Institute of Measurement and Control Systems
Prof. Dr.-Ing. C. Stiller

KSOP
Karlsruhe School of Optics & Photonics

Karlsruhe Institute of Technology

Sparse Scene Flow Segmentation for Moving Object Detection in Urban Environments

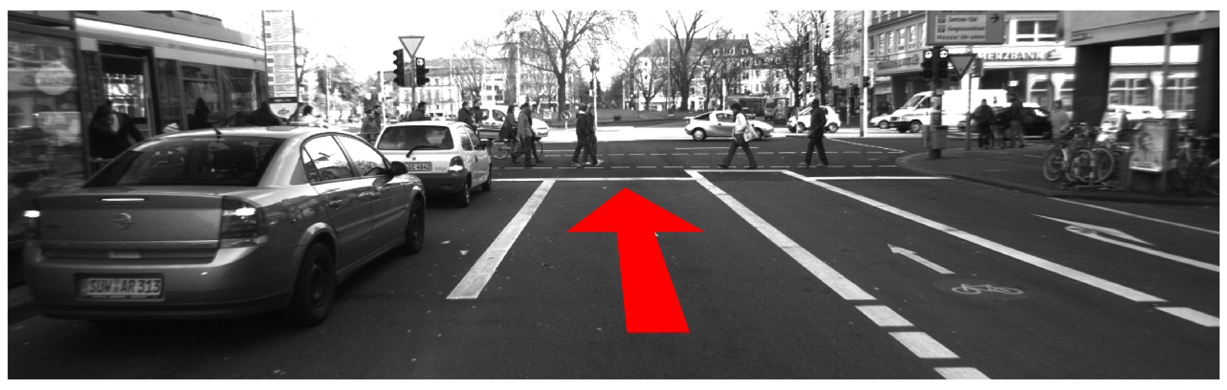
Philip Lenz, Julius Ziegler, Andreas Geiger, Martin Roser
Karlsruhe Institute of Technology (KIT)
Institute of Measurement and Control Systems
{lenz | ziegler | geiger | roser}@kit.edu

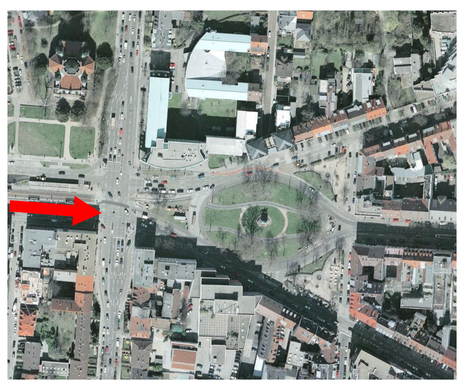
Motivation

Inner city intersections are a very demanding scenario for modern driver assistance systems.

Goal

- Perception and understanding of highly dynamic traffic scenes.
- Class-independent detection of moving object for inner city traffic scenarios.





Stereo Reconstruction

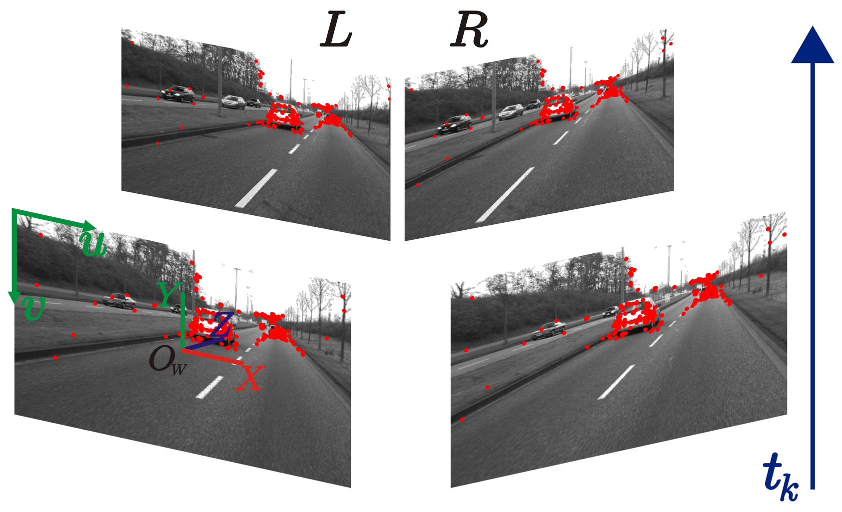
Interest points $\mathbf{x}=[u,v]^T$ are detected in two consecutive stereo image pairs.

