Video Understanding

Monique THONNAT

Projet Orion INRIA Sophia Antipolis

ORION research team

Intelligent Reusable Systems for Cognitive Vision

⇨ **Multidisciplinary team**: artificial intelligence, software engineering, computer vision

Video Understanding

- Definition:
  - real time and automated analysis of video sequences
  - video understanding= from people detection and tracking to behavior recognition
- Examples:
  - Recognition of office scenes for video communication
  - Recognition of metro scenes for visual surveillance

Video Understanding

- Behavior recognition:
  - approach based on a priori knowledge
    - model of the empty scene (3D geometry and semantics)
    - models of predefined scenarios
  - a language for representing scenarios based on combination of states and events
    - more than 20 states and 20 events can be used
  - a reasoning mechanism for real time detection of states, events and scenarios (e.g. temporal reasoning, constraints solving techniques)

Video Understanding

Applications:

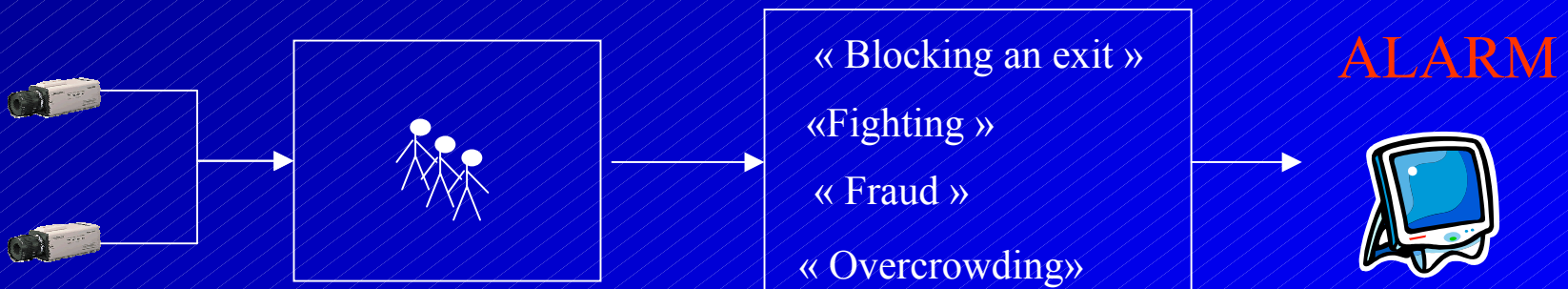
- Strong impact for visual surveillance (metro station, bank agencies, trains, airports)
  - 5 European projects (PASSWORDS, AVS-PV, AVS-RTPW, ADVISOR, AVITRACK)
  - 4 contracts with End-users companies (metro, bank, trains)
  - 2 transfer activities with Bull (Paris) and Vigitec (Brussels)

- Video understanding for video communication (Mediaspace)

Video understanding

Objective: to automate the recognition of specific human behaviors from video sequences

