CS6320: 3D Computer Vision **Prateep Mukherjee(u0876583)**

Homework 2

1 Theoretical Problems

1.1 Problem 1

For 3 cameras, point-to-point correspondence are no longer uniquely defined. Let us define u_1, u_2 and u_3 be the 3 images generated using the cameras c_1, c_2 and c_3 respectively. For uncalibrated cameras, we get the following 3 epipolar equations:

$$u_1^T \mathcal{F} u_2 = 0 \tag{1}$$

$$u_2^T \mathcal{F} u_3 = 0 \tag{2}$$

$$u_3^T \mathcal{F} u_1 = 0 \tag{3}$$

Therefore, correspondence points for every point($\Re^2$) in u_1 are given by two equations $\mathcal{F}u_2 = 0$ and $u_3^T\mathcal{F} = 0$. This is an over-determined system of linear equations which yields incoherent results.

1.2 Problem 2

- For forward translating cameras, the epipoles e and e' have the same coordinates in both the images. Since, all epipolar lines pass through epipoles, therefore the epipolar lines radiate from each epipole. Therefore, correspondence points move along these radiating lines from e. This is why we see a radially oriented intersecting planes.
- Given the disparity of the two images, depth-map $\mathbf{Z}$ can be computed by $\mathbf{Z} = -\frac{f\mathbf{B}}{d}$, where f is the focal length of the camera, $\mathcal{B}$ is the baseline shift and d is the disparity at a pixel x. For forward translating cameras, point $\mathbf{x}'$ on one camera maps to $\mathbf{x}$ on the other, such that

$$\mathbf{x}' = \mathbf{x} + \frac{K\mathbf{t}}{z}$$

where K is the calibration matrix with constant intrinsic and extrinsic parameters. Disparity $d = \mathbf{x}' - \mathbf{x}$. Therefore, the depth of world points can be computed as:

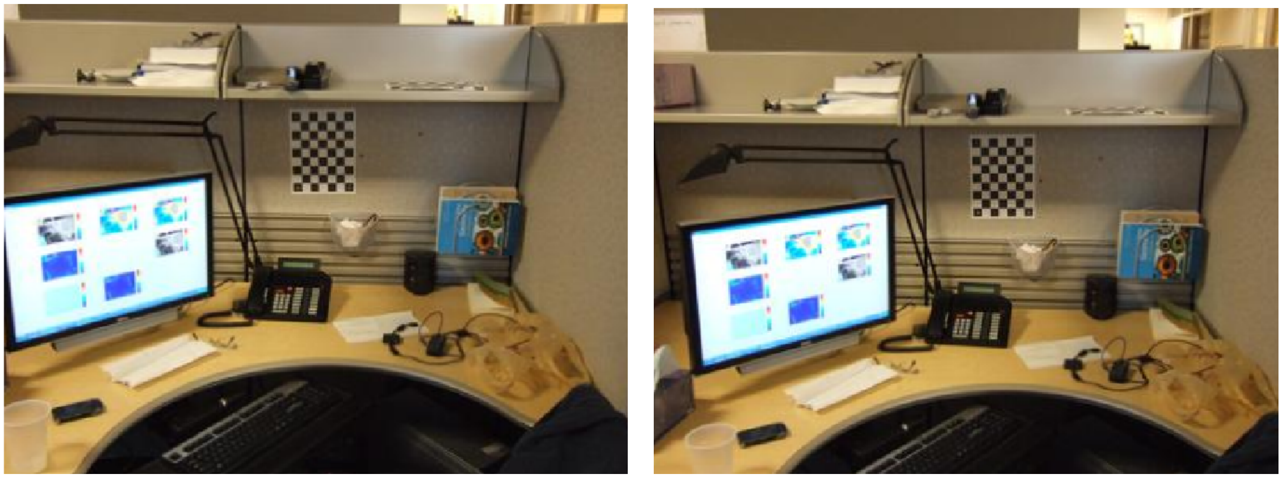
$$\mathbf{Z} = -\frac{f\mathbf{t}}{d} = -\frac{f\mathbf{t}}{\frac{K\mathbf{t}}{z}} = -C.z$$