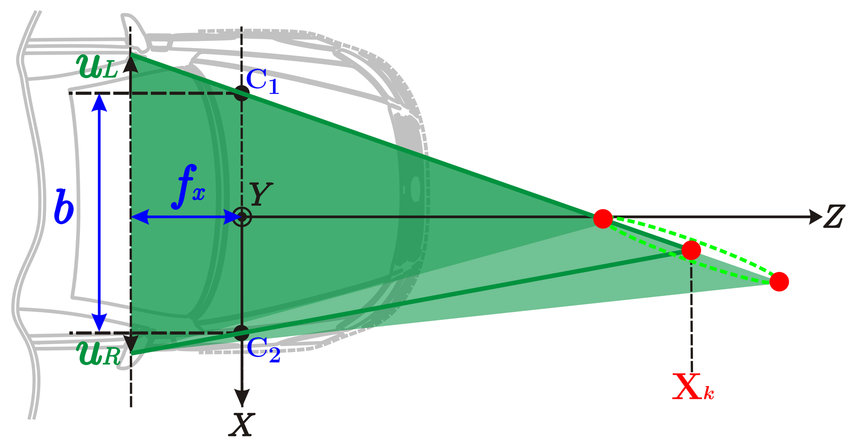
Disparities are estimated at sub-pixel accuracy for rectified images.

Error propagation leads to the quadratically growing error of the **3D position** $\mathbf{X}=[X,Y,Z]^T$ given by

$$X = \frac{(u_L - c_{u,L}) \cdot b}{d} \qquad Z = \frac{b \cdot f}{d}$$

$$Y = \frac{(v_L - c_{v,L}) \cdot b}{d}$$





System Overview

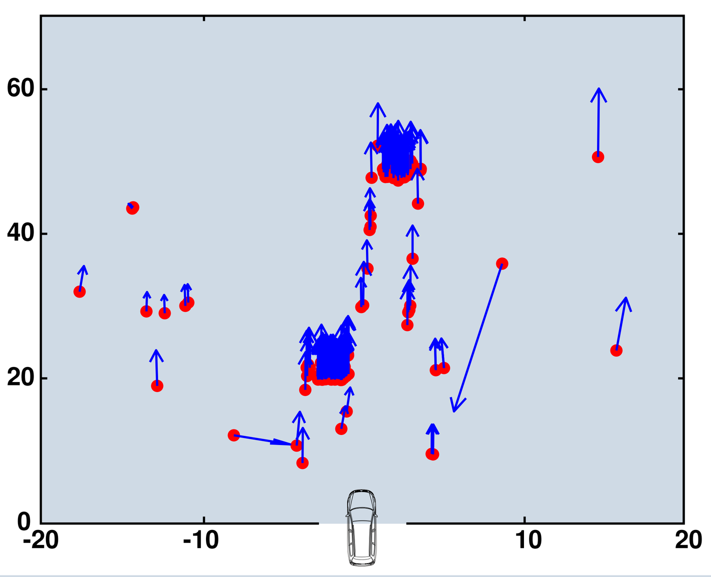
Scene Flow Description → Scene Flow Clustering → Geometry Check → Object Association → Final Detection

Scene Flow Description

Interest points are stored as tracklets.

