COMPUTER VISION AND PHOTOGRAMMETRY

STEREO

CONTENTS TODAY

- quick recap
  - Lab2
- Stereo rectification
- 3D from Stereo
- Dense stereo
- Linear Systems with SVD?

LET'S RECAP...

EPIPOLAR GEOMETRY

THE FUNDAMENTAL MATRIX F

CORRESPONDENCE CONDITION

The fundamental matrix satisfies the condition that for any pair of corresponding points $x \leftrightarrow x'$ in the two images

$$x'^{T} F x = 0 \qquad \left(x'^{T} l' = 0 \right)$$

THE FUNDAMENTAL MATRIX F

F is the unique 3x3 rank 2 matrix that satisfies $x'^TFx=0$ for all $x\leftrightarrow x'$

(i) **Transpose:** if F is fundamental matrix for (P,P'), then F^T is fundamental matrix for (P',P)

(ii) **Epipolar lines:** $l'=Fx$ & $l=F^Tx'$

(iii) **Epipoles:** on all epipolar lines, thus $e'^TFx=0, \forall x \Rightarrow e'^TF=0$, similarly $Fe=0$

(iv) **F** has 7 d.o.f. , i.e. 3x3-1(homogeneous)-1(rank2)

COMPUTING F

EQUATIONS

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

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

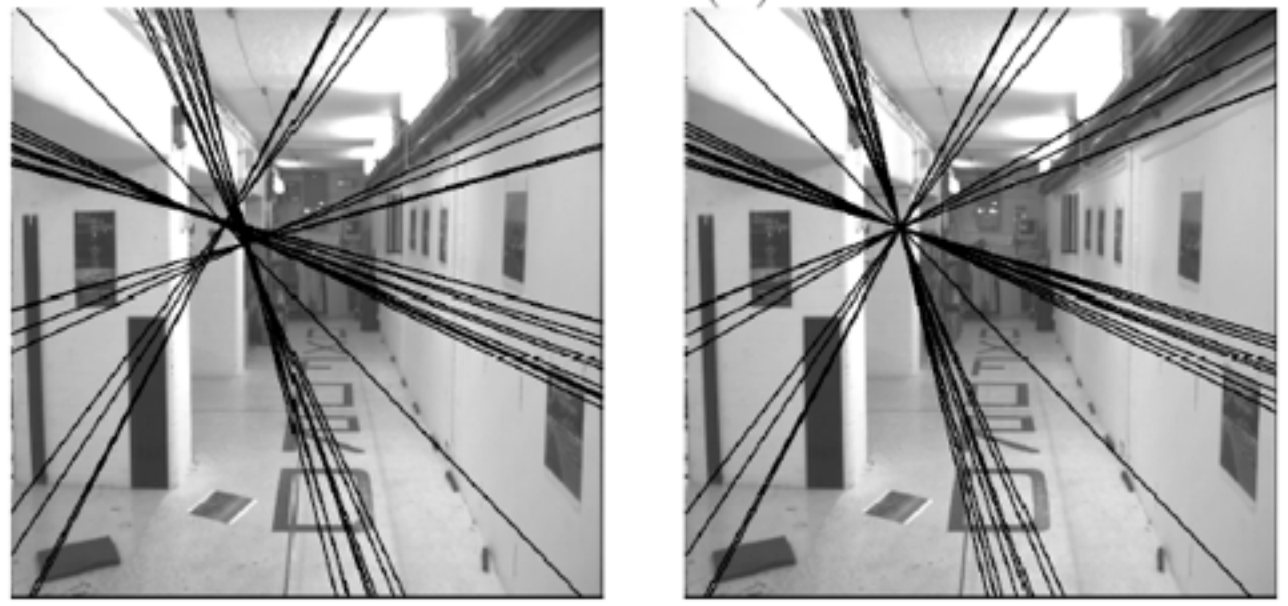
$$uu'f_{11} + uv'f_{12} + uf_{13} + vu'f_{21} + vv'f_{22} + vf_{23} + u'f_{31} + v'f_{32} + f_{33} = 0$$

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

ESTIMATING THE FUNDAMENTAL MATRIX

- 8-point algorithm
  - Least squares solution using SVD on equations from 8 pairs of correspondences
  - Enforce det(F)=0 constraint using SVD on F
- Minimize reprojection error
  - Non-linear least squares

Fundamental matrix has rank 2 : $\det(\mathtt{F}) = 0$.



