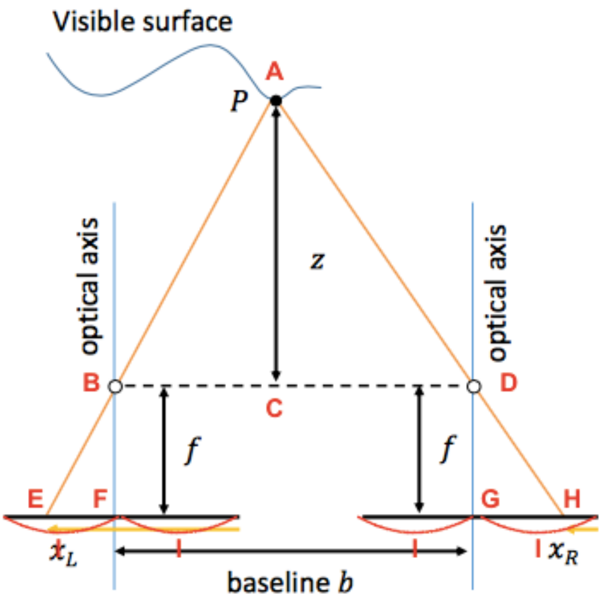
Computer Vision Homework 4

B05902115 資工四 陳建丞

Part 1: Depth from Disparity



ΔABC & ΔBEF are similar triangles

ΔACD & ΔDGH are similar triangles

Therefore,

$$
\begin{aligned}
\text{Disparity } d &= x_L - x_R \\
&= \text{EF} + \text{l} - (\text{l - GH}) \\
&= \text{EF} + \text{GH} \\
&= \text{BF}(\frac{\text{EF}}{\text{BF}} + \frac{\text{GH}}{\text{BF}}) \\
&= \text{BF}(\frac{\text{EF}}{\text{BF}} + \frac{\text{GH}}{\text{DG}}) \\
&= \text{BF}(\frac{\text{BC}}{\text{AC}} + \frac{\text{CD}}{\text{AC}}) \\
&= \frac{(\text{BF}\cdot\text{BD})}{\text{AC}} \\
&= \frac{\text{f} \cdot \text{b}}{\text{z}}
\end{aligned}
$$

Part 2: Disparity Estimation

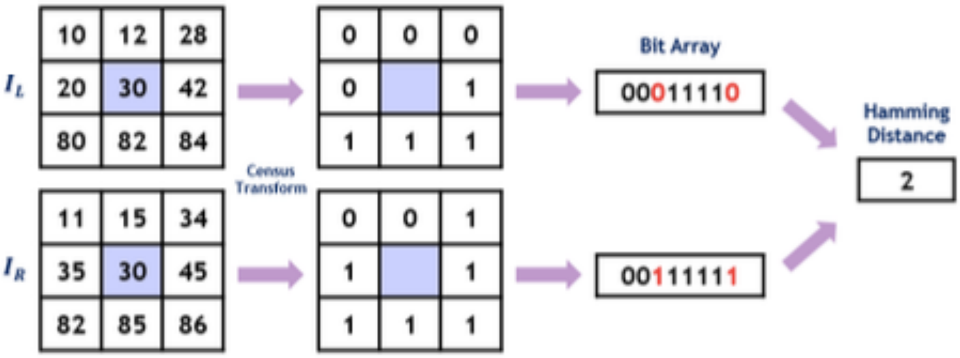
- **Implementation**
  - **cost computation**

    - Census Transform -

Census transform uses the relative intensity of input images, it is robust to intensity variations of input images. Census transform can be defined as

$$\xi(p, p') = \begin{cases} 0 & \text{if } p > p' \\ 1 & \text{otherwise} \end{cases}$$

Calculate pixel p and its 8-connected neighbors p' and get a one-dimension bit-array. The Census transform should be performed on p in left image and its corresponding pixel in the right image.



Two pixels of the Census transformed images are compared for the similarity using the Hamming distance, which is the number of bits that differ in the two bit strings.

Jaeryun Ko and Yo-Sung Ho. Stereo Matching using Census Transform of Adaptive Window Sizes with Gradient Image. IEEE 2017.

- **cost aggregation**

  - Bilateral filter -

  The cost is greatly dependent on the color intensity, some subtle change might lead to great difference. Thus, bilateral filter can smooth these small regions.

- **dispartiy optimization**

  - Winner takes all -

  For each pixel, select the disparity with minimun cost. This is a simple but effective method.

- **dispartiy refinement**

  - Left-right consistency check -

  According to the disparity values $d_{i,j}$ of pixel $p_{i,j}$ in left image, check the disparity $d'_{i-d,j-d}$ of pixel $p'_{i-d,j-d}$ in right image. Mark the pixels as following.

  $$\begin{cases} \text{Hole} & \text{if } d_{i,j} \neq d'_{i-d,j-d} \\ \text{Consisent} & \text{if } d_{i,j} = d'_{i-d,j-d} \end{cases}$$

  - Hole filling -

If a pixel is a hole after left-right consistency check, we might need to mend it. The method I use is simply find the closest none-hole left and right neighbor pixels. Use the smaller one as the disparity value of the hole.

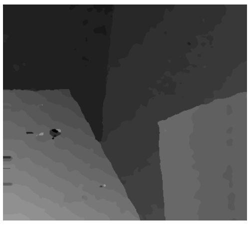
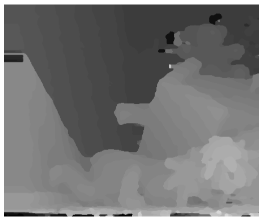
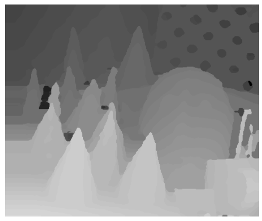
- Weighted median filter -

Implemented with a 7x7 mask

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 3 & 2 & 1 & 0 \\ 1 & 2 & 3 & 4 & 3 & 2 & 1 \\ 2 & 3 & 4 & 5 & 4 & 3 & 2 \\ 1 & 2 & 3 & 4 & 3 & 2 & 1 \\ 0 & 1 & 2 & 3 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The value of the mask is decided by the vertical distance + horizontal distance from the center. Weighted median filter performs better than original median filter because it considers the distance between neighbor pixels and center pixels. A closer pixel might have greater impact in the process.

- **Result**

Tsukuba	Venus	Teddy	Cones
			

	Tsukuba	Venus	Teddy	Cones	
Bad Pixel Ratio	6.84%	1.94%	11.12%	8.49%	7.10% (avg)
Elapsed time (sec)	35.351	63.234	139.708	139.952	378.256 (total)

- **Reference**

1. *Jaeryun Ko and Yo-Sung Ho. Stereo Matching using Census Transform of Adaptive Window Sizes with Gradient Image. IEEE 2017.*