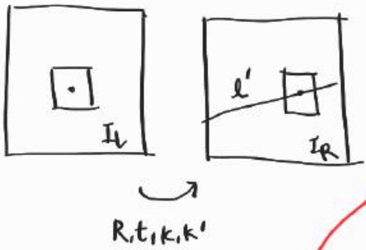
- Two-view stereo

Known baseline
(R, t, K, K')

$\xrightarrow[\text{Triangulation}]{\text{Epipolar geometry}}$ Dense 3D point cloud / disparity map (metric scale)

* As t is given at correct scale, acquired depth is accurate scale

- Basic stereo matching algorithm

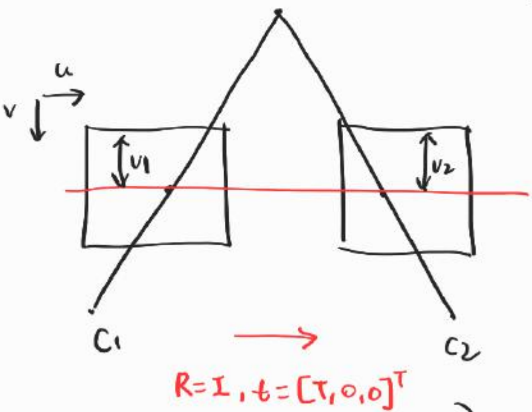


with known E/F, for every pixel in I_L

1) Find corresponding epipolar line at I_R : $l' = Fx$
2) Along l', pick the pixel that has the most similar appearance using local patch
3) Triangulate x to obtain depth

Challenge : calculating l' for every pixel is time consuming

→ Make epipolar lines as corresponding scanlines



↳ Image planes & cameras are parallel

⟺ C_1, C_2 are at same height
focal lengths are the same

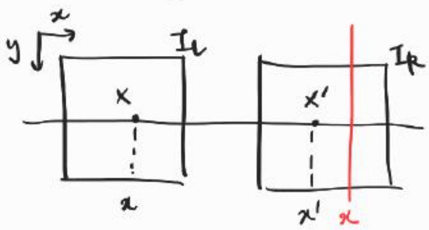
∴ epipolar lines are connected horizontal lines ($v_1 = v_2$)

Since $x'^T E x = 0$, $E = [t]_\times R$ Substitute : $E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix}$

Then, $$(u' \ v' \ 1) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = (u' \ v' \ 1) \begin{pmatrix} 0 \\ -T \\ Tv \end{pmatrix} = 0$$

$$Tv' = Tv \ \rightarrow \ v' = v$$

- Parallel images

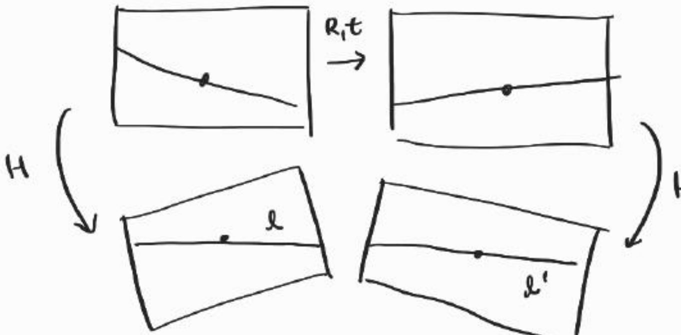


→ Disparity map : store $x - x'$ of every pixel of reference image I_L

Disparity $(x - x')$ cannot be negative! → $x > x'$

↳ Negative value implies image can see behind camera

- Non-parallel images : general case
  - ↳ Two-step process to create horizotal epipolar line
    1) Stereo calibration
       - Use Zhang's method w/ checkerboard pattern to find intrinsics (K, K')
       - Find distortion parameters and undistort image
       - Compute E using 8-point RANSAC algo.
       - Decompose E into (R, t), find exact scale using known size of checkerboard
    2) Stereo recitification
       - Correct images ($\tilde{x} = Hx$, $\tilde{x}' = H'x$) so that two images are row-aligned



General case : epipolar lines × horizontally aligned

Stereo recitification : manipulating image (H, H') s.t. l, l' horizontally aligned

* Camera position × changed!

