Department of Informatics
Technical University of Munich

Exposé on Master's Thesis

Name: Joseph M. Birkner
Matr. Nr.: 03704462
EMail: joseph.birkner@tum.de
Supervisor: Prof. Dr.-Ing. habil. Alois C. Knoll
Institute: Chair of Robotics, Artificial Intelligence and Real-time Systems
Title: Monocular 3D Traffic Perception Using HD Maps as an Auxiliary Feature
Date: July 10, 2022

1 Topic

This work is conducted within the scope of the **Providentia++** project. The project aims to provide cars with Digital Twins in the cloud, facilitating enhanced traffic prediction and micro-routing abilities for safer autonomous driving. Under the Providentia project, Sensor packages, encompassing RGB Cameras, Radar, and Lidar, are installed along major roads and intersections. We are evaluating, which minimal Sensor-suite would be suitable for a wider rollout of the sensing infrastructure under minimum cost.

The "Digital Twin" data-point of a vehicle consists of multiple components: First and foremost, position $(x|y|z)$, size $(w|h|d)$ and orientation (θ) are key variables which define a vehicle's spatial state. Since the RGB camera sensor is the cheapest among the installed package, 3D object detection from 2D video is an important area of research within Providentia.

A working initial approach for monocular 3D detection for Providentia++ was developed by XXXX in the scope of a Bachelors thesis, using 3D projection of the lower edge of 2D vehicle segments. This approach works very well for vehicles which are moving in straight lines, as the orientation can be fixed to a constant value. However, more work needs to be done for reliable monocular detection when observing traffic scenes with heavily varying vehicle orientations, especially scenes such as complex intersections.

The goal of this work is to improve the orientation estimation of turning vehicles by exploiting clues about their heading from HD maps of the observed road scene. Both map-matching and heading estimation will also benefit from considering the prior trajectory of a vehicle. The trajectory can be obtained by chaining observations of an individual vehicle across multiple video frames, thereby introducing a time component to the observation. Once bounding box estimates between frames are related through vehicle identities, it will also become possible to stabilize predictions about their spatial state via Kalmann Filters, Recurrent Neural Networks or other methods.

2 Approach

Citation: [**choset2005principles**; **thrun2005probabilistic**]

Reference: fig. 1

Figures:

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an

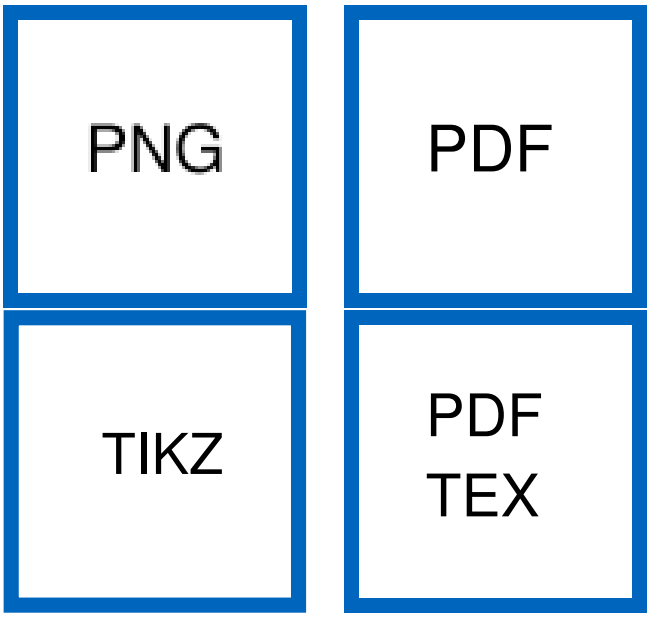

Figure 1: Image description.

impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match

the language.

3 Work Plan and necessary Resources

4 Literature