Computer Vision 2018 Fall Homework 4

B04507009 電機四 何吉瑞

Part1

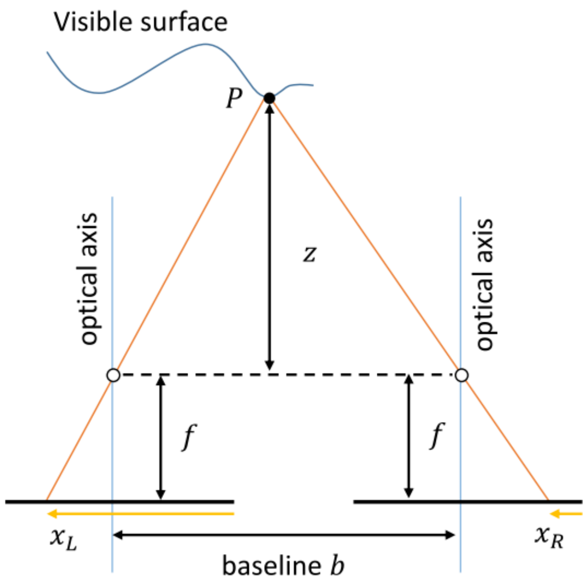


Fig. 1

(For simplicity, we say x_l and x_r are the endpoint of x_L and x_R)

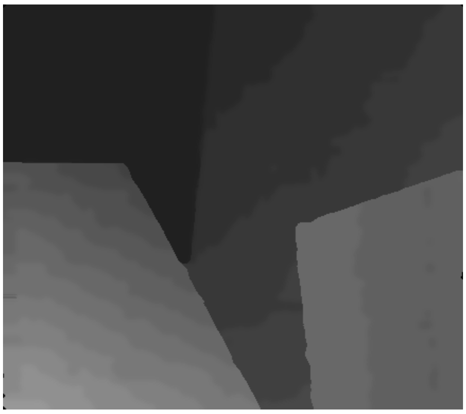
From Fig. 1, we mark the intersection between the optical axis and line segment $\overline{Px_l}$ as L. Similarly, we mark the intersection between the optical axis and line segment $\overline{Px_r}$ as R. From L, we plot a straight line parallel to $\overline{Rx_r}$ and intersects with the x_L on point Q. Since the relative position between L and the left image plane is equivalent to that between R and the right image plane, we can derive that the position of Q on the left image plane is equivalent to the position of x_r on the right image plane (recall that $\overline{LQ}$ and $\overline{Rx_r}$ are parallel to each other). That is, $\overline{x_L Q} = x_L - x_R$.

We also note that ΔPLR and $\Delta \mathrm{P}x_l Q$ are similar triangles. Therefore, we can derive that $\frac{\overline{x_L Q}}{f} = \frac{b}{z}$. By substitution, we have $x_L - x_R = \frac{f \cdot b}{z}$ and finish the proof.

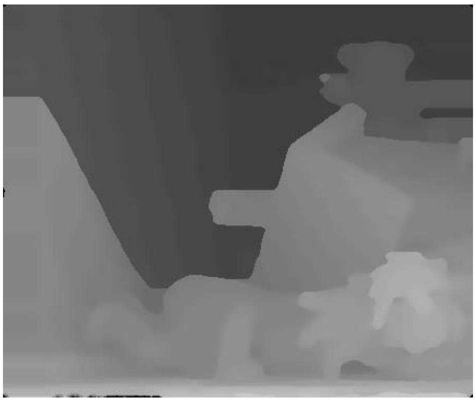
Part2

Results:

