Depth From Camera

Faculty of Engineering, Alexandria university

Authors :
Mariam Mahmoud,Sama Abdou,Ahmed Frag
Abdelrahma.n Dief , Yousef Khaled , Ahmed Gamal

Moderators:
Hassan Tamer

Abstract

In computer vision and robotics, the attempt to understand the three-dimensional structure of our surroundings is no longer an issue. The article gives a general overview of various techniques and strategies used to measure depth and obtain a three-dimensional (3D) representation of the environment using various camera configurations, including mono, stereo and RGBD cameras.

1 Introduction

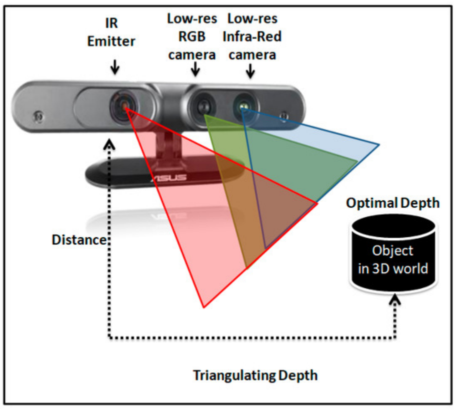
A crucial task in computer vision and several applications, including robotics, augmented reality, and 3D modelling, is depth estimation, the process of calculating the distance to objects in a scene. Active and passive approaches are the two main techniques used in depth estimation. To extract depth information from camera systems, each strategy makes use of particular methods and ideas.In this article, we go over the main differences between active and passive depth estimation methods.

2 Active approach

The DI is quickly accomplished in active approaches (such as using LIDAR sensors and RGB-D cameras). A specific kind of depth-sensing device called an RGB-D camera combines an RGB image with the corresponding depth image. Due to their low cost and power consumption, RGB-D cameras can be used in a variety of devices, including smartphones and unmanned aerial systems. The depth range of RGB-D cameras is constrained, and they have issues with specular reflections and absorbing objects. To bridge the gap between limited and dense depth maps, many depth completion strategies have been proposed.

2.1 RGB-D Cameras

Depth sensors, known as RGBD cameras, project an infrared pattern and calculate the depth from the reflected light using an infrared sensitive camer.Taking a human face as an example, the scanner emits a light pattern on the target face, and calculates the shape of the surface based on its deformation, thereby calculating the depth information of the face.Following the depth measurement, the RGB-D camera typically pairs the depth and colour map pixels in accordance with the placement of each camera during production, and outputs a one-to-one corresponding depth map and colour map. We can compute the 3D camera coordinates of the pixels, read the colour information and distance information at the same image location, and produce a point cloud.
we define an RGB-D image as an image with 3 standard color channels and an additional depth channel. The depth channel contains a measure, in meters, of the distance from the camera optical center to the scene, along the camera optical axis. We adopt the standard camera coordinate frame convention with the z axis pointing along the optical axis. We expect RGB-D images to have complete color information, and incomplete, yet very dense, depth information
An RGB-D camera will typically output a color and depth image produced by two separate cameras, RGB and infrared, each with its own intrinsic parameters and distortion coefficients. The device and its software driver might have a built-in calibration for the intrinsics of the two cameras, as well as the extrinsic pose between them. This allows for the two images to be properly undistorted, and for the depth image to be reprojected into the frame of the RGB camera to form a single unified RGB-D image



2.1.1 Hardware of RGB-D

Depth cameras are vision systems that alter the characteristics of their environment, mainly visual, in order to capture 3D scene data from their field of view. These systems have structured lighting, which projects a known pattern on the scene and measures its distortion when viewed from a different angle. This source may use a wavelength belonging to the visible field, but more commonly, will be selected from the infrared field. The more sophisticated systems implement the time of flight (ToF) technique, in which the return time of a light pulse after reflection by an object in the scene captures depth information (typically over short distances) from a scene of interest.

