People Detection
and Video Understanding

Francois BREMOND

INRIA Sophia Antipolis – STARS team
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 various sensors (especially 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 *(vandalism, bank attack, cooking, washing dishes, falling)*
- for small groups *(fighting)*
- for crowd *(overcrowding)*
- for interactions of people and vehicles *(aircraft refueling)*
Generic Platform for activity understanding

Posture Recognition

Detection → Classification → Tracking → Activity Recognition → Actions

Activity Models

Person inside Kitchen
People detection and tracking
People detection: faster R-CNN on ETHZ

Person detections with $p(\text{person} \mid \text{box}) \geq 0.8$
Motivation - Action Recognition

Hollywood dataset
Motivation - Action Recognition
UCF Sports dataset
Motivation - Action Recognition
Daily Living datasets (Rochester Univ.)
Action Recognition using Bag of Words

M. Koperski

Videos ➔ Feature detector ➔ Feature descriptor ➔ BOW model ➔ Histograms of codewords ➔ Non-linear SVM

Codeword defined as a Descriptor cluster
Violence Recognition Framework, P. Bilinski

1. **Input Video**
   - Feature Detection (Improved Dense Trajectories)

2. **Feature Description**
   - (TS, HOG, HOF, MBH)

3. **Video Representation**
   - (Improved Fisher Vectors)

4. **Classification**
   - (SVM)

**Violence**
- Pub
- Street
- Football Stadium

**Non-violence**
- School
- Movies Analysis
- Football Stadium

- Violence
- Non-violence
Gender recognition using smile: A. Dantcheva

Spatio-temporal features based on dense trajectories represented by a set of descriptors encoded by Fisher Vectors.
Toyota Smart-Home
Large scale daily living dataset
Toyota Smart-Home
Large scale daily living dataset
Issues in Action Recognition using Deep CNNs

Deep Convolutional Neural Networks (CNN)

Images

- Large Annotated data (Imagenet)
- Architecture Suitable for Images with good resolution

Videos: How to capture motion information in CNN?

- Stacking of frames

- Capture motion independently or not: several stream CNNs
  - One ConvNet to capture static (frame based) visual information.
  - Another ConvNet to capture motion information (like Optical Flow, but expensive)
  - Other Nets to capture motion on longer scales or together (Siamese)
  - Other Nets to capture object-ness.

C. Roberto de Souza, A. Gaidon, E. Vig, and A. Lopez. Sympathy for the Details: Dense Trajectories and Hybrid Classification Architectures for Action Recognition, ECCV 2016
Proposed a representation derived from human pose using Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields

The descriptor aggregates motion and appearance information along tracks of human body parts using P-CNN : Pose-based CNN Features for Action Recognition
Toyota Smart-Home
Large scale daily living dataset

Pour water for tea

Prepare tea
Activity monitoring in Greece Hospital with AD patients

Visualization of older adult performance while accomplishing the semi-guided tasks.
Conclusion - video understanding

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

Perspectives:

- Generate totally unsupervised models
- Use finer features as input for the algorithm (head, posture, facial gesture…)
- Generating language description for the activity models
- Generic activity models (cross scenes), Adaptive learning
- More semantics, emotion, mental states.

4 PhD open topics:

- Kontron: People Tracking using Deep Learning algorithms on embedded hardware
- ESI: People Re-Identification using Deep Learning
- Wildmoka: Video based Action Recognition using Deep Learning
- Nice Hospital: Uncertainty Management and Activity Recognition