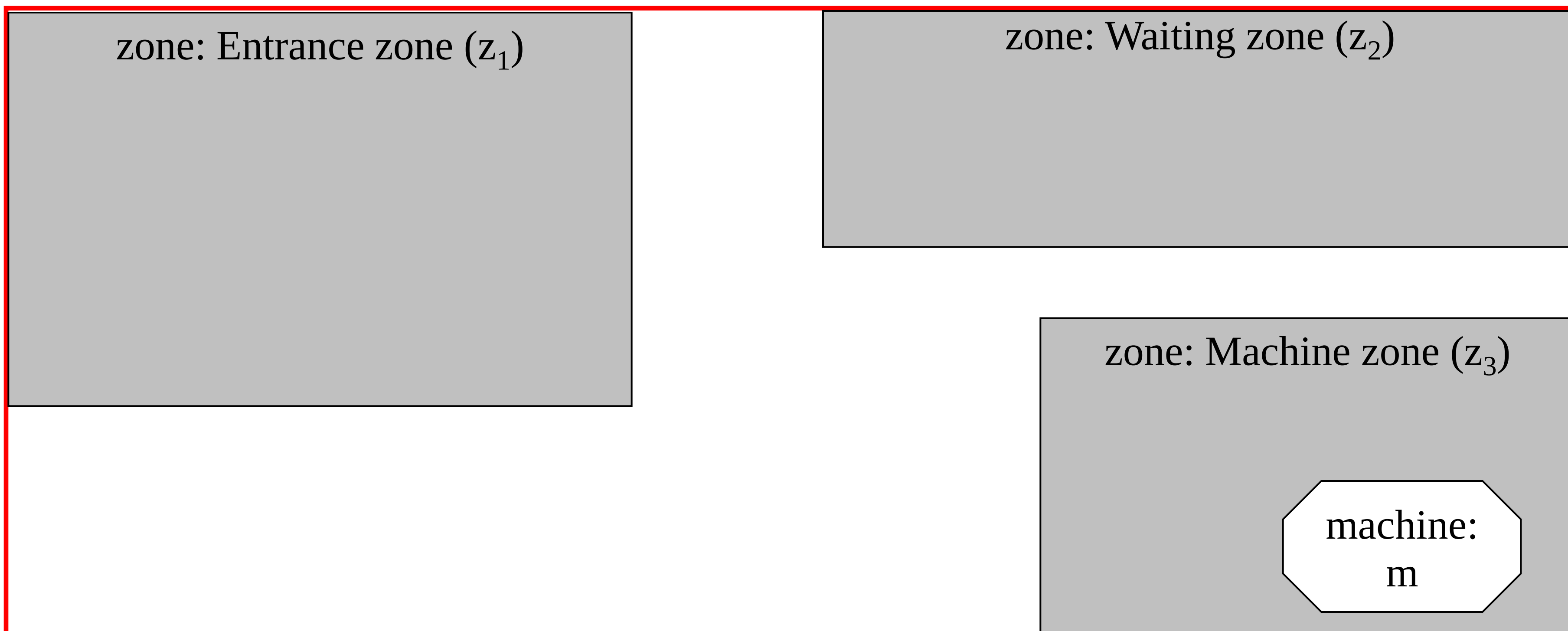
- Algorithm in two steps.



Elementary Scenario Recognition

Example: a scenario model & an observed environment

```
Scenario(Working_at_Machine,
    physical-objects(p : Person, e : Machine, z : Zone)
    constraints(
(1)         (height of p ≤ 170)
(2)         ((p in z) & (name of z = “Machine zone”))
(3)         (distance(p, e) ≤ close_distance) ) )
```



Elementary Scenario Recognition

Example: a situation

```
Scenario(Working_at_Machine,
    physical-objects(p : Person, e : Machine, z : Zone)
    constr
(1)
(2)
(3)
```

- **Problem**: [Rota, 2001] attempts all combinations of physical objects ⇒ combinatorial explosion.
- **Solution**: reorganize the knowledge represented in an elementary scenario model ⇒ **elementary scenario model compilation**.

zone: E

p_1

height = 18

height = 165

zone: Machine zone (z_3)

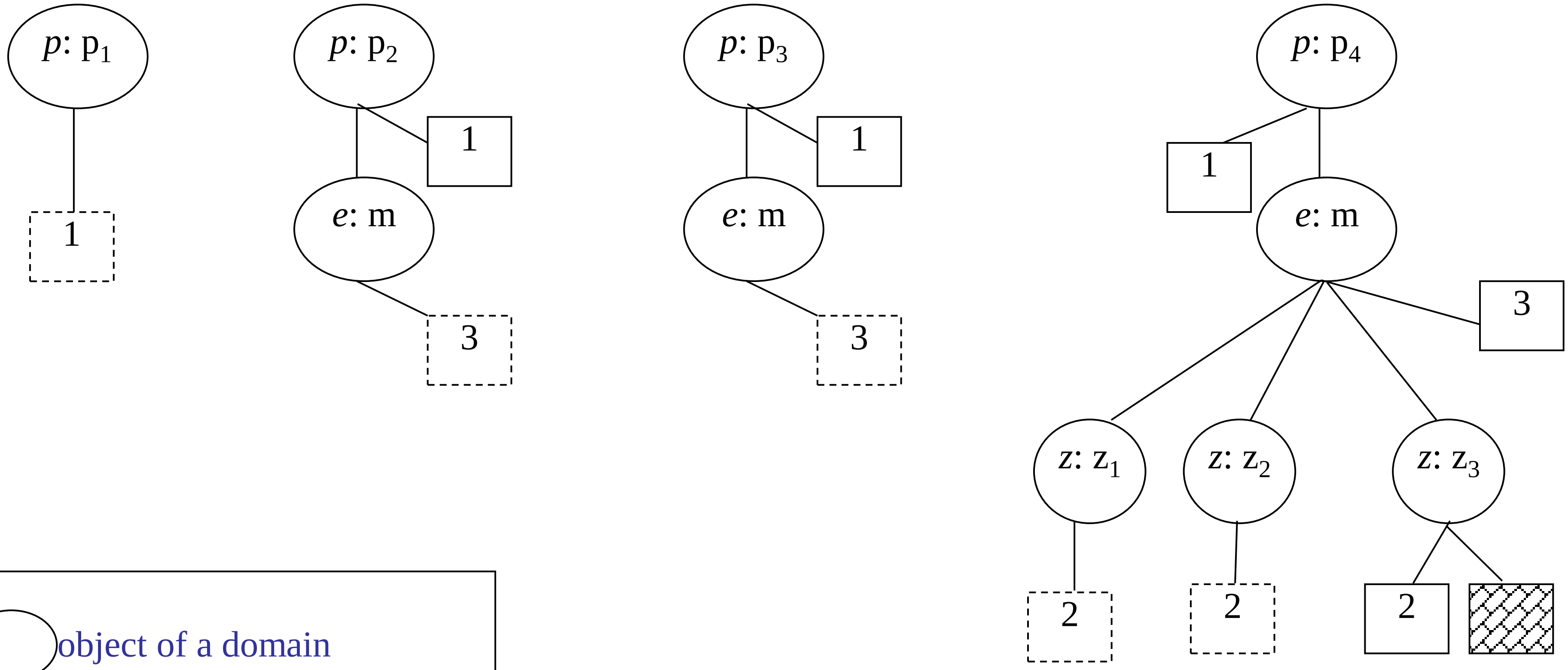
p_4

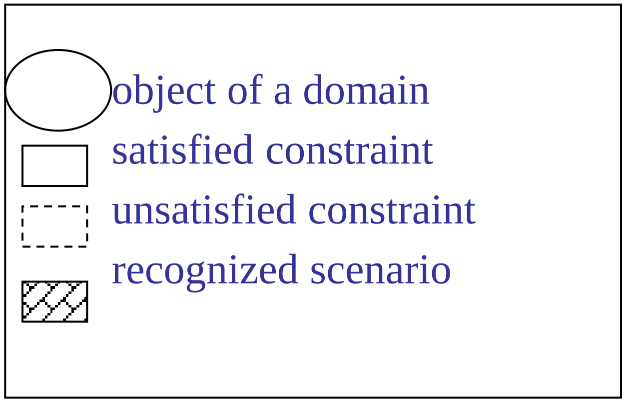
height = 170

machine: m

Recognized scenario:
Working_at_Machine(p_4, m, z_3)

Scenario Recognition: Elementary Scenario





Recognized scenario:
Working_at_Machine(p_4, m, z_3)

Scenario Recognition: Elementary Scenario

- The recognition of an **elementary scenario** model m_e consists of a loop:

1. Choosing a physical object for each physical-object variable

2. Verifying all constraints linked to this variable

m_e is recognized if all the physical-object variables are assigned a value and all the linked constraints are satisfied.

Scenario Recognition: Composed Scenario

- **Problem**:

given a scenario model $\underline{m_c = (m_1 \textit{ before } m_2 \textit{ before } m_3);}$

if a scenario instance i_3 of m_3 has been recognized

then the main scenario model m_c may be recognized.

However, the classical algorithms will try all combinations of scenario instances (already recognised) of m_1 and of m_2 with i_3

➔ a combinatorial explosion in the past.

- **Solution**:

decompose the composed scenario models into simpler scenario models in an initial (compilation) stage such as each composed scenario model is composed of two components: $\underline{m_c = (m_4 \textit{ before } m_3)}$

➔ a linear search in the past.

Scenario Recognition: Composed Scenario

Example: original “Bank_attack” scenario model

composite-event(***Bank_attack***,
  physical-objects((employee : **Person**), (robber : **Person**))
  components(
    (1) (e1 : primitive-state ***inside_zone***(employee, "Back"))
    (2) (e2 : primitive-event ***changes_zone***(robber, "Entrance", "Infront"))
    (3) (e3 : primitive-state ***inside_zone***(employee, "Safe"))
    (4) (e4 : primitive-state ***inside_zone***(robber, "Safe")))
  constraints((e2 ***during*** e1)
    (e2 ***before*** e3)
    (e1 ***before*** e3)
    (e2 ***before*** e4)
    (e4 ***during*** e3))
  alert(“Bank attack!!!”))

Scenario Recognition: Composed Scenario

Compilation: Original scenario model is decomposed into 3 new scenarios

composite-event(***Bank_attack_1***,
    physical-objects((employee : **Person**), (robber : **Person**))
    components(
    (1) (e1 : primitive-state ***inside_zone*** (employee, "Back"))
    (2) (e2 : primitive-event ***changes_zone*** (robber, "Entrance", "Infront"))
    constraints((e1 ***during*** e2)))

composite-event(***Bank_attack_2***,
    physical-objects((employee : **Person**), (robber : **Person**))
    components(
    (3) (e3 : primitive-state ***inside_zone*** (employee, "Safe"))
    (4) (e4 : primitive-state ***inside_zone*** (robber, "Safe")))
    constraints((e3 ***during*** e4)))

composite-event(***Bank_attack_3***,
    physical-objects((employee : **Person**), (robber : **Person**))
    components(
        (att_1 : composite-event ***Bank_attack_1*** (employee, robber))
        (att_2 : composite-event ***Bank_attack_2*** (employee, robber)))
    constraints(((*termination of* att_1) ***before*** (*start of* att_2)))

    alert("Bank attack!!!"))

Scenario Recognition: Composed Scenario

- A compiled scenario model m_c is composed of two components: start and termination.
- To start the recognition of m_c, its termination needs to be already instantiated.

- The recognition of a compiled scenario model m_c consists of a loop:
  1. Choosing a scenario instance for the start of m_c,
  2. Verifying the temporal constraints of m_c,
  3. Instantiating the physical-objects of m_c with physical-objects of the start and of the *termination* of m_c,
  4. Verifying the non-temporal constraints of m_c.
  5. Verifying forbidden constraints.

Scenario Recognition: Composed Scenario

	t_1	t_2	t_3	t_4
inside_zone	(P1, BC)	(P2, EZ)	(P2, IC)	(P1, Safe) (P2, Safe)
changes_zone			(P2, EZ, IC)	(P1, BC, Safe) (P2, IC, Safe)
Bank_attack_1			(P1, P2)	
Bank_attack_2				(P1, P2)
Bank_attack_3 (Bank_attack)				(P1, P2)

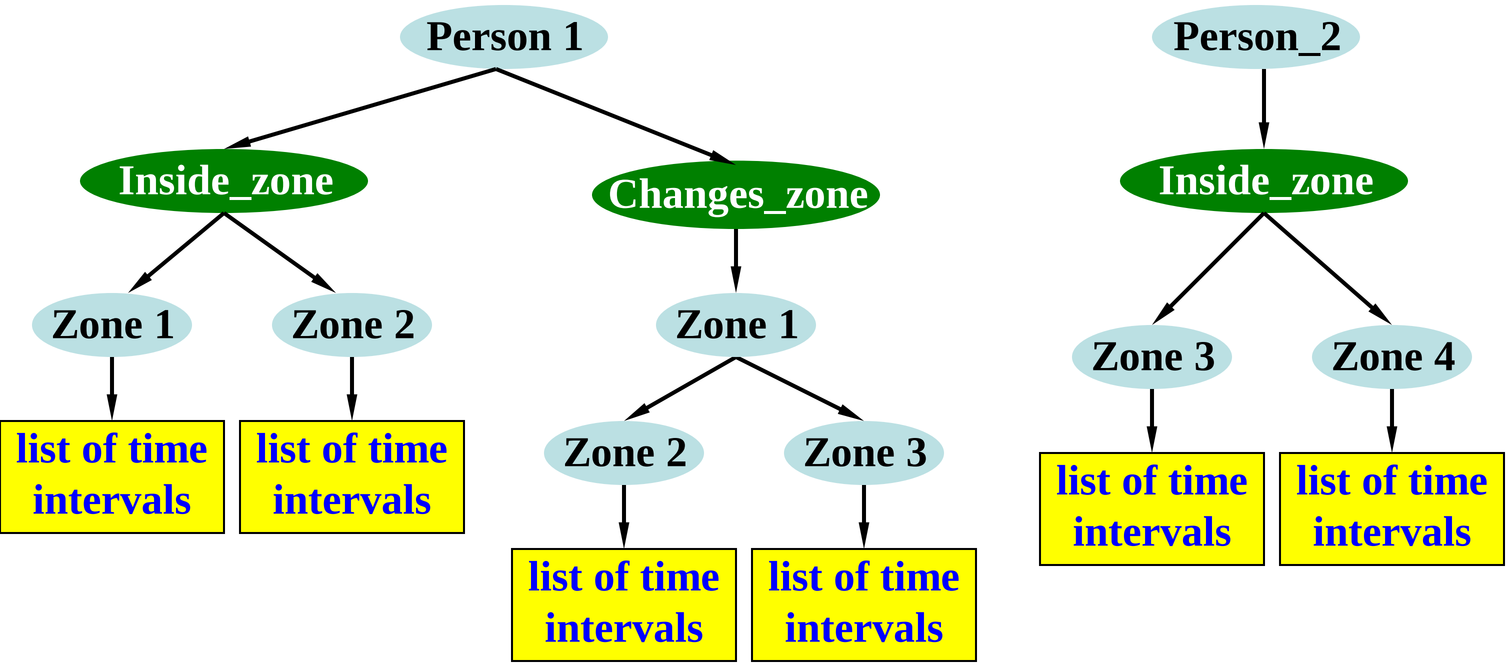
BC: Back_Counter **IC**: Infront_Counter

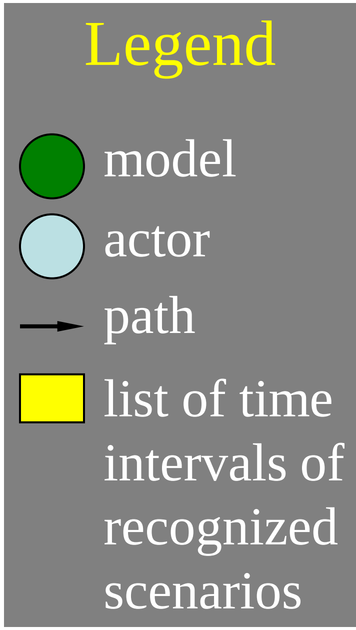
EZ: Entrance_Zone

- ♦ : the scenario instance that triggers the recognition of a composed scenario instance ended by it.
- ● the start of a composed scenario instance.

Scenario Recognition: Temporal Constraints

- The resolution of temporal constraints is improved by structuring the search domain of already recognized states, events and scenarios.





- The path (Person 1 → Inside_zone → Zone 1 →) shows the list of time intervals while Person 1 is inside Zone 1.

Scenario recognition: capacity of prediction

- Issue: in the bank monitoring application, an alert “Bank attack!!!” is triggered when a scenario “Bank_attack” is completely recognized. However, it can be too late for security agents to cope with the situation.

- Requirement: is the temporal scenario recognition method able to predict scenarios that may occur in the near future?

- Answer:
  - Yes, with some probabilities.
  - The recognition algorithm can predict scenarios that may occur by adding automatically alerts (during the compilation) to some generated partial scenario models. This task can be specified in the scenario models.

Scenario recognition : uncertainty

- Temporal precision
  - Issue: several scenario models are defined with too precise temporal constraints ⇒ they cannot be recognized with real videos.
  - Solution: we defined a temporal tolerance Δt as an integer, then all temporal comparisons are estimated using an approximation of Δt.

- Incorrect mobile object tracking
  - Issue: the vision algorithms may loose the track of several detected mobile objects ⇒ the system cannot recognize correctly scenario occurrences in several videos.
  - Solution1: experts describe different scenario models representing various situations corresponding to several combinations of physical objects.

Uncertainty Representation

Solution2: management of the vision uncertainty (likelihood):

- within predefined event models (off-line)
  - coefficients (on mobile objects and components) are provided by default.
  - Several notions of uncertainty (data, model, process) and utility.

- propagated (on-line) through the event instances
  1. mobile objects: computed by vision algorithms.
  2. primitive states (elementary):
     - a coefficient to each physical object for representing the likelihood relation between the state and each involved mobile object.
  3. events and composite states (composed):
     - a coefficient to each component for representing the likelihood relation between the event and each component.
       - defining a threshold into each state/event model for specifying at which likelihood level the given state/event should be recognized.

