Scene Understanding and Assisted Living

Francois BREMOND

INRIA Sophia Antipolis

Institut National Recherche Informatique et Automatisme

Francois.Bremond@inria.fr

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

CoBTeK,

Nice University Hospital

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)*

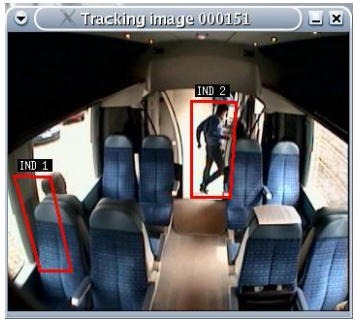
for crowd *(overcrowding)*

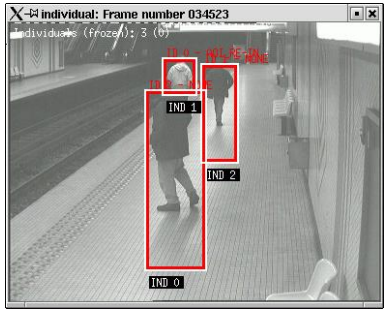
for interactions of people and vehicles *(aircraft refueling)*

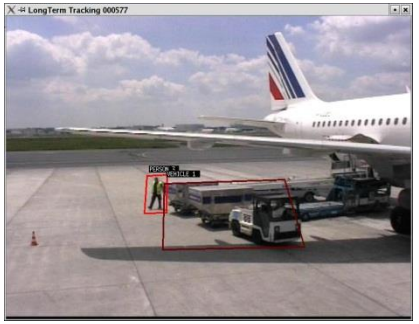
Video Understanding Applications

- 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/



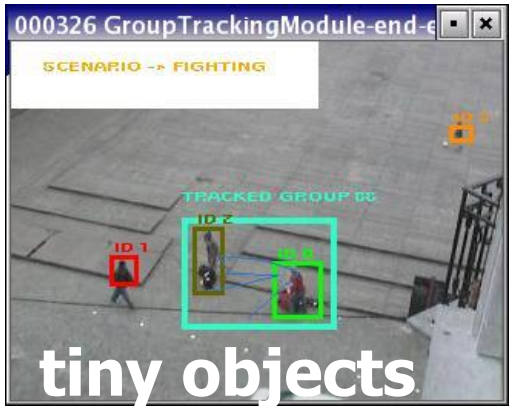




Video Understanding: Issues

Practical issues

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



Monitoring of Activities of Daily Living for Older People

- **Motivation : Increase independence and quality of life:**

- Enable older persons to live longer, autonomously in their preferred environment.
- Reduce costs for public health systems.
- Relieve family members and caregivers.

- **Approach : design sensor-based systems**
- Detecting alarming situations *(eg. Falls)*
- Assess the degree of frailty of elderly people (impact of therapies).
- Detecting changes in behavior
  (*missing activities, disorder, interruptions, repetitions, inactivity*).
- Building a video library of reference behaviors characterizing people frailty.

Monitoring of Activities of Daily Living for Older People

- 3 types of **Human Activities** of interest:
  - Activities which can be well described or modeled by users (e.g. sitting)
  - Activities which can be collected by users through positive/negative samples representative of the targeted activities (e.g. falling)
  - Rare activities which are unknown to the users and which can be observed only through large datasets (e.g. non motivated activities)

- 3 types of **methods** to Recognize Human Activities:
  - Recognition engine using hand-crafted ontologies based on a priori knowledge (e.g. rules) predefined by users
  - Supervised learning methods based on positive/negative samples to build specific classifiers for the targeted activities
  - Unsupervised (fully automated) learning methods based on clustering of frequent activity patterns to discover new activity models

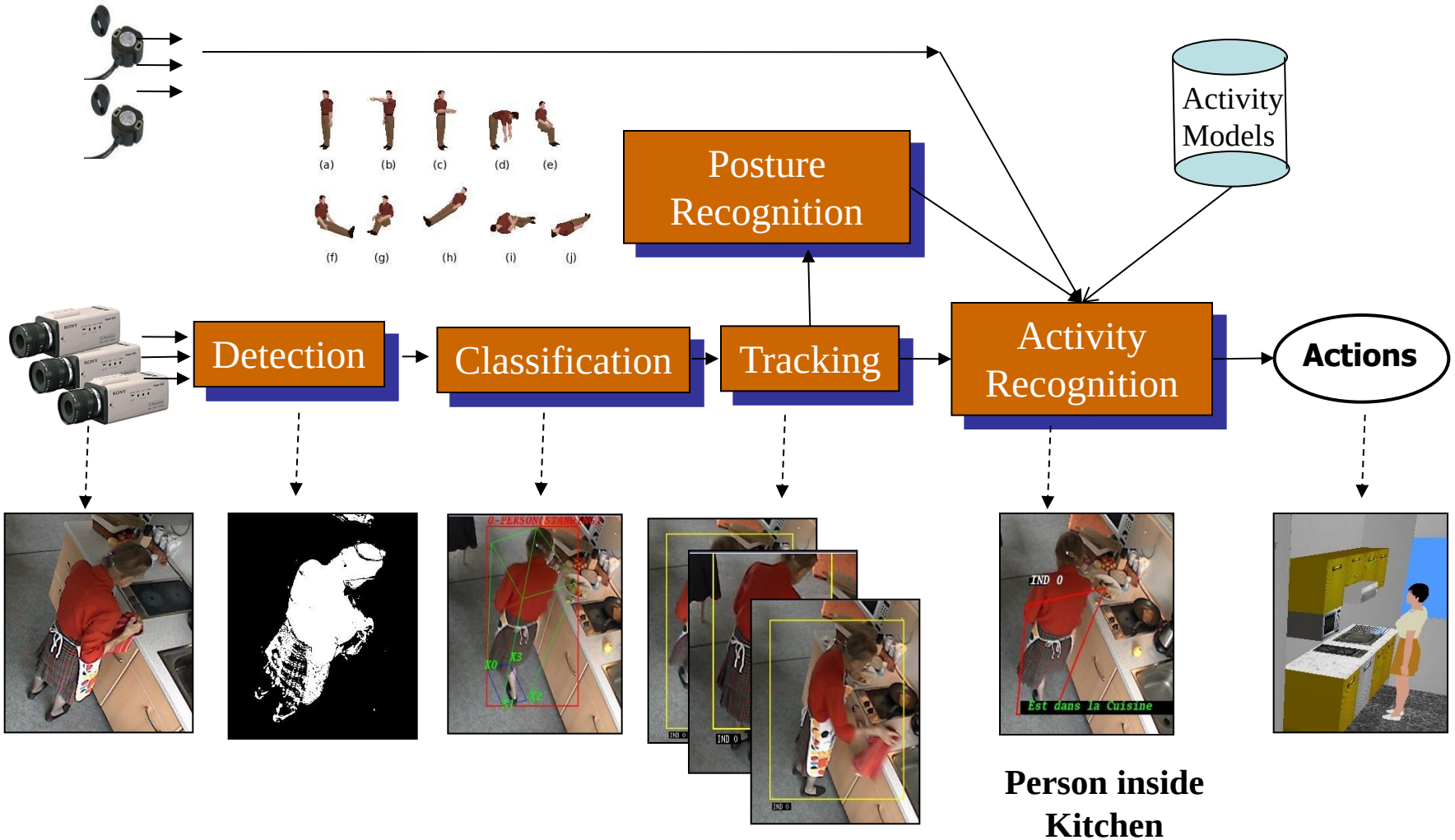
Scene Understanding and Assisted Living

Outline:

- Knowledge Representation : Scene Model
- Input of the Event Recognition process:
  - Object Detection, Tracking, Posture Recognition
- Action Recognition : Bag of Words
- Event Recognition : Knowledge
- Applications: Clinical experimentation for assisted living:
  - GerHome [2003 – 2009] : studies of older people frailty
  - Sweet-Home - CIU [2009 – 2012] : assessment of dementia
  - Dem@Care – Az@game [2012 – 2015] : autonomy of older people at home, in nursery home

- Learning Scenario Models

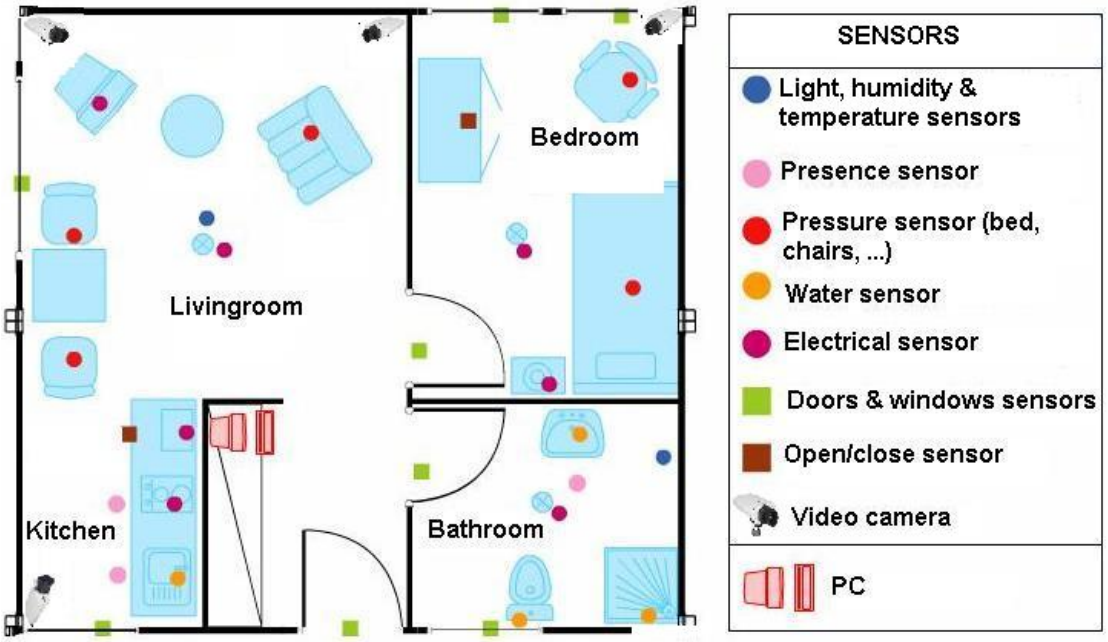
Approach: Real time semantic activity understanding

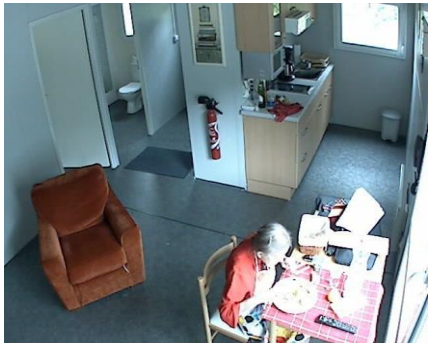


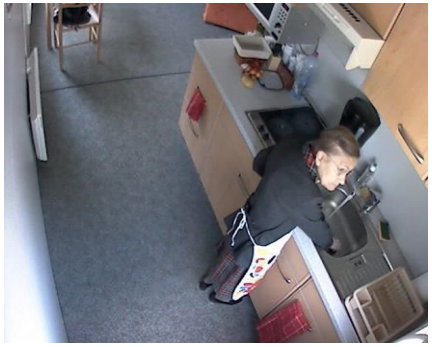
Person inside Kitchen

Knowledge Representation : 3D Scene Model

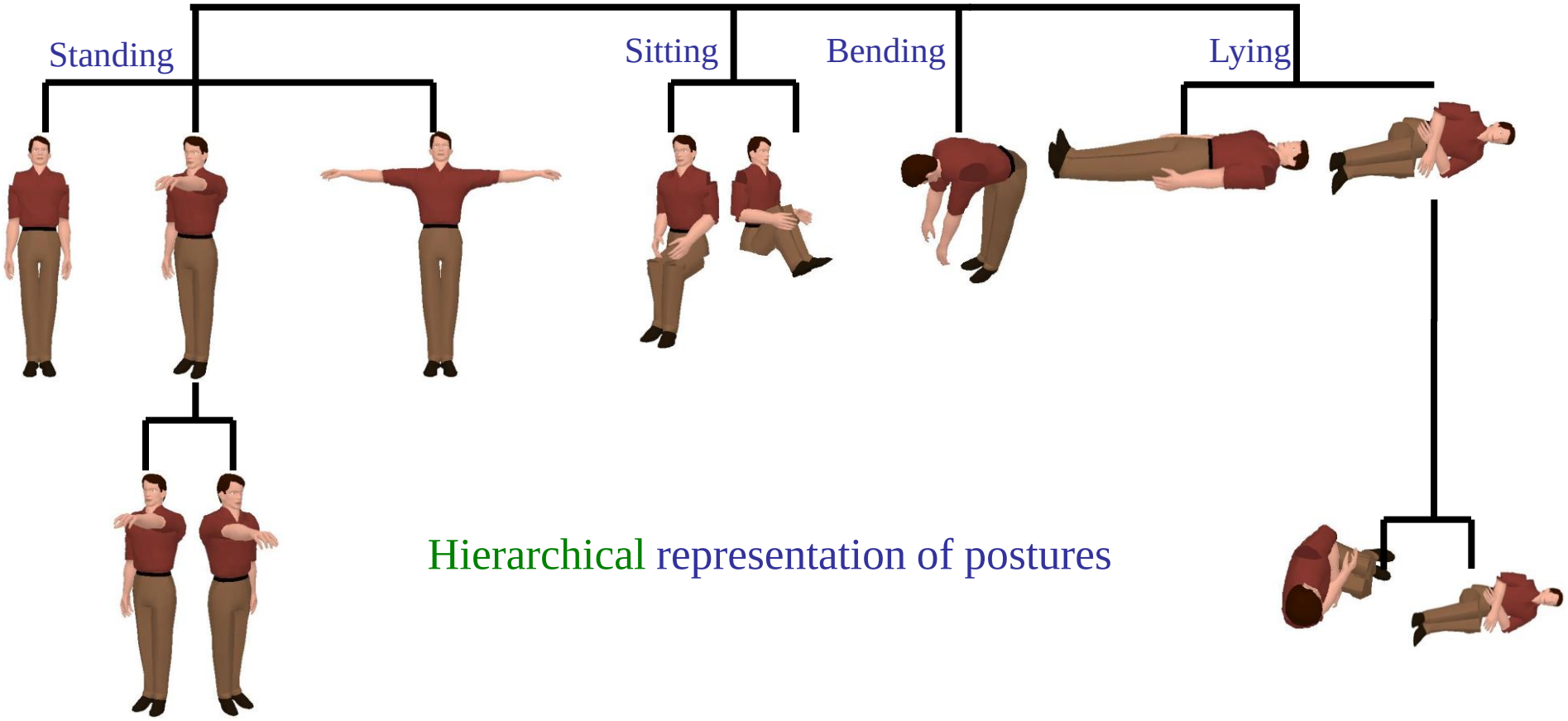
GerHome Project







Posture Recognition : Set of Specific Postures



Hierarchical representation of postures

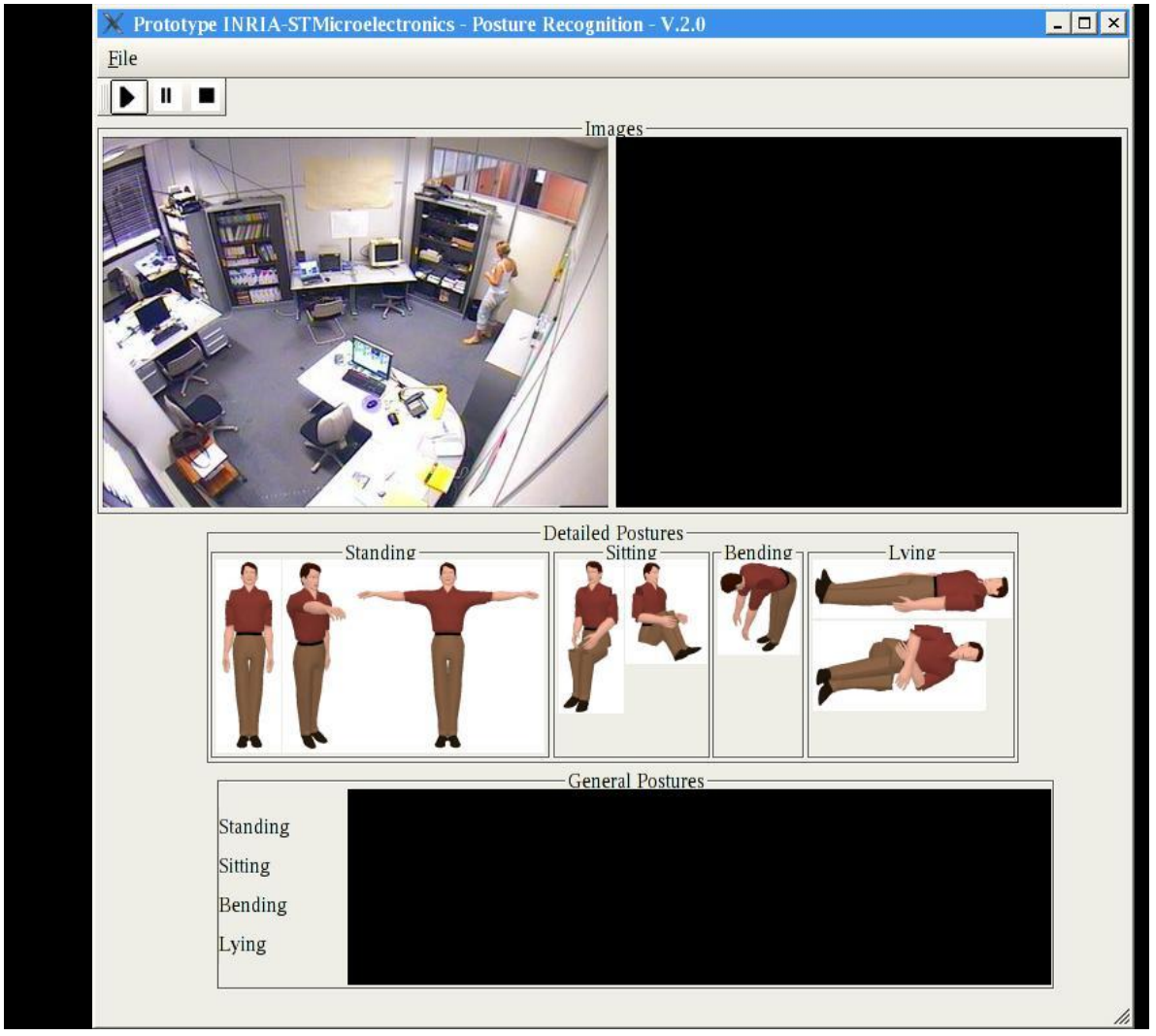
Posture Recognition : silhouette comparison

