Human Detection in Difficult Scenarios by Combining Motion and Appearance

C. Beleznai, J. Puckmayr, N. Viertl, P. Sommer

ADVANCED COMPUTER VISION
Ein Unternehmen der Austrian Research Centers

Motivation

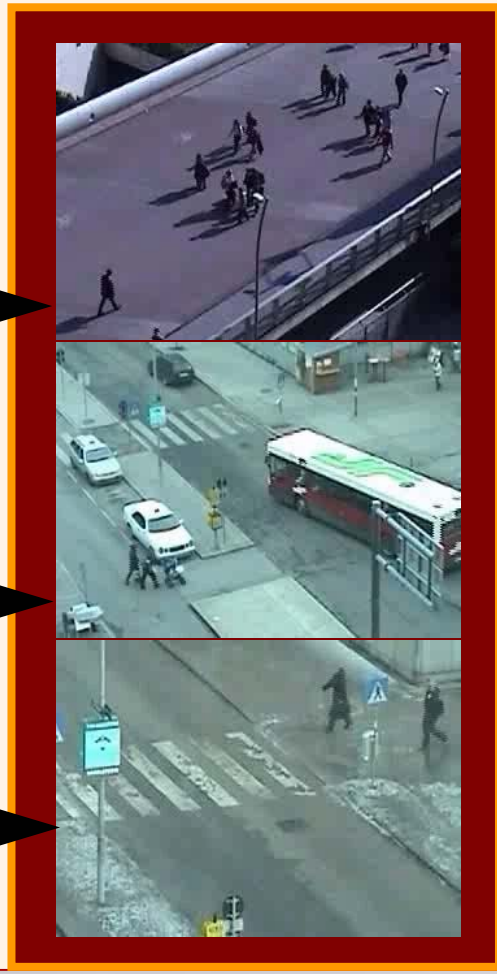
Reliable detection of moving humans by a static camera.

Detection probabilities less than one

- Occlusions
- Low contrast

False alarms by:

- Shadows
- Moving vehicles
- Shaking camera



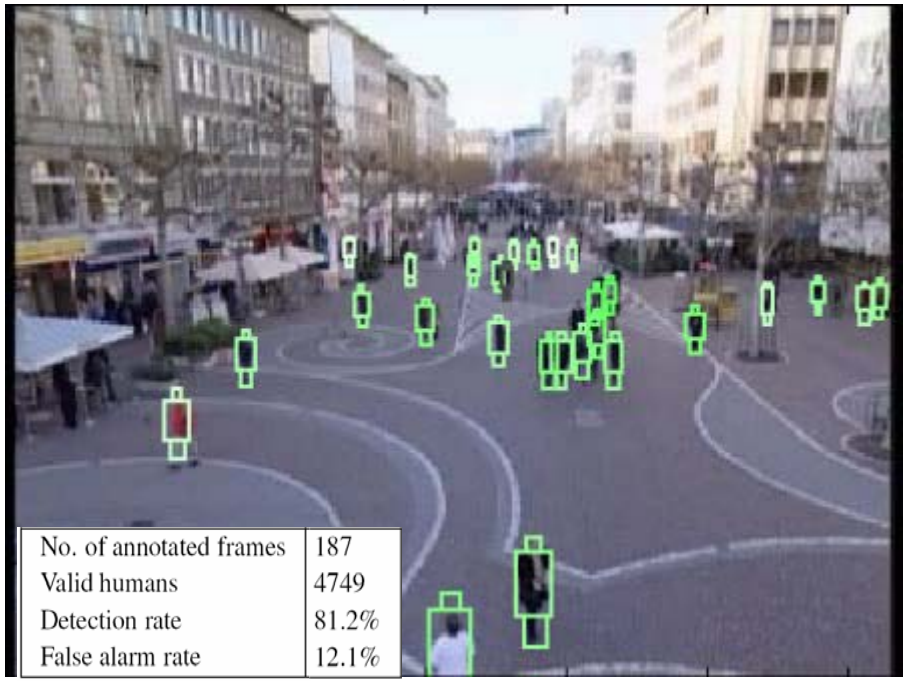
Contents

- **Detecting humans by motion detection**
  - Data-driven clustering complemented with the use of a human model.
- **Combining motion detection with statistical learning**
- **Results**
  - Crowd
  - Shadows
  - Moving vehicles, low contrast, shaking camera
- **Outlook**
  - Video data sets for evaluation / basis for cooperation

