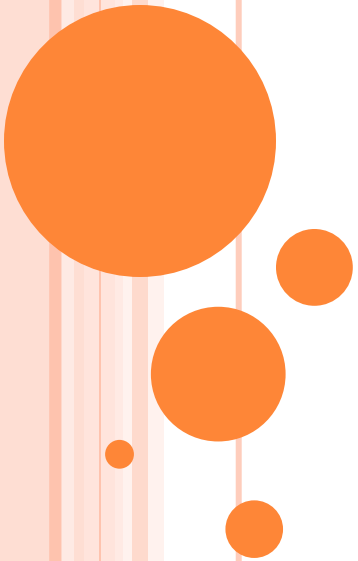
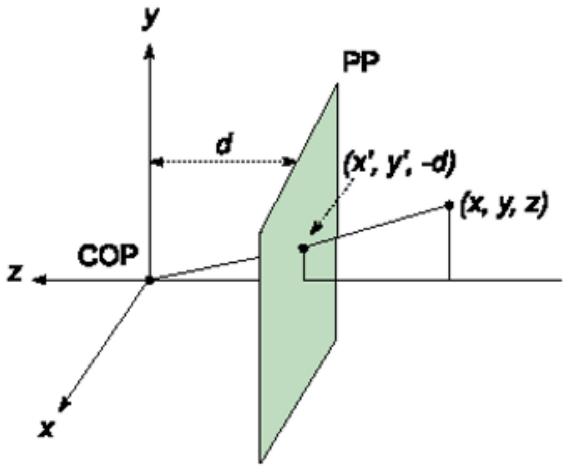


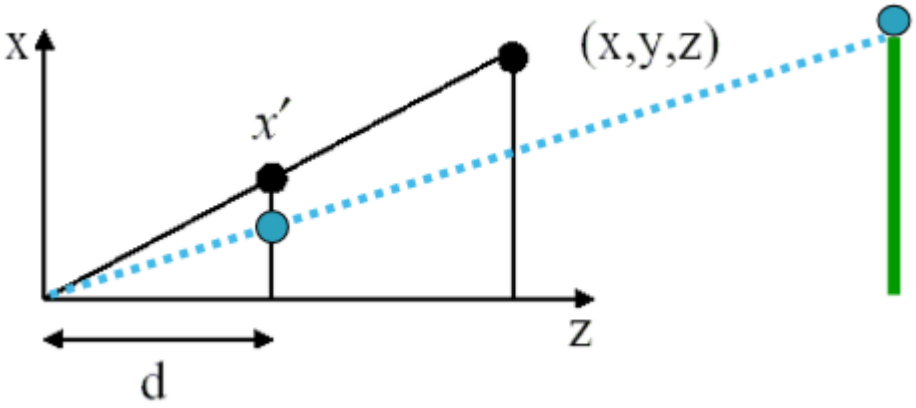
HW 2 (Due :04/24)

Stereo Depth Estimation

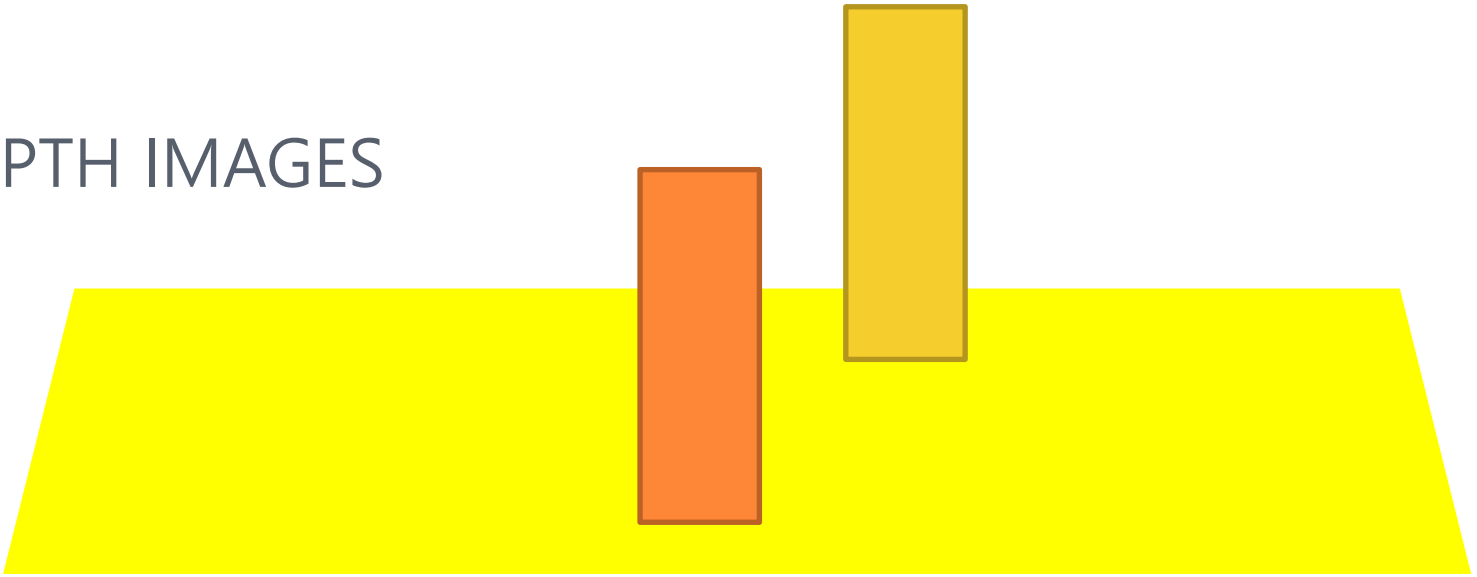
Recall: Prespective Pro

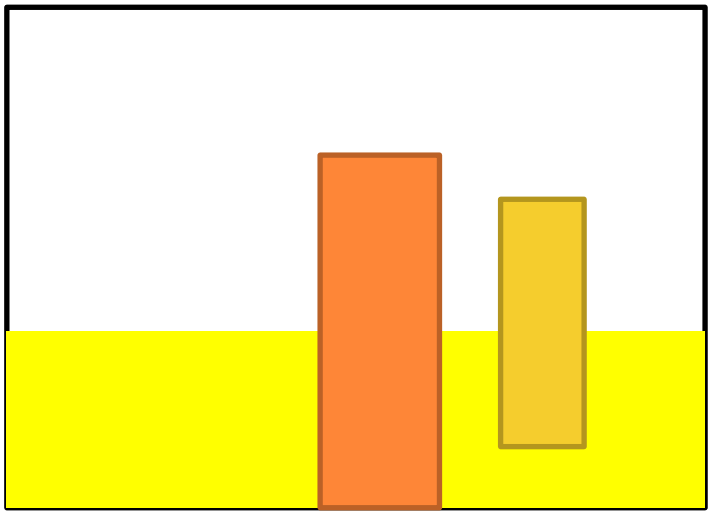


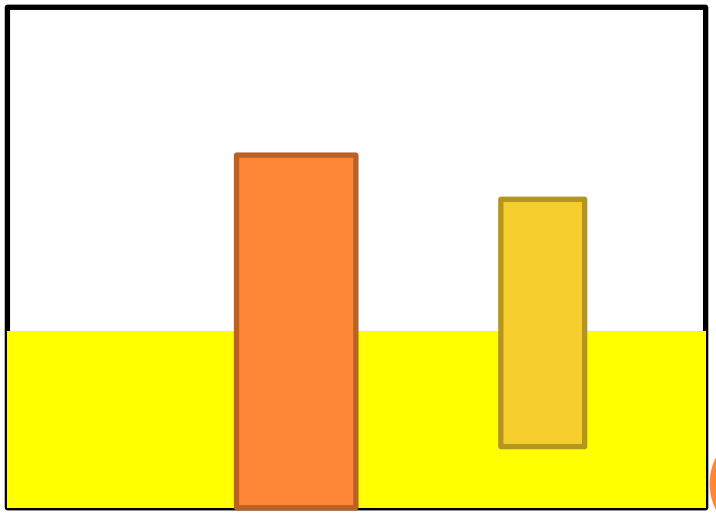
http://commons.wikimedia.org/wiki/File:Taiwan_HighSpeedRail_Train_Business_Class_Car.JPG