Uncertainty Representation

Combination of detection likelihood / confidence and utility:

PrimitiveState (**Person_Close_To_Vehicle**,
Physical Objects ((p : Person, 0.7), (v : Vehicle, 0.3))
Constraints ((p distance v $\leq$ close_distance)
(recognized if likelihood > 0.8)))

CompositeEvent (**Crowd_Splits**,
Physical Objects ((c1: Crowd, 0.5), (c2 : Crowd, 0.5), (z1: Zone))
Components ((s1 : CompositeState Move_toward (c1, z1), 0.3)
(e2 : CompositeEvent Move_away (c2, c1), 0.7))
Constraints ((e2 during s1)
(c2's Size > Threshold)
(recognized if likelihood > 0.8)))

Scenario recognition: Results

Evaluation: the experts of 20 projects in video interpretation have realized three types of tests.

- on recorded videos: to verify whether the recognition algorithm can recognize effectively scenario occurrences (correct detections).
- on live videos: to verify whether the recognition algorithm can work on a longtime interval (no false alarms).
- on recorded/simulated videos: to estimate the processing time and efficiently of the recognition algorithm.

Scenario recognition: Results
Experiment 1: recorded videos

- many sites: 2 bank agencies, several metro stations, a train and an airport…
- Bank : 27 recorded positive videos and many negative videos.
- 40 original scenario models (before the compilation): “inside_zone”, “Bank_attack”, “Vandalism”,...

	Number of tested sequences	Average number of persons/frame	Recognition rate (%)	Number of false alarms
Bank 1	10	4	80	0
Bank 2	1	2	100	0
Metros	4	2	100	0
Apron	10	1	100	0
Train	2	4	100	0

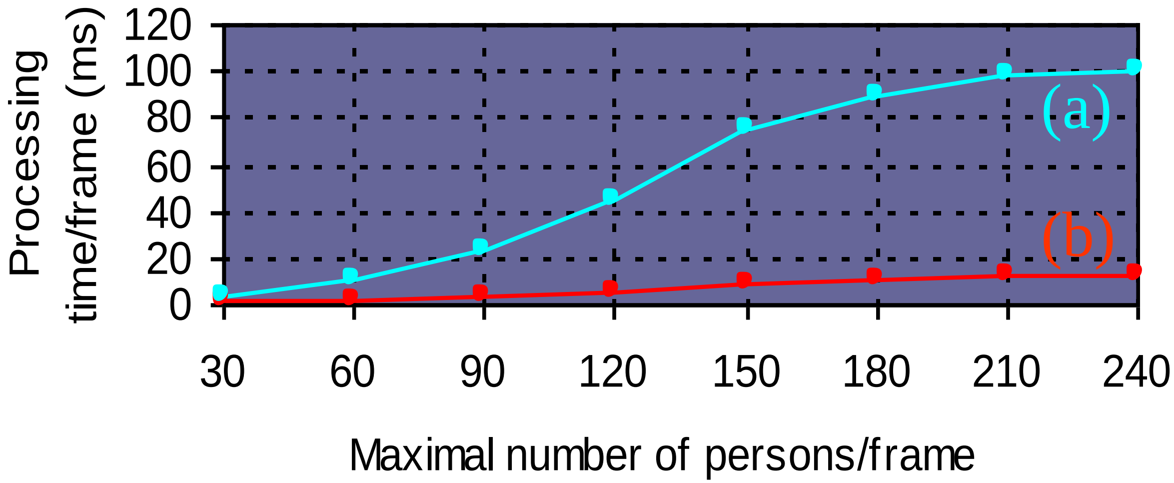
- The algorithm fails to recognize some scenario occurrences only when the vision module fails to detect the mobile objects in the scene.
- No false alarm has been reported during all the experiments.

Scenario recognition: Results
Experiment 2: live-videos

- 4 sites: 2 bank agencies, two offices, a parking and a metro station.
- 40 original scenario models (before decomposition): “inside_zone”, “Bank_attack”, “Vandalism”,...
- Results:
  - in a bank (5 days),
  - in an office (4h),
  - one week in a metro station of Barcelona,
  - in a parking (1 day)
  - the scenarios were most of the time (95%) correctly recognized (as in the first experiment) ➔ the recognition algorithm can work reliably and robustly in real-time and in a continuous mode.

Scenario recognition: Results
Experiment 3: checking the processing time

60 scenario models defined with 2 to 10 physical object variables and 2 to 10 components. The algorithms are tested on simulated videos containing up to 240 persons in the scene.



The (b) **average** and (a) **maximal** processing time/frame of the algorithm.

The composed scenario recognition algorithm is able to process up to 240 persons in the scene.

Scenario Recognition: Temporal Constraints Results

- A generic formalism to help experts model intuitively states, events and scenarios.

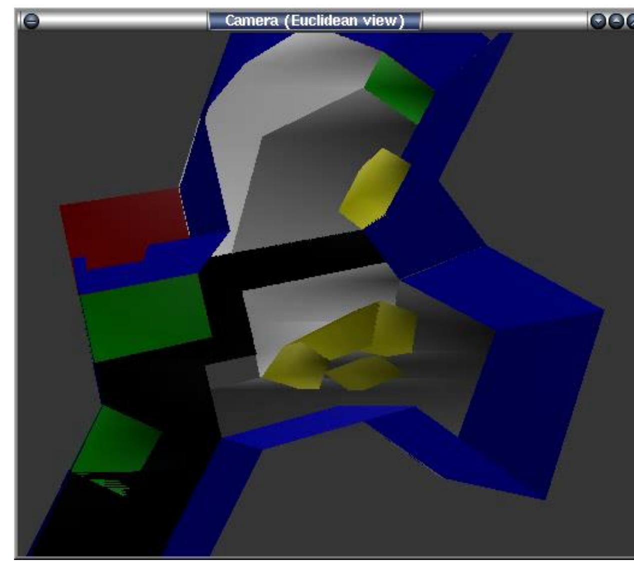
- Recognition algorithm processes temporal operators in an efficient way.
  - Linear search in the past.

- The recognition of complex scenarios (large number of actors) becomes real time.
  - Indexed Trees to structure and access the already recognized scenarios

- However,
  - uncertainty needs to be taken care
  - Scenario modeling is not always easy

Scenario recognition: Results

Bank agency monitoring in Paris (M. Maziere)



Scenario recognition: Results

Vandalism scenario example (temporal constraints) :

Scenario(vandalism_against_ticket_machine,

Physical_objects((p : **Person**), (eq : **Equipment**, ***Name***="Ticket_Machine"))

Components ((event s1: *p* **moves_close_to** *eq*)

(state s2: *p* **stays_at** *eq*)

(event s3: *p* **moves_away_from** *eq*)

(event s4: *p* **moves_close_to** *eq*)

(state s5: *p* **stays_at** *eq*))

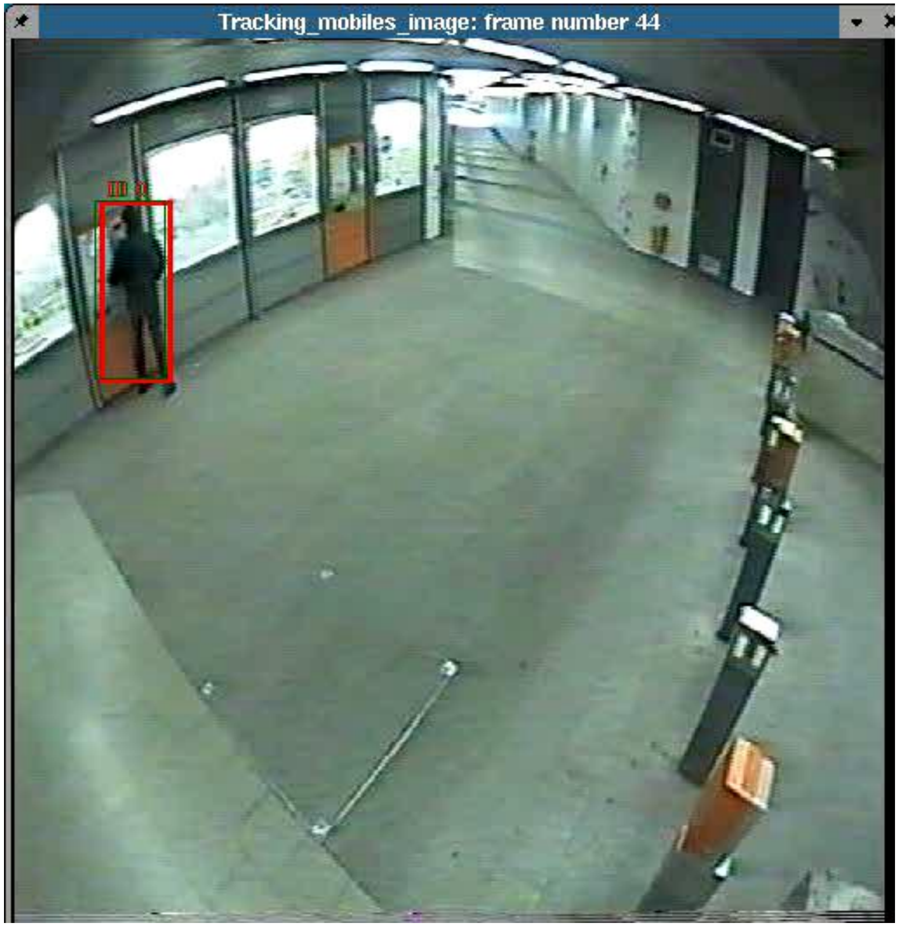
Constraints ((s1 **!=** s4) (s2 **!=** s5)

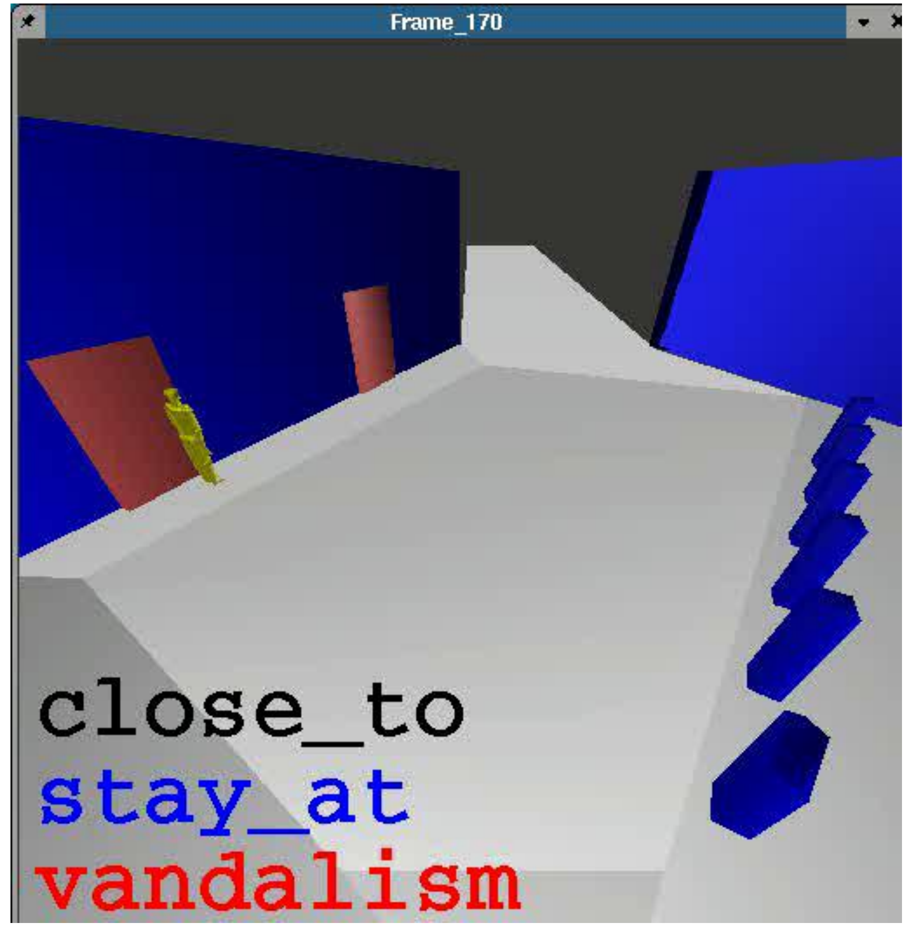
(s1 **before** s2) (s2 **before** s3)

(s3 **before** s4) (s4 **before** s5))))

Scenario Recognition: Results

Vandalism in metro in Nuremberg





Scenario recognition: Results

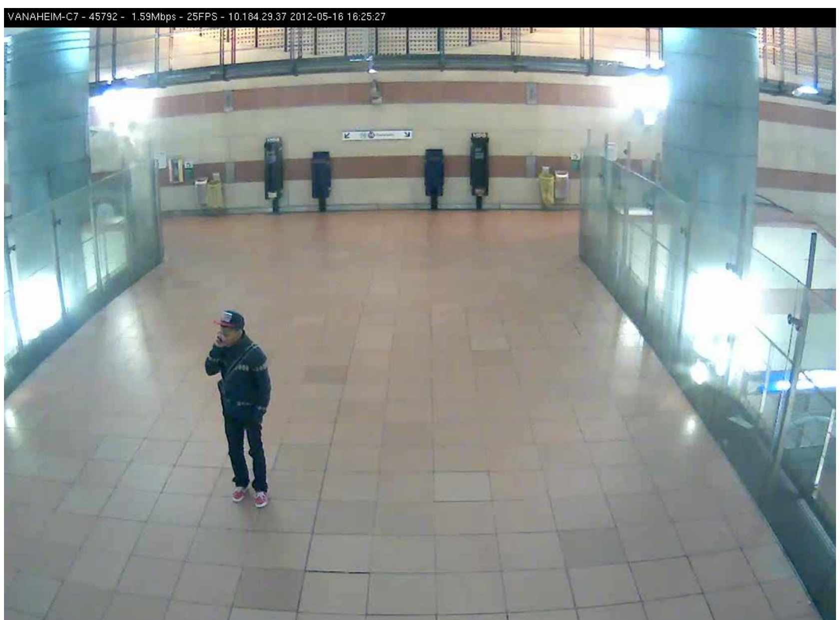
Example: a "Vandalism against a ticket machine" scenario

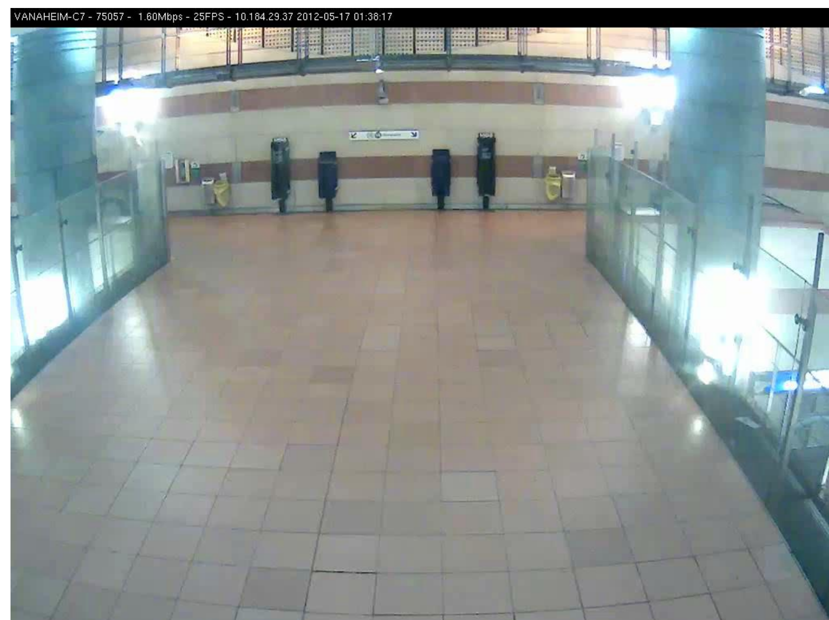
Group Scenario Detection – Paris subway

Waiting example - Erratic group example

```
PrimitiveState(in_WaitingZone2,
    PhysicalObjects((g1 : Group), (z1 : Zone))
    Constraints ((g1->Position in z1->Vertices)
    (z1->Name = WaitingZone2)
    )
    Alarm ((Level : NOTURGENT))
)
```

```
CompositeState(Erratic_Group,
  PhysicalObjects((g1 : Group))
  Components((c1 : PrimitiveState Erratic(g1)))
  Constraints((duration(c1) >= MIN_TIME_ERRATIC) )
  Alarm((Level : URGENT))
)
```





Scenario recognition:

Results Example: "Unloading Front Operation " event

- Example of the **Unloading Front Operation** (global)

CompositeEvent (UnLoading_Front_Global_Operation,
PhysicalObjects ((v1 : Vehicle), (v2 : Vehicle),
(z1 : Zone), (z2 : Zone), (z3 :Zone))
Components ((c1 : **CompositeEvent** Loader_Arrival(v1, z1, z2))
(c2 : **CompositeEvent** Transporter_Arrival(v2, z1, z3))
Constraints ((v1->SubType = LOADER)
(v2->SubType = TRANSPORTER)
(z1->Name = ERA)
(z2->Name = RF_DoorC_Access)
(z3->Name = LOADER_BackZone)
(c1 before c2)))

Scenario recognition: Results

Example: "Unloading Global Operation" event

- "Unloading Global Operation"

Scenario recognition: Results

Example: "Unloading Front Operation " event

- Example of the **Unloading Front Operation** (detailed)

CompositeEvent (UnLoading_Front_Detailed_Operation,
PhysicalObjects ((p1 : Person), (v1 : Vehicle), (v2 : Vehicle), (v3 : Vehicle),
(z1 : Zone), (z2 : Zone), (z3 :Zone), (z4 : Zone))
Components ((c1 : **CompositeEvent** Loader_Arrival(v1, z1, z2))
(c2 : **CompositeEvent** Transporter_Arrival(v2, z1, z3))
(c3 : **CompositeState** Worker_Manipulating_Container(p1, v3, v2, z3, z4)))
Constraints ((v1->SubType = LOADER)
(v2->SubType = TRANSPORTER)
(z1->Name = ERA) (z2->Name = RF_DoorC_Access)
(z3->Name = LOADER_BackZone)
(z4->Name = Behind_RF_DoorC_Access)
(c1 before c2)
(c2 before c3)))

