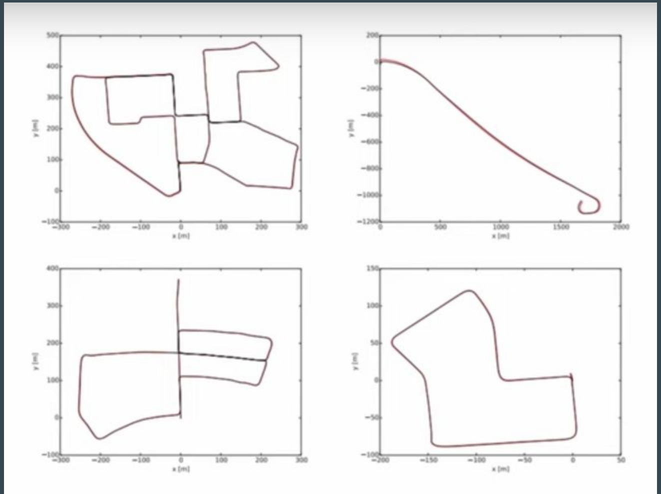
Stereo Visual Odometry

Marno Nel and Rintaroh Shima

Goal

- Estimate the 3D motion of a camera by analyzing the disparity between images captured by two cameras
- Track the camera's movement and provide a precise estimation of its trajectory in a 3D space by matching corresponding points between two camera frames



Approach/Design - Extracting Features

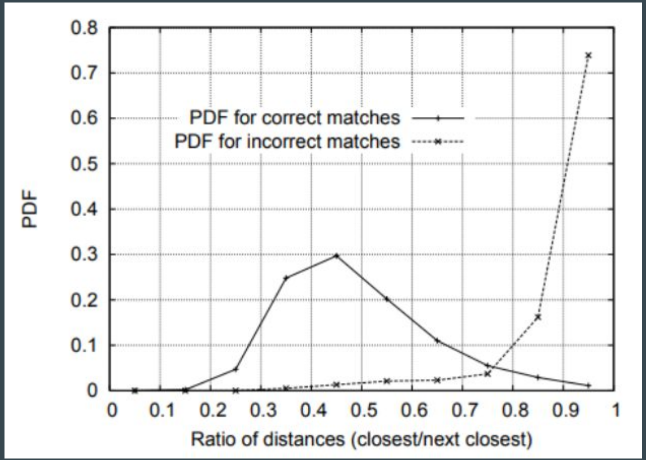
- Feature detectors used:
  - Scale-Invariant Feature Transform (SIFT)
  - Oriented FAST and Rotated BRIEF (ORB)
- SIFT
  - For offline applications, where accuracy is prioritized more than efficiency
- ORB
  - For real-time applications, such as SLAM and visual odometry, where computational efficiency is crucial

Approach/Design - Matching Features

- Brute-Force Matcher
  - For SIFT, we used L2 norm (Euclidean distance)
    - Descriptor vectors are real-valued and represent local image gradient information
  - For ORB, we used Hamming distance
    - ORB is a binary descriptor that consists of binary strings that encode the presence or absence of certain image features
- Used k-nearest neighbor to get the best matches