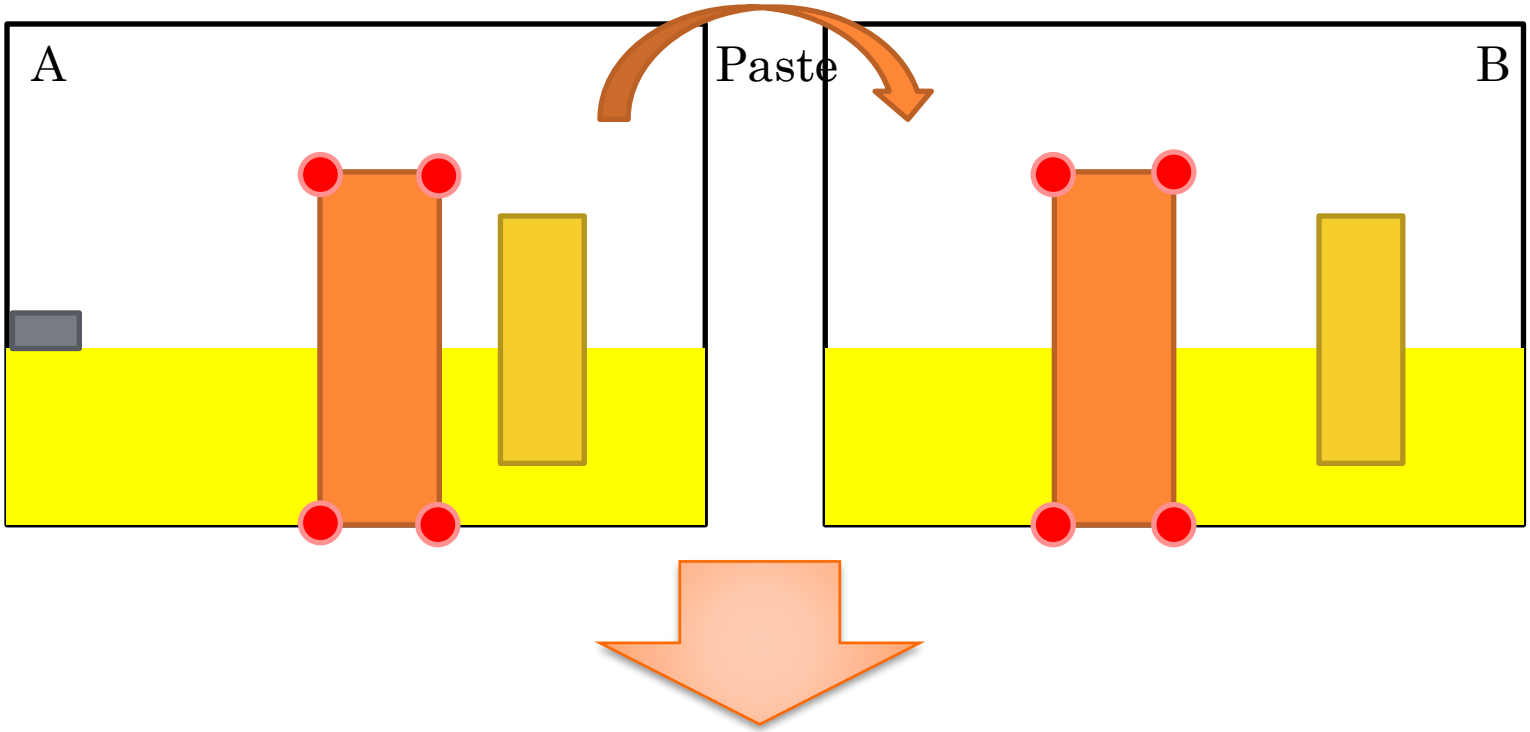


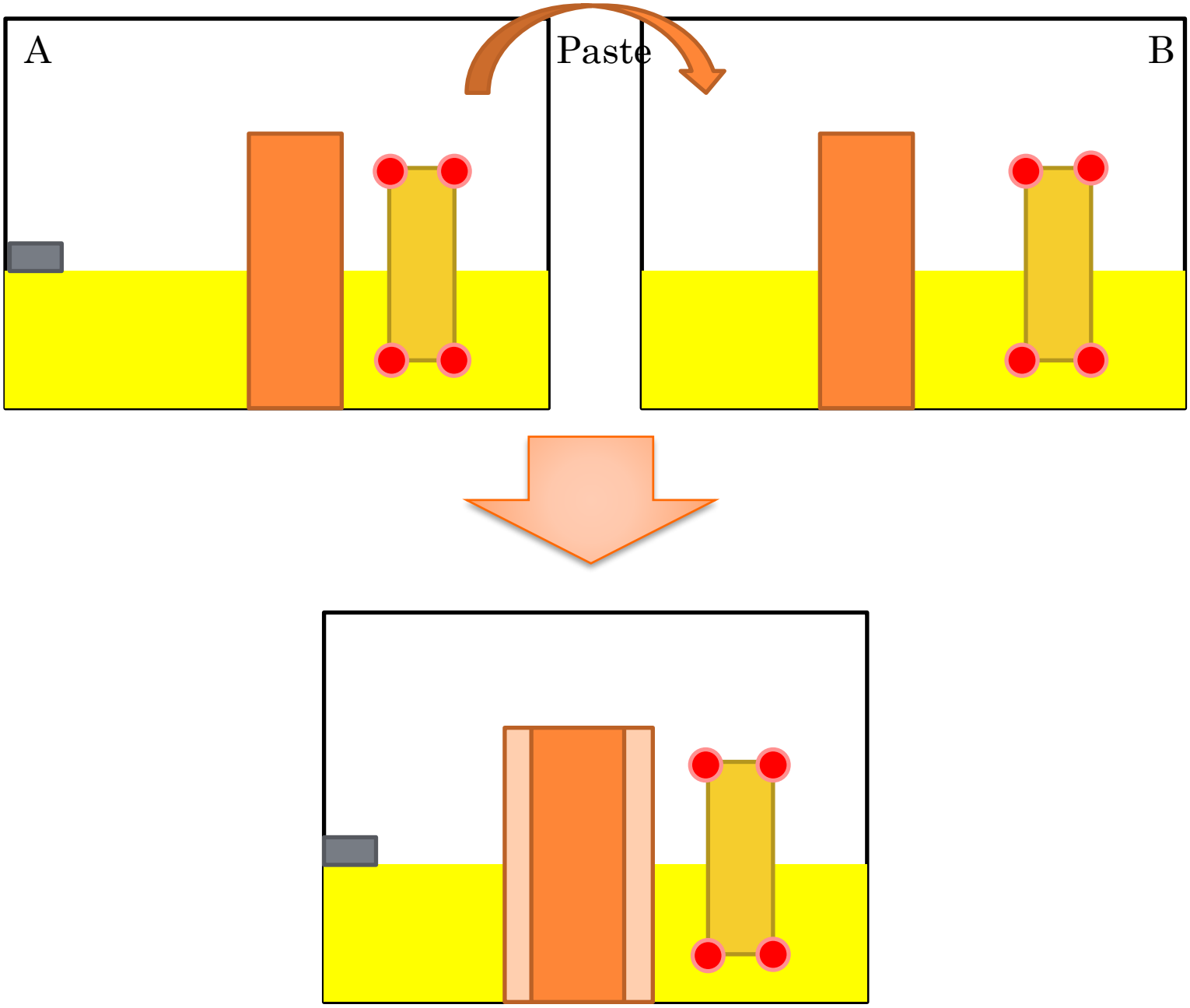
Depth images

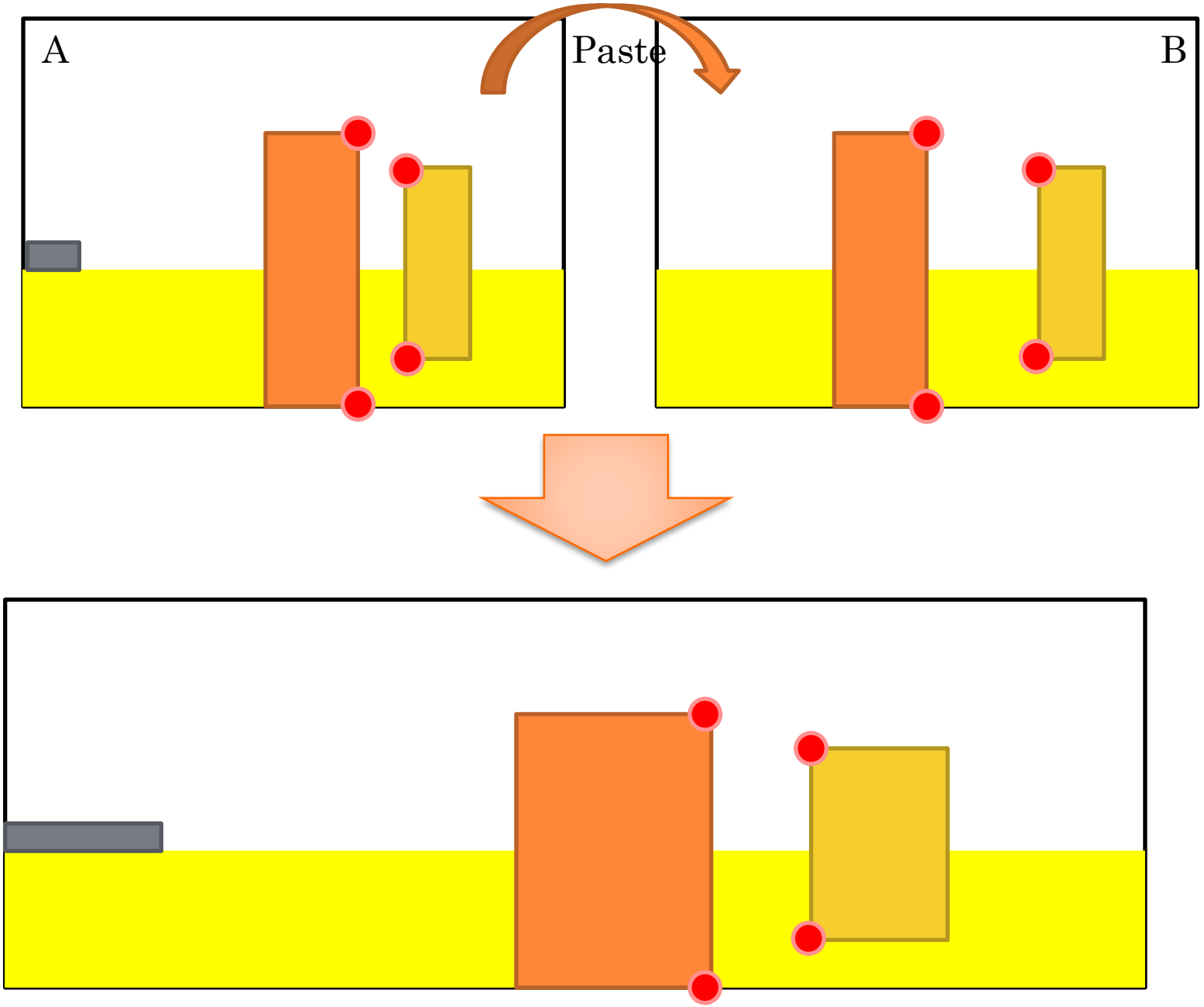







A
Paste
B


A
Paste
B


A
Paste
B

WHY?

- Image transform works on 2D image
  - Assume source/target are plane images

- For depth images, the position of point projected onto image plane depending on it’ s depth.
- Recall:

$$\begin{bmatrix} uq \\ vq \\ q \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

