Taking Multi-Object Tracking to the Next Level: People, Unknown Objects, and Carried Items

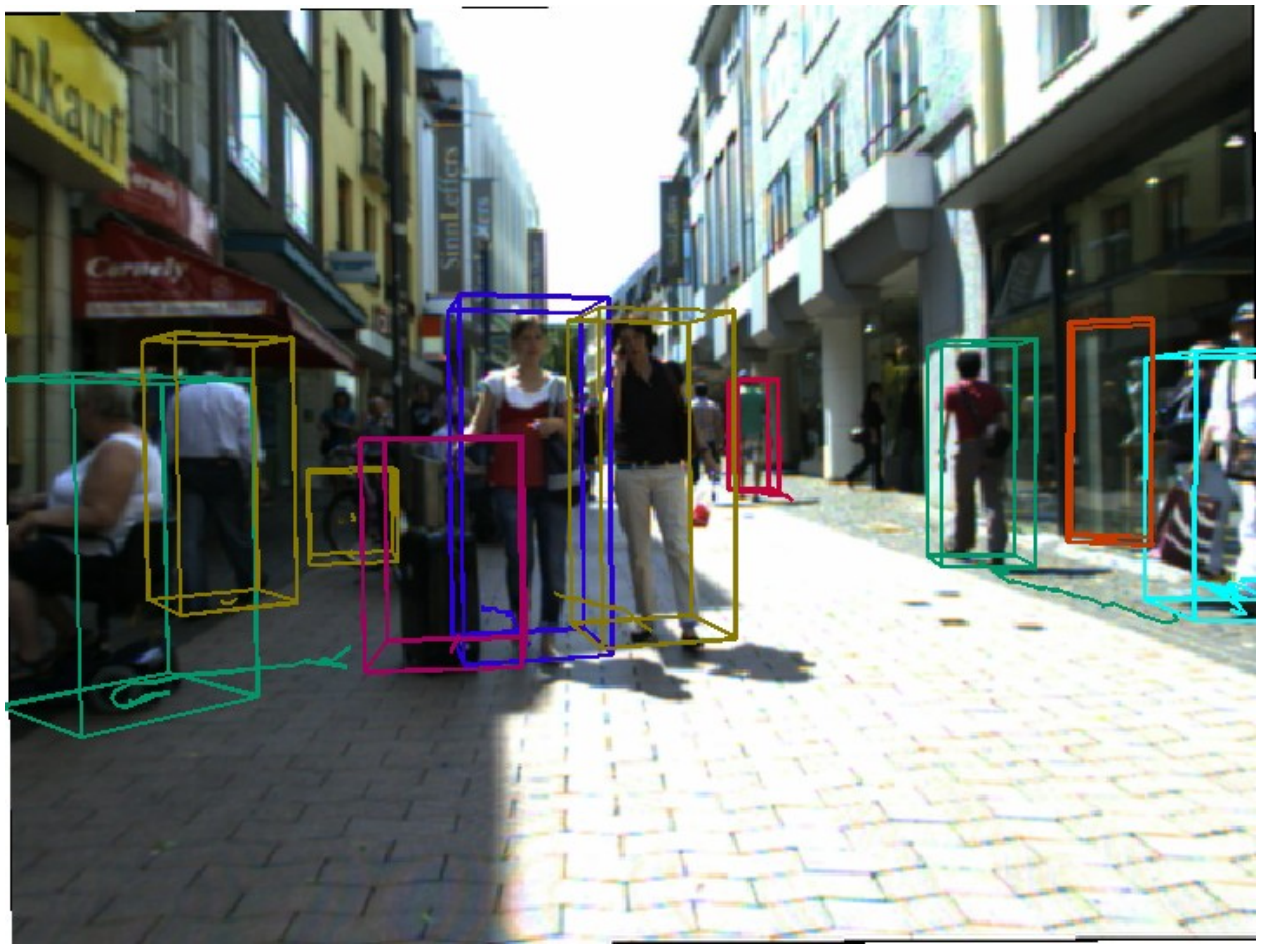
Review by Igor Bogoslavskyi

Department Of Computer Science
Albert Ludwigs University Freiburg
e-mail: bogoslai@informatik.uni-freiburg.de