Left : Uncorrected $\mathtt{F}$ – epipolar lines are not coincident.

Right : Epipolar lines from corrected $\mathtt{F}$.

Note: estimation of F (or E) is degenerate for a planar scene.

(Hartley, 1995)

THE NORMALIZED EIGHT-POINT ALGORITHM

- Center the image data at the origin, and scale it so the mean squared distance between the origin and the data points is 2 pixels
- Use the eight-point algorithm to compute $\boldsymbol{F}$ from the normalized points
- Enforce the rank-2 constraint (for example, take SVD of $\boldsymbol{F}$ and throw out the smallest singular value)
- Transform fundamental matrix back to original units: if $\boldsymbol{T}$ and $\boldsymbol{T}'$ are the normalizing transformations in the two images, than the fundamental matrix in original coordinates is $\boldsymbol{T}'^{T}\,\boldsymbol{F}\,\boldsymbol{T}$

THE SINGULARITY CONSTRAINT

$$e'^{T} F = 0 \qquad Fe = 0 \qquad \det F = 0 \qquad \operatorname{rank} F = 2$$

SVD from linearly computed F matrix (rank 3)

$$F = U\begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix}V^T = U_1\sigma_1V_1^T + U_2\sigma_2V_2^T + U_3\sigma_3V_3^T$$

Compute closest rank-2 approximation $\min\|F - F'\|_F$

$$F' = U\begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & 0 \end{bmatrix}V^T = U_1\sigma_1V_1^T + U_2\sigma_2V_2^T$$

THE ESSENTIAL MATRIX

~fundamental matrix for calibrated cameras (remove K)

$$E = [t]_\times R = R[R^T t]_\times$$

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

$$E = K'^T F K$$

5 d.o.f. (3 for R; 2 for t up to scale)

E is essential matrix if and only if
two singularvalues are equal (and third=0)

$$E = U\mathrm{diag}(1,1,0)V^T$$

$$F = K'^{-T} E K^{-1}$$
$$E = K'^T F K$$

THE ESSENTIAL MATRIX

$$\mathtt{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathtt{Z} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

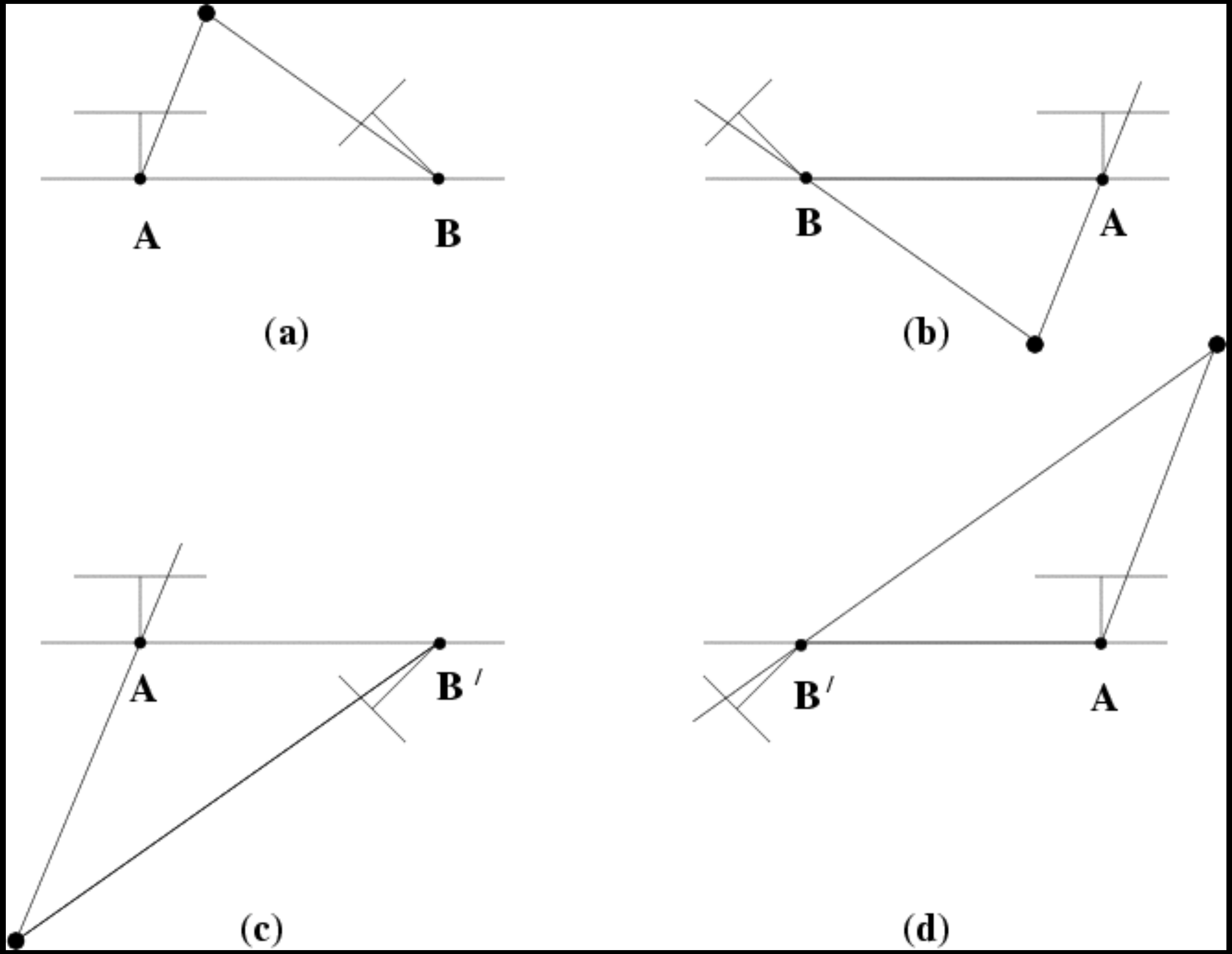
Result 9.18. *Suppose that the SVD of* $\mathtt{E}$ *is* $\mathtt{U}\,\mathrm{diag}(1,1,0)\mathtt{V}^\mathsf{T}$. *Using the notation of (9.13), there are (ignoring signs) two possible factorizations* $\mathtt{E} = \mathtt{SR}$ *as follows:*