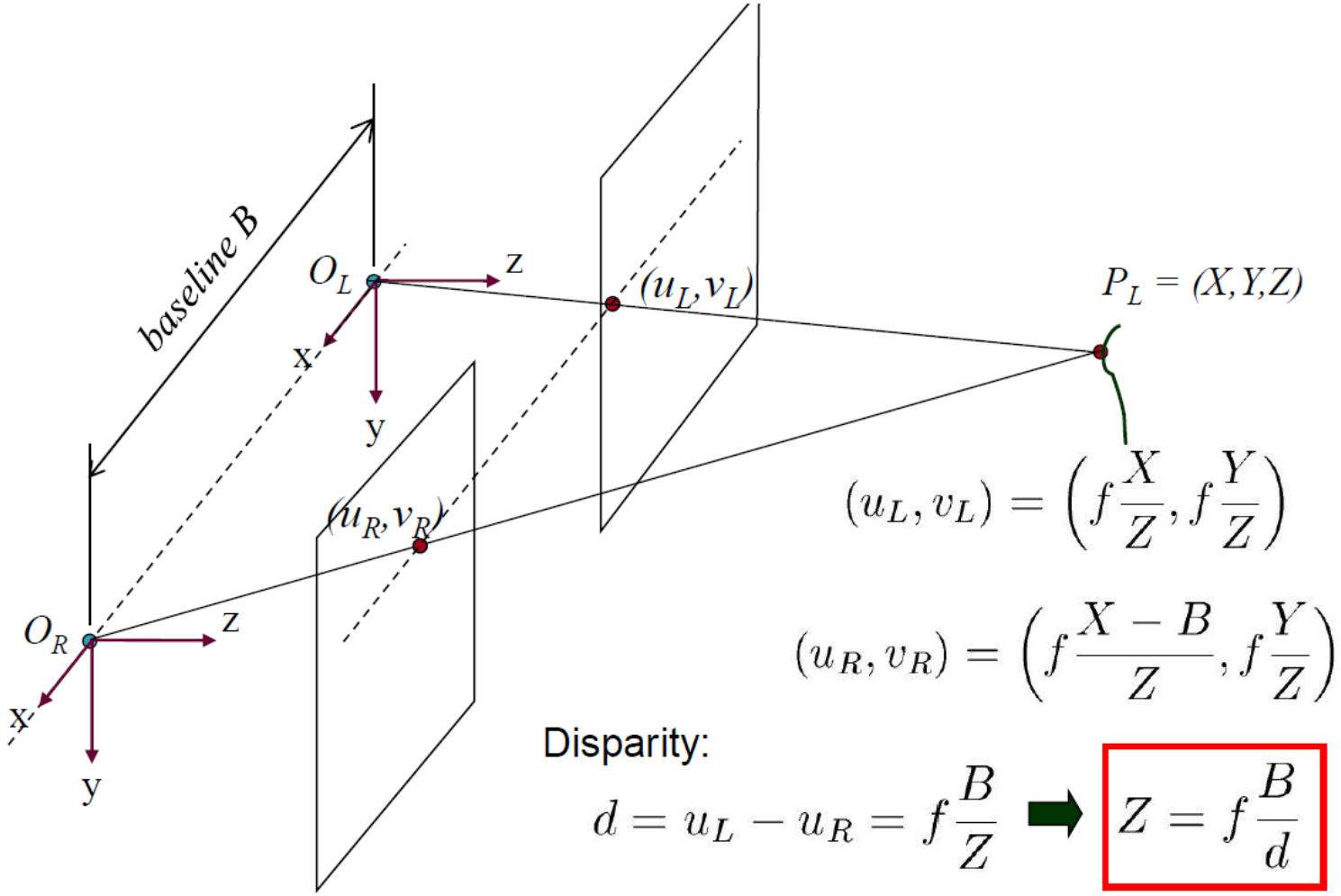
$$u = uq/q \;,\; v = vq/q$$

- 反過來說，我們可以利用p點在左右影像中的視差回推深度值!

Simplest case: rectified images

- Image planes of cameras are parallel.
- Focal points are at same height.
- Focal lengths are the same.
- Then, epipolar lines fall along the horizontal scan lines of the images

Simplest case: rectified images



$$(u_L, v_L) = \left(f\frac{X}{Z}, f\frac{Y}{Z}\right)$$

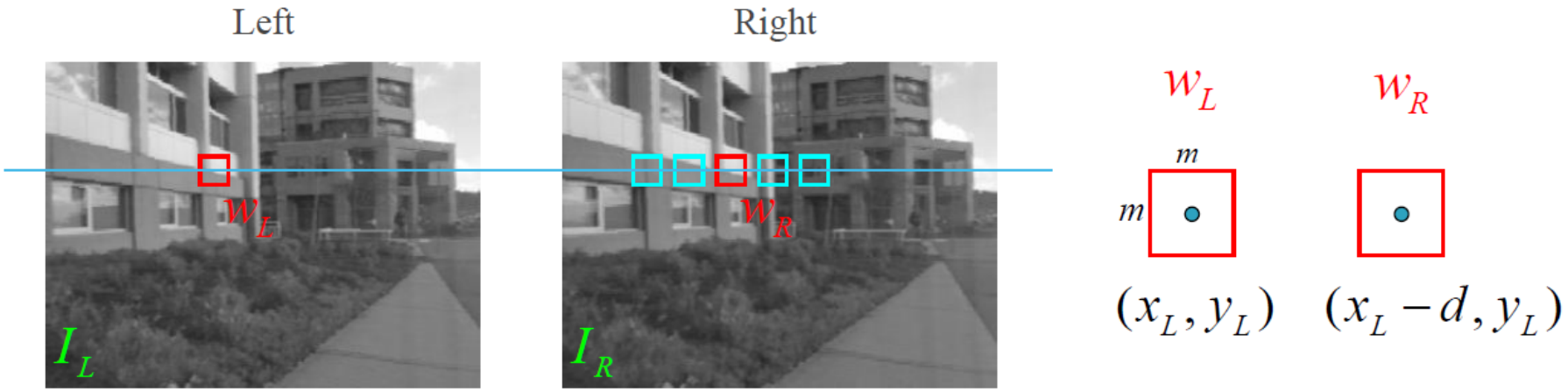
$$(u_R, v_R) = \left(f\frac{X-B}{Z}, f\frac{Y}{Z}\right)$$

Disparity:

$$d = u_L - u_R = f\frac{B}{Z} \Rightarrow \boxed{Z = f\frac{B}{d}}$$

Correspondence by correlation

SUM OF SQUARED DIFFERENCES



w_L and w_R are corresponding m by m windows of pixels.

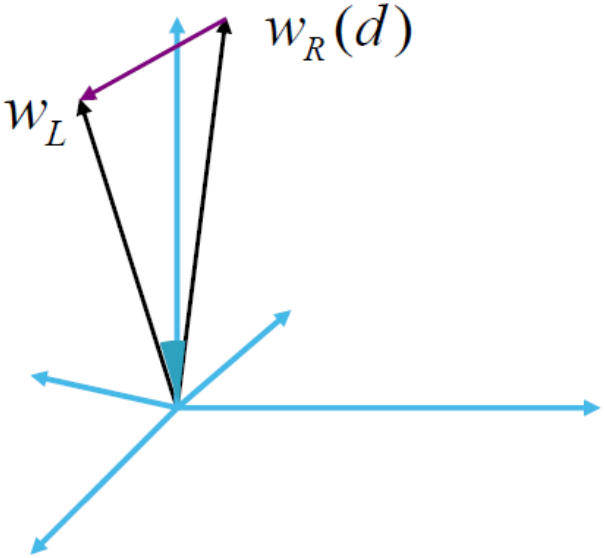
We define the window function :

$$W_m(x,y) = \{u,v \mid x - \tfrac{m}{2} \le u \le x + \tfrac{m}{2}, y - \tfrac{m}{2} \le v \le y + \tfrac{m}{2}\}$$

The SSD cost measures the intensity difference as a function of disparity :

$$C_r(x,y,d) = \sum_{(u,v)\in W_m(x,y)} [I_L(u,v) - I_R(u-d,v)]^2$$

Image Metrics



(Normalized) Sum of Squared Differences

$$C_{\text{SSD}}(d) = \sum_{(u,v)\in W_m(x,y)} [\hat{I}_L(u,v) - \hat{I}_R(u-d,v)]^2$$

$$= \left\| w_L - w_R(d) \right\|^2$$

Normalized Correlation

$$C_{\text{NC}}(d) = \sum_{(u,v)\in W_m(x,y)} \hat{I}_L(u,v)\hat{I}_R(u-d,v)$$

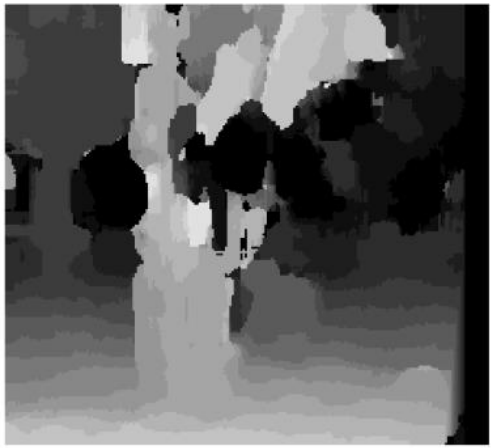
$$= w_L \cdot w_R(d) = \cos\theta$$

$$d^* = \arg\min_d \left\| w_L - w_R(d) \right\|^2 = \arg\max_d w_L \cdot w_R(d)$$

Window size

- Effect of window size.
- Some approaches have been developed to use an adaptive window size (try multiple sizes and select best match)

