# Acoustic/Lidar Sensor Fusion for Car Tracking in City Traffic Scenarios


**Hamma Tadjine, Daniel Goehring**

Hamma Tadjine
08. September 2015

IAV GmbH,
Freie Universität Berlin

# Motivation

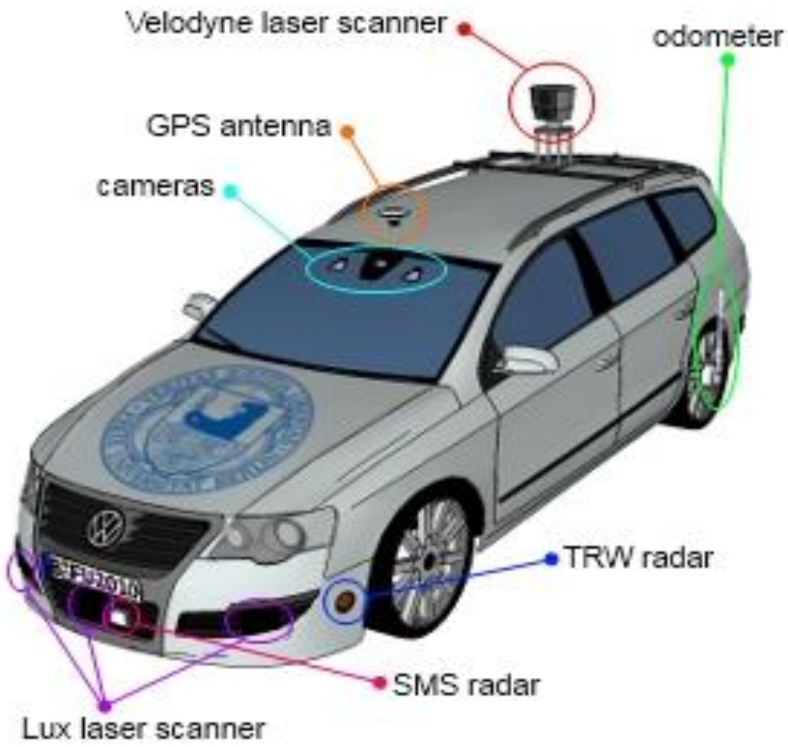
- Direction to Object-Detection: What is possible with cost-efficient microphone arrays, e.g. from Kinect?
- Fusion of multiple non-synchronized Kinect audio sensors and evaluation with data from Lidar sensors
- Application of the solution in real-world traffic scenarios

# Contribution

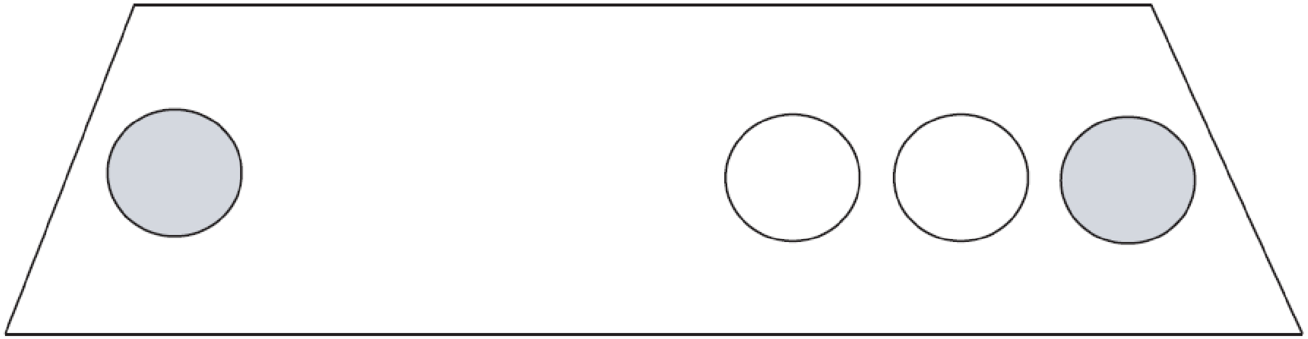
- Main components:
  - audio-based detection of objects for a single Kinect microphone array
  - creation of a representation for the belief distribution of object directions
  - combination of belief distributions of two Kinect microphone arrays
  - implementation on a real autonomous car using the OROCOS framework
  - synchronization and evaluation of the algorithm with Lidar point cloud from Ibeo Lux sensors

# Test platform

- Vehicle: VW Passat Variant, modified by VW
- Drive- and Steer-by-Wire, CAN
- Positioning system: Applanix POS LV 510
  - IMU, odometer, correction data via UMTS
- Camera systems:
  - 4 Wide angle cameras
  - 2 INKA Cameras (HellaAglaia)
  - 2 Guppy Cameras for traffic light detection
  - Continental Lane Detection
- Laser scanner:
  - IBEO Lux 6-Fusion System
  - 3D Laser scanner: Velodyne HDL 64 E
- Radar systems:
  - 2 short range (BSD 24 GHz)
  - 4 long range (ACC 77 GHz)
  - 1 SMS (24 GHz)



# Kinect sensor (Schematic)

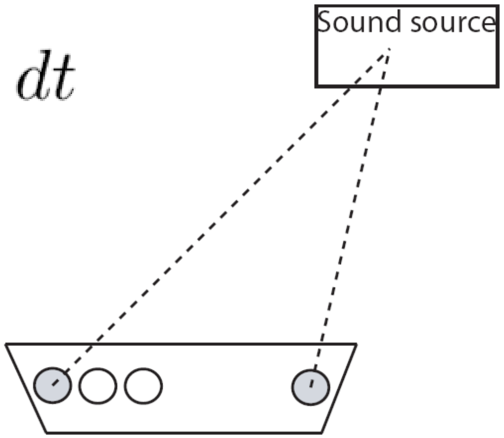


- 4 microphones, only the left and right outer microphones were used in our approach (gray circles)
- outer microphone distance: approx 22 cm