The **velocity** $\mathbf{V}$ is computed as the first order derivative of the world points $\mathbf{X}$.

$$\mathbf{V} = \frac{\Delta \mathbf{X}_{k-\Delta t_i}}{\Delta t}$$



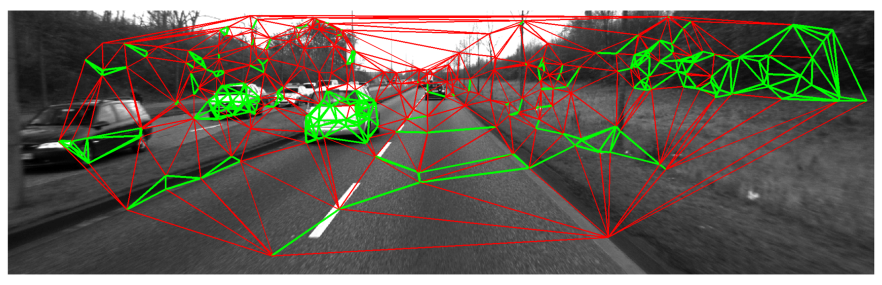
Scene Flow Clustering

Scene flow clusters describe a **similar motion**.

Removed edges of the graph exceed a threshold of the Mahalanobis distance Δ:

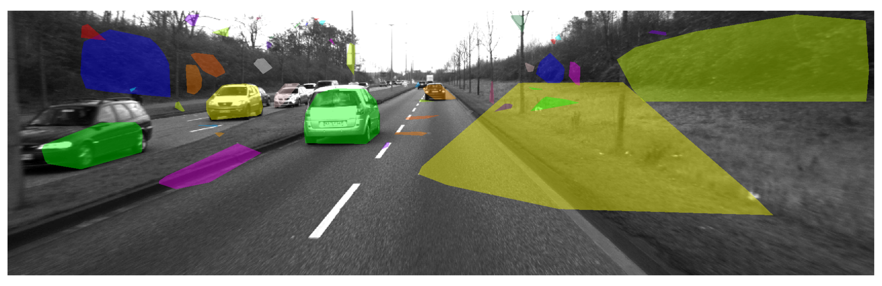
$$\Delta(\mathbf{V}_i, \mathbf{V}_j) = \sqrt{(\mathbf{V}_i - \mathbf{V}_j)\ \Sigma_{i,j}^{-1} (\mathbf{V}_i - \mathbf{V}_j)}.$$

The covariance Σ is obtained by error propagation of the 3D reconstruction.



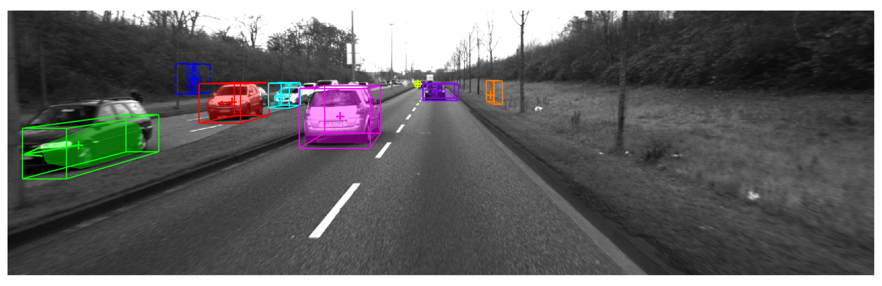
Geometry Check

Connected components describe detected objects.



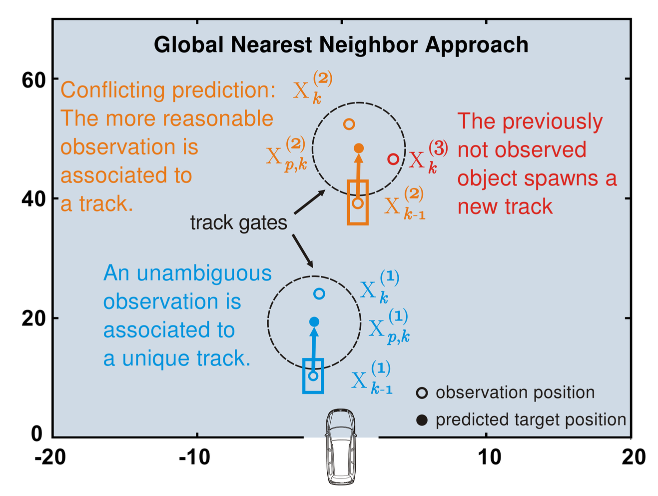
Neglected Objects:

- Exceeding reasonable dimensions.
- Not standing on the ground plane.

