SAIR

Spatial AI & Robotics Lab

CSE 473/573-A

L14: STEREO MATCHING

Chen Wang

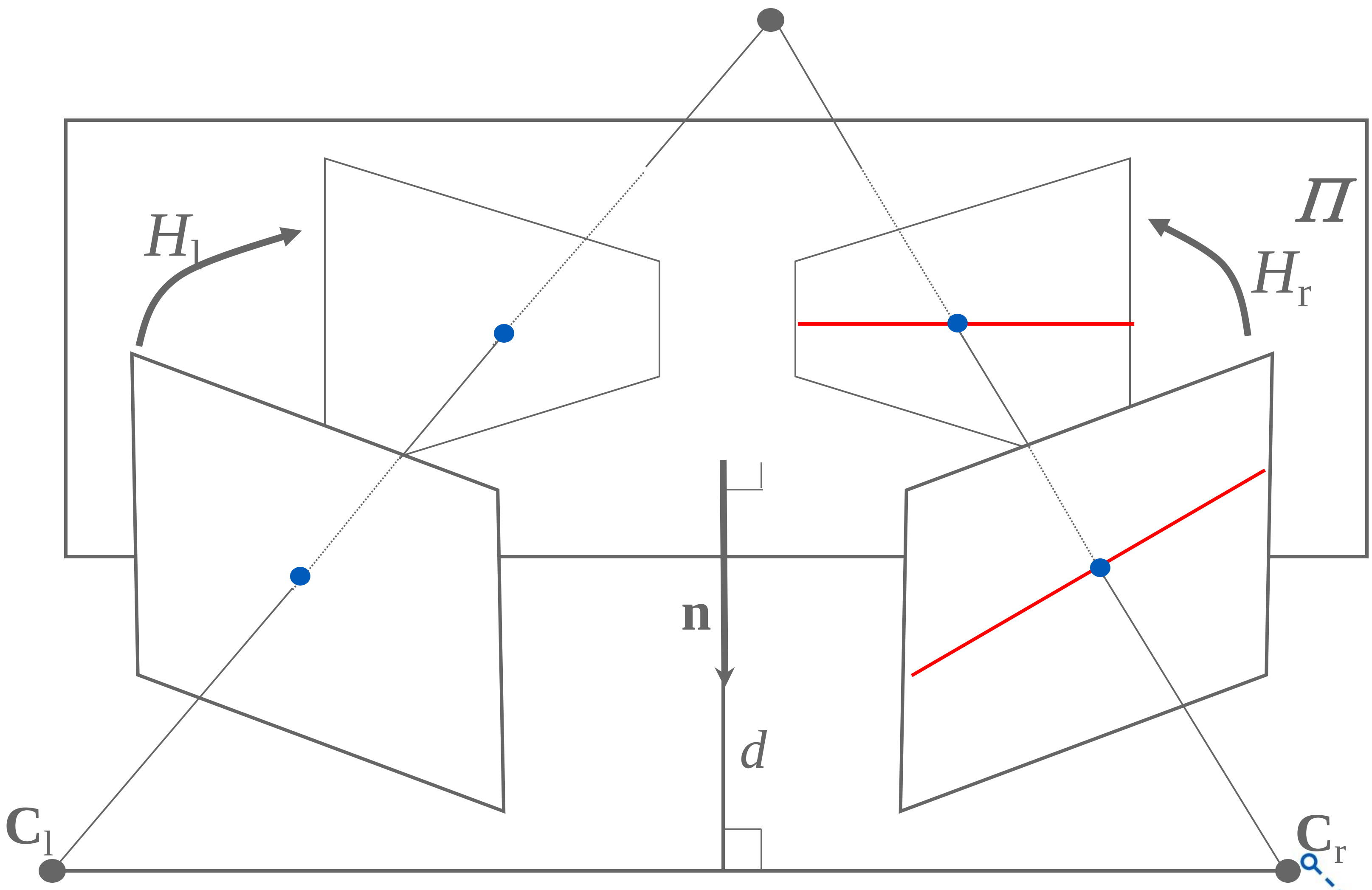
Spatial AI & Robotics Lab

Department of Computer Science and Engineering

University at Buffalo The State University of New York

Many Slides from Lana Lazebnik

Assume Rectified Stereo Images



STEREO VISION

Essential / Fundamental Matrix

Coordinates in 2-D (Recap)

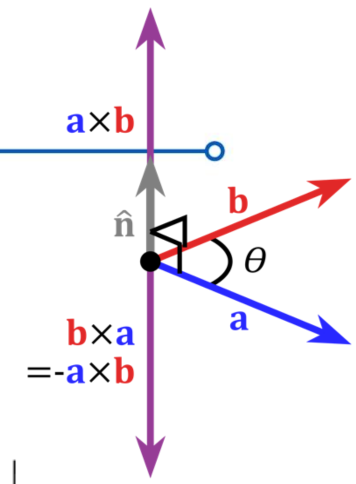
- Cartesian / homogeneous coordinates of point p

$$p = \begin{bmatrix} x \\ y \end{bmatrix} \qquad \begin{bmatrix} u \\ v \\ w \end{bmatrix} \qquad \begin{aligned} x &= u / w \\ y &= v / w \end{aligned}$$

- Homogeneous coordinate vector are equivalent if they are proportional to each other

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} \equiv \begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} \Leftrightarrow \begin{bmatrix} u \\ v \\ w \end{bmatrix} \equiv \lambda \begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} \quad \lambda \neq 0$$

Cross product (Recap)



- Cross product of two 3D vector

$$a \times b = \hat{n}\,|a|\,|b|\sin\theta$$

$$\begin{aligned}\mathbf{a}\times\mathbf{b} &= (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}) \times (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) \\ &= a_1b_1(\mathbf{i}\times\mathbf{i}) + a_1b_2(\mathbf{i}\times\mathbf{j}) + a_1b_3(\mathbf{i}\times\mathbf{k}) + \\ &\quad a_2b_1(\mathbf{j}\times\mathbf{i}) + a_2b_2(\mathbf{j}\times\mathbf{j}) + a_2b_3(\mathbf{j}\times\mathbf{k}) + \\ &\quad a_3b_1(\mathbf{k}\times\mathbf{i}) + a_3b_2(\mathbf{k}\times\mathbf{j}) + a_3b_3(\mathbf{k}\times\mathbf{k})\end{aligned}$$

$$\mathbf{a}\times\mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

- Skew-symmetric Matrix

$$\mathbf{a}\times\mathbf{b} = [\mathbf{a}]_\times \mathbf{b} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad [\mathbf{a}]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

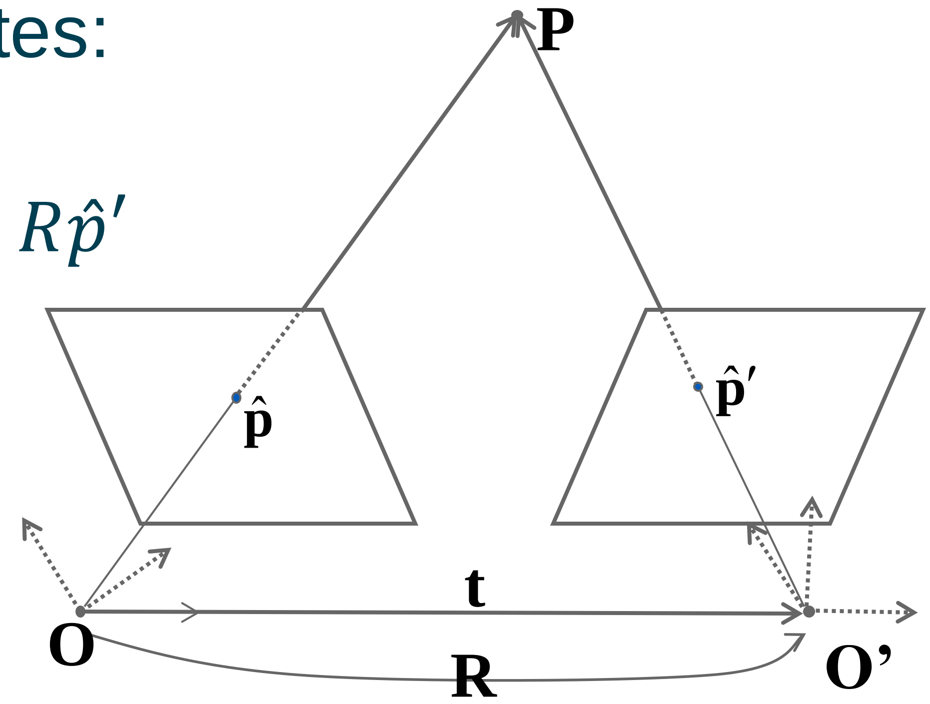
- Properties:

$$\mathrm{a}^{\mathrm{T}}(a \times b) = b^T(a \times b) = 0$$

SAIR
Spatial AI & Robotics Lab

Essential Matrix

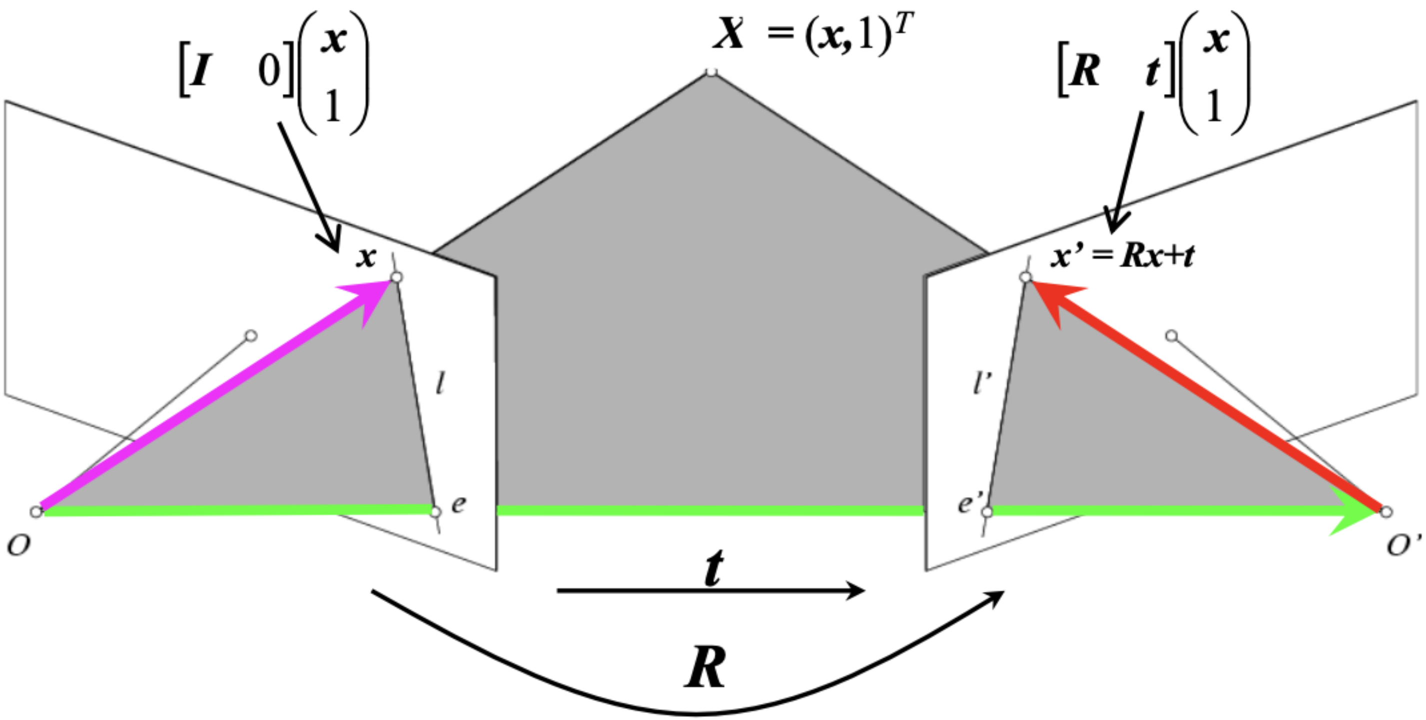
- Due to cross product properties
  - $\overrightarrow{OP} \cdot \left(\overrightarrow{OO'} \times \overrightarrow{O'P}\right) = 0$
- In homogeneous coordinates:
  - transform O to align O'
- Then direction of $\hat{p}'$ in O is $R\hat{p}'$
  - $\widehat{\boldsymbol{p}}^T(\boldsymbol{t} \times R\widehat{\boldsymbol{p}}') = 0$
  - $\widehat{\boldsymbol{p}}^T([\boldsymbol{t}]_\times R)\widehat{\boldsymbol{p}}' = 0$



$$\boxed{\begin{aligned}\hat{\mathbf{p}}^T\mathbf{E}\hat{\mathbf{p}}' &= 0\\ \mathbf{E} &= [\mathbf{t}]_\times\mathbf{R}\end{aligned}}$$

Essential Matrix

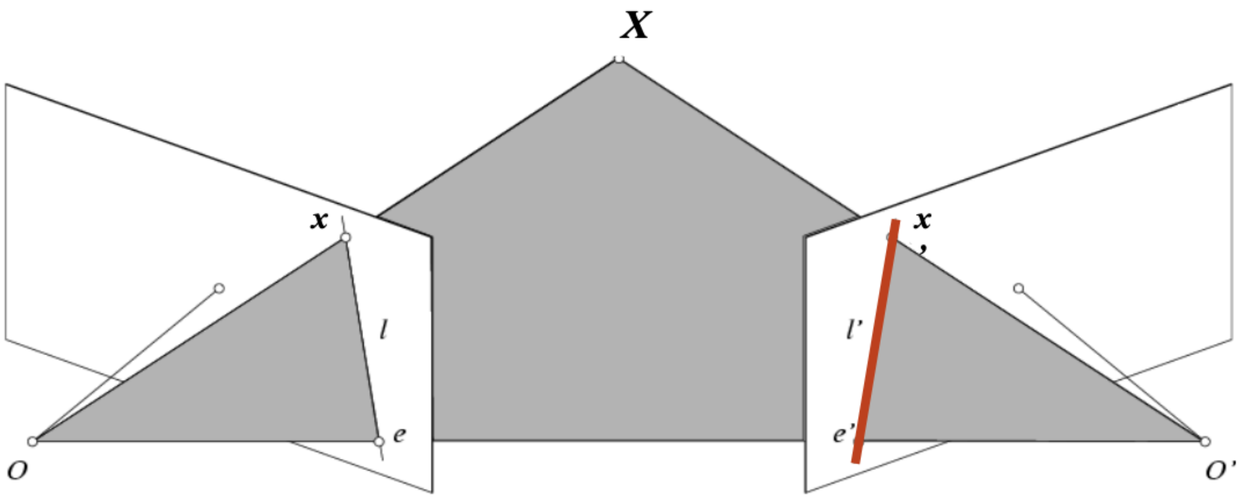
Epipolar constraint: Calibrated case



$$x' \cdot [t \times (Rx)] = 0 \Rightarrow x'^T [t_\times] R x = 0 \Rightarrow x'^T E x = 0$$

Essential Matrix
(Longuet-Higgins, 1981)

Epipolar constraint: Calibrated case



$$\boxed{x'^T E x = 0}$$

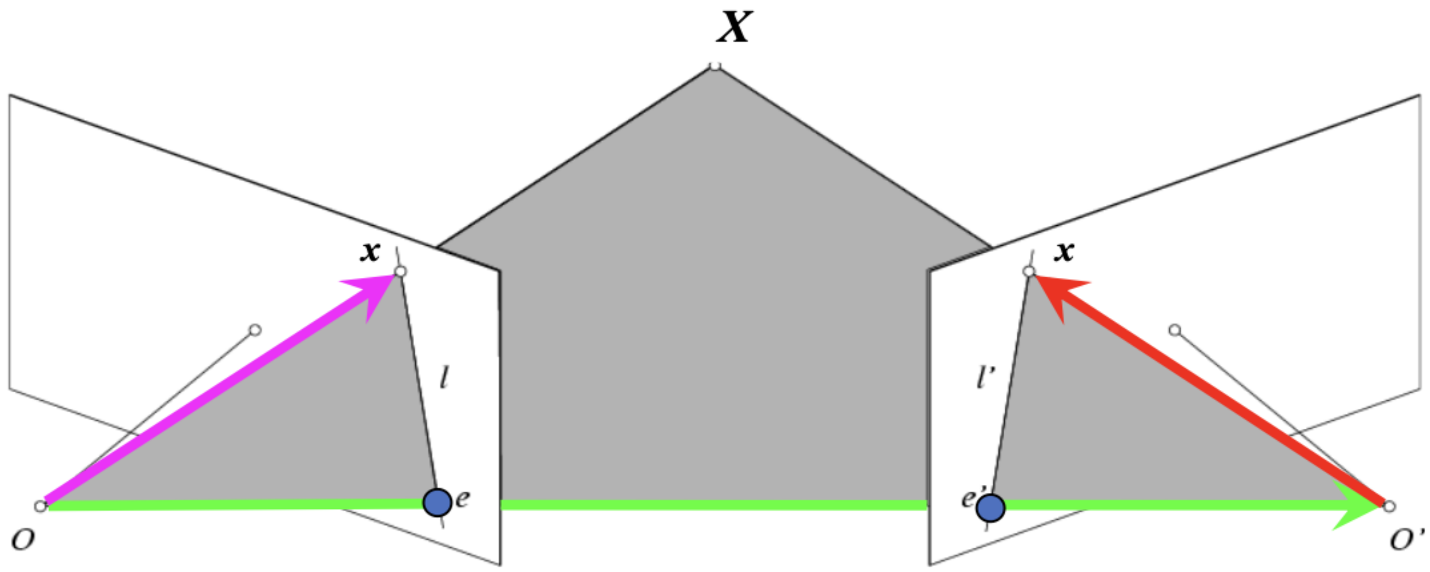
- Ex is the epipolar line associated with x' ($l' = Ex$)
- $E^T x'$ is the epipolar line associated with x ($l = E^T x'$)
- $Ee = 0$ and $E^T e' = 0$
- E is **singular** (rank two)
- $E = [t_\times]R$ has five degrees of freedom

- Recall: a line is given by $ax + by + c = 0$ or

$$\mathbf{l}^T\mathbf{x} = 0 \quad \text{where} \quad \mathbf{l} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$2T + 3R$

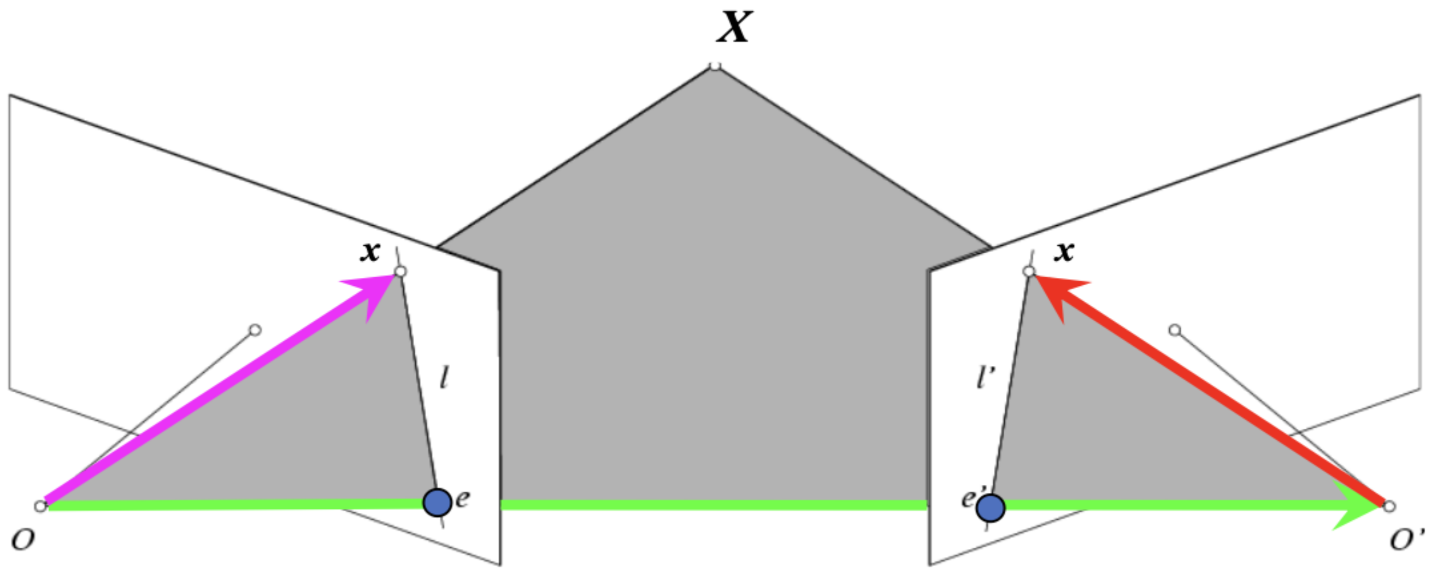
Epipolar constraint: Uncalibrated case



- The calibration matrices K and K' are unknown.
- We can write the epipolar constraint in terms of unknown normalized coordinates:

$$\hat{x}'^T E \hat{x} = 0 \qquad \hat{x} = K^{-1} x, \quad \hat{x}' = K'^{-1} x'$$

