**Computer Vision: 3D WiSe 2022/23**

**Project Report**

**Stereo Visual Odometry**

**Project group H:**

Liu, Yaxi (Data and Computer Science, MtkNr. 3769313)

Xu, Shengzhe (Data and Computer Science, MtkNr. 3769534)

Schrepfer, Jule Katharina (Physics Master, MtkNr. 3736320)

Patil, Sanmati (Scientific Computing, MtkNr. 3444150)

**Abstract**

Accurate localization of a vehicle is a fundamental challenge and one of the most important tasks of mobile robots. For autonomous navigation, motion tracking, and obstacle detection and avoidance, a robot must maintain knowledge of its position over time. Vision-based odometry is a robust technique utilized for this purpose. It allows a vehicle to localize itself robustly by using only a stream of images captured by a camera attached to the vehicle. This paper describes a stereo visual odometry algorithm for estimating frame-to-frame camera motion from successive stereo image pairs. In this work, we implemented stereo visual odometry using images obtained from KITTI Vision Benchmark Suite and presented obtained results.

**Table of Contents**

**1. Introduction**

**2. Overview**

**3. Method (Your solution)**

**4. Experiments,**

**5. Findings**

**6. Conclusion**

1. **Introduction**

In [robotics](https://en.wikipedia.org/wiki/Robotics) and [computer vision](https://en.wikipedia.org/wiki/Computer_vision), Visual Odometry is the process of determining the position and orientation of a robot by analyzing the associated camera images. The idea was first introduced for planetary rovers operating on Mars – Moravec in the early 1980s. Visual Odometry  is a technique used to localize a robot by using only a stream of images acquired from a single or multiple cameras attached to the robot. The images contain a sufficient amount of meaningful information (color, texture, shape, etc.) to estimate the movement of a camera in a static environment.

Stereo visual odometry estimates the camera's egomotion using a pair of calibrated cameras. Stereo camera systems are inherently more stable than monocular ones because the stereo pair provides good triangulation of image features and resolves the scale ambiguity.

1. **Overview**

**2.1 Dataset used**

○ KITTI Vision Benchmark Suite : The KITTI Odometry dataset was used in our project. The dataset contains 21 sequences of stereo video sequences in greyscale.

○ Calibration Files: Camera and projection matrices are provided to the user with every video sequence in the form of a calibration file. Time stamps of every frame is also provided.

○ The camera orientation is as follows and shown in Fig. 1 below:

* The X-axis is parallel to the ground and towards the right of the driver.
* The Y-axis is perpendicular to the ground and facing downwards.
* The Z-axis is is parallel to the ground and facing forward

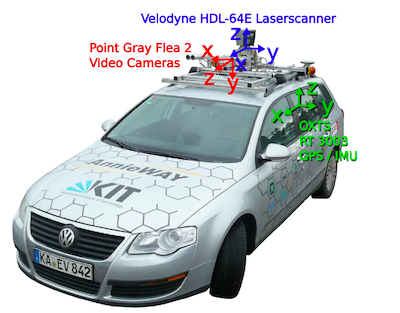


Fig. 1: KITTI dataset recording platform: VW Passat station wagon is equipped with four video cameras (two color and two grayscale cameras), a rotating 3D laser scanner and a combined GPS/IMU inertial navigation system. (Image taken from the KITTI dataset paper)

**2.2. High level approach for Visual Odometry**

The typical pipeline of Visual Odometry is as follows :

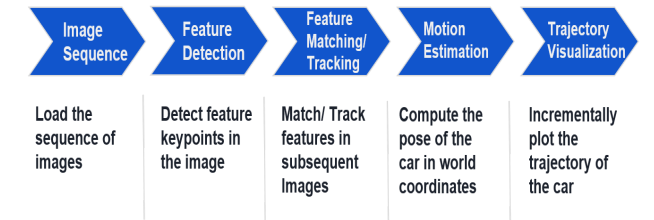


Fig. 2: Pipeline of Visual Odometry

**2.3. Our Approach**

We have implemented Stereo Visual Odometry and our approach is as follows:

1. Inputs: rectified images
2. Feature Detection: detect corner features, assign 3D vals from disparity
   * Used Harris corners and FAST features
3. Feature tracking: using sum of absolute differences generate feature scores
   * Low score indicates match
4. Match Features: use local minima from score to generate features

* Improves on state-of-the-art computation : from cubic to squared complexity

1. Find Inliers: inforce rigid world constraint to reject unlikely features
2. Estimate Motion: minimize the reprojection error and output motion