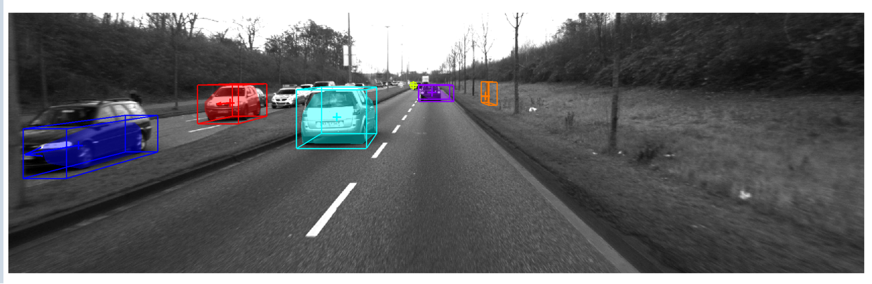


Object Association

Observation-to-Track association for unambiguous and conflicting prediction.



Final Detection



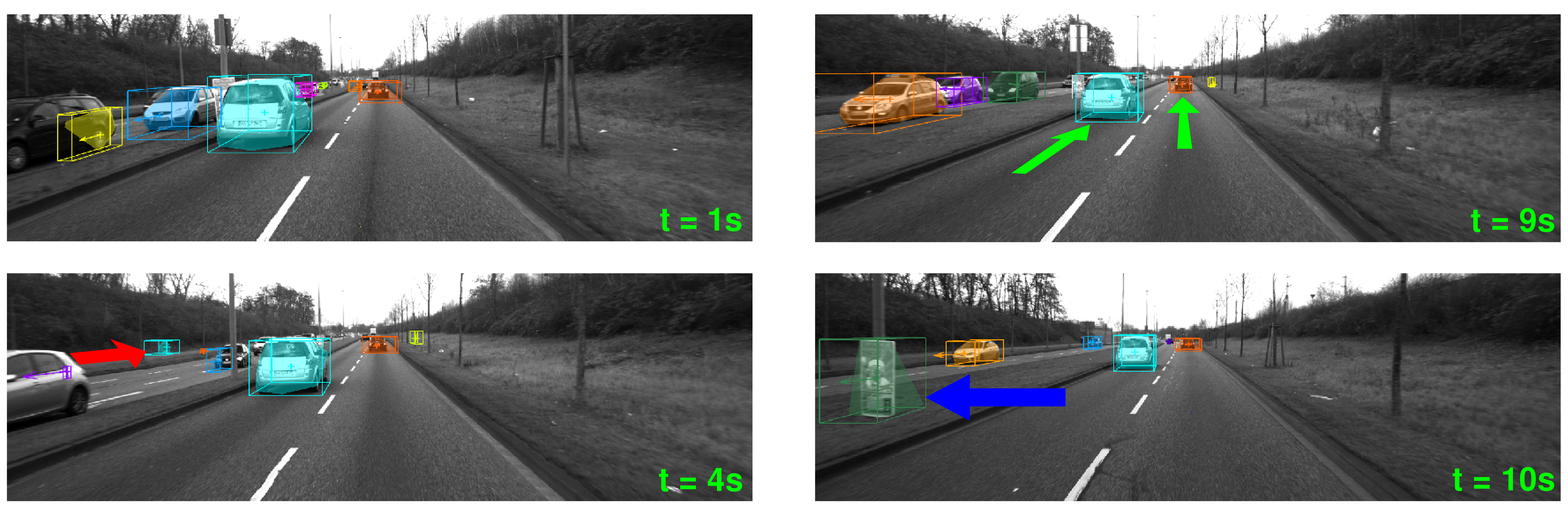
Results

Rural Road Sequences

- Continuous tracking of the **cars in front** within 15 s (10 fps).
- Detection of uncommen classes such as the **wheelchair user**.
- **Static objects** are detected since egomotion is not compensated.