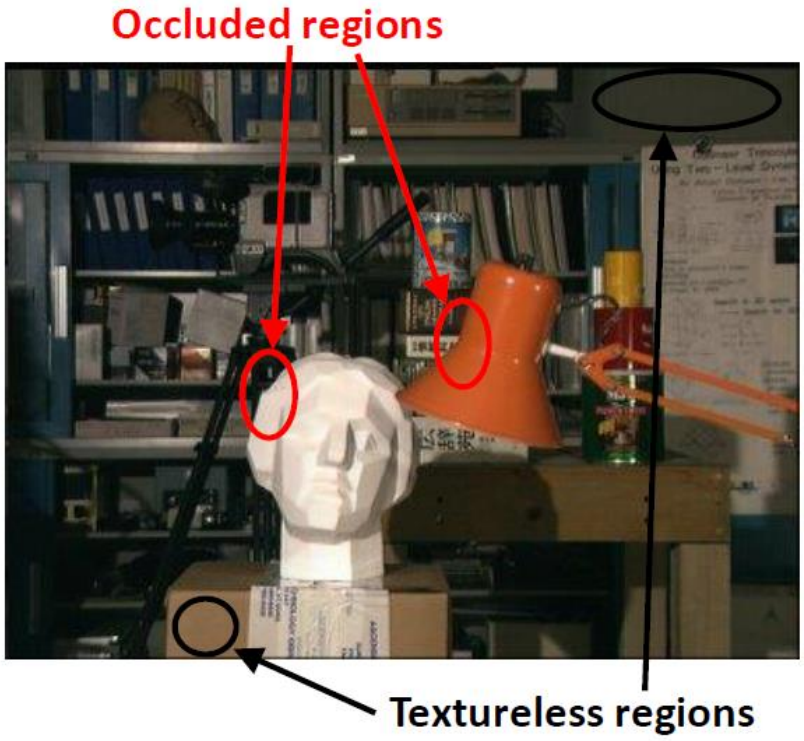
W = 3



W = 20

TWO MAJOR ROADBLOCKS

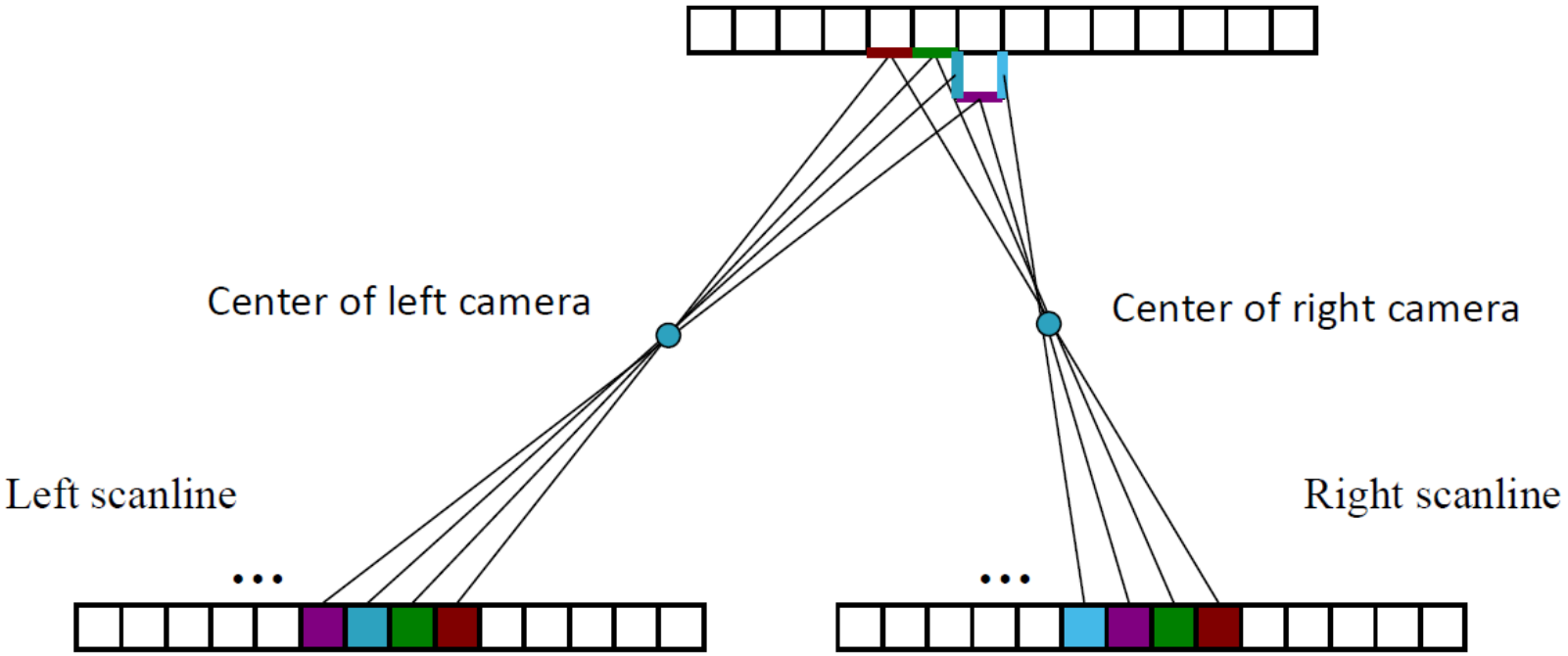
- Textureless regions create ambiguities
- Occlusions result in missing data