2 Practical Problems

2.1 Problem 4a

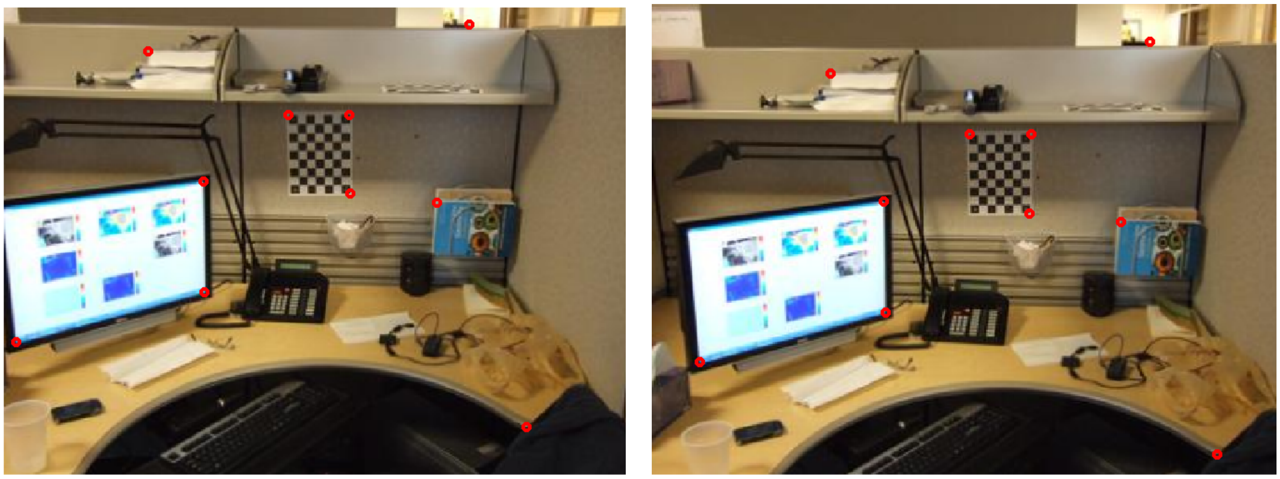
Fig 1 shows the 2 rectified stereo images used for my experiment. They are captured with two positions of the camera. Firstly an image(Fig. 1a) is taken. Then, I translate for some distance to the left to take the second image(Fig. 1b).

Next, I generate matching points on both the images. These points were all picked manually. Fig. 2 shows the colored points on both images.
The fundamental matrix $\mathcal{F}$ is built from these points. Since $\mathcal{F}$ has only 8 free parameters(setting $\mathcal{F}_{33} = 1$), we need only 8 of the above points. A random sampling is done to pick 8 points in this



(a) (b)

Figure 1: Image of a desk space captured from two positions of the camera. **(a)** Right view; **(b)** Left view.



(a) (b)

Figure 2: Images in Fig. 1 overlayed with matching points(in red). A total of 10 points were selected.

assignment. A cross-validation study amongst all picked points to pick the "best" 8 points can be done, which would give the best epipolar lines. The remaining points on one image are used to compute the epipolar lines on the other. Fig. 3 shows the different experiments with the remaining points on left and right images.

Fig. 3 shows that the epipolar lines do pass through corresponding points.

In order to compute the epipole of the left camera, eigen vector of $\mathcal{F}$ is considered which lies in its null-space. Therefore, we seek u such that $\mathcal{F}u = \bar{0}$. This is done by considering the eigen decompostion for $\mathcal{F}' \times \mathcal{F}$. Then, I taken the eigen vector with the smallest eigen value, and normalize it to get the pixel coordinates of the epipole in homogenous coordinates. The epipole for the left image is computed at $(250.47, 117.22)$. Fig. 4 shows the epipole on the left image.

