Actions in Context: System for people with Dementia

Àlex Pardo, Albert Clapés, Sergio Escalera and Oriol Pujol

recerCaixa

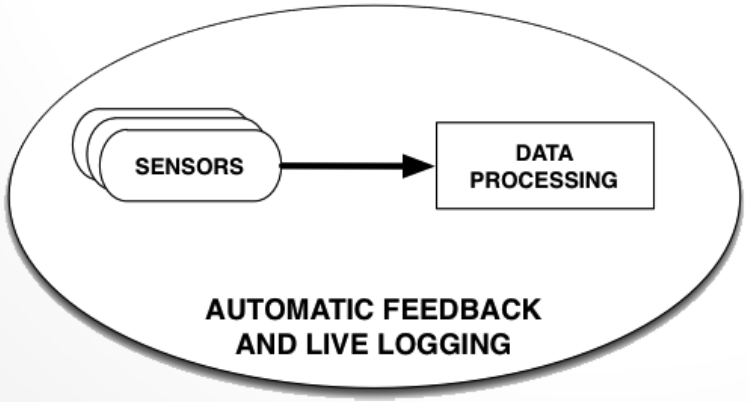
CVC
Centre de Visió per Computador

UB Universitat de Barcelona

PBA

Motivation

- Complex Systems are usually studied as networks where each node is usually associated to unimodal data
- We hypothesize, the future of complex systems is linked with the analysis of higher level information such as citizen behaviour.
- More complex information must be extracted and analysed
- In our proposal, a single node is a smart environment that fuses a Personal Area Network with Computer Vision systems.



Case of Study: Improved Autonomy for People with Dementia

Outline

- Introduction
- System
- Validation
- Conclusion

Introduction

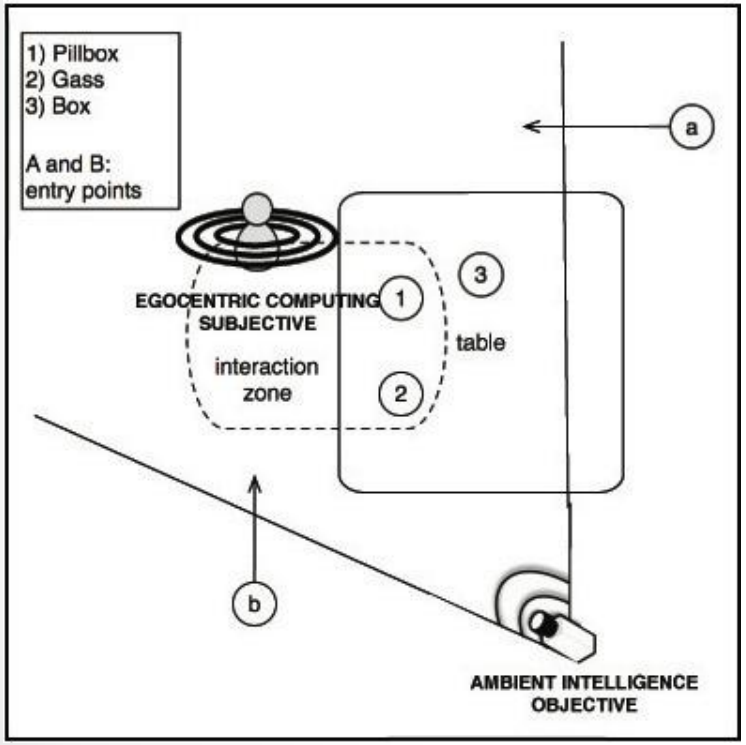
- Contextualization
  - Number of people with dementia increasing
  - Alzheimer disease hinders the autonomy of the patient
  - Constant attention

