Depth estimation methods

Estimating Depth from Cameras: Methods and Approaches

In the field of computer vision, estimating depth from cameras is a fundamental task that enables us to perceive the three-dimensional structure of the world. By accurately estimating depth, we can create immersive virtual reality experiences, enable autonomous navigation, and facilitate augmented reality applications. In this article, we will explore different methods and approaches to estimate depth using various types of cameras, including Mono Camera, Stereo Camera, and RGBD Camera.

Mono Camera

A Mono Camera, also known as a single-camera system, captures images using a single lens. Estimating depth from a mono camera is a challenging task since it lacks the ability to perceive depth directly. However, there are several techniques that can be used to infer depth information from a mono camera.

Motion-based Approaches

One common approach to estimating depth from a mono camera is by utilizing the motion cues in the scene. By tracking the movement of objects or the camera itself over consecutive frames, it is possible to estimate depth information. This can be achieved through optical flow algorithms or by using visual odometry techniques that estimate camera motion and scene structure simultaneously.

Depth from Defocus

Depth from defocus is another technique used with mono cameras to estimate depth information. It leverages the variation in the blur of objects at different distances from the camera. By analyzing the amount of defocus in the image, it is possible to infer the depth of objects. This approach typically requires knowledge of the camera's focal length and aperture settings.

Stereo Camera

A Stereo Camera consists of two or more cameras arranged side by side, mimicking the human binocular vision system. By capturing images from slightly different viewpoints, stereo cameras can estimate depth by triangulating corresponding points between the left and right images.

Triangulation using 2 cameras

As seen in the image, we have 2 identical cameras left camera and right camera displaced by the horizontal baseline b. where the z-depth is inversly proportional to the disparity(uL-ur)

Trigonometric functions

Trigonometric Functions By analyzing the angular size of an object within the camera's field of view and the focal length of the camera lens, trigonometry helps us calculate the distance to the object. The fundamental trigonometric relationship involved is the tangent function, as it relates the object's height or width to the distance, forming the basis of distance estimation algorithms.

Stereo Correspondence

Stereo correspondence algorithms match pixels or features between the left and right images to establish correspondences. These correspondences are then used to compute the disparity, which represents the horizontal shift of pixels between the two images. By knowing the disparity, the depth can be estimated using the triangulation principle.

Structure from Motion

Structure from Motion (SfM) is a technique that estimates the 3D structure of a scene by jointly estimating camera motion and the 3D positions of points in the scene. By using multiple camera views, SfM can reconstruct a dense 3D point cloud, which provides depth information. SfM algorithms typically require feature matching and bundle adjustment techniques to refine the 3D reconstruction.

RGBD Camera

An RGBD Camera, such as the Microsoft Kinect or Intel RealSense, combines a traditional RGB camera with an additional depth sensor. These cameras provide both color and depth information for each pixel in the image, enabling more accurate depth estimation compared to mono or stereo cameras.

Time-of-Flight

One common technology used in RGBD cameras is Time-of-Flight (ToF) sensing. ToF cameras emit a light signal and measure the time it takes for the signal to bounce back from objects in the scene. This information is used to compute the depth at each pixel, providing a dense depth map of the scene.

Structured Light

Another approach used in RGBD cameras is structured light. In this method, a pattern of light is projected onto the scene, and the deformation of the pattern on the objects is captured by the camera. By analyzing the deformation, the depth at each pixel can be estimated.

Conclusion

Estimating depth from cameras is a crucial task in computer vision, enabling a wide range of applications. In this article, we explored different methods and approaches for estimating depth using Mono Camera, Stereo Camera, and RGBD Camera. While mono cameras require innovative techniques to infer depth indirectly, stereo cameras leverage the principle of triangulation to estimate depth. RGBD cameras, on the other hand, provide direct depth measurements, enabling more accurate and dense depth estimation. The choice of camera and depth estimation method depends on the specific application requirements, accuracy needs, and available hardware resources.

Advancements in depth estimation techniques and camera technologies continue to push the boundaries of 3D perception, making it possible to create more immersive and interactive experiences in fields such as virtual reality, robotics, and augmented reality.