# Signal shift calculation via cross-correlation

- For continuous signals f and g holds:

$$(f \star g)(\tau) = \int_{-\infty}^{\infty} f(t) \cdot g(t+\tau) \ \ dt$$




- For discrete signals f and g holds:

$$(f \star g)[n] = \sum_{-\infty}^{\infty} f[m] \cdot g[m+n]$$

- We are interested in the delay ***n*** between the two discrete microphone signals:

$$n_{delay} = \max_{n} (f \star g)[n]$$

# Time delay between to Microphoness

- the two microphones provide audio signals with a sampling rate of 16.8 kHz

- the time difference for a signal approaching the two microphones is

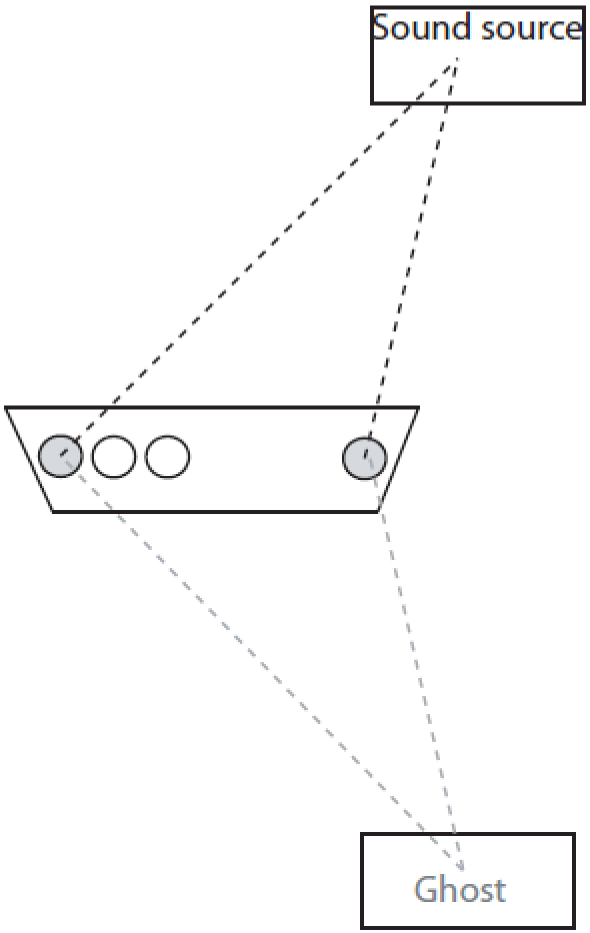
$$\Delta t = n_{delay} \cdot 16.8kHz$$

, which translates, given the speed of sound (340 meters per second), into a distance difference of:

$$\Delta s = n_{delay} \cdot 16.8kHz \cdot 340\frac{m}{s}$$

# Time delay between to Microphoness

- the two microphones have a distance of 0.22 meters (base distance)
- given the **base distance**, and the **signal shift**, for an **assumed distance** of the object (far away, e.g. 25 m) we have a **defined triangle**
- we can calculate the **angle** to the object w.r.t. symmetry
- on a plane, two solutions remain



# Distribution of possible angles to object

- sampling frequency is limited to 16.8 kHz

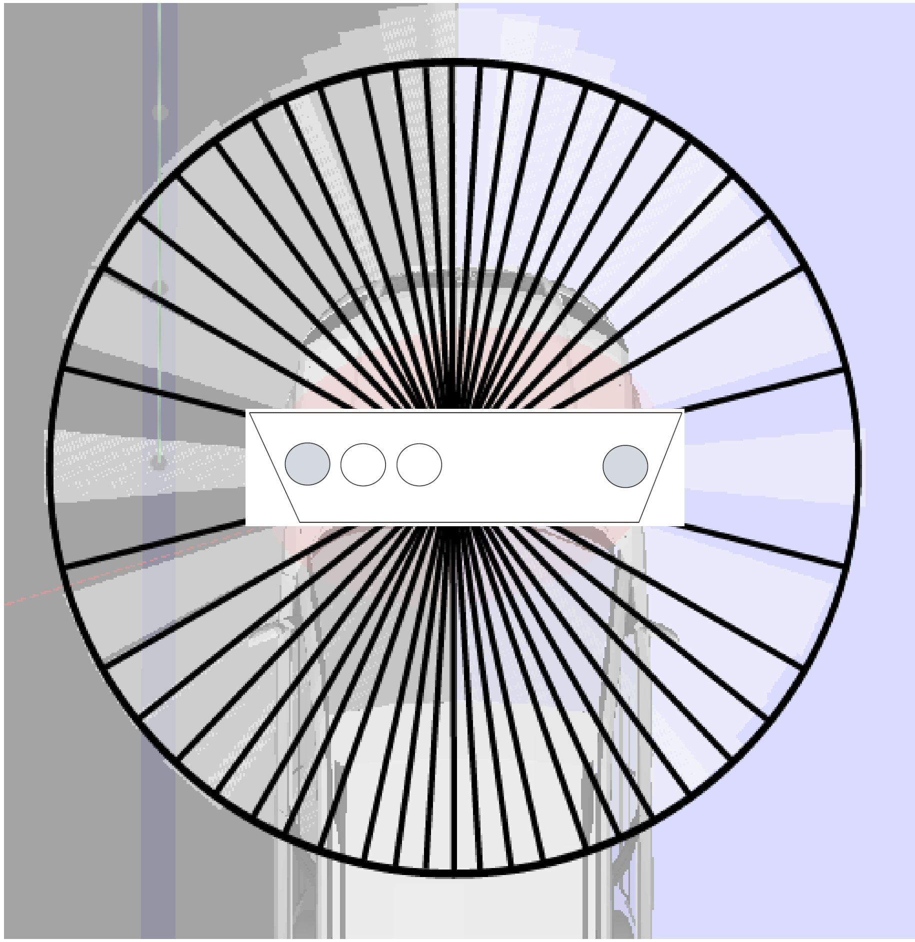
- for a given base distance of the two microphones (0.22 m) and a given signal shift, we can have only:

$$2 * 0.22 \text{ m} * 16.8 \text{ kHz} / 340 \text{ m} \approx 22$$

possible discrete outcomes for angular directions
→ approx. 46 different angular segments (with symmetry)