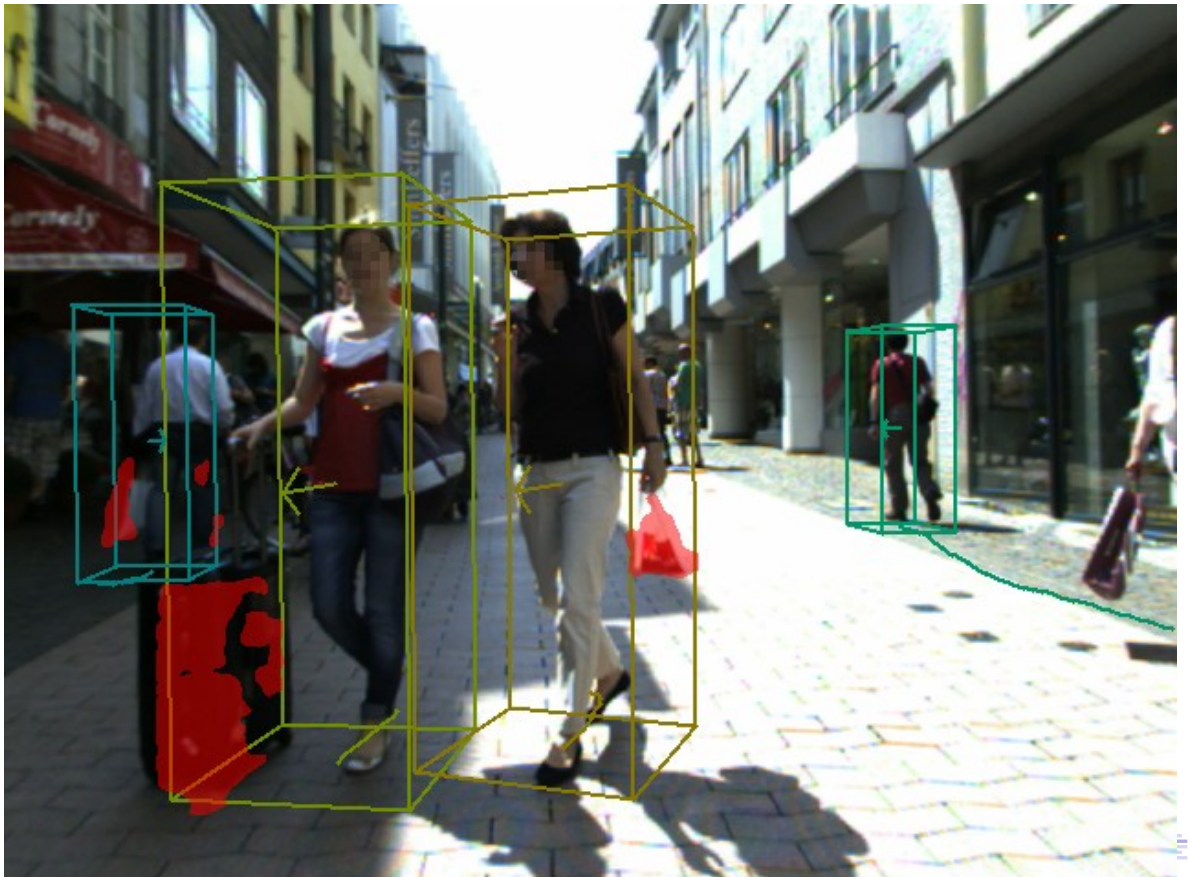
2.02.2013

Motivation

- Tracking objects in the moving scene is an important task in mobile robotics



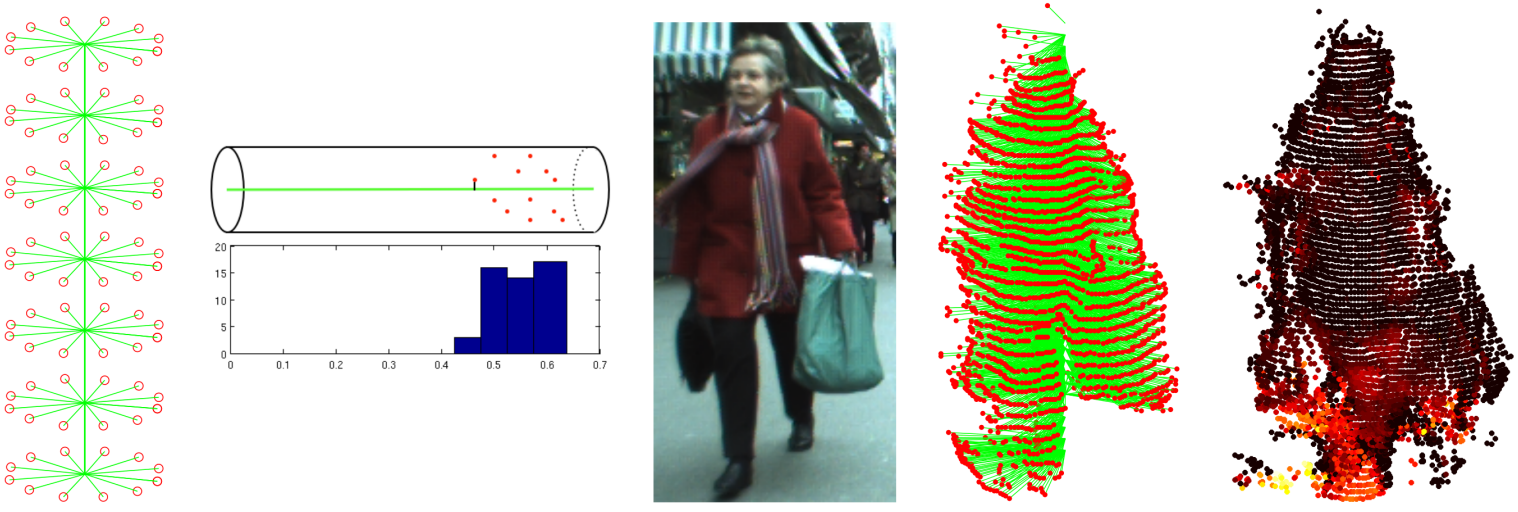
- Previously only *tracking-by-detection* approaches. These need pre-trained detector models.
- It is important to recognize and track other objects in peoples' surroundings.
- Methods that can detect and track also novel object types and learn models for them on-line are needed.



