Shape from Stereo

Objectives

1. Align two images of a scene along their epipolar line
2. Find matching points in a scene pair using feature matching algorithms
3. Derive depth information from image disparity

Stereometry

How far a scene point is from a camera can be estimated using two images of the same scene taken by the same camera. The second image should be taken at a short displacement from the original position of the camera. Alternatively, the scene can be captured by two identical cameras spaced slightly apart. This estimated distance from the camera is referred to as “range” or “depth” in 3D imaging.

Consider the diagram below:

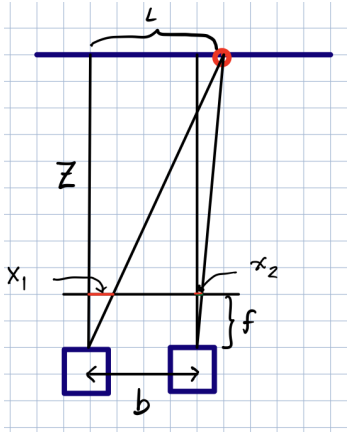

Figure 1. Stereometry setup

Two identical cameras with focal length f are placed a distance b apart. A scene point (red dot) is at a range Z from the cameras. If we use the Z line from the left camera as the origin of the x-axis, the scene point is at a distance L from the x-axis. In the image frame of each camera, the red dot is seen to be at a distance x1 for the left camera, and x2 from the right camera.

By similar triangles we see that $\frac{L}{Z} = \frac{x_1}{f}$. Equivalently, we see that $\frac{L-b}{Z} = \frac{x_2}{f}$. Substituting $L = \frac{Zx_1}{f}$ and solving for Z we find that $Z = \frac{bf}{x_1 - x_2}$.

The term (x1-x2) is known as disparity. The larger the disparity, the nearer the point, while the smaller the disparity the farther the point.

Two conditions must be met to reconstruct the scene in 3D:

1. Both cameras must have their y-axes aligned. If we will use only one camera, then the f is the same. We must move the camera through a measured distance b. If at the second position the camera was rotated, the second scene must be transformed to align with the first.
2. We must have estimates of Z for many points on the scene.

We can use functions in MATLAB's Computer Vision Toolbox to satisfy these conditions. Given matching points in both scenes, we can compute the Fundamental Matrix which relates the transformation of points in the first scene to the second. We can then use the estimated Fundamental Matrix to align both scenes.

Matching points can be automated using local feature detectors. In computer vision, there are existing feature match detectors such as SURF (speeded up robust features) or SIFT (scale-invariant feature transform).

Procedure

1. Take a picture of a colorful, textured object using your camera. **Remember to turn off autofocus to avoid changing f.**
2. Move your camera horizontally by 1 or 2cm and retake a picture of the object. **Make sure to measure the camera displacement accurately!**
3. Load the two images in Matlab, convert to gray.
4. Find matching points and estimate the Fundamental Matrix from these matching points.
5. Rectify and align both images using the Fundamental Matrix.
6. Compute the disparity from each point on the image.
7. Compute depth from the disparity map.

Please see stereotry.mlx for the sample cod

Due May 25, 2022