Real world

Virtual world

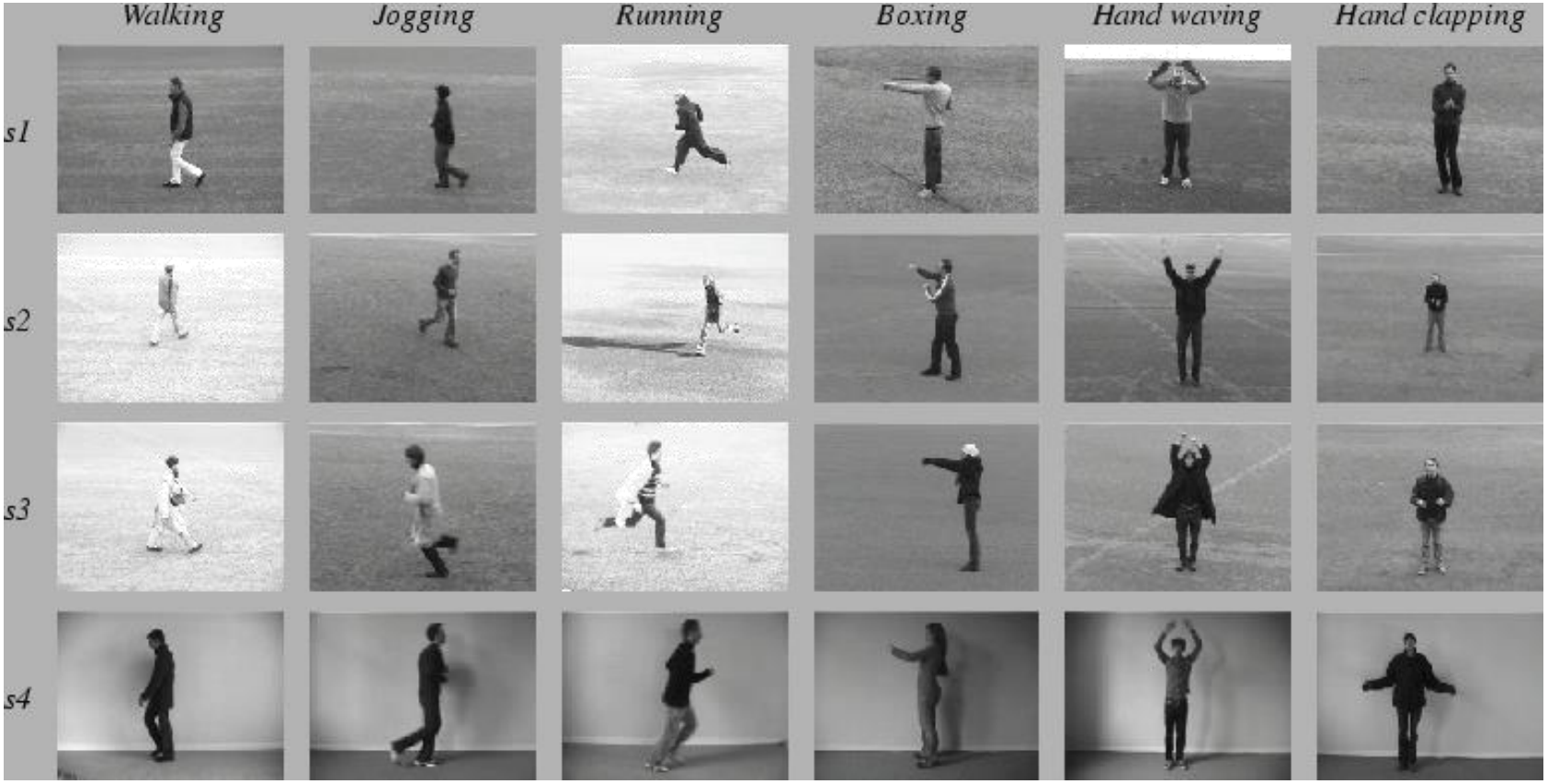
Generated silhouettes

Posture Recognition : results



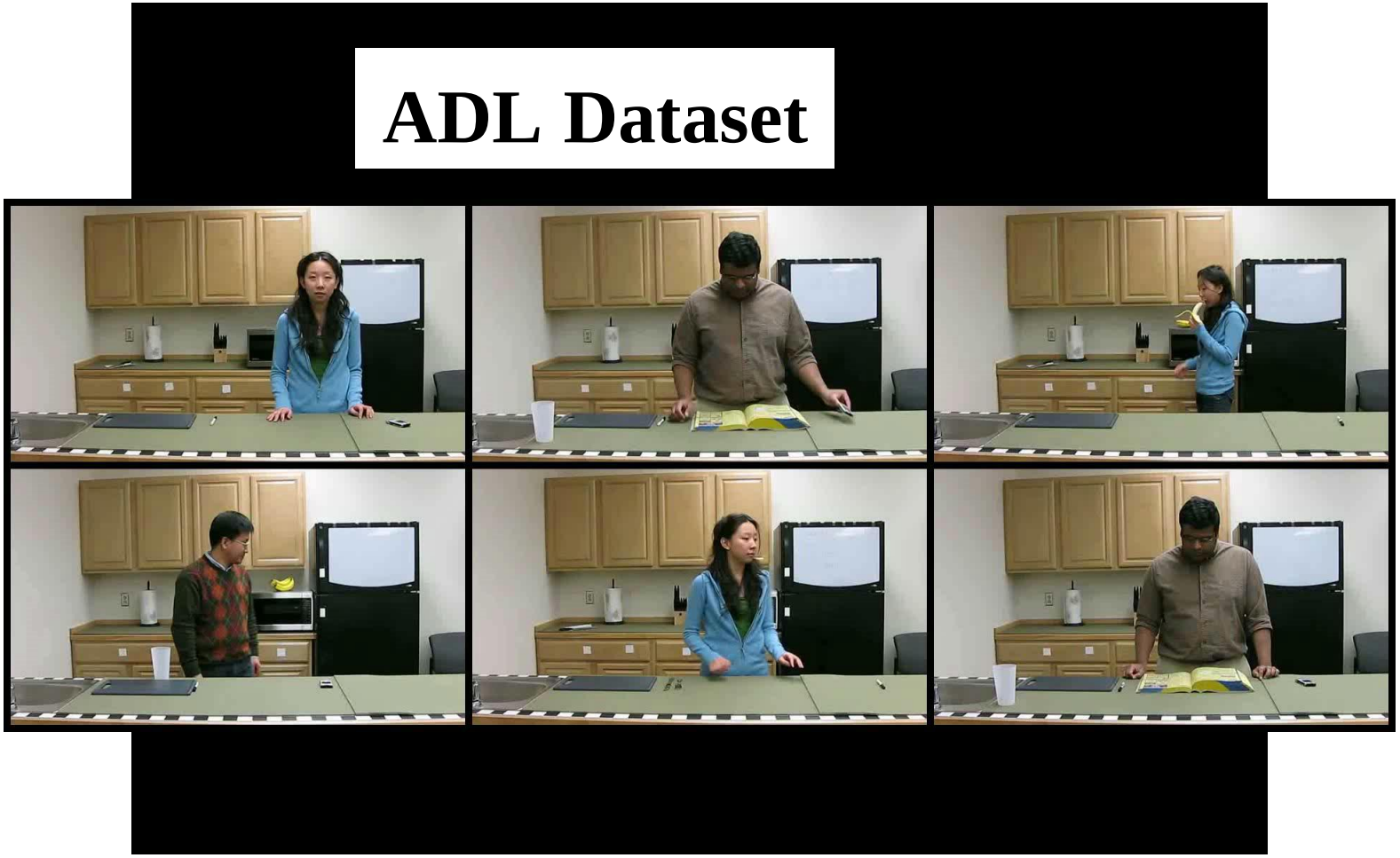
Posture Recognition : automaton

Action Recognition (MB. Kaaniche, P. Bilinski)



Type of gestures and actions to recognize

Action Recognition (P. Bilinski)

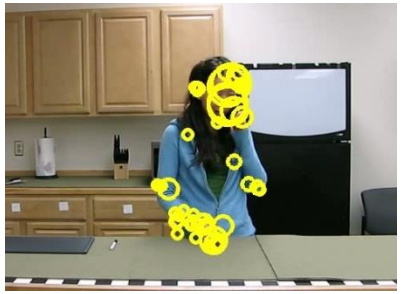


Type of gestures and actions to recognize

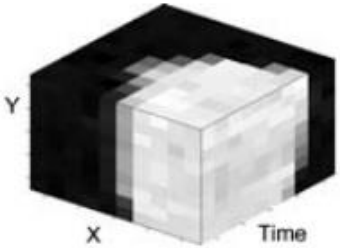
Action Recognition Algorithms

Videos

Point detector



Point descriptor



All feature vectors

BOW model

Codebook generation

Bag-of-words model

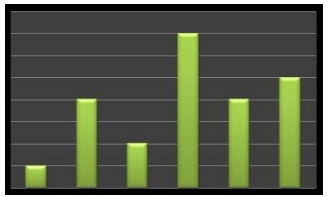
Offline Learning:

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

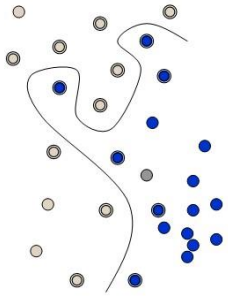
Online recognition:

Non-linear SVM

All testing feature vectors → Assignment to the closest codeword →



Histogram of codewords →



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]

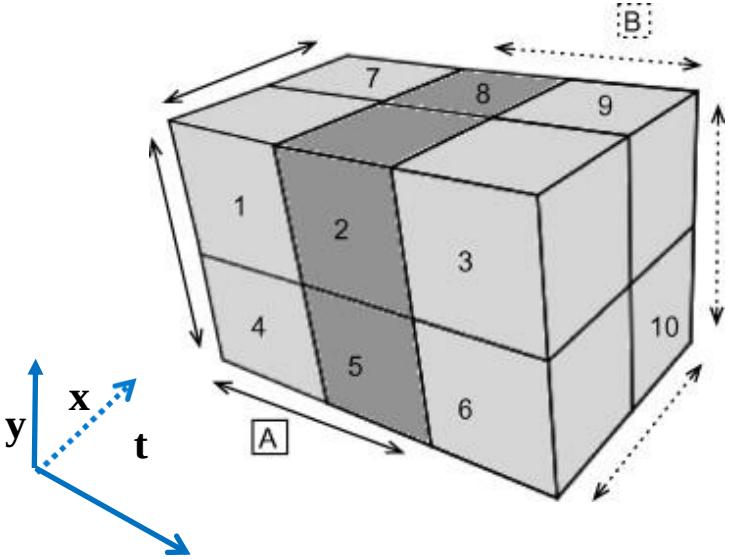
Stipx2 and 1 Person out validation = test

Beyond BOW (Piotr BILINSKI)

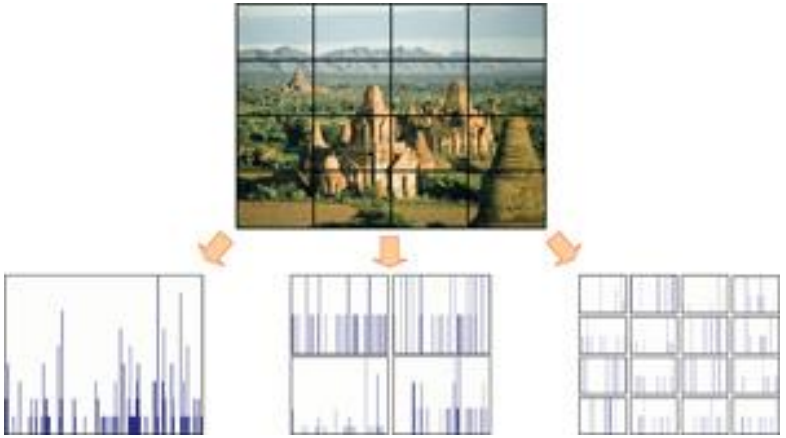
- The BOW model does not consider spatial and temporal information about the features

- It ignores:
  - Relative position of features
  - Local density of features
  - Local pairwise relationships among the features
  - Information about the space-time order of features

BOW – limitations – SOA solutions

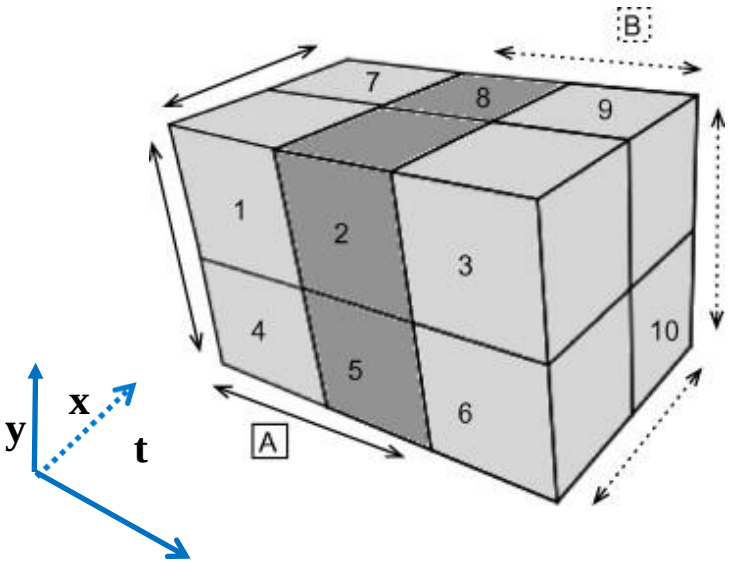


Spatio-temporal grids

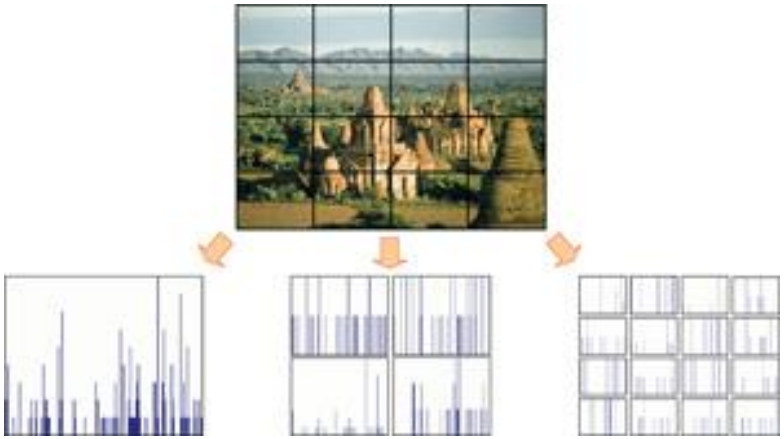


Multi-scale pyramids

BOW – limitations – SOA solutions



Spatio-temporal grids

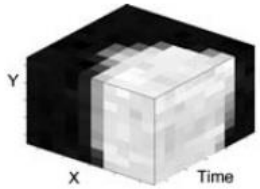


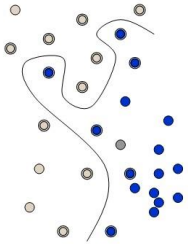
Multi-scale pyramids

These methods are still limited in terms of detail description providing only a coarse representation

Goal: Overcoming BOW limitations

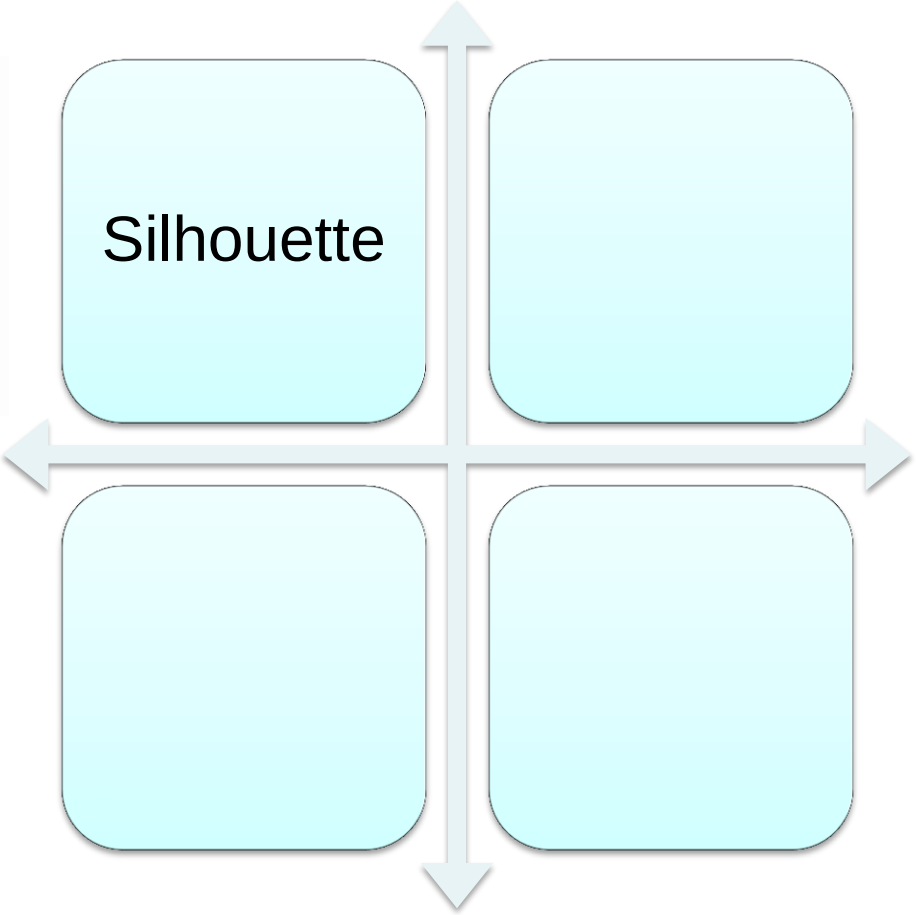
Feature Extraction → Action Modeling → Action Classification





Solution: designing complex features (e.g. contextual features) encoding spatio-temporal distribution of commonly extracted features (e.g. STIP).

SOA – Common Features



SOA – Common Features

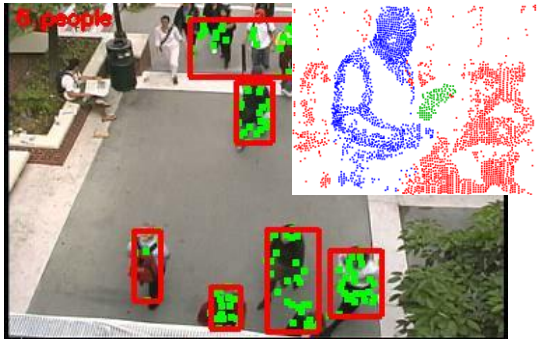
Silhouette

Require precise algorithms – difficult, especially in real-world videos

SOA – Common Features

Silhouette

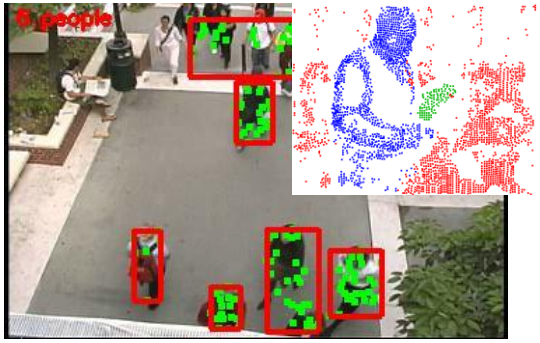
Motion Trajectories



SOA – Common Features

Silhouette

Motion Trajectories



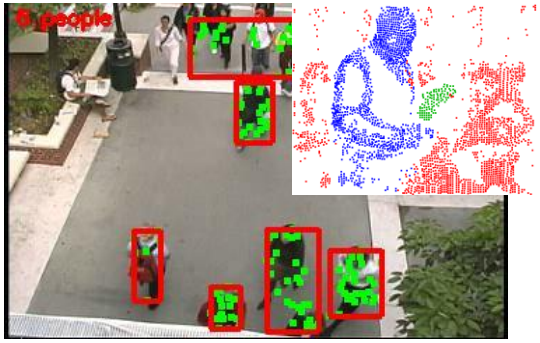
Require: feature point / object tracking

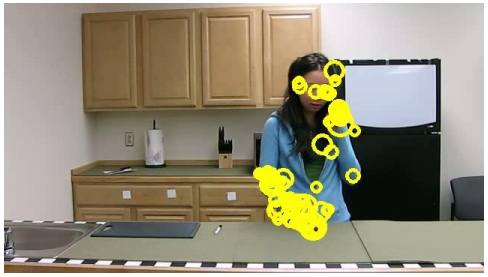
Difficult due to: illumination variability, occlusions, noise, low discriminative appearance or drifting problems

SOA – Common Features

Silhouette

Motion Trajectories

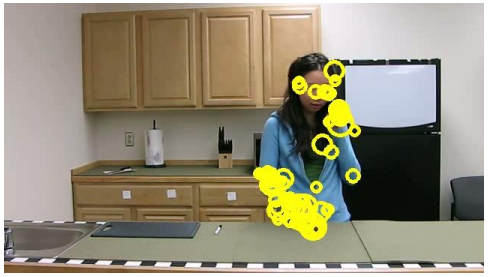




STIP

SOA – Common Features

- Popular, good performance
- Do not require segmentation, object localization and tracking
- Able to capture both visual and motion appearance
- Robust to viewpoint, scale changes and background clutter
- Fast to process
- Capturing only local spatio-temporal information

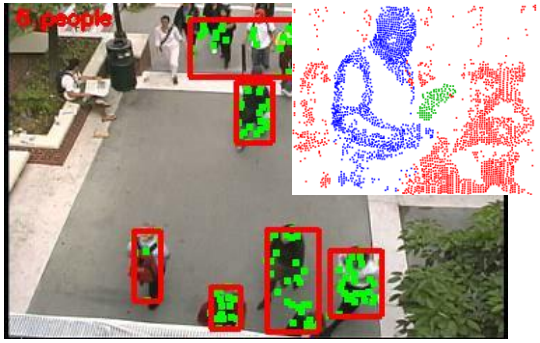


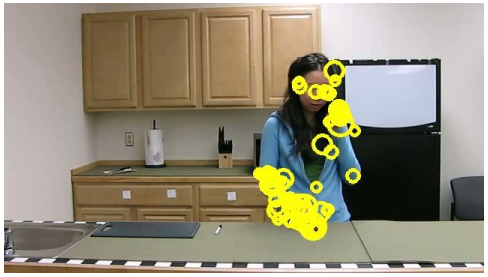
STIP

SOA – Common Features

Silhouette

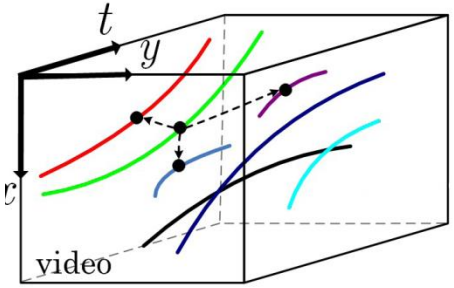
Motion Trajectories





STIP

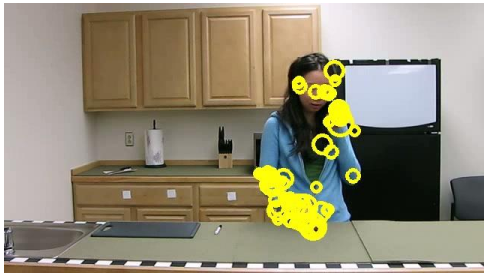
Contextual Features



SOA – Common Features

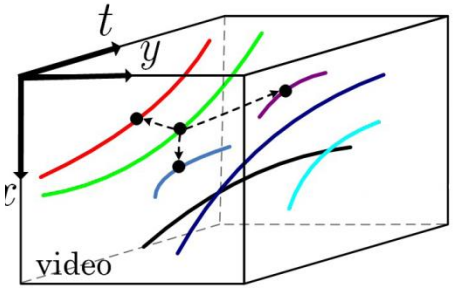
Contextual Features can:

- **enhance the discriminative power of features** (hand vs feet)
- capture scene information (goal: scene independent)
- capture human-object interactions (goal: no object detector)
- **capture figure-centric statistics** – to capture spatio-temporal distribution of features on larger neighbourhood (to overcome the limitations of the BOW)

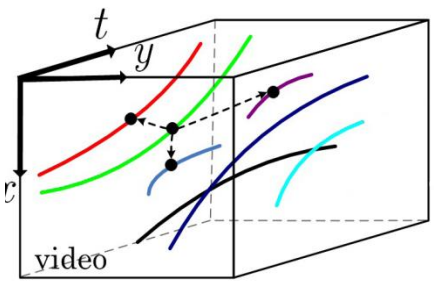


STIP

Contextual Features



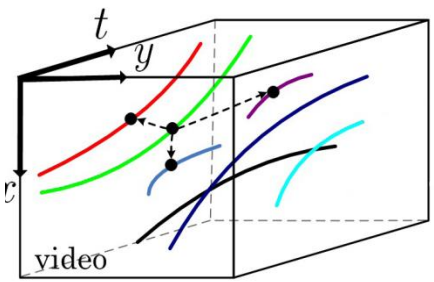
SOA – CF – Local figure-centric stat.



Capture types of trajectories

Sun et al. ‘09

SOA – CF – Local figure-centric stat.

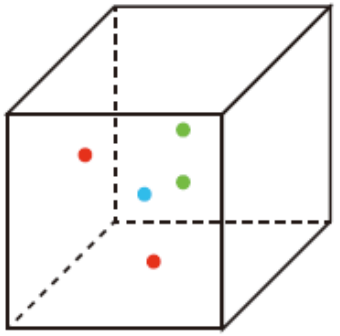


Capture types of trajectories

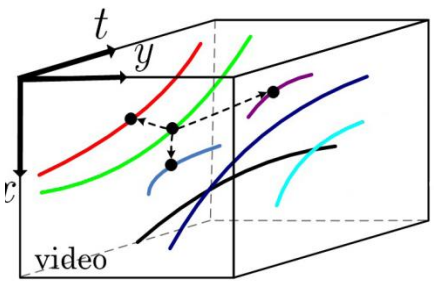
Sun et al. ‘09

Capture types of STIP

Wang et al. ‘11



SOA – CF – Local figure-centric stat.