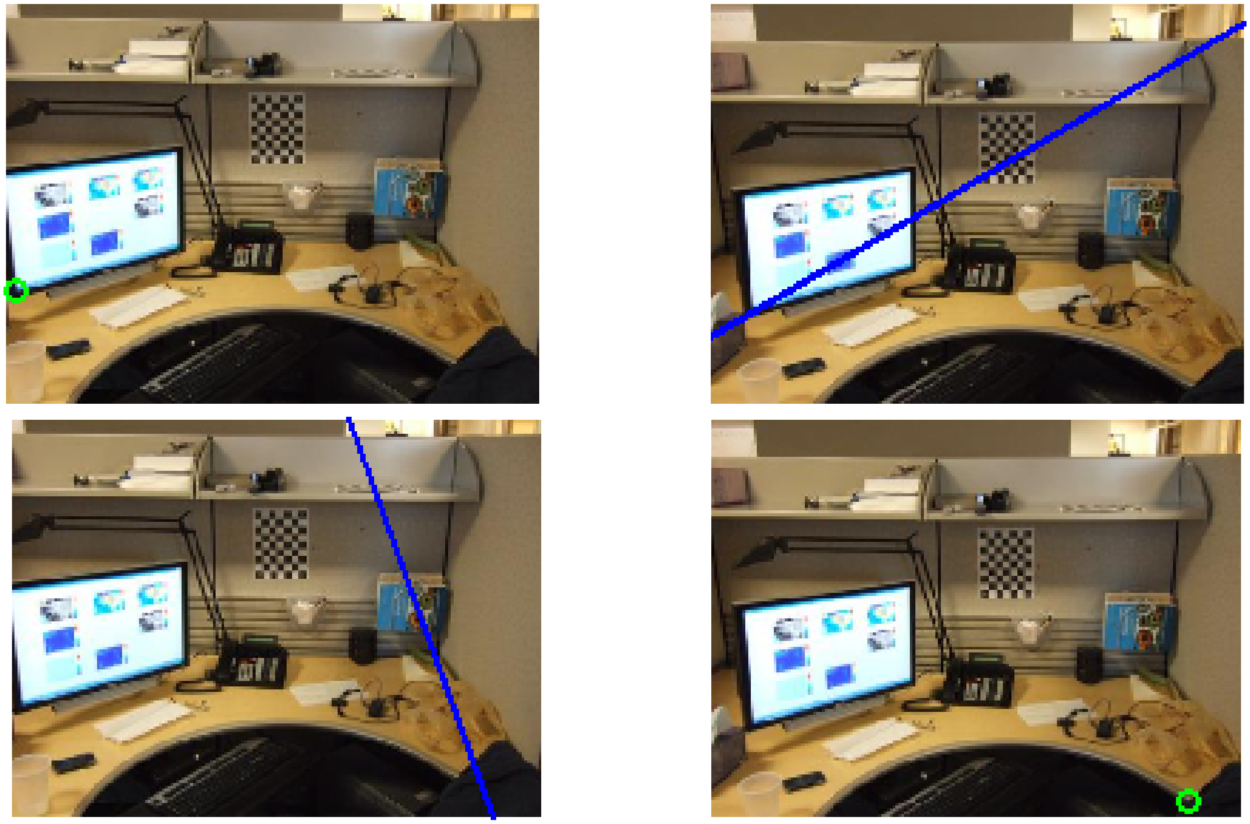


Figure 3: The image pair shown with points(in green) and epipolar line(in blue) . **Top** Epipolar line drawn on right image for corresponding point on left. **Bottom** Epipolar line drawn on left image for corresponding point on right.

2.2 Problem 4b

I take the classical tsukuba image from middlebury dataset for my disparity map experiment. Fig. 5 shows the pair of images used. The calibration details of the camera is provided at here. The original images(2880×1988) are scaled down to 384×288 resolution for ease of computation.

As input parameter to my code, I get user-specified values for *window_size*, which is used as region of interest for computing normalized cross-correlation(NCC). Thereafter, I shift the left image and compute NCC for every window patch. The minimum value of NCC is chosen as the value for every pixel. I do this for the other image as well.

Finally, in order to get a good disparity image, I adopt a **winner takes all** strategy. For every pixel, I compute the minimum disparity from left and right images, and label each pixel with this disparity. Fig. 6 shows the disparity maps for left and right images.