Problem Description

- The problem of detecting novel objects is not trivial.
- To do that one has to answer an even harder question - *what is an object*.
- This itself involves segmenting object from the video stream input.

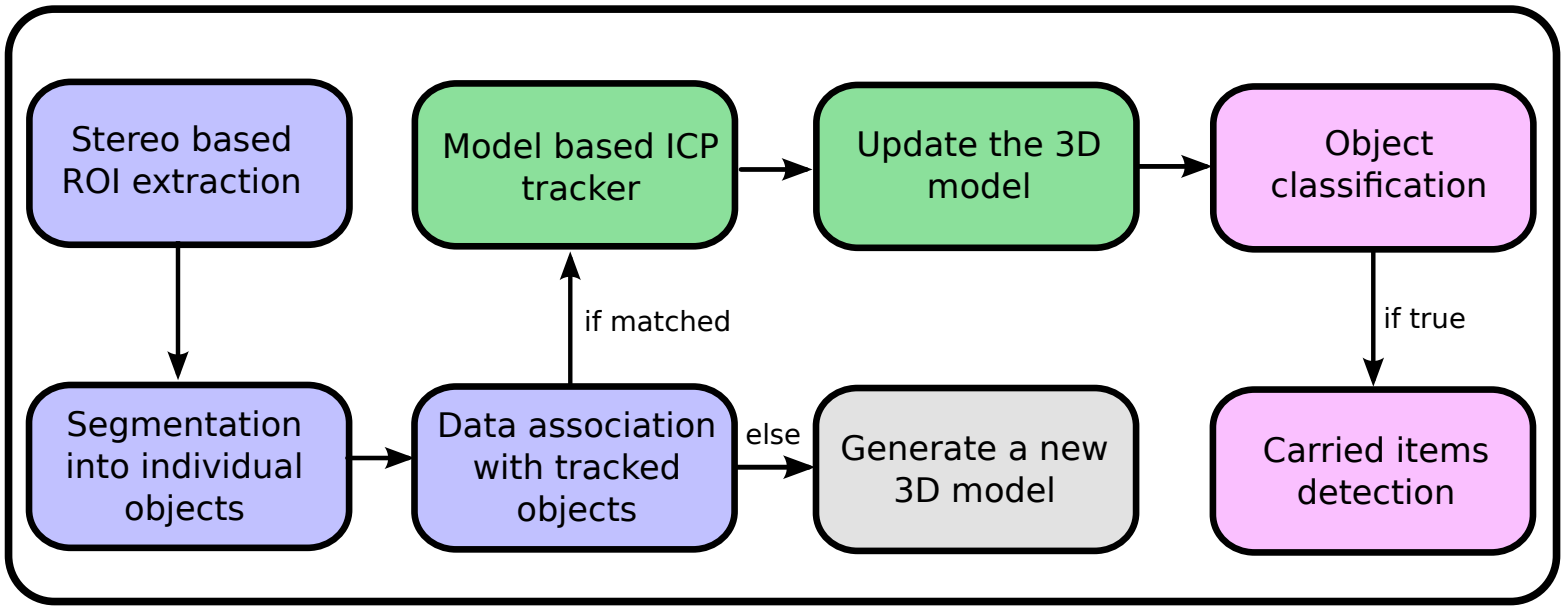
3D Model Representation (Generalized Christmas Tree)



(a) (b) (c) (d) (e)

- The model is composed of a vertical axis and several layers of equally spaced horizontal rays.
- Along each ray, 3D points are stored in distance histogram.

