SYDE 671 Assignment 5

Pascale Walters

Q1: Briefly describe triangulation (using images if you like). Why can't we find an absolute depth for each point when we don't have calibration information for our cameras?

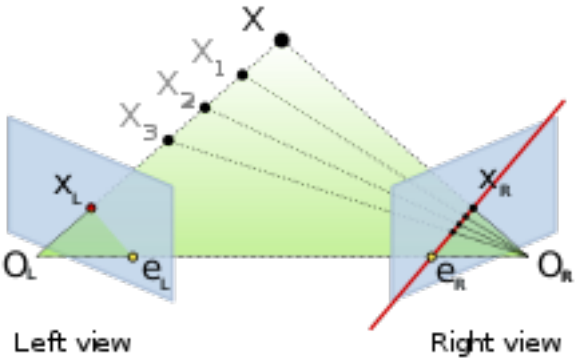
Triangulation refers to determining the 3D location of a point based on its 2D position in at least two images. If the calibration information for the cameras are known, it becomes a matter of solving a geometric problem to find the 3D location. The absolute depth cannot be determined without calibration information for the cameras because the effective size of the pixels must be known.

Q2: Why does rectification simplify matching features across our stereo image pair? What information do we need to know to rectify our image pair?

Image rectification refers to projecting images onto a common plane. Once the images have been transformed, all epipolar lines are horizontal. This simplifies matching features across the pair as only a search along a horizontal line, rather than a diagonal one, needs to be performed. To perform image rectification, the camera calibration matrices (rotation and translation) are required.

Q3: What does it mean when the epipolar lines: a) radiate out of a point on the image plane, b) converge to a point outside of the image plane, or c) intersect at more than one point?

a) There is translation and rotation of the cameras relative to each other such that the baseline is contained within the cameras' respective fields of view, as in the figure below. The two cameras are positioned relative to each other such that the two epipoles e_L and e_R are on the two image planes. A special case of this would be when the epipoles are in the same position in both images when the camera has translated forwards (focus of expansion).



https://en.wikipedia.org/wiki/Epipolar_geometry#/media/File:Epipolar_geometry.svg

b) There is translation and rotation of the cameras relative to each other such that the baseline (the line that connects the two camera centres) is not contained within the cameras' respective fields of view. In the above figure, e_L and e_R would not be on the two image planes.

c) In the case of multiple epipoles, there are likely more than two cameras with multiple baselines.

Q4: Suppose that we have the following three datasets of an object of unknown geometry:

(a) A video circling the object;

(b) A stereo pair of calibrated cameras capturing two images of the object; and

(c) Two images we take of the object at two different camera poses (position and orientation) using the same camera but with different lens zoom settings.

For each scenario:

1. Explain if we can calculate the essential matrix, fundamental matrix, or both;
2. State an advantage and disadvantage of using each setup for depth reconstruction; and
3. Name an application scenario for each of the different setups.

Both the essential and fundamental matrices can be calculated in the case of the video circling the object because the intrinsic camera parameters remain constant and the scene geometry is known. In the case of the stereo pair of calibrated cameras, both the essential and fundamental matrices can be calculated because the scene and cameras are calibrated and all parameters are known. Only the fundamental matrix can be calculated in the case of different intrinsic camera parameters.

An advantage of using video for depth reconstruction is that many frames can be quickly collected using only one camera. Disadvantages are that the images cannot be captured at the same time which may cause the appearance of the object to change and that it may be difficult to control the path and position of the camera. An application of a video circling the object is a CT scan.

In the case of the stereo pair, the conditions can be tightly controlled, but it can require a time-consuming setup. An application is in the academic setting for motion capture. The precise data can be used for applications such as pose estimation.

With two images taken with the same camera but with different poses, the pair can be captured quickly, as there is not much precision required. However, this can be a disadvantage as it becomes impossible to calculate the essential matrix. This setup can be used for depth reconstruction in the wild, such as for an augmented reality application on a smartphone.