$$\mathtt{S} = \mathtt{UZU}^\mathsf{T} \quad \mathtt{R} = \mathtt{UWV}^\mathsf{T} \quad \textit{or} \quad \mathtt{UW}^\mathsf{T}\mathtt{V}^\mathsf{T}\ . \tag{9.14}$$

Result 9.19. *For a given essential matrix* $\mathtt{E} = \mathtt{U}\,\mathrm{diag}(1,1,0)\mathtt{V}^\mathsf{T}$, *and first camera matrix* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, *there are four possible choices for the second camera matrix* $\mathtt{P}'$, *namely*

$$\mathtt{P}' = [\mathtt{UWV}^\mathsf{T} \mid +\mathbf{u}_3] \text{ or } [\mathtt{UWV}^\mathsf{T} \mid -\mathbf{u}_3] \text{ or } [\mathtt{UW}^\mathsf{T}\mathtt{V}^\mathsf{T} \mid +\mathbf{u}_3] \text{ or } [\mathtt{UW}^\mathsf{T}\mathtt{V}^\mathsf{T} \mid -\mathbf{u}_3].$$

FOUR POSSIBLE RECONSTRUCTIONS FROM E



(ONLY ONE SOLUTION WHERE POINTS IS IN FRONT OF BOTH CAMERAS)

TRIANGULATION

$$
\begin{aligned}
x_j &= \frac{p_{00}^{(j)} X + p_{01}^{(j)} Y + p_{02}^{(j)} Z + p_{03}^{(j)} W}{p_{20}^{(j)} X + p_{21}^{(j)} Y + p_{22}^{(j)} Z + p_{23}^{(j)} W} \\
y_j &= \frac{p_{10}^{(j)} X + p_{11}^{(j)} Y + p_{12}^{(j)} Z + p_{13}^{(j)} W}{p_{20}^{(j)} X + p_{21}^{(j)} Y + p_{22}^{(j)} Z + p_{23}^{(j)} W},
\end{aligned}
$$

$$
\begin{aligned}
x(\mathbf{p}^{3\top}\mathbf{X}) - (\mathbf{p}^{1\top}\mathbf{X}) &= 0 \\
y(\mathbf{p}^{3\top}\mathbf{X}) - (\mathbf{p}^{2\top}\mathbf{X}) &= 0
\end{aligned}
$$

$$
\mathtt{A} = \begin{bmatrix} x\mathbf{p}^{3\top} - \mathbf{p}^{1\top} \\ y\mathbf{p}^{3\top} - \mathbf{p}^{2\top} \\ x'\mathbf{p}'^{3\top} - \mathbf{p}'^{1\top} \\ y'\mathbf{p}'^{3\top} - \mathbf{p}'^{2\top} \end{bmatrix}
$$

