

Depth Estimation Methods

Moustafa Esam, Team 1

September 15, 2023

1 Introduction

Depth estimation is an essential task in computer vision required for applications such as simultaneous localization and mapping, object detection, and semantic segmentation. It involves reconstructing a 3D model from 2D images, often using advanced techniques to organize depth information. This can be achieved using 2D-sensored cameras. In this report, we provide an overview of depth estimation using optical camera configurations, including stereo cameras, monocular cameras, and RGBD cameras.

2 Monocular Camera

Mono cameras are single-view cameras that capture images using a single lens and an image sensor, providing 2D images. However, depth information can be estimated from these cameras using various techniques.

2.1 Visual Odometry

Visual odometry involves identifying the depth and location of objects by analyzing a sequence of images. It estimates depth by tracking the camera's motion through consecutive frames and analyzing how the scene moves across different viewpoints. This technique is valuable for mapping and navigation.

2.2 Structure From Motion (SfM)

SfM is a technique that constructs 3D models of objects by analyzing consecutive 2D frames, producing 3D models similar to LiDAR. It uses triangulation and other image analysis methods to process a series of 2D frames and output the 3D model.

2.3 Convolutional Neural Networks (CNNs)

Deep learning techniques, such as CNNs, involve training models on datasets like NYU or KITTI to identify depth maps from single images.

3 Stereo Camera

Stereo cameras consist of two lenses separated by a known baseline, capturing two images of the scene from slightly different angles, which provides depth estimation cues. Some commonly used techniques include matching and triangulation.

3.1 Matching

Matching techniques aim to find corresponding pixels between two images.

3.1.1 Stereo Matching

Stereo matching relies on the disparity between images taken from two cameras separated by a known horizontal distance. By finding matching pixels in the two images, depth can be accurately estimated, though this technique is sensitive to lighting conditions.

3.1.2 Semi-Global Matching (SGM)

SGM is a stereo matching algorithm that considers global information in the disparity map, minimizing the cost function and resulting in improved accuracy.

3.2 Triangulation

Triangulation is a technique similar to stereo matching but relies on calculations to determine depth after extracting disparity from two images. It is less sensitive to lighting conditions.

4 RGBD Camera

An RGBD camera is an RGB camera that also provides depth data, constructing a depth map using a 3D depth sensor such as a time-of-flight or stereo sensor.

- Direct Depth Measurement: RGBD cameras use time-of-flight or structured light to directly measure depth, providing accurate depth maps.
- Depth From Stereo: These cameras use stereo sensors for depth data, combining data from the two sensors to offer highly accurate measurements.
- Real-time 3D Mapping: RGBD cameras can create real-time 3D models of objects and scenes since they are sensor-based.

5 Conclusion

Estimating depth from cameras is a fundamental step in obtaining a 3D model of the scene or world. Monocular, stereo, and RGBD cameras each offer different approaches with unique characteristics for depth estimation, serving various application requirements. While monocular cameras are lightweight and provide real-time data, stereo cameras are known for their high accuracy. RGBD cameras combine the best of both, providing both color and depth data. As technology continues to advance, these methods will become more precise, introducing new possibilities in the fields of robotics and computer vision.

6 Referencces

Learnopencv article
Paperswithcode article
Springer articles