Epipolar constraint: Uncalibrated case



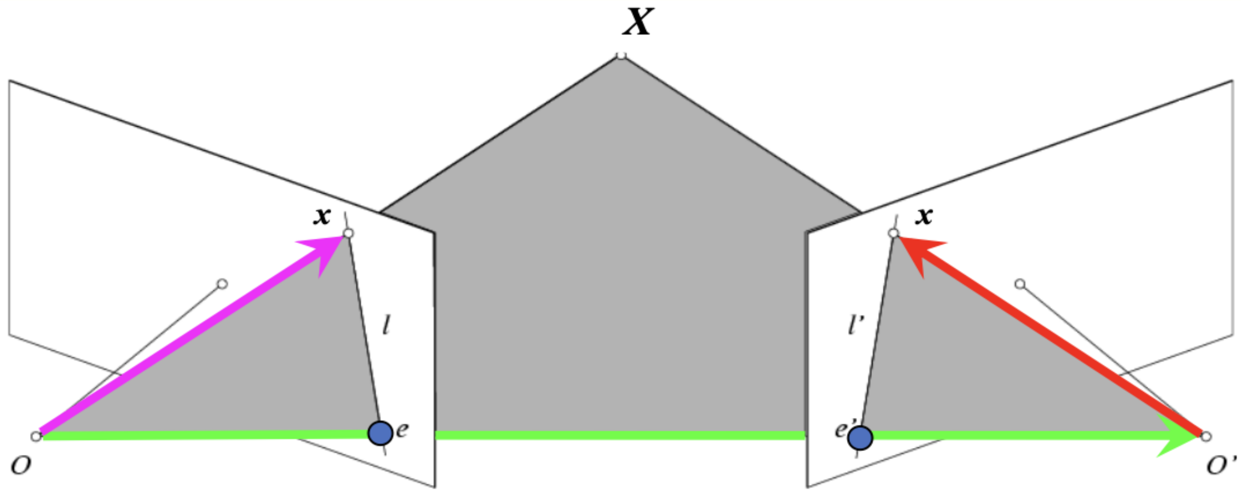
$$\hat{x}'^{T} E \hat{x} = 0 \quad \Rightarrow \quad x'^{T} F x = 0 \quad \text{with} \quad F = K'^{-T} E K^{-1}$$

$$\hat{x} = K^{-1} x$$

$$\hat{x}' = K'^{-1} x'$$

Fundamental Matrix
(Faugeras and Luong, 1992)

Epipolar constraint: Uncalibrated case



$$\hat{\boldsymbol{x}}'^{T} \boldsymbol{E} \hat{\boldsymbol{x}} = 0 \Rightarrow \boldsymbol{x}'^{T} \boldsymbol{F} \boldsymbol{x} = 0 \quad \text{with} \quad \boldsymbol{F} = \boldsymbol{K}'^{-T} \boldsymbol{E} \boldsymbol{K}^{-1}$$

- Fx is the epipolar line associated with x' ($l' = Fx$)
- $F^T x'$ is the epipolar line associated with x ($l = F^T x'$)
- $Fe = 0$ and $F^T e' = 0$
- F is singular (rank two)
- F has seven degrees of freedom $(5E + 2K)$

Estimating the Fundamental Matrix : 8-point algorithm

- Solve a system of homogeneous linear equations
  - a. Write down the system of equations

$$\boldsymbol{x}^T F \boldsymbol{x}' = 0$$

$$uu'f_{11} + uv'f_{12} + uf_{13} + vu'f_{21} + vv'f_{22} + vf_{23} + u'f_{31} + v'f_{32} + f_{33} = 0$$

$$\mathrm{A}\boldsymbol{f} = \begin{bmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu_v' & u_nv_n' & u_n & v_nu_n' & v_nv_n' & v_n & u_n' & v_n' & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ \vdots \\ f_{33} \end{bmatrix} = \mathbf{0}$$

  - b. Solve f from $Af = 0$ using SVD (refer to Project 1).

```
[U, S, V] = svd(A);
f = V(:, end);
F = reshape(f, [3 3])';
```

Q & A

- Why do we need 4 points for homography but 7/8 points for fundamental matrix calculation?
  - In the case of fundamental matrix, each point relates to only one constraint, while in homograph, each point is related to two constraints.
- Can we use 7 points solve fundamental matrix?
  - In fact, the fundamental matrix only has 7 degrees of freedom. In this case, the rank-2 constraint must be enforced during the computations.

Estimating the Fundamental Matrix

- 8-point algorithm
  - Least squares solution using SVD on equations from 8 pairs of correspondences.
- 7-point algorithm
  - least squares to solve for null space (two vectors) using SVD and 7 pairs of correspondences.
  - Solve for linear combination of null space vectors that satisfies $\det(F) = 0$
- Minimize reprojection error
  - Non-linear least squares
- Note: estimation of F (or E) is degenerate for a planar scene.

SAIR
Spatial AI & Robotics Lab

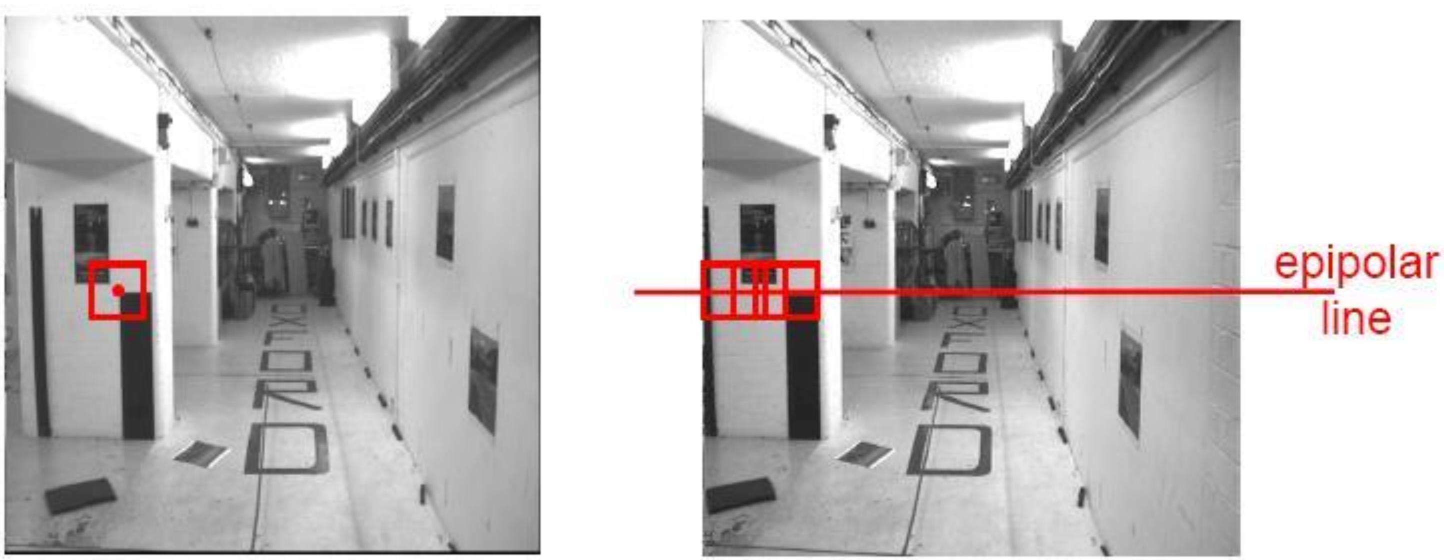
Correspondence Search

- Other “soft” constraints (To cover)
  - 1. Similarity
  - 2. Uniqueness
  - 3. Disparity gradient
  - 4. Ordering

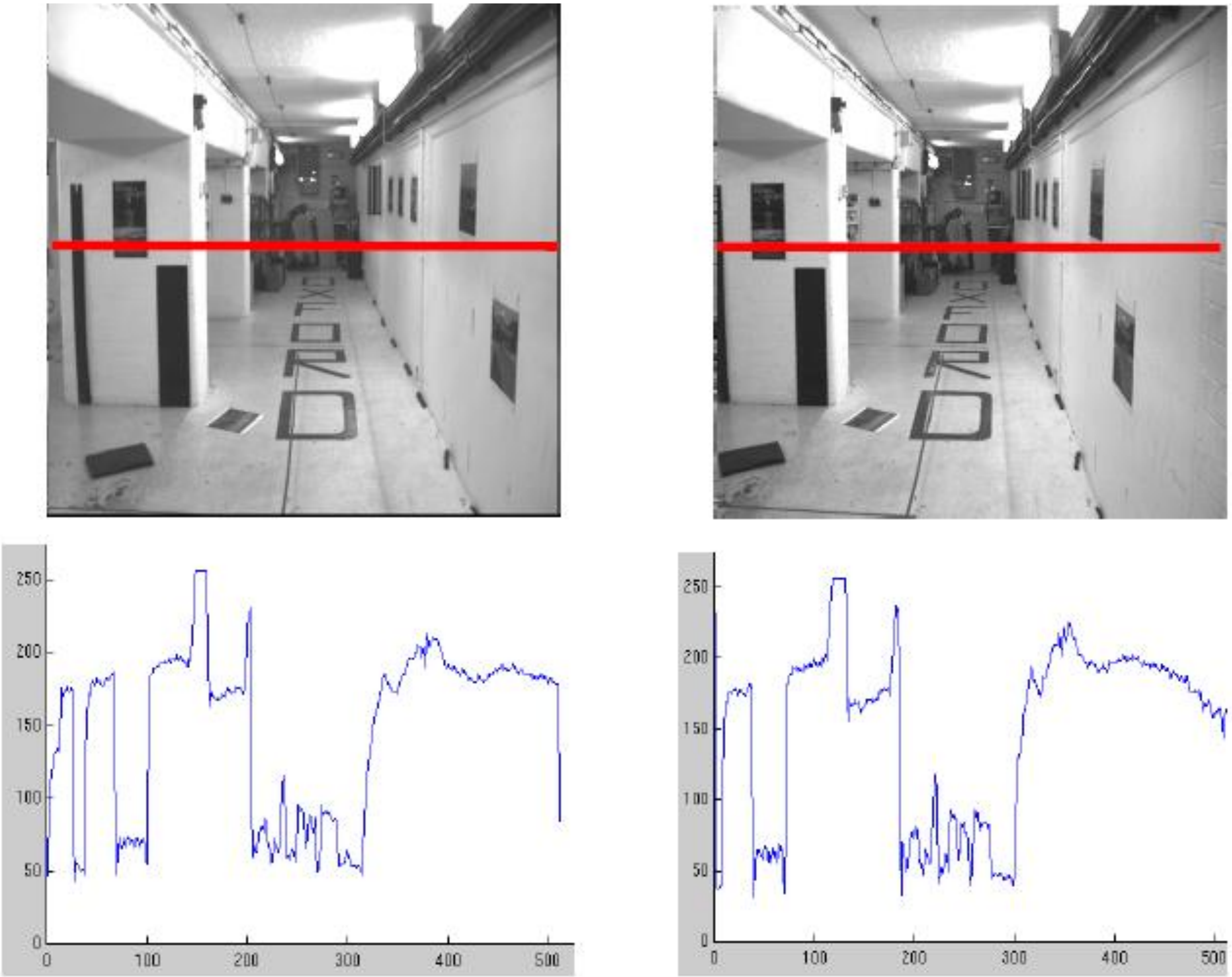
- To find matches in the image pair, we will assume
  - Most scene points visible from both views
  - Matched regions are similar in appearance

Dense correspondence search

- Neighborhoods of corresponding points are similar in intensity patterns.