---> *Interpretation of the videos from pixels to alarms*



Cameras

People detection and tracking

Behavior recognition

Interface for alarm management

Video Understanding

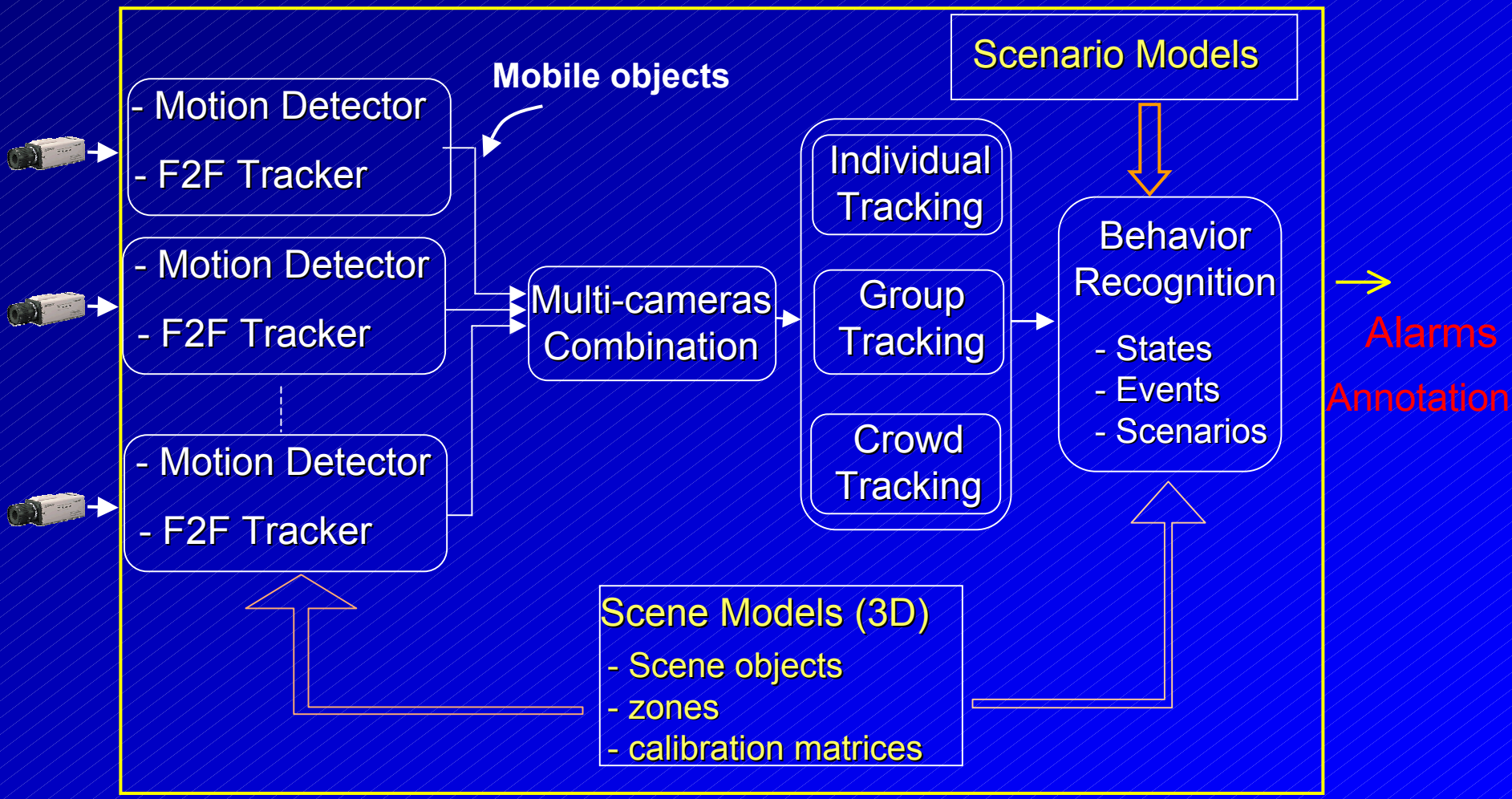
- Typical application:

  European project **ADVISOR:**

  Annotated **D**igital **V**ideo for **I**ntelligent **S**urveillance and **O**ptimised Retrieval

- Intelligent system of video surveillance in metros
  Problem : 1000 cameras but few human operators
- Automatic selection in real time of the cameras viewing abnormal behaviours
- Automatic annotation of recognised behaviors in a video data base using XML