2.1.2 Methods using in RGB-D cameras

1-Stereo vision involves using two or more cameras to capture images from slightly different viewpoints, similar to how our eyes perceive depth. The depth is calculated based on the parallax (displacement) between corresponding points in the two images.
Disparity: The disparity (d) between corresponding points in the left (L) and right (R) images is calculated:

$$d = x\_L - x\_R$$

where x_L is the x-coordinate of a point in the left image and x_R is the x-coordinate of the same point in the right image.
Depth Calculation Once the disparity is known, the depth Z can be calculated using the formula

$$Z = f * \frac{B}{d}$$

where:
Z is the depth of the point.
f is the focal length of the camera.
B is the baseline (distance between the two cameras).
d is the disparity. **2-Time-of-Flight (ToF):**Time-of-flight cameras measure the time it takes for a light pulse to travel to an object and back. The depth is calculated based on the speed of light and the time-of-flight of the pulse
Depth Calculation: The depth (Z) is calculated using the formula:

$$Z = \frac{c * t}{2}$$

where:
Z is the depth.
c is the speed of light.
t is the time-of-flight of the light pulse.

3 Passive Approach

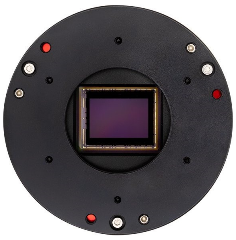
illuminates the scene while in the latter case, the scene illumination is provided by the ambient light. Examples of active range-finding techniques that involve sending a controlled energy beam and detection of the reflected energy include ultrasonic and optical time-of-flight estimation. Contrived lighting-based approaches include striped and grid lighting and Moire fringe patterns. Passive range-finding techniques are image-based methods. Monocular image-based techniques include texture gradient analysis, photometric methods, occlusion cues, and focusing- and defocusing-based ranging. Methods based on motion or multiple relative positions of the camera include reconstruction from multiple views, stereo disparity analysis, and structure from motion. Most of these methods are actually geometric triangulation systems of different kinds. In fact, almost every circumstance that includes at least two views of a scene is potentially exploitable as a range finder. Most of the active ranging techniques have little to do with the human visual system. Their purpose is neither to model nor to imitate the biological vision processes, but rather to provide an accurate range map to be used in a given application. Passive methods have a wider range of applicability since no artificial source of energy is involved, and natural outdoor scenes fall within this category. In light of the fascinating power of humans in inferring 3-D structure of objects from visual images, a great deal of effort has been directed towards this end. Passive techniques are particularly appropriate for military or industrial applications where security or environmental constraints preclude the use of light sources such as lasers or powerful incandescent projectors. However, active ranging methods based on structured lighting sources or time-of-flight measuring devices are certainly acceptable in indoor factory environments.

3.1 Stereo cameras

Stereo cameras are passive sensors that are used to determine depth. Similar to how our eyes perceive depth, they use passive techniques to calculate depth information by taking two or more pictures of a scene from marginally different angles. The disparity (shift) between corresponding points in the images caused by these various points of view is used to determine the depth of objects in the scene.
Here's how stereo cameras work in a simplified way:
Image Capture: A stereo camera captures two or more images (usually grayscale) of the same scene from slightly offset positions. These positions simulate the separation between our eyes.
Correspondence Matching: The camera software or algorithm identifies corresponding points or features in the left and right images. These are points in the scene that are visible in both images, but their positions differ due to the stereo baseline (distance between the camera lenses or sensors).
Disparity Calculation: The disparity, which is the horizontal shift between the positions of corresponding points in the left and right images, is calculated for each matched pair of points. Larger disparities correspond to objects that are closer to the camera, while smaller disparities correspond to objects that are farther away. **Depth Estimation:** Using the disparity information and the known geometry of the stereo camera setup (baseline distance and focal length), the depth of each point in the scene can be calculated. There are various mathematical equations, such as the triangulation formula, that relate disparity and depth.
To sum up, Stereo cameras are passive because they do not emit any active signals, such as laser beams or structured light patterns (used in active depth sensors like LiDAR or structured light cameras). Instead, they rely on the natural ambient light and the inherent texture and features of the scene to calculate depth. This makes them well-suited for a wide range of applications, including computer vision, robotics, and 3D reconstruction, without the need for additional illumination or active sensing technologies.