Intensity profiles



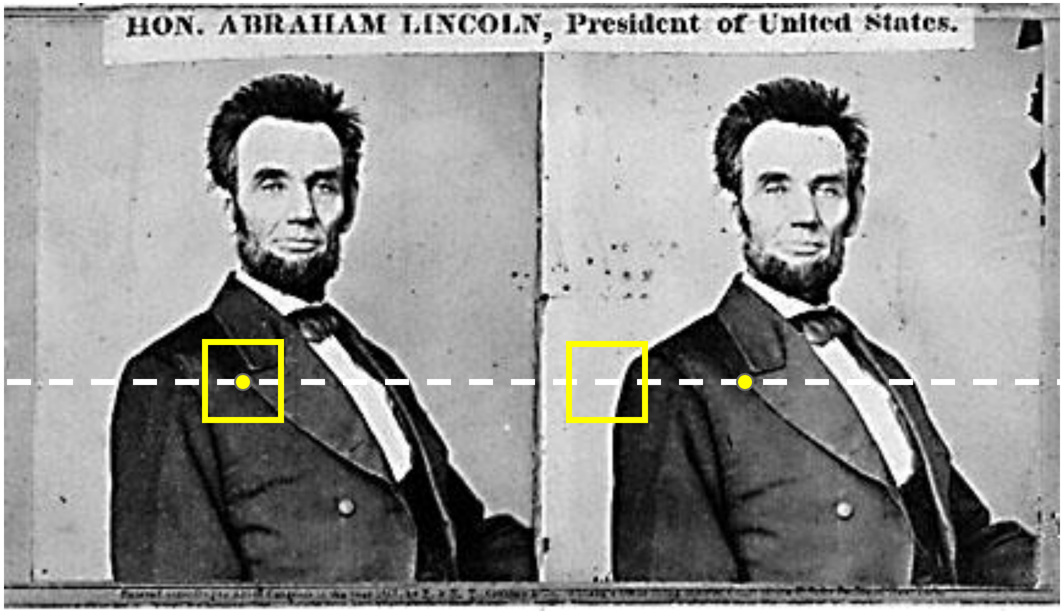
Intensity profiles

- Clear correspondence between intensities, but also noise and ambiguity

Source: Andrew Zisserman

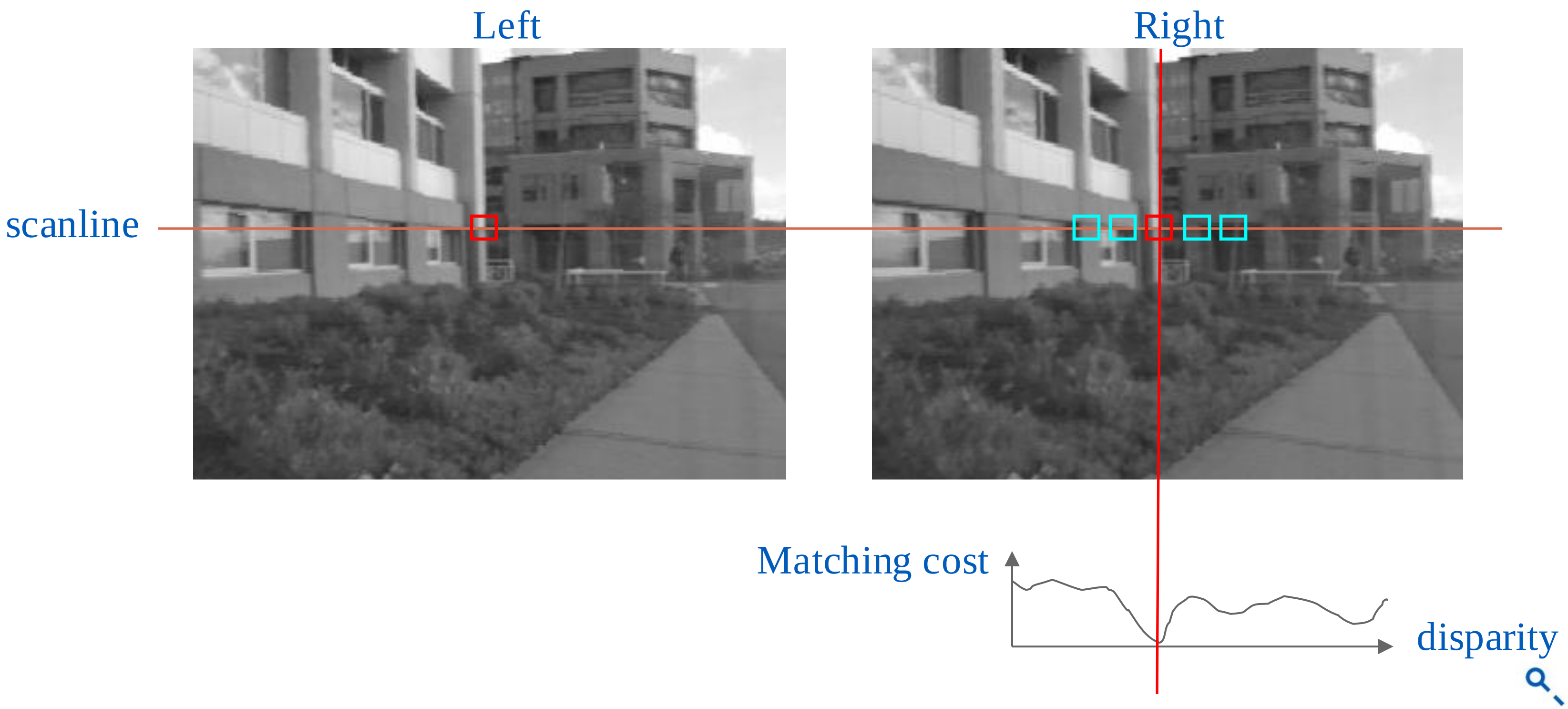
Dense correspondence search

- For each epipolar line
  - For each pixel / window in the left image
    - Compare with every pixel / window on same epipolar line
    - Pick position with minimum match cost
      - SSD, normalized correlation

