Towards Deep Understanding of the Vulnerable Road User

Fabian Flohr
Principal Engineer, Daimler AG, Ulm, Germany

Mercedes-Benz
The best or nothing.
Why do we need vehicle automation

Vehicle Automation

Motives
Different Reasons for Vehicle Automation

The Disruption Case

Growing on-demand mobility

Mercedes-Benz
The Disruption Case: Forms of On-Demand Mobility

**Car Sharing**
You are the driver

- Flexible, no driver, affordable
- Limited coverage area, parking issues

**Taxis / "Ubers"**
Chauffeured

- Versatile, convenient, anywhere
- Costly, needs driver
The Disruption Case  Automated On-Demand Mobility

<table>
<thead>
<tr>
<th>Car Sharing</th>
<th>Robocabs</th>
<th>&quot;Ubers&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>You are the driver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexible, no driver</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited coverage area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-driving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eventually the best of both worlds?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chauffeured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convenient, anywhere</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs driver</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mercedes-Benz
Different Reasons for Vehicle Automation

- **The Disruption Case**: Growing on-demand mobility
- **The Mobility Case**: Mobility for those who cannot drive
- **The Comfort Case**: Automation to meet customer demands
- **The Safety Case**: Automation to improve traffic safety
The Safety Case

1.24 million road traffic deaths occur every year worldwide.

Every 2 hours a pedestrian is killed.

Every 8 minutes a pedestrian is injured.

#1 cause of death among those aged 15-29 years.

Global status report on road safety, WHO 2013

Mercedes-Benz
How an automated vehicle works:

Vehicle Automation Technology
Vehicle Automation Components

**Sensors**
What an autonomous vehicle does not "see", it is likely to drive into

**Backend**
Data provisioning and collection as well as operator commands

**Autonomous Vehicle "Brain"**
Deciding which course of action is right depending on the situation

**Redundant Actuators**
Anything a driver would do needs to be controlled electronically

Mercedes-Benz
**Processing Sensor Data**

**Dense Spatial Data**
- Completeness of data under all conditions

**Classification**
- Deep learning for semantic scene labeling
Deep CNNs Semantic Stixel World

Efficient Scene Labeling using Convolutional Neural Networks

Semantic Labeling: 2 Mio labeled points

Stereo Matching: 2 Mio 3D-points

3D Stixel Representation: 1.000 Stixel

Semantic Stixel: 1000 Stixel

Mercedes-Benz
Deep CNNs: Region-based Inference

Instance-based Object Detection using Convolutional Neural Networks

Mercedes-Benz
Feeding the Brain More Data

Complexity
- 25 classes
- 2Mpixel stereo image pairs

Variability
- 50 German cities

Number of annotated images
- 5’000 fine
- 20’000 coarse

Metadata
- Stereo
- Video
- GPS
- Odometry

Benchmark
- Scene labeling
- Object detection

www.cityscapes-dataset.net

Feeding the Brain More Data
Tsinghua-Daimler Cyclist Detection Benchmark

<table>
<thead>
<tr>
<th>Dataset Statistics</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
<th>Non-VRU</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># image frames</td>
<td>9741</td>
<td>1019</td>
<td>2914</td>
<td>1000</td>
<td>14674</td>
</tr>
<tr>
<td># cyclist BBs</td>
<td>16202</td>
<td>1314</td>
<td>4657</td>
<td>0</td>
<td>22173</td>
</tr>
<tr>
<td># pedestrian BBs</td>
<td>0 (ignored)</td>
<td>1541</td>
<td>7401</td>
<td>0</td>
<td>8942</td>
</tr>
<tr>
<td># other rider BBs</td>
<td>0 (ignored)</td>
<td>190</td>
<td>1105</td>
<td>0</td>
<td>1295</td>
</tr>
<tr>
<td># total BBs</td>
<td>16202</td>
<td>3045</td>
<td>13163</td>
<td>0</td>
<td>32410</td>
</tr>
</tbody>
</table>

X. Li, F. Flohr, Y. Yang, H. Xiong, S. Pan, K. Li and D.M. Gavrila, “A New Benchmark for Vison-Based Cyclist Detection”, IV 2016.
Where vehicle automation gets complicated:

Understanding the VRU
Processing Sensor Data

Dense Spatial Data
Completeness of data under all conditions

Classification
Deep learning for semantic scene labeling

Motion/Intention Analysis
Movement calculation and intention estimation
**Intent Recognition**  Will the Pedestrian(s) Cross?

Difficult? Indeed, pedestrian crossing might be only detected when already underway in this way.

N. Schneider and D.M. Gavrila. Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study. *Proc. German Conf. on Patt. Recognition 2013*
Point kinematics can be augmented with the image motion of the detected object.

Augmented visual features allow machine learning algorithms to predict pedestrian crossing ~200 ms earlier.


Mercedes-Benz
Intent Recognition: Will the Pedestrian(s) Cross?

Gaussian Process Dynamical Systems

Offline
Learning of a low-dimensional manifold (latent space) to represent optical flow pedestrian images (measurement space), incl. dynamical model

Online
Particle filtering in latent space. Likelihood by mapping particles onto measurement space and comparing with measured dense optical flow

Augmented visual features allow machine learning algorithms to predict pedestrian crossing ~200 ms earlier.

Intent Recognition Will the Pedestrian(s) Cross?

A human driver relies heavily on context cues to anticipate how a traffic situation will evolve.

Various context cues influence the pedestrian motion


Mercedes-Benz
Intent Recognition Will the Pedestrian(s) Cross?
Dynamic Bayesian Networks with Context Cues

Various context cues influence the pedestrian motion

Mercedes-Benz
Intent Recognition Pedestrian Torso & Head Orientation
CNNs for Instance Detection and Orientation Regression

Robust object orientation is estimated together with bounding box detection and regression.

Joint detection and pose estimation
Mercedes-Benz
Head and torso estimation takes into account physical constraints.

For clarity, results are only shown for one pedestrian at a time.

Robust head and torso estimation used as context information

Mercedes-Benz
Intent Recognition Will the Pedestrian(s) Cross?
Dynamic Bayesian Networks with Context Cues

A human driver relies heavily on context cues to anticipate how a traffic situation will evolve.

Various context cues influence the pedestrian motion

Mercedes-Benz
Real-time Multi-Person 2D Pose Estimation Using Part Affinity Fields

Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh
Carnegie Mellon University
Intent Recognition Do We Need to Know More?
Human Pose Estimation for Gesture Recognition and Intention Prediction

- **Appearance**
  - Riders / pedestrians have to be distinguished
  - Classification of special VRU types e.g. police/fire/military uniforms, safety clothing e.g. vest (road worker, accident), school guards

- **Implicit**
  - VRU intention features which are extracted based on human body language;
    - e.g. direction of motion, body / head orientation, gate cycle

- **Explicit / Defined context, specific**
  - Gestures which are defined by alphabets and have a specific meaning and defined context
    - e.g. Cyclist alphabet, Police signal alphabet, Road workers (with/without sign), School guard gestures

- **Explicit / Undefined context, fuzzy**
  - Gestures which can be performed by all VRU types and are often fuzzy and in a not completely observable context
    - e.g. please pass by, please stop, call a taxi
Thank You Very Much For Your Attention!