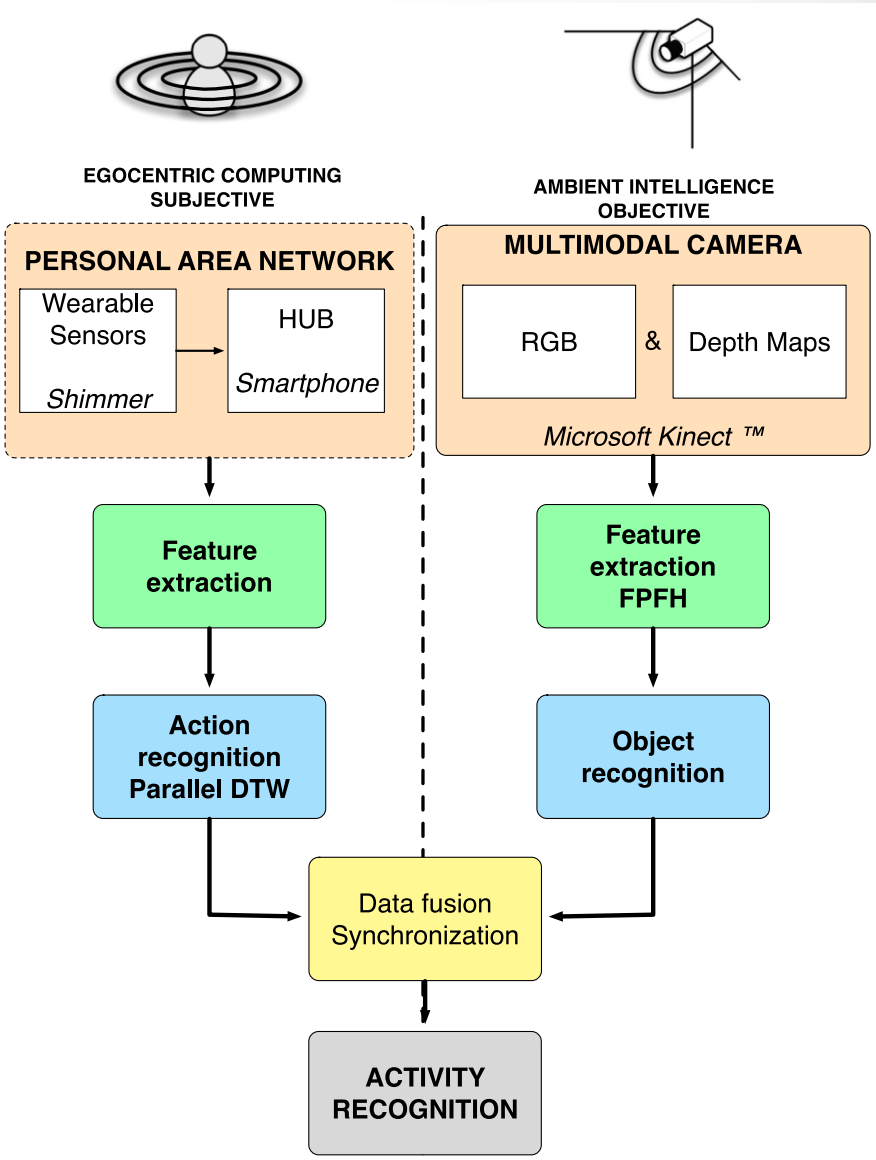
Image by Jon Díez Supat

Introduction

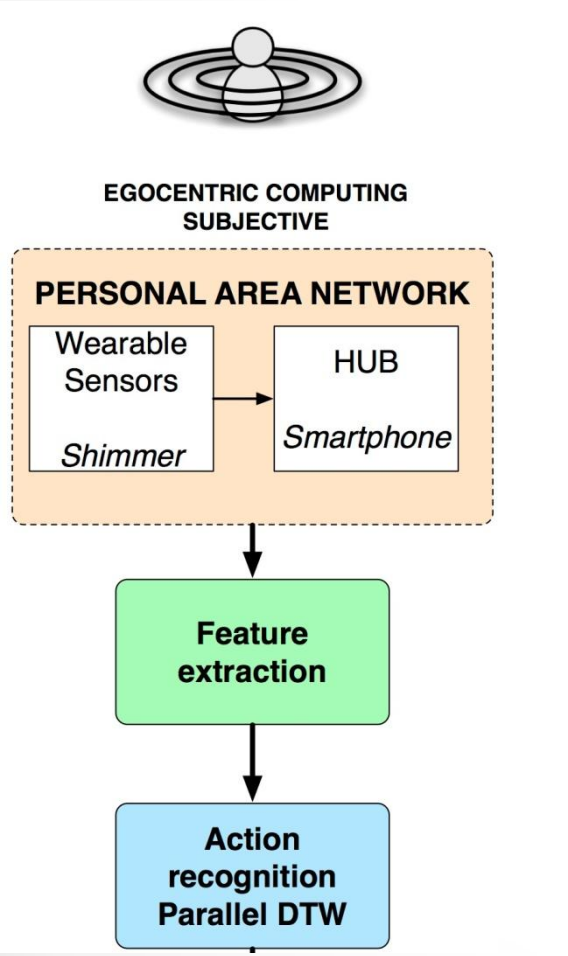
- Proposal
  - Multi-modal fusion pipeline for real-time:
    - User and object detection
    - Object identification
    - Gesture recognition
    - User-object relationship recognition

System


1) Pillbox
2) Gass
3) Box
A and B:
entry points
a
b
1
2
3
table
interaction
zone
EGOCENTRIC COMPUTING
SUBJECTIVE
AMBIENT INTELLIGENCE
OBJECTIVE


EGOCENTRIC COMPUTING
SUBJECTIVE
AMBIENT INTELLIGENCE
OBJECTIVE
PERSONAL AREA NETWORK
Wearable
Sensors
Shimmer
HUB
Smartphone
MULTIMODAL CAMERA
RGB
&
Depth Maps
Microsoft Kinect ™
Feature
extraction
Feature
extraction
FPFH
Action
recognition
Parallel DTW
Object
recognition
Data fusion
Synchronization
ACTIVITY
RECOGNITION