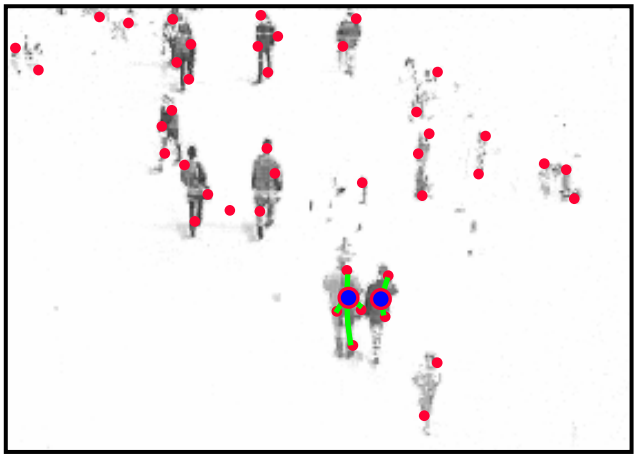
Motion-based human detection

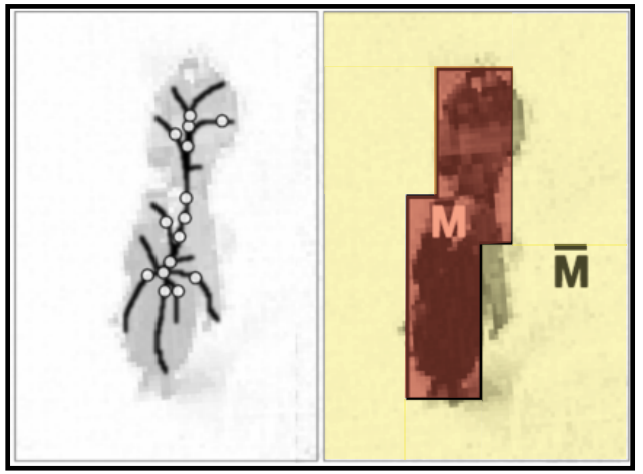
■ Fast clustering of the difference image

- Mean Shift procedure using integral images.
- Model-based validation of hypothesized configurations:
  - Removing spurious detections
  - Occlusion handling



No. of annotated frames	187
Valid humans	4749
Detection rate	81.2%
False alarm rate	12.1%





C. Beleznai, B. Frühstück and H. Bischof,
"*Tracking Multiple Humans using Fast Mean Shift Mode Seeking*",
PETS 2005 Workshop

