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

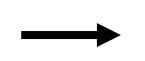- Action classification: assigning an action label to a video clip

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

Action recognition - tasks

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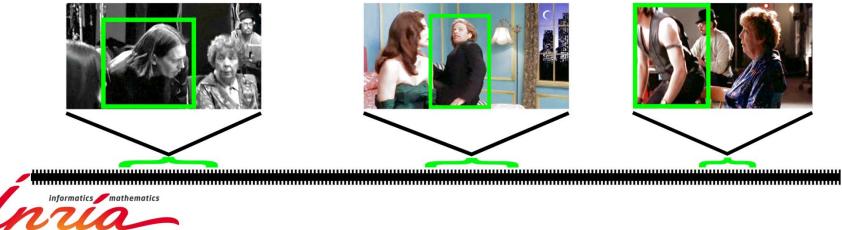

Making sandwich: present
Feeding animal: not present
...

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

Action recognition - tasks

- Action localization + interaction with an object

[Prest et al., PAMI 13]

Why automatic video understanding?

Huge amount of video is available and growing daily

BBC Motion Gallery

ina — TV-channels recorded since 60's

You Tube — >34K hours of video upload every day

CCTV SURVEILLANCE CAMERA — ~30M surveillance cameras in US => ~700K video hours/day

Inria — informatics mathematics

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
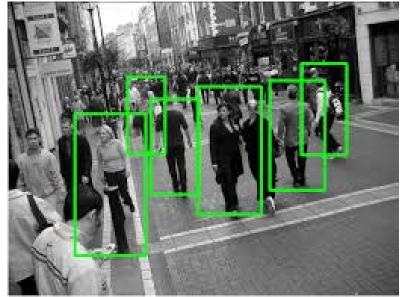smoking in movies

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

Machine visual perception - applications

- Complete description (story) of a video

As the headwaiter takes them to a table they pass by the piano, and the woman looks at Sam. Sam, with a conscious effort, keeps his eyes on the keyboard as they go past. The headwaiter seats Ilsa...

Machine visual perception - applications

- Complete description (story) of a video

As the headwaiter takes them to a table they pass by the piano, and the woman looks at Sam. Sam, with a conscious effort, keeps his eyes on the keyboard as they go past. The headwaiter seats Ilsa...

Machine visual perception - applications

- Complete description (story) of a video

As the headwaiter takes them to a table they pass by the piano, and the woman looks at Sam. Sam, with a conscious effort, keeps his eyes on the keyboard as they go past. The headwaiter seats Ilsa...

informatics mathematics
Inria

Machine visual perception - applications

- Complete description (story) of a video

As the headwaiter takes them to a table they pass by the piano, and the woman looks at Sam. Sam, with a conscious effort, keeps his eyes on the keyboard as they go past. The headwaiter seats Ilsa...

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

Playing Instrument   Reading   Taking Photo   Riding Horse   Walking

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

Action recognition – approaches

- Motion information necessary to disambiguate actions

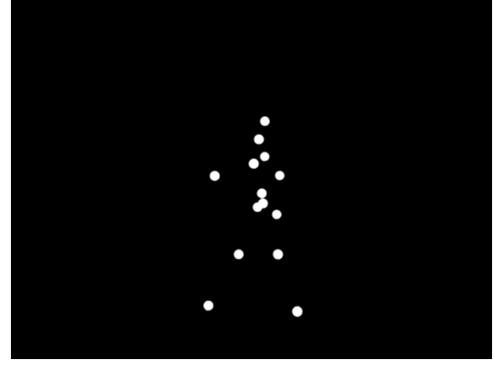Open or close door?

- 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