System
Egocentric Computing

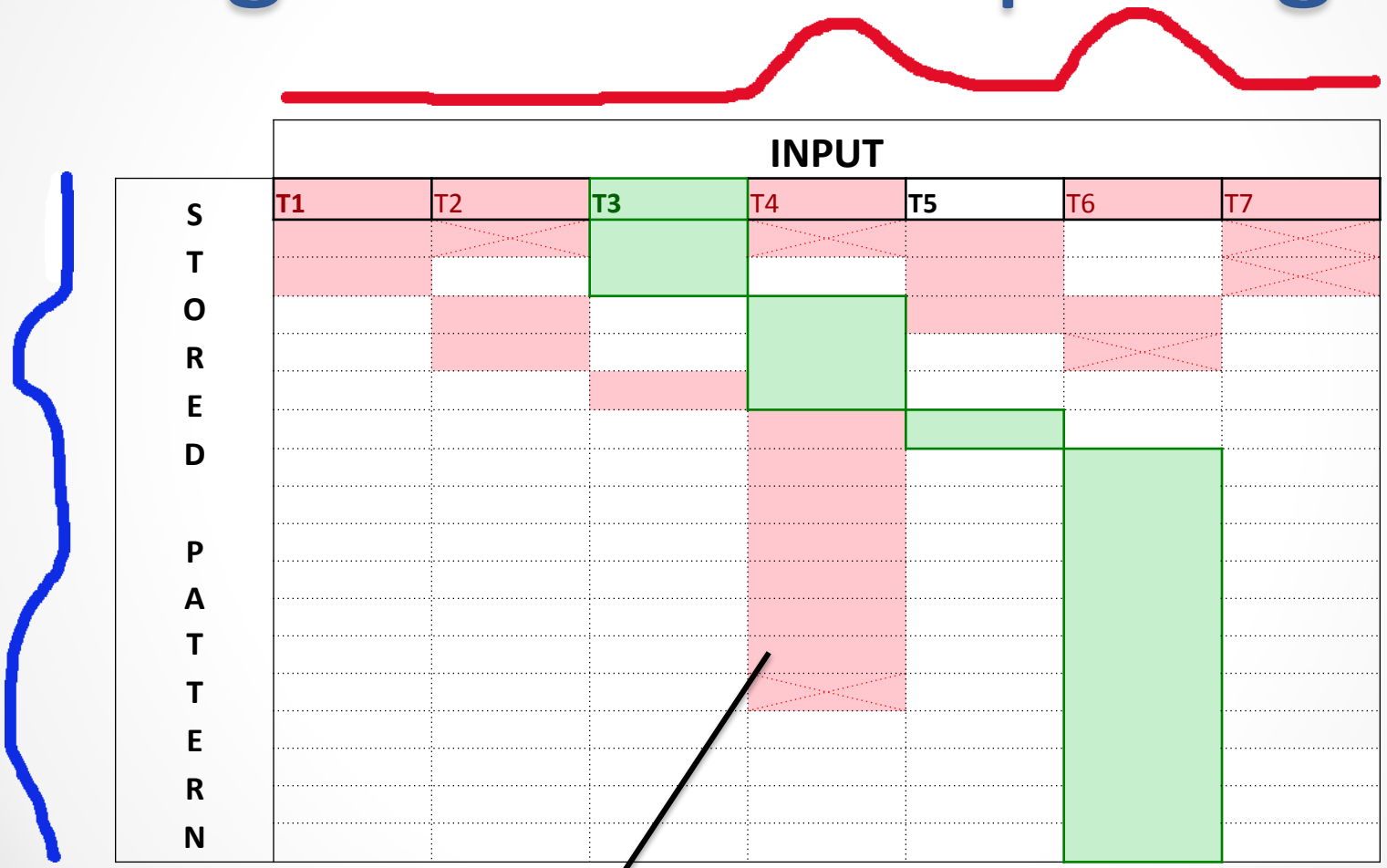


Data obtained by the sensor is transmitted to the smartphone using Bluetooth in order to send it to the server.

Extract features from sensor: raw angular speed and accelerometer values and and its energies.

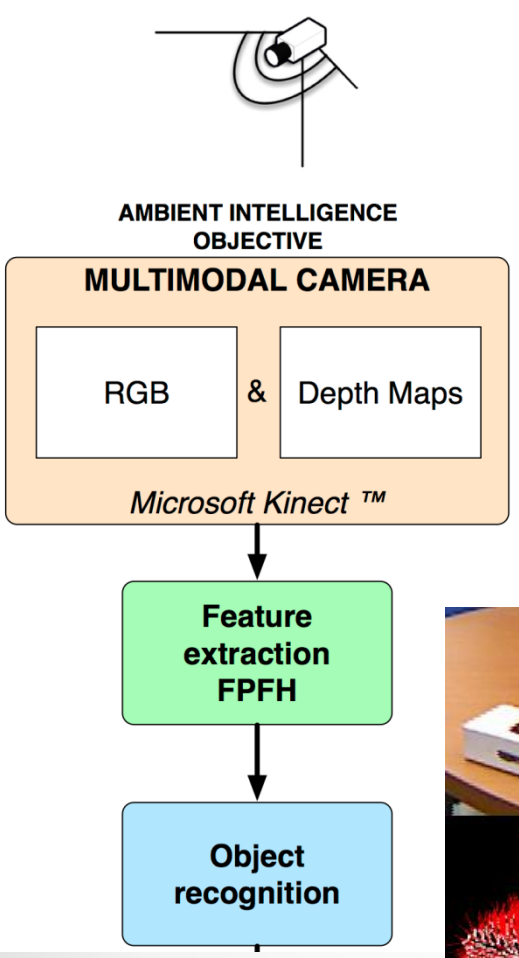
Apply a Parallelized Dynamic Time Warping algorithm to detect gestures using a single thread for each new possible DTW recognition path.

