Shirong Zheng

CSC47100-E

Prof. Zhu

11/17/2017

Assignment 4

1. (Stereo- 30 points) Estimate the accuracy of the simple stereo system (Z = f B/d, Figure 3 in the lecture notes of stereo vision) assuming that the only source of noise is the localization of corresponding points in the two images, which means the error in estimating d. Please derive (10 points) and discuss (20 points) the dependence of the error in depth estimation of a 3D point as a function of (1) the baseline width B, (2) the focal length f, (3) stereo matching error, and (4) the depth of the 3D point, Z.

Hint: Take the partial derivatives of Z with respect to d, assuming that both B and f and constant parameters.

$$Z(d) = f\frac{B}{d} \Rightarrow d = \frac{fB}{Z} \Rightarrow d^2 = \frac{(fB)^2}{Z^2}$$

$$\frac{\partial Z}{\partial d} = \frac{\partial (f\frac{B}{d})}{\partial d} = \frac{fB\,\partial(\frac{1}{d})}{\partial d}$$

$$= \frac{-\frac{fB}{\partial^2}\partial d}{\partial d} = \frac{-fB}{\partial^2}$$

$$|\partial Z| = \left| -\frac{fB}{d^2}\partial d \right|$$

$$= \frac{fB}{d^2}\partial d \leftarrow \text{function of distance}$$

$$= \frac{fB}{\frac{(fB)^2}{Z^2}}\partial d = \frac{Z^2}{fB}\partial d$$

(1) The baseline width B increase, then the error increase.
(2) The focal length with f increase, then the error decrease.
(3) The stereo matching error increase, then the error increase.
(4) The depth of the 3D point, Z, increases, then the error increase.

2. (Motion- 40 points) Could you obtain 3D information of a scene by viewing the scene by a camera rotating around its optical center (10 points)? Discuss why or why

not(10 points). What about translating the camera along the direction of its optical axis (10 points)? Explain. (10 points)

That the camera has to physically move and it can be rotated in order to gather 3D information. Another way of gathering 3D information is if the object is in motion. It could be said 3D Motion can be characterized by a rotation matrix and a translation matrix.

No, since the 3D motion field equation as:

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \underbrace{\frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}}_{\text{Rotation part: no depth information}} + \underbrace{\frac{1}{Z}\begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix}\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}}_{\text{Translation part: depth Z}}$$

While in the condition of pure rotation, which means Translation part =0, then

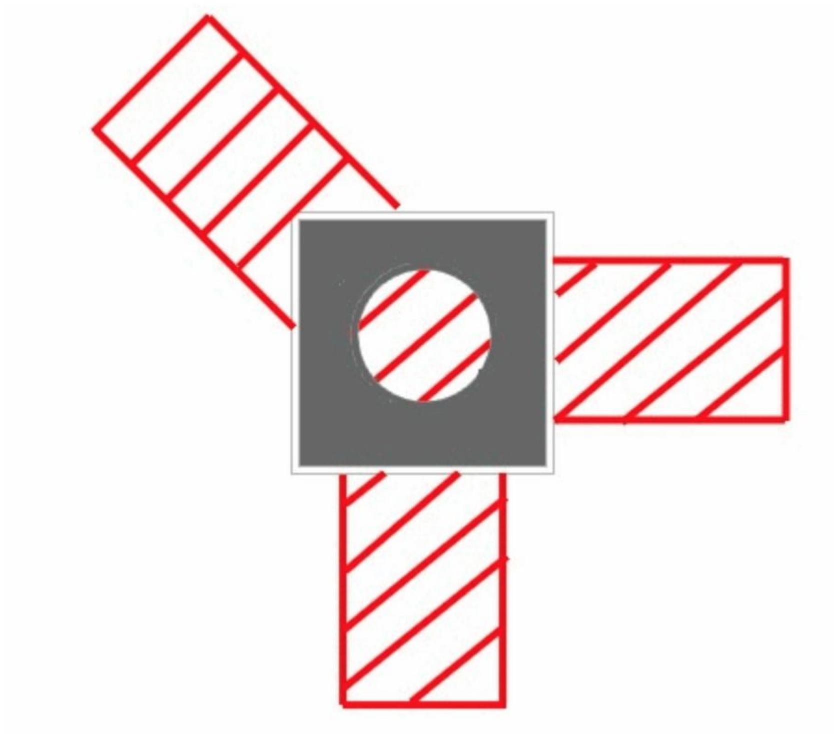
$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{f}\begin{pmatrix} xy & -(x^2+f^2) & fy \\ y^2+f^2 & -xy & -fx \end{pmatrix}\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}.$$

So, there is no depth information, which means there is no Z value inside the formula, and it has no 3D information. This is the equation of rotational part . As we can see, there is no depth information. This is because we do not have a Z variable anywhere here.