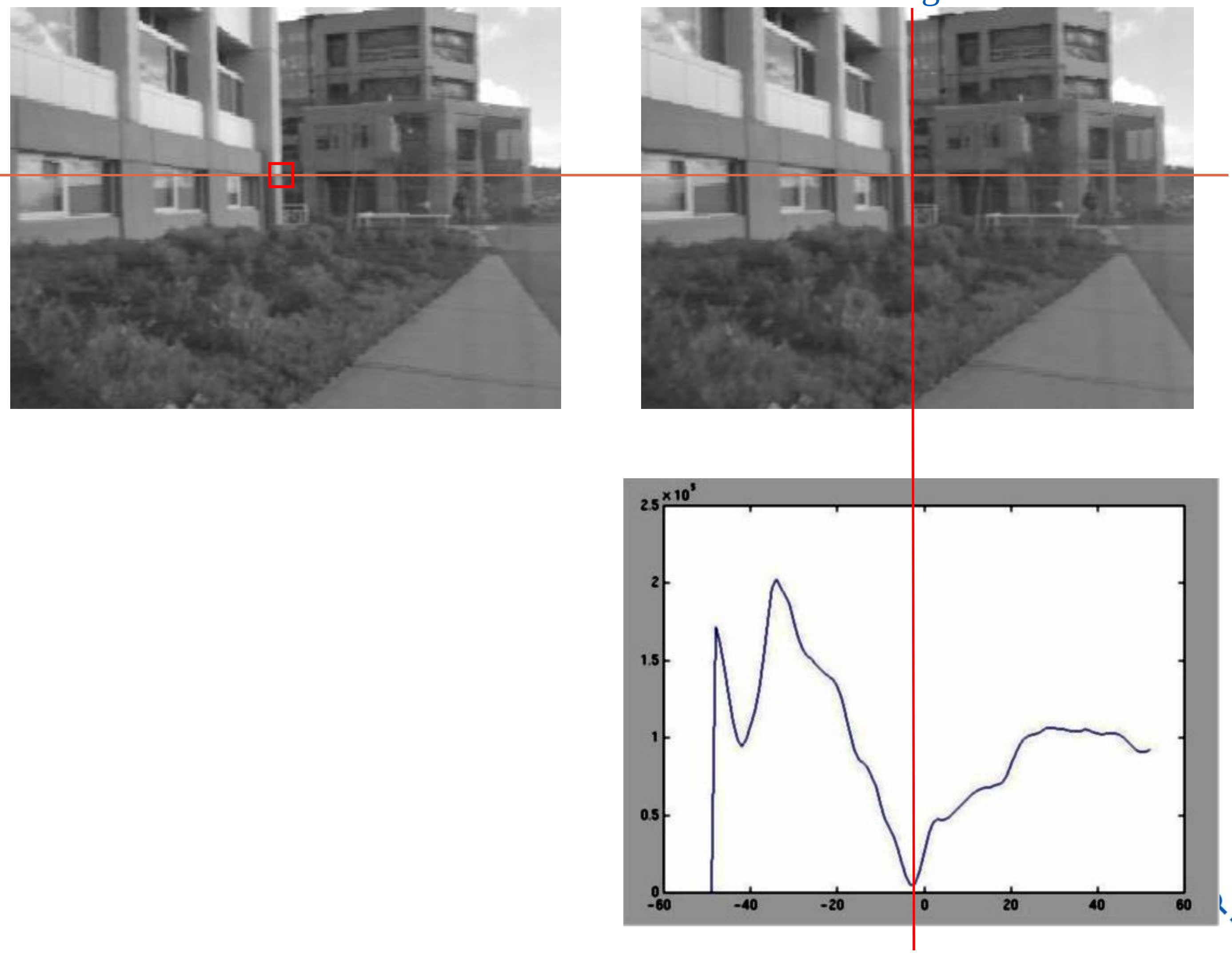


Similarity constraints

- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation



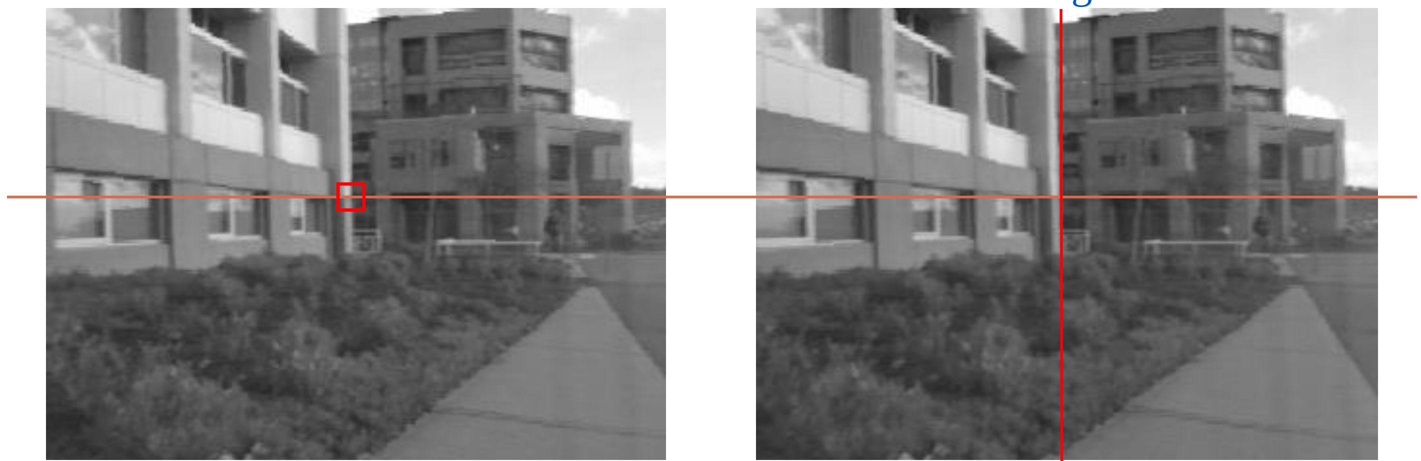
Similarity constraints: SSD

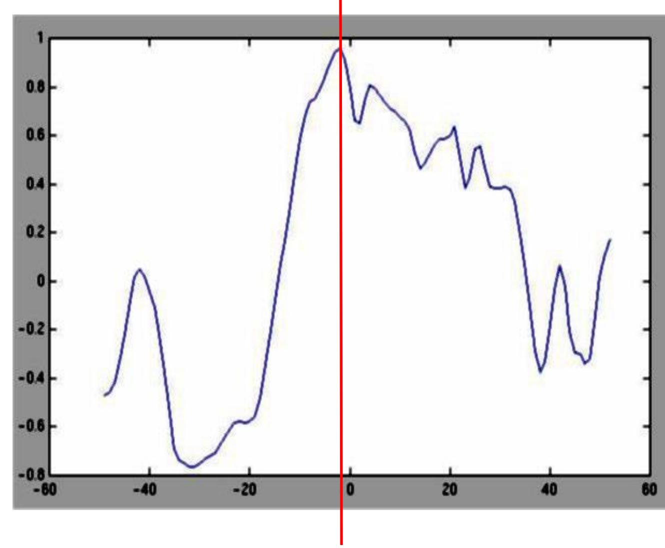


Similarity constraints: Norm. Corr.