System Egocentric Computing



$$M(i,j) = min(M(i-1,j), M(i,j-1), M(i-1,j-1)) + d(p_i, s_j)$$

System
Ambient Intelligence

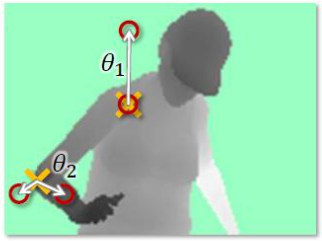


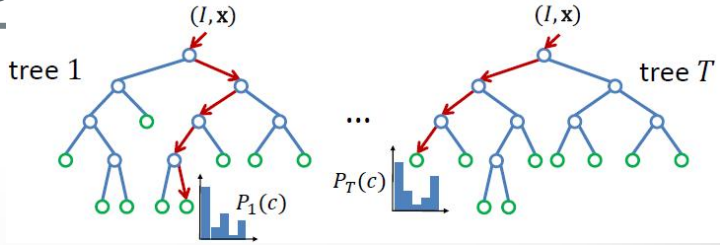
RGB and Depth values are sent to the computer.

A Gaussian distribution per pixel is learnt in order to subtract the background.

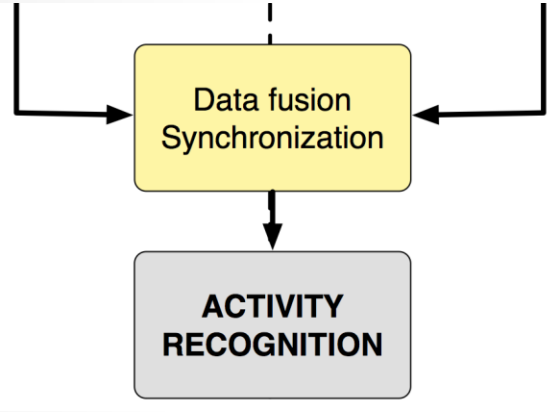
Users are detected an tracked by using Random Forest classifier.

Fast P … scribe the ol … *K*-NN … ects in the … tions.