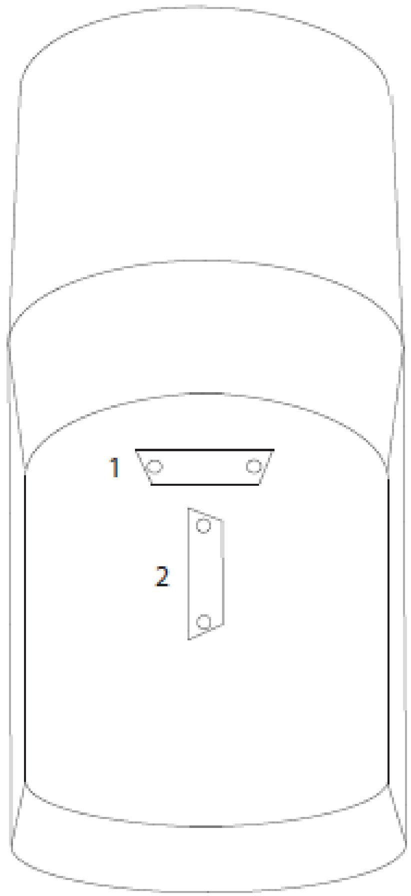
# Angular segment distribution

- different segments (46) cover different angular intervals
- each segment can be interpreted as a **belief cell** for an object in an angular direction interval
- radius for each segment will represent cross-correlation value (belief)



# Combination of Kinect sensors

- Symmetry disambiguation on a plane can be achieved with two Kinect (each 2 microphones),
- Both devices  are rotated by 90 degrees towards each other
- Kinect 1 can distinguish between left and right but not between front and rear direction
- Kinect 2 can distinguish between front and rear but not between left and right direction



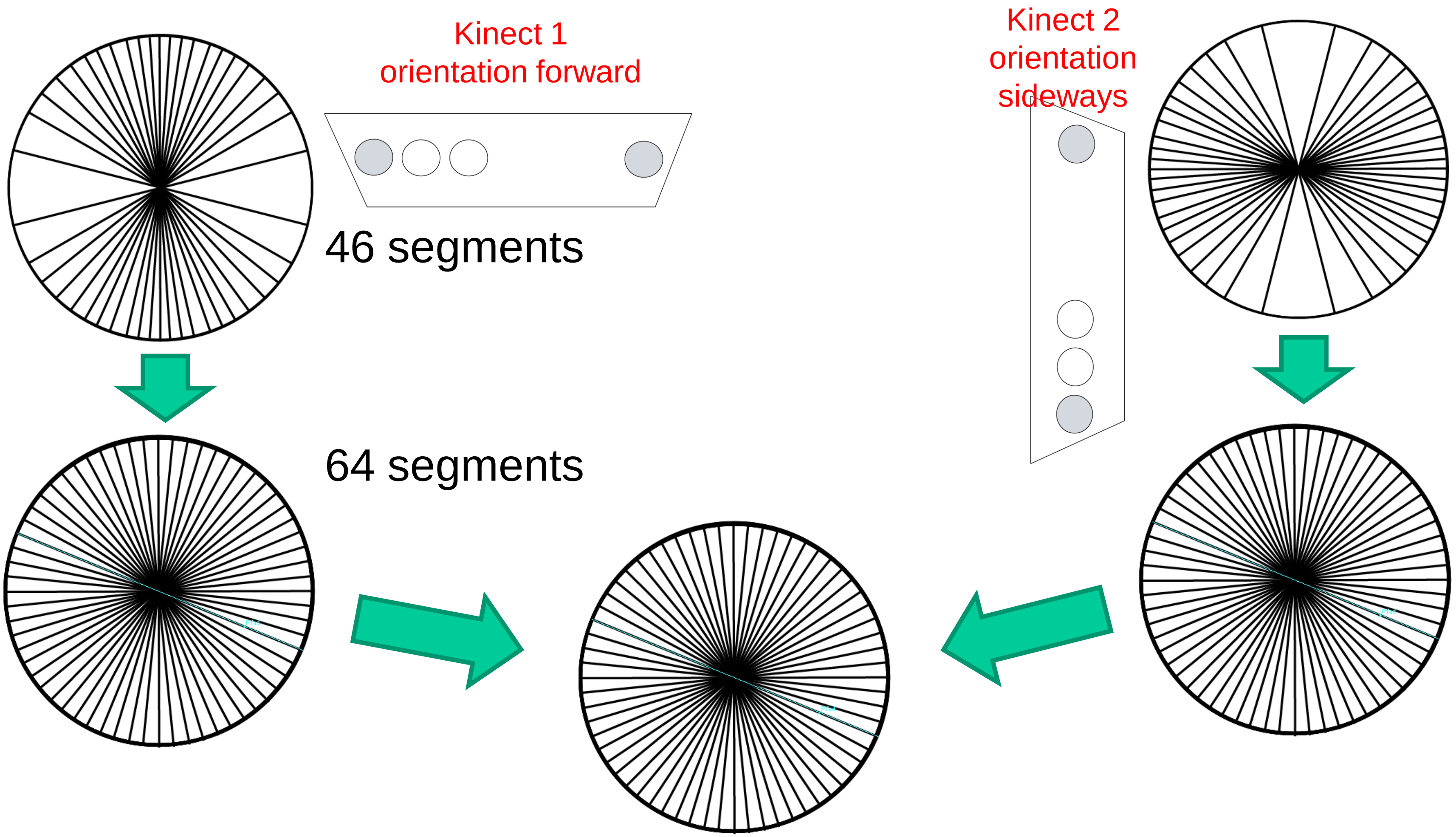
# Combination of Kinect sensors

- Symmetry disambiguation on a plane can be achieved with two Kinect microphone pairs, which are oriented by 90 degrees towards each other
- For fusion, we subsampled the two non-equally spaced histograms into two qually spaced histograms