Tracking-Before-Detection Approach

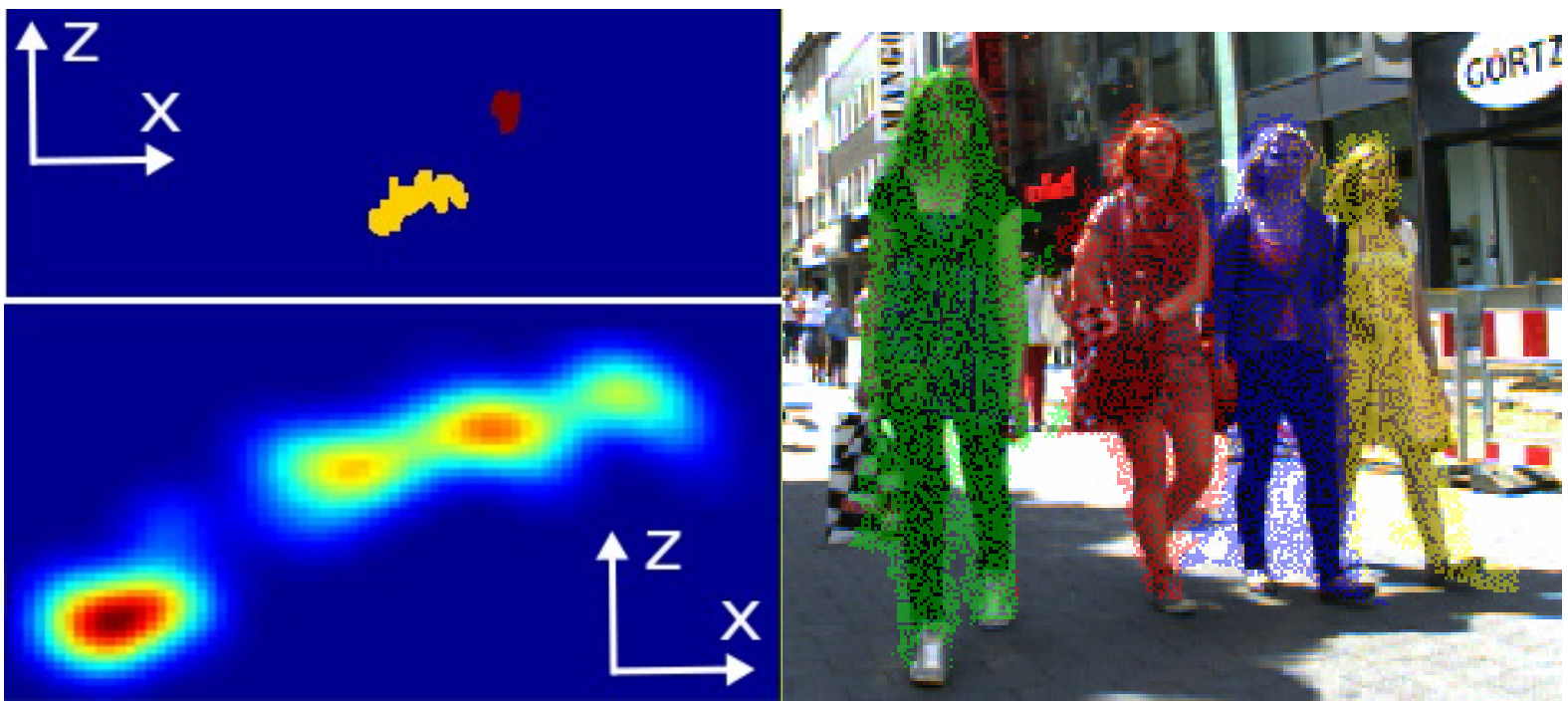


Extracting Regions of Interest (ROIs)

- Project the 3D points within 2m height corridor onto the ground plane.
- Take 2D histogram of these points.
- Remove regions that continuously extend beyond height of 2m to exclude walls and other elevated objects.
- Threshold the histograms and group to connected components via 8-neighborhood.

Segmentation Into Individual Objects

- Problem: people walking close to each other are still connected in the ground projection.
- Smooth and segment the original histogram via Quick Shift algorithm.
- The result is a segmentation of the ROIs into individual objects.



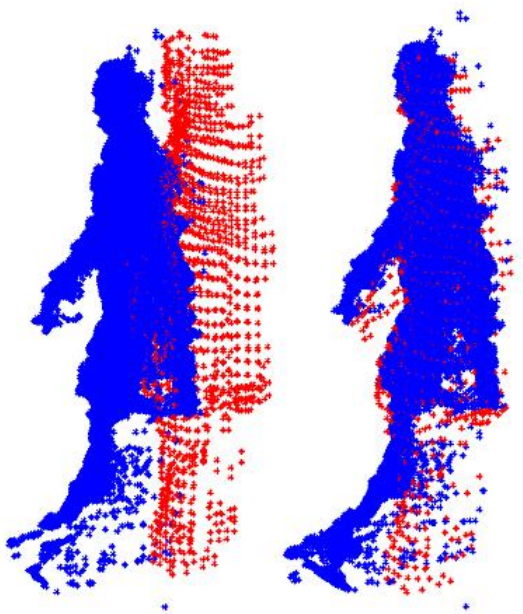
Data Association With Tracked Objects

Associating ROIs with existing tracks.

- Match newly segmented ROIs to each track's ROI from previous frame.
- Match is successful if the over-union of the ROIs' ground projection footprints is over 50%.
- Start a new track for all ROIs that cannot be associated.