Solve the sytem with the same methods as before

FROM EPIPOLAR GEOMETRY TO CAMERA CALIBRATION

- Estimating the fundamental matrix is known as “weak calibration”
- We know the calibration matrices of the two cameras, we can estimate the essential matrix: $E = K^T F K'$
- The essential matrix gives us the relative rotation and translation between the cameras, or their extrinsic parameters

GIVEN 2D CORRESPONDENCES

- We know how to calibrate a camera given 3D-2D points
  - Get P
  - From P we know hot to get K, R, C
- We can compute the Fundamental Matrix, and the Essential Matrix
- We can compute 3D points for this correspondence
  - If we have K, we get the actual 3D points
- This are sparse 3D points

STEREO DENSE RECONSTRUCTION

SPECIAL CASE: DENSE BINOCULAR STEREO

- Fuse a calibrated binocular stereo pair to produce a depth image

IMAGE 1

IMAGE 2

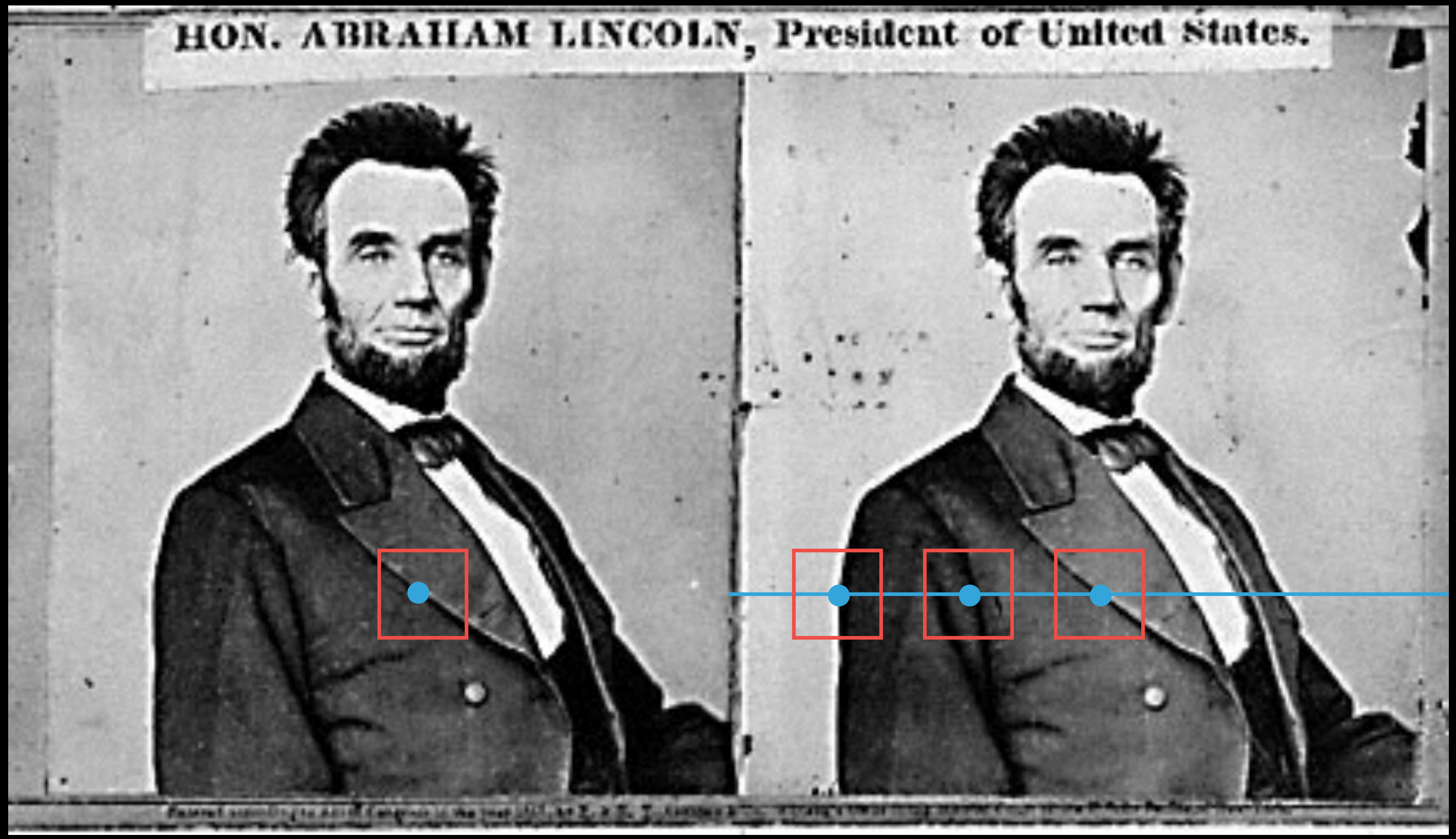
DENSE DEPTH MAP

MANY OF THESE SLIDES ADAPTED FROM STEVE SEITZ AND LANA LAZEBNIK

BASIC STEREO MATCHING ALGORITHM

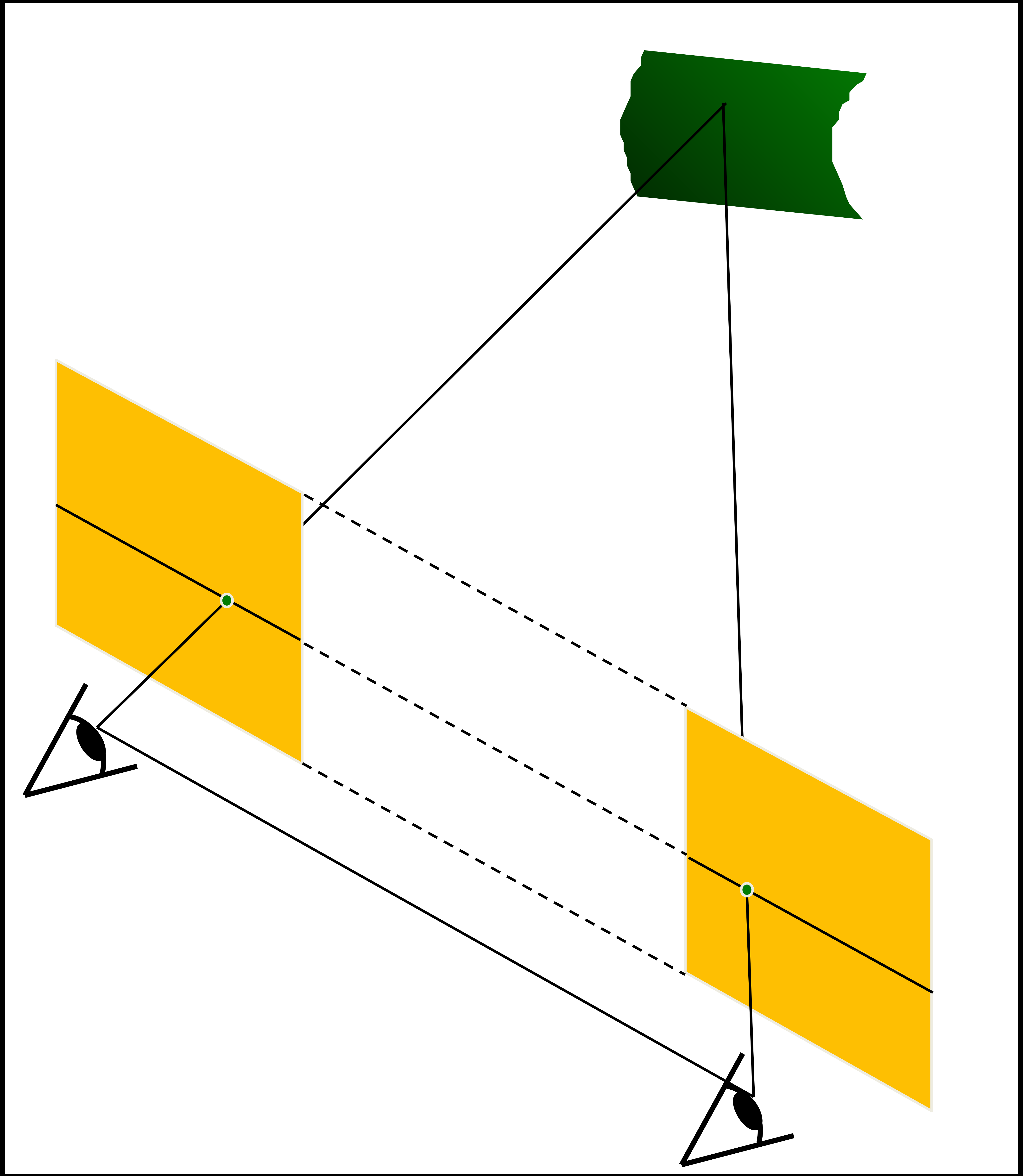
- For each pixel in the first image
  - Find corresponding epipolar line in the right image
  - Examine all pixels on the epipolar line and pick the best match
  - Triangulate the matches to get depth information