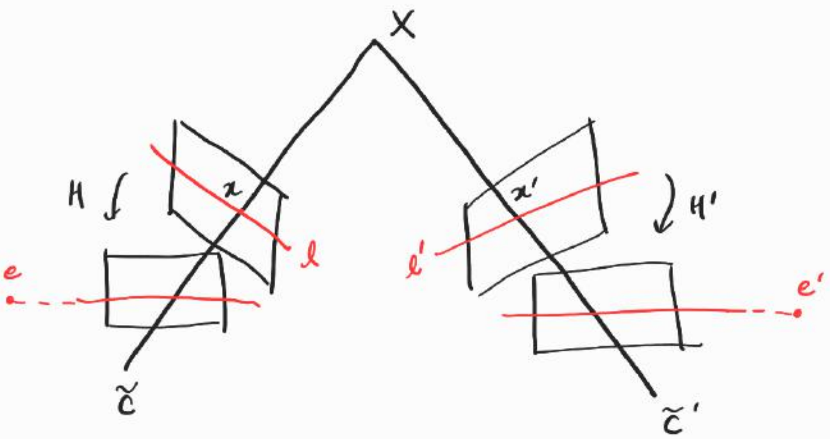
∴ 2 properties
- All epipolar lines are horizontal
- Corresponding points have equal vertical coordinates

Recall epipole is intersection of epipolar lines.
Since epipolar lines are parallel, e and e' mapped to infinity point $[1\ 0\ 0]^T$

↳ $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = He = H'e'$

- Stereo recitification

$\tilde{C}, \tilde{C}'$: inhomogenous camera center



$P = K[I|0]$

$P' = K'[R|t]$

- Steps

1) Compute normalized epipoles $\hat{e}, \hat{e}'$

$$e = Pc' = P\begin{bmatrix}\tilde{c}' \\ 1\end{bmatrix} = K[I|0]\begin{bmatrix}-R^Tt \\ 1\end{bmatrix} = -KR^Tt = -K\tilde{c}' \rightarrow \hat{e} = \tilde{c}'$$

$$e' = P'c = P'\begin{bmatrix}\tilde{c} \\ 1\end{bmatrix} = K'[R|t]\begin{bmatrix}0 \\ 1\end{bmatrix} = K't \rightarrow \hat{e}' = t$$

2) Compute H s.t. $H(\hat{e}) = [1\ 0\ 0]^T$

Good choice is pure rotation matrix

$$H = \begin{bmatrix}R_1^T \\ R_2^T \\ R_3^T\end{bmatrix} \quad \text{where} \quad R_1 = \frac{\tilde{c}'}{\|\tilde{c}'\|} \qquad \because H\tilde{c}' \text{ first element} = \frac{\tilde{c}'^T\tilde{c}'}{\|\tilde{c}'\|} = 1$$

$$R_2 = \frac{[-\tilde{c}_y, \tilde{c}_x, 0]^T}{\sqrt{\tilde{c}_x^2 + \tilde{c}_y^2}}, \quad R_3 = R_1 \times R_2 \qquad \because R \text{ is orthogonal matrix}$$

3) Compute H' s.t. $H'(\hat{e}') = [1\ 0\ 0]^T$

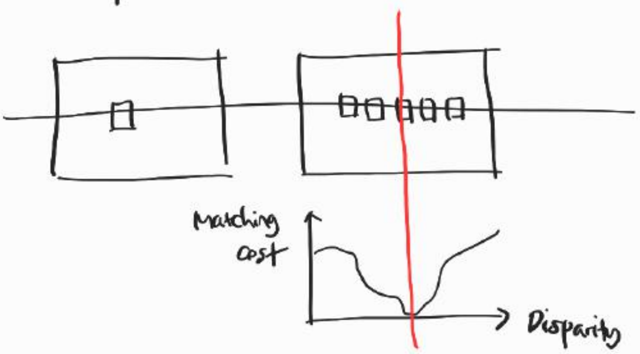
Good choice is $H' = HR$

4) Apply Hx, $H'x'$ of image pair

In practice, this gives hole in $Hx/H'x'$ image (as all pixels are not mapped)

$\therefore$ Find H^{-1}, H'^{-1}, then for each pixel of $Hx/H'x'$ images, find corresponding pixel value of original image

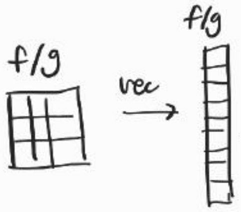
- Correspondence Search



- Types of matching cost

1) Normalized Cross correlation

f/g $\xrightarrow{\text{vec}}$ f/g



$\nearrow$ $\bar{f}/\bar{g}$: mean intensities

$\searrow$ σ_f/σ_g : std w/ $\bar{f}, \bar{g}$

Then,

$$\rho_{NCC}(f,g) = \frac{(f-\bar{f})(g-\bar{g})}{\sigma_f\sigma_g} \in [-1, 1]$$

