Project 5 Questions

Questions

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?

A1: The position of one point in space can be deduced relative to two reference points. Specifically, it can be deduced from two reference angles and the distance between them. Rigorously, consider 3 points in space making $\triangle ABC$. Let B, C be reference points and A be the point of triangulation. Let $\overline{BC}$ be the baseline with length l. An altitude can be dropped down from A to a point, T, on the baseline; altitude $\overline{AT}$ is perpendicular to the baseline and has length d. This creates two right triangles $\triangle ATB, \triangle ATC$. The base legs of these traingles, $\overline{BT}, \overline{TC}$ have lengths that sum to l. The remaining leg is the altitude, so they share length d. Consider $\tan B, \tan C$ which give the proportions of d to each of the remaining legs, l_n for $n = 1, 2$. The lengths of the remaining leg could then be calculated as the quotient of d and the proportion $\frac{d}{l_n}$, which implies:
$l = \frac{d}{\tan B} + \frac{d}{\tan C}$.
This can be rearranged and simplified to $d = l\frac{\sin B \ \sin C}{\sin{(B+C)}}$ Recap: we used the distance between two points and the angles of the rays to a third point to calculate d the distance from the third point to the baseline.
In a camera scene with two cameras and a point they are both looking at, lacking calibration information would imply not having information about the length between the cameras, l in the previous model. Without an absolute l it becomes a free variable which you can scale to different solutions for d. So, instead of having an absolute depth, one can pick an l to fix; This would give a relative depth which would be useful if there were other points triangulated with the same calibration information.

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

A2: Rectification of a stereo image pair produces a new coordinate system where corresponding points in the images now share y values in the rectified plane. This constrains the feature matching process to search on a line in the rectified plane (which can be tranformed back to a line on the corresponding image plane). Rectifying an image pair involves utilizing their fundamental matrix, F, which contains information about both the intrinsics and extrinsics of the cameras. If we have this information, just multiplying with F suffices rectifying the image; otherwise, ground-truth corresponding points are required to estimate F.

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?

A3:

(a) It implies that one camera was in front of the other, causing the epipole to be in the field of view of the camera. Since the epipolar lines all intersect at one point in the image they organize radially; radial epipolar lines infer that corresponding points will either be more inwards or outwards with respect to the epipole.

(b) If the epipole is outside the field of view of the camera, then the point where the lines intersect is too, so the epipolar lines on the image plane will have related slopes in that they all intersect at the epipole. Accordingly, they will appear slanted towards eachother; a cutout of the outside of radial response, like a rectangular patch off the perimeter of a portabello.

(c) This infers that when the epipolar line were calculated, the estimation of the fundamental matrix was not rank 2. To solve this, constraint the fundamental matrix so that $\det(F) = 0$

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

(a) A video circling the object;

(b) An 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.

A4:

(a) The fundamental matrix could be calculated using corresponding points in the different image planes at different moments in the video. Pairs of corresponding points all the way around the object allow for the whole surface of the object to be reconstucted. Circling the object also allows for robostness towards occlusion, allowing for more reconstruction in areas that might not be visible from just two frames (though the reconstructions would not have absoute depth). This would be useful for creating a way for people to make 3d reconstructions of objects using videos from their phones. Such a reconstruction would be useful if one wanted to share the geometry of an object by digital means.

(b) Since the cameras are calibrated, both of the matrices would be able to be calculated. This would allow for an absolute meausure of depth, and so an application it would be useful for is to deduce how far an object in a scene is.

(c)