- Simplest case: epipolar lines are scanlines
  - When does this happen?

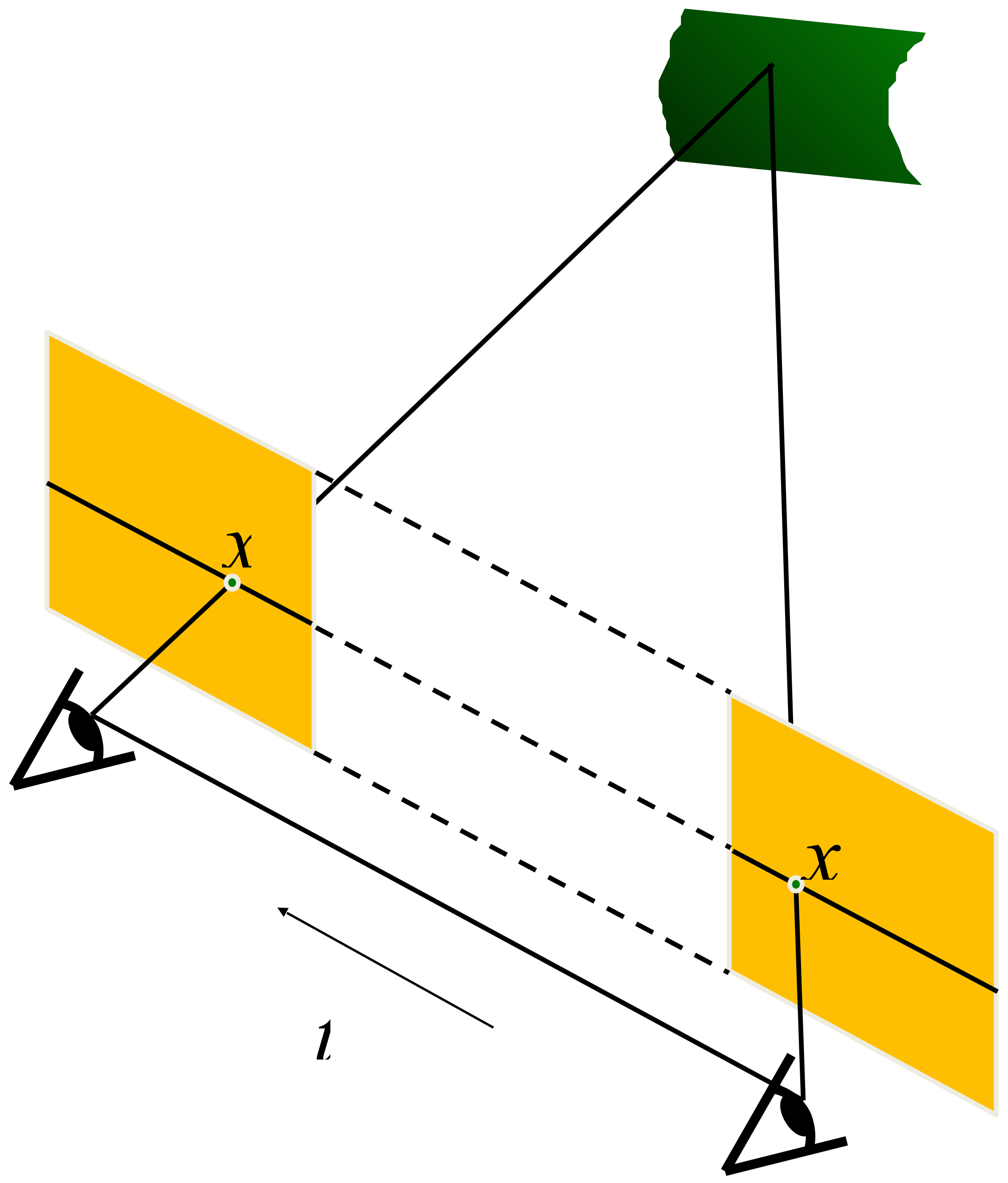


SIMPLEST CASE: PARALLEL IMAGES

- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths are the same
- Then, epipolar lines fall along the horizontal scan lines of the images



