A Multi-Modal System for Road Detection and Segmentation

Xiao Hu\textsuperscript{1} / Sergio A. Rodríguez F.\textsuperscript{1} / Alexander Gepperth\textsuperscript{2}

\textsuperscript{1}University Paris Sud - \textsuperscript{2}ENSTA ParisTech

IEEE Intelligent Vehicles Symposium
Dearborn, June 2014
Motivation

Identify the drivable area in structured environments is a key task for Intelligent and Autonomous Vehicles:

- Helps to
  - Define navigation strategies
  - Constrain obstacle research space
  - Adapt object tracking dynamics
Problem statement

Road detection based on illumination invariant images

Model-based Classifier defined in a shadow-invariant feature space [ALVAREZ2011]\(^1\)

**Method basis**
- Online chromaticity selection for road surface
- Robust to shadows and illumination changes

**Assumptions**
- Road model updated from samples at each frame

---

Motivation

Intended to improve road detection:

- Monocular road detection
  - Needs road feature initialization
  - Enlarges functional range (in high traffic situations)
  - Helps to cope with overexposure
Outline

1. Introduction
   - Problem statement

2. Perception System
   - System set-up
   - Strategy
   - Ground Plane Detection
   - Data Combination

3. Experimental Results

4. Conclusions
Modalities of the Perception System

KITTI Vehicle Set Up:

- Color camera (passive)
  - Facing front covering 90° horizontal
  - Resolution 1242 × 375
- HD LIDAR (active)
  - Velodyne: 64 layers over 360°
- Vehicle speed measurements

Assumptions:

- Synchronized data is considered
- Intrinsic camera parameters
- Extrinsic parameters (Cam-HD LIDAR)
Perception Strategy

- HD Multi-layer LIDAR
- RGB Camera

Road Boundary Detection
Identification of the navigable space

Ground Plane Estimation
Ground surface model

Ground Points Extraction
Classification of 3D points lying on the ground surface

Illumination-invariant Image Conversion
RGB space projection into log-chromaticity space

Probability Map
Bayesian framework for data fusion

Image Segmentation
Pixel classification using road features

Outline
- Detect and introduce scene structure constraints into image segmentation
- Bayesian framework to combine information
- Experimental validation and performance quantification
Scene Structure Detection

Road Boundary Detection

- Low part layers, from HD LIDAR data, are projected to 2D plane ($X^L - Y^L$)
- Histogram orthogonal to ego driving direction
- Estimates are temporally filtered dealing with histogram dilution (e.g. turns)

Plane Estimation

Identification of candidate LIDAR points lying on the ground plane

**Assumption:**
Road boundaries stand on the ground plane
- Cell based representation according to their $X^L$ coordinate
- Robust plane fitting using lowest points of cells
Scene Structure Detection (II)

Ground Points Extraction

- 3D classification w.r.t detected ground plane
- **Criterion**: orthogonal distance threshold (small pitch changes)
- **Assessment indicator**: Ground point set cardinality, $N_g, t \ll N_g, t_{t-1}$
  - Unreliable estimate: Ground points prediction through a constant speed model, $P_{t|t-1} = A \cdot P_{t-1}$

Sergio A. Rodriguez F.
Probabilistic framework

Probability Map

- **Sky removal** [HELD2012]²
  - Assumption: Road flatness
  - Projection of a ground point \( P_\infty \) onto the image plane
  \[
  \text{pixel}_\infty = M_C \cdot P_\infty
  \]  

- **Parametric Gaussian Model Update**
  - Parametric Gaussian function \( F(\mu, \sigma) \) updated every image
  - Representative training data: projected LIDAR ground points

Probabilistic framework

Classification and Segmentation

- **Road similarity map**
  - Illumination-invariant features scored through the updated model
  
  \[
  \text{prob}_{x,y} = \exp\left(-\frac{(I_{\theta_{x,y}} - \mu)^2}{2\sigma^2}\right)
  \]  

  (2)

- **Model based classification** [PETR2008]³

  \[
  \text{pixel} = \begin{cases} 
  \text{road} & \text{if } \text{prob}_i > T_{\lambda} \\
  \text{non-road} & \text{otherwise}
  \end{cases}
  \]  

  (3)

  - Probability map errors and outliers? (e.g. grassland)
  - Constraint: Region with the most road pixels included
  - Morphological operator: flood-fill algorithm

Performance Evaluation

KITTI dataset
- Comparison of proposed method vs. [ALVAREZ2011]
- Matlab implementation: ~600ms per frame
- Use cases: nearby objects, illumination changes

Parameters:
Histogram resolution for road boundary detection: 0.15m
Ground plane RANSAC threshold: 0.15m
Model based classification of road similarity map: 0.68
Experimental Results

Quantitative Evaluation

- **KITTI-UMM dataset [Training set]:** Collection of 95 images from several videos
- Pixel-wise measures
- No temporal filtering was used (fair comparison)
- \(F_{\text{max}}\) is obtained for a given \(T_x\)

<table>
<thead>
<tr>
<th>Pixel-wise measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>(\frac{TP}{TP + FP})</td>
</tr>
<tr>
<td>Recall</td>
<td>(\frac{TP}{TP + FN})</td>
</tr>
<tr>
<td>F1-measure</td>
<td>(\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}})</td>
</tr>
<tr>
<td>(F_{\text{max}})</td>
<td>(\text{argmax}(\text{F1-measure}))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>(F_{\text{max}})</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>81.84 %</td>
<td>72.66 %</td>
<td>93.68 %</td>
</tr>
<tr>
<td>V-disparity (^4)</td>
<td>80.64 %</td>
<td>80.35 %</td>
<td>80.93 %</td>
</tr>
<tr>
<td>BL (^5)</td>
<td>76.17 %</td>
<td>65.02 %</td>
<td>91.95 %</td>
</tr>
</tbody>
</table>

- \(F_{\text{max}}\) improvement
- False detections are increased if ground detection fails
  - Temporal filtering for ground plane detection


Conclusions

- A multi-modal based road detection approach was proposed and experimentally validated.
- Ground detection was used to better initialize image road segmentation.
- The performance of the approach was compared to the state-of-the-art using a public dataset.
- The obtained results show that merging color information and geometrical context improves the detection performance in terms of F1-measure.

Future works

- Multiple object tracking aided by scene structure detection (i.e. context).
Thank you for your attention!