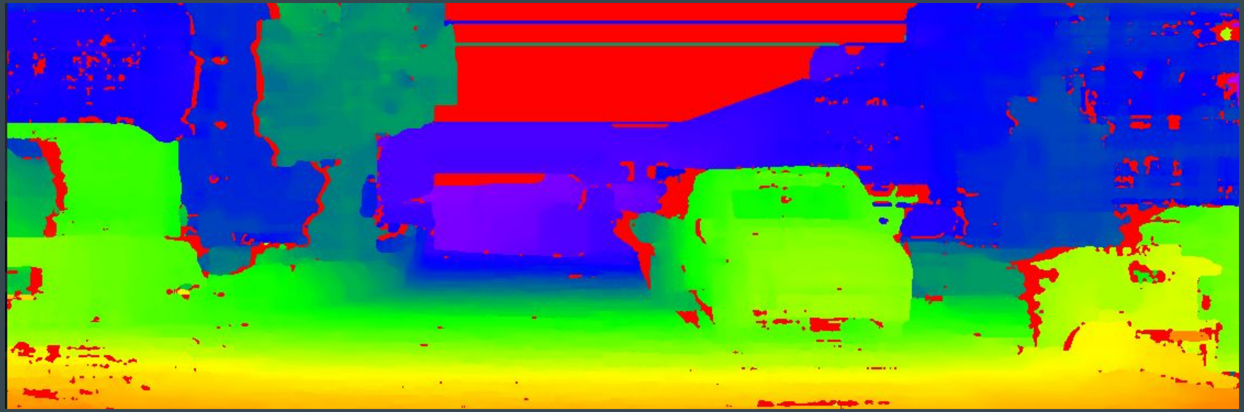
Approach/Design - Lowe's Ratio Test

- Technique used to filter out ambiguous matches
- Compares the distance between the best and second-best matches for a given feature descriptor and accepts the match only if the ratio of their distances is below a certain threshold



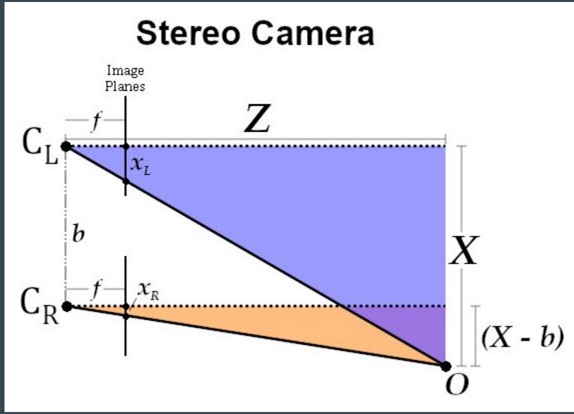
Approach/Design - Generating Disparity Map

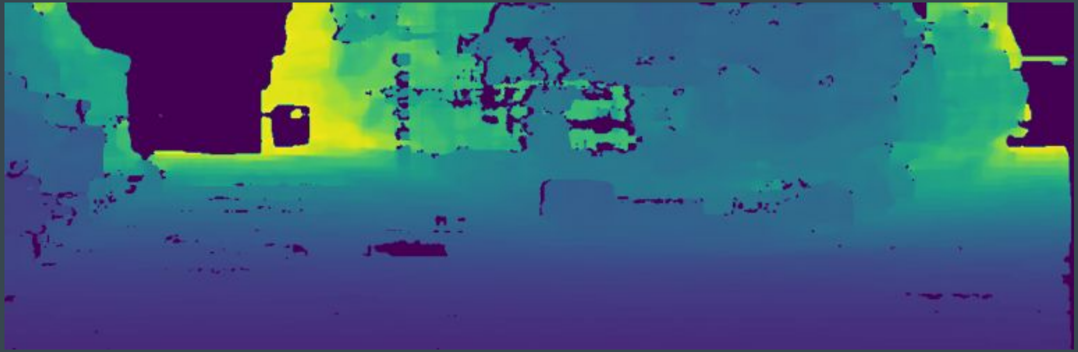
- Compute dense disparity map from a pair of stereo images
  - Stereo Block Matching (StereoBM)
    - Produces reasonably accurate disparity maps
    - Computationally inexpensive
  - Stereo Semi-Global Block Matching (StereoSGBM)
    - Produces higher-quality disparity maps by incorporating global optimization
    - Relatively slower compared to StereoBM



Approach/Design - Generating Depth Map

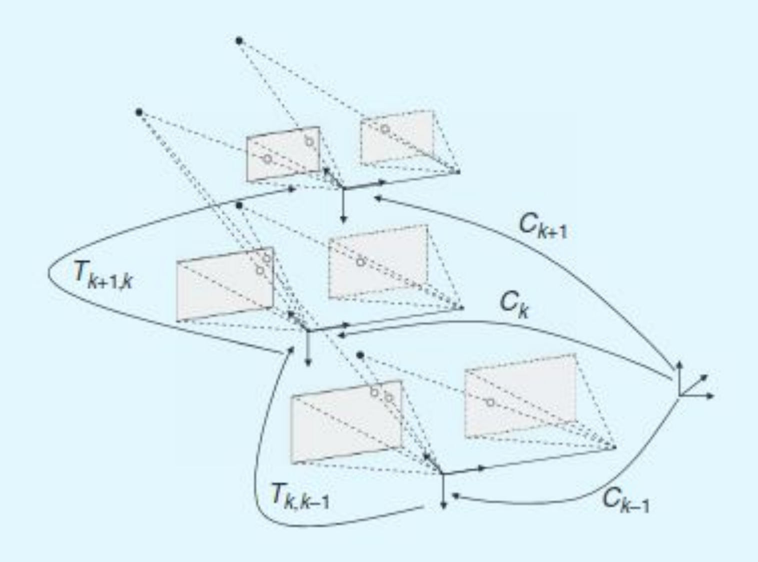
- Extract the focal length and baseline from the projection matrix
- Using similar triangles, we can get the following equation to compute the depth:
    - f * b = Z * d -> Z = f * b / d
    - Where f is the focal length, b is the baseline, d is the disparity, and Z is the depth