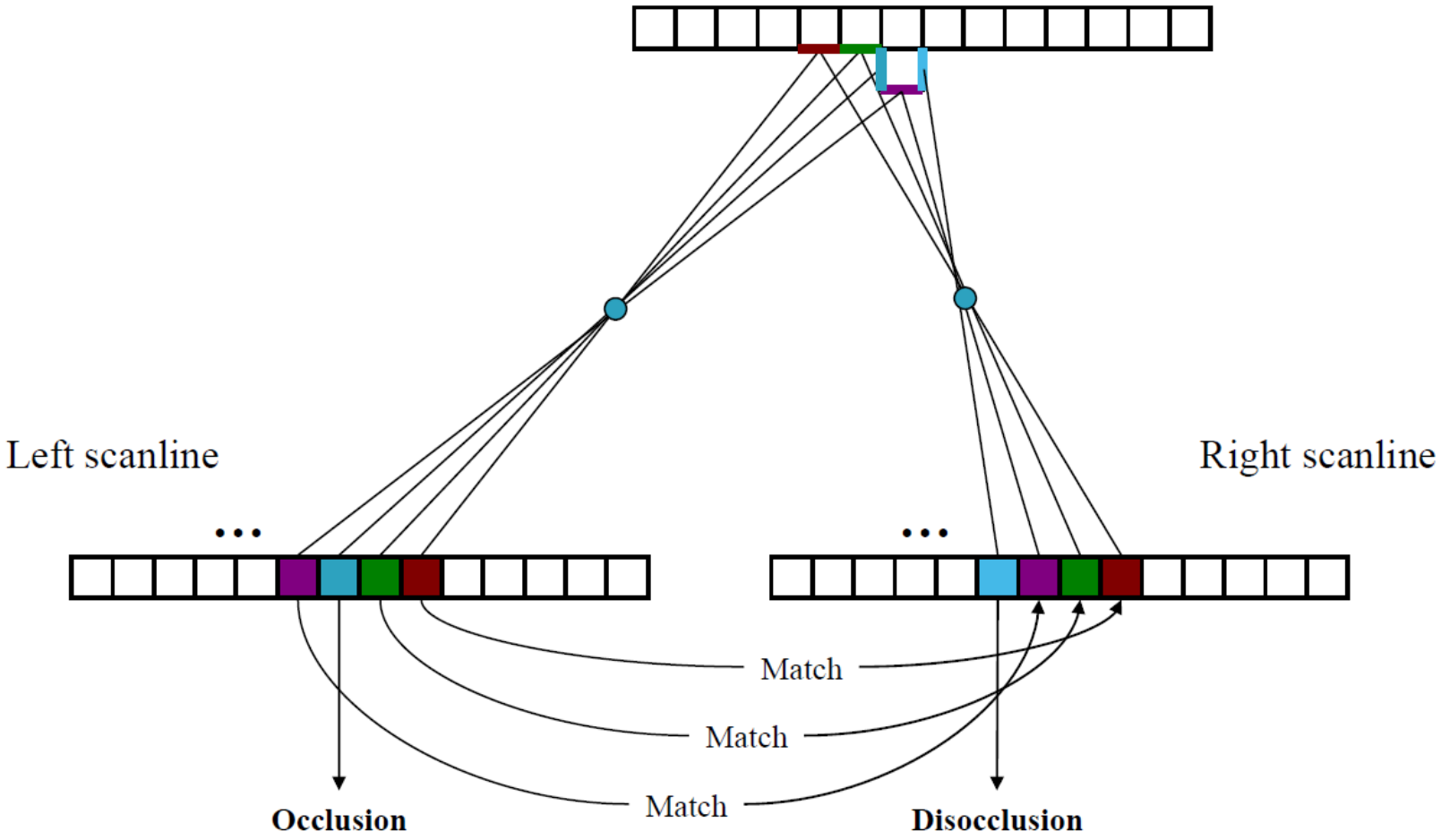
Dealing with ambiguities and occlusion

- Ordering constraint:
  - Impose same matching order along scanlines.
- Uniqueness constraint:
  - Each pixel in one image maps to unique pixel in other.
- Can encode these constraints easily in dynamic programming.

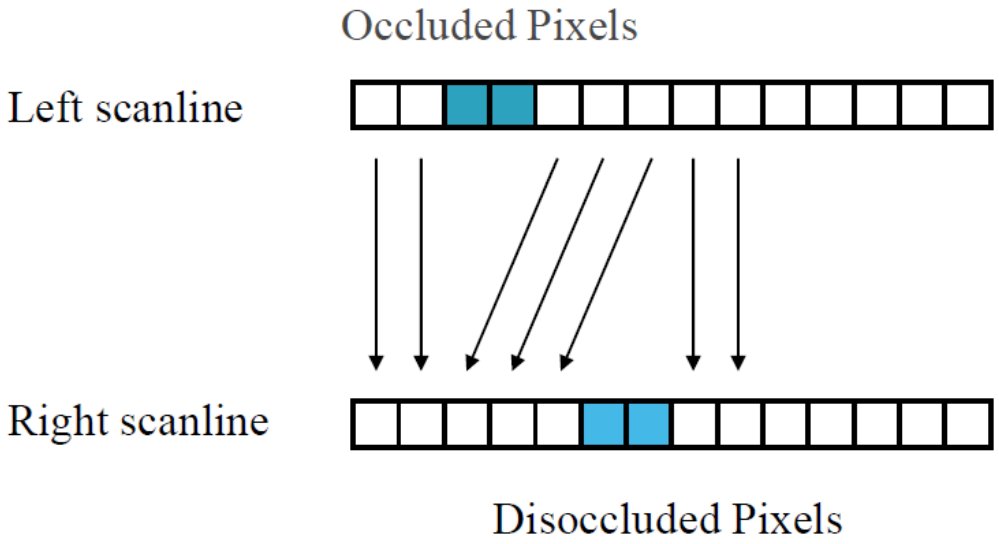
Pixel-based stereo



Stereo correspondences



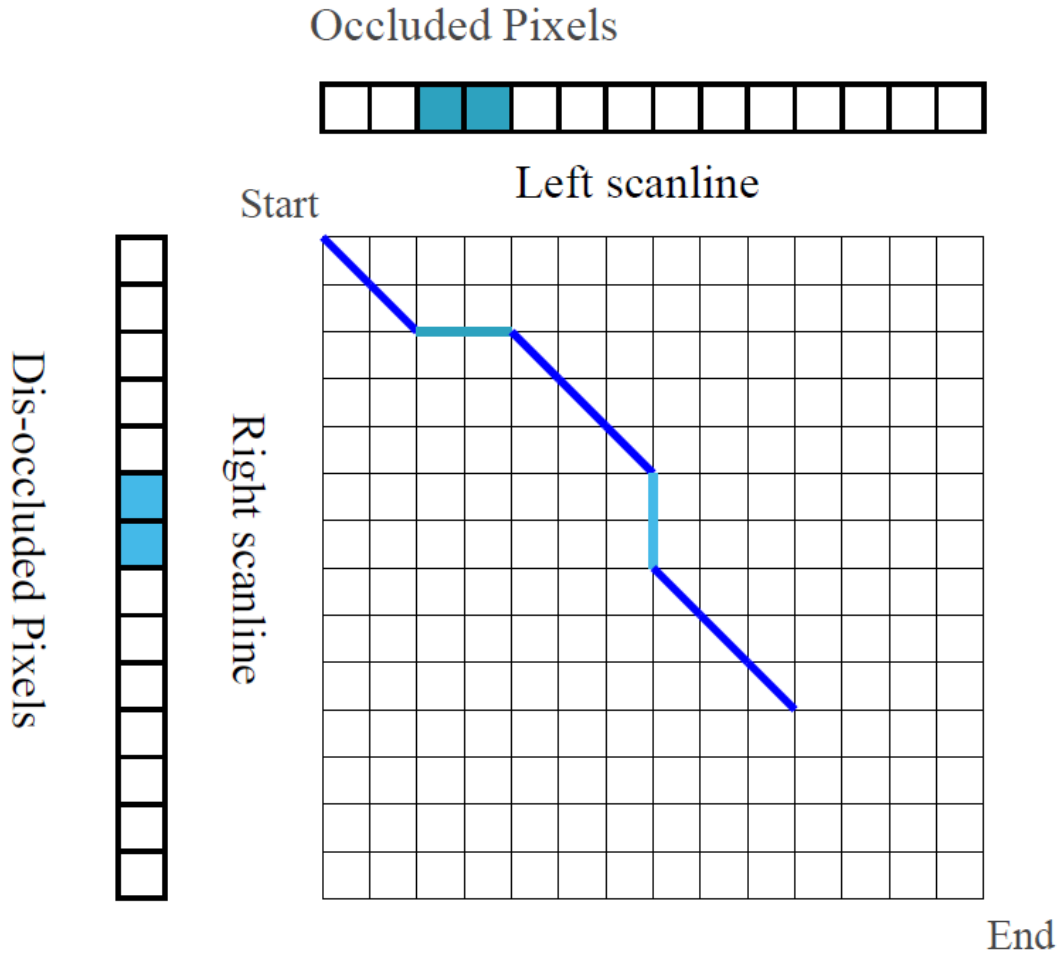
Search over correspondences



Three cases:

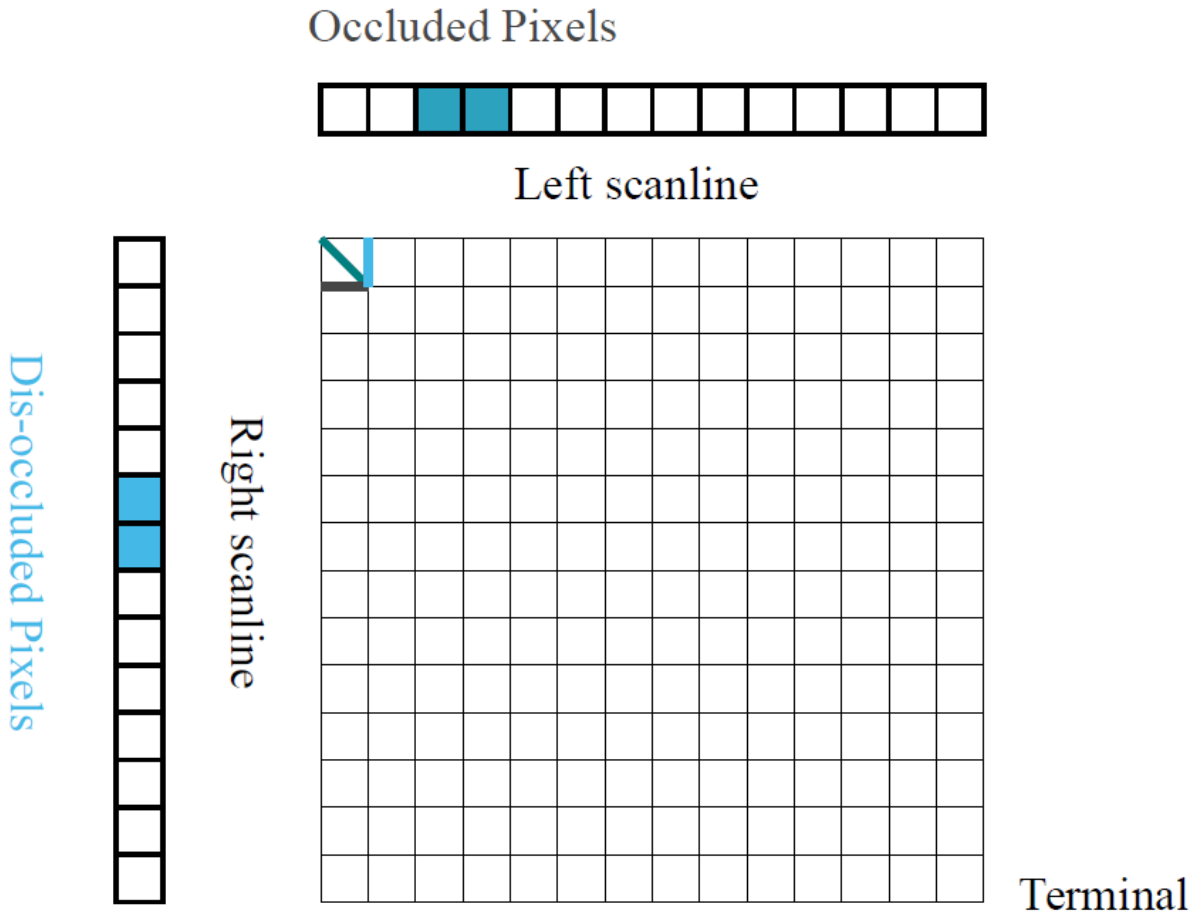
- Sequential – cost of match
- Occluded – cost of no match
- Disoccluded – cost of no match

Matching with dynamic programming

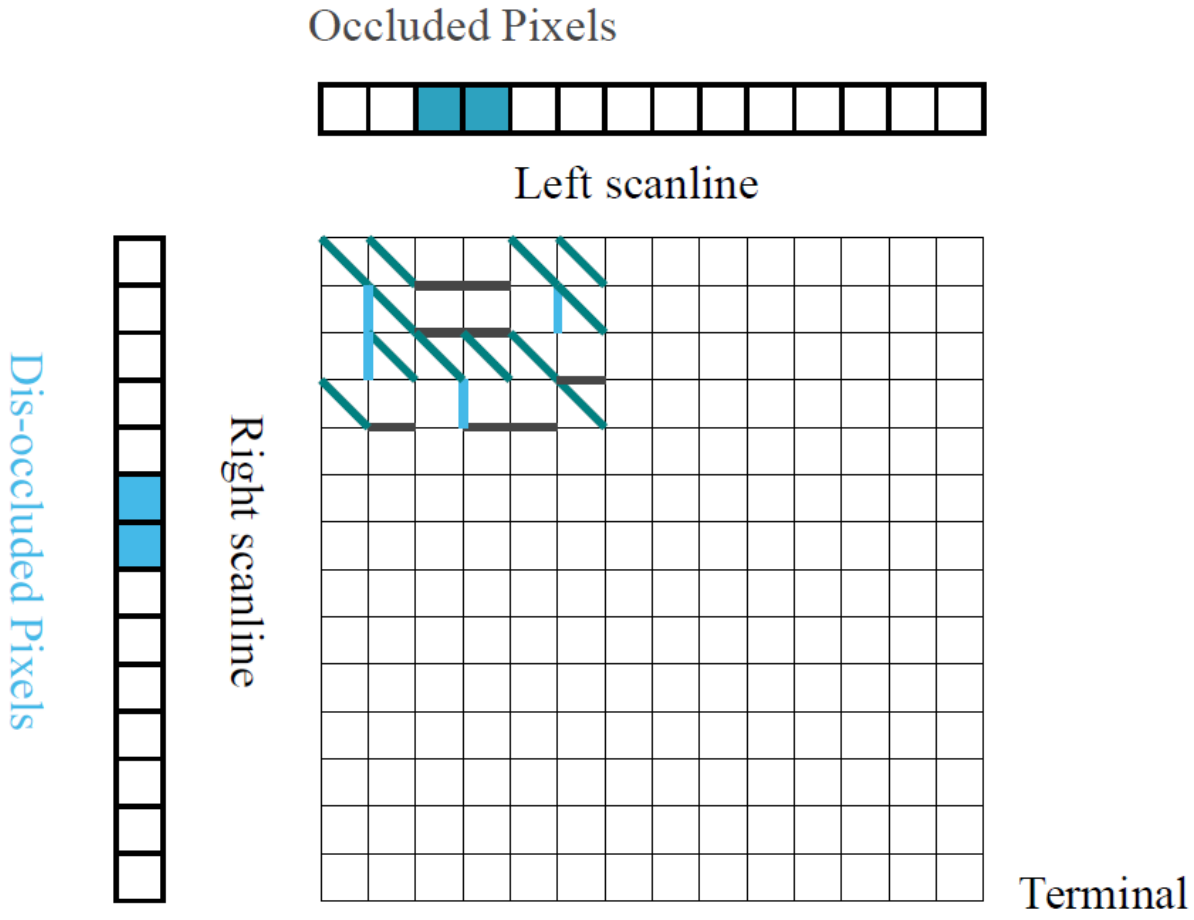


Dynamic programming yields the optimal path through grid. This is the best set of matches that satisfy the ordering constraint