$\uparrow\rho \rightarrow \uparrow$ similarity of f and g

⊕ Invariant to gain and bias e.g. lighting, material reflection

accuracy

⊖ fails when lacking texture variance along patches

2) Sum of Squared Differences

$\rho_{SSD}(f,g) = \|f-g\|^2$ i.e. L2 distance ∴ ↑ρ → ↓ similarity

↳ Since it's unbounded, normalized version : $\rho_{SSD}(f,g) = e^{-\frac{\|f-g\|^2}{\sigma^2}} \in [0,1]$

⊖ Sensitive to outliers

↳ solution : normalize patch intensities $\rho_{NSSD}(f,g) = \left\| \frac{f-\bar{f}}{\sigma_f} - \frac{g-\bar{g}}{\sigma_g} \right\|^2$

* $\rho_{NSSD}(f,g) = 2(1-\rho_{NCC}(f,g))$

3) Sum of Absolute Differences

$\rho_{SAD}(f,g) = \|f-g\|_1$ i.e. L1 distance

⊕ Robust to outliers

⊖ Sensitive to bias and gain

4) Mutual Information

$MI(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$ i.e. Dependency b/w r.v. X and Y

$\rho_{MI}(f,g) = -MI(f,g)$ ∴ ↑ρ cost = f,g are dissimilar

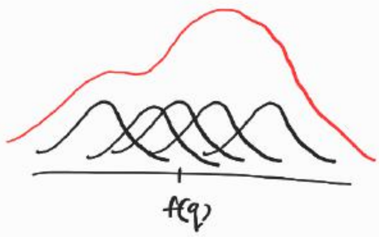
- Calculating $p(x,y)$ – Parzen window method

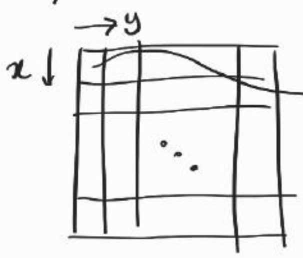
$p(x,y) = \frac{1}{|\Omega|} \sum_{q \in \Omega} K(f(q)-x,\, g(q)-y)$, $x,y \in [0,255]$ for gray scale

where $K(\cdot,\cdot)$ is 2d kernel density function e.g. 2d Gaussian

e.g. 1D case

$\sum_q K(f(q)-x)$: superposition of gaussians



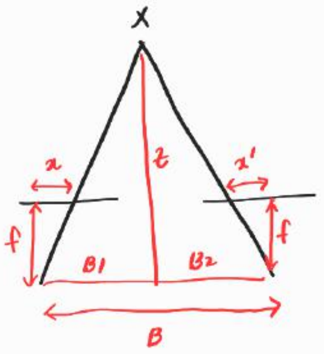


Discrete (x,y)

$P(X=1, Y=1)$

Then, $P(X) = \sum_Y P(X,Y)$ and $P(Y) = \sum_X P(X,Y)$ marginalization

- Depth from Disparity



Baseline = T after recitification

Using similar triangle property:

$$\frac{x}{f} = \frac{B_1}{z}, \quad \frac{-x'}{f} = \frac{B_2}{z}$$

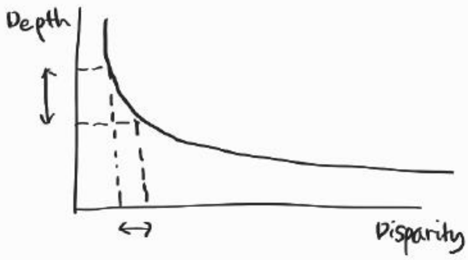
By adding two equations:

$$\frac{x-x'}{f} = \frac{B_1+B_2}{z}$$

- disparity : $x - x' = \frac{(B_1+B_2)f}{z} = \frac{Bf}{z}$
- depth : $z = \frac{(B_1+B_2)f}{x-x'} = \frac{Bf}{x-x'}$

$\therefore z \propto \frac{1}{x-x'}$: depth and disparity are inversely proportional

$x - x' > 0$ ($x > x'$) : disparity is positive value $\because z = \frac{Bf}{x-x'}$ where $z, B, f > 0$



when disparity is small i.e. stereo cameras are close

Small noise → huge Δ in measured depth ∴ ↑ error

- Block Matching

Above search method : independent block/patch comparisons → results blocky depth maps

- Large window size : ↑ detail, ↑ noise
- Small window size : ↓ detail (↑ smoothness)

- Failures of correspondence match
  - Textureless surfaces → uniform error
  - Occlusions, repetitions → multi-modal (peak) error
  - Non-Lambertian (view-dependent) surfaces e.g. glass

- Scanline optimization stereo

To overcome above failure cases, add a constraint that supports smoothness

↓ neighboring pixels should have similar disparities (depths)

  - Disparities w/ bound

For every pixel p, find discrete disparities from other image's scanline s.t.