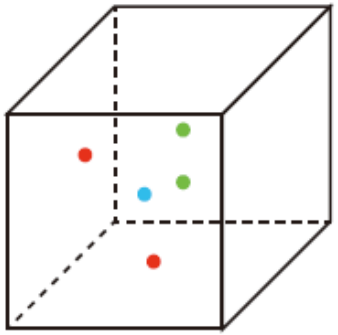


Capture types of trajectories

Sun et al. '09

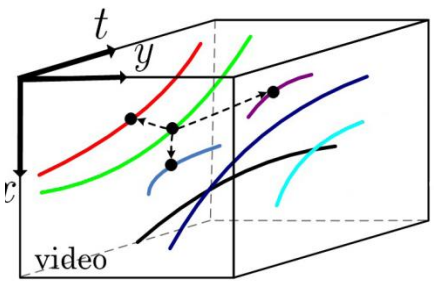
Capture types of STIP

Wang et al. '11



Capture local statistics (+)
Ignore local pairwise relationships between features (-)

SOA – CF – Local figure-centric stat.

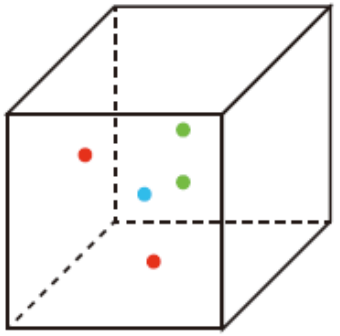


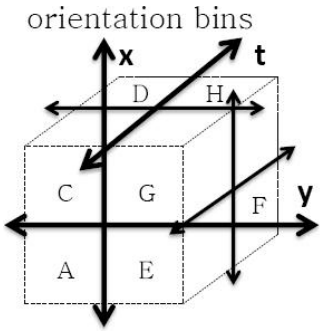
Capture types of trajectories

Sun et al. ‘09

Capture types of STIP

Wang et al. ‘11

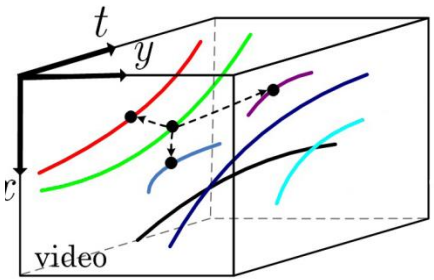




Capture orientation between STIP

Kovashka et al. '10

SOA – CF – Local figure-centric stat.

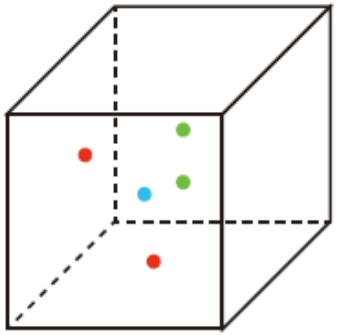


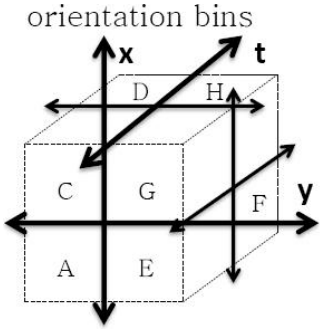
Capture types of trajectories

Sun et al. ‘09

Capture types of STIP

Wang et al. ‘11



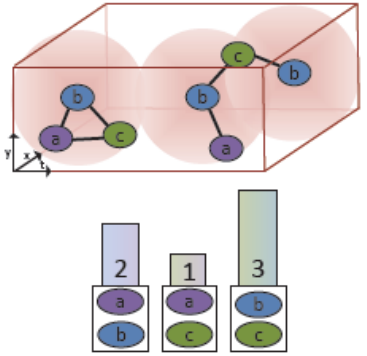


Capture orientation between STIP

Kovashka et al. '10

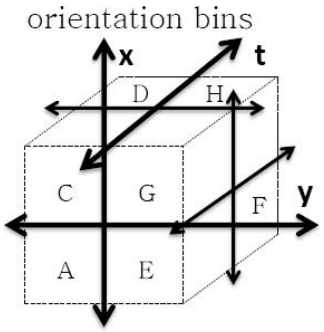
Capture pairwise co-occurrening STIP

Banerjee et al. '11



SOA – CF – Local figure-centric stat.

Capture local statistics (+)
Capture relationships between local features (+)
Limited ignoring spatio-temporal order of features (-)

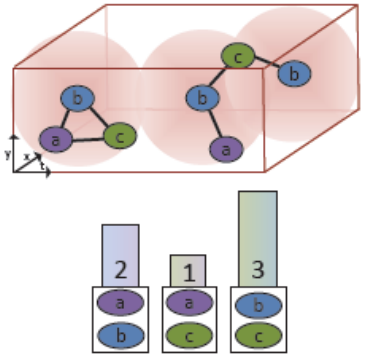


Capture orientation between STIP

Kovashka et al. '10

Capture pairwise co-occurrening STIP

Banerjee et al. '11



Conclusion on SOA limitations

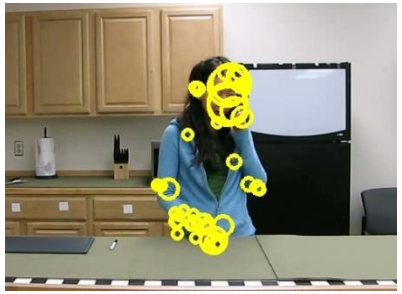
- Recent methods:
  - have focused on capturing global and local statistics of features
  - mostly ignore relations between the features
    - especially, spatio-temporal order of features

- Our goal is to propose a novel representation of CF:
  - overcoming limitations of BOW, i.e. capturing:
    - Global statistics of features
    - Local statistics of features
    - Pairwise relationship between features
    - Order of local features
  - to enhance the discriminative power of features and improve action recognition performance

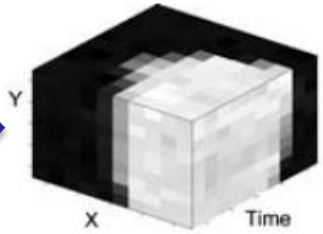
Contextual Statistics of Space-Time Ordered Features for Human Action Recognition (Piotr BILINSKI)

Videos

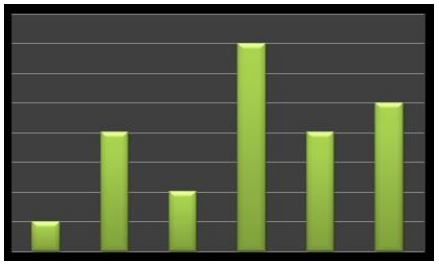
Feature detector



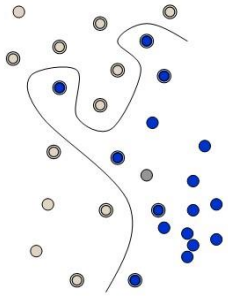
Feature descriptor



BOW model

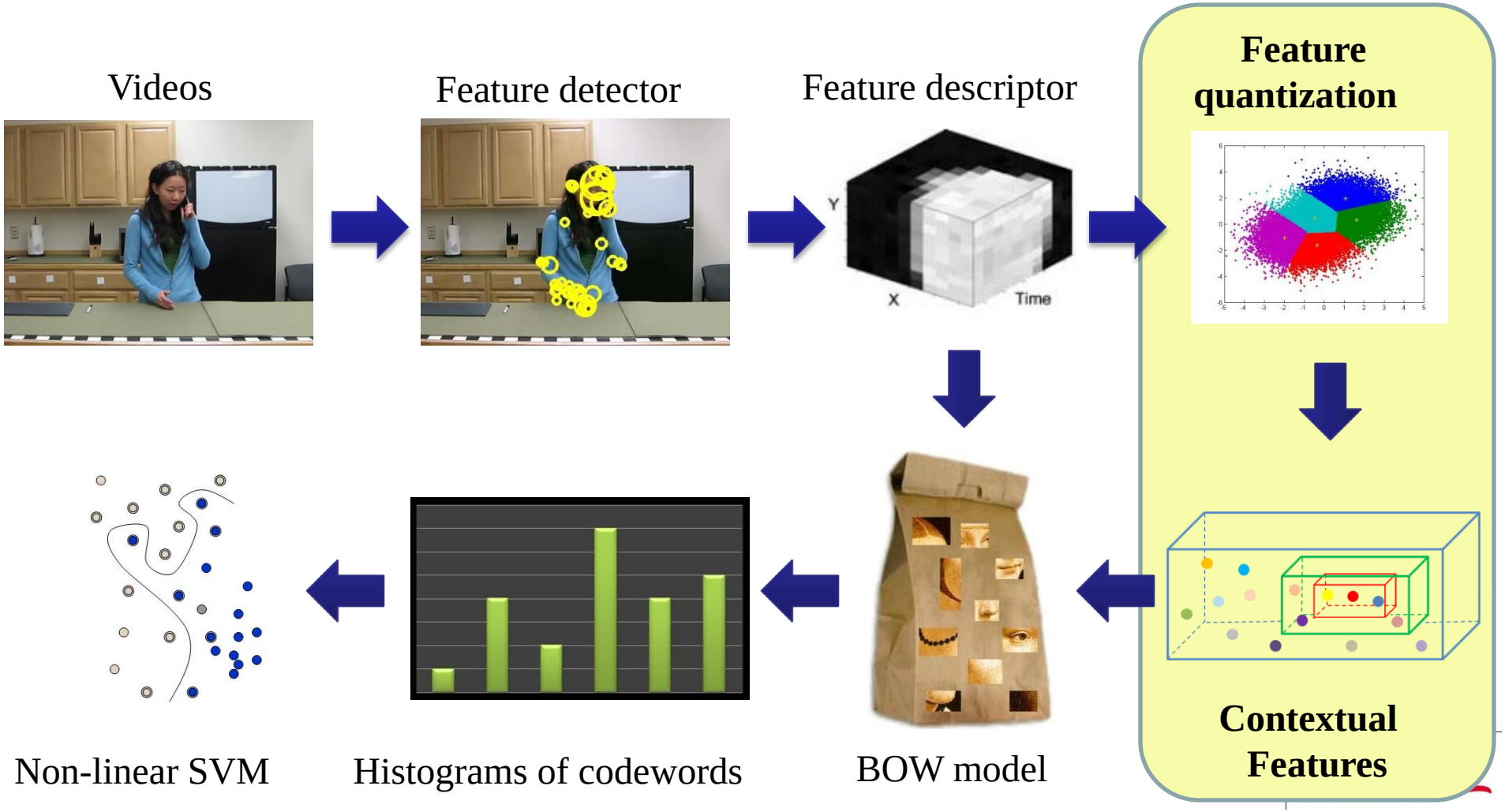


Histograms of codewords



Non-linear SVM

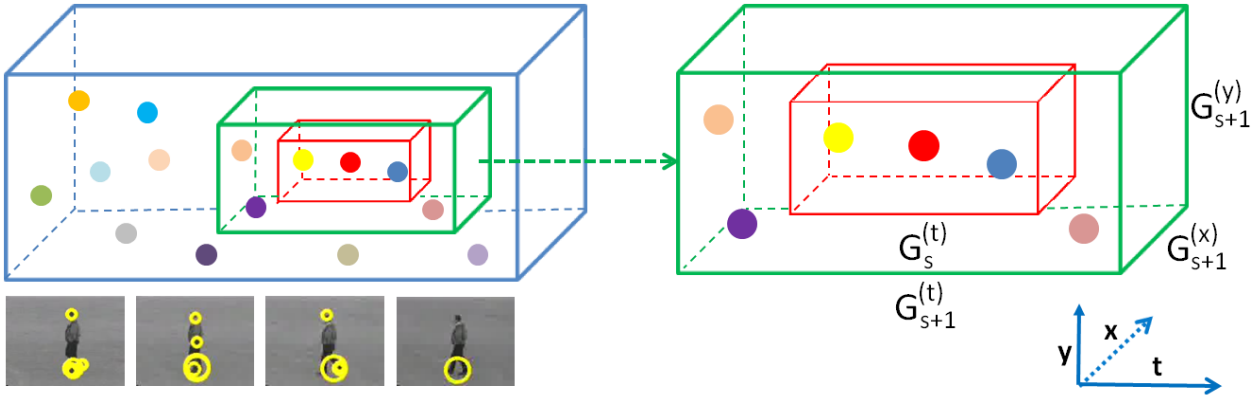
Overview of our approach



Contextual Features

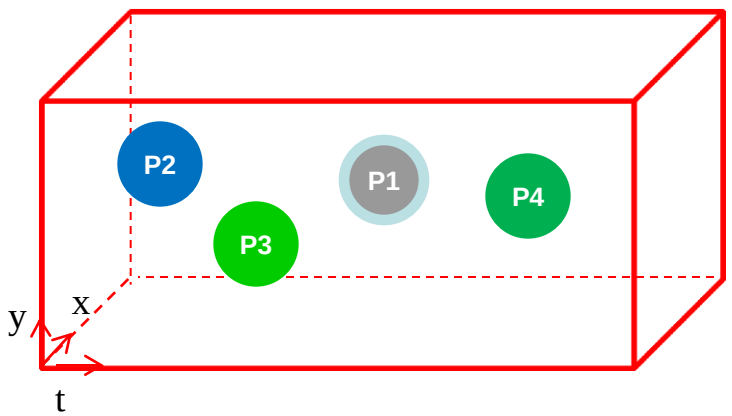
Video

Quantized local features
(features assigned to visual worlds)



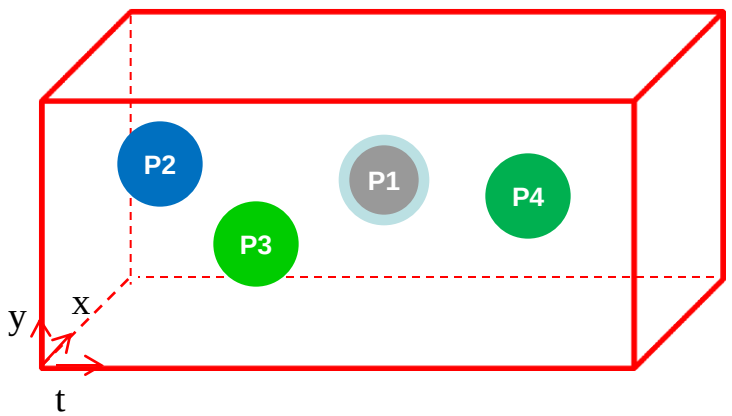
Multi-scale figure-centric neighbourhoods

Space-Time Ordered CF



$$\mathfrak{M}_{3,s}^{(F)} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Space-Time Ordered CF

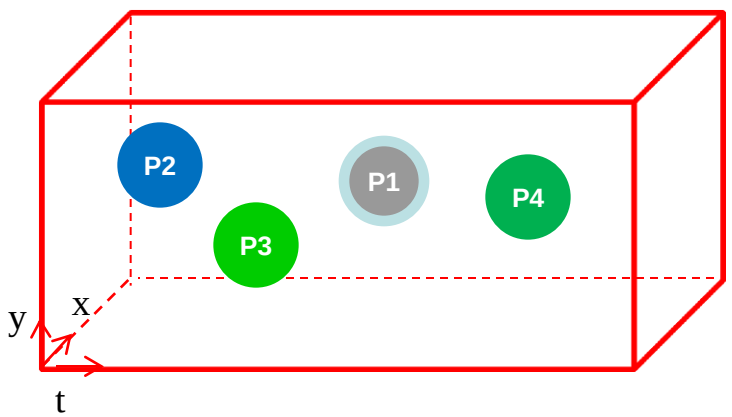


$$\mathfrak{M}_{3,s}^{(F)} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Problems:

- Features have different size
- Big feature vector (e.g. 10^3 features in video, 10^6 DB)

Space-Time Ordered CF

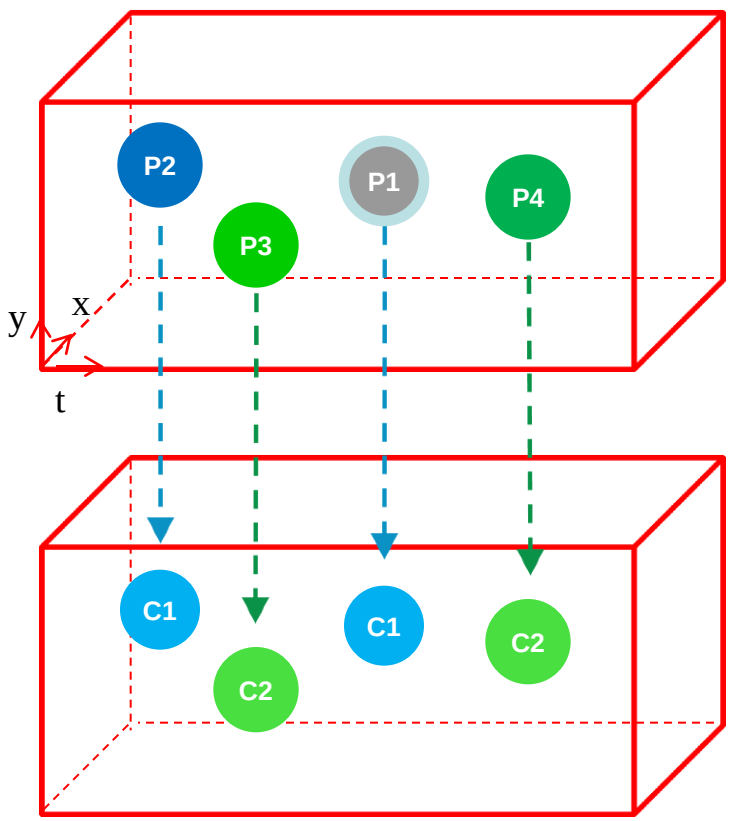


$$\mathfrak{M}_{3,s}^{(F)} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Solution – use point to codebook mapping and represent the same matrix using visual codewords:

- Features will be of equal size
- We will decrease the size of the CF features

Space-Time Ordered CF

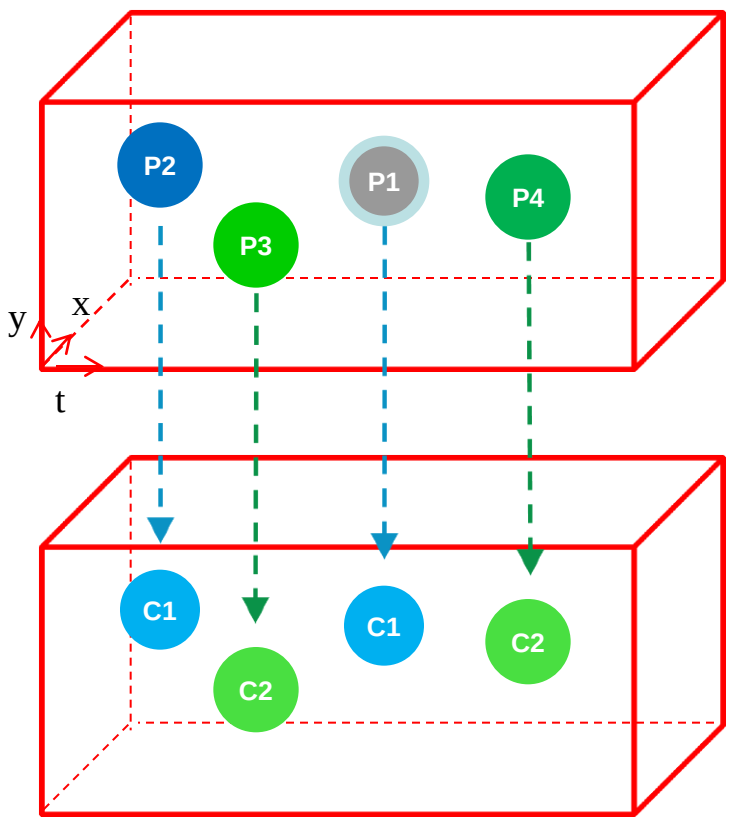


$$\mathfrak{M}_{3,s}^{(F)} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

using point to codebook mapping

$$\mathfrak{M}_{3,s}^{(C)} = \begin{array}{c} \\ C_1 \\ C_2 \end{array} \begin{array}{c} \begin{array}{cc} C_1 & C_2 \end{array} \\ \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \end{array}$$

