Human Model and Motion Based 3D Human Action Recognition in Multiple View Scenarios

EUSIPCO 2006

C. Canton, J.R. Casas, M.Pardàs

Technical University of Catalonia, Barcelona, Spain

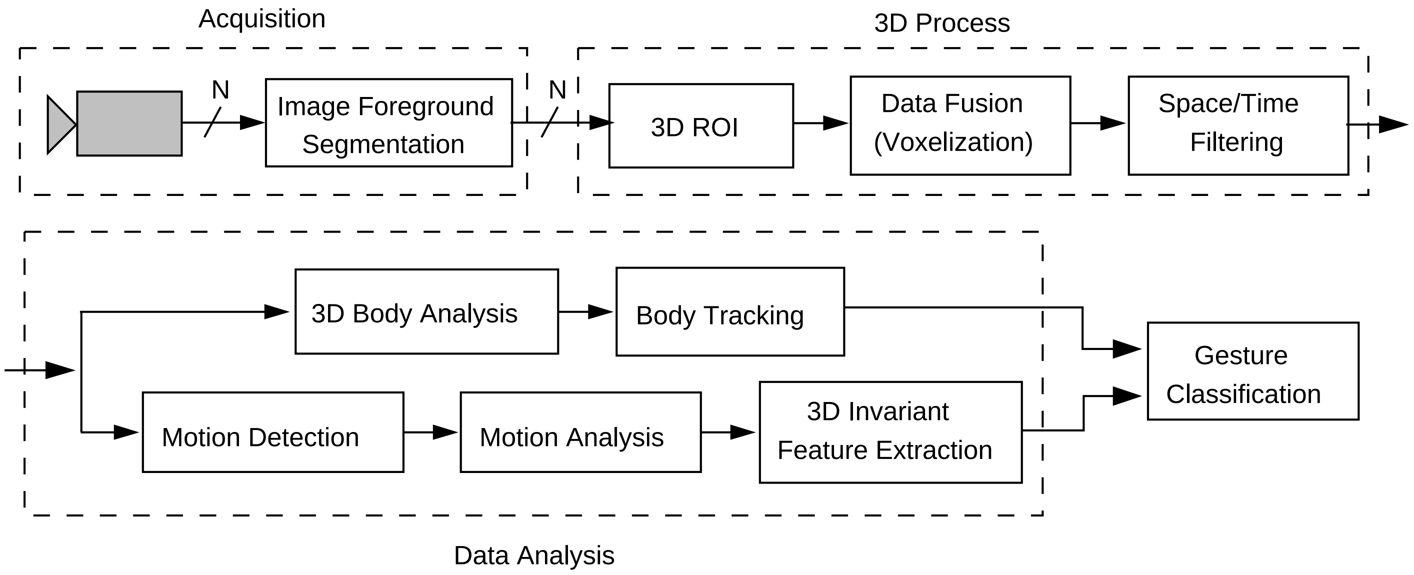
Outline

- Objectives
- System overview
- 3D Data Analysis & Fusion
- Motion Analysis & Feature Extraction
- Human Model Analysis & Feature Extraction
- Classification
- Conclusions & Future Work