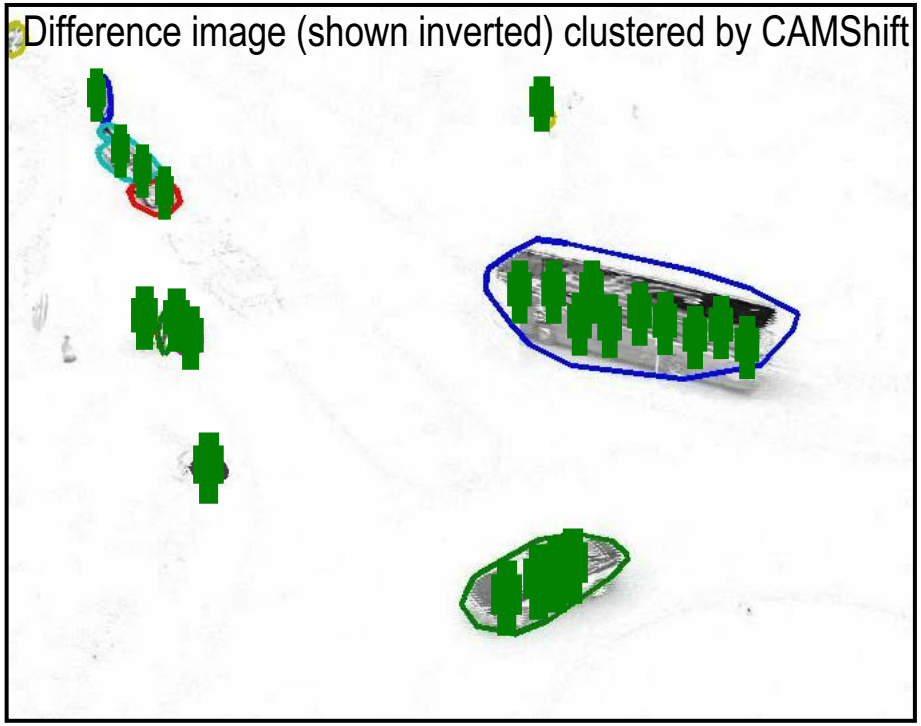
Dual-hypothesis propagation

Problem:

Model-based human detection is not discriminative enough.

Ambiguity:

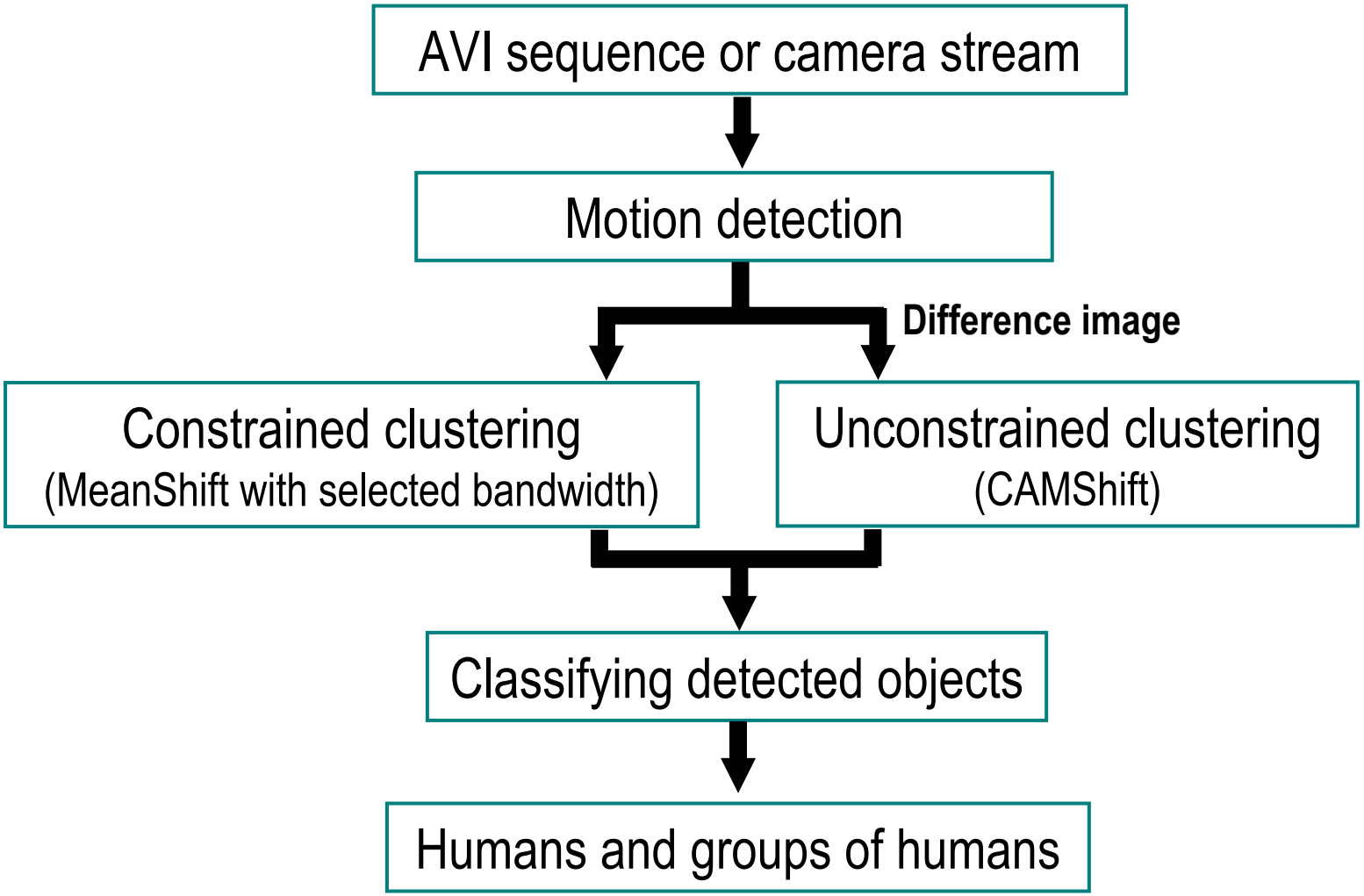
- **Group of humans**
- **Vehicles, motion clutter**