Video Understanding Platform

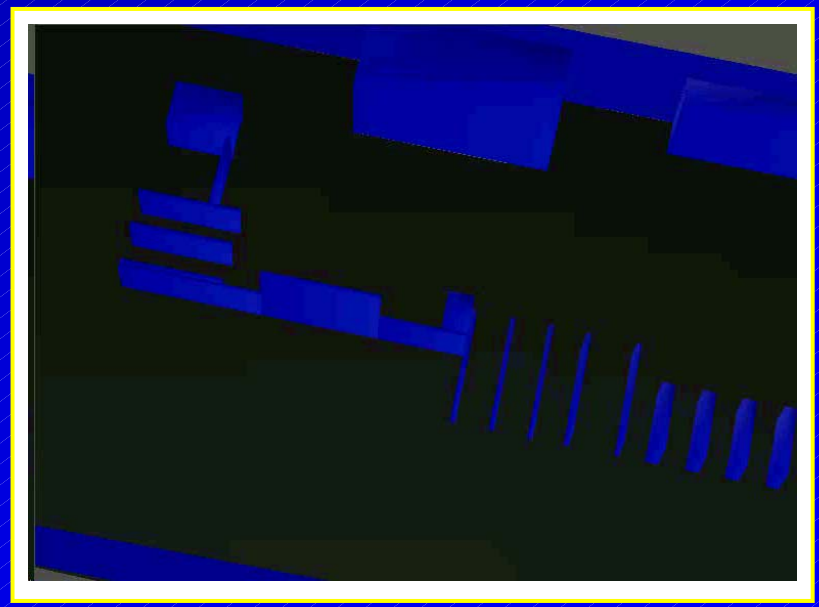


Video Understanding: 3D Scene Model

- 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

Video Understanding: Scene Model

- Barcelona Metro Station Sagrada Famiglia mezzanine (cameras C10, C11 and C12)



Video Understanding

- *States, Events* and *Scenarios :*
  - *State:* a spatio-temporal property involving one or several actors on a time interval

    Ex : « close», « walking», « seated»
  - *Event:* a significant change of states

    Ex : « enters», « stands up», « leaves »
  - *Scenario:* a long term symbolic application dependent activity

    Ex : « fighting», « vandalism»

Video Understanding

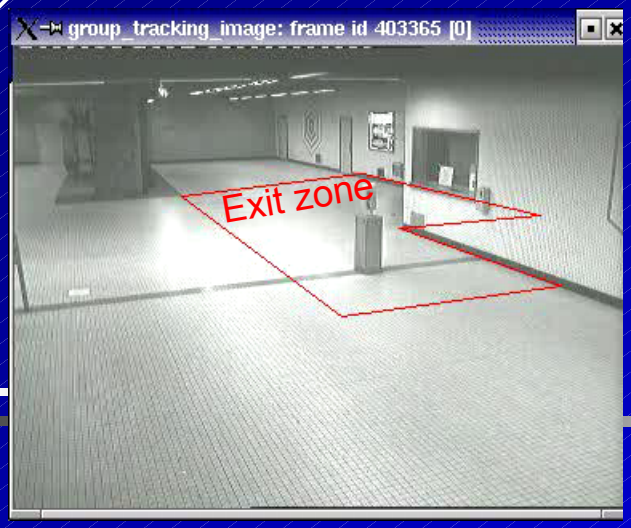
- Several types of States :
  - posture → *{lying, crouching, standing}*
  - direction → *{towards the right, towards the left, leaving, arriving}*
  - speed → *{stopped,walking, running}*
  - distance/object → *{close, far}*
  - distance/person → *{close, far}*
  - posture/object → *{seated, any}*
- Several types of Events :
  - 1 person : *falls down, crouches down, stands up, goes right side, goes left side, goes away, arrives, stops, starts running*
  - 1 person & 1 zone : *leaves , enters*
  - 1 person & 1 equipment : *moves close to,sits on, moves away from*
  - 2 persons : *moves close to, moves away from*

Scenario Recognition

- We use **several formalisms** to recognise states, events and scenarios :
  - specific routines
  - classification
  - finite state automaton
  - propagation of temporal constraints