Match and Update the Model

- Once the ROI is matched to previous frames, we want to match and update the model.
- Match the GCT from current frame to GCT from previous frames via ICP.
- After the model were matched we want to update current model to store as much as possible information about the tracked object.



(a) (b)

Object Classification and Tracking

- Pass the newly generated track to person/non-person detector.
- Evaluate the detector only for a small region around the back-projected segmented 3D points.
- No further classification needed for current object.

Pedestrian Model

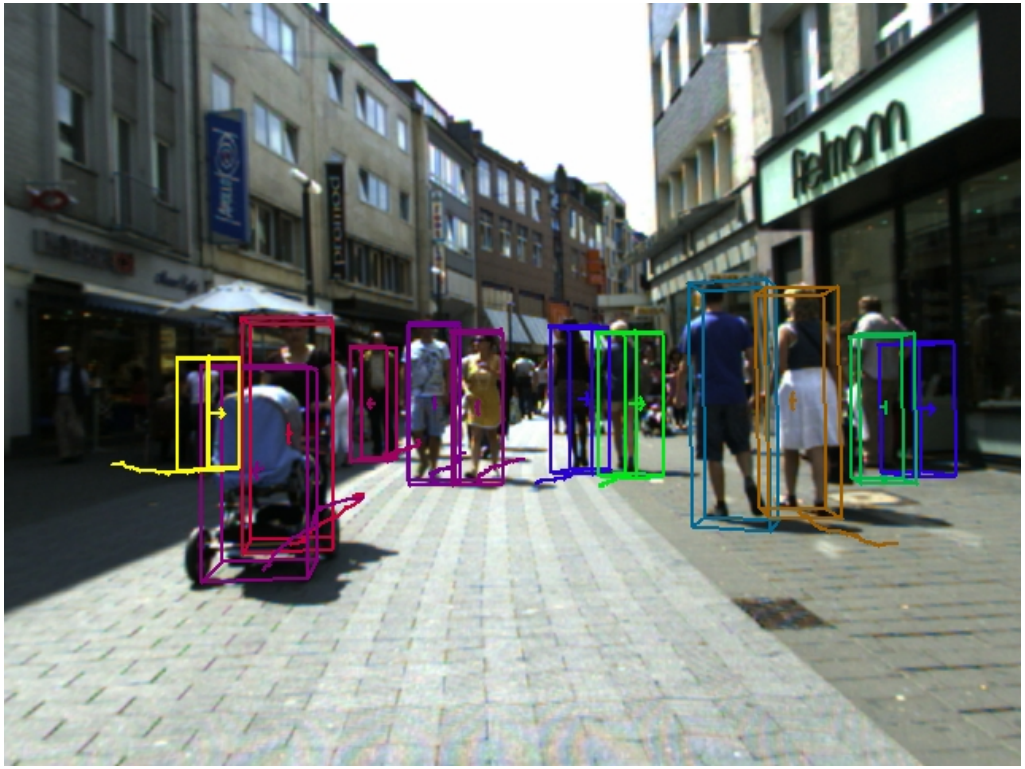
- The 3D model is constantly refined to provide as much detail as possible.
- Compare the on-line model to a learned statistical shape of pedestrians (right image)
- Detect deviations, that cannot be explained by variation in GCT model.

Carried Item Segmentation

Carried item segmentation is based on Conditional Random Fields.

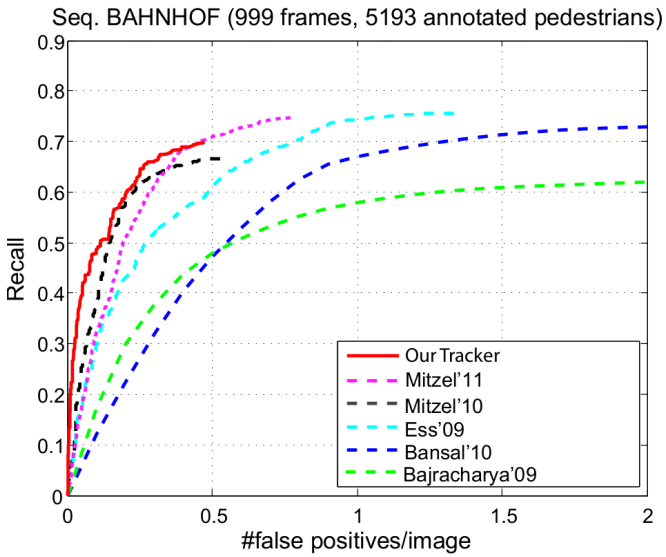
- Unary potentials are based on the Bhattacharyya distance between the distance histograms of the on-line tracked and learned model rays.
- A weighting function in a direction orthogonal to the ground plane was added to unary potential as a prior that carried items are usually not in the leg area.
- Pairwise potentials - contrast-sensitive Potts model based on image colors was used.