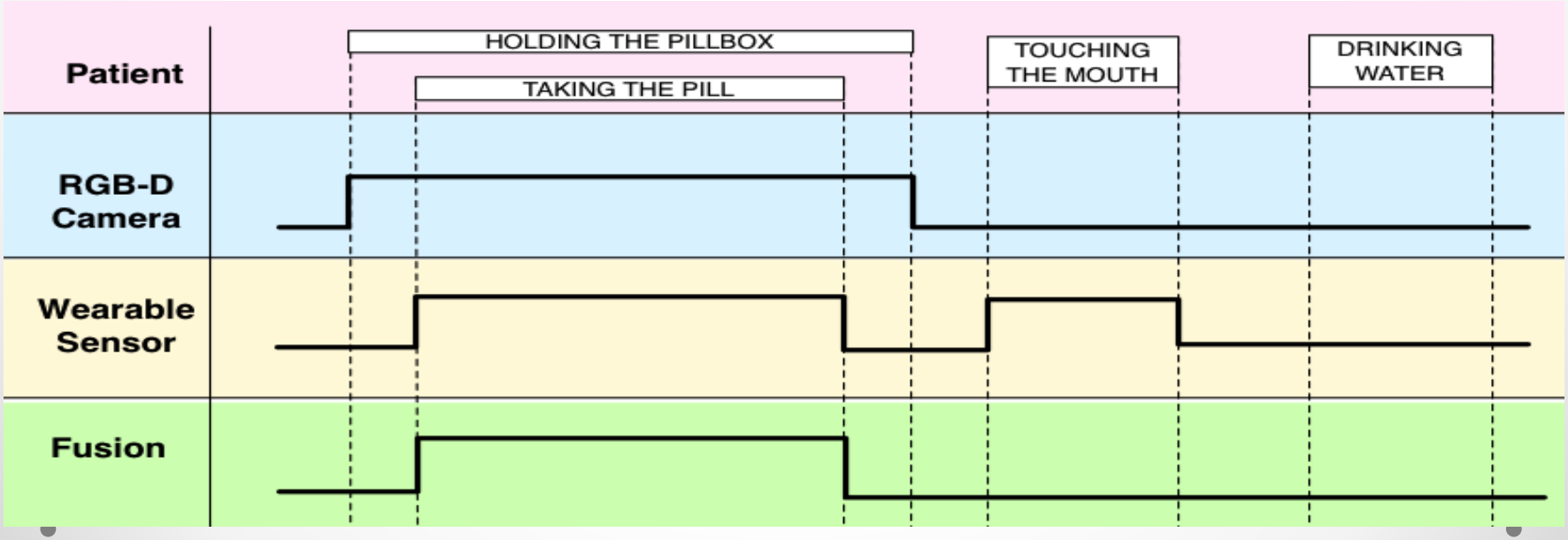


System Fusion



The results obtained in both modalities are synchronized by the use of timestamps.

The intersection between object interaction and gesture recognition defines the activity the user is performing.



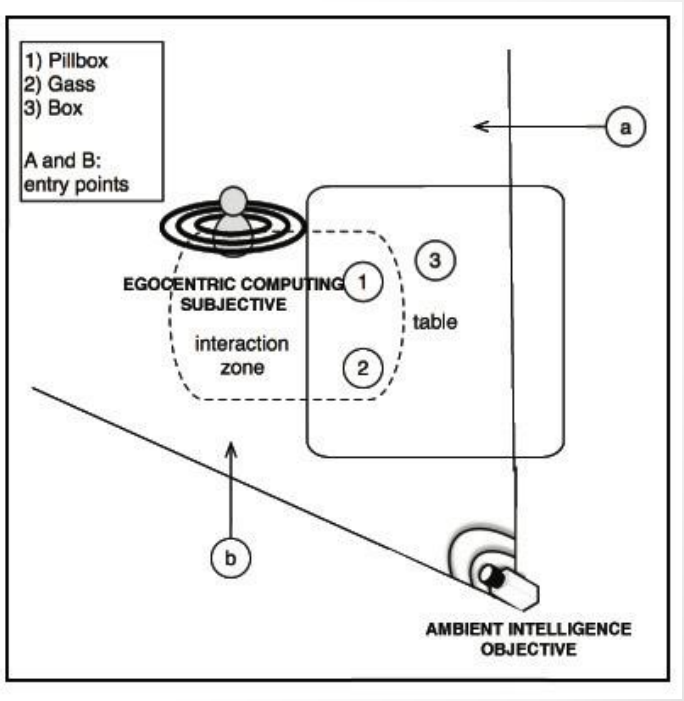
Results

- Hardware
  - Microsoft ® Kinect ™
  - Wearable sensor Shimmer ®
  - Smartphone Samsung GT-I9250
  - Quad-core processor and 8Gb of RAM memory computer

- Settings
  - Parallel Dynamic Time Warping threshold is set using leave-one-out cross validation
  - Weights used on Parallel DTW are experimentally set
  - 400 frames are used to learn a background model
  - Pixels segmented are those with a greater than $\delta = 1.15$ standard deviations