Scenario recognition: Results

Parked aircraft monitoring in Toulouse (F Fusier)

- "Unloading Front Operation"

Scenario recognition: Results

Example: “Aircraft Arrival Preparation ” event

- Aircraft Arrival Preparation (involving the GPU)

SCENARIO AIRCRAFT_ARRIVAL_PREPARATION_SCENARIOS

Vehicle: GPU

Person: Handler

Zones: ERA, GPU_Access, Arrival_Preparation

Dynamic Zone: GPU_Door

Vehicle_Arrived_In_ERA

Gpu_Enters_Gpu_Access_Area

Gpu_Stopped_In_Gpu_Access_Area

Handler_Gets_Out_Gpu

Handler_From_Gpu_Deposites_Chocks_Or_Stud

Scenario recognition: Results

Example: "Tow Tractor Arrival" event

- Tow Tractor Arrival

Scenario recognition: Results

Example: “vandalism_against_window” event

CompositeEvent(***vandalism_against_window***,

PhysicalObjects((vandal : Person)), (w : Equipment))

Components((vandalism_against_window_VIDEO :
CompositeEvent vandal_close_to_window(vandal, w))
(vandalism_against_window_AUDIO :
CompositeEvent tag_detected_close_to_person(vandal)))

Constraints((vandalism_against_window_VIDEO ***during***
vandalism_against_window_AUDIO))

Alarm(AText("Vandalism against window")
AType("URGENT")))

Scenario recognition: Results

Example: "Scratch & theft in a train" scenarios

Scenario recognition: Results

Example: a "Disturbing people in a train" scenario

1^{st} experiment : Multi-sensor Scenario recognition

Example of “Taking meal” event model

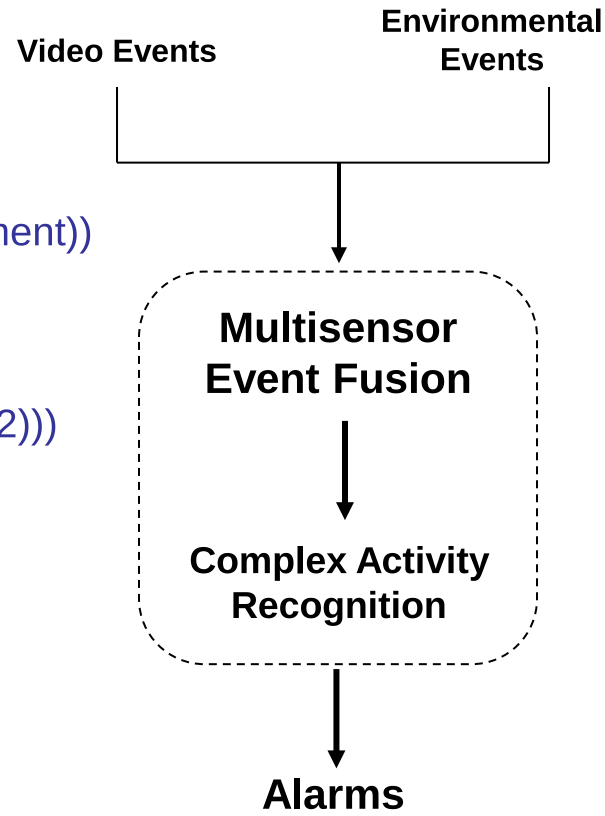
CompositeEvent (**M_TakingMeal**,

PhysicalObjects ((p : Person), (z1 : Zone), (z2 : Zone), (eq1 : Equipment))

Components ((c1 : **PrimitiveState** V_InLivingroom(p, z1))
(c2 : **PrimitiveState** V_CloseToTable(p, eq1))
(c3 : **CompositeState** M_PersonSittingAtDinningtable(p, z2)))

Constraints ((z1’s **Name** = Livingroom),
(z2’s **Name** = Dinningtable)
(eq1’s **Name** = table),
(c2 **Duration** >= threshold1),
(c2 **During** c1),
(c3 **During** c2),
(c3 **Duration** >= threshold2))

Alert (”Person is taking a meal”, ”NOTURGENT”)



Multi-sensor Scenario recognition

- **Language combining multi-sensor information**

Activity (Use Fridge,

Physical Objects ((p: Person), (Fridge: Equipment), (Kitchen: Zone))

Components ((c1: Inside zone (p, Kitchen))

(c2: Close_to (p, Fridge))

(c3: Bending (p)

(c4: Opening (Fridge))

(c5: Closing (Fridge)))

Detected by video camera

Detected by contact sensor

Constraints ((c1 before c2)

(c3 during c2)

(c4:time + 10s < c5:time)))

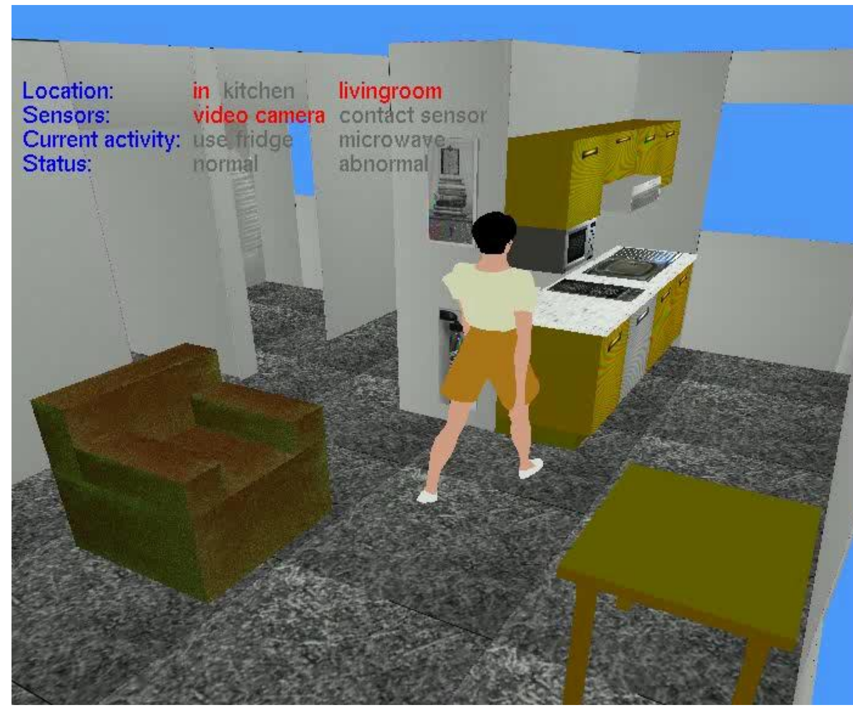
Recognition of the "Prepare meal" event

- The person is recognized with the posture "standing with one arm up", "located in the kitchen" and "using the microwave".

Visualization of a recognized event in the Gerhome laboratory

Recognition of the "Resting in living-room" event

- The person is recognized with the posture "sitting in the armchair" and "located in the living-room".



Visualization of a recognized event in the Gerhome laboratory

Event recognition results

- 14 elderly volunteers have been monitored during 4 hours (total: more than 56 hours).
- Recognition of the “Prepare meal” event for a 65 old man

Event recognition results

- Recognition of the “Having meal” event for a 84 old woman

Discussion about the obtained results

+ Results of recognition of 6 daily activities for 5*4=20 hours

Activity	GT	TP	FN	FP	Precision	Sensitivity
Use fridge	65	54	11	9	86%	83%
Use stove	177	165	11	15	**92%**	**94%**
Sitting on chair	66	54	12	15	78%	82%
Sitting on armchair	56	49	8	12	80%	86%
Prepare lunch	5	4	1	3	**57%**	**80%**
Wash dishes	16	13	3	7	65%	81%

- Errors occur at the border between living-room and kitchen
- Mixed postures such as bending and sitting due to segmentation errors

Discussion about the obtained results

+ Good recognition of a set of activities and human postures (video cameras)

Activity	GT	TP	FN	FP	Precision	Sensitivity
Use fridge	65	54	11	9	86%	83%
Use stove	177	165	11		92%	94%
Sitting on chair	66	54	12	15	78%	82%
Sitting on armchair	56	49	8	12	80%	86%
Prepare lunch	5	4	1	3	57%	80%
Wash dishes	16	13	3	7	65%	81%

Bag on chair

2 instances of the event

Cold meal

- Errors occur at the border between living-room and kitchen
- Mixed postures such as bending and sitting due to segmentation errors

Recognition of a set of activities comparing two elderly people

Activity	Used sensor (s)	Elderly people 1 (64 years) Activity duration (min:sec) Mean (m1)	Elderly people 1 (64 years) Activity duration (min:sec) Total	Elderly people 1 (64 years) Nb inst (n1)	Elderly people 2 (85 years) Activity duration (min:sec) Mean (m2)	Elderly people 2 (85 years) Activity duration (min:sec) Total	Elderly people 2 (85 years) Nb inst (n2)	Normalized Difference NDA= |m1-m2|/(m1+m2)	Normalized Difference NDI= |n1-n2| / (n1+n2)
Use fridge	Video + contact	**0:12**	2:50	**14**	**0:13**	1:09	**5**	**4 %**	**47 %**
Use stove	Video + power	**0:08**	4:52	**35**	**0:16**	27:57	**102**	**33 %**	**49 %**
Use upper-cupboard	Video + contact	**0:51**	21:34	**25**	**4:42**	42:24	**9**	**69 %**	**47 %**
Sitting on chair	Video + pressure	**6:07**	73:27	**12**	**92:42**	185:25	**2**	**87 %**	**71 %**
Entering the living-room	Video	**1:25**	25:00	20	**2:38**	35:00	13	**30 %**	21 %
Standing	Video	0:09	30:00	**200**	0:16	12:00	**45**	28 %	**63 %**
Bending	Video	**0:04**	2:00	**30**	**0:20**	5:00	**15**	**67 %**	**33 %**

Table 2: Monitored activities, their frequency (n1 & n2), mean duration (m1 & m2) and total duration for 2 volunteers staying in the GERHOME laboratory for 4 hours; NDA=Normalized Difference of mean durations of Activities=|m1-m2|/ (m1+m2); NDI=Normalized Difference of Instances number=|n1-n2|/(n1+n2); possible differences in behavior of the 2 volunteers are signified in bold

Recognition of a set of activities comparing two elderly people

Activity	Used sensor (s)	Elderly people 1 (64 years) Activity duration (min:sec) Mean (m1)	Total	Nb inst (n1)	Elderly people 2 (85 years) Activity duration (min:sec) Mean (m2)	Total	Nb inst (n2)	Normalized Difference NDA= |m1-m2|/(m1+m2)	NDI= |n1-n2| / (n1+n2)
Use fridge	Video + contact	**0:12**	2:50	**14**	**0:13**	1:09	**5**	**4 %**	**47 %**
Use stove	Video + power	**0:08**	4:52	**35**	**0:16**	27:57	**102**	**33 %**	**49 %**
Use upper-cupboard	Video + contact	**0:51**	21:34	**25**	**4:42**	42:24	**9**	**69 %**	**47 %**
Sitting on chair	Video + pressure	**6:07**	73:27	**12**	**92:42**	185:25	**2**	**87 %**	**71 %**
Entering the living-room	Video	**1:25**	25:00	20	**2:38**	35:00	13	**30 %**	21 %
Standing	Video	0:09	30:00	**200**	0:16	12:00	**45**	28 %	**63 %**
Bending	Video	**0:04**	2:00	**30**	**0:20**	5:00	**15**	**67 %**	**33 %**

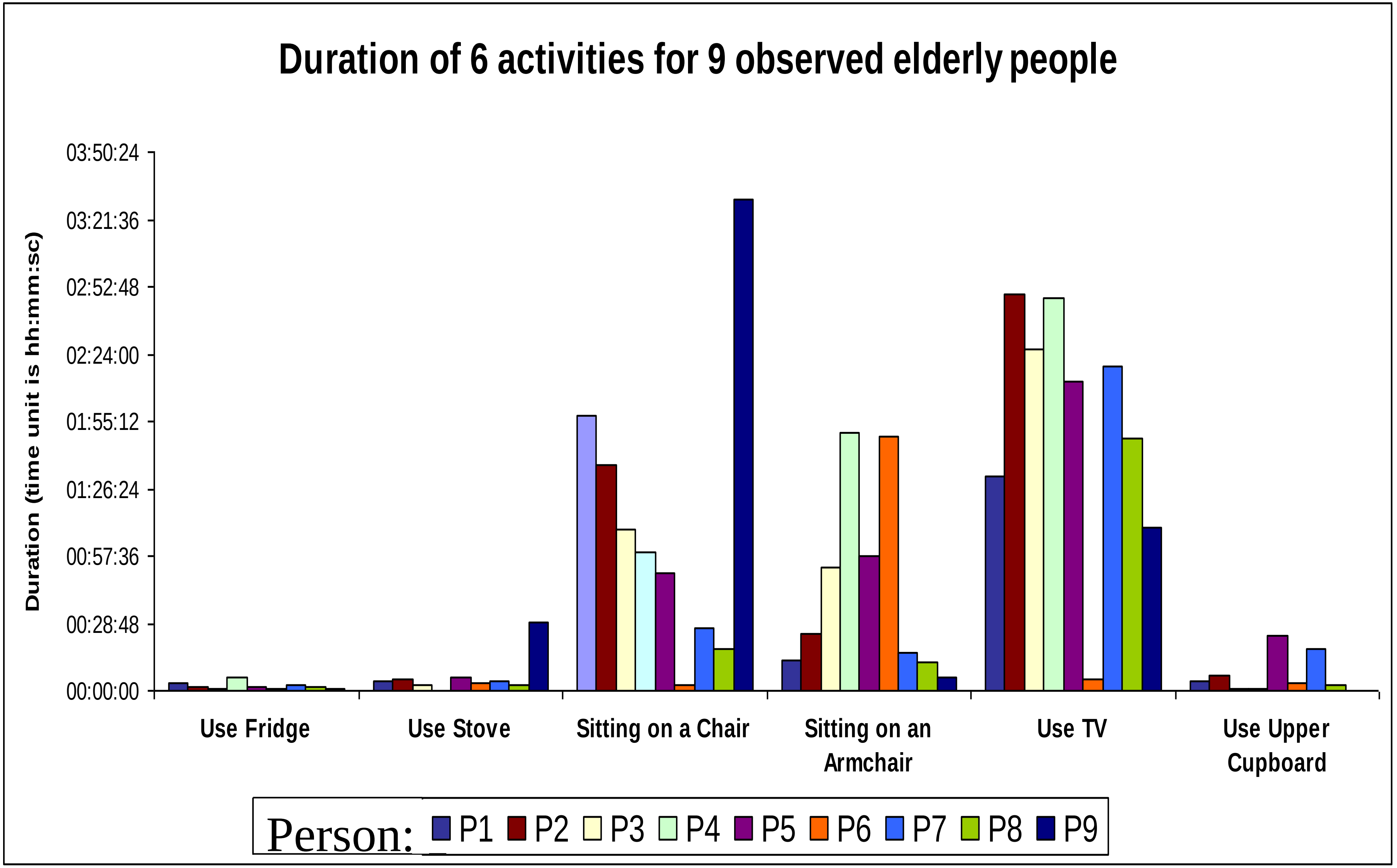
Table 2: Monitored activities, their frequency (n1 & n2), mean duration (m1 & m2) and total duration for 2 volunteers staying in the GERHOME laboratory for 4 hours; NDA=Normalized Difference of mean durations of Activities=|m1-m2|/ (m1+m2); NDI=Normalized Difference of Instances number=|n1-n2|/(n1+n2); possible differences in behavior of the 2 volunteers are signified in bold

Recognition of a set of activities comparing two elderly people

Activity	Used sensor (s)	Elderly people 1 (64 years) Activity duration (min:sec) Mean (m1)	Elderly people 1 (64 years) Activity duration (min:sec) Total	Elderly people 1 (64 years) Nb inst (n1)	Elderly people 2 (85 years) Activity duration (min:sec) Mean (m2)	Elderly people 2 (85 years) Activity duration (min:sec) Total	Elderly people 2 (85 years) Nb inst (n2)	Normalized Difference NDA= |m1-m2|/ (m1+m2)	Normalized Difference NDI= |n1-n2| / (n1+n2)
Use fridge	Video + contact	**0:12**	2:50	**14**	**0:13**	1:09	**5**	**4 %**	**47 %**
Use stove	Video + power	**0:08**	4:52	**35**	**0:16**	27:57	**102**	**33 %**	**49 %**
Use upper-cupboard	Video + contact	**0:51**	21:34	**25**	**4:42**	42:24	**9**	**69 %**	**47 %**
Sitting on chair	Video + pressure	**6:07**	73:27	**12**	**92:42**	185:25	**2**	**87 %**	**71 %**
Entering the living-room	Video	**1:25**	25:00	20	**2:38**	35:00	13	**30 %**	21 %
Standing	Video	0:09	30:00	**200**	0:16	12:00	**45**	28 %	**63 %**
Bending	Video	**0:04**	2:00	**30**	**0:20**	5:00	**15**	**67 %**	**33 %**

Table 2: Monitored activities, their frequency (n1 & n2), mean duration (m1 & m2) and total duration for 2 volunteers staying in the GERHOME laboratory for 4 hours; NDA=Normalized Difference of mean durations of Activities=|m1-m2|/ (m1+m2); NDI=Normalized Difference of Instances number=|n1-n2|/(n1+n2); possible differences in behavior of the 2 volunteers are signified in bold

Evaluation and results



2nd experiment : CMRR in Nice Hospital
Screening of AD patients

- **Medical staff & healthy younger**
  - 22 people (more female than male)
  - Age: ~ 25-35 years
  - Medical staff
  - 1 video camera, Actiwach