Scenario Recognition : Automaton

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

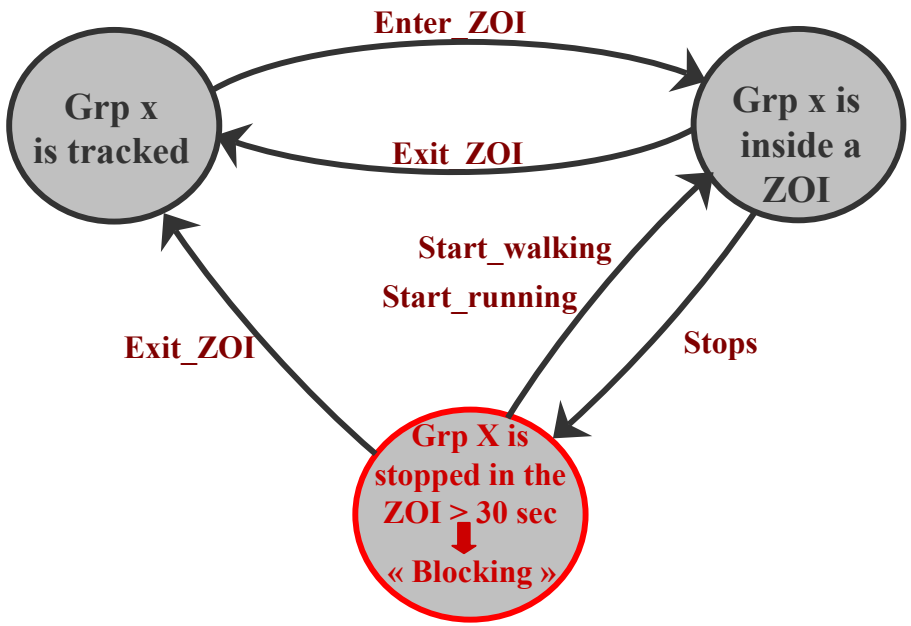


Mobile objects Detection

Group Tracking

Blocking

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



Results

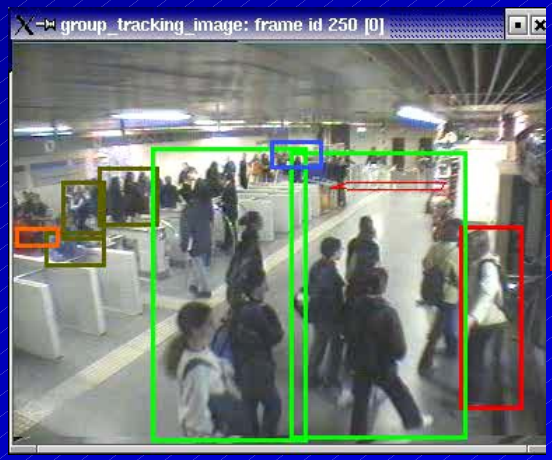
- **Examples : Brussels and Barcelona Metros**

Group behavior

Blocking

Group behavior

Fighting



Crowd behavior

Overcrowding

Individual behavior

Jumping over barrier

Scenario Recognition : Temporal constraints

- A temporal scenario is constituted by three parts :
  - Physical_objects: people, physical objects and interesting zones relative to the scenario.
  - Components : a set of states, events or scenarios
  - Constraints : a set of constraints on the physical objects and on the components involving them. *Use of symbolic, logical, spatial and temporal constraints including Allen's interval algebra operators*.

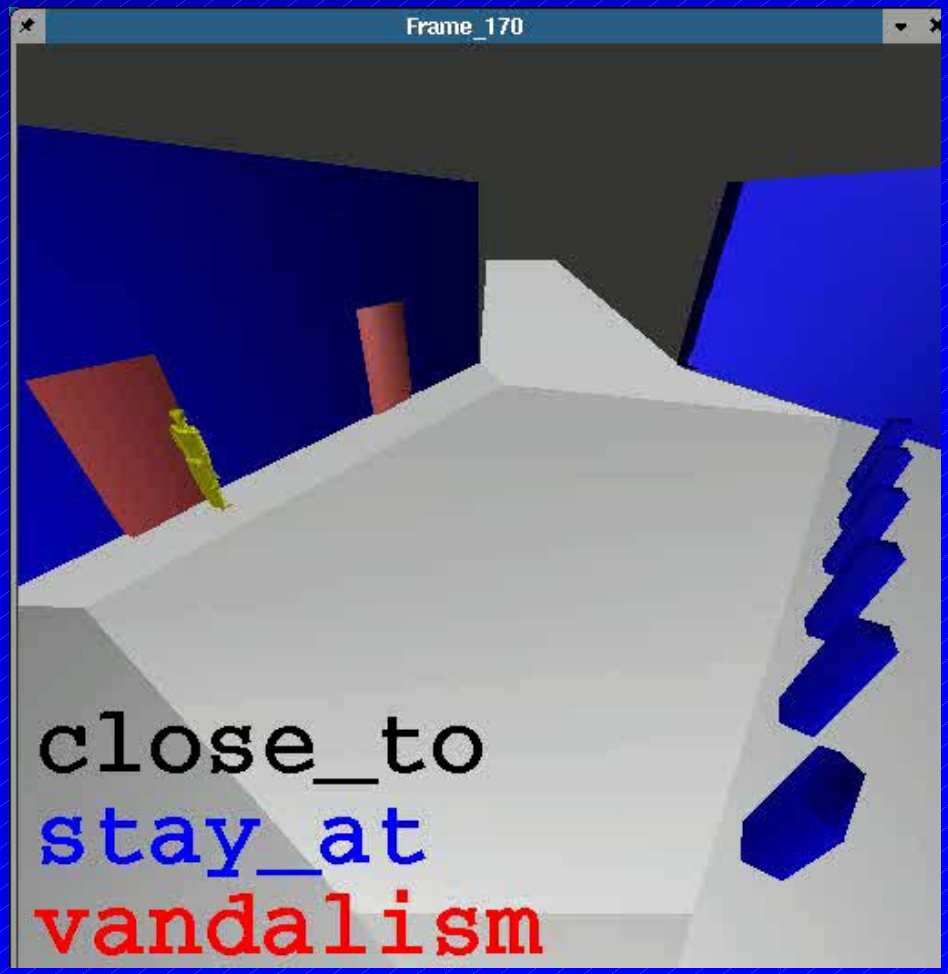
Scenario Recognition : Temporal constraints