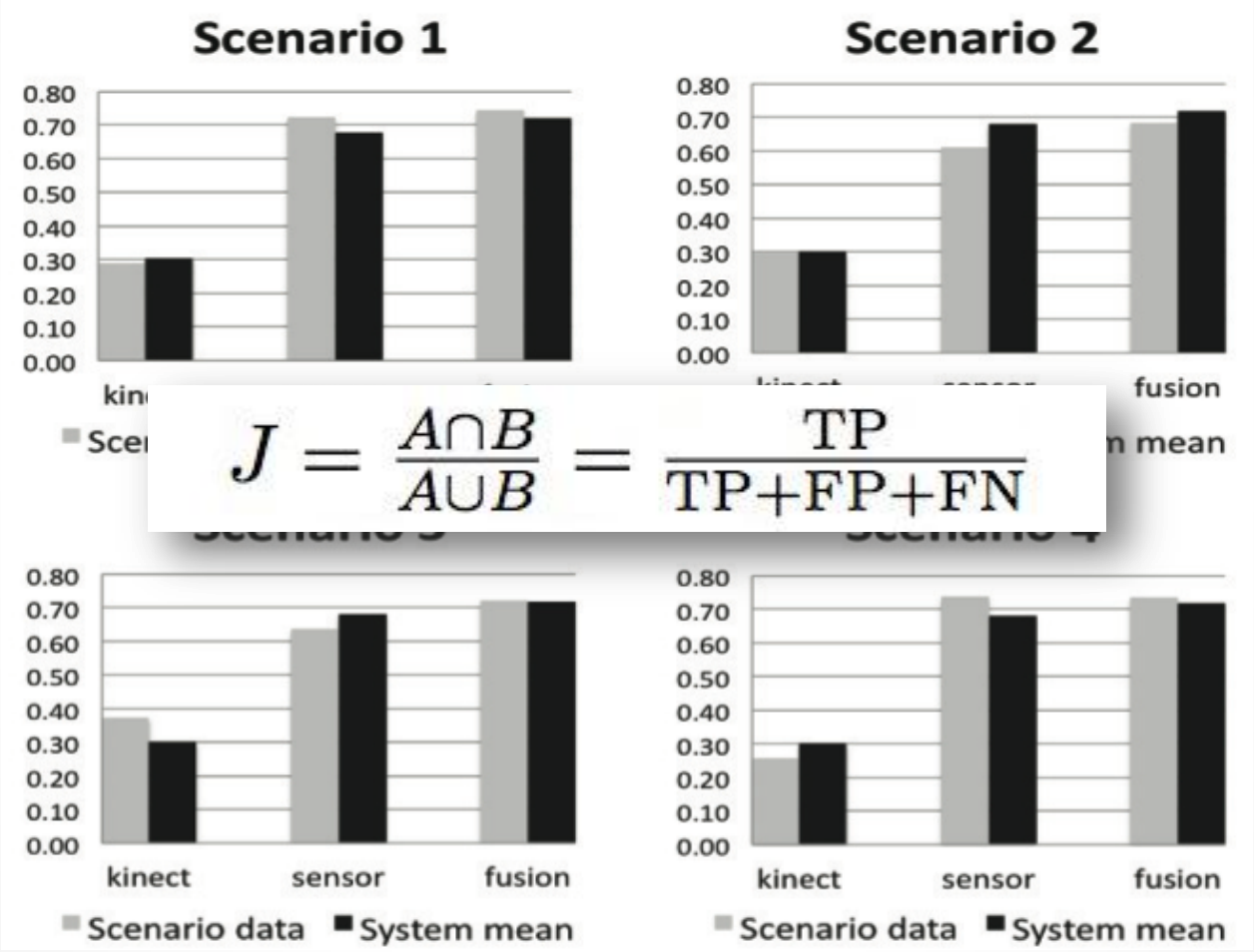
Results

- Dataset
  - 13 different multi-modal synchronized recordings
  - 4 different types of scenes
  - 3 different objects on the table
  - In every scenario, two users appear
  - The patient always takes the pill and drinks water
  - The orientation of the patient changes in order to test occlusion cases