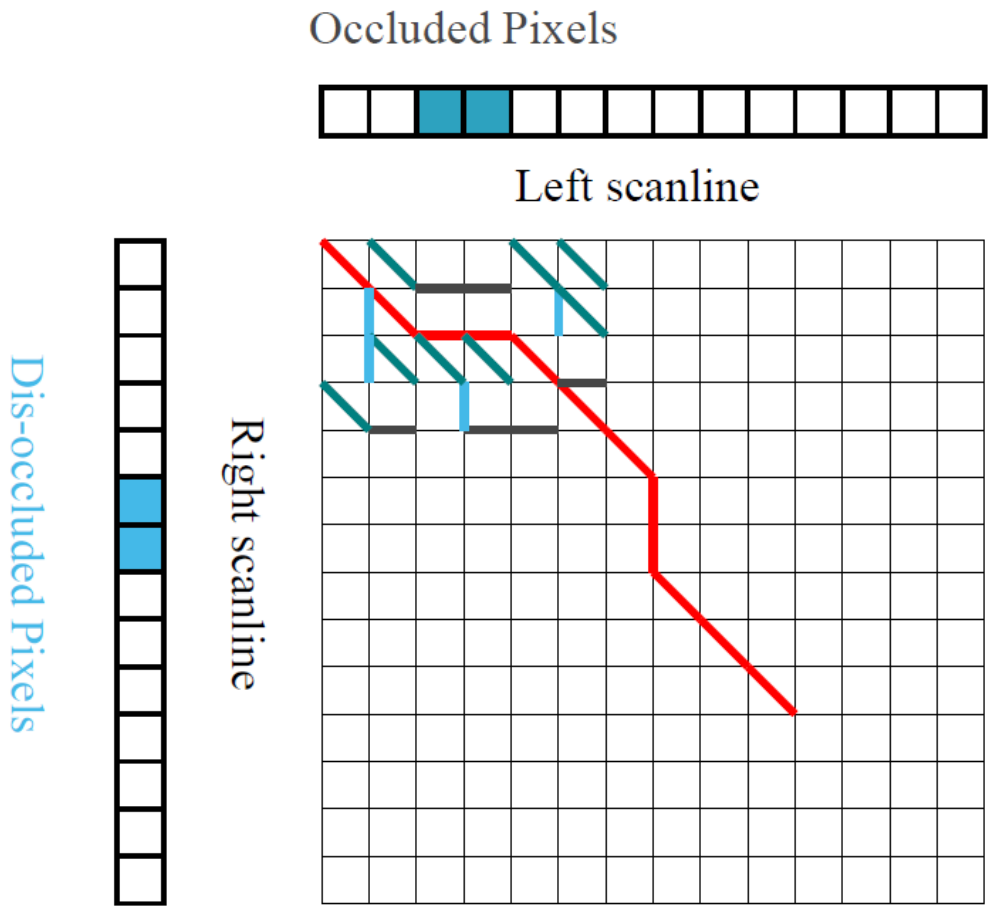
Matching with dynamic programming



Matching with dynamic programming

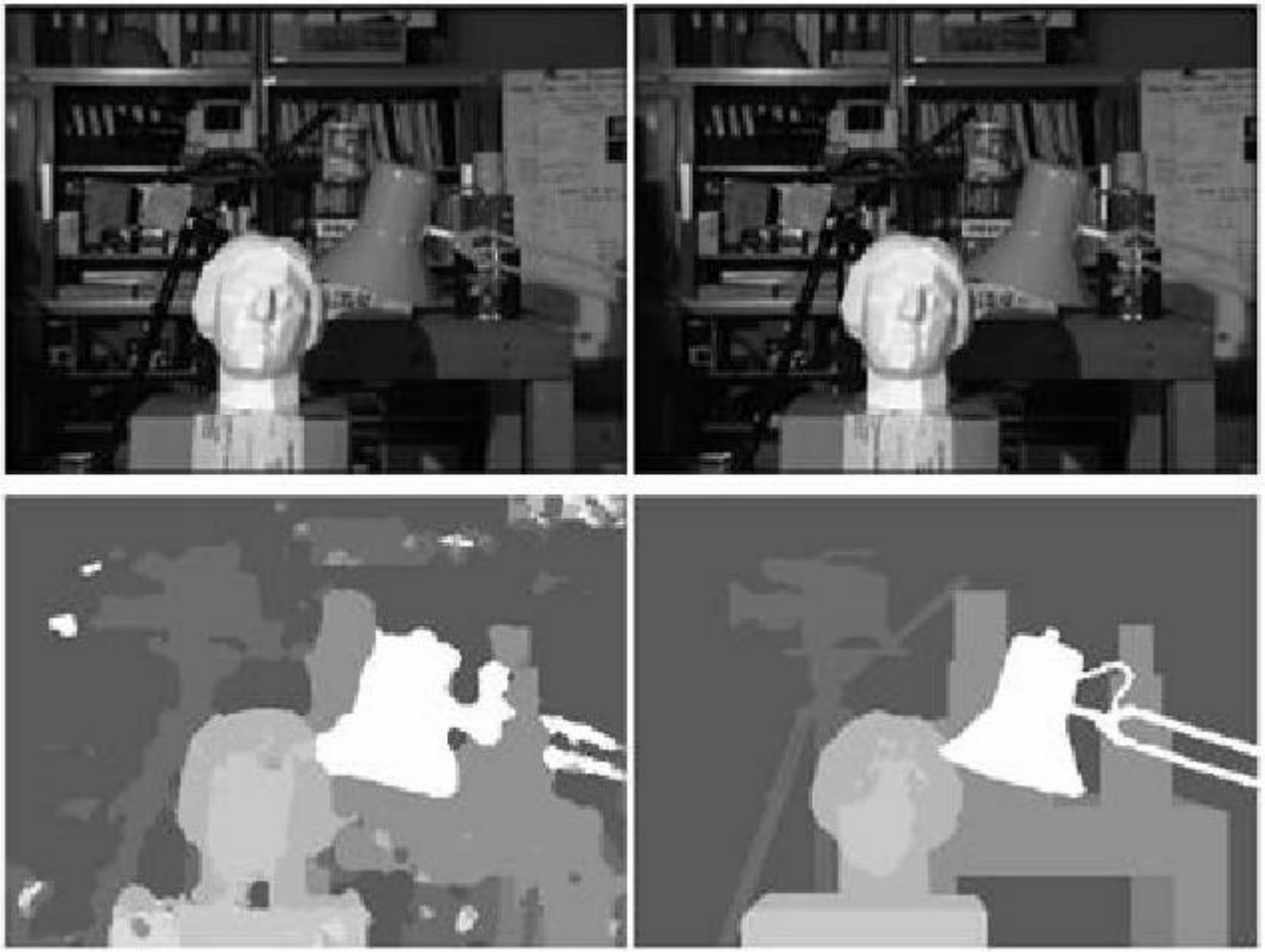


Matching with dynamic programming



Scan across grid computing optimal cost for each node given its upper-left neighbors.
Backtrack from the terminal to get the optimal path.

Results



Requirements

- Due: 04/24(SUN) 23:59:59
- 輸入：Two stereo image pairs
  - https://vision.middlebury.edu/stereo/data/
- 輸出：depth map (像上次一樣的純文字檔讓我可以用matlab打開看深度)、disparity map
- Bonus：拿自己拍攝的影像建立深度值
  - 注意相機位置要平行放置
  - Well callibrated

評分標準

- 輸出Disparity Map　80%
- 輸出Depth Map　20%

- 自己拍攝影像並成工輸出Depth Map　20%