Approach:

1. **Data-driven clustering (CAMShift)**
2. **Model-based human detection**
3. **Classifying the output of motion detection** *{human, other}*
4. **Region-based statistics** *{group of humans, other}*

Combining motion detection with classification

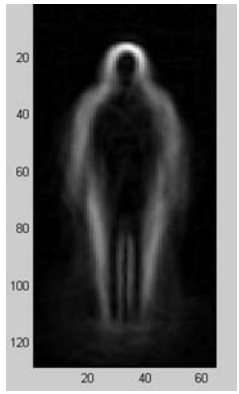


Classifier-based human detector

- **Features:**
  - Integral orientation histograms
- **Classifier:**
  - Cascade of weak classifiers trained by AdaBoost

First 16 eigenvectors

Mean edge strength



Detection and tracking results – Visual Surveillance Sequence A

Sequence VS_A

- 3930 frames
- 25 fps
- 720 × 576 pixels
- human size:
  22-32 pixels

Difficulties:

- shadows cast by humans
- crowded situations
- moving flag
- moving vehicles

Detection and tracking results – Visual Surveillance Sequence B

Sequence VS_B

- 4773 frames
- 25 fps
- 720 × 576 pixels
- human size:
  18-34 pixels

Difficulties:

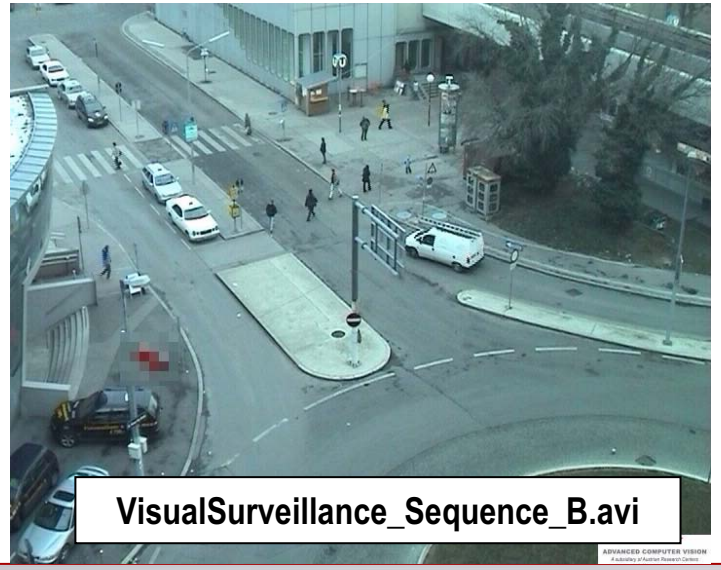
- crowded situations
- moving vehicles
- reflections

Results: Shaking camera

Results: low (human - background) contrast

Video data

- **Ground truth will be soon available**
  - Annotation according to the CAVIAR ground-truthing scheme
  - Bounding box positions
  - Trajectories
- **Testing your algorithms on the data**



Conclusion

- **Promising detection and tracking improvements for challenging data:**
  - No ghost objects.
  - Eliminating detections caused by shadows.
  - Spurious detections by motion clutter greatly reduced.
  - Operating in real-time (8-12 fps, 3 GHz PC)
- **Quantitative analysis is needed to assess improvements.**