Space-Time Ordered CF



$$\mathfrak{M}_{3,s}^{(F)} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

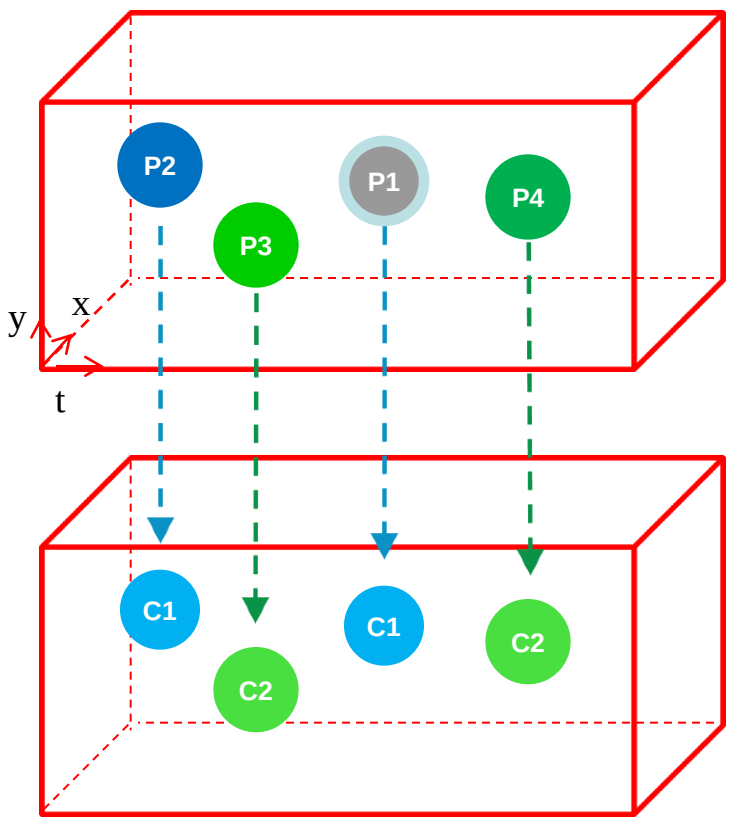
using point to codebook mapping

$$\mathfrak{M}_{3,s}^{(C)} = \begin{array}{c} \\ C_1 \\ C_2 \end{array} \begin{array}{c} \begin{array}{cc} C_1 & C_2 \end{array} \\ \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \end{array}$$

reshape to a single dim. vector

$$\mathfrak{CF}_{3,s} = [1, 3, 1, 1]^T$$

Space-Time Ordered CF



$$\mathfrak{M}_{3,s}^{(F)} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

using point to codebook mapping

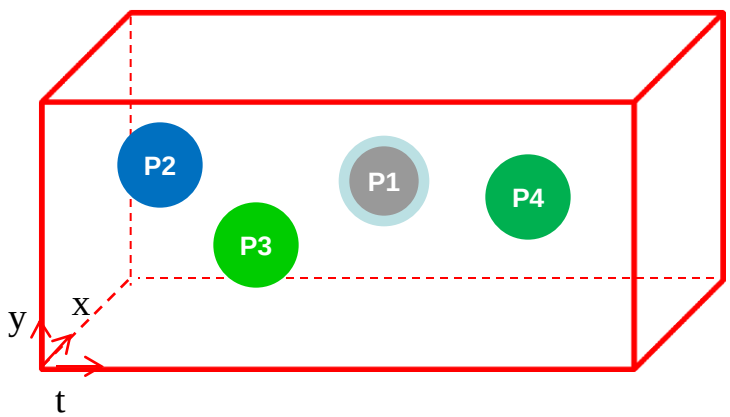
$$\mathfrak{M}_{3,s}^{(C)} = \begin{array}{c} \\ C_1 \\ C_2 \end{array} \begin{array}{c} \begin{array}{cc} C_1 & C_2 \end{array} \\ \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \end{array}$$

reshape to a single dim. vector

$$\mathfrak{CF}_{3,s} = [1, 3, 1, 1]^T$$

PCA/LDA

Space-Time Ordered CF



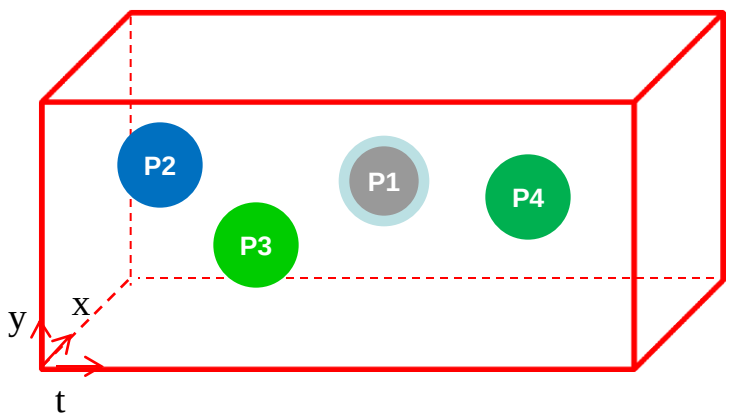
$$\mathfrak{M}_{3,s}^{(F)} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Good results can be obtained with small codebook size:

- Banerjee et al. AVSS'11 – obtain SOA results using small codebook size
- Our experiments on the WEIZMANN dataset:

20	25	30	35	40	50	60	70	80	100	1000
66.7	**71.1**	70	**74.4**	83.3	80	84.4	**87.8**	82.2	83.3	**87.8**

Space-Time Ordered CF



$$\mathfrak{M}_{3,s}^{(F)} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & P_3 & P_4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Good results can be obtained with small codebook size:

- Banerjee et al. AVSS'11 – obtain SOA results using small codebook size
- Our experiments on the WEIZMANN dataset:

20	25	30	35	40	50	60	70	80	100	1000
66.7	**71.1**	70	**74.4**	83.3	80	84.4	**87.8**	82.2	83.3	**87.8**

To reduce the processing time we use small codebooks

KTH - Results

s1	s2	s3	s4	**s1-s4**
98.67%	95.33%	92.62%	98.00%	**96.16%**

s1 – outdoors, s2 – outdoors with scale variation
s3 – outdoors with different clothes, s4 – indoors

Method	Year	Accuracy
Liu *et al.* [21]	2009	93.8%
Wu *et al.* [37]	2011	94.5%
Kim *et al.* [12]	2007	95.33%
Wu *et al.* [36]	2011	95.7%
Lin *et al.* [10]	2011	95.77%
Our method		**96.16%**

ADL - Results

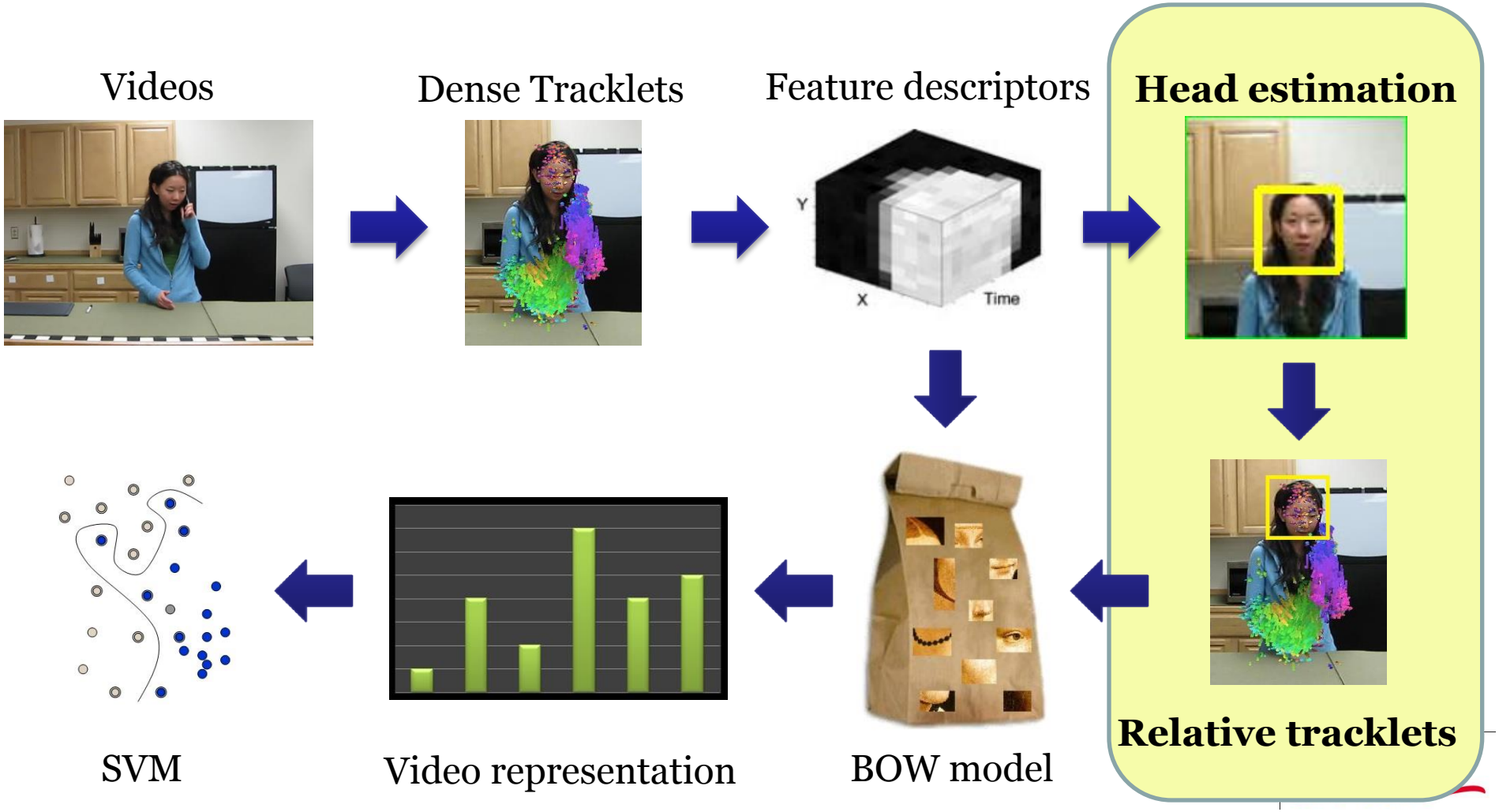
Method	Year	Recognition Rate (%)
Matikainen *et al.* [24]	2010	70%
Satkin *et al.* [29]	2010	80%
Banabbas *et al.* [4]	2010	81%
Raptis *et al.* [28]	2010	82.67%
Messing *et al.* [25]	2009	89%
Wang *et al.* [34]	2011	96% (93.8% for KTH)
Our method		**93.33%**

STIPx2 and 1 Person out validation <> test

Relative Dense Tracklets for Human Action Recognition

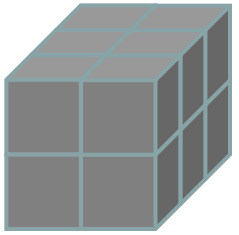
- We propose relative tracklets, introducing spatial information to the bag-of-words model.
- The tracklets computation is based on their relative positions according to the central point of our dynamic coordinate system.
- As this central point, we choose the center of a head to provide camera invariant description.
- We focus on home care applications, thus human head can be used as a reference point for computing our relative features.

Overview of the approach



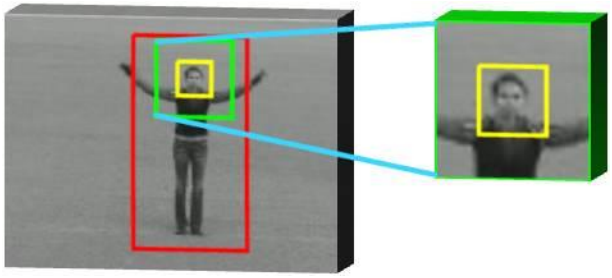
Tracklet Descriptors

- Shape Multi-Scale Tracklet (SMST) Descriptor
  - encodes a local motion pattern of a tracklet as its displacement vectors normalized by the sum of the magnitudes of these displacement vectors.
- HOG and HOF descriptors:
  - encode appearance around tracklets.
  - For each tracklet we define a grid (2×2×3).
  - For each cell of a grid we compute a histogram.
  - HOG – capture local visual appearance.
  - HOF – capture local motion appearance.

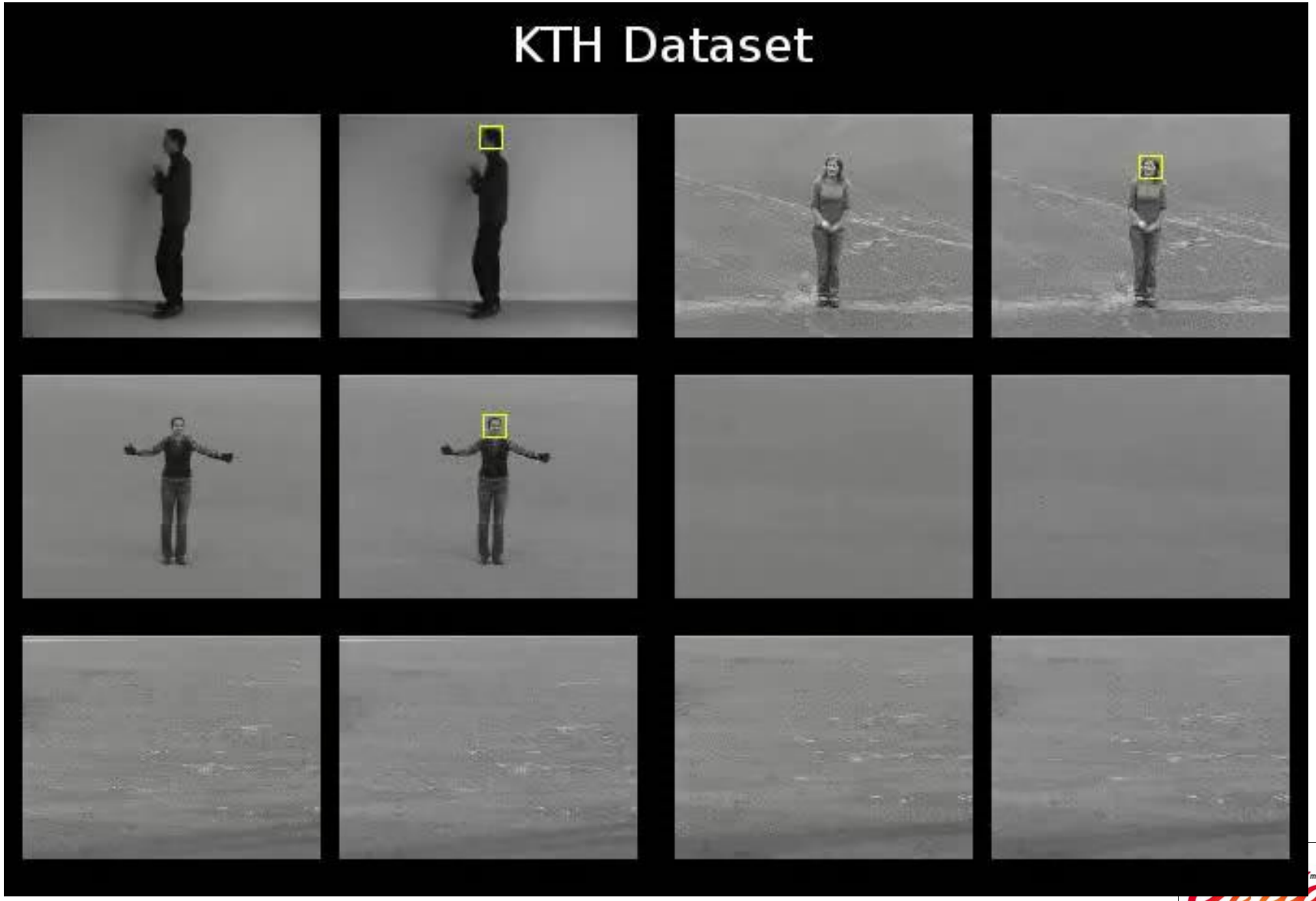


- Relative Multi-Scale Tracklet (RMST) Descriptor
  - encodes shape of a tracklet with respect to the estimated head trajectory.
- Combined Multi-Scale Tracklet (CMST) Desc.
  - Combination of SMST and RMST.

KTH Action Dataset – Results



KTH Action Dataset – Results



KTH Action Dataset – Results

Official Split:

Method	s1	s2	s3	s4	overall
SMST	98.15%	88.89%	88.89%	90.74%	91.67%
RMST	96.30%	88.89%	87.04%	92.60%	91.21%
CMST	**98.15%**	**92.59%**	**92.59%**	**96.30%**	**94.91%**

LOOCV:

Method	s1	s2	s3	s4	overall
SMST	98.00%	92.00%	93.29%	94.67%	94.49%
RMST	98.67%	89.33%	95.30%	96.67%	94.99%
CMST	**99.33%**	**93.33%**	**97.32%**	**98.67%**	**97.16%**

s1 – outdoors, s2 – outdoors with scale variation, s3 – outdoors with different clothes, s4 – indoors

KTH Action Dataset – Results

Official Split:

Method	Year	Acc.
Laptev *et al.*	2008	91.8%
Yuan *et al.*	2009	93.3%
Zhang *et al.*	2012	94.1%
Wang *et al.*	2011	94.2%
Gilbert *et al.*	2011	94.5%
Kovashka *et al.*	2010	94.53%
Becha *et al.*	2012	94.67%
Our method		**94.91%**

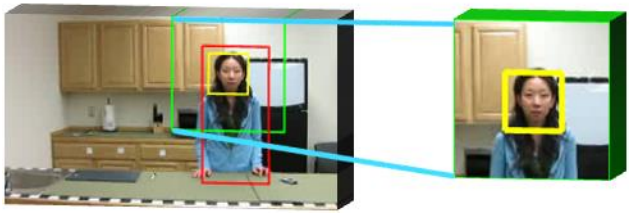
LOOCV:

Method	Year	Acc.
Liu *et al.*	2009	93.8%
Ryoo *et al.*	2009	93.8%
X. Wu *et al.*	2011	94.5%
Kim *et al.*	2007	95.33%
Zhang *et al.*	2012	95.5%
S. Wu *et al.*	2011	95.7%
Lin *et al.*	2011	95.77%
Our method		**97.16%**

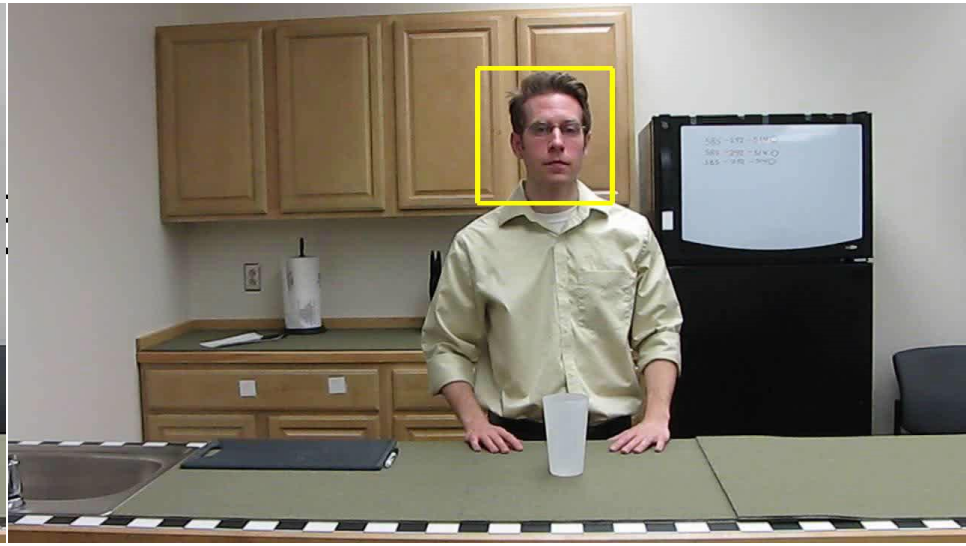
LOOCV:

Method	s1	s2	s3	s4	overall
Wu *et al.*	96.7%	91.3%	93.3%	96.7%	94.5%
Lin *et al.*	98.83%	94.00%	94.78%	95.48%	95.77%
Our method	**99.33%**	**93.33%**	**97.32%**	**98.67%**	**97.16%**

ADL Dataset
University of Rochester Activities of Daily Living – Results

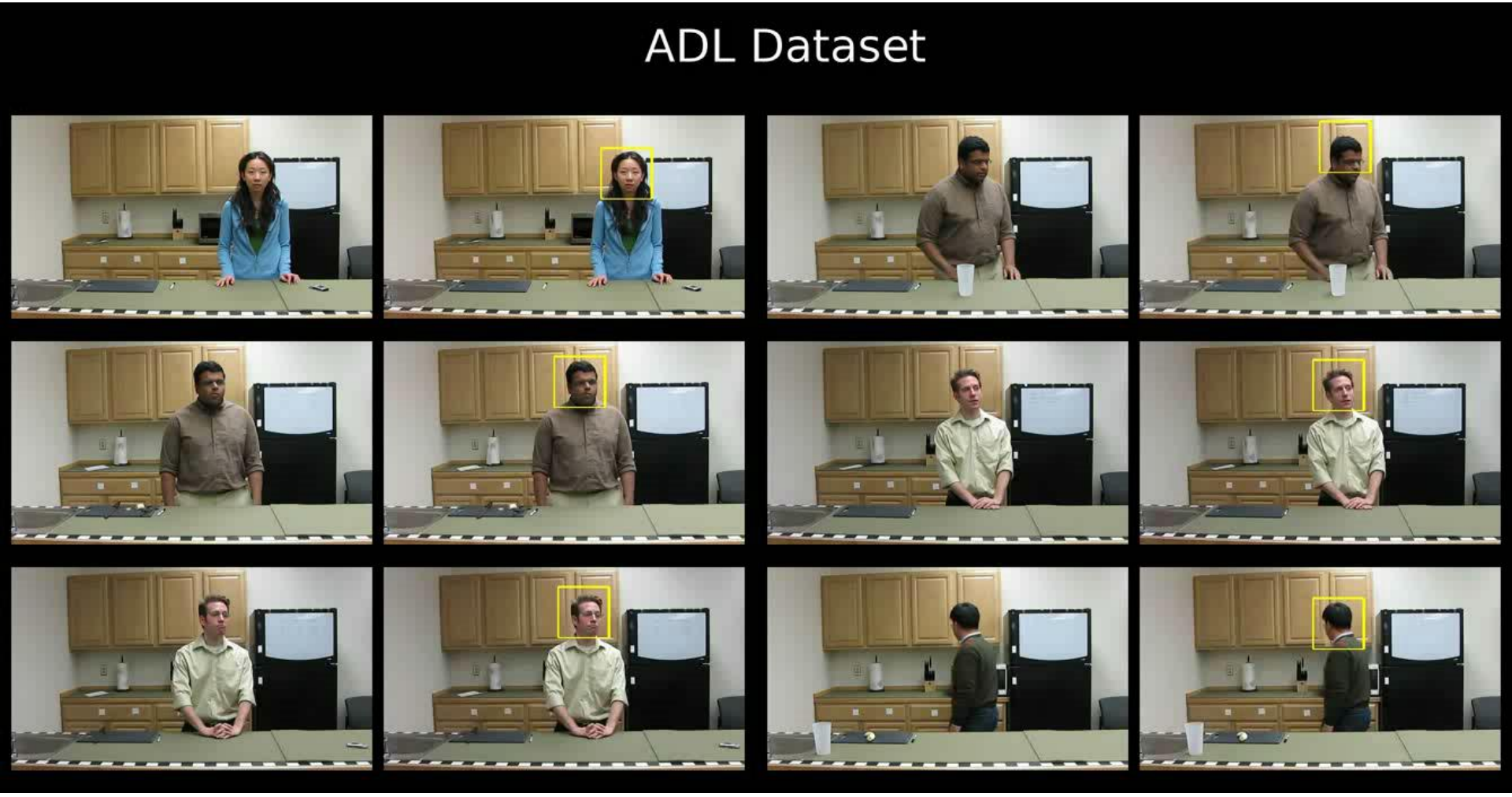


ADL Dataset – Results



ADL Dataset – Results

Action Recognition using ADL: Benchmarking video dataset



ADL Dataset – Results

Method	Accuracy
SMST	76.67%
RMST	78.67%
CMST	**88.00%**
CMST + HOG-HOF	**92%**

Method	Accuracy
Matikainen *et al.*	70%
Satkin *et al.*	80%
Banabbas *et al.*	81%
Raptis *et al.*	82.67%
Messing *et al.*	89%
Wang *et al.*	96% (93.8% for KTH)
Our method	**92%**

Head + Tracklet

Hospital Action Dataset

8 actions (semi-guided): playing cards, matching ABCD sheets of paper, reading, sitting down and standing up, turning back, standing up and moving ahead, walking back and forth.

55 patients.

Spatial resolution: 640×480.

Frame rate: 20 fps.