3.2 mono cameras

Mono cameras are very similar to OSC cameras in that they share similar body designs, and have a sensor and cooling circuitry. What makes Mono cameras fundamentally different is that the sensor has no filters applied to its surface. No Bayer Pattern is used at all. Every pixel sees the full spectrum of light that the sensor is capable of responding to. When you expose a subframe with just the camera, you get an image that collects photons from across the spectrum. While this allows the sensor to maximize the capture of photons, it does so with no color information - all color detail is lost. This situation can be improved by the use of filters. Putting a filter in front of the sensor shapes the light being measured. Placing a red filter in front of the sensor would allow you to selectively measure light from only the red part of the spectrum. If I used separate Red, Green, and Blue filters to create a series of subframes, I could stack the frames from each filter to create an R, G, and B master image, which to then be combined to form a color image. I can play further games by using precision filters that carefully target the light I want to measure. This can work to eliminate the light that you don't want to measure (i.e. light pollution), ensuring that the light I am measuring is coming from the selected target. While a filter drawer can be affixed to your imaging chain, allowing you to swap out one filter at a time, it is much more common and convenient to use a filter wheel that can hold 7 or 8 filters. The filter wheel can then be rotated to select the filter desired at any given point.



3.2.1 Depth Estimation

in monocular (or "mono") cameras is a more complex task compared to stereo cameras, as monocular cameras capture images from a single viewpoint and lack the direct depth information provided by stereo disparities. Instead, depth estimation in monocular cameras often involves using computer vision techniques and additional information

$$Z = \frac{f * h}{p}$$

where :

Z is the depth of the point in the scene.
f is the focal length of the camera lens (in meters).
h is the height of the object in the image (in meters).
p is the height of the object in pixels in the image.

3.2.2 Pros and cons for Mono Cameras

Pros

Flexibility: Can do broadband RGB capture as well as very broad L captures and can do Narrowband capture Allows for mixes of NB and Broadband signal to create unique looks
Many filters to choose from, different frequency targets, bandwidths, and quality levels available
Very Efficient: All pixels used to full benefit for each filter exposure and filter can maximize light being captured
Narrowband can reject all but most important wavelengths for nebulae, dramatically increasing contrast for some objects
No dependence upon Bayer interpolations
Processing Flexibility
Each color filter can be pre-processed and processed as needed to maximize quality and Color combinations and blend methods are rich and varied
Easier to address color channel specific noise or gradients, which yields - according to many - yielding cleaner backgrounds
Robustness and Better able to deal with gradients from Light Pollution and Moonlight (with Narrowband)

Cons

More Expensive: The cost of a complete set of filters can be very high and the cost of the filter wheel.
Larger Form Factor: The resulting Camera/Filter wheel combination is bulkier and less streamlined and The resulting Camera/Filter wheel combination is slightly heavier and contributed to Z-axis balance issues
More Complicated to Use and Greater setup complexity - need to mount camera, filter wheel, and filters
Need to capture more flavors of subs to create a color image - no ability to capture a color image with a single sub and Subs for Narrowband will require longer exposures
Additional calibration files will be required for each color filter used and Focus adjustment needed after filter swap You need a complete set of subs for each filter used in order to process the image
Mono is cable of producing better images but only if the user understands and is able to process the image data appropriately
Mosaics can multiply the amount of processing necessary

4 Conclusion

Depth estimation from cameras is a critical task for understanding the 3D structure of the world. In this article, we explored different methods and cameras for depth estimation, including the Mono Camera, Stereo Camera, and RGBD Camera. Each approach has its advantages and limitations, and the choice depends on the specific requirements of the application. By understanding these techniques, we can leverage depth information to enable advanced computer vision tasks and create immersive experiences.