- The value of each non-uniform belief cell is assigned to (split into) the uniform belief cells covered (fully or partially covered)
- Combination of both kinect belief distributions via

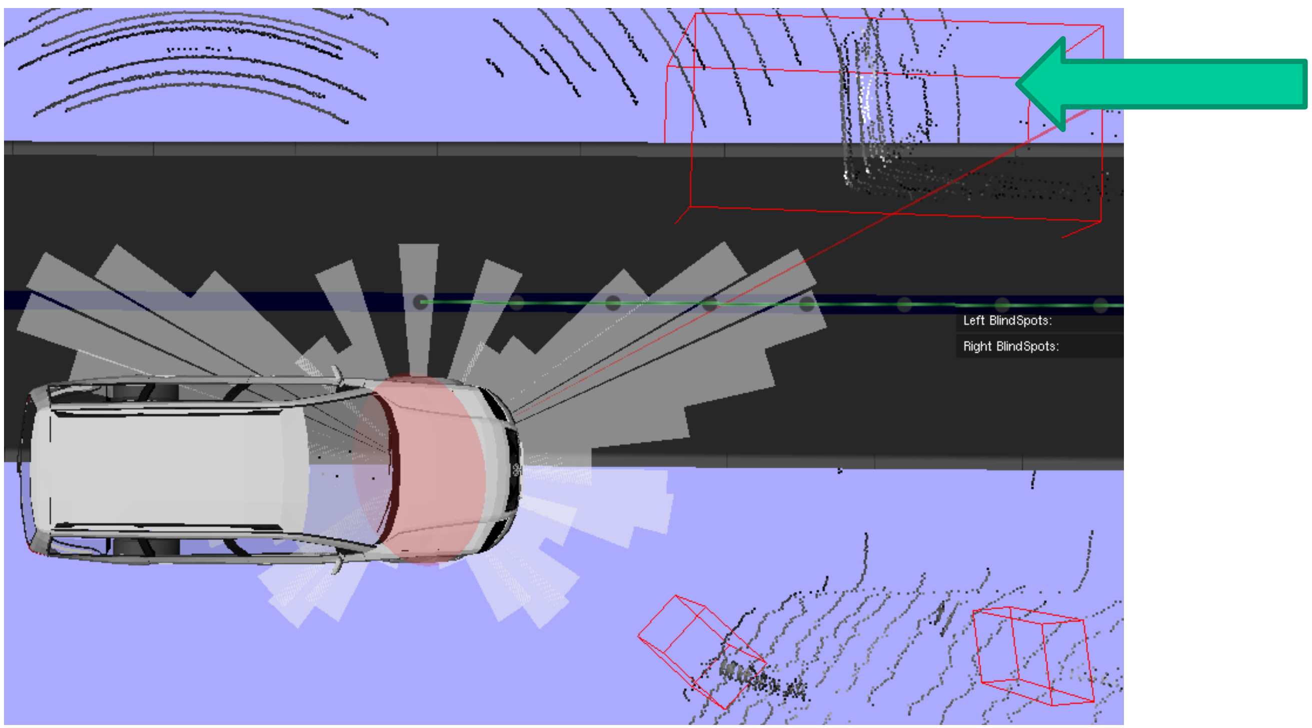
***cell-wise multiplication***

# Subsampling of belief cells and fusion for two Kinect sensors



# Traffic Example

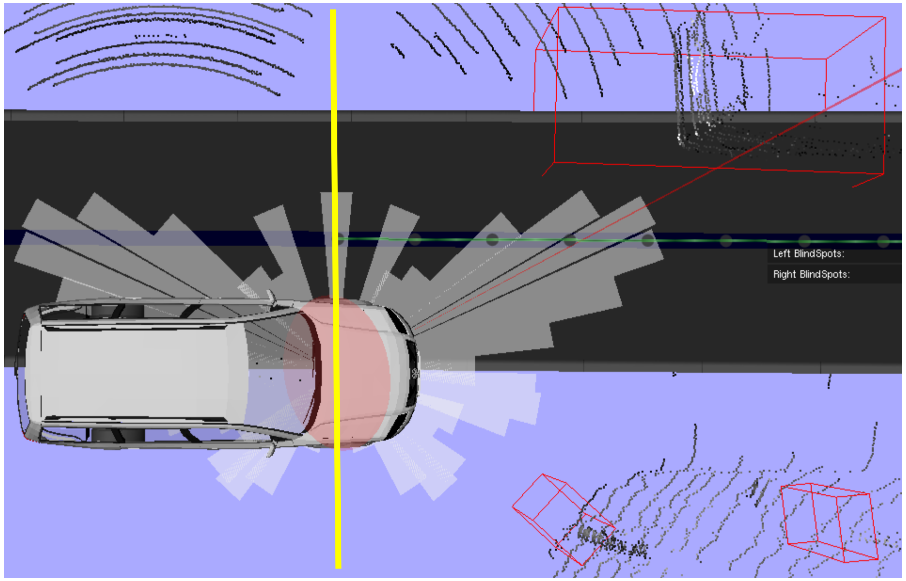
Kinect facing to the front, length of each (non-equal) angular segment represents angular belief