SPECIAL CASE OF FUNDAMENTAL MATRIX



Epipolar constraint:

$$x^T E x' = 0, \quad E = [t_\times]R$$

$$R = I \qquad t = (T, 0, 0)$$

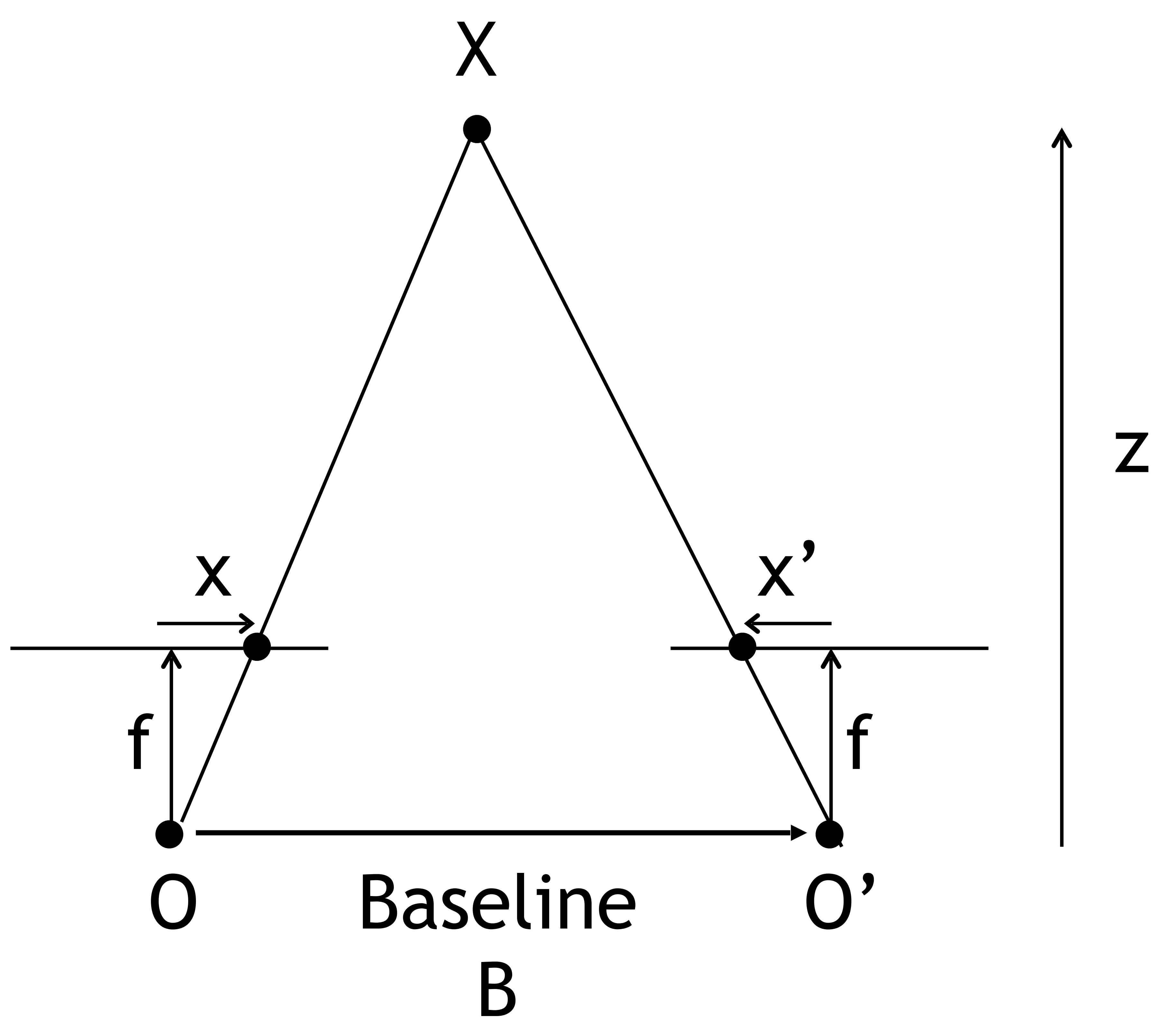
$$E = [t_\times]R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix}$$

$$\begin{pmatrix} u & v & 1 \end{pmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0 \qquad \begin{pmatrix} u & v & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -T \\ Tv' \end{pmatrix} = 0 \qquad Tv = Tv'$$

The y-coordinates of corresponding points are the same!