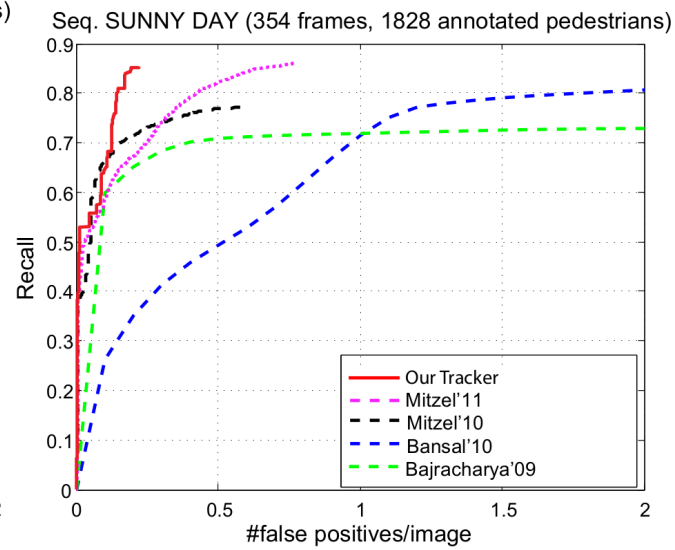
Datasets

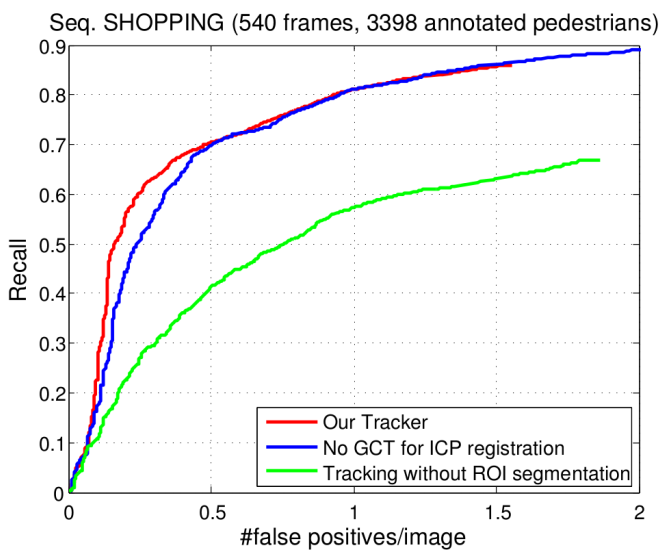


- Three different datasets captured from a stereo camera setup were used:
  - BAHNHOF: 999 frames with 5193 pedestrian annotations.
  - SUNNY DAY: 354 frames with 1867 pedestrian annotations.
  - SHOPPING: over 540 frames with 3398 pedestrians annotations.