Passing car, Lidar data

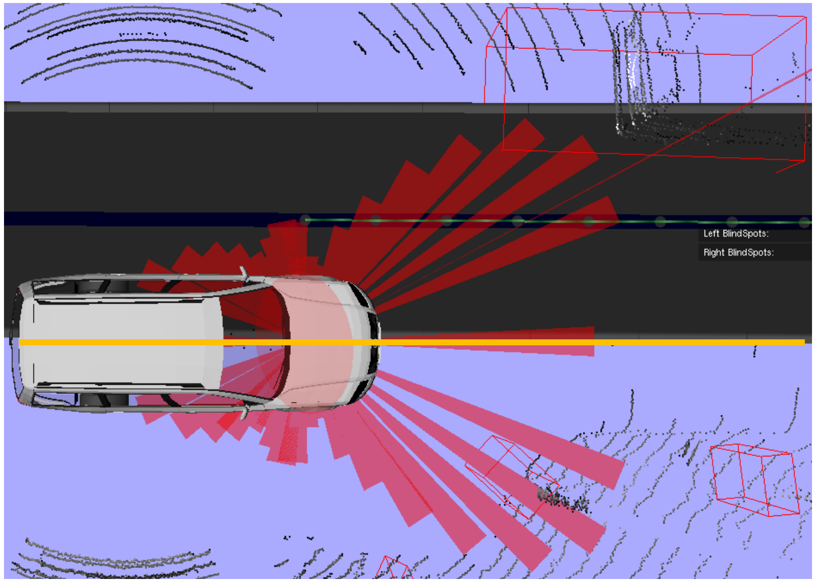
# Traffic Example, Step by Step

Kinect facing to the front



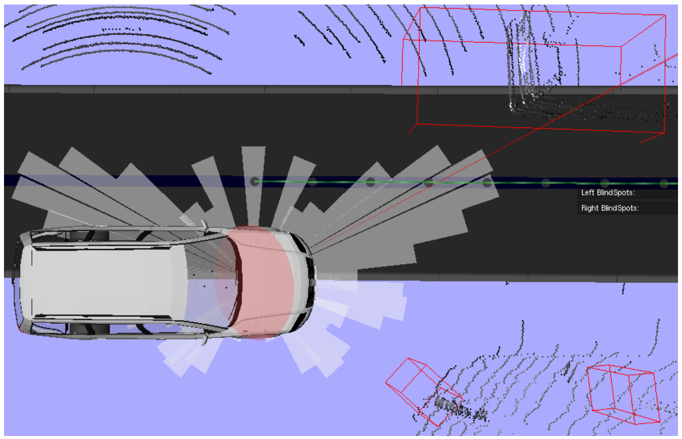

front/rear symmetry
(yellow axis)

Kinect facing to the side

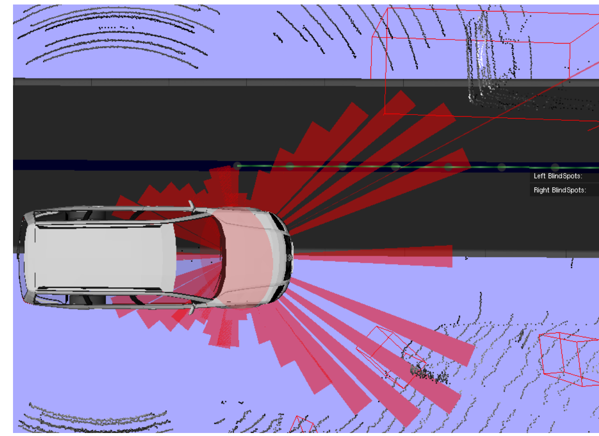


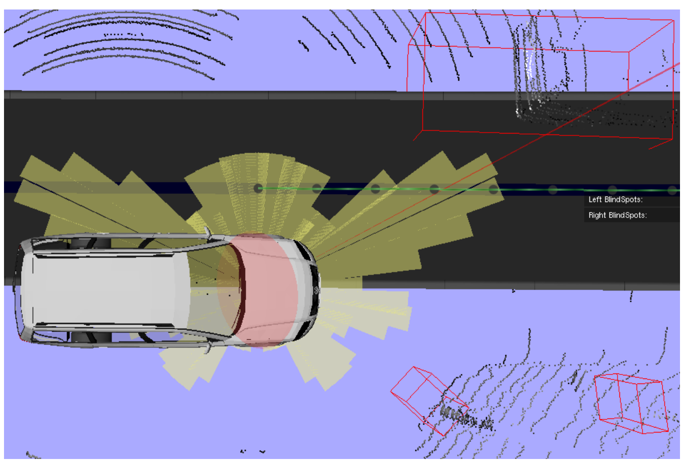

left/right symmetry,
(orange axis)