Challenges: different shapes, sizes, genders and ethinicities of people, occlusions, and multiple people (sometimes both patient and doctor are visible).

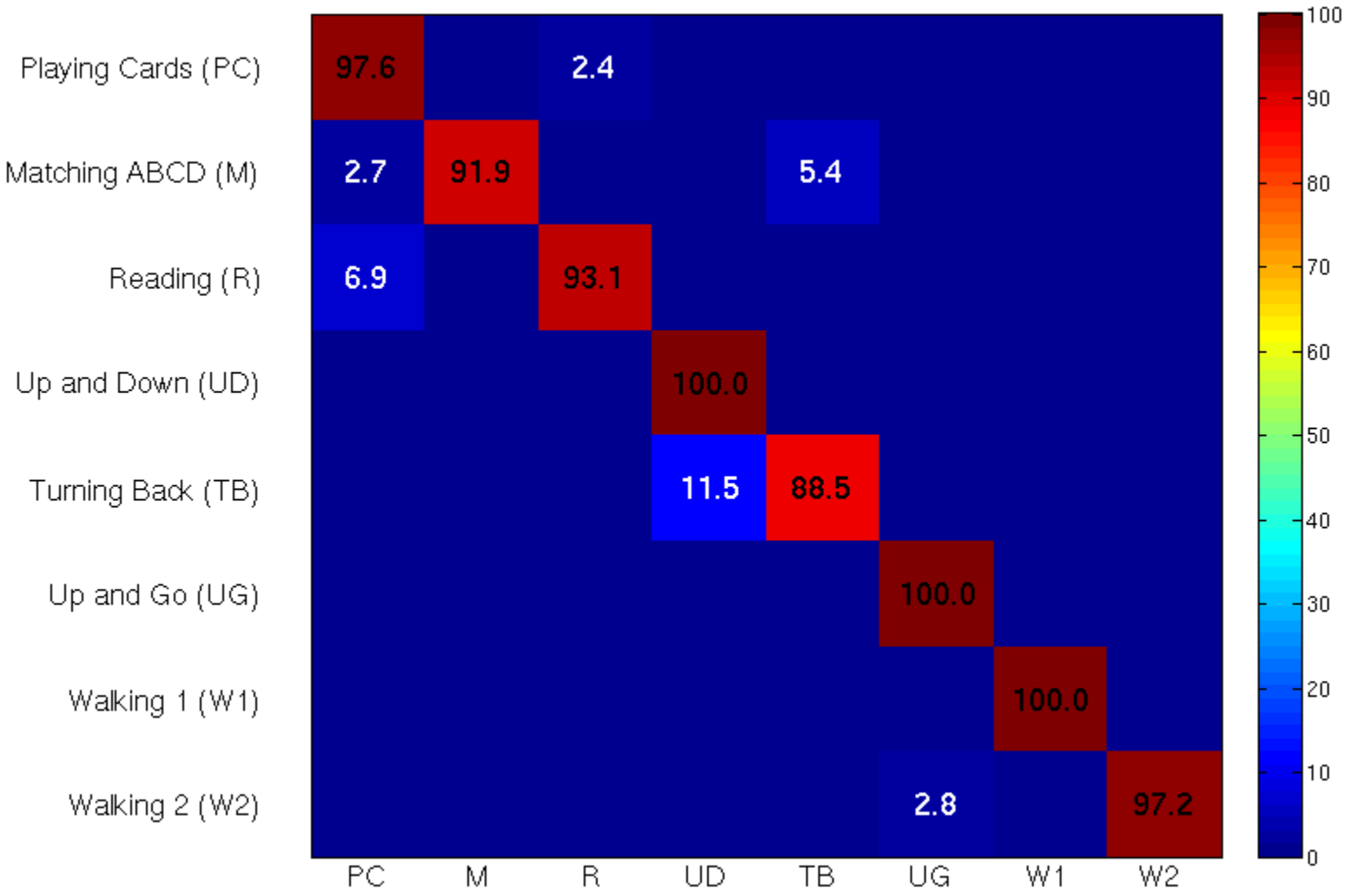
Evaluation Scheme: 5-people-fold cross-validation.

Action Recognition using Nice hospital video dataset

Hospital Action Dataset – Results

Method	Accuracy
SMST	86.3%
RMST	87.2%
CMST	**91.5%**
CMST + HOG-HOF	**93.0%**

Hospital Action Dataset – Results 1

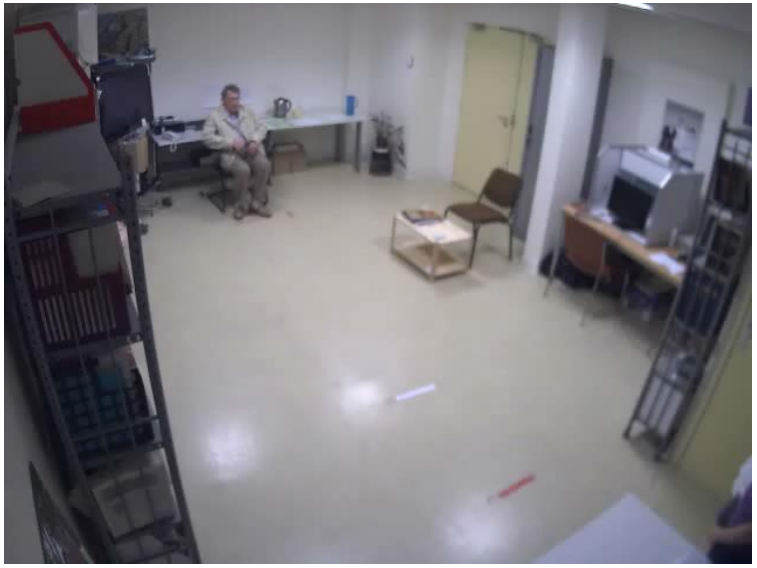


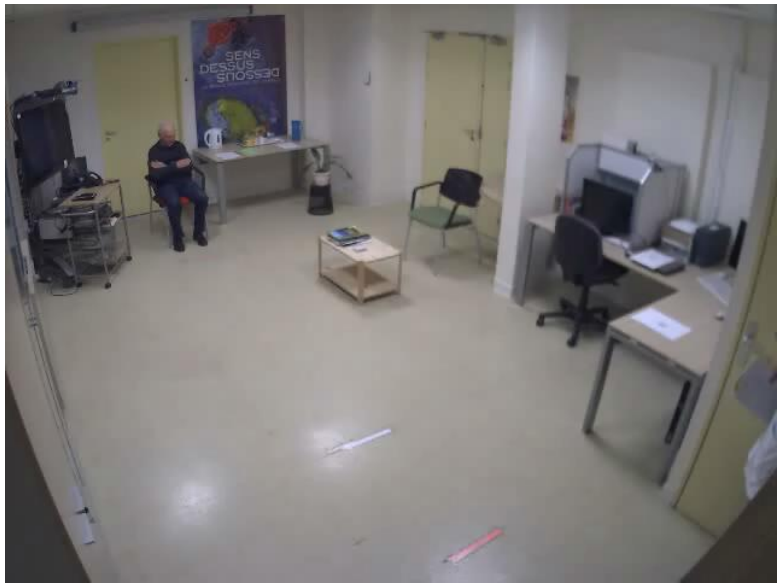
Issues in Action Recognition

- 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...

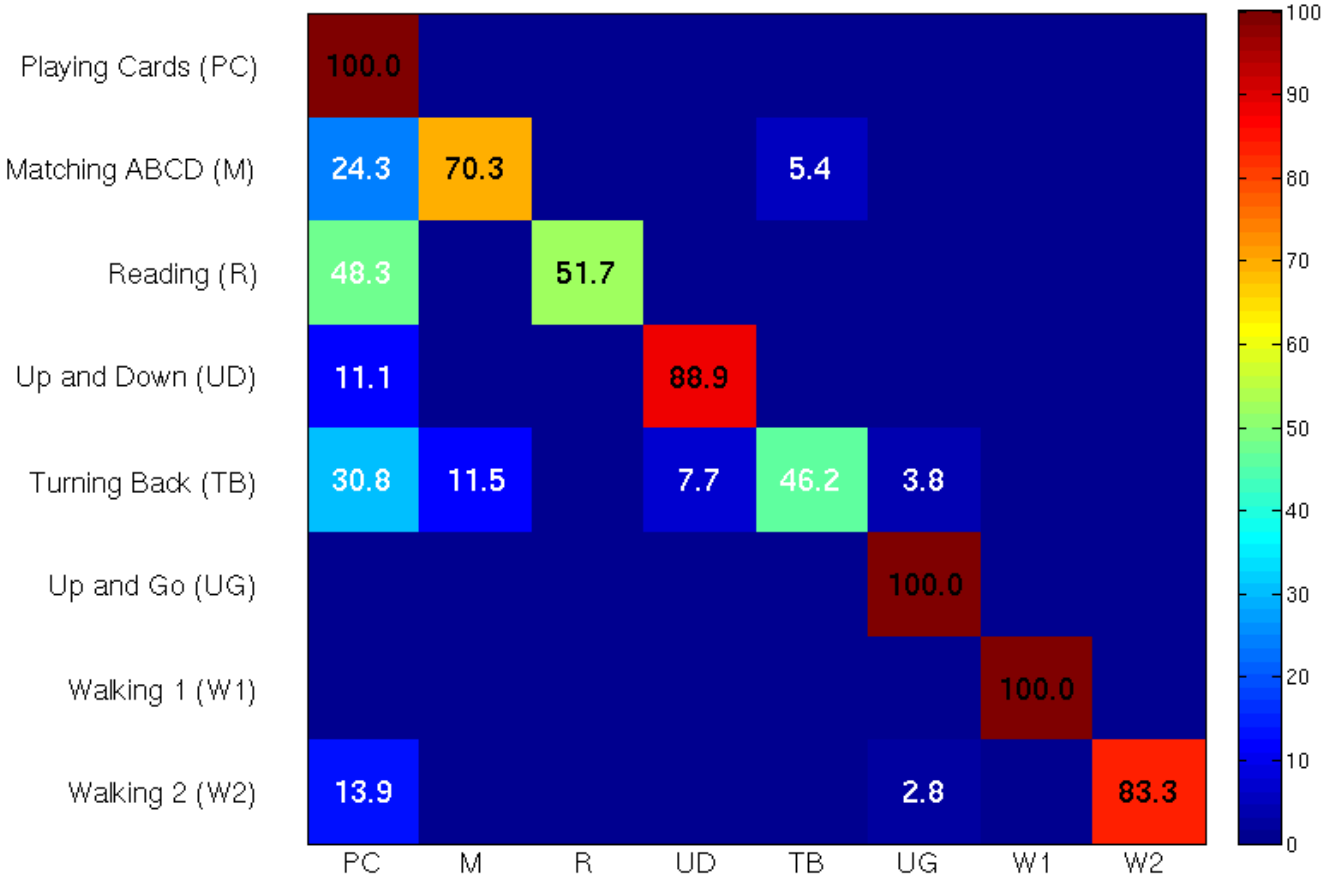
Issues in Action Recognition

- Finer actions, more discriminative (NC, MCI, AD)





Hospital Action Dataset – Results 2



AD versus NC

Playing Cards	69%
Up and Go	66%
Reading	44%

Monitoring of Activities of Daily Living

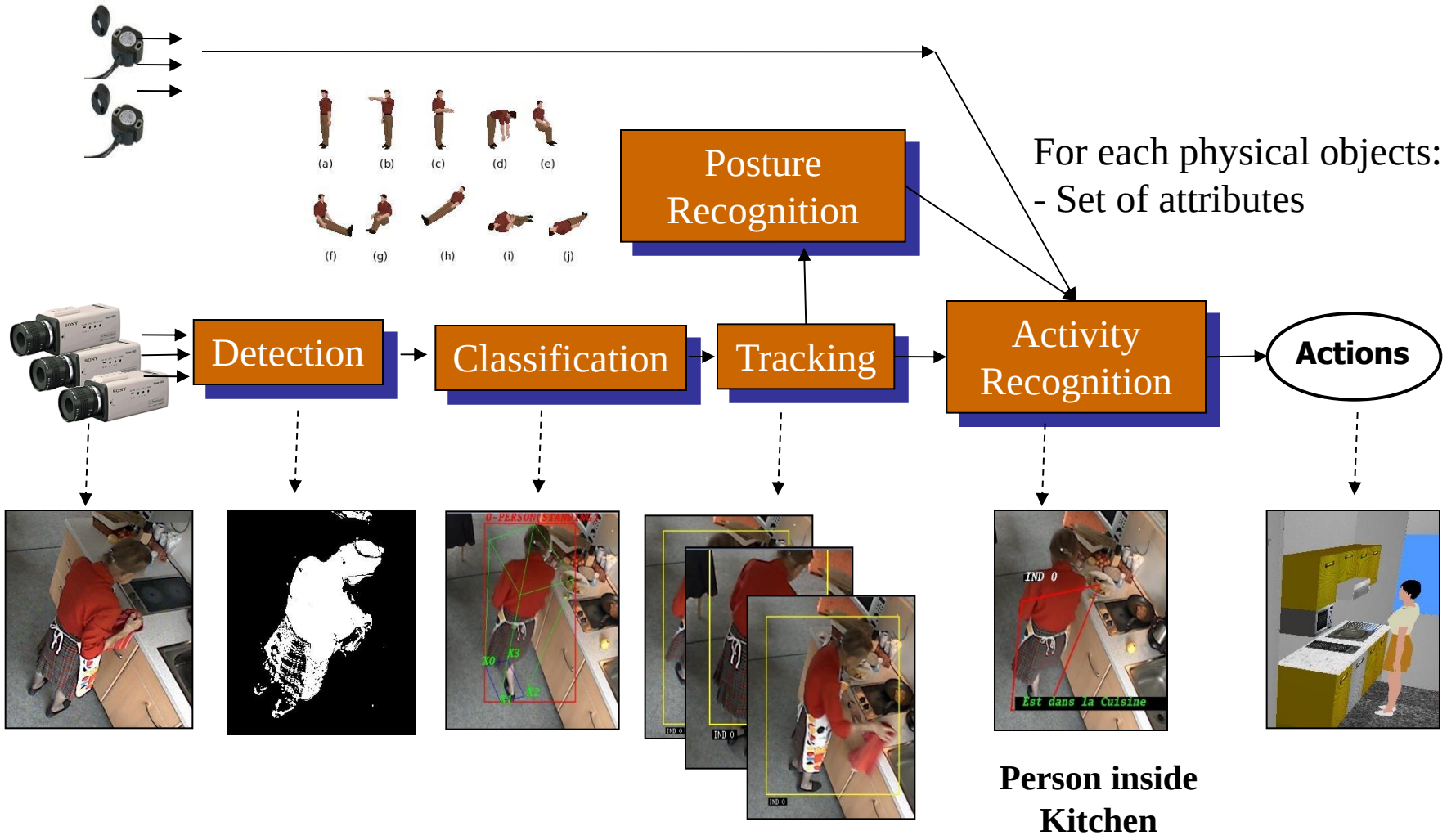
Motivation : Increase independence and quality of life:

- Enable older adults to live longer, autonomously in their preferred environment.
- Reduce costs for public health systems.
- Relieve family members and caregivers.

Technical Goals : design sensor-based systems

- Detecting alarming situations *(eg. Falls)*
- Assess the degree of frailty of older adults (impact of therapies).
- Detecting changes in behavior profile
  (*missing activities, disorder, interruptions, repetitions, inactivity*).
- Building a video library of reference behaviors characterizing people frailty.

Approach: Real time semantic activity understanding



1st experiment : Gerhome laboratory

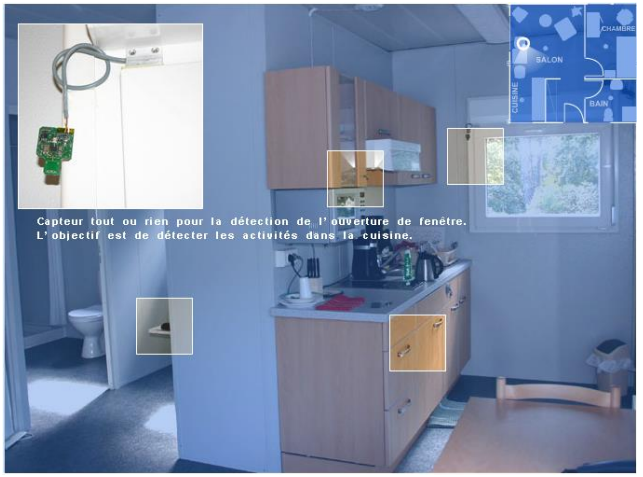
GERHOME (Gerontology at Home) : homecare laboratory (built in 2005)

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

- Experimental site in CSTB (Centre Scientifique et Technique du Bâtiment) at Sophia Antipolis http://gerhome.cstb.fr
- **Partners:** INRIA, CSTB, CHU-Nice, CG06…

Sensors installed in Gerhome laboratory

Contact sensor in the cupboard door



Contact sensor in the window

Pressure sensor underneath the legs of armchair

Video camera in the living-room

Event Recognition based on Knowledge

An 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

A language to model complex events

- **Language combining multi-sensor information**

EVENT (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)))

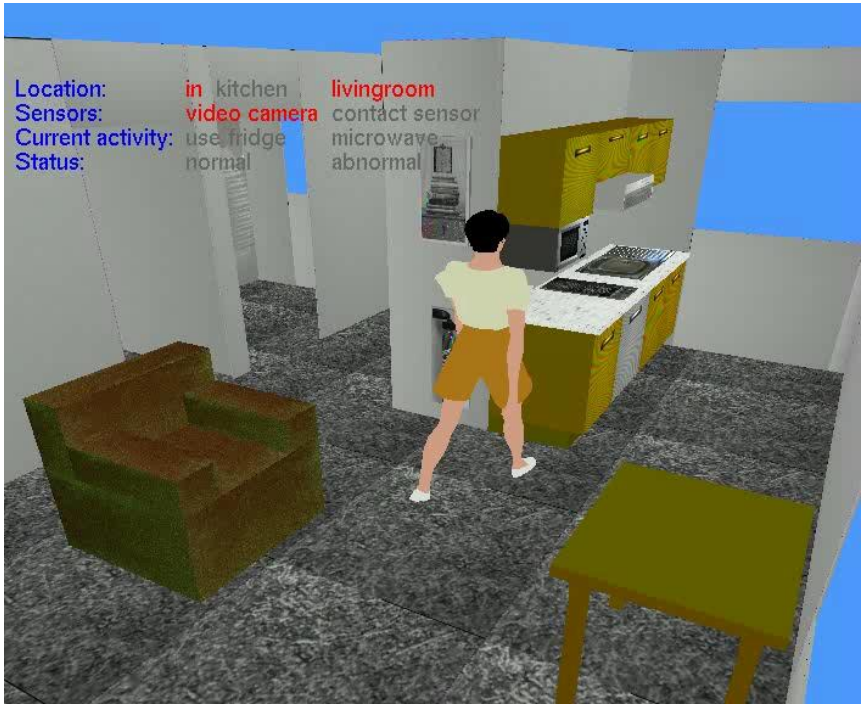
Recognition of the “Prepare meal” event

- 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	Bag on chair	**92%**	**94%**
Sitting on chair	66	54	12	15	78%	82%
Sitting on armchair	56	49	8	12	80%	86%
		Cold meal		2 instances of the event		
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

Recognition of a set of activities comparing two elderly people

		Elderly people 1 (64 years)			Elderly people 2 (85 years)			Normalized Difference	
Activity	**Used sensor (s)**	Activity duration (min:sec)		Nb inst (n1)	Activity duration (min:sec)		Nb inst (n2)	NDA= |m1-m2|/ (m1+m2)	NDI= |n1-n2| / (n1+n2)
		Mean (m1)	Total		Mean (m2)	Total			
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

		Elderly people 1 (64 years)			Elderly people 2 (85 years)			Normalized Difference	
Activity	**Used sensor (s)**	Activity duration (min:sec)		Nb inst (n1)	Activity duration (min:sec)		Nb inst (n2)	NDA= |m1-m2|/ (m1+m2)	NDI= |n1-n2| / (n1+n2)
		Mean (m1)	Total		Mean (m2)	Total			
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

		Elderly people 1 (64 years)			Elderly people 2 (85 years)			Normalized Difference	
Activity	**Used sensor (s)**	Activity duration (min:sec)		Nb inst (n1)	Activity duration (min:sec)		Nb inst (n2)	NDA= |m1-m2|/ (m1+m2)	NDI= |n1-n2| / (n1+n2)
		Mean (m1)	Total		Mean (m2)	Total			
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

2nd experiment : Sweet-Home Objectives

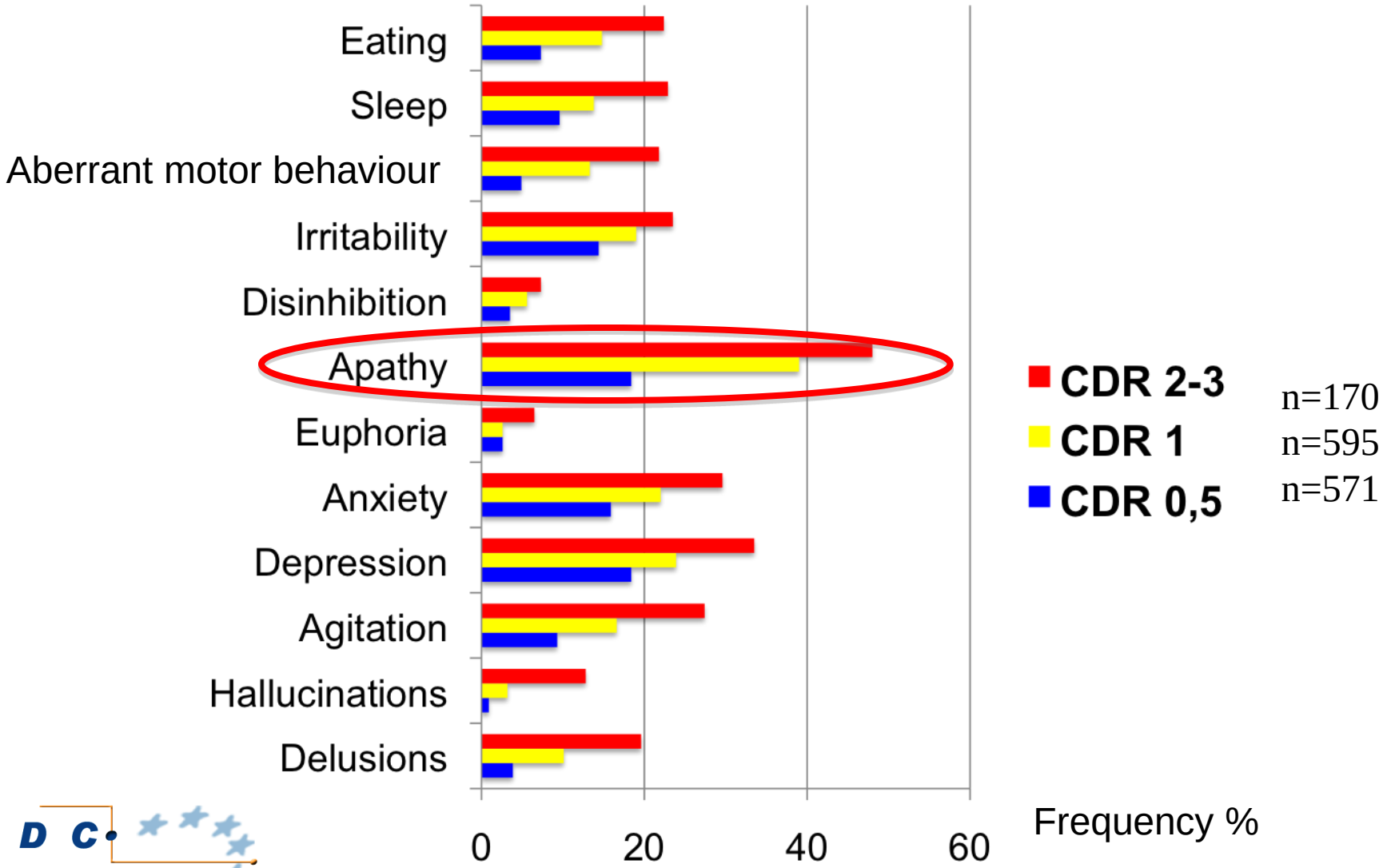
State Of the Art: Changes of behaviors and daily activities have been found in the early stage of Alzheimer Disease (AD) and may serve as signals for early diagnosis of AD.