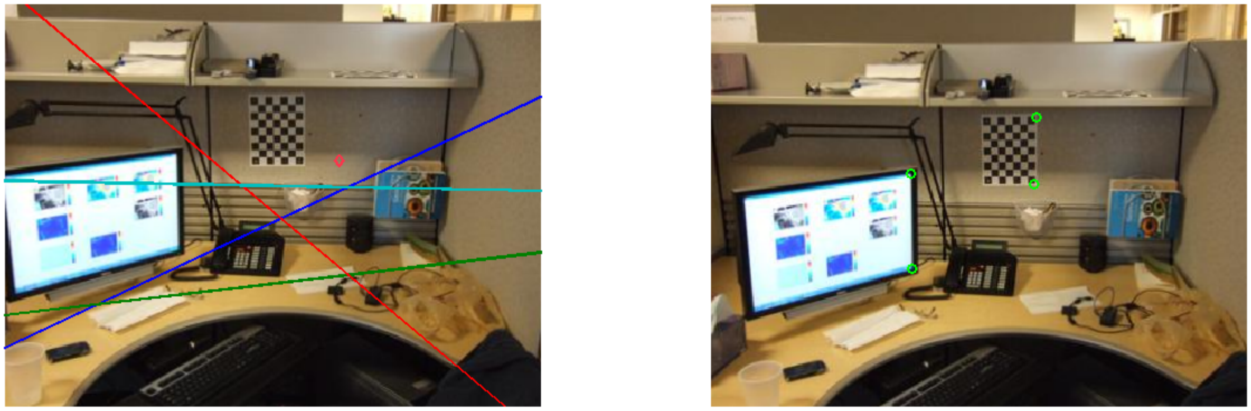
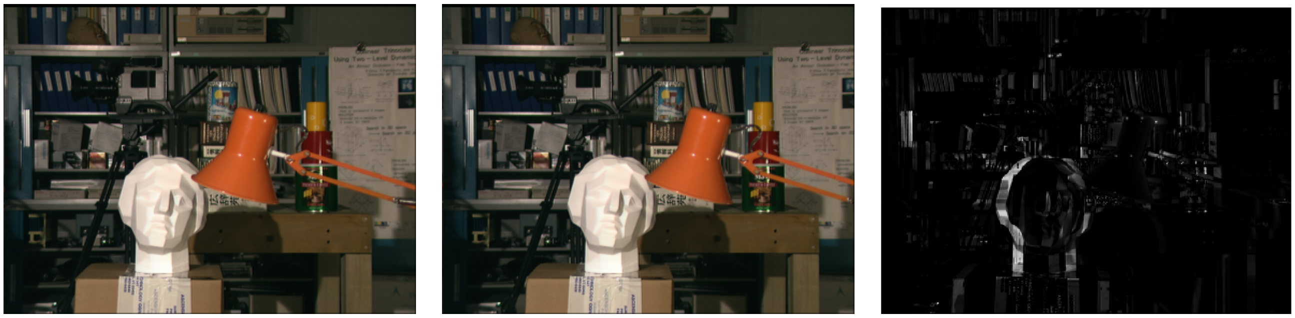
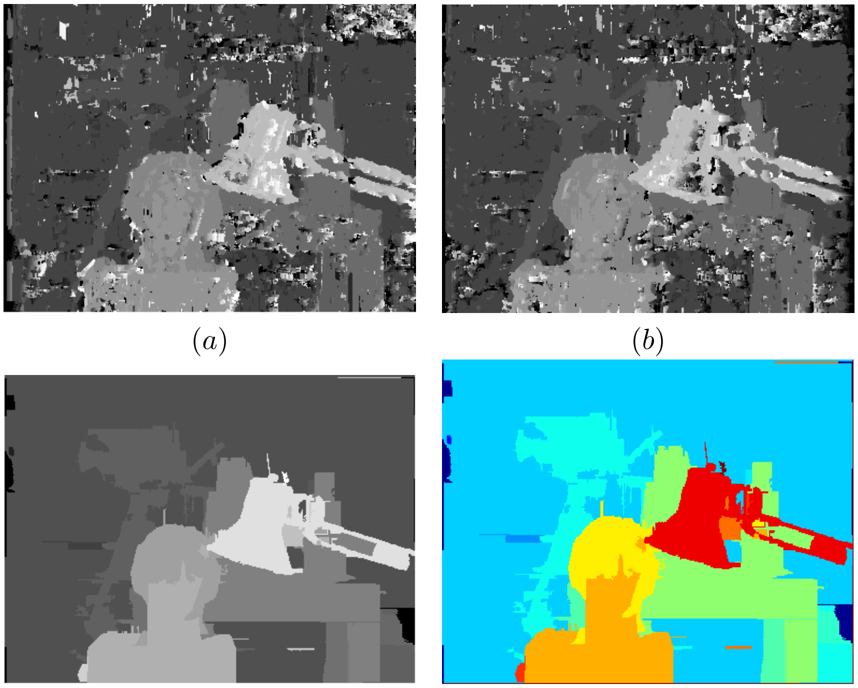