# Traffic Example, Step by Step

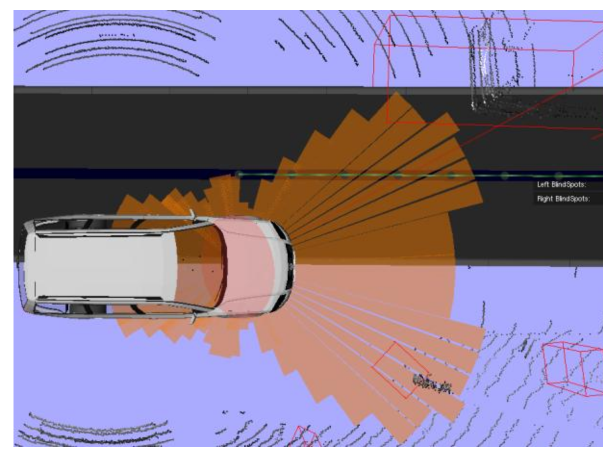



nonuniform

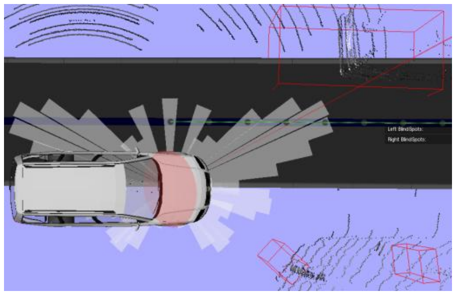


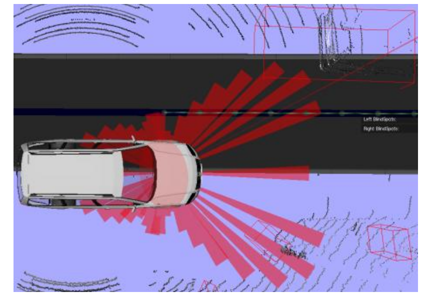

uniform segments

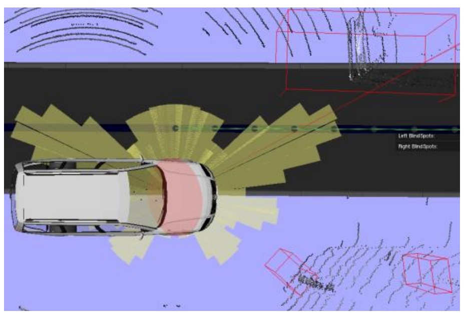


# Traffic Example, Step by Step

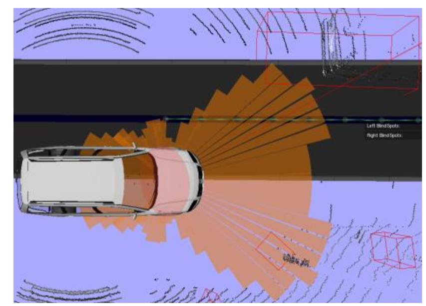


nonuniform

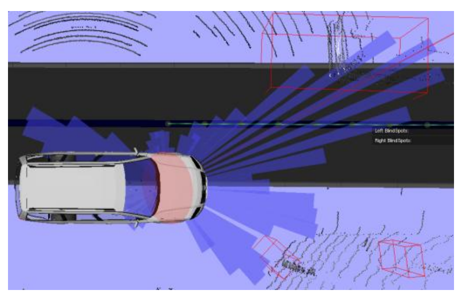




uniform
segments



data fusion,
no symmetries

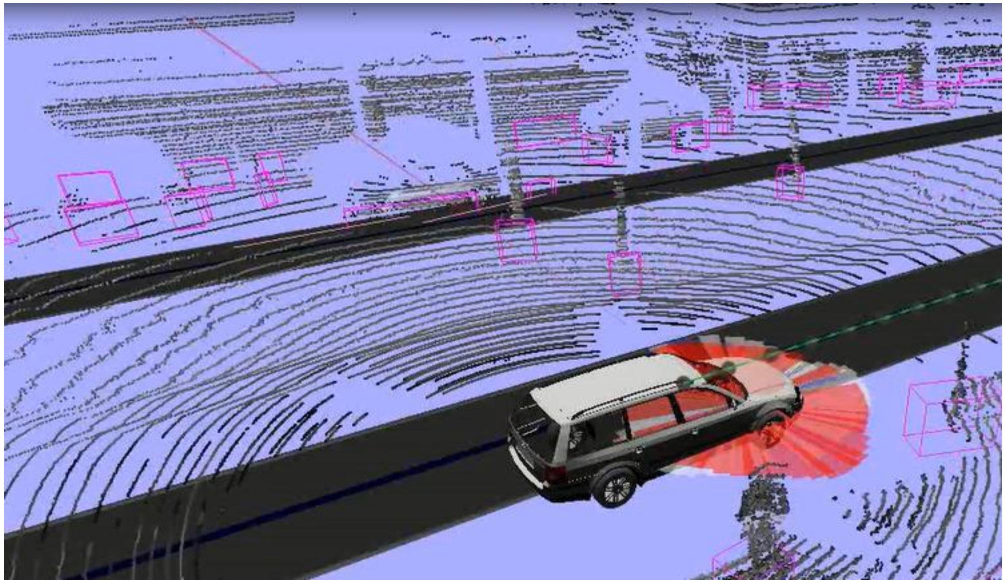


# Object Direction calculation

- What is the angle to the object?
  - After fusion, angular segment with the highest value wins (maximum likelihood)
  - Drawback: only one direction possible
  - For multiple objects it would be possible to search for **multiple** large angular segments (not too close to each other)

# Experimental setup

- Approach was tested in our autonomous car in a real traffic situation
- driving the car created much wind noise → car was parked on the side of the read, passing vehicles were detected