Sweet-Home Objectives:

- Study the correlation between
  - gold standard scales (MMSE, NPI, …) and
  - the BPSD, IADL, gait information and cognitive capability (i.e. memory) of AD patients.
- Identify potential significant indicators for quantifying this correlation and for AD early diagnosis : discriminating the AD patients and NC normal controls.
- Develop a software platform to compute these indicators and for multiple sides to cooperate on the finding of best AD treatment.

BPSD: Behavioral and Psychological Symptoms of Dementia
IADL: Instrumental Activities of Daily Living

Neuropsychiatric symptoms become more frequent with the disease progression



Sweet Home

EADC European Alzheimer's Disease Consortium

Method, Sites and Scenario

- **Method : Event Recognition Platform**
  - **Video** (France) / **audio Sensors** (Vietnam)
  - **Inertial Sensors** (Taiwan)
  - **Participants in Protocol1 : 57 AD patients + 70 NC Normal Control**

- **3 Sites for the Experimental Assessment**
  - **2 Indoor (France&Taiwan) and 1 Outdoor (Taiwan) Sites**

- **6 Scenarios**
  - **Step A: Directed Activities**
    - A1: 7 indoor Activities (France&Taiwan)
    - A2: Walking 40 Meters outdoor (Taiwan)
  - **Step B: Semi Directed Activities**
    - B1: 10 indoor Activities (France&Taiwan)
    - B2: 2 tests for Prospective Memory (Taiwan)
  - **Step C: Free Indoor Activities (France)**
  - **Step D: Test for Sense of Direction outdoor (Taiwan)**

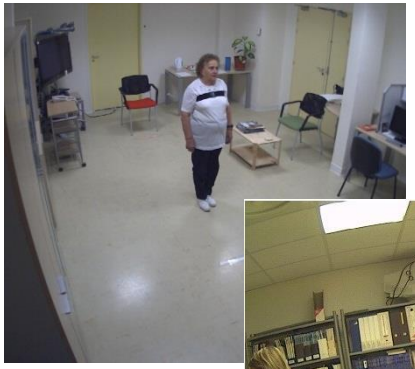
Experimental Results

Step A1: Directed Activities

- ***Balance testing***
  - (S1) Side by side stand, one's feet together
  - (S2) Semi tandem stand, stand with the side of the heel of one foot touching the big toe of the other foot
  - (S3) Tandem stand, with the heel of one foot in front of and touching the toes of the other foot
  - (S4) Participant stands on one foot (right then left foot), eyes open, for 10 seconds or less if he has difficulties
- ***Speed of walk testing (S5)***
  - The examiner asks the participant to walk through the room, from the opposite side of the video camera for 4 meters and then to go back.
- ***Repeated chair stands testing***
  - (S6) The examiner asks the participant to make the first chair stand, from sat to stand position without using his arms. The examiner will then ask the participant to do the same action 5 times in a row.
  - (S7) Time Up & Go test(TUG)

Monitoring Activities at Nice Hospital and in Taiwan

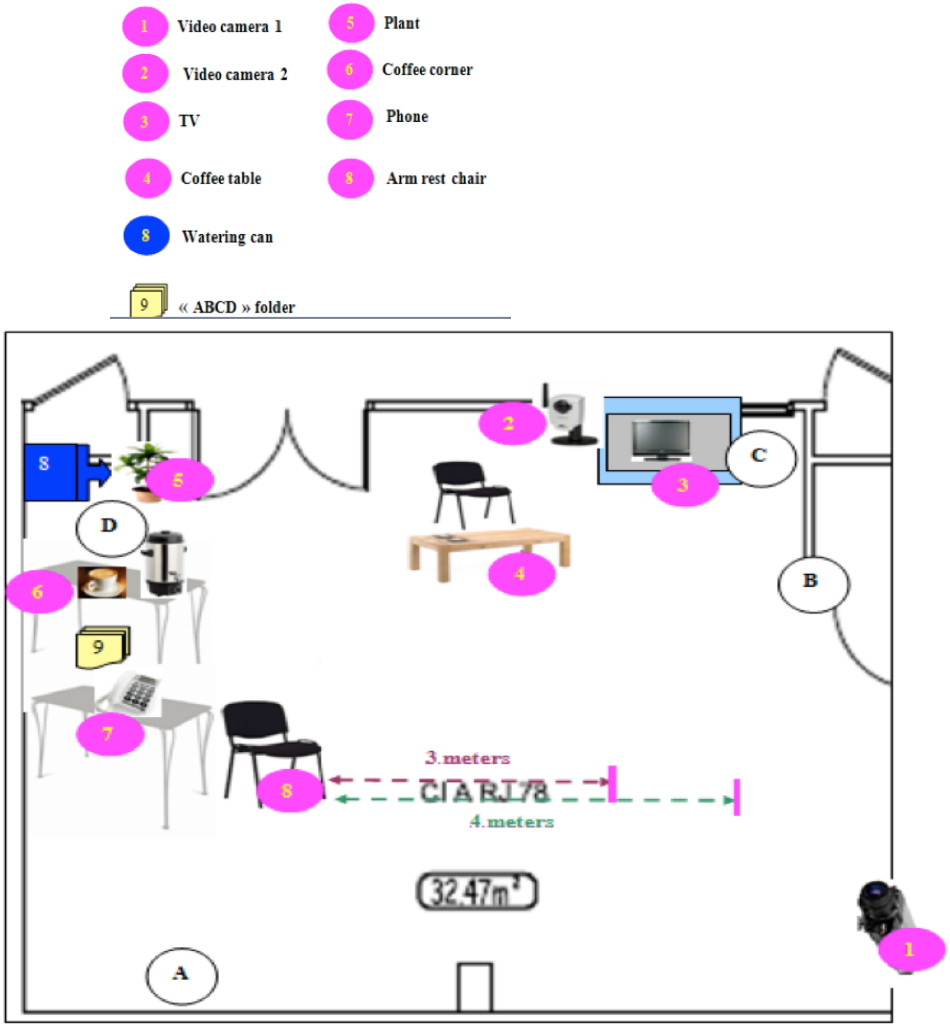
- **Medical staff & healthy younger**
  - 22 people (more female than male)
  - Age: ~ 25-35 years
  - Medical staff
- **Older persons (normal control)**
  - 20 (woman & man)
  - Age: ~ 60-85 years
- **Alzheimer patients**:
  - 21 AD people (woman & man)
  - 19 MCI (mild cognitive impairment) and mixed
  - Age: ~ 60-85 years

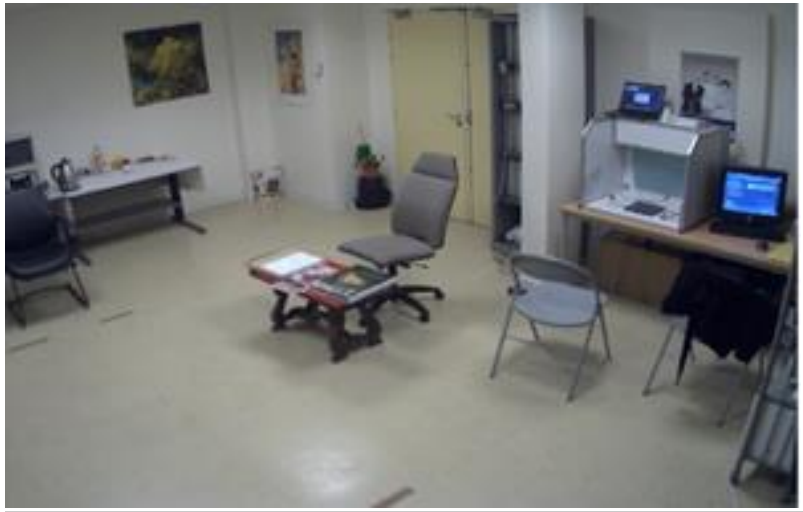


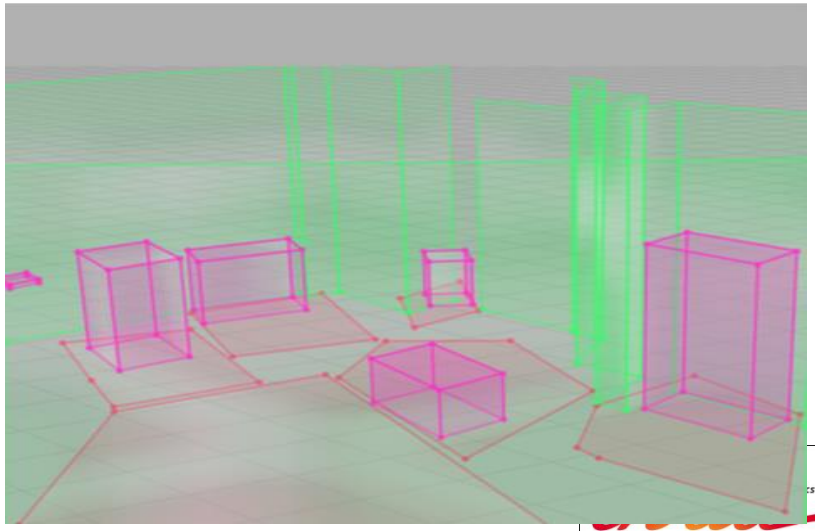
- **Activities monitored by various sensors**:
  - 2D RGB video cameras,
  - 3D RGBD video cameras,
  - inertial sensors : Actiwach/ motionPod
  - Stress sensors (impedance)
  - Microphones

- **3 Medical Protocols**
  - Protocol1: ~1year (2010-2011) **36** (18NC/6MCI/12AD) persons recruited
  - Protocol2: ~1year(2011-2012) **79** (29NC/36MCI/14AD) persons recruited
  - Protocol3: start on 06/2012 - **150** (50NC/50AD/50MCI) persons expected

CMRR in Nice Hospital
Screening of AD patients

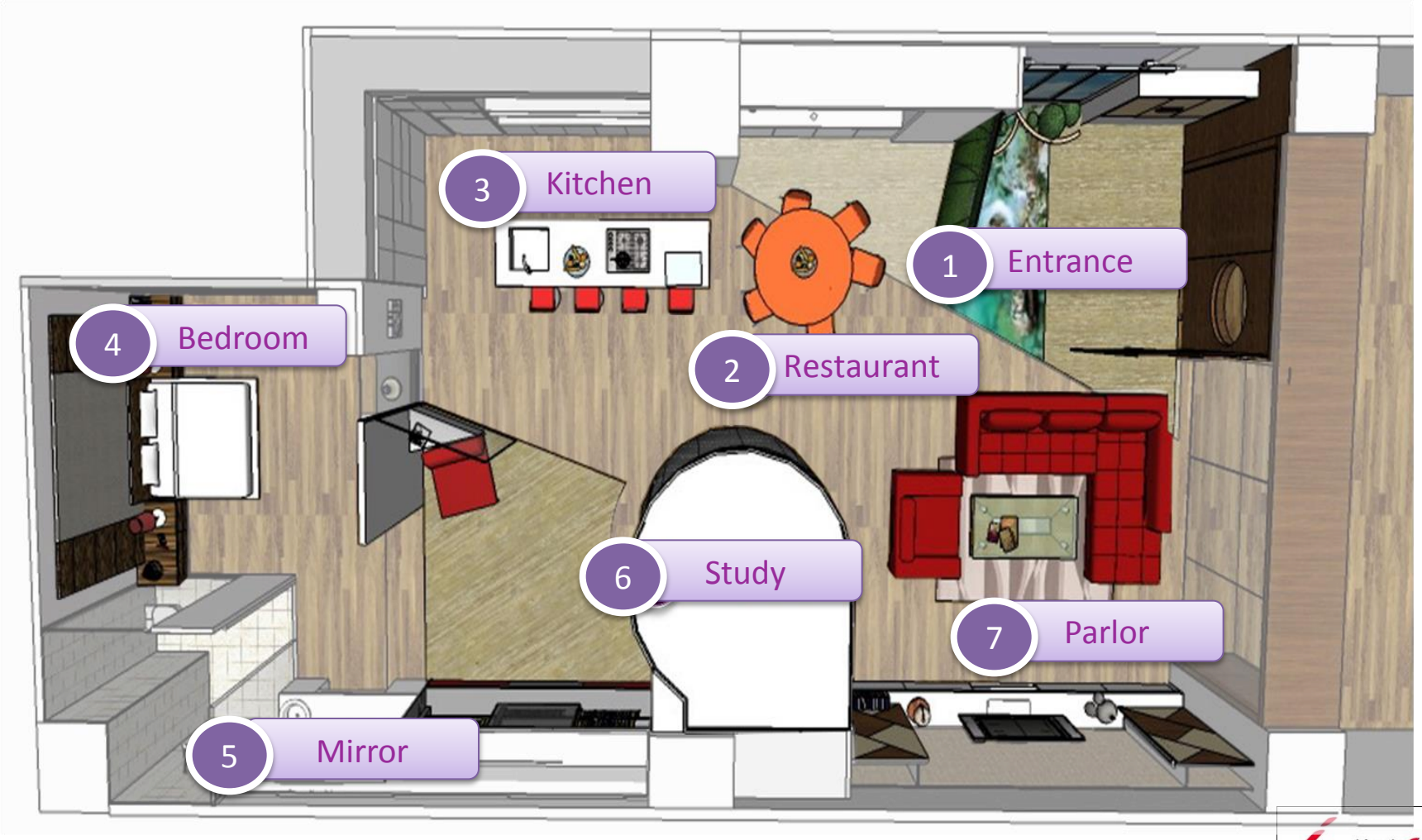






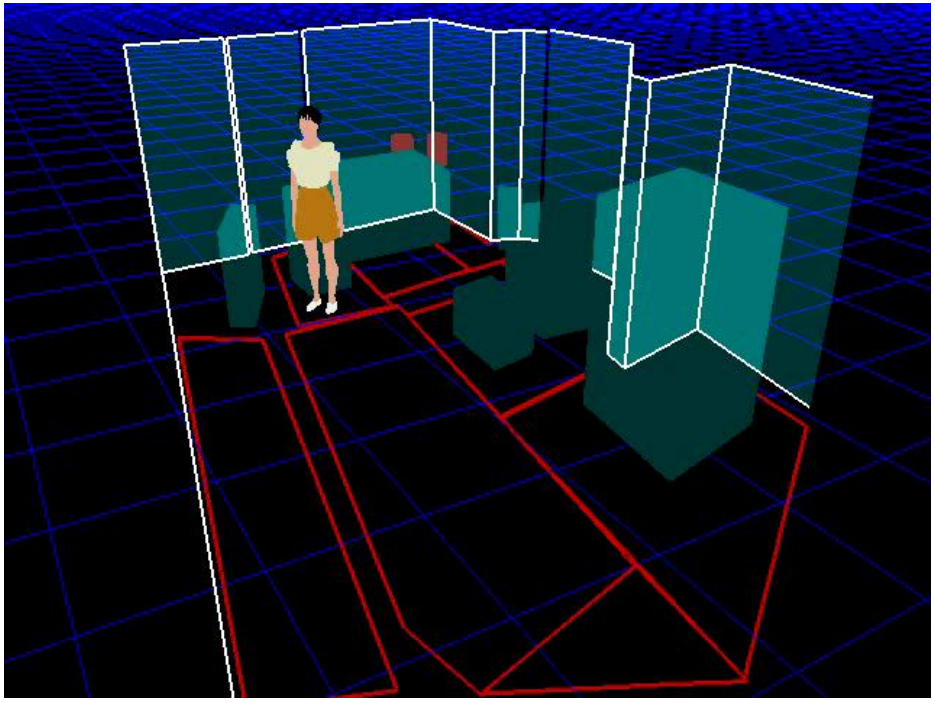
2nd site for the Assessment: apartment

Indoor Experimental Environment in Taiwan



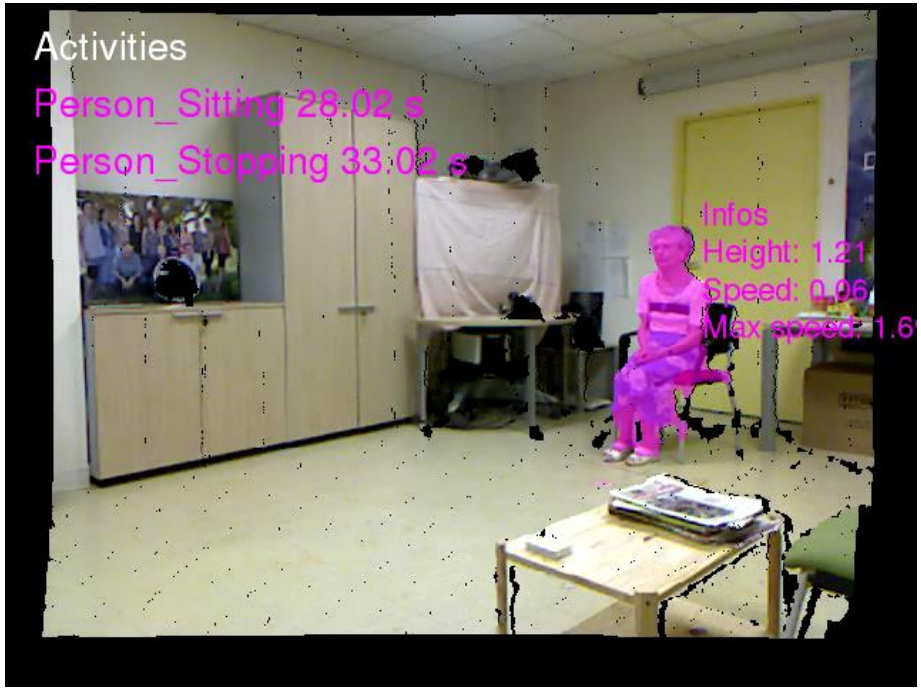
Activity monitoring in Nice Hospital with AD patients

Recognition of the “stand-up” activity.



Activity monitoring in Nice Hospital with AD patients

Recognition of the “stand-up & walking” activity.



Activity monitoring in Nice Hospital with AD patients

Recognition of 5 Physical Tasks (guided activities) with **RGB-D Camera**.

Approach Physical Task	Native Algo. Sensitivity	Native Algo. Precision	Using SDK Sensitivity	Using SDK Precision
Balance Test	100.0	89.28	92.30	100.0
Up and Go	80.0	90.90	100.0	81.25
Repeated Transfer	92.0	57.50	100.0	81.25
WS. Outward	100.0	83.33	100.0	96.29
WS. Return	96.0	82.75	92.30	85.71
Average	94.35	81.25	96.92	88.11

N: 24 participants, ~ 6 min. Each, Total: 120 events. Sens.: Sensitivity Index, eq.1; WS: Walking Speed Test.

Activity monitoring in Nice Hospital with AD patients

Prototype Accuracy on **Physical Tasks** (guided activities)

Comparison of Proposed System – Descriptive Models

2D-RGB Camera x **RGB-D Camera**

APPROACH	**RGB-D CAMERA**		**RGB CAMERA**	
Physical Task	Sensitivity	Precision	Sensitivity	Precision
Repeated Transf.	100.00	90.90	75.00	100.00
Up and Go	93.30	90.30	91.66	100.00
Balance Test	100.0	100.00	95.83	95.83
WS.Outward	100.0	90.90	91.66	100.00
WS. Return	90.00	100.00	87.50	95.45
Average	**96.60**	**94.20**	**88.33**	**98.33**

N: 29 participants, ~ 6 min. each, Total: ~150 events. Sens.: Sensitivity index, eq.1; WS: Walking Speed test

Activity monitoring in Nice Hospital with AD patients

Prototype Accuracy on **IADLs** (semi-guided activities)

Proposed System – Descriptive Models – **2D-RGB Camera**

IADLs	Sensitivity	Precision
Using Phone	72.83	85.50
Watching TV	80.00	71.42
Using Desk	92.72	58.62
Preparing Tea/ Coffee	90.36	69.44
Using Pharmacy Basket	100.00	88.09
Watering plant	100.00	64.91
Reading	71.42	69.76
Average Performance	**86.76**	**72.53**

N: 29 participants, ~ 15 min each, Total: 435 min, ~232 events.