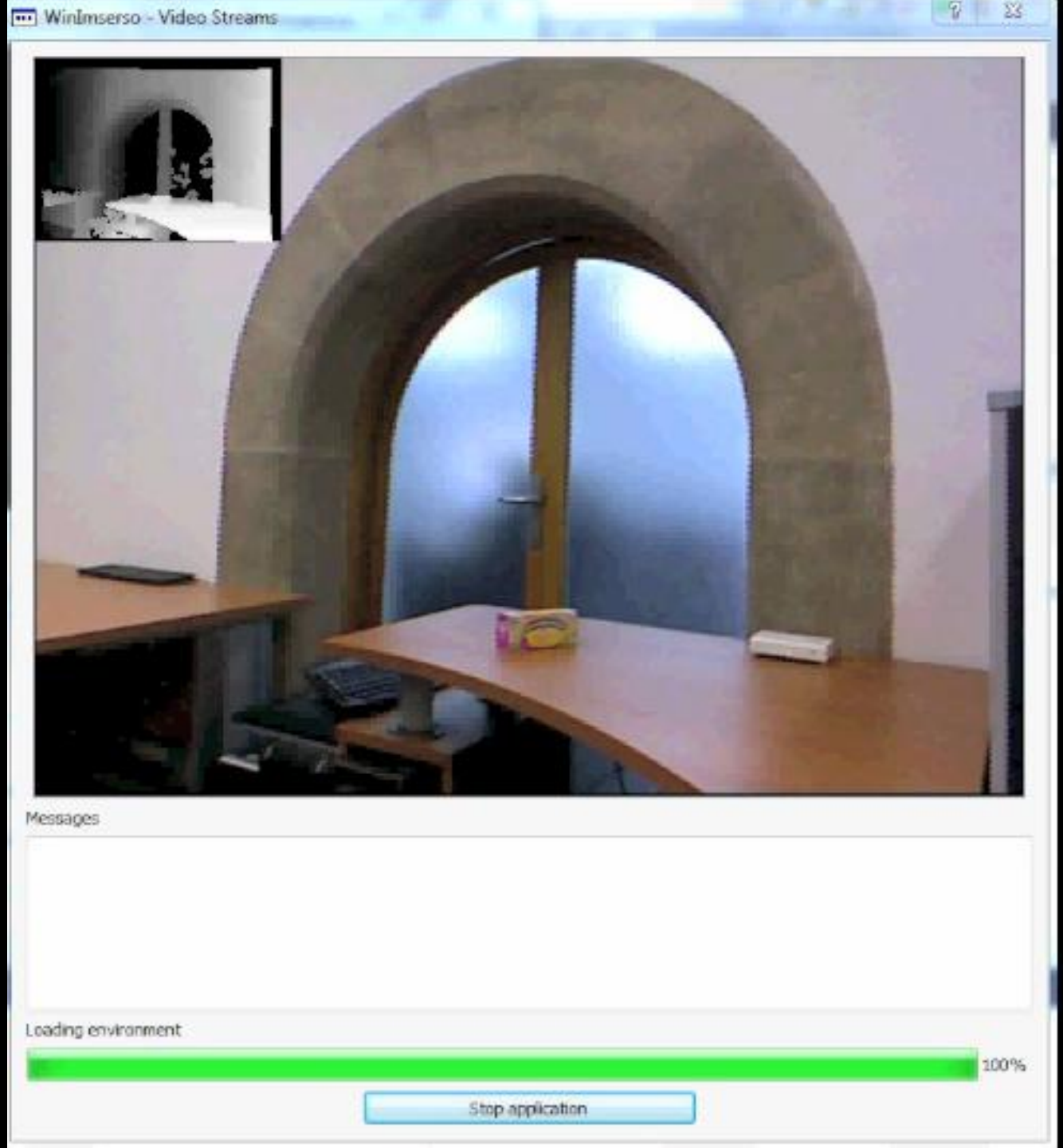
Results

Validation



$$J = \frac{A \cap B}{A \cup B} = \frac{\mathrm{TP}}{\mathrm{TP+FP+FN}}$$

$$J = \frac{A \cap B}{A \cup B} = \frac{\mathrm{TP}}{\mathrm{TP+FP+FN}}$$


WinImserso - Video Streams
Messages
Loading environment
100%
Stop application

Conclusion

- Fusing ambient and egocentric modalities
  - Fast and robust system
  - Capable of recognizing everyday activities with different objects
  - Natural conditions of indoor scenes
  - Feasible real-time processing
- Future work
  - Add more objects and gestures
  - Enlarge the network:
    - Use smartphone built-in to detect actions
    - Use a second Kinect in order to reduce occlusions
  - Implement real cases in the system in order to analyse its real significance
  - We believe in a multidisciplinary research based on Complex Networks, Citizen Sensoring and Behaviour Analysis disciplines.

Acknowledgements:
This work has been partly supported by RECERCAIXA 2011 Ref. REMEDI and TIN2009-14404-C02.