Approach/Design - Estimating Motion

- Get the depth from the depth map for all the features
- Using the equations below, calculate the x and y coordinates from the pixel coordinates of the features
  - x = (u - c_x) * z / f_x
  - y = (v - c_y) * z / f_y
  - Where (u, v) represents the pixel coordinates, (c_x, c_y) represents the optical center of the image, z represents the depth, and (f_x, f_y) represents the focal length
- Use the Perspective-n-Point RANSAC algorithm to obtain the rotation matrix and translation vector of the camera between two consecutive frames



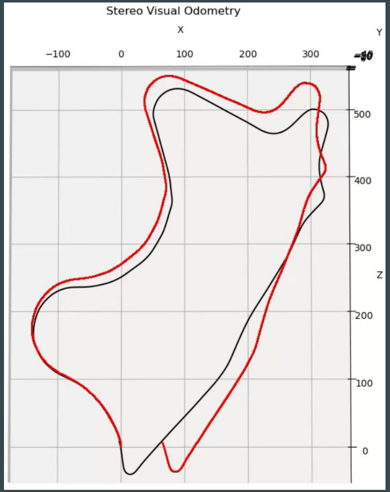
Results for KITTI Dataset 09

Time Comparison:

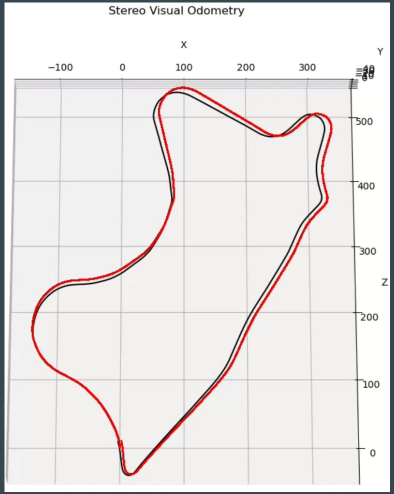
- Ground truth = 2 min and 45 sec
- ORB, Brute-Force, distance ratio of 0.6, StereoBM = 5 min and 50 sec
- SIFT, Brute-Force, distance ratio of 0.45, StereoSGBM = 10 min and 35 sec
    - ORB is 55% more efficient than SIFT

2D Endpoint Accuracy Comparison:

- Ground truth coordinates = [-1, 8] m
- ORB coordinates = [65, 8] m
- SIFT coordinates = [2, 10] m
    - SIFT is 94.5 % closer in Euclidean distance to ground truth than ORB



ORB



SIFT

Improvements

- Sensor fusion
  - Integrate other sensors, such as IMU or LiDAR to improve accuracy and robustness
  - Kalman filtering or particle filtering can be applied to combine the data from multiple sensors effectively
- Visual SLAM and loop closure detection
  - By identifying previously visited locations in the scene, it can help mitigate cumulative drift errors and improve the global consistency of the estimated camera trajectory

References

[1] D. Scaramuzza and F. Fraundorfer, "Visual Odometry [Tutorial]," in IEEE Robotics & Automation Magazine, vol. 18, no. 4, pp. 80-92, Dec. 2011, doi: 10.1109/MRA.2011.943233.

[2] F. Fraundorfer and D. Scaramuzza, "Visual Odometry : Part II: Matching, Robustness, Optimization, and Applications," in IEEE Robotics & Automation Magazine, vol. 19, no. 2, pp. 78-90, June 2012, doi: 10.1109/MRA.2012.2182810.