If we look at this from a practical perspective, we can see that if we were to take a camera and just rotate it, all we would get is a panoramic image. It would just be very wide image. There would be no depth information because we do not have 2 reference points to compare. For translation along the optical axis, there is no depth information. This is because you are simply changing the focal length if you move the camera forward or backwards. You need to have some distance in the X or Y direction in order to capture the 3D information.

We see that translating the image across the camera's optical axis does not give us useful information because it is just a zoomed in version of the original image. The right side image is the original. If we zoom in, which is the same as translating across the camera's optical axis, we get the image on the left. The left side image is almost same as the right side.

3. (Motion- 10 points) Explain that the aperture problem can be solved if a corner is visible through the aperture.



We can see that there is a circle in the middle. Through this circle, we see a slash, moving in the "lower right" direction. However, if you want to create this visual effect, there are three possibilities. The first possibility is that a "horizontal" note is drawn towards "right" but there is a "slash" on the note, so there is an illusion when looking through the aperture. The second possibility is that a "straight" note moves "directly below," but because there is a "slash" on the note, it still feels rightward and downward when viewed through the aperture. The third possibility is a piece of paper with a "straight line" on it, but moving "to the lower right" so that when viewed through the aperture, it looks the same as the two moving directions seen before.

Therefore, although the three strips of paper have different "moving directions" and even the "streaks" of the strips are different, the illusion that the streaks move in the same direction can be observed through the apertures. This is the difference between "local" and "global" visual processing. Our vision system locally may have the illusion of aperture problems, but when we look at the universe as a whole, we analyze the different directions of movement of the three strips of paper.

The aperture problem states that the motion of a spatial image cannot be determined if you see if through a small aperture. Because we cannot determine the direction the image is going. The motion of a 2D feature, such as a corner is very obvious. With a corner we can see exactly which direction it is going in. If we see a corner, then we are able to determine the direction it is going in. Therefore, solving the aperture problem requires us to find a corner in the image. It might appear to be going in a certain direction but we cannot be sure. For example, an edge may appear to be moving diagonally, but it will be moving up and down.

4. (Stereo Programming - 20 points + 20 bonus points) Use the image pair (Image 1, Image 2) for the following exercises.

(1). Fundamental Matrix (20 points). - Design and implement a program that, given a stereo pair, determines at least eight point matches, then recovers the fundamental matrix (10 points) and the location of the epipoles (5 points). Check the accuracy of the result by measuring the distance between the estimated epipolar lines and image points not used by the matrix estimation (5 points). Also, overlay the epipolar lines of control points and test points on one of the images (say Image 1- I already did this in the starting code below). Control points are the correspondences (matches) used in computing the fundamental matrix, and test points are those used to check the accuracy of the computation.

Hint: As a first step, you can pick up the matches of both the control points and the test points manually. You may use my matlab code (FmatGUI.m) as a starting point - where I provided an interface to pick up point matches by mouse clicks. The epipolar lines should be (almost) parallel in this stereo pair. If not, something is wrong either with your code or the point matches.

The epipolar geometry is the intrinsic projective geometry between two views. It is independent of scene structure, and only depends on the cameras' internal param- eters and relative pose. The fundamental matrix F encapsulates this intrinsic geometry.

For the fundamental matrix, I generated the A matrix and used SVD of A. Then I created the Fundamental Matrix using the A matrix. The Fundamental Matrix is the one shown below.

```
-1.2666714e-06  -2.9192004e-06   1.8504652e-03
  4.1636665e-06   1.2685225e-06  -2.9888410e-03
 -1.6975784e-03   7.1168573e-05   9.9999238e-01
```

The image of the epipolar lines