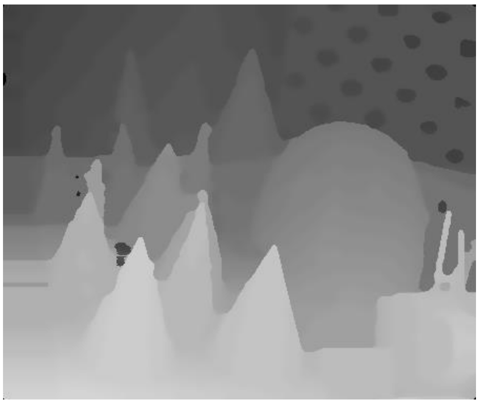
Bad Pixel Ratio: 3.56%



Bad Pixel Ratio: 0.48%



Bad Pixel Ratio: 10.10%



Bad Pixel Ratio: 7.93%
Average Bad Pixel Ratio: 5.52%

Methods:

Matching Cost:
I use Census cost as the based approach. For every pixel in the left image, I crop a window centered at it. The window size is (27, 27). Likewise, for each pixels in the right image, I crop the corresponding window. Then, I randomly select 4096 pairs of coordinates $(y_{j1}, x_{j1}, y_{j2}, x_{j2})$, where j ranges from 1 to 4096. For each window w_i, a 4096-bit string $r(w_i)$ is generated:

$$r(w_i) = \left[w_i\left(y_{j1}, x_{j1}\right) > \ w_i\left(y_{j2}, x_{j2}\right)\right] \ \ 0 < j \leq 4096$$

Given two windows w_i and w_k, we can estimate the matching cost between them by calculating the Hamming distance between $r(w_i)$ and $r(w_k)$.

Cost Aggregation:
I tried the method proposed by Zhang et al [1], but it didn't improve the performance. It may because I adopt different methods in other stages. So I skip this part in my algorithm.

Cost Optimization:
I used scanline optimization, which alleviates the matching ambiguities by adopting an optimizer with smoothness constraints and moderate parallelism. The scanline optimization includes four independent processes. Each process is performed in a distinct direction. Given a scanline direction $\mathbf{r}$, we can modify the cost $\mathrm{C}(\mathbf{p}, \mathrm{d})$ at pixel $\mathbf{p}$ and disparity d by the following equation:

$$\mathrm{C}(\mathbf{p}, \mathrm{d}) \ \leftarrow \ \mathrm{C}(\mathbf{p}, \mathrm{d}) + \min\left(\mathrm{C}_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, \mathrm{d})\right), \mathrm{C}_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, \mathrm{d} \mp 1) + P_1,$$

$$\min_k \mathrm{C}_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, \mathrm{k}) + P_2, \min_k \mathrm{C}_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, \mathrm{k}),$$

where P_1 and P_2 are parameters that penalizes the disparity changes of neighboring pixels. The resulting cost volume is the average of the optimized cost volume in four scanline directions.

Cost Refinement:
I used validity check [3] in this stage. We first generate a validity map to see the

erroneous pixels in disparity map. We first calculate the disparity for both left and right image. According to the calculated left disparity map, I warp the left disparity map to the right disparity map. If the difference between the warped left pixel and corresponding right pixel is larger than 1, we say that pixel is not valid. The non-valid pixel usually occur on the edge or in the presence of occlusion. For every pixel that is non-valid in the disparity map, we replace it with a neighboring valid pixel. Finally, I apply the median filter to the disparity map.

Reference:

[1] R. Zabih and J. Woodfill, "Non-Parametric Local Transforms for Computing Visual Correspondance," *Proc. European Conf. Computer Vision*, pp. 151-158, May 1994.

[2] Zhang et al, "Cross-based local stereo matching using orthogonal integral images," *CSVT* 2009.

[3] H. Hirschmu¨ller, "Stereo Vision Based Mapping and Immediate Virtual Walkthroughs," PhD dissertation, School of Computing, De Montfort Univ., Leicester, U.K., June 2003.

[4] X. Mei, X. Sun, M. Zhou, H. Wang, X. Zhang, "On building an accurate stereo matching system on graphics hardware," IEEE *International Conference on Computer Vision Workshops*, pages 467–474, 2011.