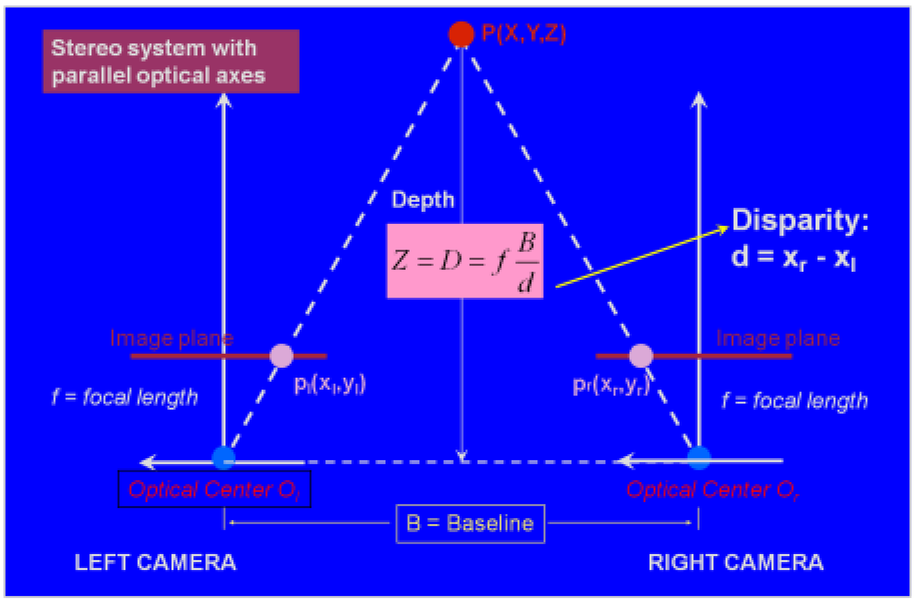
Assignment 2 - Stereo and Motion

Brandon Chin (ID: 2938)

Writing Assignments

1. Estimate the accuracy of the simple stereo system (Figure 3) assuming that the only source of noise is the localization of corresponding points in the two images. Discuss the dependence of the error in depth estimation as a function of disparity error, depth itself, the baseline width and the focal length, respectively.

Hint: Z = fB/d; Take the partial derivatives of Z with respect to d, B, f respectively.



The figure above shows a diagram of a simplified stereo vision system where both cameras are positioned perfectly parallel to each other and have the exact same focal length.

To determine the position of a point in a disparity image, we can use the equation

$$Z = fB/d$$

- Z = the depth, or distance along the camera z-axis
- f = the focal length
- B = the baseline width between the two cameras
- d = the disparity, or distance between two projected points

To determine the depth error, we can calculate the partial derivatives of Z with respect to the variable of uncertainty. Assuming the only uncertainty is in the disparity d, the depth error is equal to

$$\partial Z = (Z^2/fB)\ \partial d$$

If the uncertainty is in the baseline width, the depth error with respect to the baseline is equal to

$$\partial Z = (f/d)\ \partial B$$

Finally, if the uncertainty is in the focal length, the depth error with respect to the focal length is equal to

$$\partial Z = (B/d)\ \partial f$$

- Depth error is inversely proportional to the baseline width, meaning a larger baseline width will provide better depth accuracy but a smaller field of view (FOV)
- Depth error is inversely proportional to the focal length, meaning a larger focal length will provide a better depth accuracy but a smaller FOV
- Depth error is proportional to the square of the depth, meaning the nearer a point is, the greater the accuracy (*ie.* the smaller the uncertainty).

2. Could you obtain 3D information of a scene by viewing the scene by a camera rotating around its optical center? (Show why or why not) What about moving the camera along its optical axis?

The motion field equation under rotation around the camera's optical center can be written as

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}$$

Since Z is not included in the equation, no 3D information is carried. The disparity between any two images under this rotation is 0. In order to be able to extract any 3D information, some translational operation must be applied instead. If we move the camera along its optical axis however, 3D information can be obtained because translational images can be captured at different time frames.

3. Discuss at least 2 similarities and at least 2 differences between stereo vision and visual motion.

The similarities between stereo vision and visual motion are:

- that both use the translational difference between two or more images to extract information about depth
- that 3D information is derived similarly by measuring the disparity between points; points that are further away move smaller amounts than points that are closer

The differences between stereo vision and visual motion are:

- that a stereo vision system requires two cameras each capturing a slightly different image of the same scene, whereas a system of visual motion only requires one moving camera that is capturing multiple images of the same scene at different times.
- that regarding correspondence, the disparities between consecutive frames are much smaller due to dense temporal sampling

4. Give five examples where humans use stereo and motion, and Google another five examples that use machine vision algorithms of stereo and motion.

Five examples in which humans use stereo and motion are:

- driving a car
- designing a 3D model
- reaching out to touch/grab something
- pouring a drink
- stepping over an obstacle

Five example that use machine vision algorithms of stereo and motion are:

- Tracking the speed of a car
- Motion estimation / video compression (*e.g.* MPEG)
- robot navigation (*e.g.* UAVs, Mars rovers)
- Feature extraction from aerial surveys (*e.g.* contour maps, 3D building mapping)
- Facial recognition