Objectives

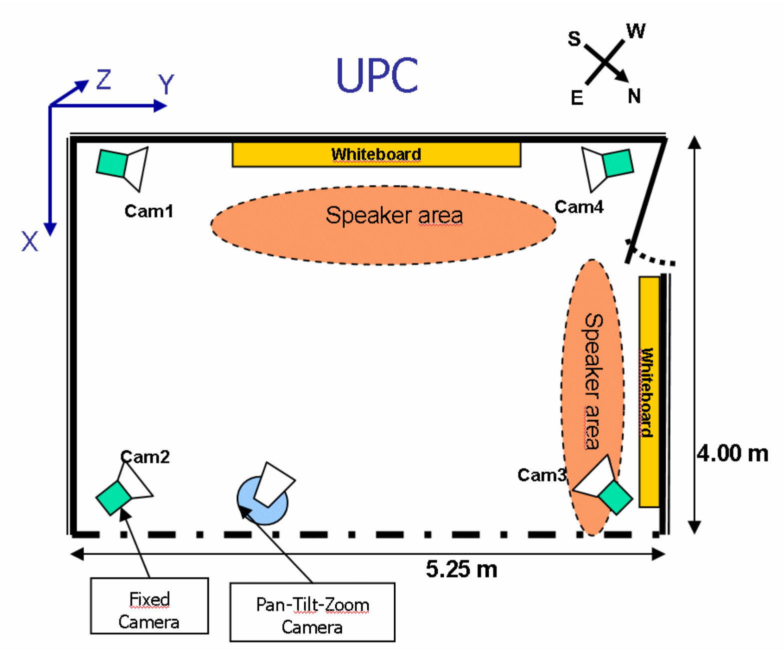
- Aim: Detect and analyze gestures of several people inside a space provided with multiple calibrated cameras
- Recognize a small set of actions (8) by using the data regarded by all cameras
- Tackle typical problems arisen when working with only one camera (occlusions, scale precision,…) by exploiting redundancy among cameras
- Exploit motion and the underlying human body structure of the action perform to enhance robustness

System Overview



UPC application: Smart Rooms

- Human activity monitoring and interpretation of the actions performed in in indoor environments surveyed by multiple fixed cameras



Data sample

3D Data Analysis & Fusion (I)

- Apply a robust Bayesian correspondence algorithm and tracking to detect spatial regions of interest among the foreground segmentation of all views.

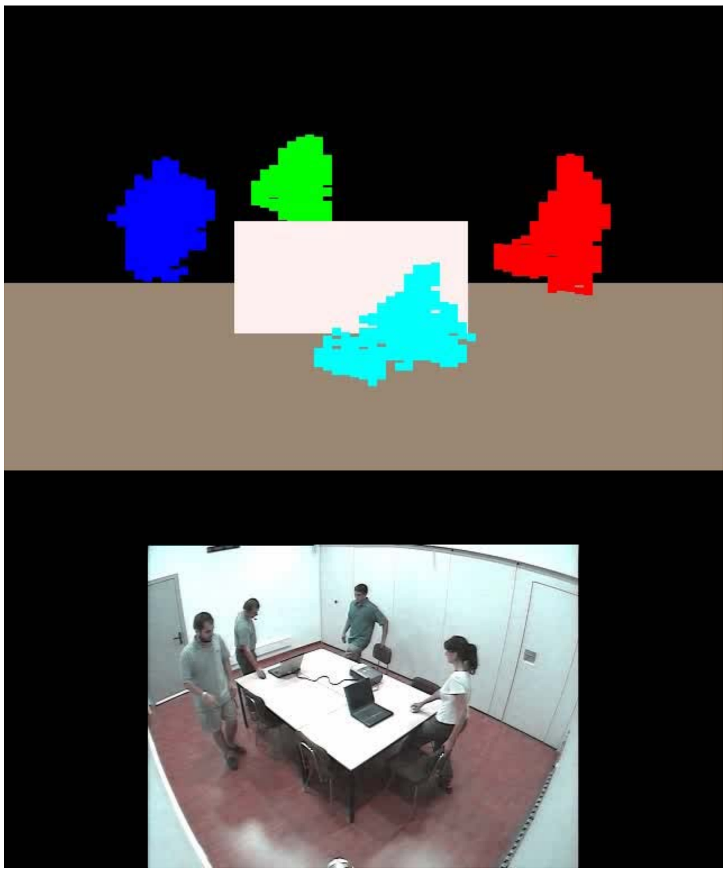
3D Data Analysis & Fusion (II)