Pedestrian Tracking Performance





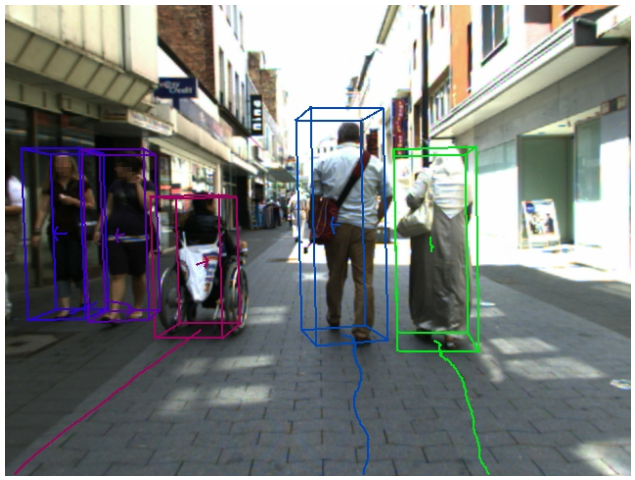


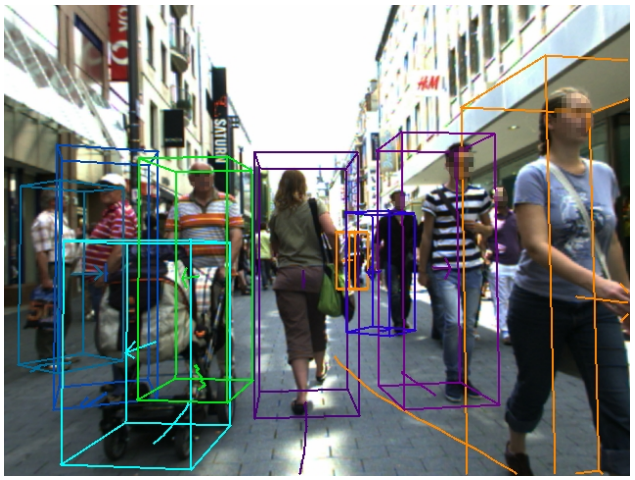
Object Tracking Performance

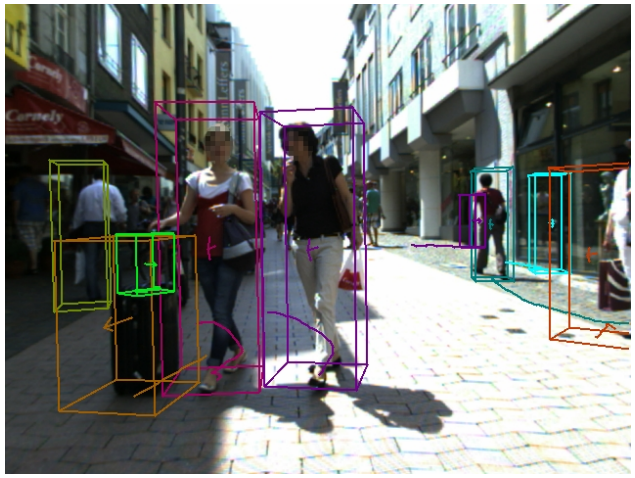
Object types	mostly tracked ($> 80\%$)	nearly tracked ($> 50\%$)	not tracked ($< 20\%$)
Child stroller	5	0	0
Walking aid	0	1	0
Suitcase	1	1	0
Wheelchair	2	0	0
Garbage bin	6	1	0
Bicycle	2	3	4
Advertising rack	0	2	3
Ticket dispenser	1	1	0

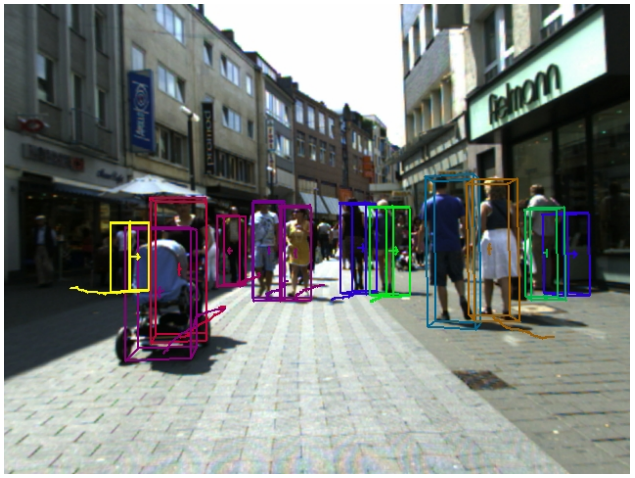
- Tracking the unknown objects listed in Tab. 8
- 6060 frames of video material.