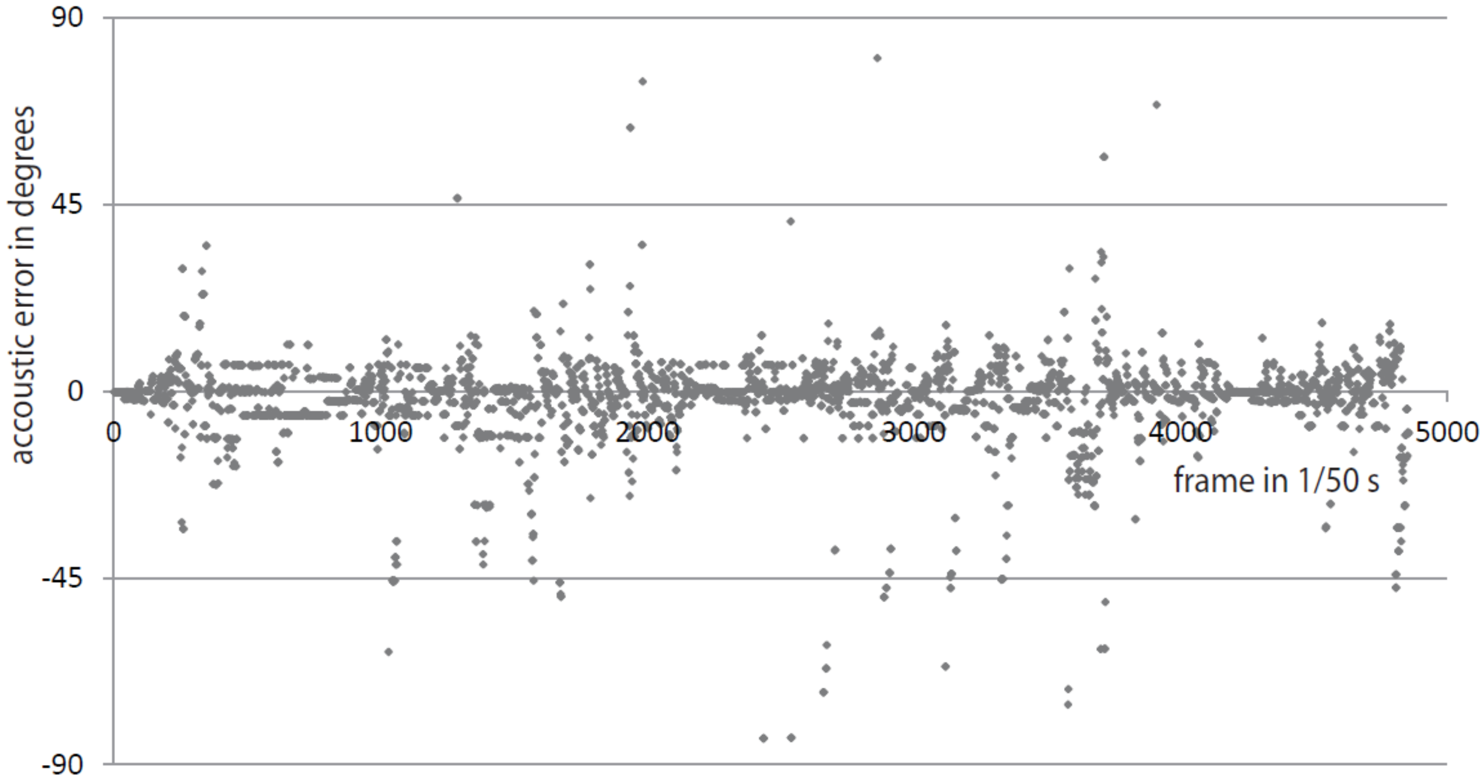
# Experimental evaluation

- Lidar scanner from Ibeo Lux (6 scanners) used to evaluate accuracy of sound source localization
- Idea: compare the angle calculated using audio data with the closest angle of moving obstacles from using Lidar
- Lidar objects were clustered and tracked from point cloud data

# Demo: Video 1 and 2

# Experimental results

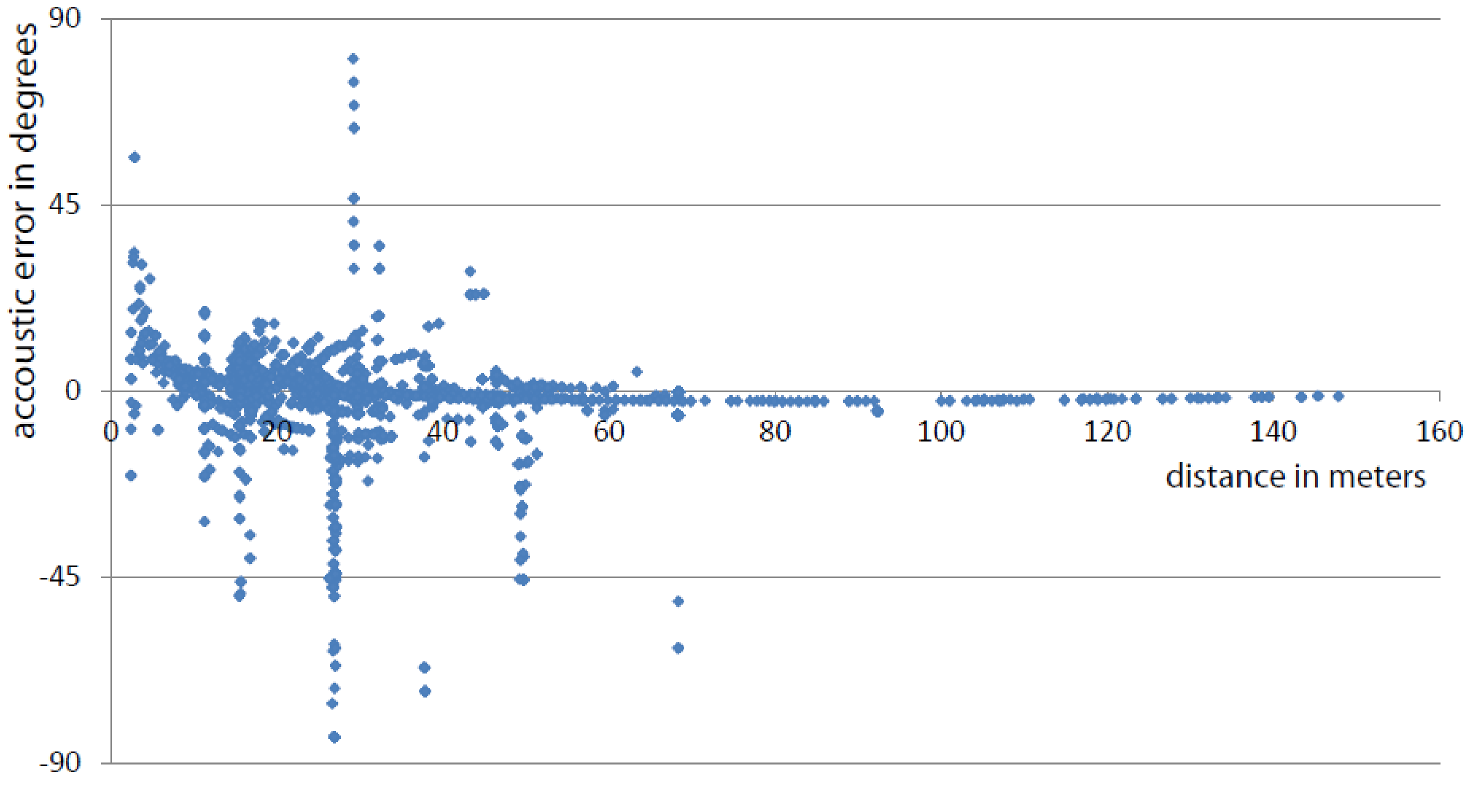
Angular error over time w.r.t. Lidar data