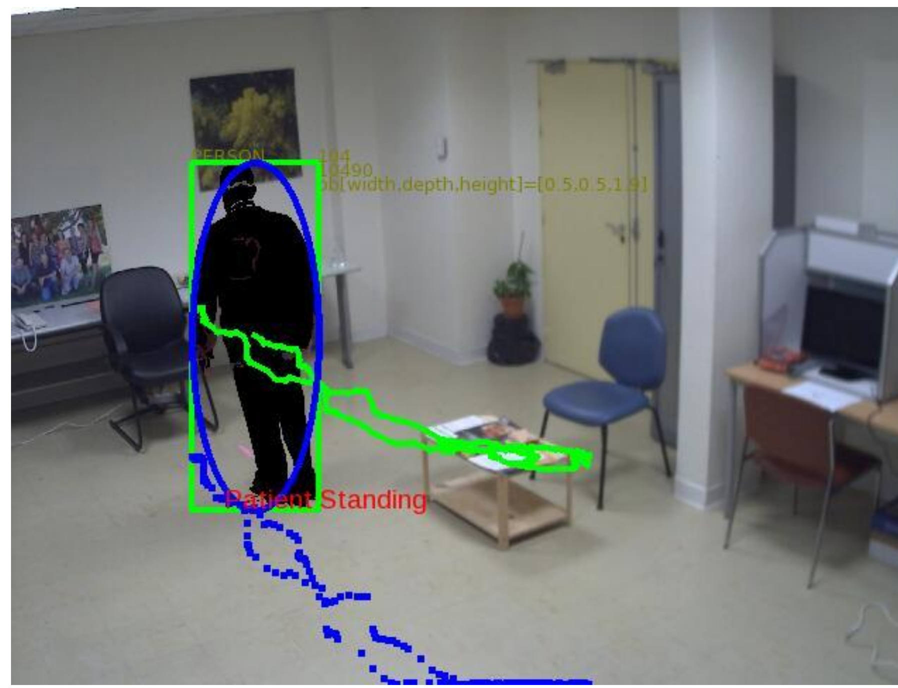
- **Older persons**
  - 20 (woman & man)
  - Age: ~ 60-85 years
  - 2 video cameras
  - Actiwach/ motionPod

- **Alzheimer patients**:
  - 21 AD people (woman & man)
  - 19 MCI (mild cognitive impairment) and mixed
  - Age: ~ 60-85 years
  - 2 video cameras
  - Actiwach/ motionPod

Reconnaissance d’un protocole au CM2R - CoBTeK

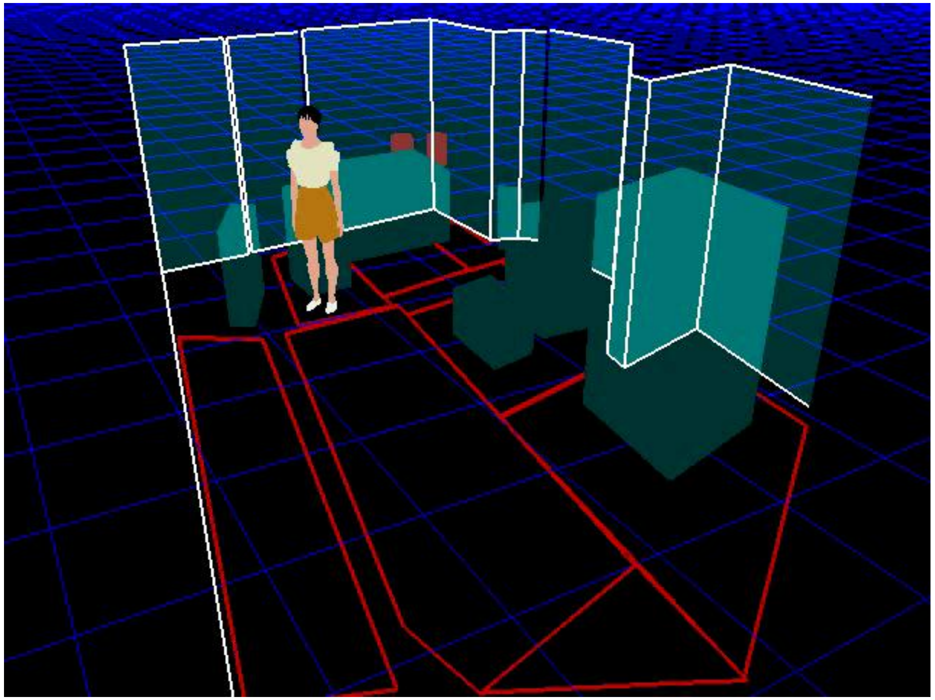
Centre Mémoire de Ressources et de Recherche du CHU Nice

Reconnaissance de l’activité “stop and go” et «sit down» en utilisant le capteur vidéo au CM2R.



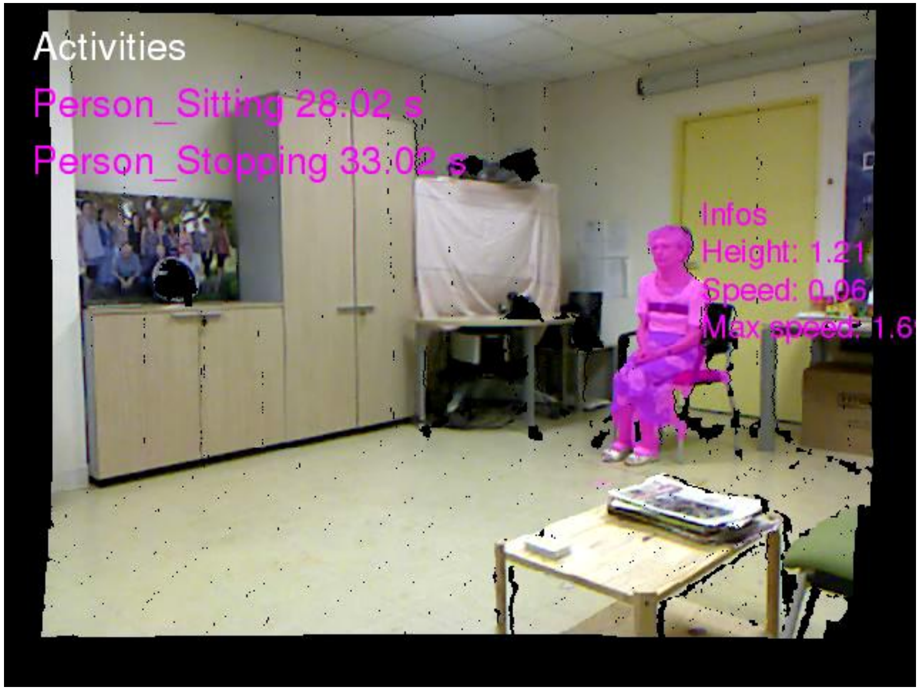
Activity monitoring in Nice Hospital with AD patients

Recognition of the "stand-up" activity.



Activity monitoring in Nice Hospital with AD patients

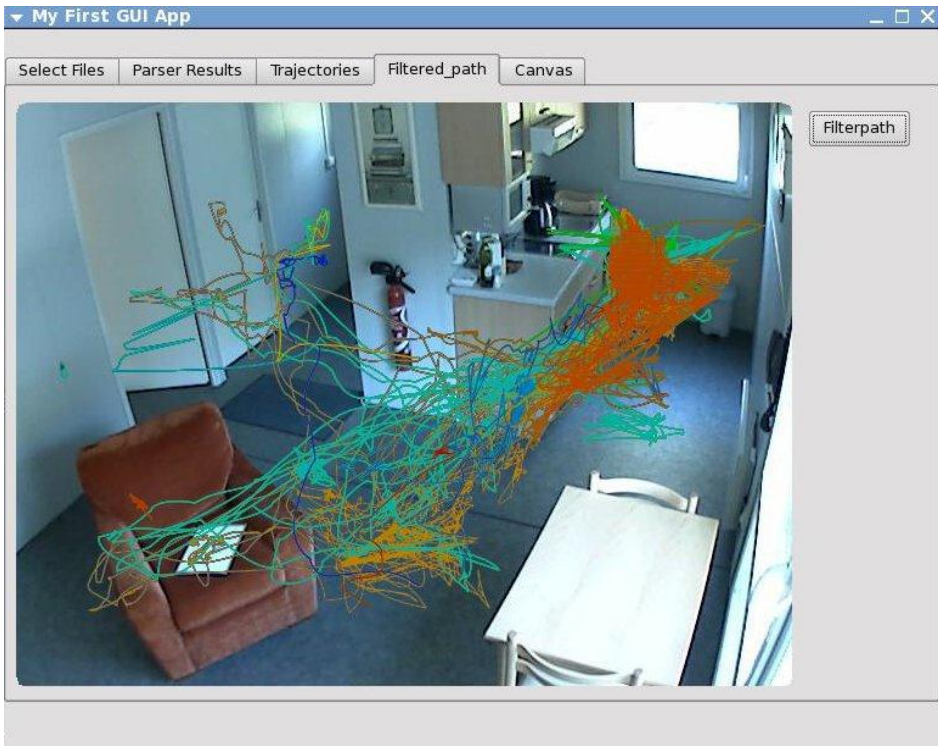
Recognition of the "stand-up & walking" activity.



Inria informatics mathematics

Learning Scenario Models : scene model
(G. Pusiol)

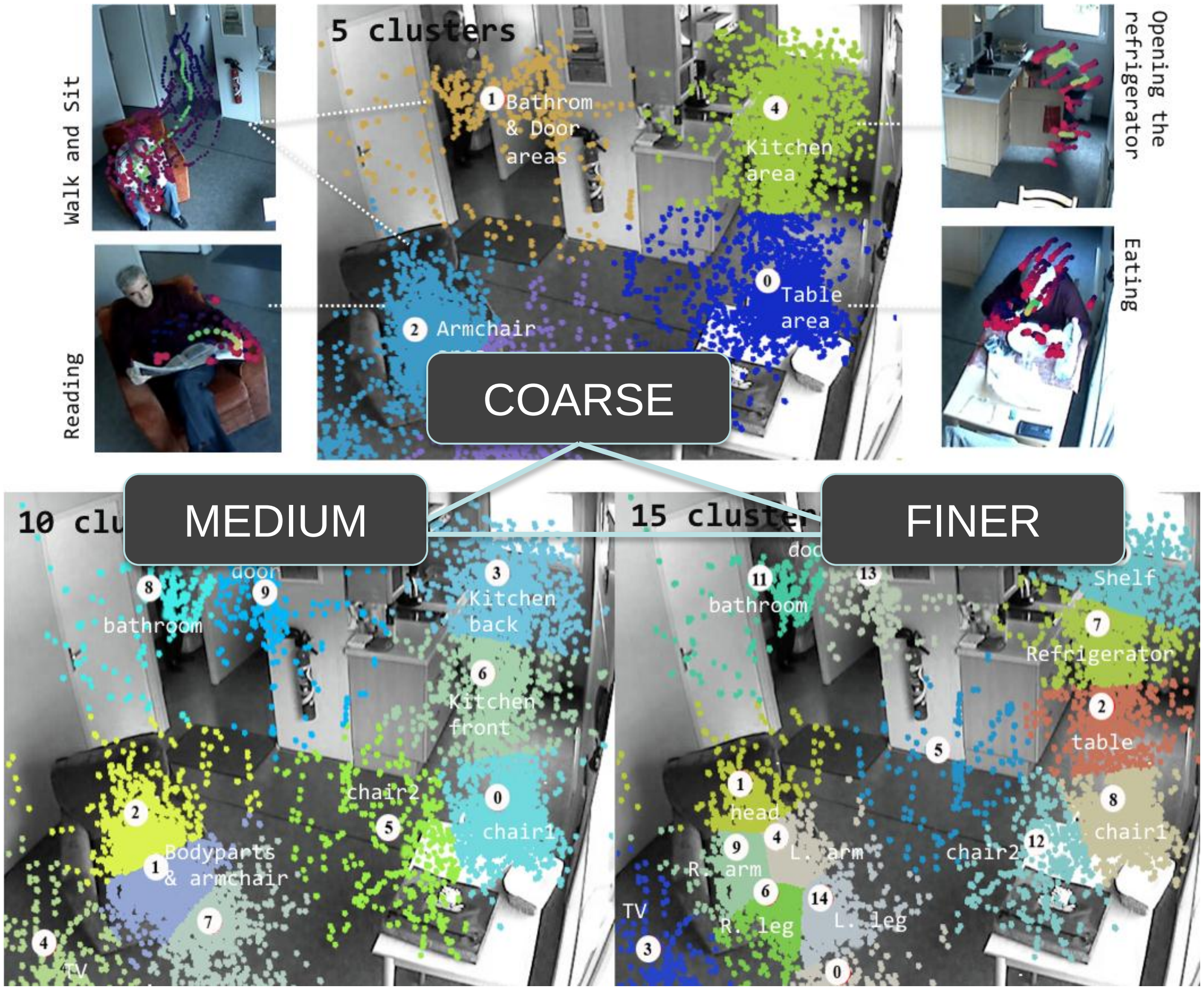
Localization of the person during 4 observation hours



Walked distance = 3.71 km

Stationery positions of the person

Learning Scenario Models : scene model

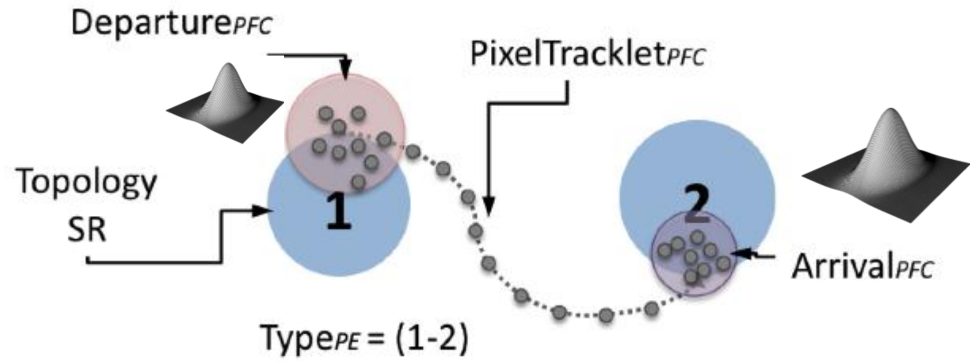


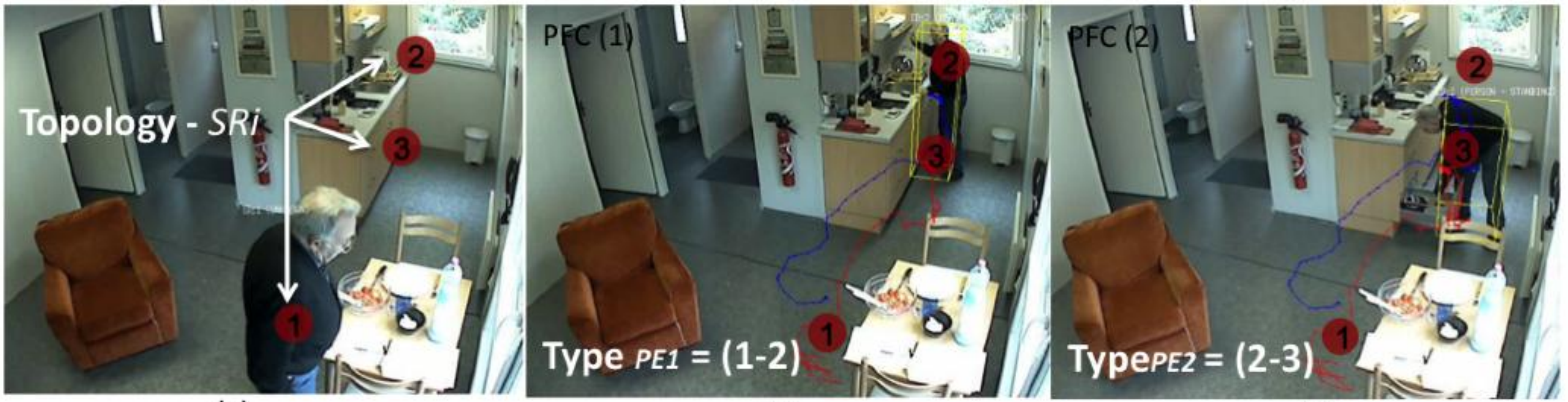
Topologies are important because is where the reasoning is

The **Scene Model = 3 Topologies**: **Multi-Resolution**.

Learning Scenario Models : Primitive Events

$$Type_{PE} = (Start \rightarrow End)$$





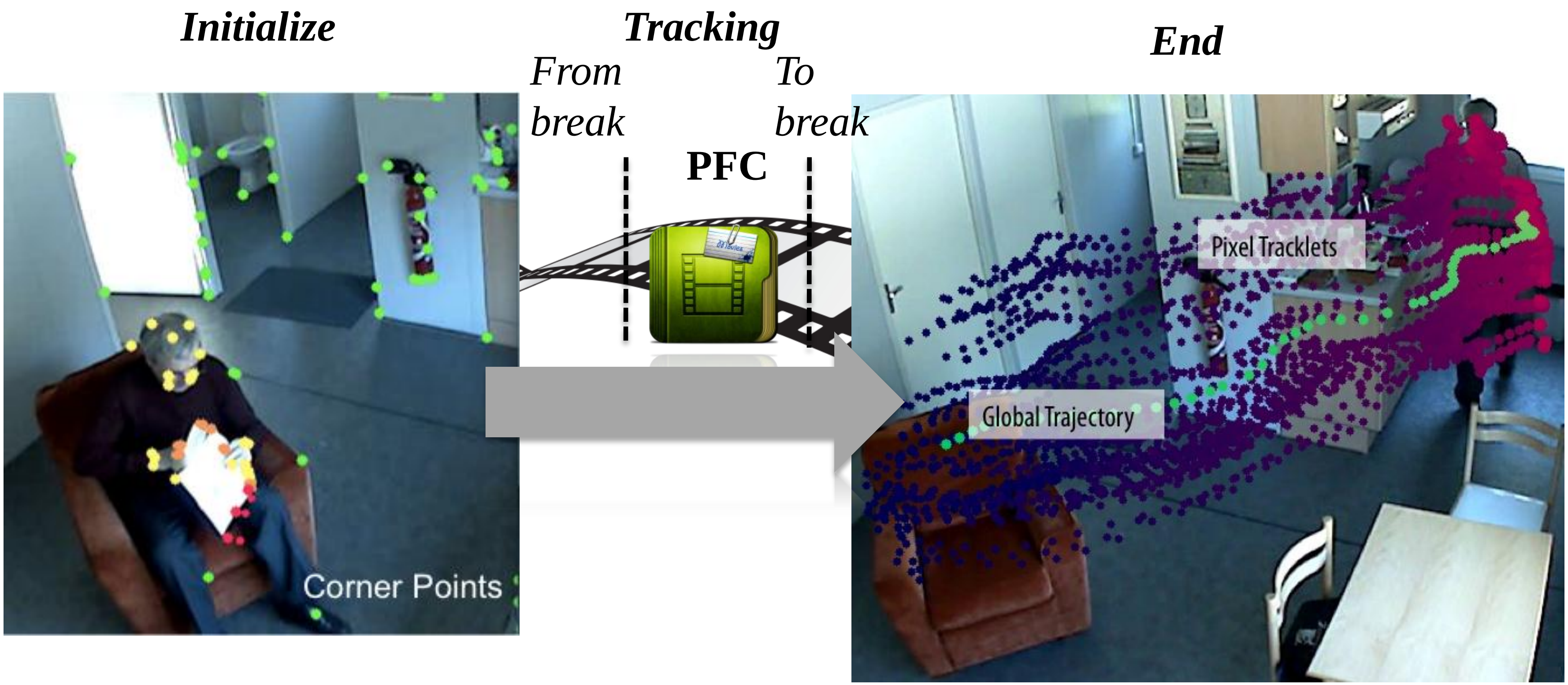
(a) (b) (c)

Primitive Event : global object motion between 2 zones.