$\{(x,y), (\tilde{x}, y)\} : x - \tilde{x} < \alpha$

- Cost function

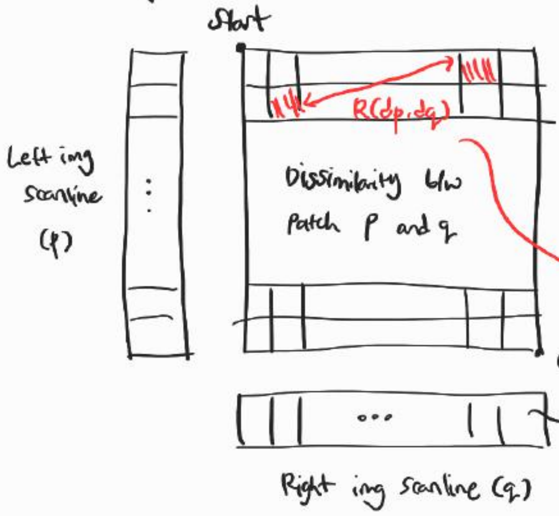
$$E(d) = \underbrace{\sum_p D(p, d_p)}_{} + \underbrace{\sum_{q \in N} R(d_p, d_q)}_{}$$

where $R(d_p, d_q) = \begin{cases} 0 & \text{if } d_p = d_q \\ R_1 & \text{if } |d_p - d_q| = 1 \\ R_2 & \text{if } |d_p - d_q| > 1 \end{cases}$ $(R_1 < R_2)$

↳ Smoothness constraint comparing disparity of neighboring pixels q along the scanline

↳ Same as block-matching cost

↓ Solved by dynamic programming



→ Original block-matching problem:
find lowest dissimarity of each row (= patch of left image)

↓ New problem

find shortest path from start to end

→ Neigboring pixels in left image should have similar disparity ≡ close columns of matrix

→ Patch along Scanline

↓ By applying bound to disparity

→ Discarded disparity ($x - \tilde{x} \geq \alpha$)

⇒ Resulting disparity map creates streaking artifacts (instead of blocky effects)

∵ Not considering relationships b/w different scanlines within an image

- Semi-Global Matching (SGM)

$$E(d) = \sum_p D(p, d_p) + \sum_{q \in N'} R(d_p, d_q)$$

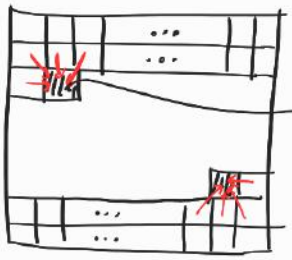
→ from neighboring pixels $q \in N$ along scanline to $q \in N'$ in any direction

↓ Extreme case: global matching
$q \in N''$ be the whole pixels in the image

problem: NP-complete

requires graph-cut (alpha-expansion) method to solve a problem

• 8-direction algorithm



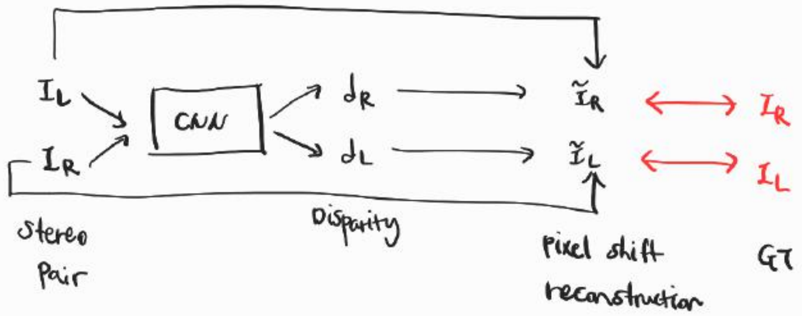
→ Aggregate cost of pixel p = $S(p,d)$ [forward pass (4-dir), backward pass (4-dir)]

$= \sum_r L_r(p,d)$

$\therefore$ Disparity at pixel p : $d^*(p) = \underset{d}{\operatorname{argmin}}\, S(p,d)$

sub-pixel disparity calculated by bilinear interpolation

• Single-view stereo : Deep learning



Use disparity to shift original image :