Approaching cars are detected

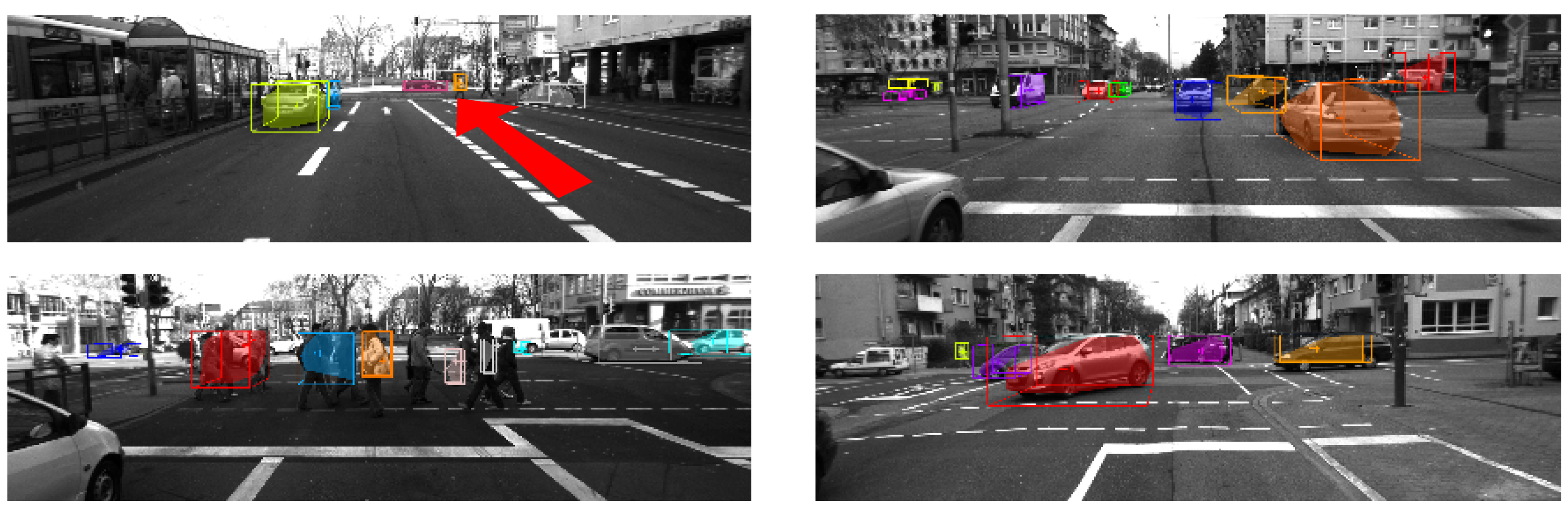
- within 3 time steps at a distance up to 25 m.
- within 5 time steps for a greater range of up to 60 m.



Inner City Intersection Scenarios

- **Pedestrians** are detected in a range up to 30 m.
- For the far range up to 60 m larger objects are detected as well.
- Turning or partly occluded cars are detected in a single frame.

- Similarly moving groups of pedestrians are detected as one object.
- Observation-to-Track association fails for sharply turning objects.



Conclusion

We presented a novel approach for object detection for challenging inner city traffic scenarios.

- Computationally sparse interest points.
- 3D description of moving objects in the current environment.
- Class-independent detection also of uncommon objects.

Future Work

Future work will include:

- More sophisticated multiple target tracking to reduce false detections and handle (partly) occluded objects.
- Egomotion compensation using visual odometry.
- Inclusion of a motion model to consider object rotations.

Karlsruhe Institute of Technology

The KSOP funding is provided by the German Research Foundation (DFG)