Advantage:

The topology regions and primitive events are semantically understandable.

Learning Scenario Models: Local tracklets



1. Initialize sparse KLT points

2. Track the points during the whole PFC - pyramidal KLT - [Bouguet 2000]

3. Filter with the global tracker

4. Re-initialize for a new PFC (means short errors)

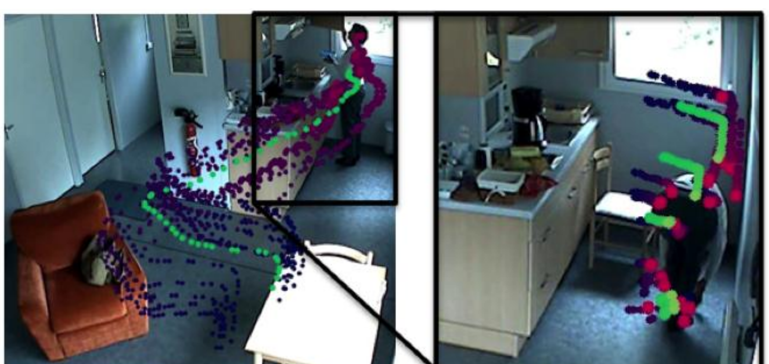
Learning Scenario Models: Local tracklets

Goal: Get semantic describable **main motion** of the body parts parts from perceptual information. *(i.e. complement the global spatial description)*

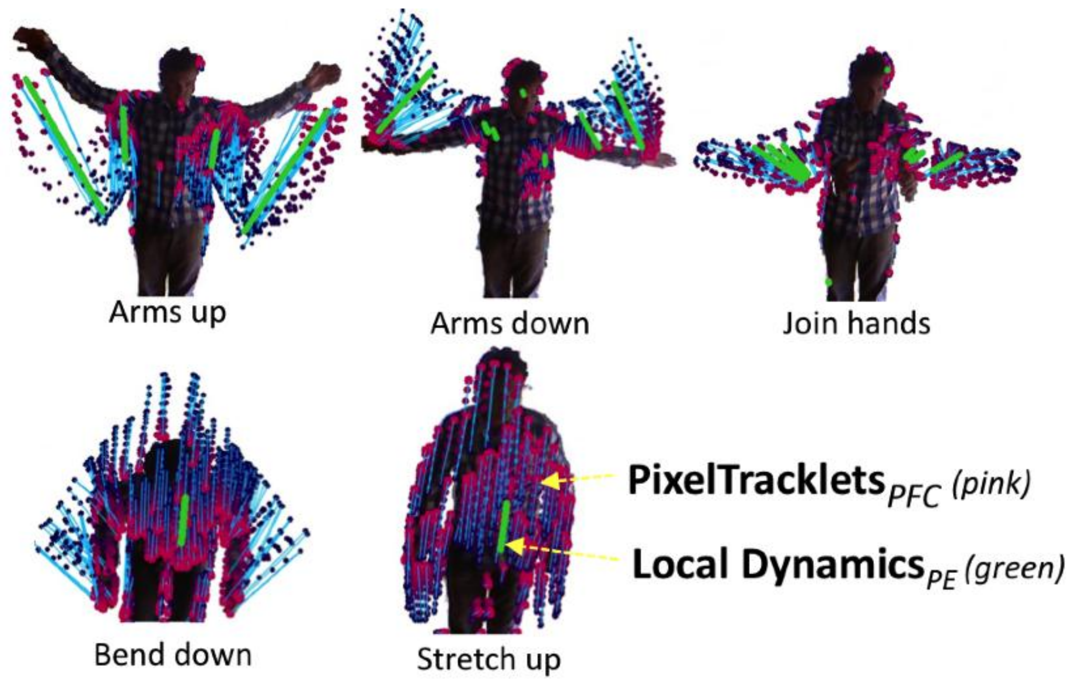
$LocalDynamics_{PE}$ = **Clustering** (Mean Shift) the Pixel tracklets

Trick: *Adapt the bandwidth dynamically to the walked distance.*

$$h = \|PFC_{Departure}.\mu - PFC_{Arrival}.\mu\| * C$$



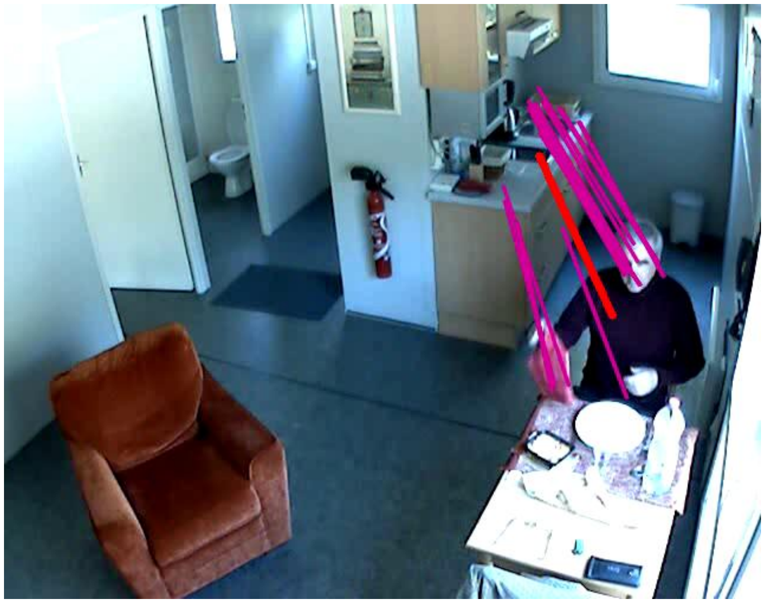
(a) (b)



WE GET and USE THE GREEN LINES

Learning Scenario Models: Local tracklets

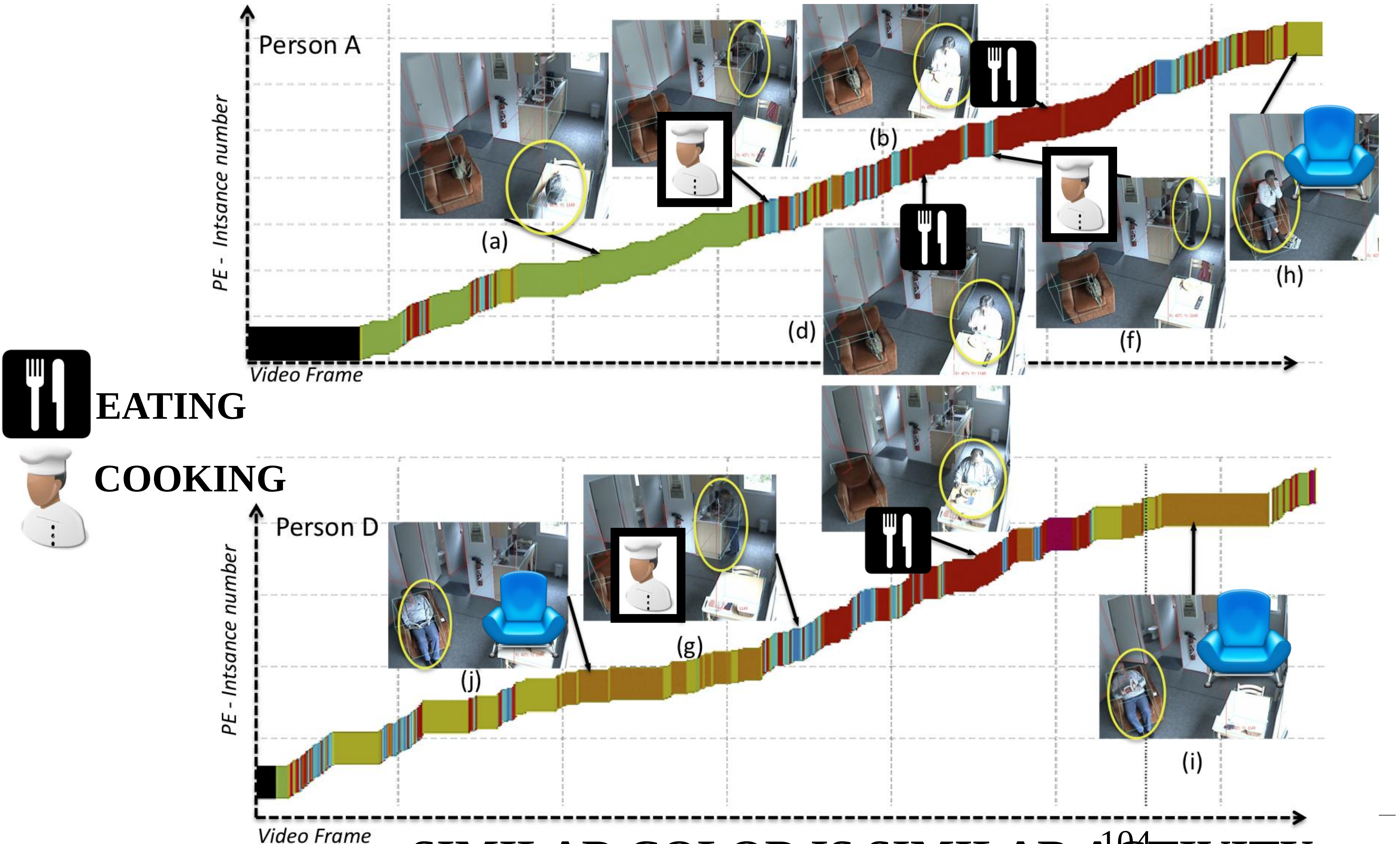
EXAMPLE



SURF & SIFT: slower to compute

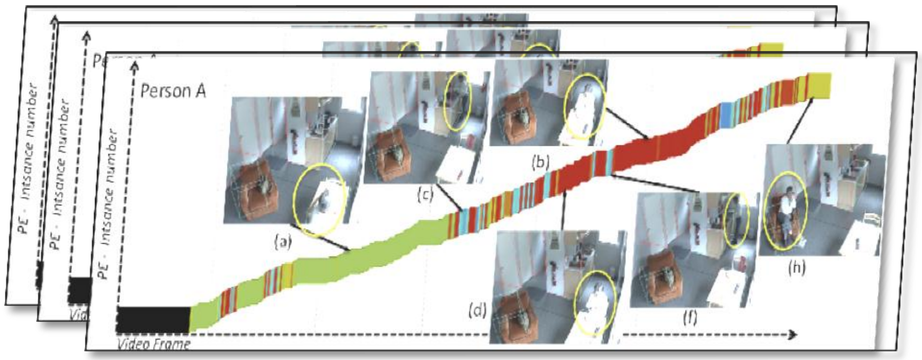
Primitive Events **Results:**

Each PE is colored by its type



SIMILAR COLOR IS SIMILAR ACTIVITY

Activity Discovery: Find start/end of interesting activity and classify them



Input: Sequences of PE

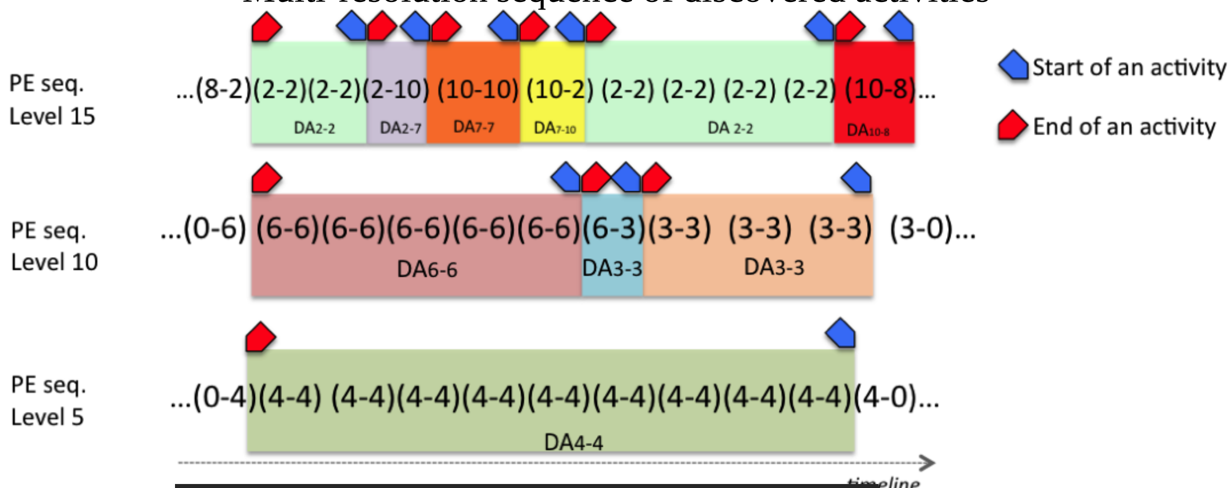
3 RESOULUTIONS

Group/extract by patterns

$$Stay_{A-A} = (A \rightarrow A)^+$$

$$Change_{A-B} = (A \rightarrow B), A \neq B$$

Multi-resolution sequence of discovered activities



-Easy to understand

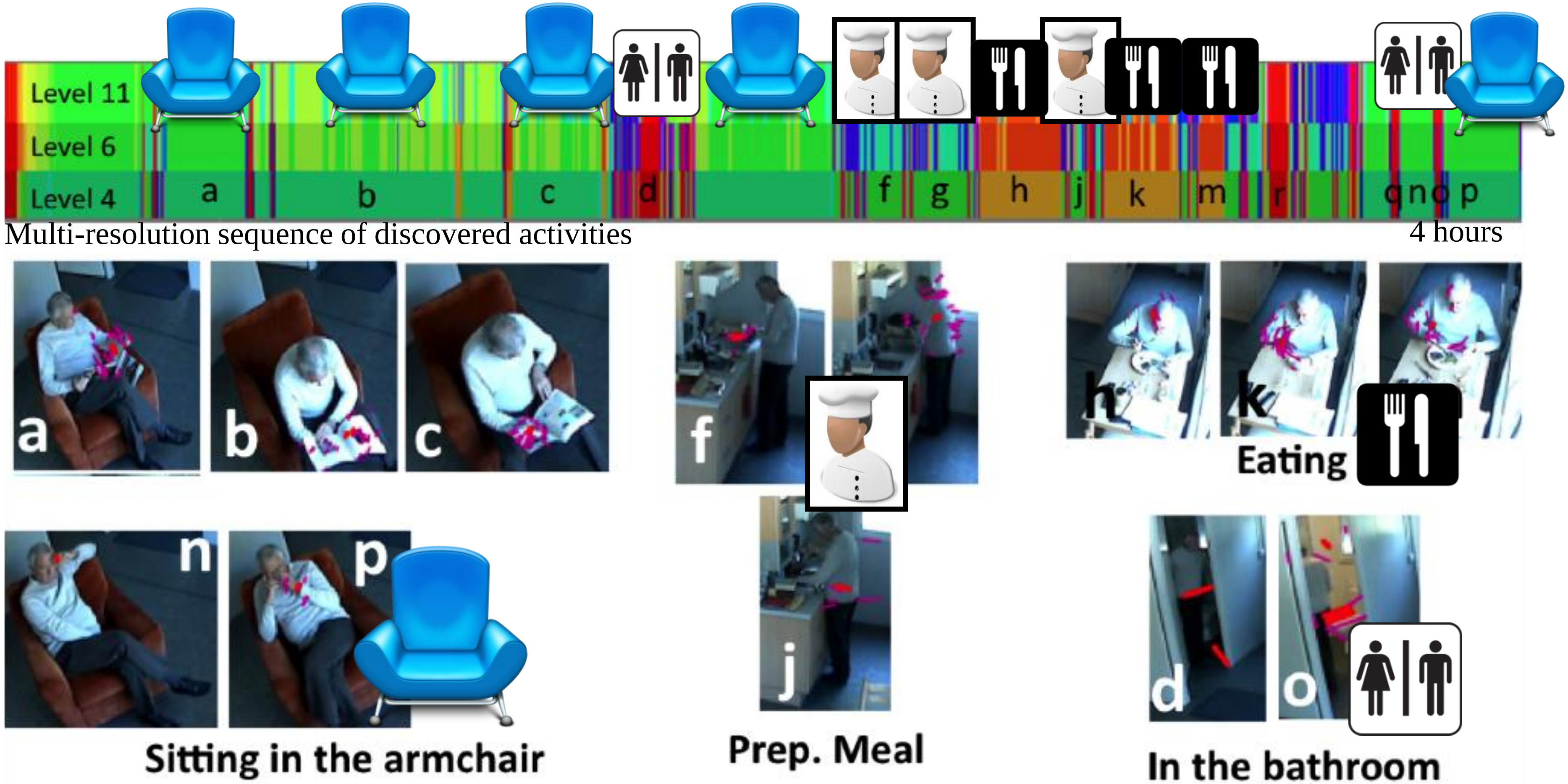
-Non parametric and Deterministic

-The basic patterns can describe complex ones

DA = Discovered Activity *(color = DA type)*

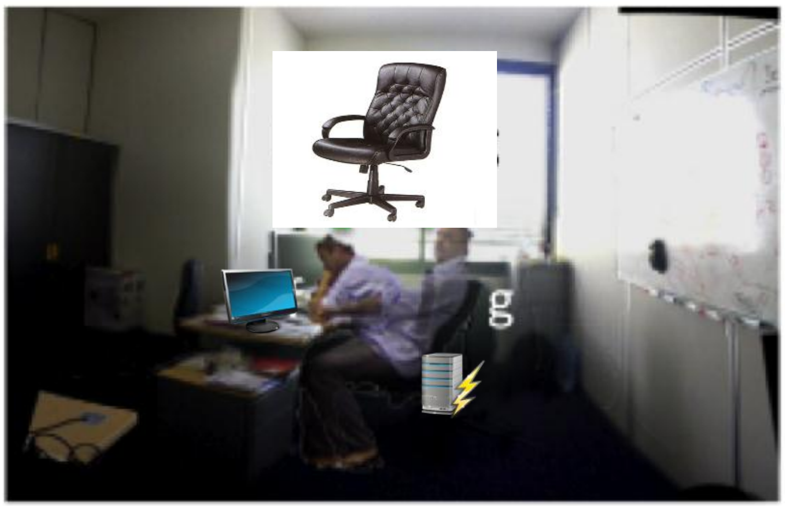
Discovery Results:

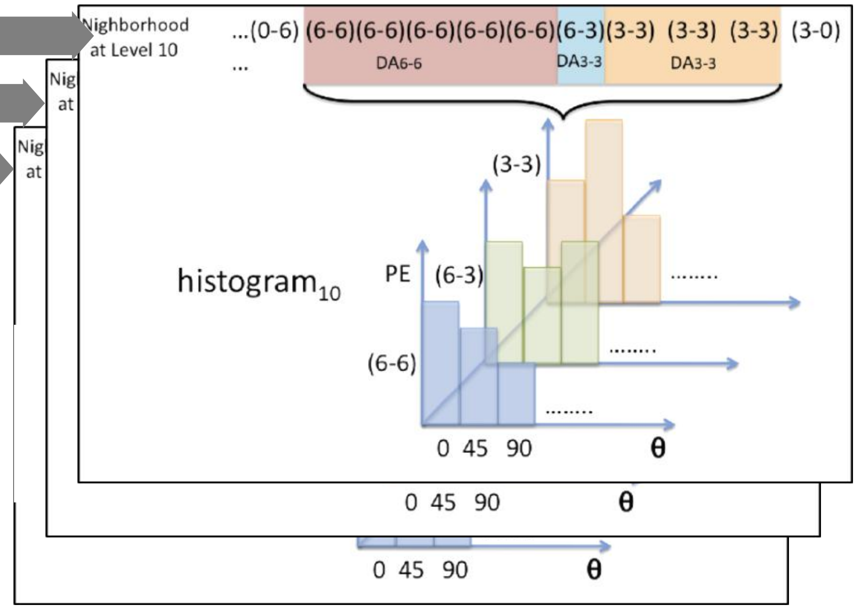
Similar color is similar Discovered Activity



Activity Models *Histograms of Multi-resolutions (HM)*

Is a set of 3 histograms. Each histogram has 2 dimensions. Containing global and local descriptions of the DAs.





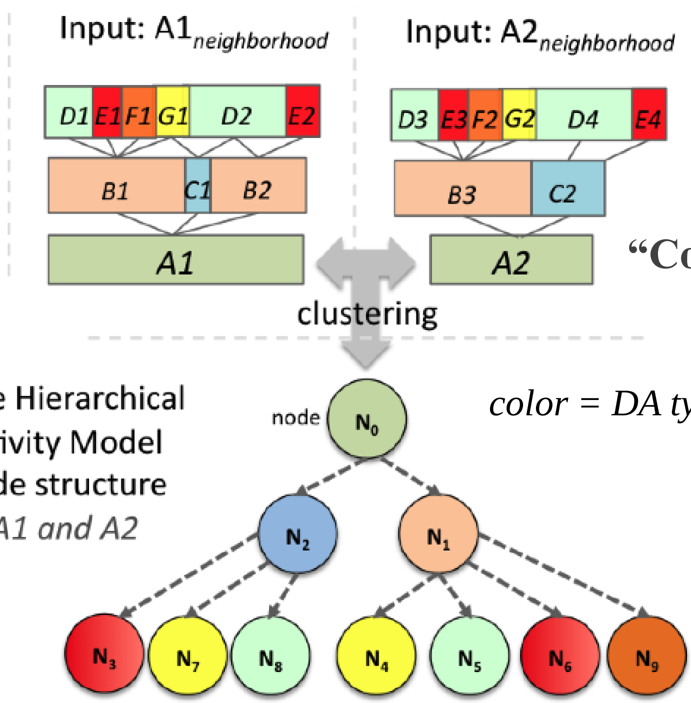
"Coding at Chair"

$$H_i(type_{PE}, \theta) = \sum LDD_N(\theta) \quad : \{\forall N \mid type_N = type_{PE}\}$$

Activity Models: Hierarchical Activity Models (HAM)

Building Nodes

Input **Training Neighborhoods** of a target activity



"Coding at Chair"

The Hierarchical Activity Model node structure *of A1 and A2*

color = DA type

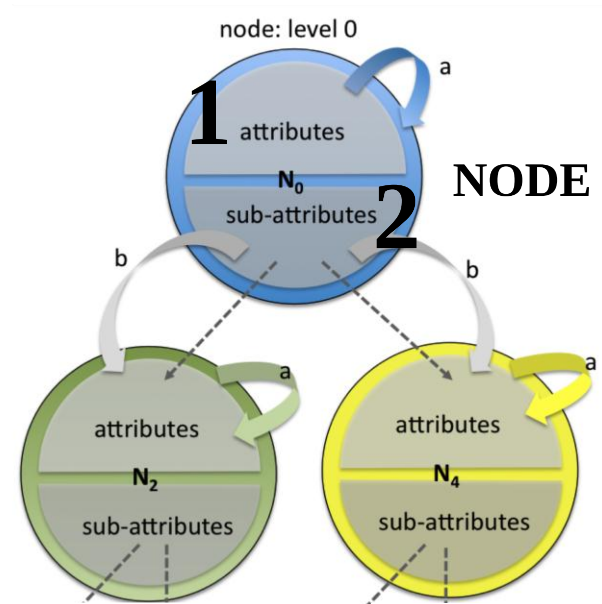
Tree of Nodes

A node N is a set of discovered activities {DA1,DA2...,DAn} where all DAs are at the same resolution level and are of the same type (color)

*A node is composed of **two elements***

1 Attributes

2 Sub-attributes



SUB-NODE **SUB-NODE**

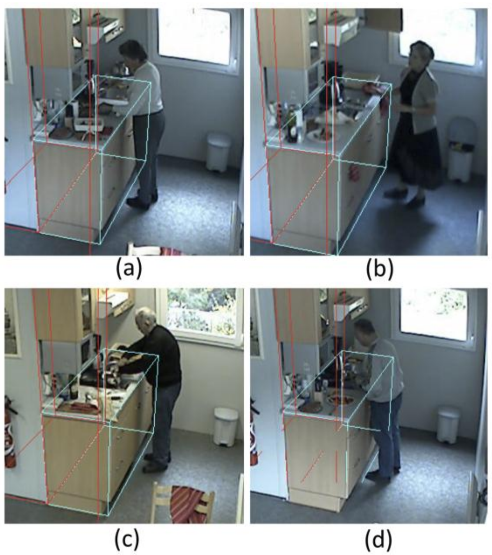
Results

5 targeted activities to be recognized

“Sitting in the armchair”
“Cooking”
“Eating at position A”
“Sitting at Position B”
“Going from the kitchen to the bathroom”.

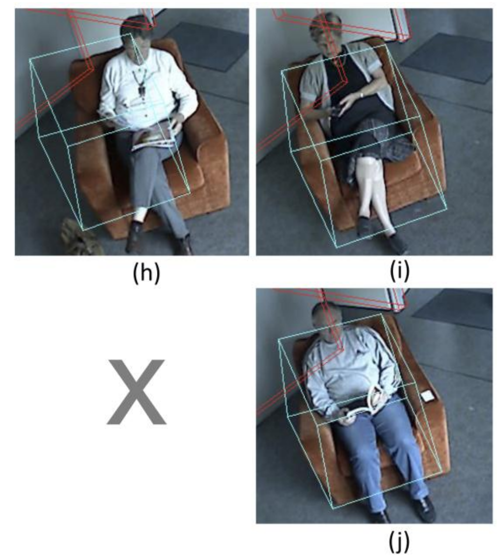
Scene logical Regions

4 Test Persons



“Cooking”

“Eating at Position A”



“Sitting in the Armchair”

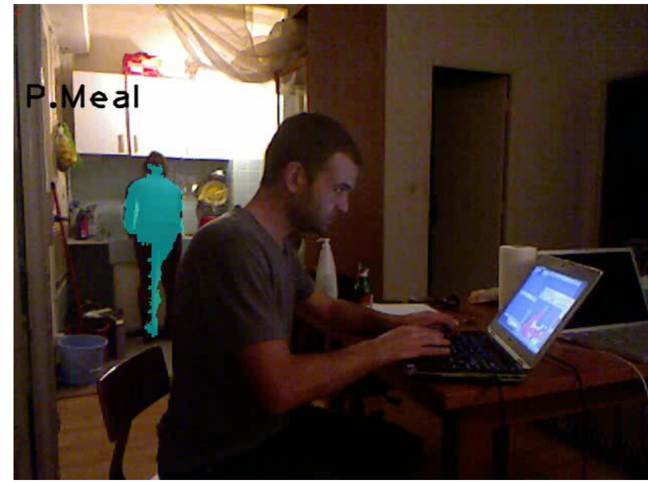
Evaluation

Results: RGB-D Multiples Persons

(i) Sitting.far (red)

(a) Sitting.near

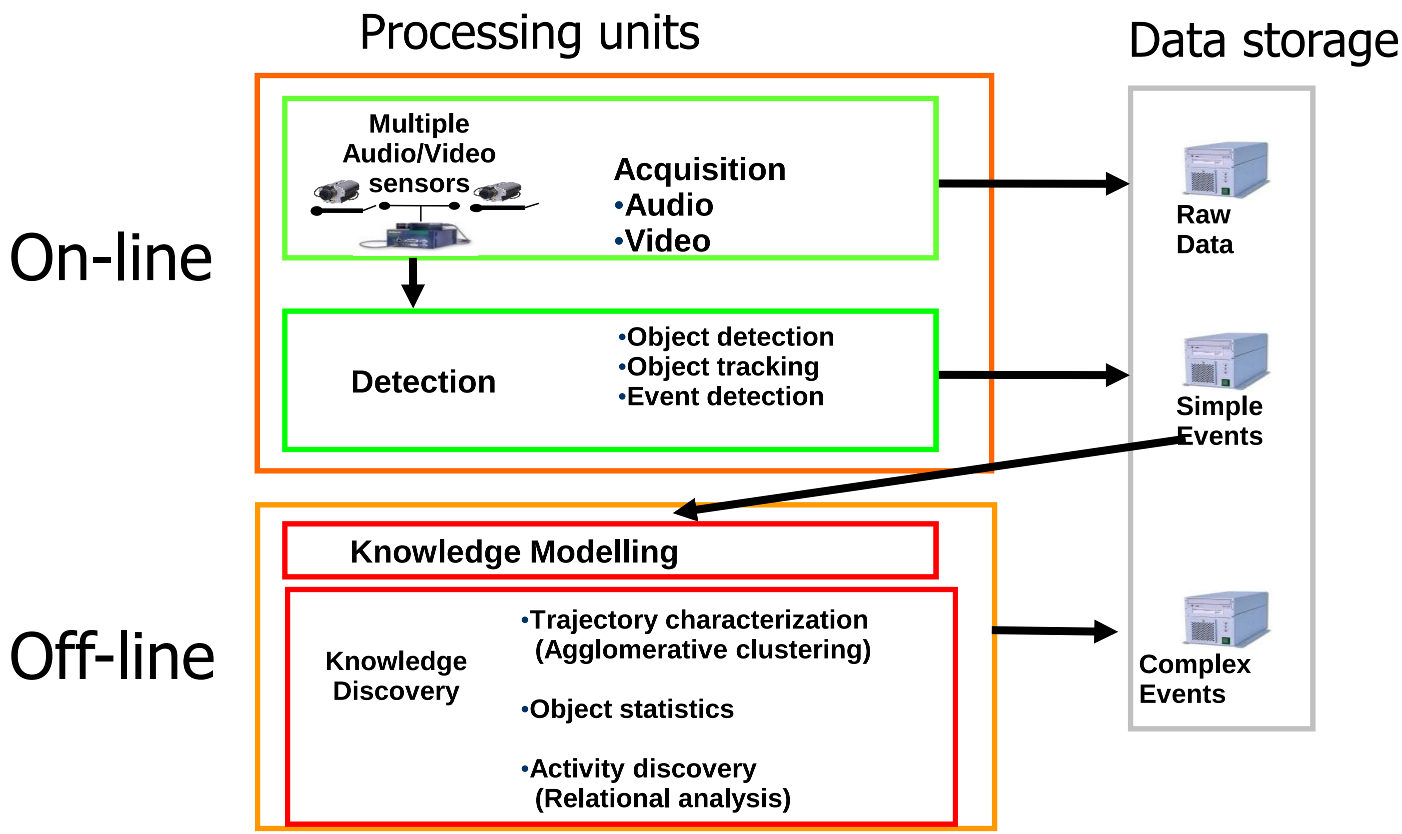
(a) Preparing Meal, Interaction.TrashCan



Tracked object:	silhouette center		
	TP	FP	FN
Preparing.Meal	10	0	0
Interacting.Trashcan	2	1	1
Sitting.near	3	0	0
Sitting.Far	3	0	1

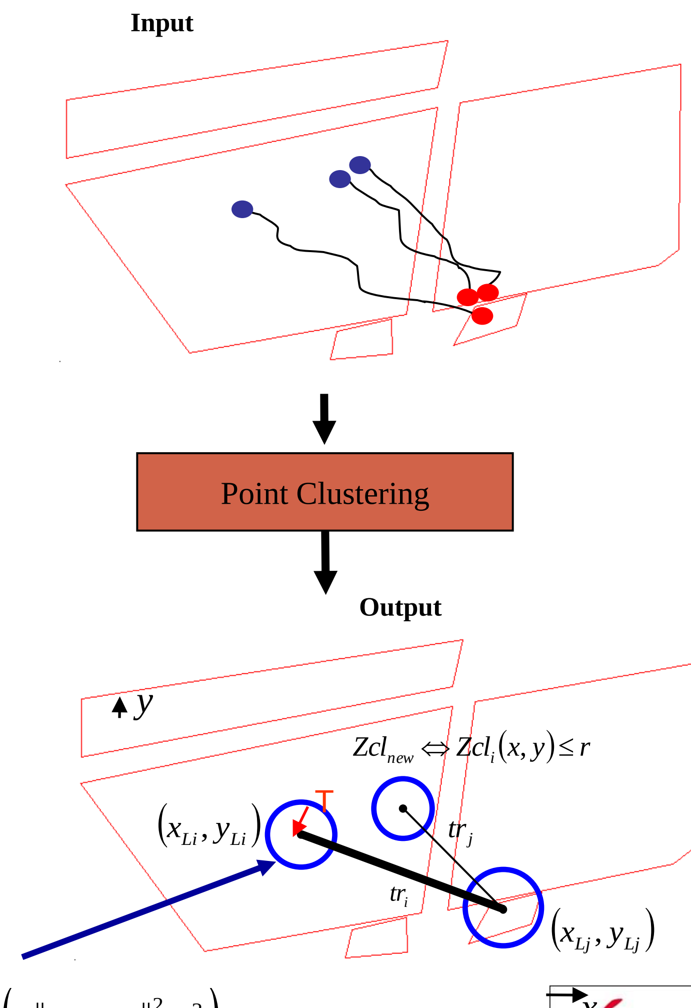
Video Understanding : Knowledge Discovery (E. Corvee, JL. Patino_Vilchis)

- CARETAKER/VANAHEIM: European projects to provide an efficient tool for the **management of large multimedia** collections.