Figure 4: Epipole on the left images shown as diamond.



(*a*) (*b*) (*c*)

Figure 5: (a,b) : **Tsukuba** stereo images from middlebury vision dataset; (c) Difference image where bright regions show the shift.



(a)

(b)

(c)

(d)

Figure 6: Disparity maps for **(a)** left and **(b)** right images. **(c)** Final disparity using the **winner takes all** strategy. **(d)** Disparity map in rgb.

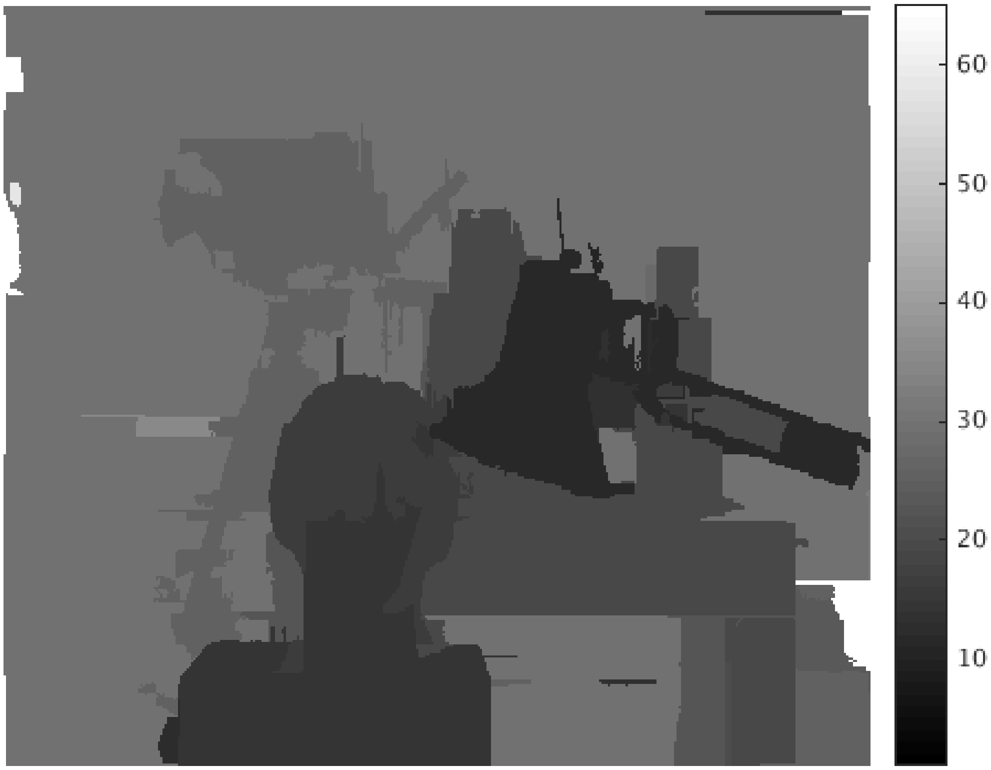


Figure 7: Z-depth shown with baseline shift = 176.252. Lighter regions represent far-away.