Experiment 2 : Conclusion

Experiments in France and Taiwan have similar results : Alzheimer Disease's patients can be objectively characterized by several criterions based on activities :

- Step A:
  - Lower balance
  - Short gait length and gait frequency, irregular gait cycle
- Step B & C:
  - Lower ability in daily living, Experience apathy
  - Cognitive frailty and lower memory
- Step D:
  - Lower sense of direction

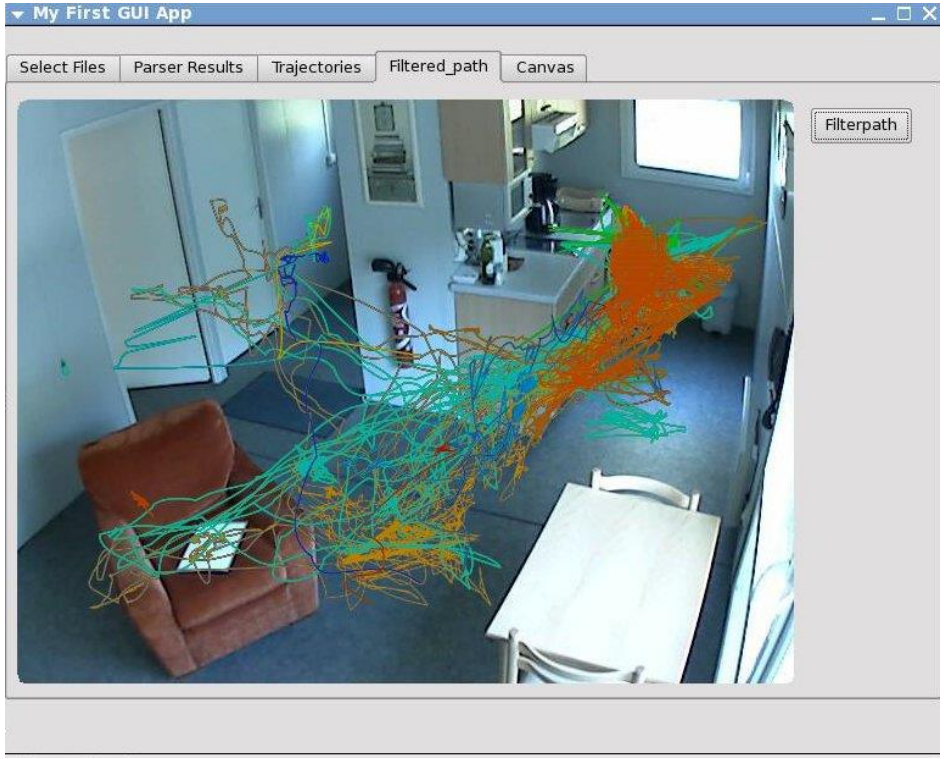
Next : Evaluate the correlation between dementia and the screening factors in the early stage.

- comparing AD/MCI and MCI/NC.

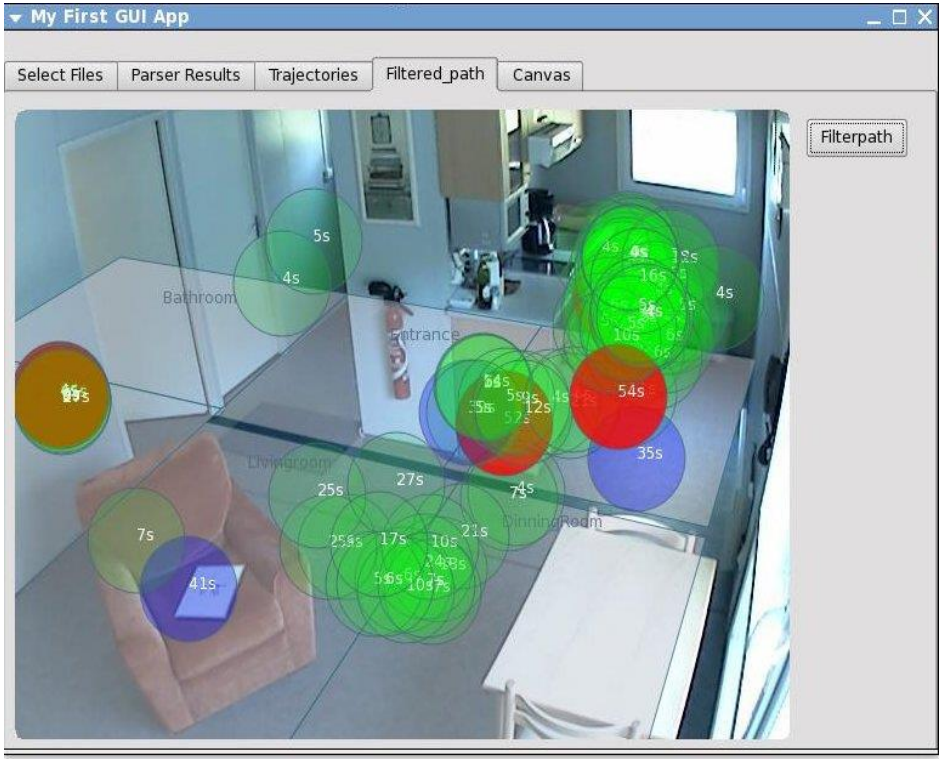
Learning Scenario Models : scene model

(G. Pusiol)

Localization of the person during 4 observation hours

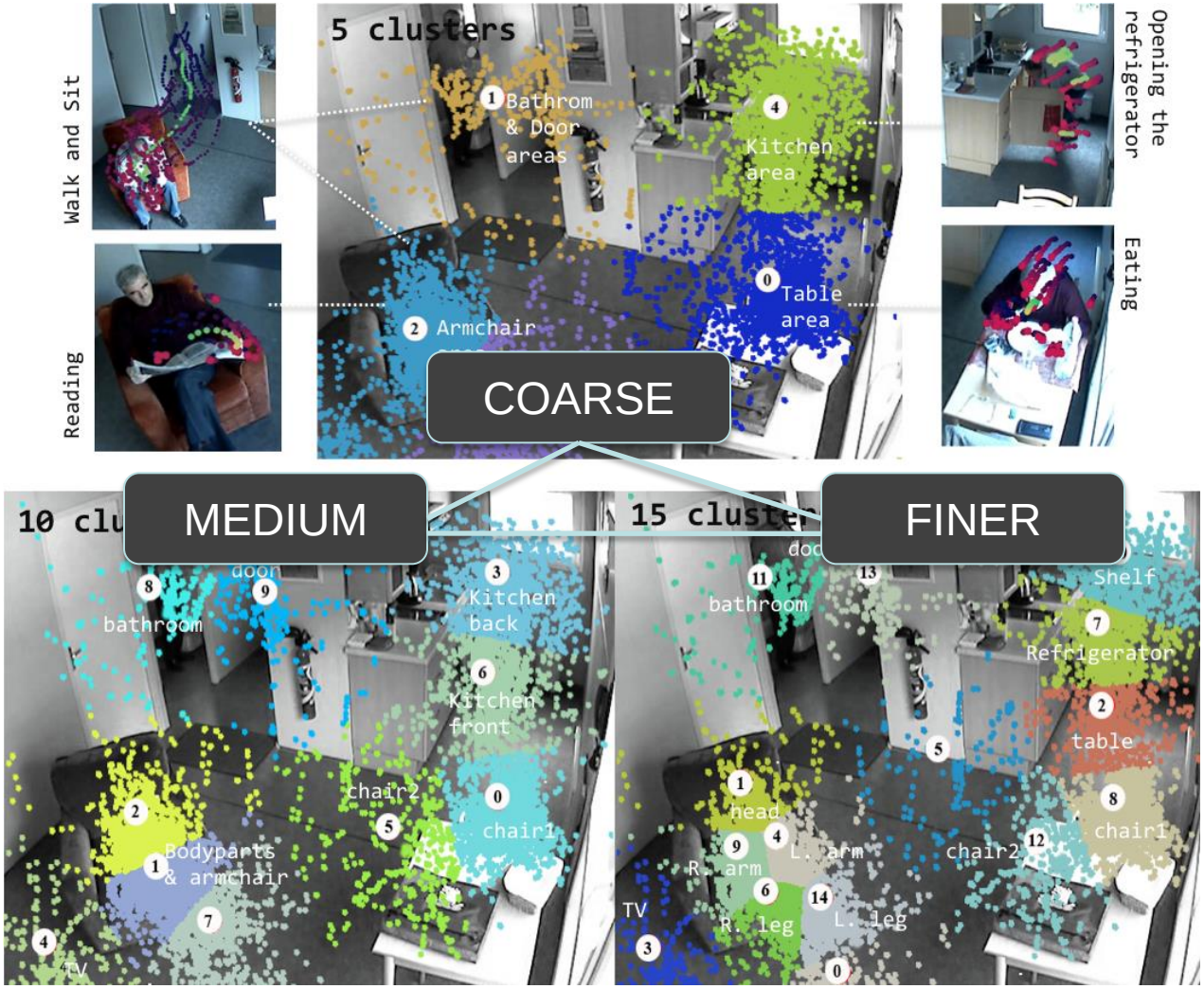


Stationery positions of the person



Walked distance = 3.71 km

Learning Scenario Models : scene model

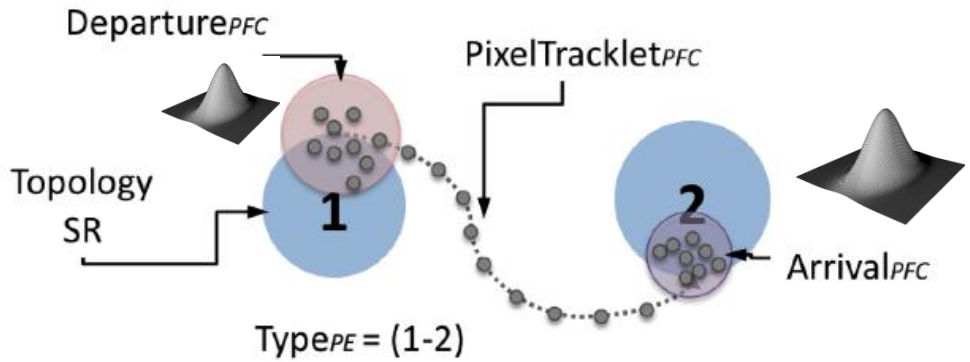


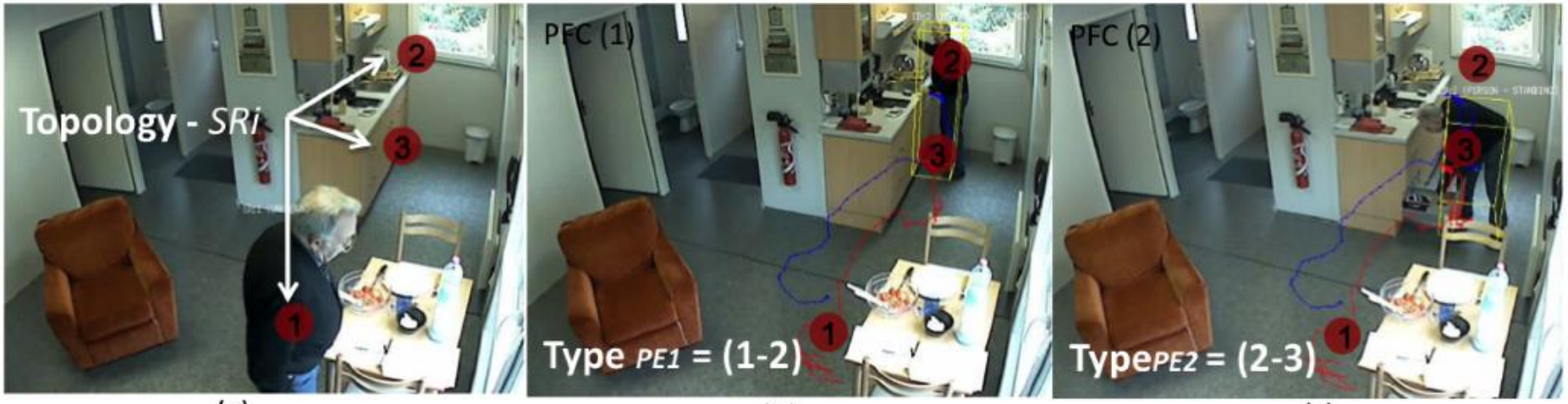
Topologies are important because is where the reasoning is

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

Learning Scenario Models : Primitive Events

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



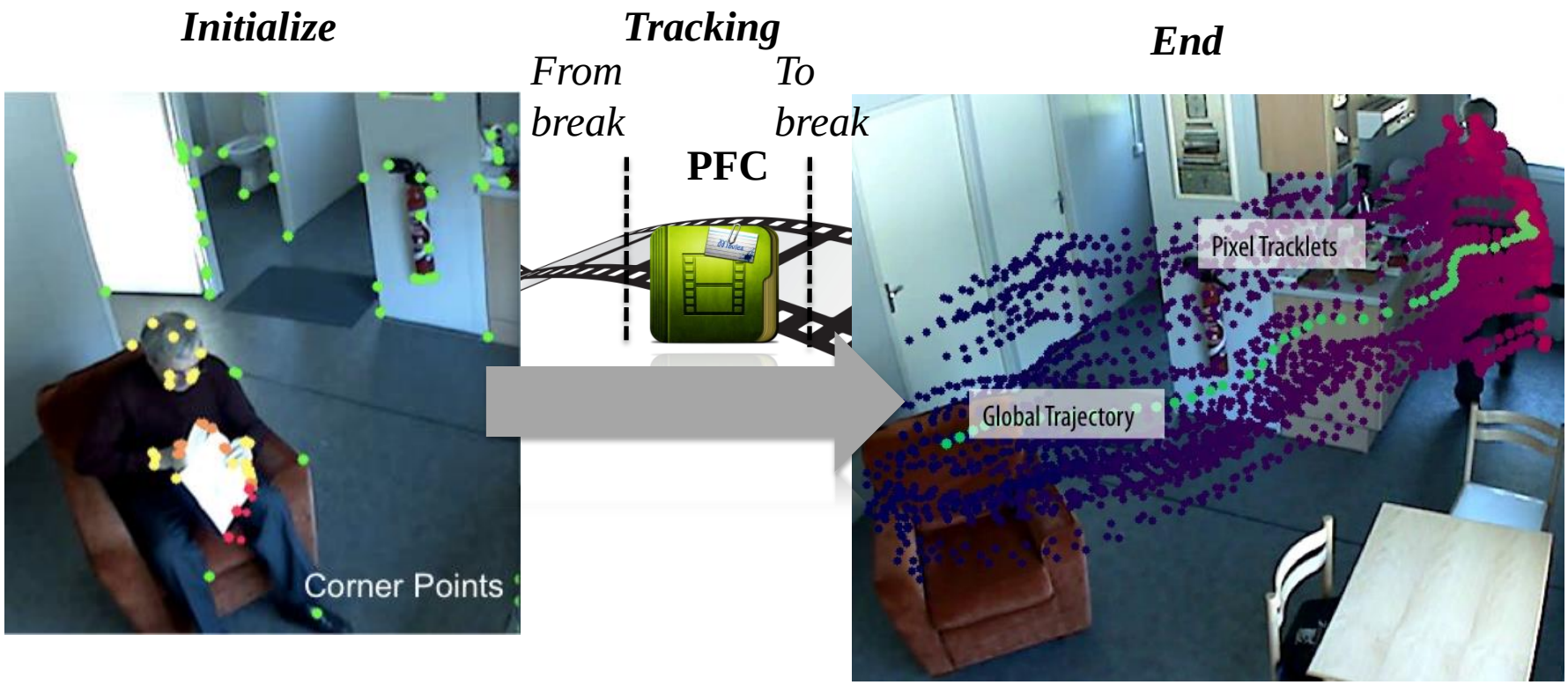


(a) (b) (c)

Primitive Event : global object motion between 2 zones.

Advantage:
The topology regions and primitive events are semantically understandable.

Learning Scenario Models: Local tracklets

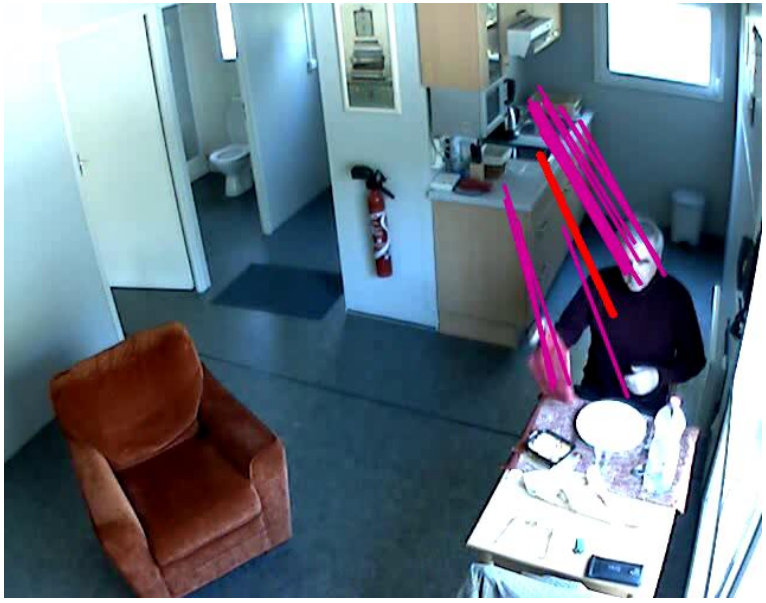


Associating Primitive Events with Local Dynamics:

1. *Initialize sparse KLT points*
2. *Track the points during the whole PE - pyramidal KLT*
3. *Filter with the global tracker*

Learning Scenario Models: Local tracklets

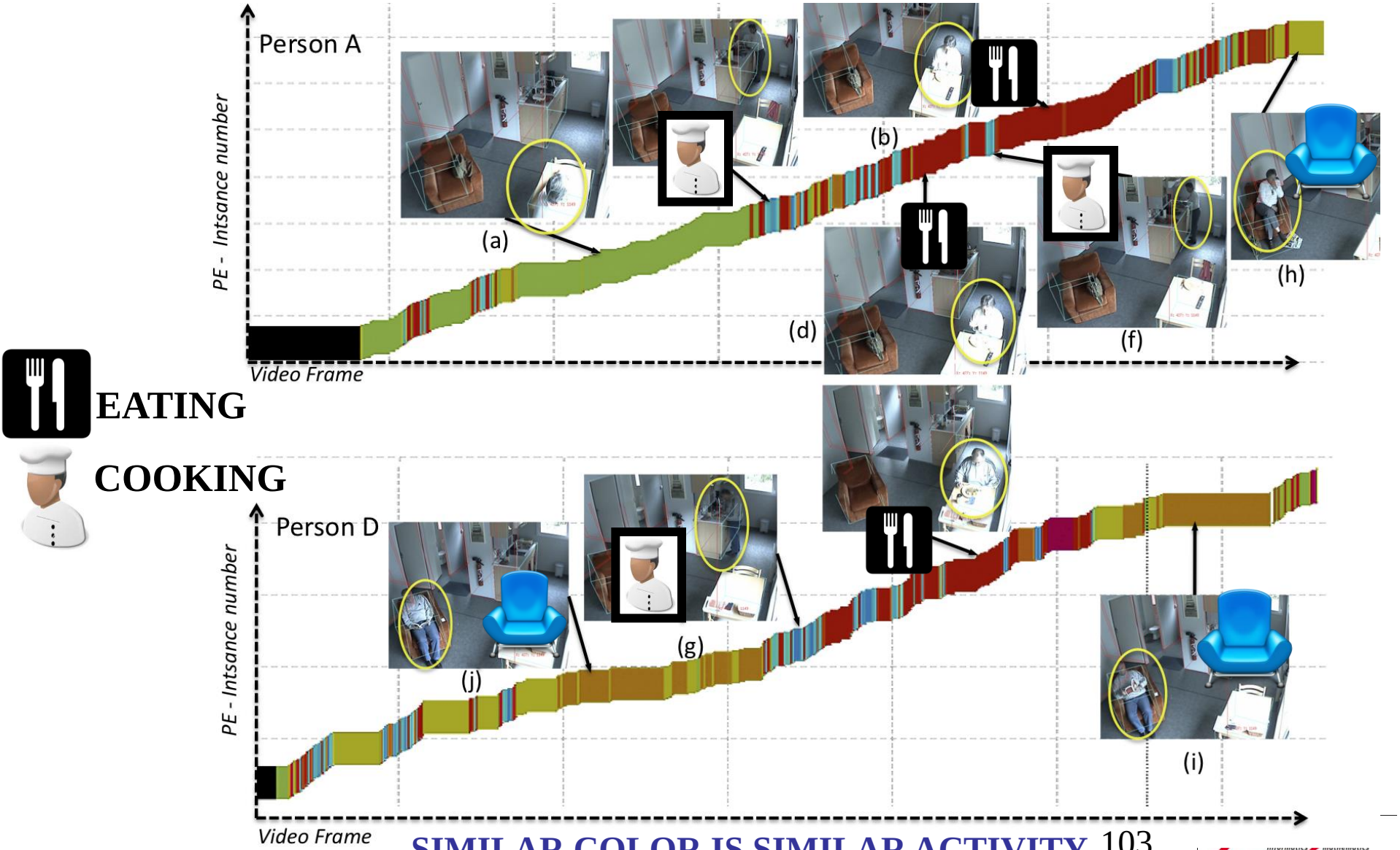
Example of Local Dynamics



SURF & SIFT: slower to compute

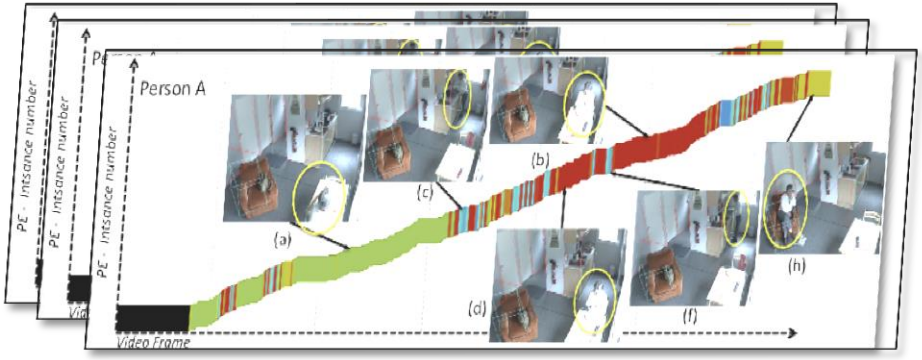
Primitive Events Results:

Each PE is colored by its type



SIMILAR COLOR IS SIMILAR ACTIVITY

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



Group PEs by types

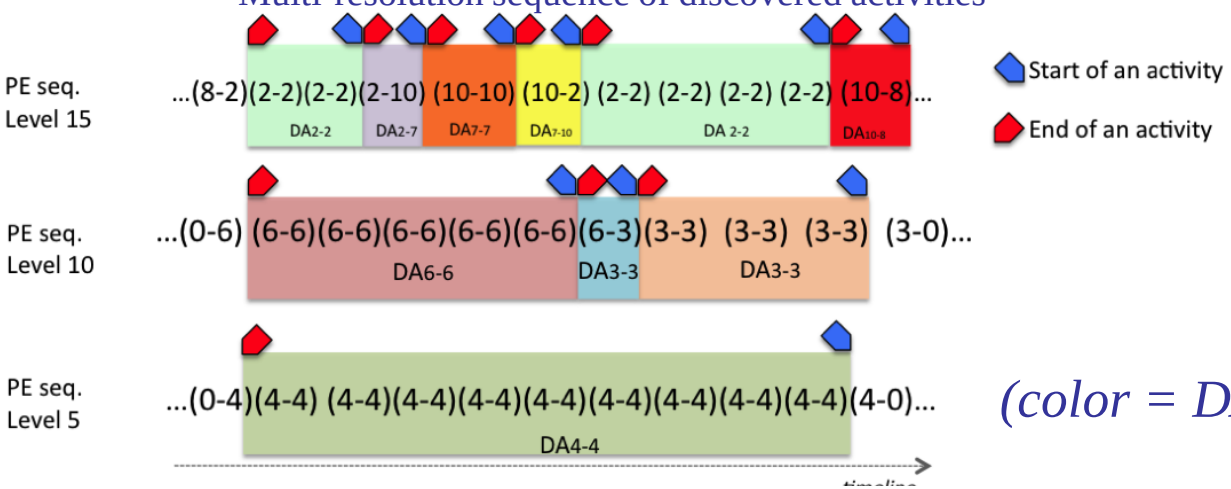
$$Stay_{A-A} = (A \rightarrow A)^{+}$$

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

Input: Sequences of PE

3 RESOLUTIONS

Multi-resolution sequence of discovered activities



-Easy to understand

-Non parametric and Deterministic

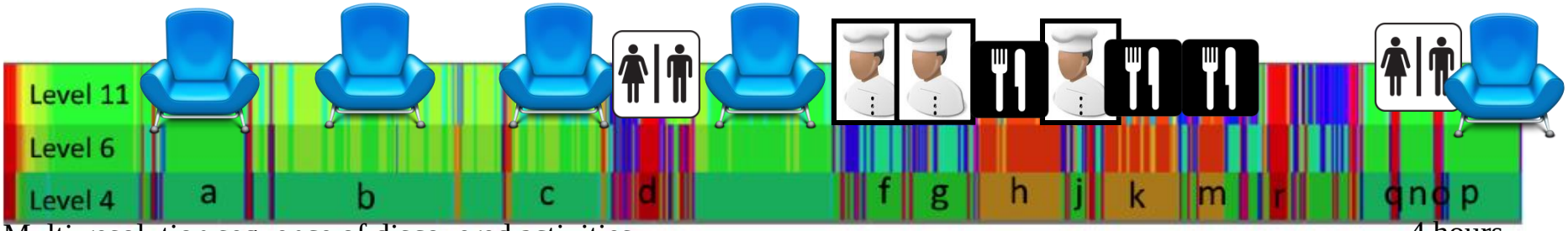
-The basic patterns can describe complex ones

(color = DA type)

DA = Discovered Activity

Discovery Results:

Similar color is similar Discovered Activity

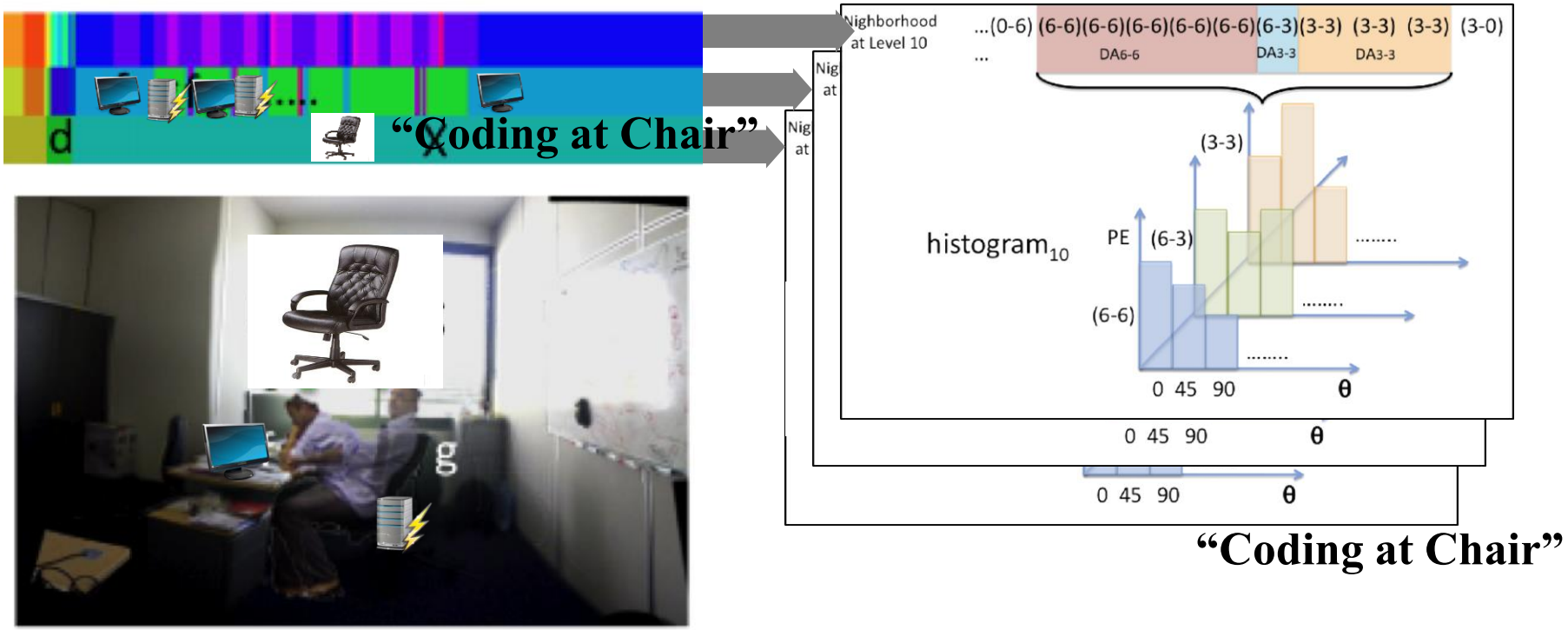


Multi-resolution sequence of discovered activities

4 hours

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

Activity Model : is a set of 3 histograms.
Each histogram has 2 dimensions. Containing global and local descriptions of the DAs.

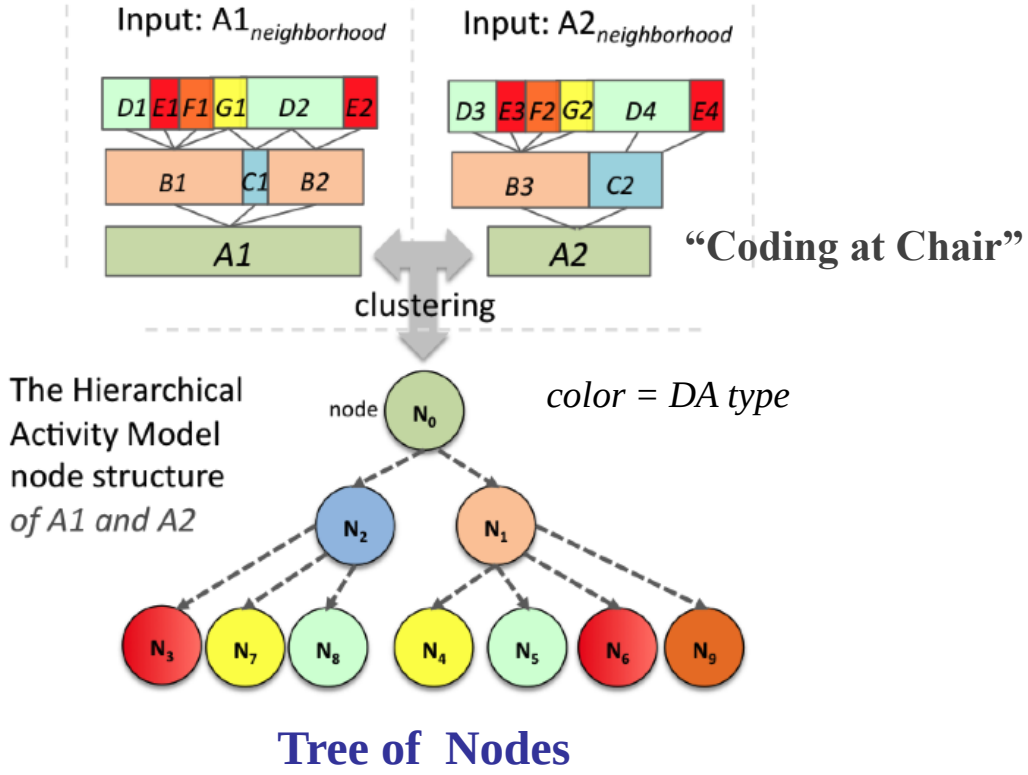


$$H_i(type_{PE}, \theta) = \sum LDD_N(\theta) \quad : \{\forall N \mid type_N = type_{PE}\}$$

Activity Models: Hierarchical Activity Models (HAM)

Building Nodes

Input **Training Neighborhoods** of a target activity

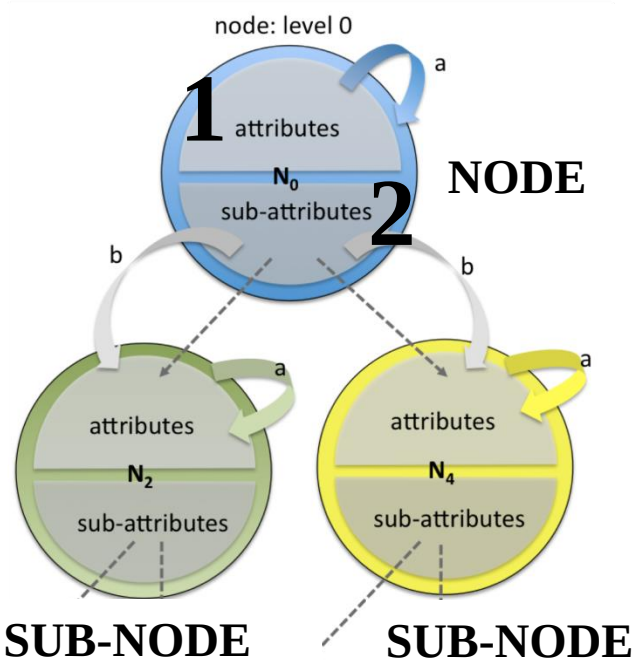


A **node N** is a set of discovered activities {DA1,DA2...,DAn} where all DAs are at the same resolution level and are of the same type (color)

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

1 Attributes

2 Sub-attributes



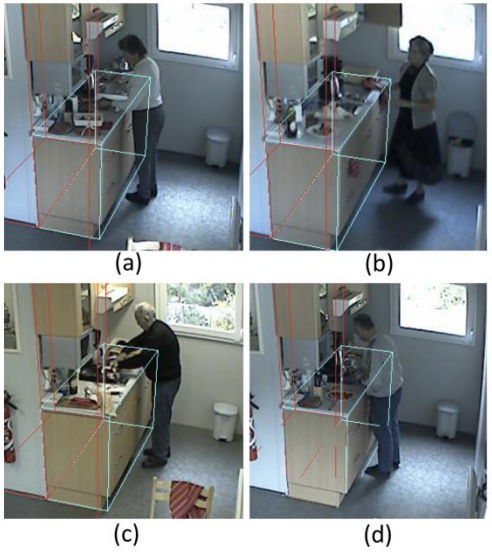
Results

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".

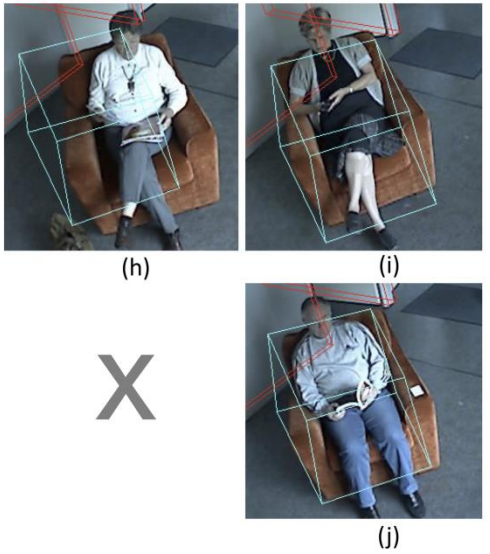
Scene logical Regions

4 Test Persons



"Cooking"

"Eating at Position A"



"Sitting in the Armchair"

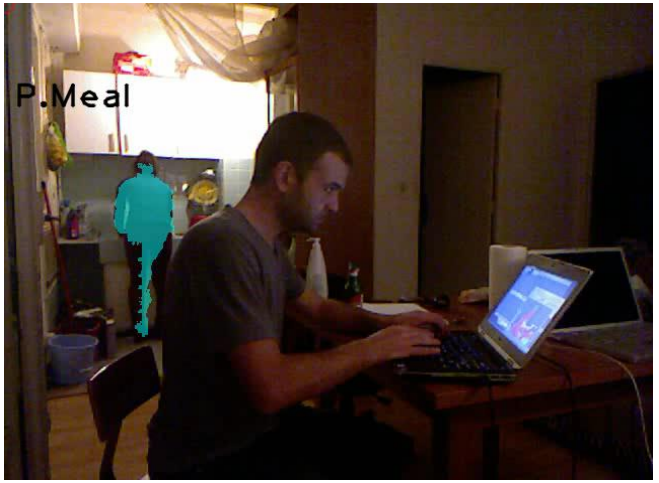
Evaluation

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

New European FP7 Project Dem@Care

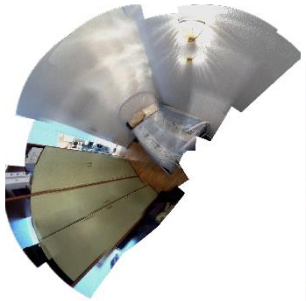
Future Work: Pilot2 @Nursing-Home (France, Ireland) & Pilot3 @Home (Ireland, Sweden):

- Context:
  - Monitoring of the **5 functional areas**: **Sleep** (diurnal/nocturnal), **ADL/IADLs**, **Physical Exercise**, **Social Interaction**, **Mood**

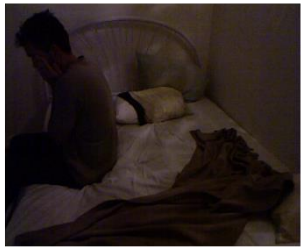
- Clinical Motivations : autonomy
  - Clinician benefits: Maintain comprehensive views of the status and progress of PwD's health in order to increase the **early detection** rate of **functional decline and other disorders** in older persons
  - PwD/Caregiver benefits: Receive **adaptive feedback** and **personalized support**

- Expected sensors (to be specified):
  - Video camera: RGB ambient (Axis®)/embedded (GoPro®) video camera, SenseCam®(Image, ambient light, T°C), RGBD video camera (Kinect®)
  - Audio: Ambient and embedded microphone
  - Accelerometers/Physiological sensors: BodyMedia SenseWear Pro3® (Skin conductance, 2D accelerometer, T°C), Philips DTI-2®(Skin conductance, 3D accelerometer, T°C, ambient light), Wireless Inertial Measurement Unit devices (accelerometry, gyroscope data)
  - Environmental sensors: Power consumption, Presence sensor

Use Cases in daily life - Sleep



(a) The scene



(b) Camera POV

(c) Person segmentation (RGB)

Discovery Results: *Sleeping*

Similar color is similar Discovered Activity

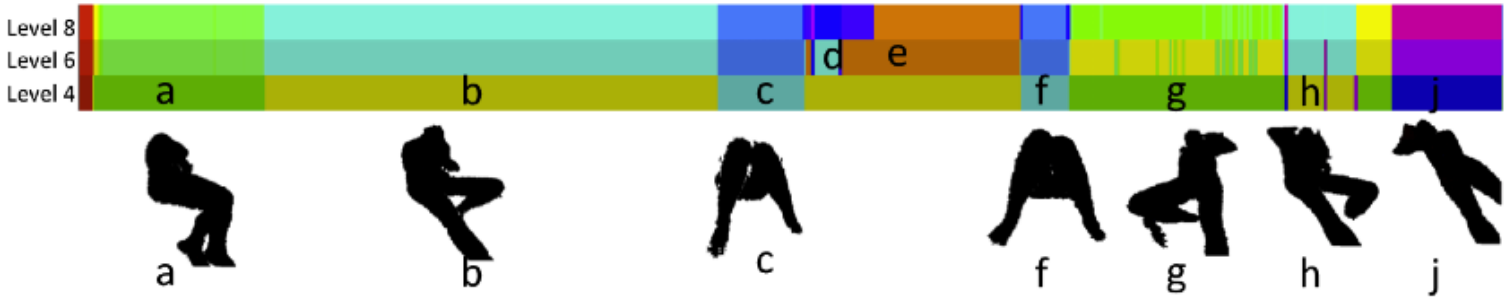
Level = Resolution = number of topology regions

d to e is the motion of the arms towards the head.

6 Hours

Multi-resolution sequence of discovered activities

color = DA type



Note the correspondence between the images and the colors (eg. c=f).

Conclusion - video understanding

A **global framework** for building real-time video understanding systems:

- Hypotheses:
  - mostly fixed cameras
  - 3D model of the empty scene
  - behavior models
- Results:
  - **3 types of Activity Monitoring Systems** to measure levels of everyday activities: from hand-craft to (un)supervised learned models of activity.
- 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….)

Conclusion for Assistive Living

Key advance : ICT software performance still needs to be measured

- Bracelets (wandering), fall detectors, serious games, low techs…
- Activity monitoring systems to measure levels of everyday activities.

Key perspectives : diagnosis, protection, engagement, empowerment

- Medical research, education : complete knowledge on AD, ageing through behavioural studies.
- Assessment : to understand behavioural disorders (sleeping disorders, apathy), frailty, disease burden. Reasons for going to institutions?
- Tools for personalised coaching, care : links between behavioural disorders and their causes: corrective actions, carer training.
- Engagement : social interaction, initiate activities, stimulation (serious games).

Limitations:

- User-center systems : large variety of people, environment…
- ICT software : reliable, accurate, autonomous
- Local companies : Installation and maintenance of large variety of sensors

Conclusion

International Journals

1. R. Romdhane, E. Mulin, A. Derreumeaux, N. Zouba, J. Piano, L. Lee, I. Leroi, P. Mallea, R. David, M. Thonnat, F. Bremond and P.H. Robert, Automatic Video Monitoring system for assessment of Alzheimer's Disease symptoms, *JNHA - The Journal of Nutrition, Health and Aging* Ms. No. JNHA-D-11-00004R1 - 2011.
2. C. Crispim, V. Joumier, Y-L Hsu, M-C Pai, P-C Chung, A. Dechamps, P.H. Robert and F. Bremond Alzheimer's patient activity assessment using different sensors, *Gerontechnology 2012;* 11(2):266-267; 2012, Best Paper Award of ISG*ISARC2012.

International Conferences

1. R. Romdhane, B. Boulay, F. Bremond and M. Thonnat. Probabilistic Recognition of Complex Event. In the *8th International Conference on Computer Vision Systems*, ICVS 2011, Sophia Antipolis, France on the 20th-22nd of September 2011.
2. G. Pusiol, F. Bremond and M. Thonnat. Unsupervised Discovery, Modeling and Analysis of long term Activities. In the *8th International Conference on Computer Vision Systems*, ICVS 2011, Sophia Antipolis, France on the 20th-22nd of September 2011.
3. J. Joumier, R. Romdhane, F. Bremond, M. Thonnat, E. Mulin, P.H. Robert, A. Derreumeaux, J. Piano and L. Lee. Video Activity Recognition Framework for assessing motor behavioural disorders in Alzheimer Disease Patients. In the *International Workshop on Behaviour Analysis*, Behave 2011, Sophia Antipolis, France on the 23rd of September 2011.
4. C. Crispim-Junior, V. Joumier, Y.L. Hsu, M.C. Pai, P.C Chung, A. Dechamps, P. Robert and F. Bremond. Sensors data assessment for older person activity analysis. *In the proceedings of the ISG*ISARC 2012 Conference (International Society for Gerontechnology, the 8th world conference in cooperation with the ISARC, International Symposium of Automation and Robotics in Construction)*, Eindhoven, The Netherlands, 26 to 29 June 2012.
5. P. Bilinski and F. Bremond. Contextual Statistics of Space-Time Ordered Features for Human Action Recognition. *The 9th IEEE International Conference On Advanced Video and Signal Based Surveillance (AVSS 12)*, Beijing on 18-21 September 2012.
6. C. Crispim-Junior, V. Joumier and F. Bremond. A Multi-Sensor Approach for Activity Recognition in Older Patients. *In the proceedings of the Second International Conference on Ambient Computing, Applications, Services and Technologies.* , AMBIENT 2012. Barcelona, September 23-28, 2012.
7. P. Bilinski and F. Bremond. Statistics of Pairwise Cooccurring Local SpatioTemporal Features for Human Action Recognition. *The 4th International Workshop on Video Event Categorization, Tagging and Retrieval (VECTaR 2012)* , ECCV 2012 Workshops, Firenze, Italy, October 13, 2012.

Are we addressing End-user needs?

There are several end-users in homecare:

- Doctors (gerontologists):
  - Frailty measurement (depression, …)
  - Alarm detection (falls, gas, dementia, …).
- Caregivers and nursing home:
  - Cost reduction: no false alarm and reduction employee involvement.
  - Employee protection.
- Persons with special needs, including young children, disabled and older people:
  - Feeling safe at home.
  - Autonomy: at night, lighting up the way to bathroom.
  - Improving life: smart mirror, summary of user day, week, month, in terms of walking distance, TV, water consumption.
- Family members and relatives:
  - Older people safety and protection.
  - Social connectivity.

Practical Problems and Solutions

Problems	Solutions
Privacy confidentiality and ethics: video (and other data) recording, processing and transmission.	No video recording and transmission, only textual alarms.
Acceptability for older people	User empowerment.
Usability	Easy ergonomic interface (no keyboard, large screen), friendly usage of the system.
Cost effectiveness	The right service for the right price, large variety of solutions.
Legal issues, no certification	Robustness, benchmarking, on site evaluation
Installation, maintenance, training, interoperability with other home devices	Adaptability, X-Box integration, wireless, open standards (OSGI, …)
Research financing	Insurances, Companies or Governments : France (lobbies), Europe (not organized), US, Asia.