Left Right

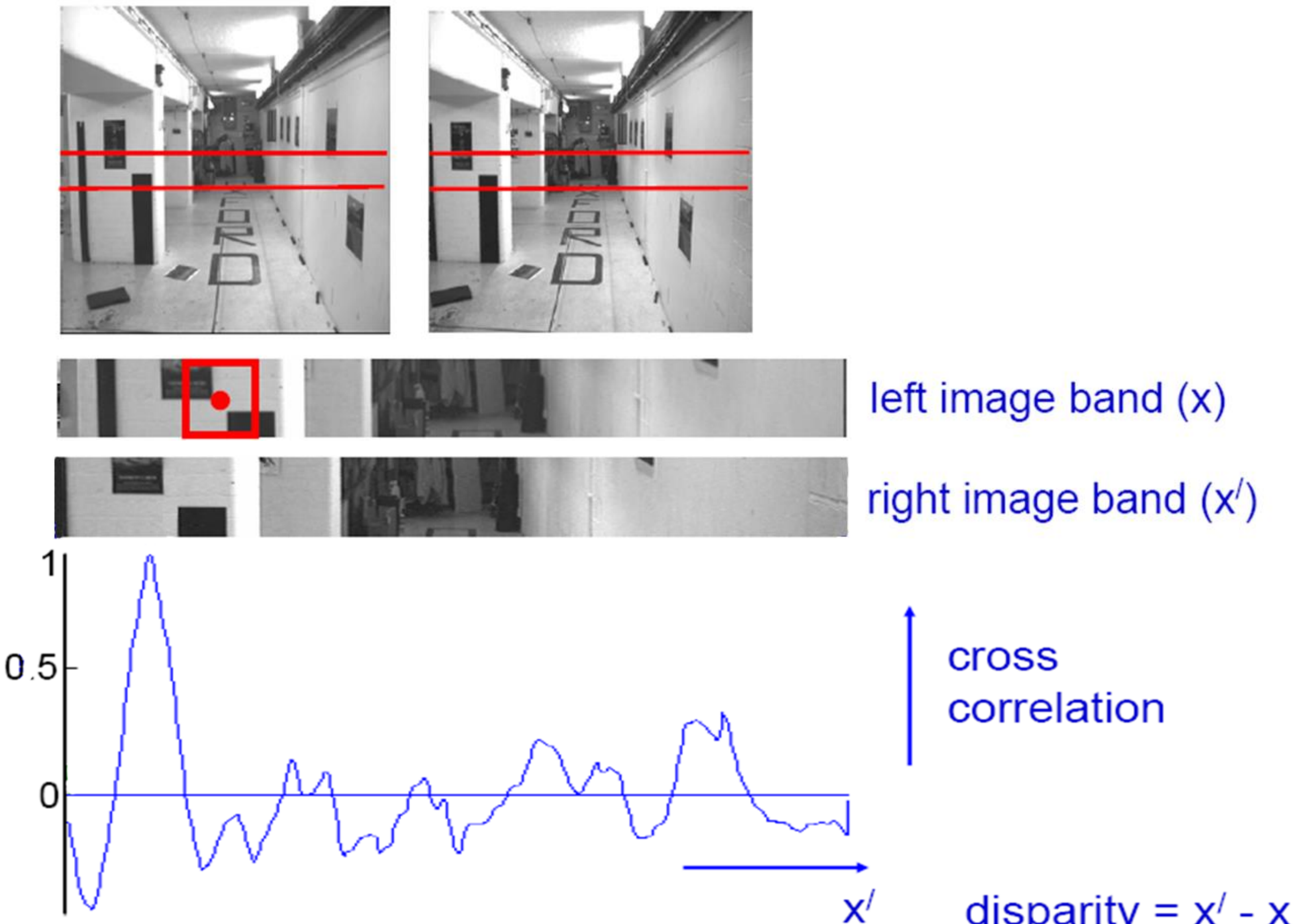
scanline



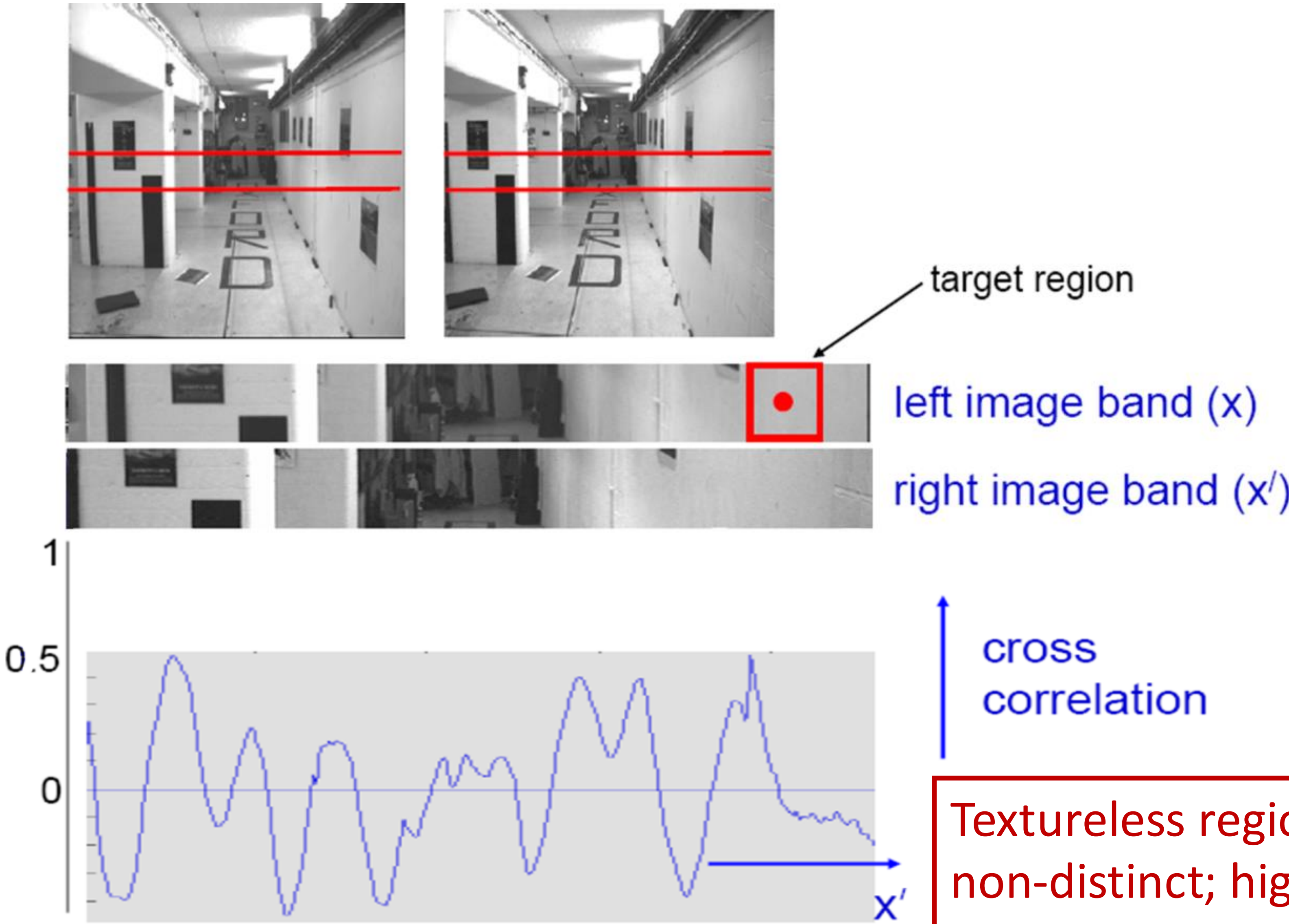


Norm. corr

Correlation-based window matching



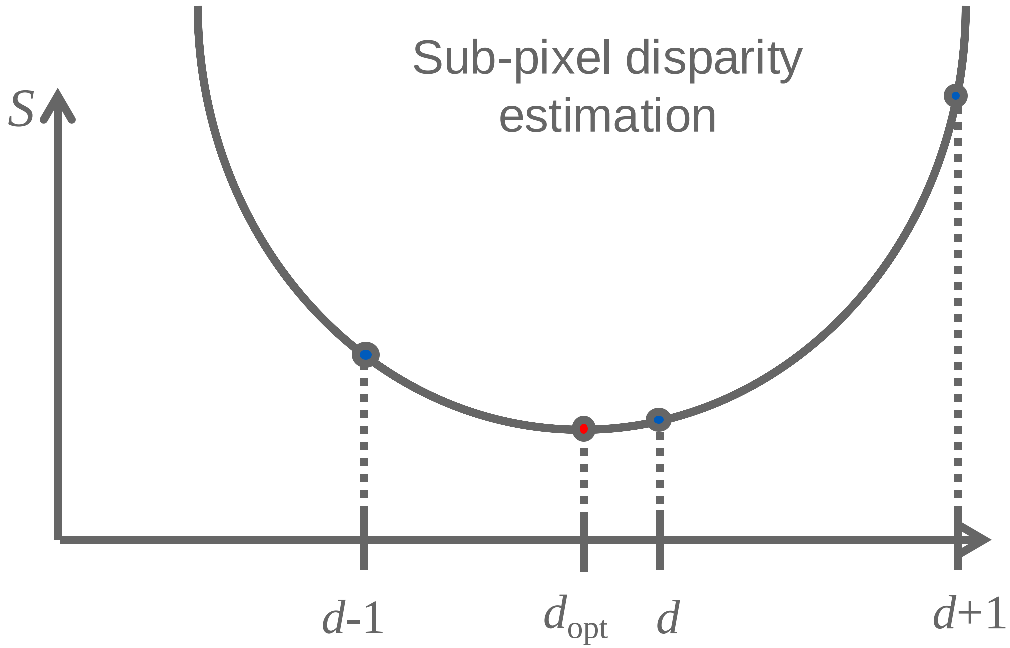
Correlation-based window matching



Textureless regions are non-distinct; high ambiguity for matches.

Sub-pixel disparity estimation

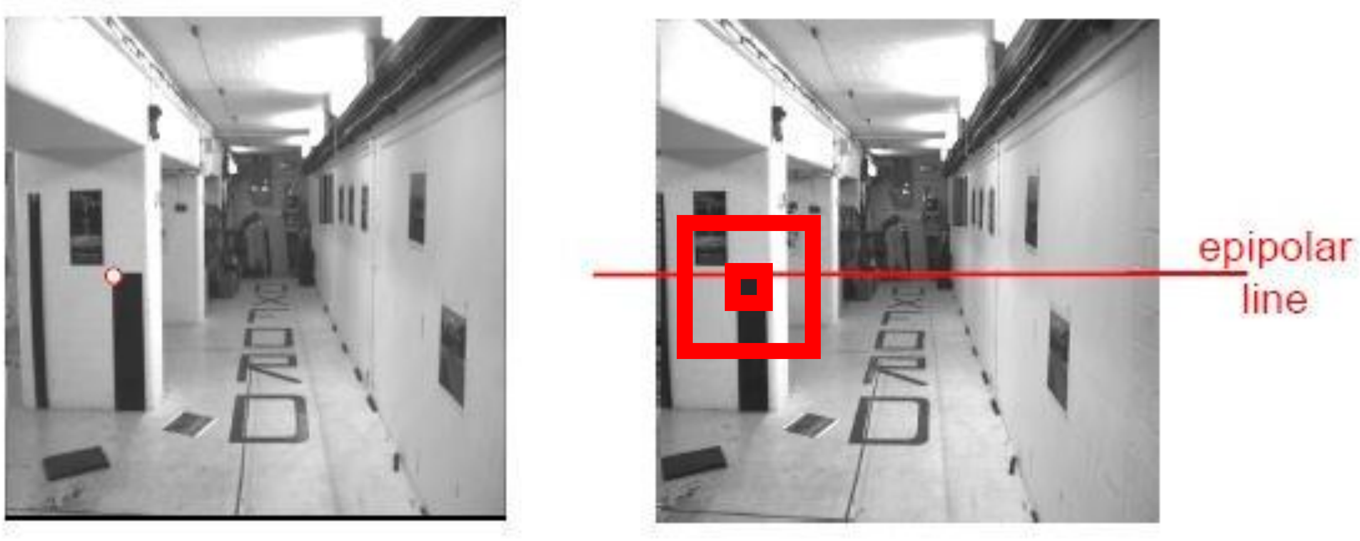
- Let S be the SSD



- $S(d) = ad^2 + bd + c$
  - $S(0) = \ c$
  - $S(1) = a + b + c$
  - $S(-1) = a - b + c$
- Solving this, we obtain:
  - $a = (\ S(1) + S(-1) - 2\ S(0))/2$
  - $b = (S(1) - S(-1))/2$
  - $c = S(0)$
- $S'(d) = 2ad + b = 0$

$$d_{opt} = \frac{(S(-1) - S(1))}{2(S(1) + S(-1) - 2S(0))}$$

Effect of window size

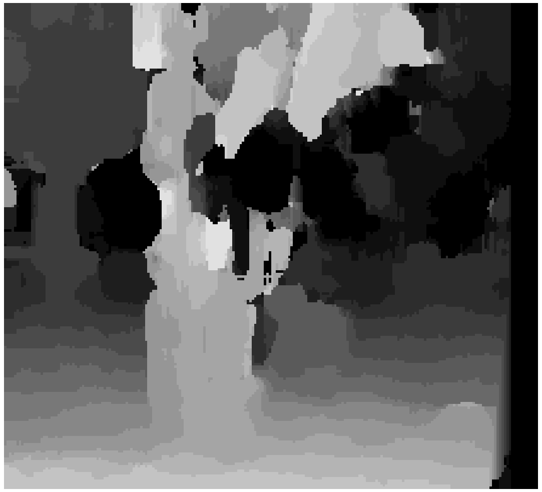


Source: Andrew Zisserman

Effect of window size

- large enough to have sufficient intensity variation
- small enough to contain only pixels with about the same disparity.

W = 3



W = 20

from Li Zhang

Uniqueness constraint

- Up to one match in right image for every point in left image

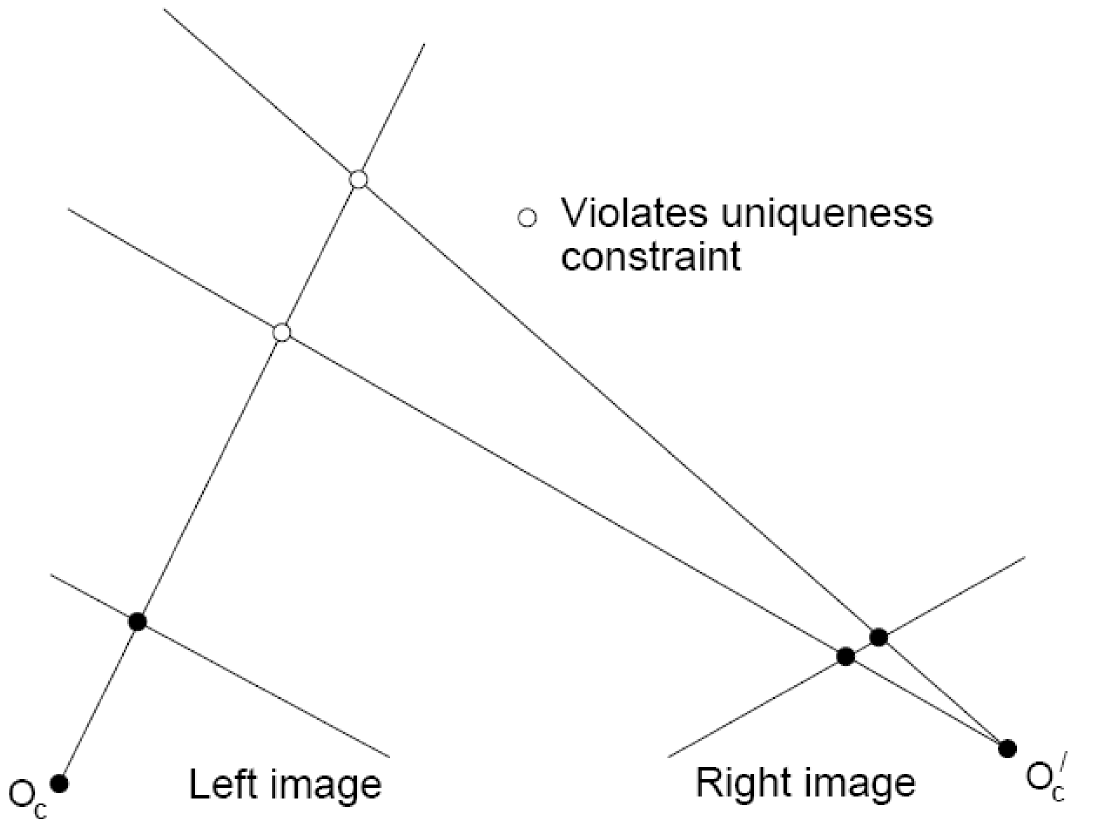
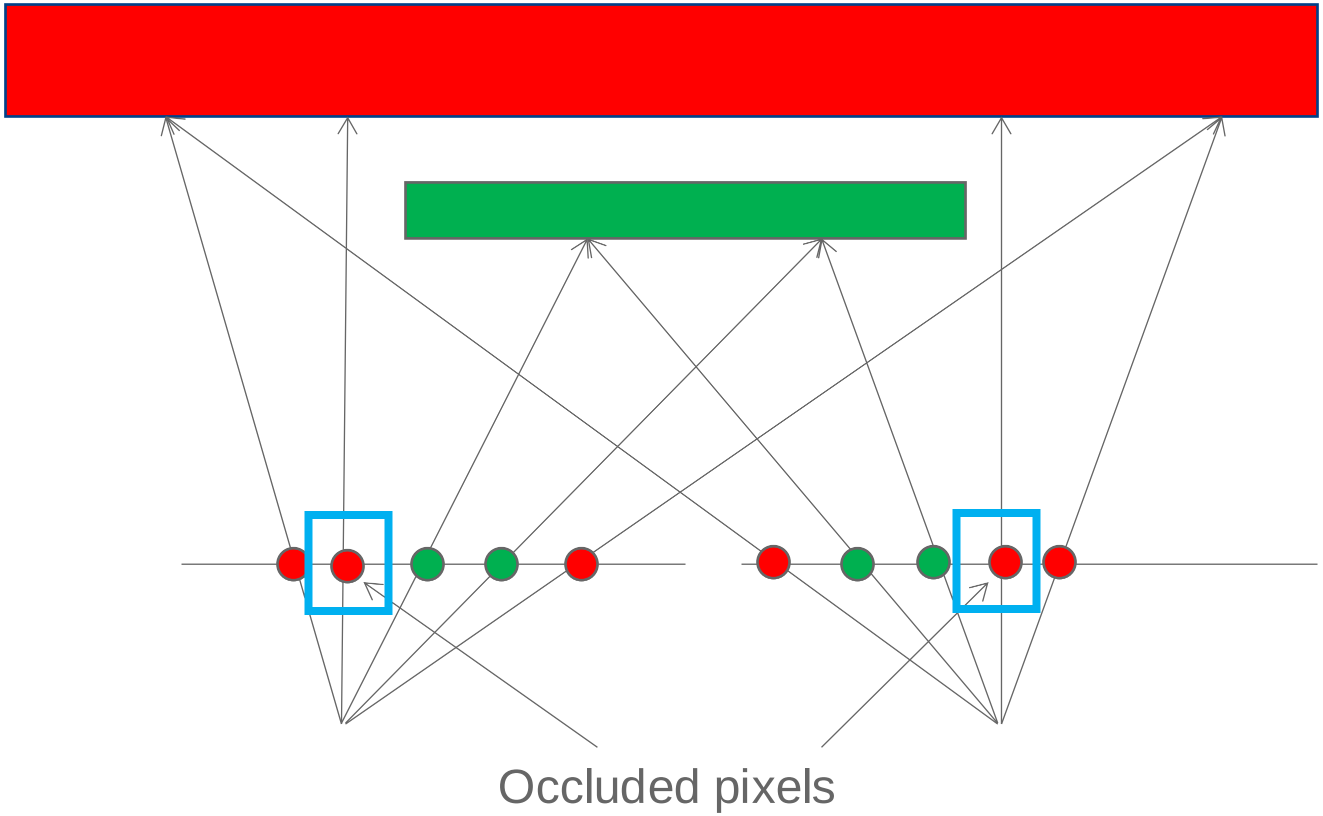