- Define a data fusion process:

$$\Omega(x,t) = \left\{ I_n(x,t), \beta^k(x,t), R(\cdot) \right\}$$

  that take into account:
  - Segmented images $I^n(x,t)$ from all cameras
  - 3D Regions of Interest $B^k(x,t)$
  - A generic data fusion method $R(\cdot)$
- Space/time spurious voxel filtering

3D Data Analysis & Fusion (III)



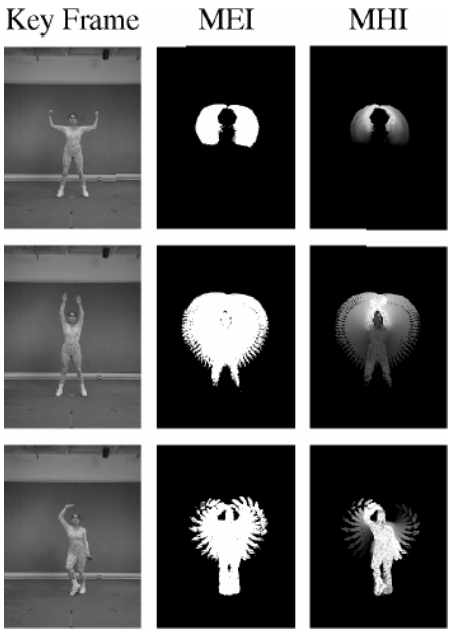
Motion Analysis (I) - Objective

- Define a 3D motion representation extending *Bobick et al.* Motion History Image (MHI) and Motion Energy Image (MEI) to volumes

- Extract robust features over these representations to perform classification

Motion Analysis (II)

- *Bobick et al.* addressed the problem of monocular human gesture recognition by defining a space representation of motion: Motion History Image and Motion Energy Image



Drawbacks:

- **View dependent**
- **Sensitive to occlusions**

Motion Analysis (III)

- Motion History Volume (**MHV**) and Motion Energy Volume (**MEV**) are defined over the voxel reconstruction of the space

MEV

MHV

Motion Analysis (V)

- Bobick overcome occlusions by computing features over *N* different views of the same gesture

INFORMATION FUSION AT FEATURE LEVEL

- We overcome occlusions by generating a data model from all information coming from all views and then computing features

INFORMATION FUSION AT DATA LEVEL

Motion Classification Features

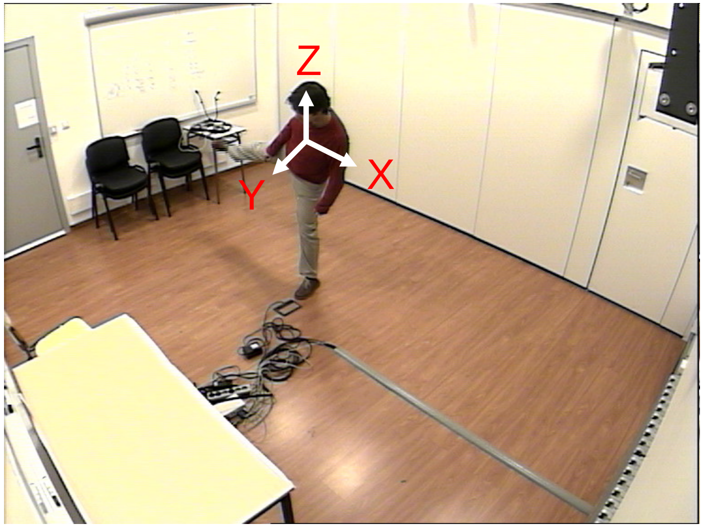
- Informative features derived from low level representations of motion (MHV, MEV) are required
- Features must be invariant to translation, scaling and rotation and robust
- 3D invariant statistical moments constructed through *Lo et al.* (*PAMI* 1989) method proved to be very suitable as features
- The feature vector describing motion is formed by 10 invariant statistical moments (5 computer over MHV and 5 over MEV)