Pedestrian/Object Tracking Examples



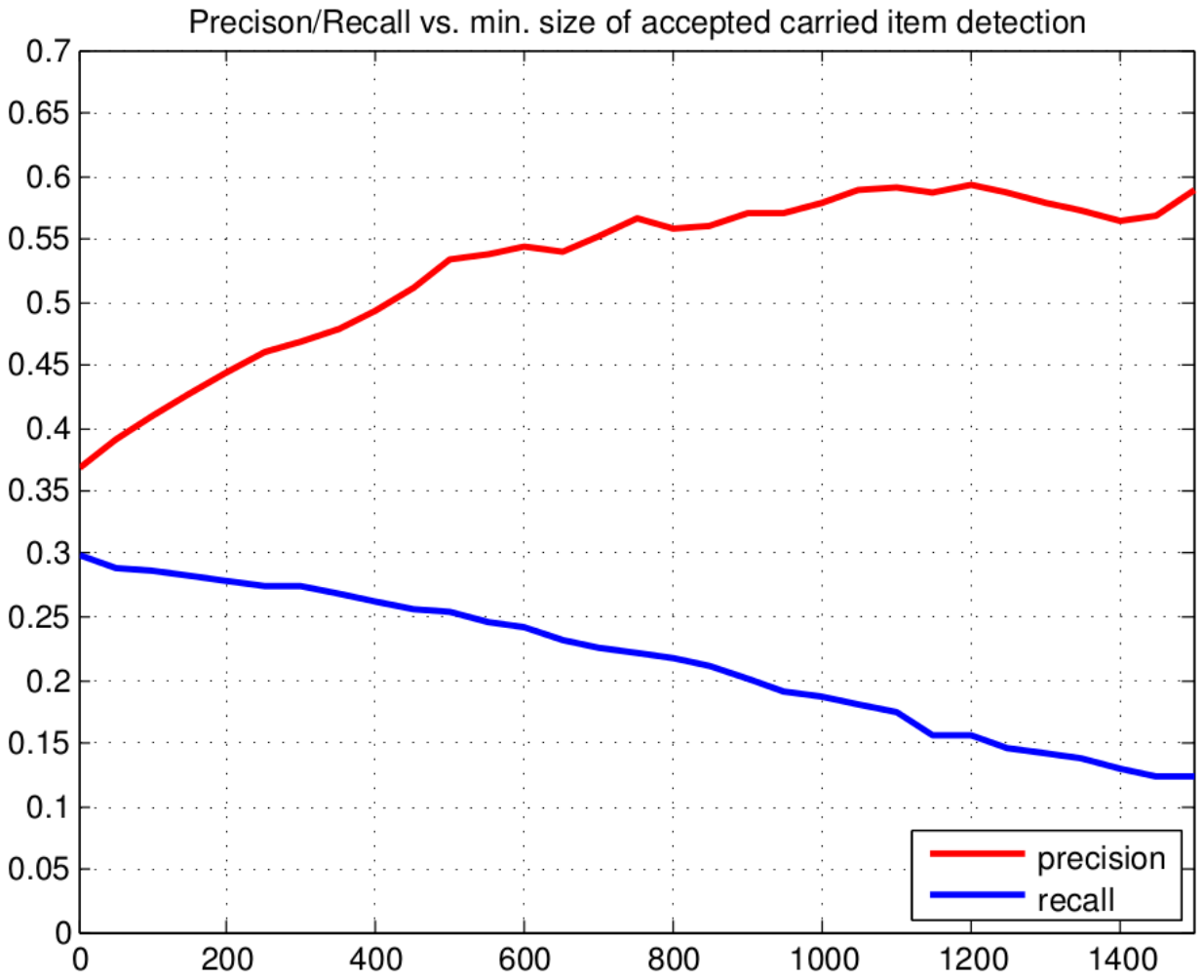




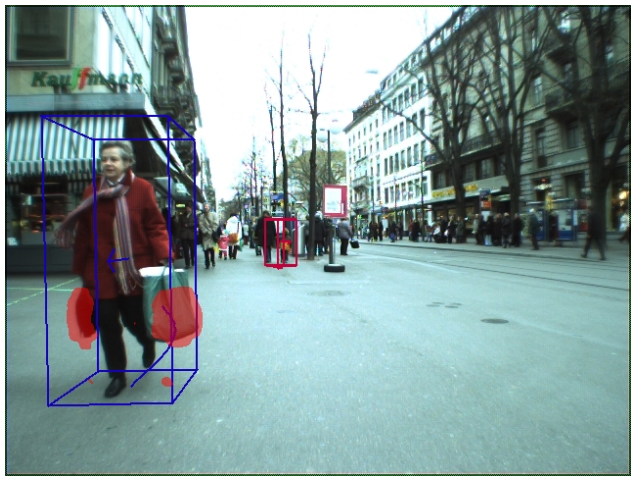


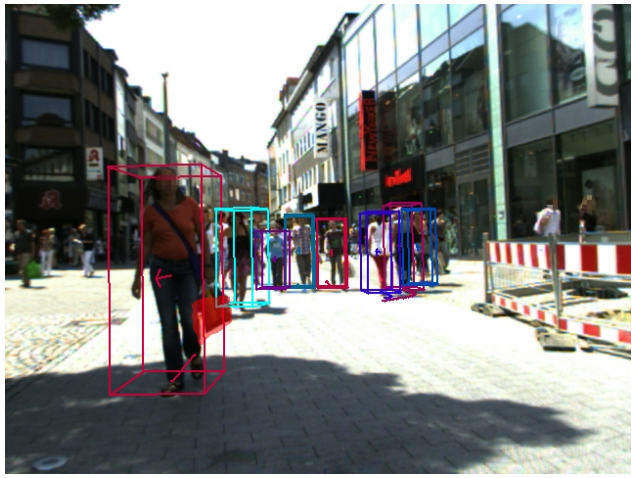
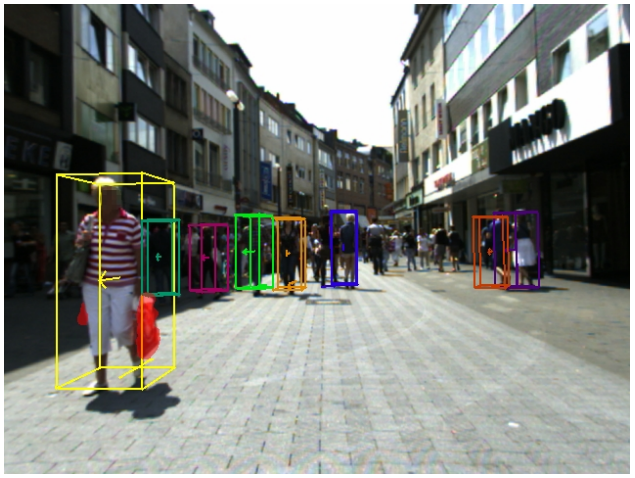
Carried Item Detection Performance

Comparison of segmentation results with the labeled data.



Carried Item Detection Examples



Conclusion

- The approach is based on *tracking-before-detection* paradigm as opposed to older *detection-before-tracking* one.
- This leads to the possibility to detect and track unknown objects.
- The presented 3D model allows not only achieving state-of-the-art tracking performance but also analyzing the shape of tracked person in more detail to detect carried items while running the system on-line.

Questions?