Online learning of zones : Point Clustering

trajectories → Tracklet calculation → Stop points → 2.Point clustering → Discovered Zones



$$Zcl_i(x,y) = \exp\left(-\|x - x_{Li}\|^2 T^2\right)\exp\left(-\|y - y_{Li}\|^2 T^2\right)$$

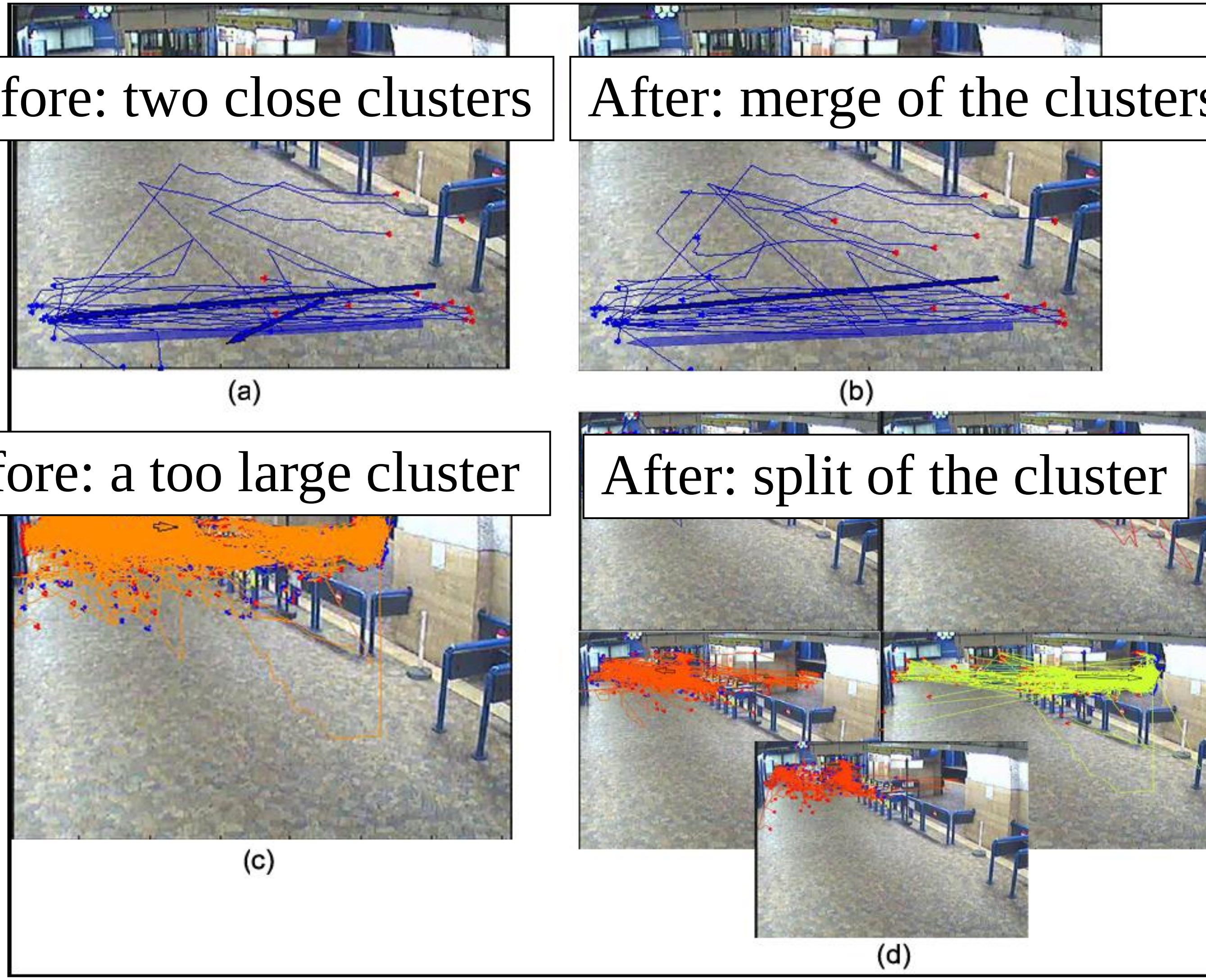
Trajectory feature weight optimization : Results

Before: two close clusters

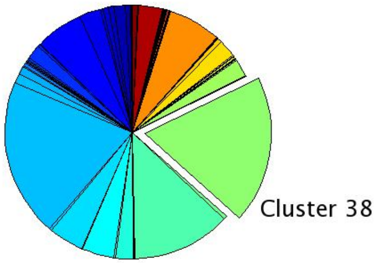
After: merge of the clusters

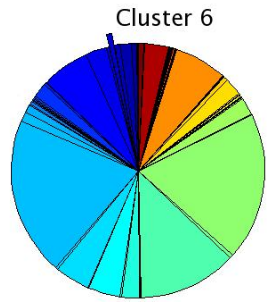
Before: a too large cluster

After: split of the cluster



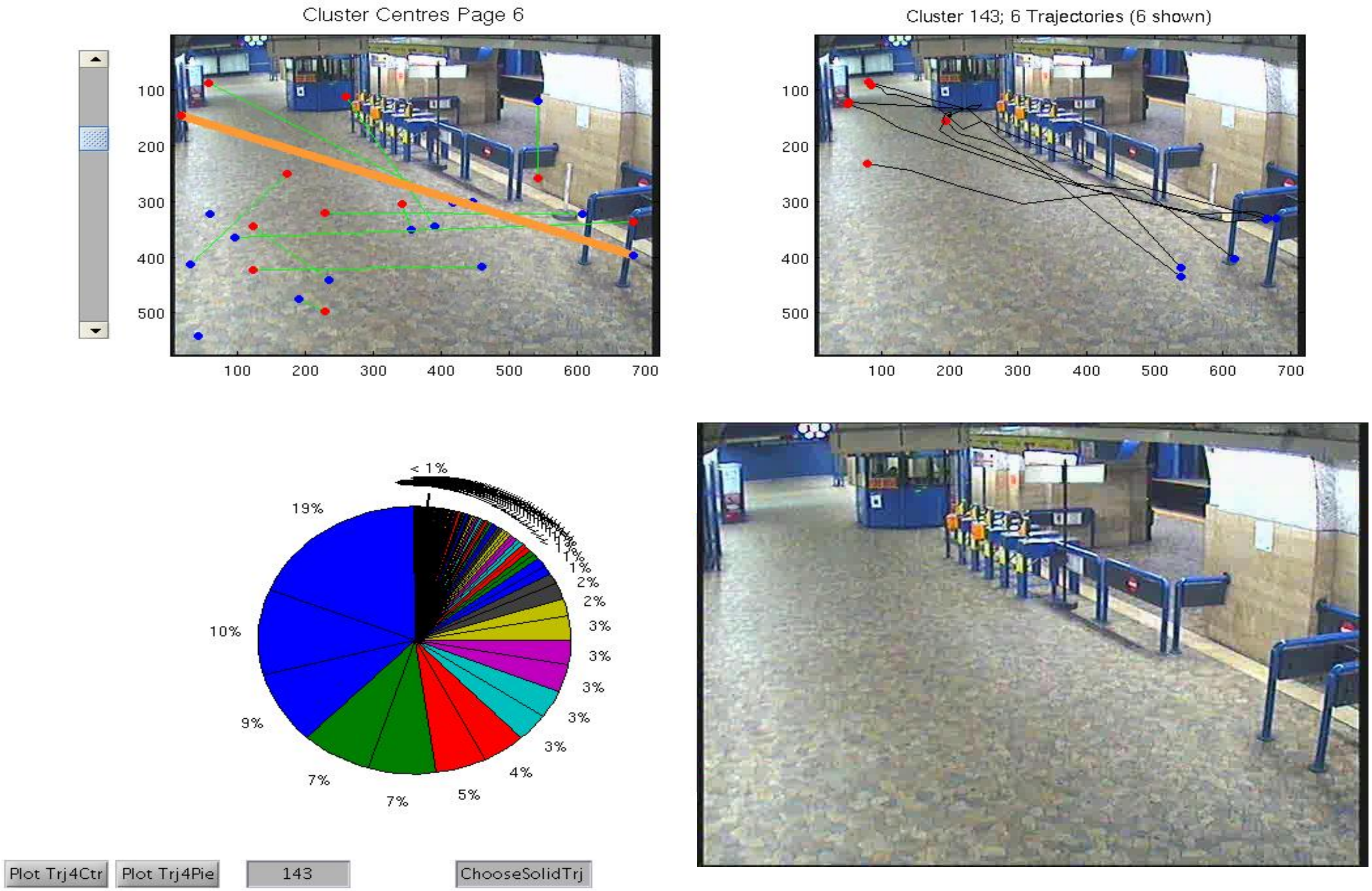
Results : Trajectory Clustering



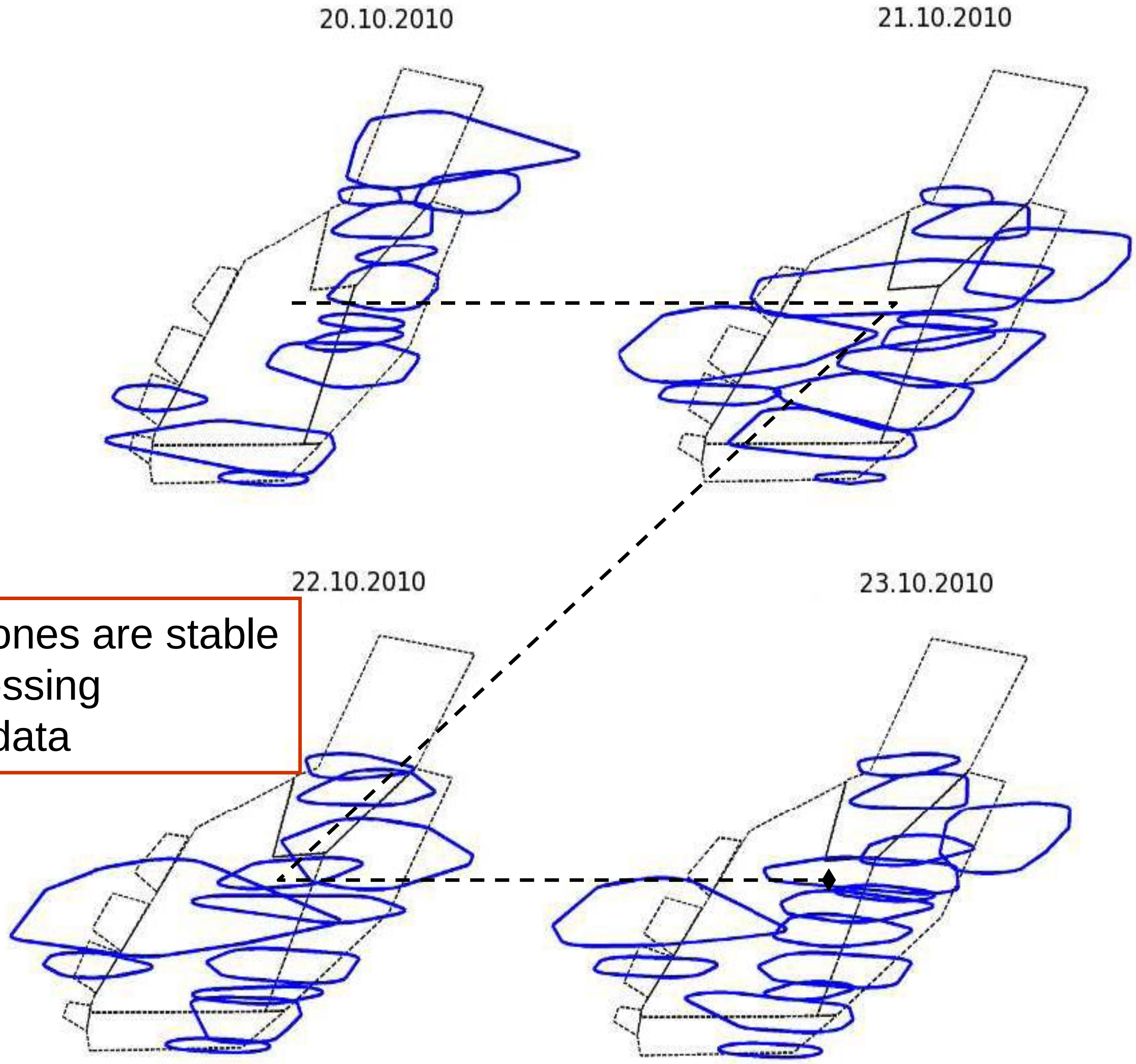


	Cluster 38	Cluster 6
Number of objects	385	15
Object types	types: {'Crowd'} freq: 385	types: {'Person'} freq: 15
Start time (min)	[0.1533, 48.4633]	[28.09, 46.79]
Duration (sec)	[0.04, 128.24]	[2.04, 75.24]
Trajectory types	types: {'4' '3' '7'} freq: [381 1 3]	types: {'13' '12' '19'} freq: [13 1 1]
Significant event	types: {'void '} freq: 385	types: {'inside_zone_Platform '} freq: 15

Trajectory Clustering: rare events in Roma subway



Online learning of zones



Learned zones are stable after processing long term data

Online learning of events trough time

16_25_06 (Wednesday)			16_00_01 (Thursday)			16_00_00 (Saturday)		
rank	(%)	Event	rank	(%)	Event	rank	(%)	Event
1	31.46	at zone Turnstiles	1	29.74	at zone Turnstiles	1	28.33	at zone Turnstiles
2	9.79	at zone Entrance2	2	9.86	at zone Entrance2	2	10.08	at zone Entrance2
3	7.86	zone Entrance2 to zone Turnstiles	3	8.61	zone Entrance2 to zone Turnstiles	3	7.85	zone Entrance2 to zone Turnstiles
4	4.89	zone Turnstiles to zone Entrance2	4	4.64	zone Turnstiles to zone Entrance2	4	5.47	zone Turnstiles to zone Entrance2
5	4.83	zone Turnstiles to zone Centre hall	5	5.04	at zone Centre hall	5	4.55	zone Entrance1 to zone Turnstiles
6	3.72	zone Centre hall to zone Turnstiles	6	3.91	zone Entrance1 to zone Turnstiles	6	3.84	zone Centre hall to zone Turnstiles
7	3.45	at zone Centre hall	7	4.15	zone Turnstiles to zone Centre hall	7	4.69	at zone Centre hall
8	3.31	zone Entrance1 to zone Turnstiles	8	3.75	zone Centre hall to zone Turnstiles	8	3.77	zone Turnstiles to zone Centre hall
9	2.48	zone Turnstiles to zone Entrance1	9	2.60	zone Turnstiles to zone Entrance1	9	2.41	zone Turnstiles to zone Entrance1
10								e Entrance1
11		zone Turnstiles						e Vending machine2
12	1.79	at zone Vending machine1	12	1.57	zone Vending machine1 to zone Turnstiles	12	1.65	at zone Vending machine1
13	1.51	at zone Vending machine2	13	1.31	zone Vending machine1 to zone Centre hall	13	1.55	zone Vending machine1 to zone Turnstiles
14	1.51	zone Vending machine1 to zone Centre hall	14	1.43	at zone Vending machine1	14	1.51	zone Entrance2 to zone Centre hall

Four simple events are the most frequently occurring;
The frequency of occurrence of other events changes slightly.

Online learning of activities trough time

set001			set002			set008		
1	20.13	at zone Turnstiles	1	17.08	at zone Turnstiles	1	17.86	at zone Turnstiles
2	11.21	at zone Entrance2	2	10.23	at zone Entrance2	2	10.93	at zone Entrance2
3	5.98	zone Entrance2 to zone Turnstiles	3	6.30	zone Entrance2 to zone Turnstiles	3	6.10	zone Entrance2 to zone Turnstiles
4	4.13	at zone Turnstiles;at zone Turnstiles	4	3.24	zone Entrance1 to zone Turnstiles	4	3.92	zone Turnstiles to zone Entrance2
5	3.16	zone Turnstiles to zone Entrance2	5	3.24	at zone Turnstiles;at zone Turnstiles	5	3.70	at zone Turnstiles;at zone Turnstiles
6	2.61	zone Entrance1 to zone Turnstiles	6	2.90	zone Turnstiles to zone Entrance2	6	3.53	zone Entrance1 to zone Turnstiles
7	2.29	at zone Entrance1	7	2.26	at zone Entrance1	7	1.85	at zone Entrance1
8	2.18	zone Turnstiles to zone Centre hall	8	1.74	zone Centre hall to zone Turnstiles	8	1.74	zone Centre hall to zone Turnstiles
9	1.74	zone Centre hall to zone Turnstiles	9	1.95	zone Turnstiles to zone Centre hall	9	1.48	zone Turnstiles to zone Entrance1
10	1.52	zone Turnstiles to zone Entrance1	10	1.42	at zone Centre hall	10	1.63	zone Entrance2 to zone Turnstiles;at zone Turnstiles
1[illegible]	[illegible]	[illegible] Turnstiles	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] all
12	1.31	at zone Centre hall	12	1.42	zone Entrance2 to zone Turnstiles;at zone Turnstiles	12	1.06	zone Entrance1 to zone Turnstiles;at zone Turnstiles
13	1.31	zone Vending machine1 to zone Turnstiles	13	1.36	zone Entrance2 to zone Centre hall	13	1.54	zone Turnstiles to zone Centre hall
14	1.20	at zone Turnstiles;zone Turnstiles to zone Entrance2	14	1.00	zone Entrance1 to zone Turnstiles;at zone Turnstiles	14	1.01	at zone Turnstiles;zone Turnstiles to zone Entrance2

The most frequently occurring activities correspond to three simple events.