Figure from Gee & Cipolla 1999

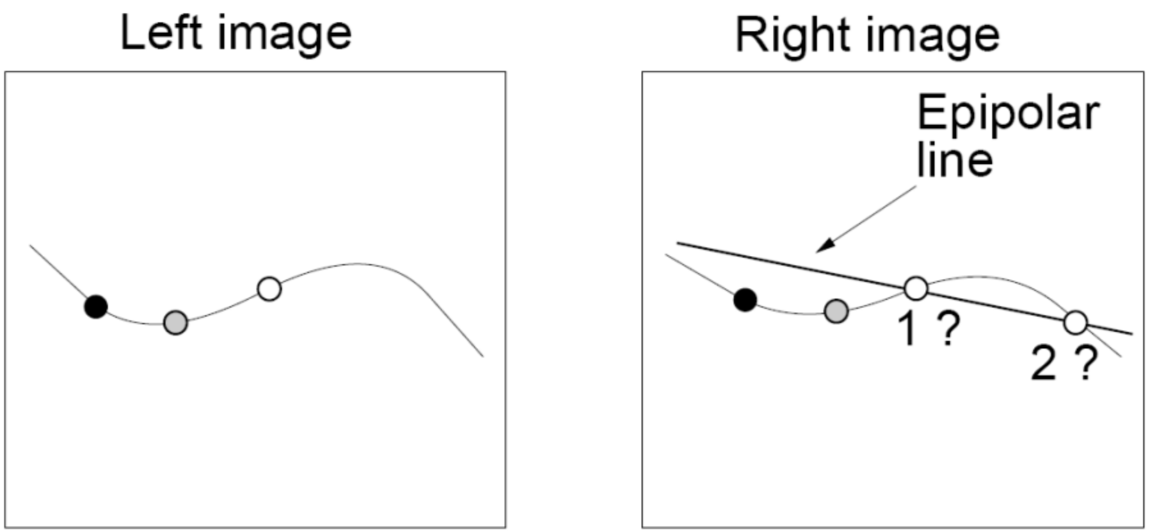
Problem: Occlusion

- Uniqueness says "up to one match" per pixel
- What if there is no match?



Disparity gradient constraint

- Assume piecewise continuous surface, so we want disparity estimates to be locally smooth



Given matches ● and ◎, point ○ in the left image must match point 1 in the right image. Point 2 would exceed the disparity gradient limit.

Figure from Gee & Cipolla 1999

Ordering constraint

- Points on **same surface** (opaque object) will be in same order in both views

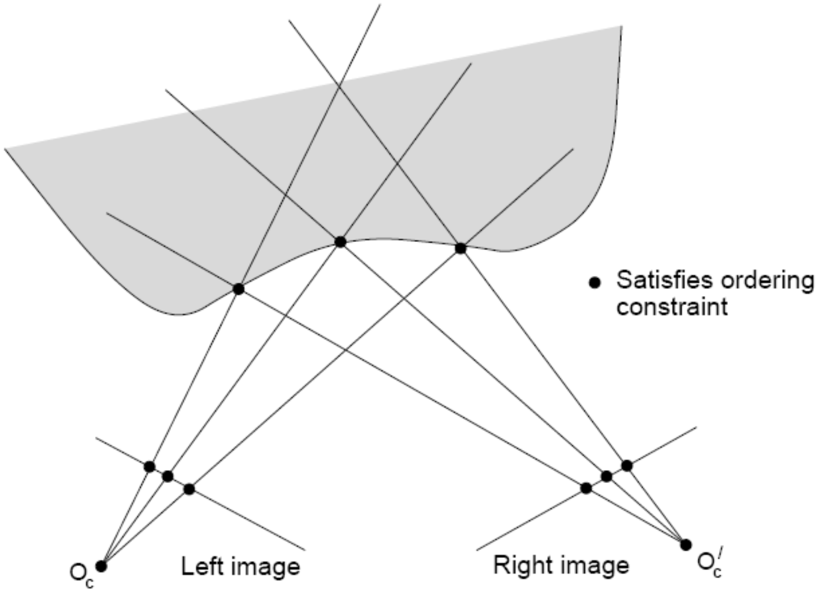
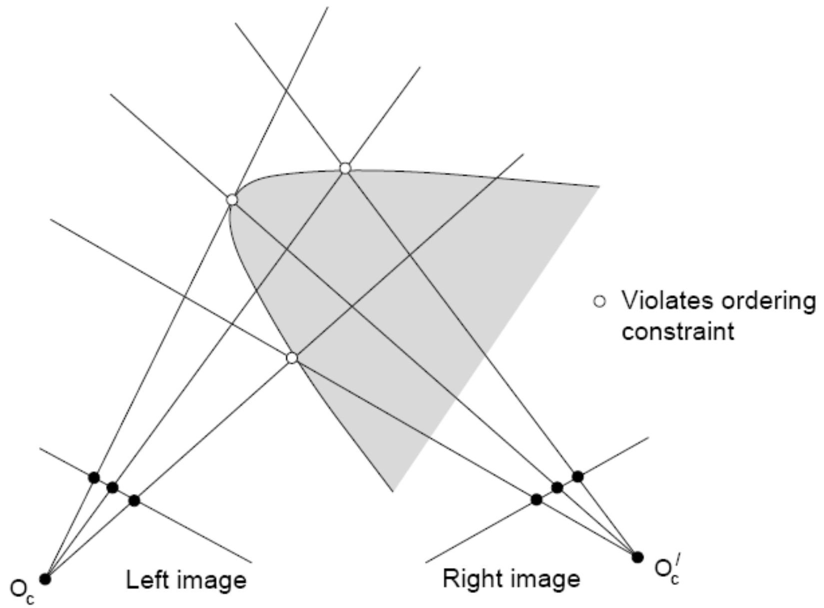
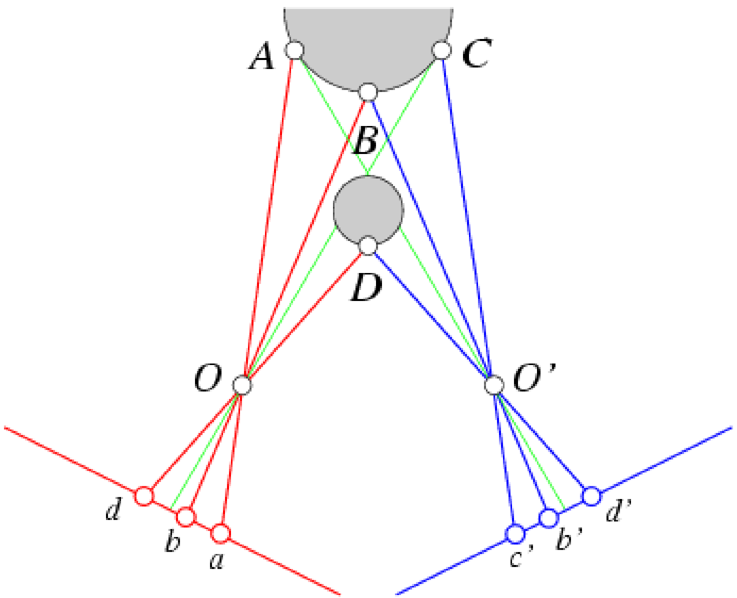


Figure from Gee & Cipolla 1999

Ordering constraint

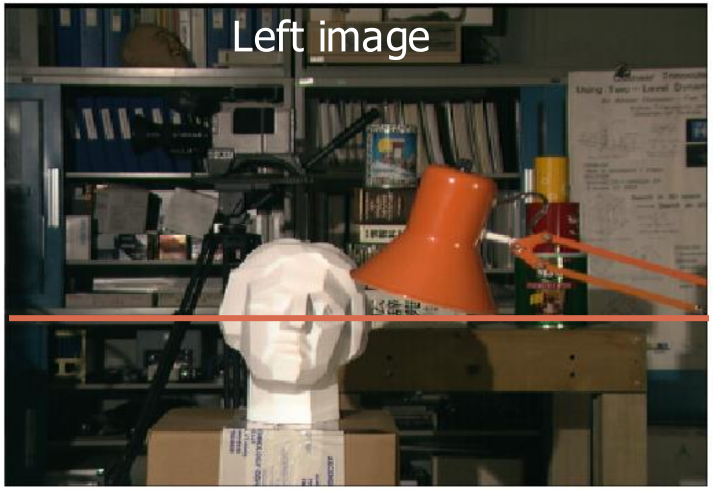
- Won’t always hold, e.g., consider transparent object, or an occluding surface