angular error standard deviation: 10.3 degrees

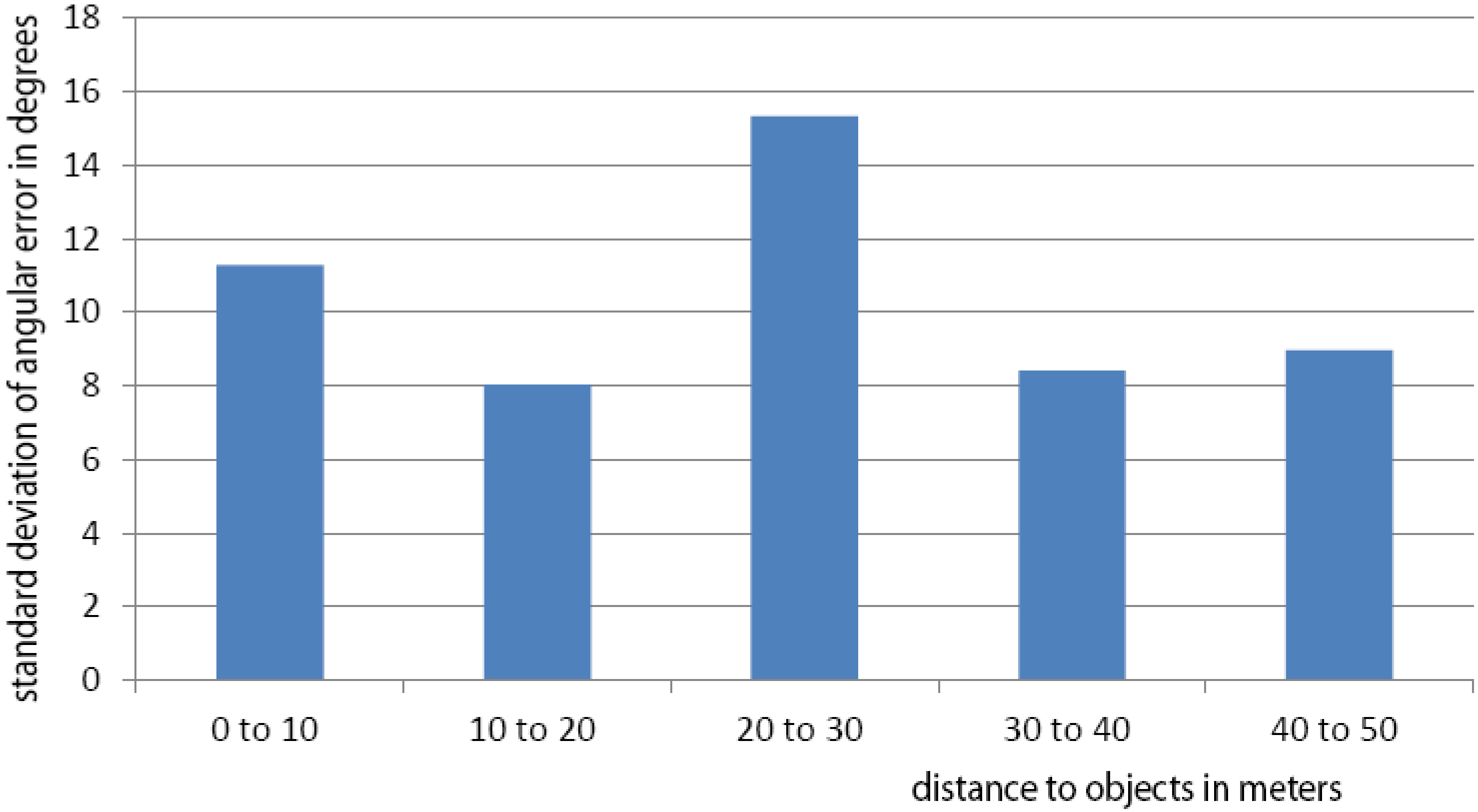
# Experimental results (contd.)

Angular error over distances w.r.t. Lidar data



# Experimental results (contd.)

## Angular error over distances



# Experimental results (contd.)

- for close objects it is usually hard to tell the exact angle, due to their size – therefore the error for more distant objects was often smaller than for close ones

- other inaccuracies were caused by sound reflections on houses and trees close to the street

- errors caused by limited sound velocities in combination with high velocities of cars could not be measured → city traffic 50 km/h

# Conclusion

- we presented an approach to calculating angles to objects using accoustic data from 2 Kinect microphone pairs
- showed how data from two non-synchronized devices can be combined using subsampled and uniform angular interval segments
- Detected angles were assigned and compared to real world Lidar data

- Approach was implemented on a real autonomous car robotics modular framework (OROCOS) and tested in a real world traffic situation

# Future Work

- Current challenges – wind noise while driving
- How to keep track of multiple sources
- How to handle sound reflections, e.g. from buildings, trees, etc.

# Future Work (contd.)

- How can the signal intensity used for distance estimation

- Band pass filters / FFT can help to select specific signals, e.g. emergency vehicles with certain signal horn frequencies and signal patterns – furthermore try to detect alternation between two frequencies

- Use tracking in combination with doppler effect to estimate velocities while vehicle is passing (change in frequency)