Online learning : Most common activities

at Turnstiles

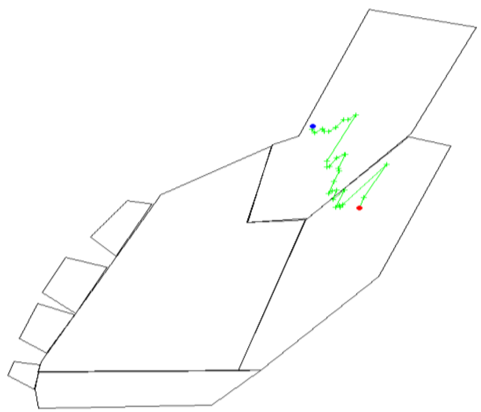
zone South Entry to zone Turnstiles

at zone South Entry

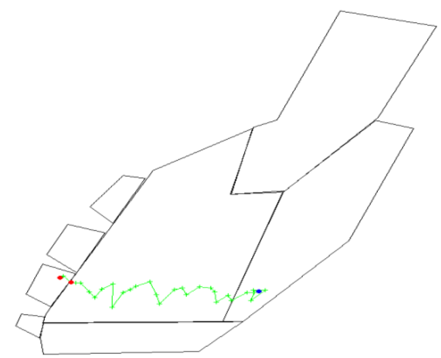
zone North Entry to zone Turnstiles

Online learning : Most common activities

zone Turnstiles to zone South Entry



zone Vending machine1 to zone Turnstiles



zone Vending machine2 to Turnstiles

Online learning : Rare activities

Loitering (from tracking error)

at zone Turnstiles;at zone Turnstiles;at zone Turnstiles

Loitering: people talking then entering the station

at zone Centre hall;at zone Centre hall;at zone Centre hall;at zone Turnstiles

VM not working (from tracking error)

at zone Centre hall;zone Centre hall to zone Vending machine2;zone Vending machine2 to zone Vending machine1

VM not working

zone South Entry to zone Centre hall;zone Centre hall to zone Vending machine2; zone Vending machine2 to zone Vending machine1;zone Vending machine1 to zone Turnstile

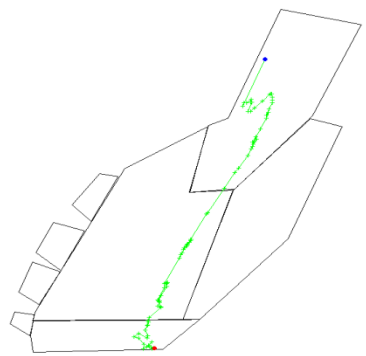
Online learning : Rare activities

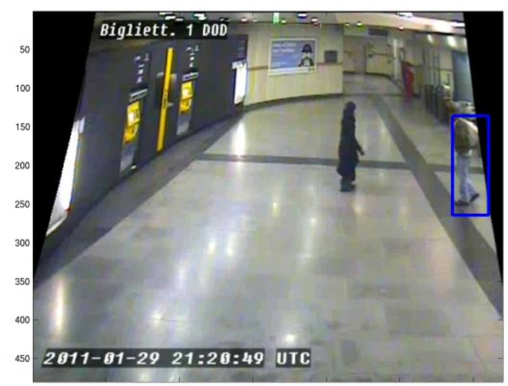
going through the station

zone North Entry to zone South Entry

tickets do not work; need new tickets

zone Turnstiles to zone Centre hall;zone Centre hall to zone Vending machine2;zone Vending machine2 to zone Centre hall;zone Centre hall to zone Turnstiles





Online learning : Loitering activities

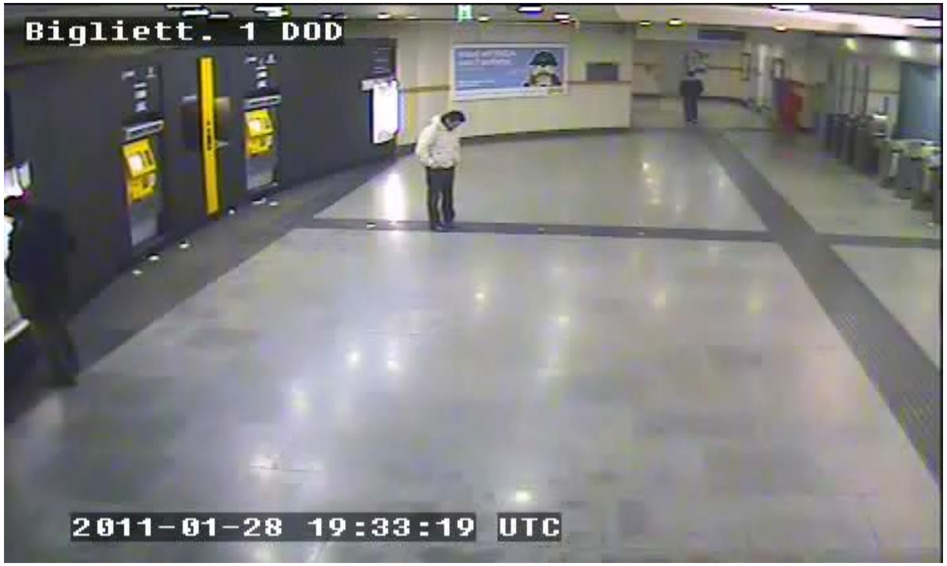
Look for object O
where

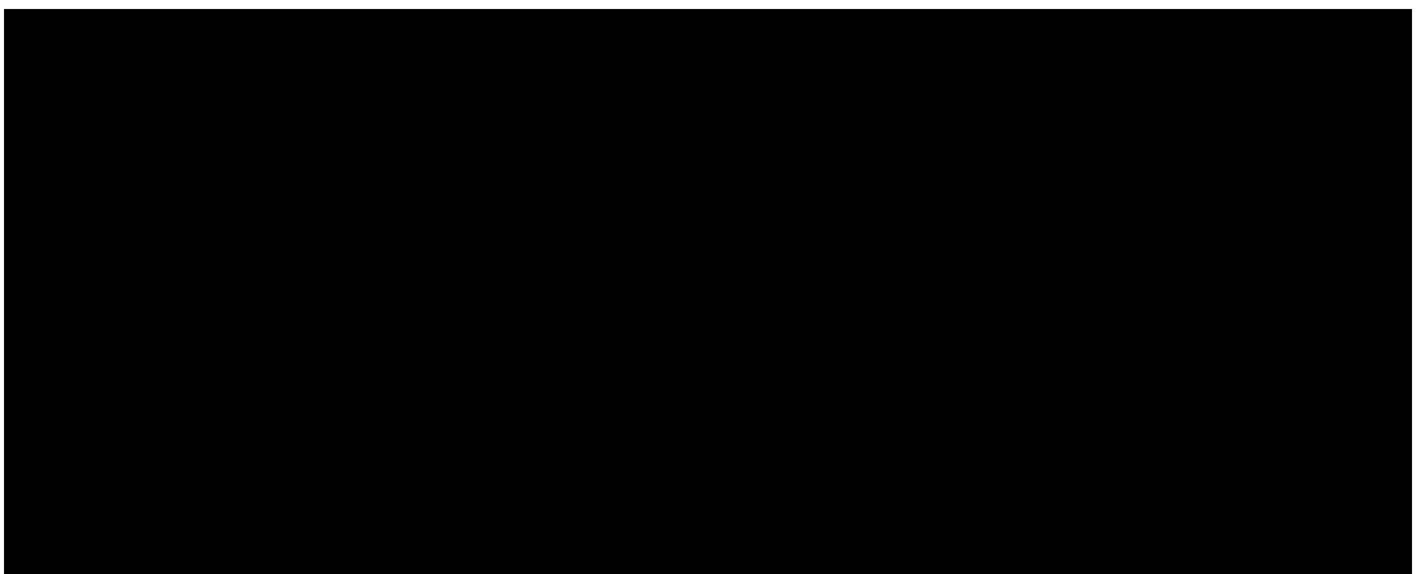
O.Zone.avg_speed < (M.global_avg_speed – M.global_stddev_speed/2)
and
O.path_length > (M.avg_path_length + M.stddev_path_length)
and
O.walked_zones_nb > 8

M : average on objects tracked on 8 hours

Loitering	GT #	TP #	FP #	Precision	Recall
2011-01-29T18_00_01-1262318	1	1	0	0.45	1
2011-01-28T20_00_01-1763039	2	2	3		
2011-01-28T20_00_01-1467943	2	2	2		
2011-01-30T20_00_00	0	0	0		
2011-02-01T20_00_00	0	0	~ 1		

Online learning : Loitering activities





Video Understanding: Conclusion

Global approach integrating all video understanding functionalities

while focusing on the easy generation of dedicated systems based on

- cognitive vision: *4D analysis (3D + temporal analysis)*
- artificial intelligence: *explicit knowledge (scenario, context, 3D environment)*
- software engineering: *reusable & adaptable platform (control, library of dedicated algorithms)*

⇨ **Extract and structure knowledge (invariants & models) for**

- Perception for video understanding (perceptual, visual world)
- Maintenance of the 3D coherency throughout time (physical world of 3D spatio-temporal objects)
- Event recognition (semantics world)
- Evaluation, control and learning (systems world)

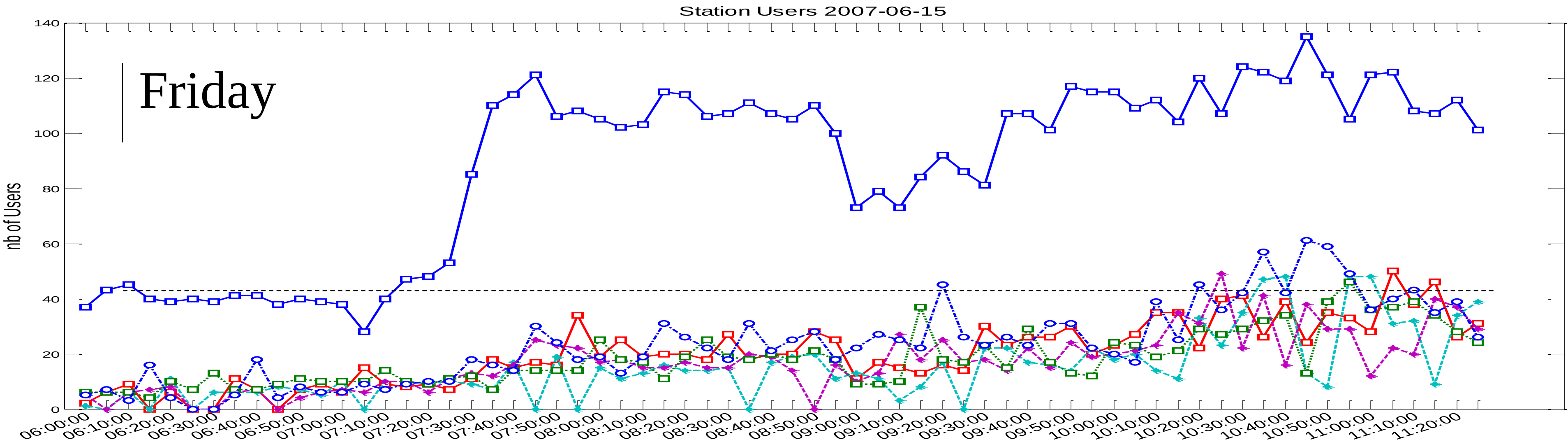
Conclusion

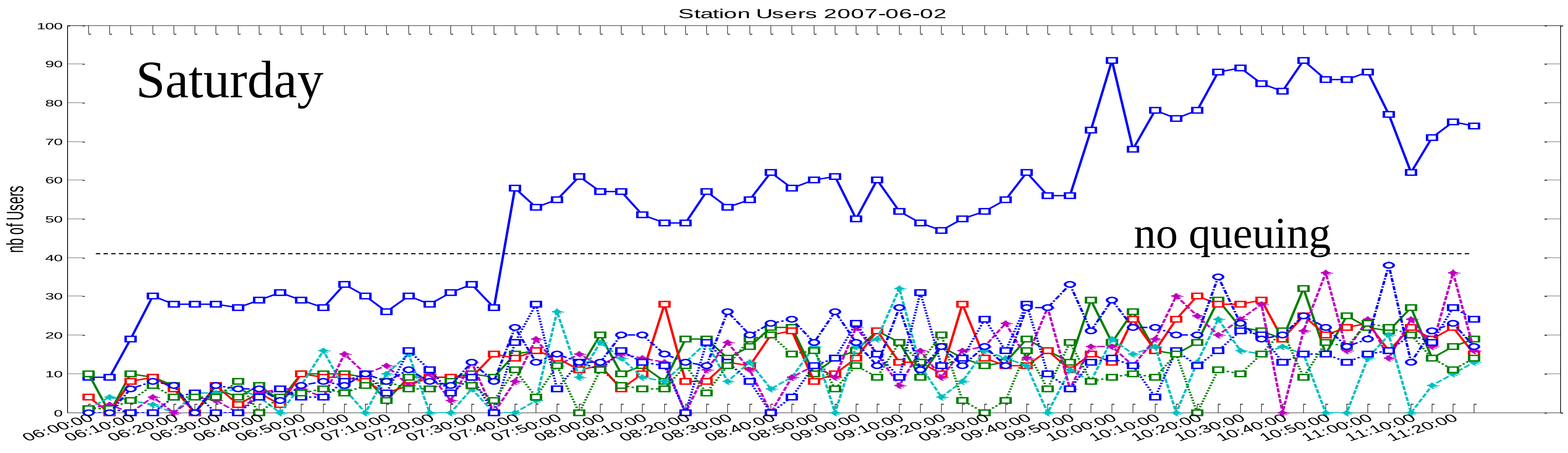
A **global framework** for building video understanding systems:

- Hypotheses:
  - mostly fixed cameras
  - 3D model of the empty scene
  - predefined behavior models
- Results:
  - Video understanding real-time **systems** for Individuals, Groups of People, Vehicles, Crowd, or Animals …
- Perspectives:
  - Finer human shape description: *gesture models, face detection*
  - Design of learning techniques to complement a priori knowledge:
    - visual concept learning
    - scenario model learning
  - Scaling issue: managing large network of heterogeneous sensors (cameras, PTZ, microphones, optical cells, radars….)

Event detection examples

Trajectory Clustering: two day analysis

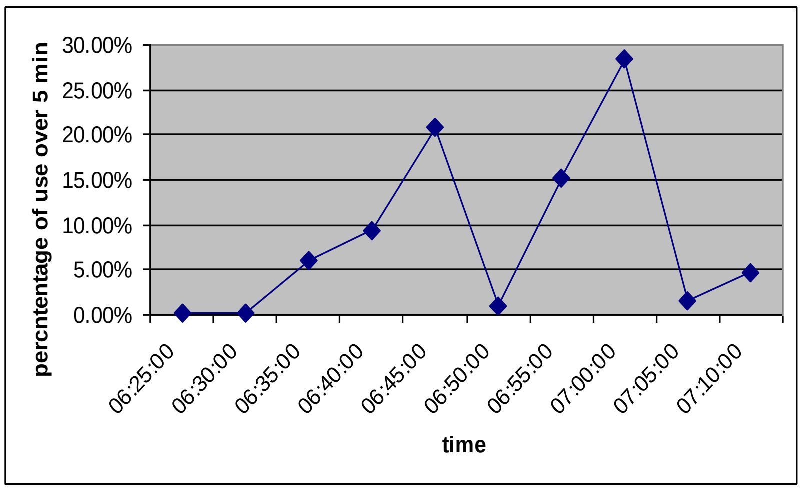


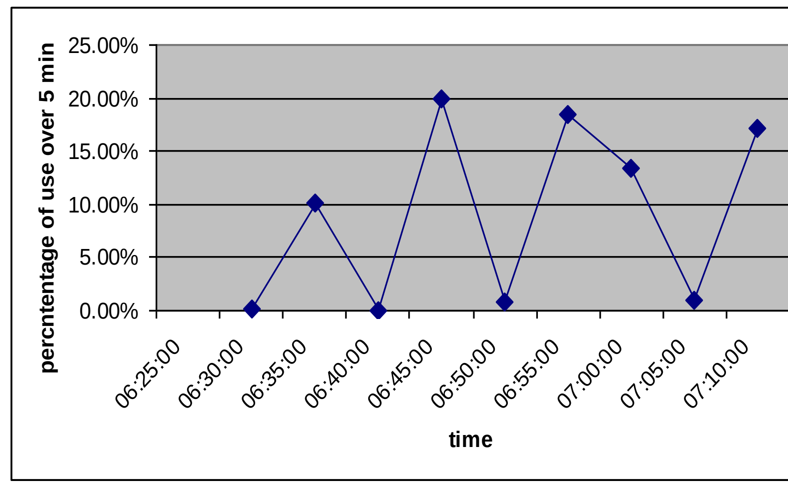


Contextual Object Analysis

Vending Machine 2

Vending Machine 1



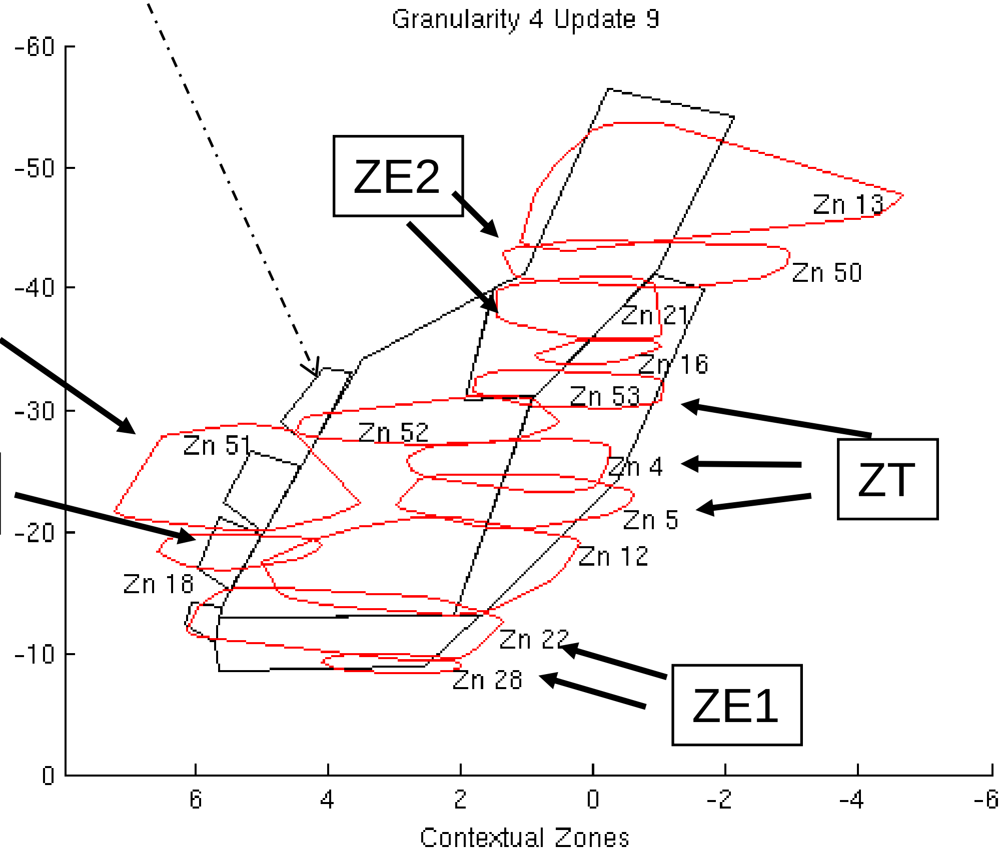


With an increase of people, there is an increase on the use of vending machines

3.1 Results and Evaluation : 13 last learned zones (several days)

13 Matching Zns	Missing Zns
7 GT Zns	1 GT Zn (Poster2 Zn)

*Matching between zones calculated from their intersection



Video Understanding