Towards Fully-automated Driving

Challenges and Potential Solutions

Dr. Gijs Dubbelman
Mobile Perception Systems
EE-SPS/VCA

tue-mps.org
Mobile Perception Systems

tue-mps.org

- 6 PhDs, 1 postdoc, 1 project manager, 2 software engineers
Automotive Projects
Highly Automated Driving (HAD)

- SAE level 4 and above
  - The driver is no longer the safety fallback
  - Including dynamic urban environments
  - Including adverse weather conditions
State-of-the-Art

Autoware

Google

UBER

Ford
State-of-the-Art Uber
How Does It Work
State-of-the-Art

• **Digital rail road**: nominal vehicle trajectory is predefined in the map.

• **Absolute positioning**: RTK-GPS aided by FOG IMU and 3D point-cloud localization

• **Map as a sensor**: road and static object geometry are contained in the map

• **Dynamic objects**: detect and track from vision, RaDAR, and LiDAR
Challenges

- **Scalability:** creating HAD maps for absolute positioning is extremely labor intensive.
- **Timeliness:** frequent updates are required to keep map fresh.
- **Costly hardware:** RTK-GPS, FOGs, and dedicated LiDARS.
- **Limited robustness:** adverse weather and lighting conditions and highly-dynamic traffic scenes
Challenges

- SAE Level 4+ & Scalable

- SAE Level 2
  - Personally owned
  - Scalable

- SAE Level 4
  - Robot Taxi
  - Not scalable
Our Research

1. Research scalable HAD map technologies
   • From absolute to relative positioning
   • Distributed Semantic-SLAM techniques to automatically create and update maps
   • Crowdsource visual observations

2. Research AI to make vehicle less dependent on HAD map
   • Deep learning and real-time inference
   • From pattern recognition to reasoning
Our Research

Research scalable HAD map technologies

- Stereo Visual Odometry
- Pose-chain based SLAM
- Pose-graph based GNSS-odometry fusion
- Distributed SLAM using crowdsourcing
Stereo Visual Odometry
Pose-chain SLAM

- GPS (27 km)
- VO (19268 poses, 20 closures, RMSE: 504 m)
- COP-SLAM (68 millisecond, RMSE: 49 m)
- G²O (7100 millisecond, RMSE: 46 m)
Pose-chain SLAM

[Graphs and charts showing comparison between different SLAM methods]
Pose-chain GNNS-Odometry Fusion
## Pose-chain GNNS-Odometry Fusion

![Map with pose-chain GNNS-Odometry Fusion results](image)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Max.</th>
<th>Acc.</th>
<th>Prec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>66.291</td>
<td>0.489</td>
<td>2.653</td>
</tr>
<tr>
<td>2</td>
<td>21.799</td>
<td>0.874</td>
<td>2.491</td>
</tr>
<tr>
<td>3</td>
<td>49.068</td>
<td>0.715</td>
<td>2.293</td>
</tr>
<tr>
<td>4</td>
<td>4.828</td>
<td>1.089</td>
<td>1.025</td>
</tr>
<tr>
<td>5</td>
<td>16.776</td>
<td>0.678</td>
<td>1.141</td>
</tr>
<tr>
<td>6</td>
<td>3.672</td>
<td>0.336</td>
<td>0.954</td>
</tr>
<tr>
<td>7</td>
<td>7.775</td>
<td>0.782</td>
<td>1.177</td>
</tr>
<tr>
<td>8</td>
<td>18.037</td>
<td>0.805</td>
<td>1.264</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>23.531</strong></td>
<td><strong>0.721</strong></td>
<td><strong>1.625</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Max.</th>
<th>Acc.</th>
<th>Prec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.282</td>
<td>0.462</td>
<td>0.956</td>
</tr>
<tr>
<td>2</td>
<td>14.173</td>
<td>0.788</td>
<td>2.339</td>
</tr>
<tr>
<td>3</td>
<td>14.015</td>
<td>0.727</td>
<td>1.679</td>
</tr>
<tr>
<td>4</td>
<td>4.846</td>
<td>1.092</td>
<td>1.133</td>
</tr>
<tr>
<td>5</td>
<td>4.266</td>
<td>0.653</td>
<td>1.010</td>
</tr>
<tr>
<td>6</td>
<td>3.117</td>
<td>0.320</td>
<td>0.928</td>
</tr>
<tr>
<td>7</td>
<td>5.297</td>
<td>0.798</td>
<td>1.295</td>
</tr>
<tr>
<td>8</td>
<td>5.367</td>
<td>0.796</td>
<td>1.346</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>7.170</strong></td>
<td><strong>0.705</strong></td>
<td><strong>1.336</strong></td>
</tr>
</tbody>
</table>

**Improvement w.r.t. GNSS (%)**

| Improvement w.r.t. GNSS (%) | 69.528 | 2.282 | 17.794 |
Distributed SLAM using Crowdsourcing

Pose-chain data
- Chain of 6D poses
- Traffic sign loc./types
- Lane boundary loc./types

SLAM back-end
- Graph optimization
  1. COP-SLAM
  2. G2O

Multi-vehicle map

Pose-chain data

SLAM front-end
- Visual odometry
- Vehicle data (GPS-IMU)
- Traffic sign / lane detector
- Key-frame descriptors
  - Stored locally
  - Send on request

3G

SLAM front-end
- Visual odometry
- Vehicle data (GPS-IMU)
- Traffic sign / lane detector
- Key-frame descriptors
  - Stored locally
  - Send on request