- Vandalism scenario description :

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))))

Results

Vandalism in metro (Nuremberg)



Results

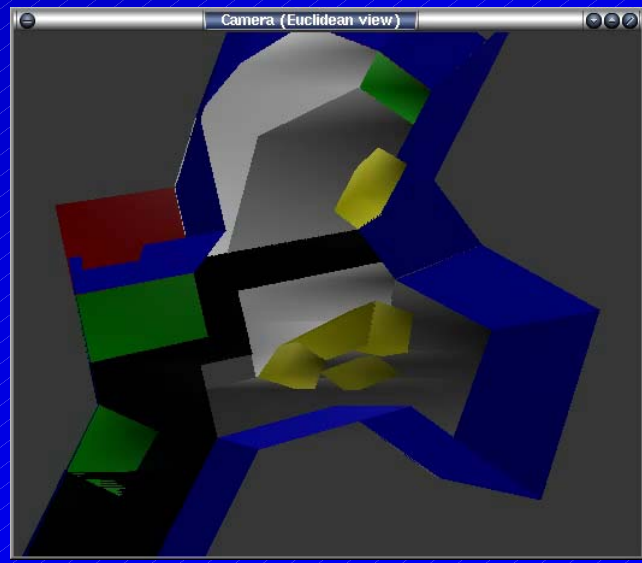
- Recognition of five behaviors: "Blocking", "Fighting", "Jumping over barrier", "Vandalism" and "Overcrowding"
- Tested on 50 metro sequences (10 hours)
- True positive per sequence: 70% ("Fighting") to 95% ("Blocking")
- False positive per sequence: 5% ("Fighting", "Jumping over barrier") to 0% (others)

Results for Bank Monitoring

Bank attack scenario description :

scenario Bank_attack_one_robber_one_employee
physical_objects:
((employee : **Person**), (robber : **Person**), z1: **Back_Counter**, z2: **Entrance_Zone**, z3: **Front_Counter**, z4: **Safe**, d: **Safe_door**)
components:
(State c1 : **Inside_zone**(employee, z1))
(Event c2 : **Changes_zone**(robber, z2,z3))
(State c3 : **Inside_zone**(employee, z4))
(State c4 : **Inside_zone**(robber, z4)))
constraints :
((c2 ***during*** c1) (c2 ***before*** c3)
(c1 ***before*** c3) (c2 ***before*** c4)
(c4 ***during*** c3)
(d **is** open))

Video Understanding for bank surveillance



Conclusion

- Hypotheses:
  - fixed cameras
  - 3D model of the empty scene
  - predefined behavior models
- Results:
  - Behavior understanding for Individuals, Groups of people or Crowd
  - an operational language for video understanding (more than 20 states and events)
  - a real-time platform (5 to 25 frames/s)

Conclusion

Future work:

- Learning techniques to compute optimal set of parameters and to dynamically configure the platform
- Recognition of scenarios based on the recognition of postures
- New application domains : Aware House, Health, Teaching,…

Video Understanding

Animated videos generated from video understanding