$\tilde{I}_R = I_L(d_R)$, $\tilde{I}_L = I_R(d_L)$

then compare it w/ original image

$\tilde{I}_R \leftrightarrow I_R$, $\tilde{I}_L \leftrightarrow I_L$

• Training loss ($I \leftrightarrow \tilde{I}$)

$L = \lambda$(Appearance Matching Loss) + μ(Disparity Smoothness Loss) + ω(Left-Right Disparity Consistency Loss)

• Multi-view Stereo

Given multi-view images and camera poses, find depth of all images

$\therefore$ Sparse 3D reconstruction → Dense 3D reconstruction

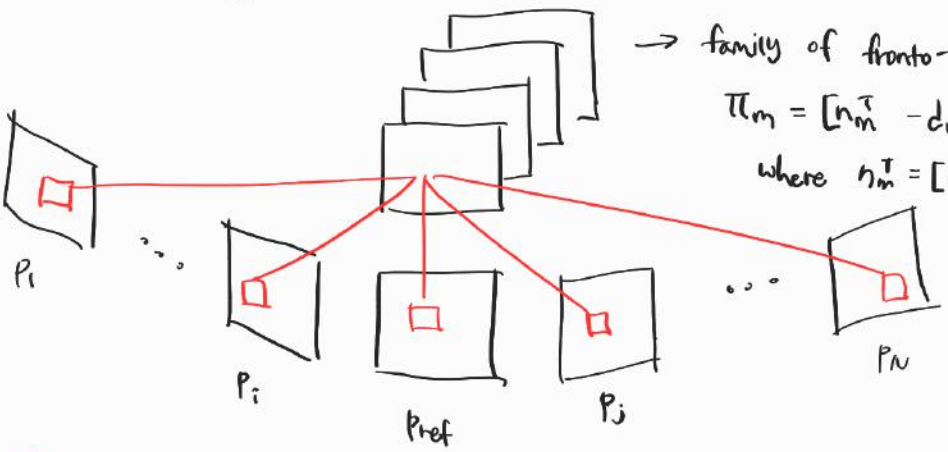
• Plane sweeping algorithm

Given M 3D planes (used for depth tests) & N+1 camera positions,

assume projection matrices P_k w/o radial distortion :

$P_k = K_k[R_k^T \; -R_k C_k]$ for camera $k = 1, 2, \cdots, N$

$P_{ref} = K_{ref}[I \; 0]$



→ family of fronto-parallel sweeping depth planes

$\Pi_m = [n_m^T \; -d_m]$ for $m = 1, 2, \cdots, M$

where $n_m^T = [0 \; 0 \; 1]^T$ and $d_m \in [d_{near}, d_{far}]$ bounded

Planar mapping homography b/w P_{ref} and P_k

- Homography

$$H_{\pi_m, p_k} = K_k \left(R_k^T + \frac{R_k^T C_k n_m^T}{d_m} \right) K_{ref}^{-1}$$

Then (x,y) at P_{ref} maps to (x_k, y_k) at P_k :

$$[\tilde{x}, \tilde{y}, \tilde{w}]^T = H_{\pi_m, p_k} [x, y, 1]^T$$

$$x_k = \tilde{x}/\tilde{w}, \quad y_k = \tilde{y}/\tilde{w}$$

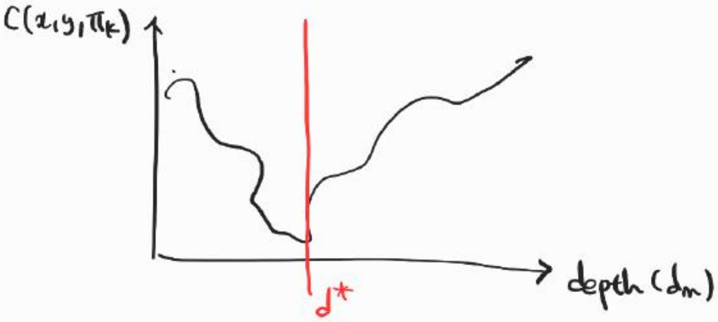
- Matching step

To find matching depth map π_k of (x,y) at P_{ref},

For every pixel (x,y) and depth plane π_k :

$$C(x,y,\pi_k) = \sum_{k=0}^{N-1} \sum_{(i,j)\in w} \left| I_{ref}(x-i, y-j) - \beta_k^{ref} I_k(x_k - i, y_k - j) \right|$$

w: patch around (x,y) β : gain ratio



Finally, depth is intersection b/w π_m and backprojected ray

$$Z_m(x,y) = \frac{-d_m}{[x,y,1] K_{ref}^{-T} n_m}$$

* Quality of depth map ↑ as # of views ↑