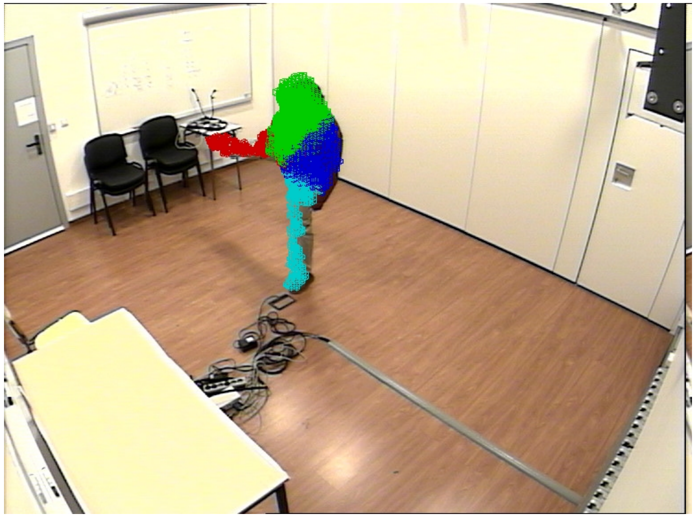
Human Model Analysis (I)

- Objective: Extract position of body parts (arms and legs) through a classification of the voxel reconstruction of the scene

Human Model Analysis (II)

- Compute and track centroid and covariance matrix of the voxel representation of the person under study. Classify all voxels as belonging to one of the categories right/left-arm/leg





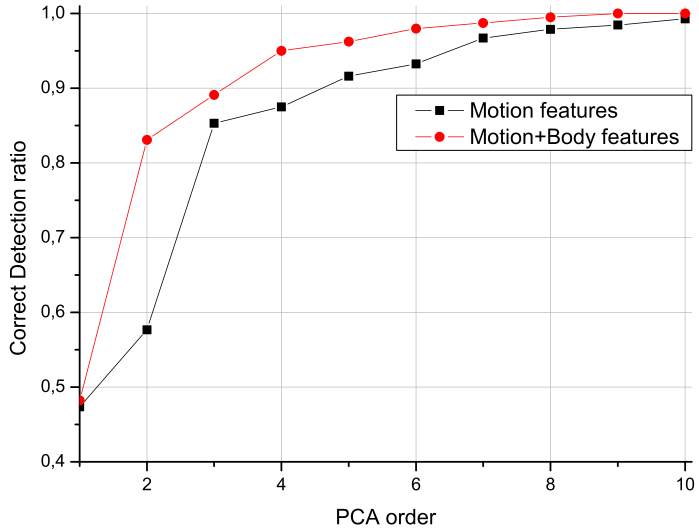
Human Body Features

- 4 body features are defined as the relative amount of motion (MHV) in each body part

Classification

- 14 dimension feature vector is constructed putting together motion and human model features
- PCA is applied to data in order to acchieve dimension reduction
- In the classification stage, a simple Bayes classifier produced good classification results (more sophisticated classification methods are still pending to be tested)
- 8 classes are tested in the classification obtaining an average error probability of *p(error)=0.0154*

Results



Confusion Matrix

	ω_0	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7
ω_0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ω_1	0.0	-	0.0	0.006	0.0	0.0	0.0	0.0
ω_2	0.0	0.0	-	0.010	0.0	0.0	0.0	0.0
ω_3	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0
ω_4	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0
ω_5	0.0	0.0	0.0	0.0	0.107	-	0.0	0.0
ω_6	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0
ω_7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-

Conclusions

- A multi-view human gesture analysis technique based on a data fusion scheme is presented
- Performance results with the feature fusion approach to this problem by *Bobick et al.* are similar <u>but</u> occlusion can be better handled by our data fusion method and complexity is considerably reduced
- Body model information increases performance and provides human limb position information

Future research

- Features derived from the incoming 3D data are under research (Zernike moments, Fourier descriptors,…)
- More sophisticated classification schemes are being tested to increase correct detection rates

THANK YOU FOR YOUR ATTENTION