

Action recognition

Cordelia Schmid

INRIA Grenoble



Action recognition – examples

- Short actions, i.e. answer phone, shake hands

answer phone



hand shake



Hollywood dataset

Action recognition – examples

- Activities/events, i.e. birthday party, grooming an animal

Birthday party



Grooming an animal



TrecVid Multi-media event detection task (MED)

Action recognition - tasks

- Action classification: assigning an action label to a video clip



Making sandwich: present
Feeding animal: not present

...

Action recognition - tasks

- Action classification: assigning an action label to a video clip



Making sandwich: present
Feeding animal: not present
...

- Action localization: search locations of an action in a video



Action recognition - tasks

- Action localization + interaction with an object



Why automatic video understanding?

Huge amount of video is available and growing daily

BBC Motion Gallery



TV-channels recorded since 60's



>34K hours of video upload every day



~30M surveillance cameras in US
=> ~700K video hours/day

Why automatic video understanding?

- Query for videos in professional Archives and YouTube
- Analyze and describe content of videos



Education: How do I
make a pizza?



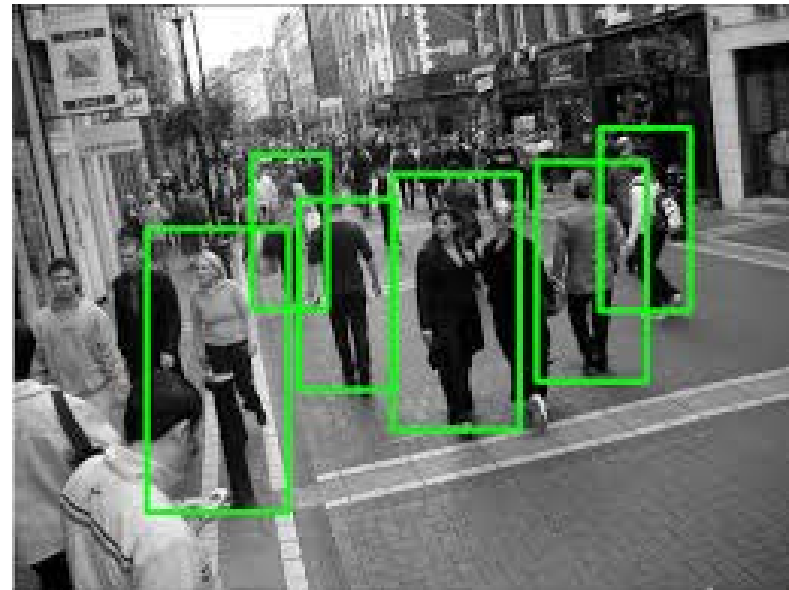
Sociology research:
Influence of character
smoking in movies

Why automatic video understanding?

- Car safety & self-driving and video surveillance
 - Detection of humans (pedestrians) and their motion, detection of unusual behavior



Courtesy Volvo



Courtesy Embedded Vision Alliance

Action recognition - difficulties

- Large variations in appearance
 - Viewpoint changes
 - Intra-class variation
 - Camera motion

Variation in appearance: viewpoint change



Variation in appearance: intra-class variation



Variation in appearance: camera motion



Action recognition - difficulties

- Large variations in appearance
 - Viewpoint changes
 - Intra-class variation
 - Camera motion
- Manual collection of training data is difficult
 - Many action classes, rare occurrence
 - Pose and object annotation often a plus
- Action vocabulary is not well defined
 - What is the action granularity?
 - How to represent composite actions?

Action recognition – approaches

- Action recognition from still images
 - Human pose + interaction with objects
 - Extract key frames from video



Results on PASCAL VOC 2010 Human action classification dataset [Prest et al., PAMI 2012]

Action recognition – approaches

- Motion information necessary to disambiguate actions

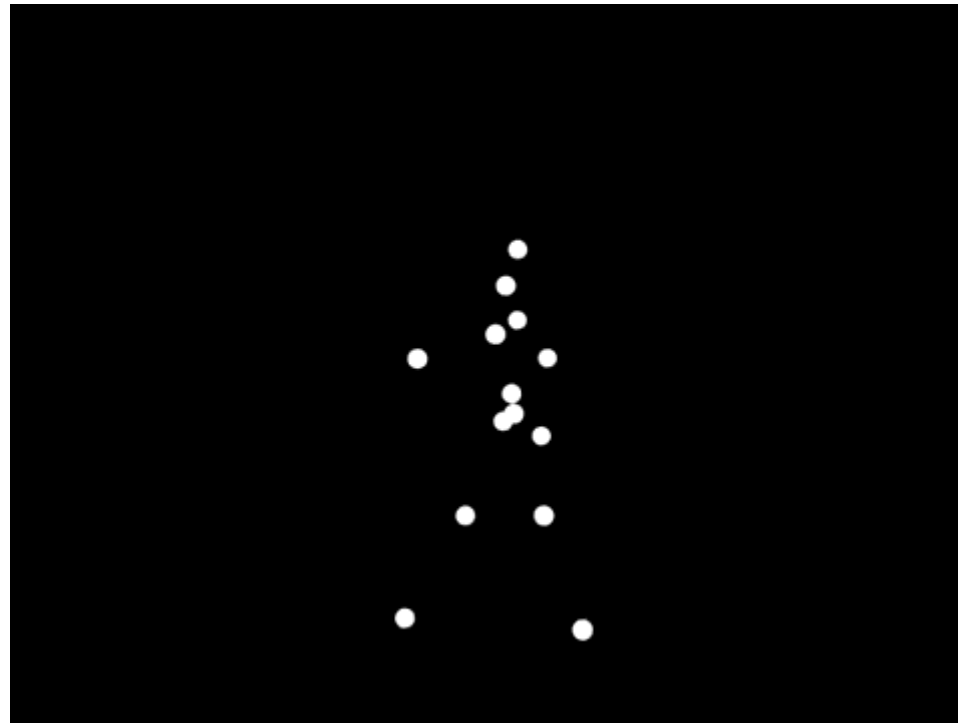


Open or close door?

- Motion often sufficient by itself

Motion perception

- Gunnar Johansson [1973] pioneered studies on sequence based human motion analysis
- Moving light displays enable identification of motion, familiar people and gender



male walker

Overview

- Optical flow
- Video classification
 - Bag of spatio-temporal features
- Action localization
 - Spatio-temporal human localization
- Action description based on human pose
 - Human pose description based on CNN features