3G

SLAM front-end
- Visual odometry
- Vehicle data (GPS-IMU)
- Traffic sign / lane detector
- Key-frame descriptors
  - Stored locally
  - Send on request

3G
Our Research

Research AI to make vehicle less dependent on HAD map

- Probabilistic Computer Vision
  - Ego-lane boundary detection
  - Stereo obstacle detection

- Semantic Scene Understanding
  - Hierarchical F-CNNs
  - End-to-end learning of semantic occupancy grids
Ego-lane Boundary Detection

DETECTING EGO LANE 703
VP -5,70 L -370,240 S 4,6
Stereo Obstacle Detection
Stereo Obstacle Detection
Semantic Scene Understanding

• **Aim:** reuse computation for multiple classifiers

• **Our task:** Semantic Scene Segmentation

• **Goal:** Extend the number of classes, without an extra labeling effort and by maximally reusing feature computation layers
Approach: Hierarchical Network

Datasets tree:
- Primary dataset: high-level classes
  - traffic sign, road, vehicle, ...
- Auxiliary dataset 1: sign subclasses
  - keep right, yield, roundabout, ...
- Auxiliary dataset 2: road subclasses
  - ego lane, lane markings, ...
- Auxiliary dataset 3: vehicle subclasses
  - bus, car, truck, ...
- Auxiliary dataset 4: car subclasses
  - per image annotations:
    - car models, brands, ...

Classifiers tree:
- Primary classifier
- Auxiliary classifier 1
- Auxiliary classifier 2
- Auxiliary classifier 3
- Auxiliary classifier 4

Shared Fully Convolutional Feature Extractor

classifier block = adaptation layers + softmax classifier
Recent results

Training on 3-level hierarchy with 3 datasets → inference results for 108 classes

Cityscapes

L2 results

Mapillary Vistas

L3 results
Scene Parsing of 108 classes in real-time
Quantitative Comparison

- Goal: Train traffic sign sub-classes on GTSDB and test on Cityscapes
- Compare a flat classifier with our hierarchical classifiers

<table>
<thead>
<tr>
<th>Classifier type</th>
<th>Flat</th>
<th>Flat</th>
<th>Hierarchical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained on</td>
<td>$\text{EV}_s$</td>
<td>$\text{CV}_c$</td>
<td>$\text{GV}_c$</td>
</tr>
<tr>
<td>Tested on</td>
<td>$\text{CV}_c$</td>
<td>$\text{GV}_c$</td>
<td>$\text{GV}_s$</td>
</tr>
<tr>
<td>Mean Accuracy (%)</td>
<td>69.4</td>
<td>59.2</td>
<td>7.8</td>
</tr>
<tr>
<td>Mean IoU (%)</td>
<td>60.4</td>
<td>-</td>
<td>5.4</td>
</tr>
</tbody>
</table>

The hierarchical classifier approach does better than a flat classifier approach, even when the flat classifier is trained on the target dataset

- Nvidia TitanX GPU
- Tensorflow
- 52 ms @ 512 x 706
- ~20 fps
End-to-end Occupancy Grids

![Diagram of End-to-end Occupancy Grids]

**TABLE I**

<table>
<thead>
<tr>
<th>Method</th>
<th>mean IoU</th>
<th>f.w. IoU</th>
<th>disparity</th>
<th>camera calibration</th>
<th>RGB image</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Disp.</td>
<td>80.8</td>
<td>93.2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fixed-plane Assum.</td>
<td>49.2</td>
<td>63.4</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ours</td>
<td>59.0</td>
<td>76.3</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
</tbody>
</table>

**TABLE II**

**ROBUSTNESS EVALUATION W.R.T. VEHICLE LOCAL DYNAMICS.**

<table>
<thead>
<tr>
<th>Perturbation</th>
<th>mean IoU</th>
<th>f.w. IoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>No perturbation</td>
<td>59.0</td>
<td>76.3</td>
</tr>
<tr>
<td>± 1.5° pitch</td>
<td>54.5</td>
<td>71.8</td>
</tr>
<tr>
<td>± 3.0° pitch</td>
<td>45.1</td>
<td>61.5</td>
</tr>
<tr>
<td>± 5° roll</td>
<td>54.1</td>
<td>71.9</td>
</tr>
<tr>
<td>± 10° roll</td>
<td>45.8</td>
<td>63.6</td>
</tr>
</tbody>
</table>
TU/e MPS Research Vehicle

tue-mps.org
Vehicle Integration
Anticipatory systems: Anticipate on possible future events from sensory data from the past.
Future Research Challenges

Extending the prediction horizon for the behavior of road users in highly dynamic urban traffic situations from 1 second to several seconds.

- Too complex for rule-based approaches
- Too few training samples, due to combinatorics, for learning-based approaches

Requires novel approaches for spatiotemporal and inter-object reasoning
AI Paradox

- Majority of adults:
  - Cannot beat the Chess world champion
  - Can drive a car safely
- Artificial Intelligence
  - Can beat the Chess world champion
  - Cannot drive a car safely

Driving a car safely is more challenging than beating the Chess world champion
Conclusion

- Current approaches for automated driving either do not reach SAE level 4 or are not scalable

- We require scalable HAD mapping technology
  - Crowdsourcing with Distributed SLAM

- We require improved AI that can anticipate
  - From pattern recognition to reasoning
Questions