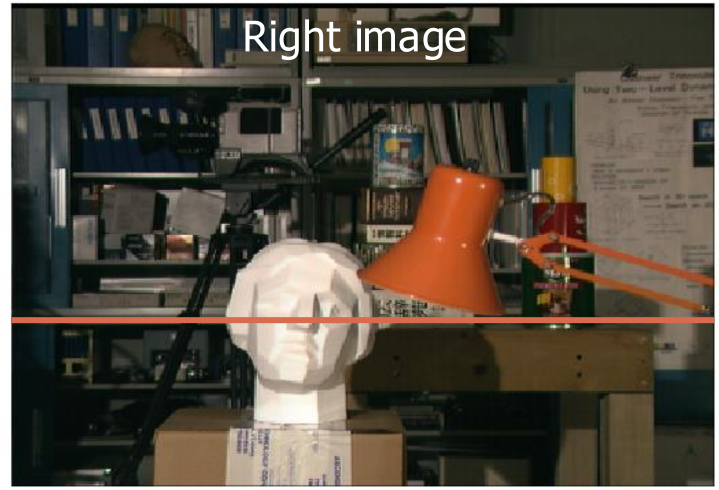


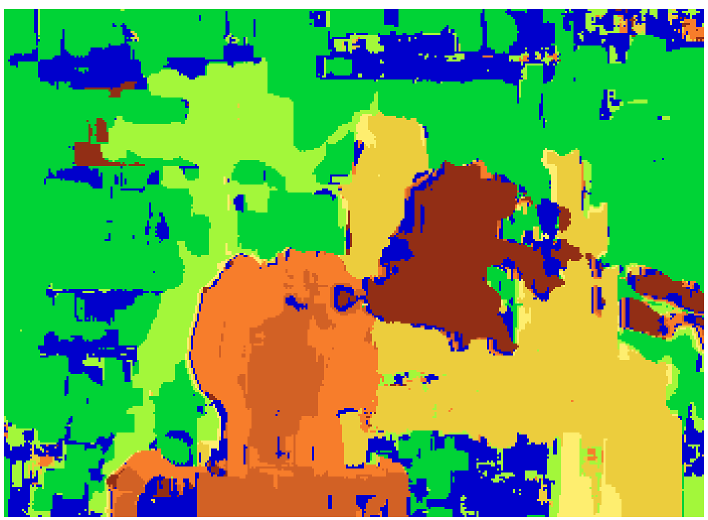


Figures from Forsyth & Ponce

Results with window search







Window-based matching
(best window size)

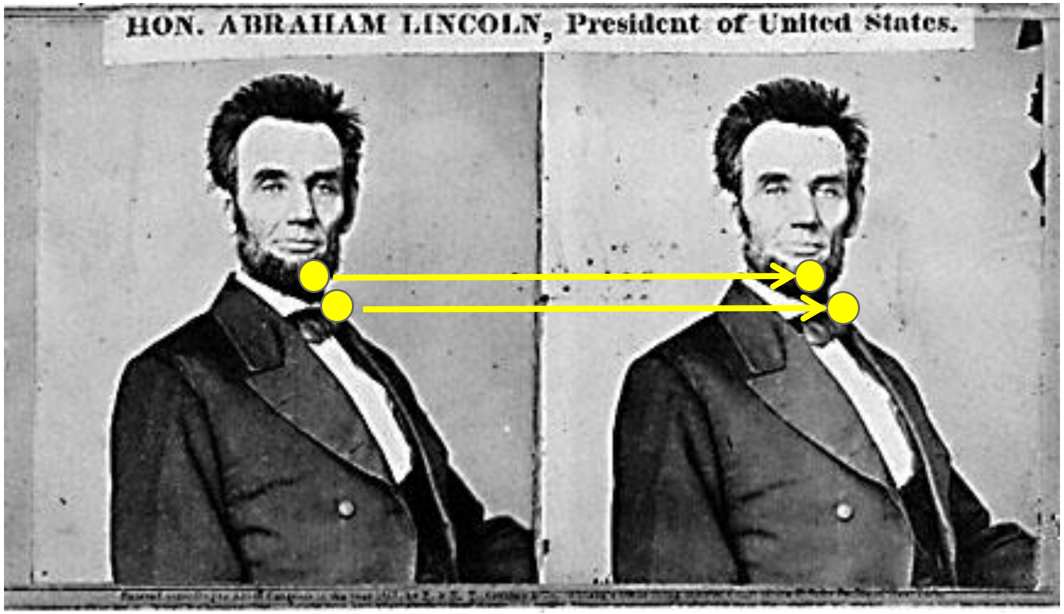
Ground truth

Better solutions

- Beyond individual correspondences estimation
- Optimize correspondence assignments jointly
  - Scanline at a time (DP)
  - Full 2D grid (graph cuts)

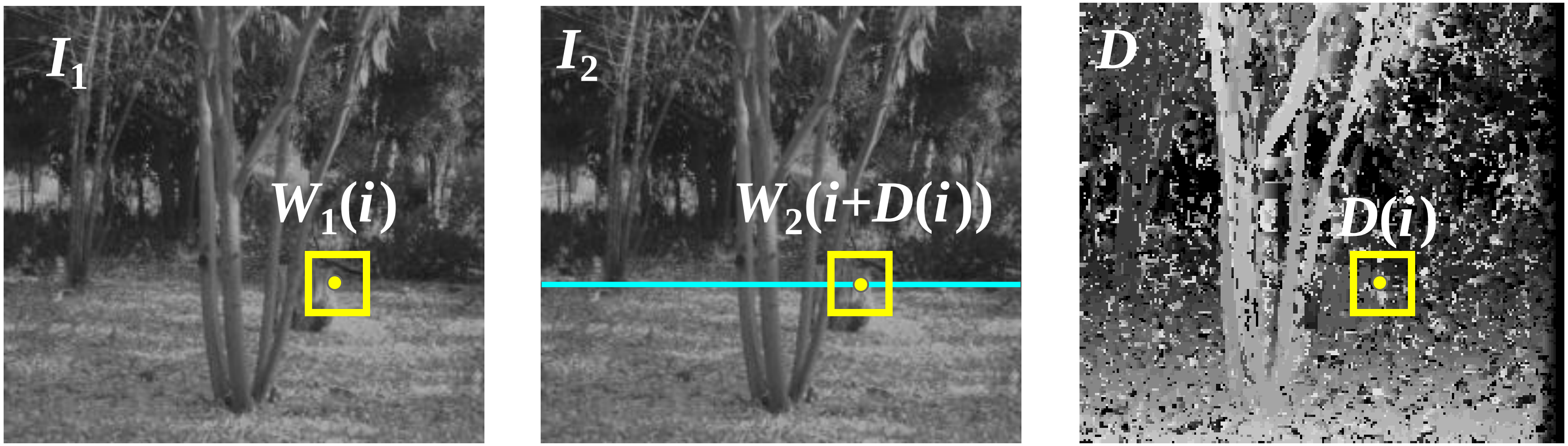
Stereo as energy minimization

- What defines a good stereo correspondence?
  - Match quality
    - Want each pixel to find a good match in the other image
  - Smoothness
    - Adjacent pixels often move about the same amount.



Stereo matching as energy minimization

- Energy functions of this form can be minimized using *graph cuts*



$$\boxed{E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)}$$

$$\boxed{E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2}$$

$$\boxed{E_{\text{smooth}} = \sum_{\text{neighbors}\, i,j} \rho\left(D(i) - D(j)\right)}$$

Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 2001

Better results…

Graph cut method

Ground truth

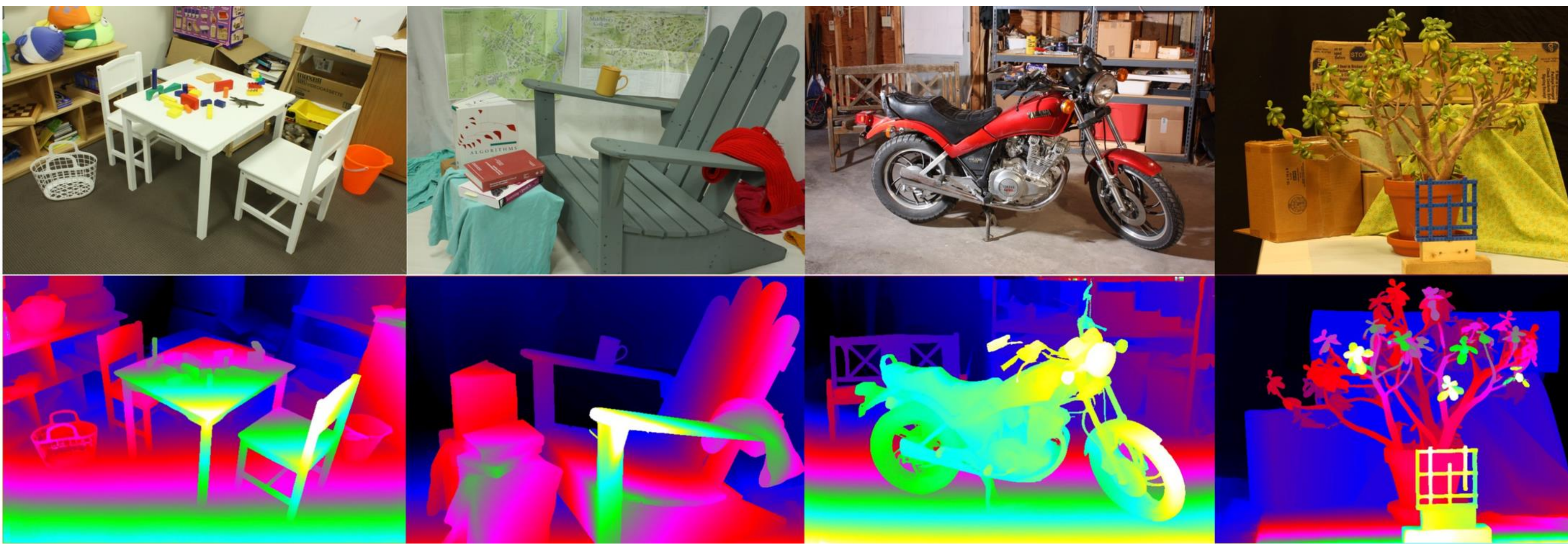
SAIR
Spatial AI & Robotics Lab

Challenges

- Low-contrast
  - Textureless image regions
- Occlusions
- Violations of brightness constancy
  - e.g., specular reflections
- Really large baselines
  - Foreshortening and appearance change
- Camera calibration errors

Data-driven Stereo Matching

- Data-driven estimated terms in energy minimization.
  - Disparity Proposal Network
  - Neural message passing



SAIR
Spatial AI & Robotics Lab

Tongfan Guan, Chen Wang, Yu-hui Liu, Neural Markov Random Field for Stereo Matching, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.