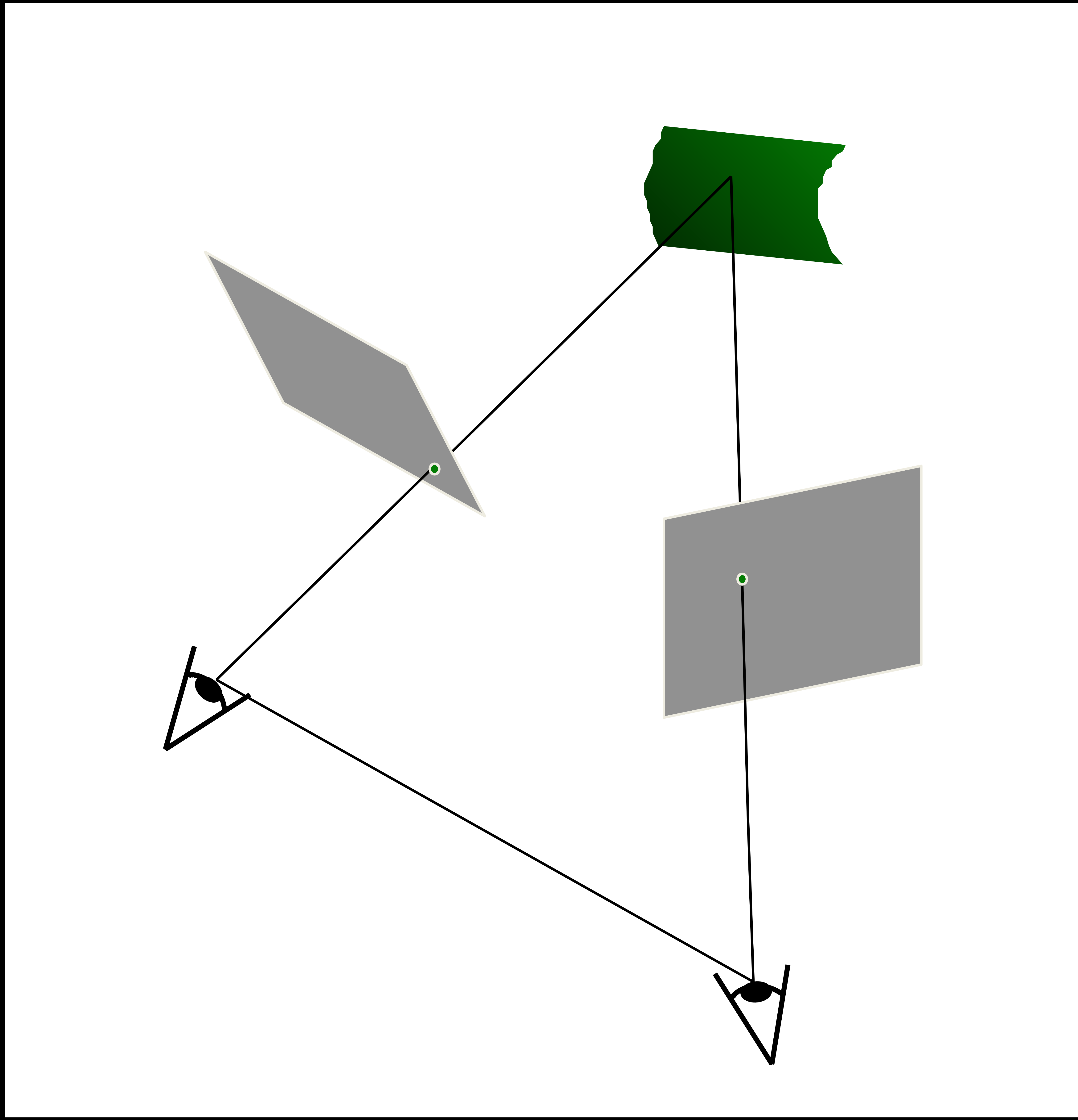
DEPTH FROM DISPARITY

- Disparity is inversely proportional to depth!



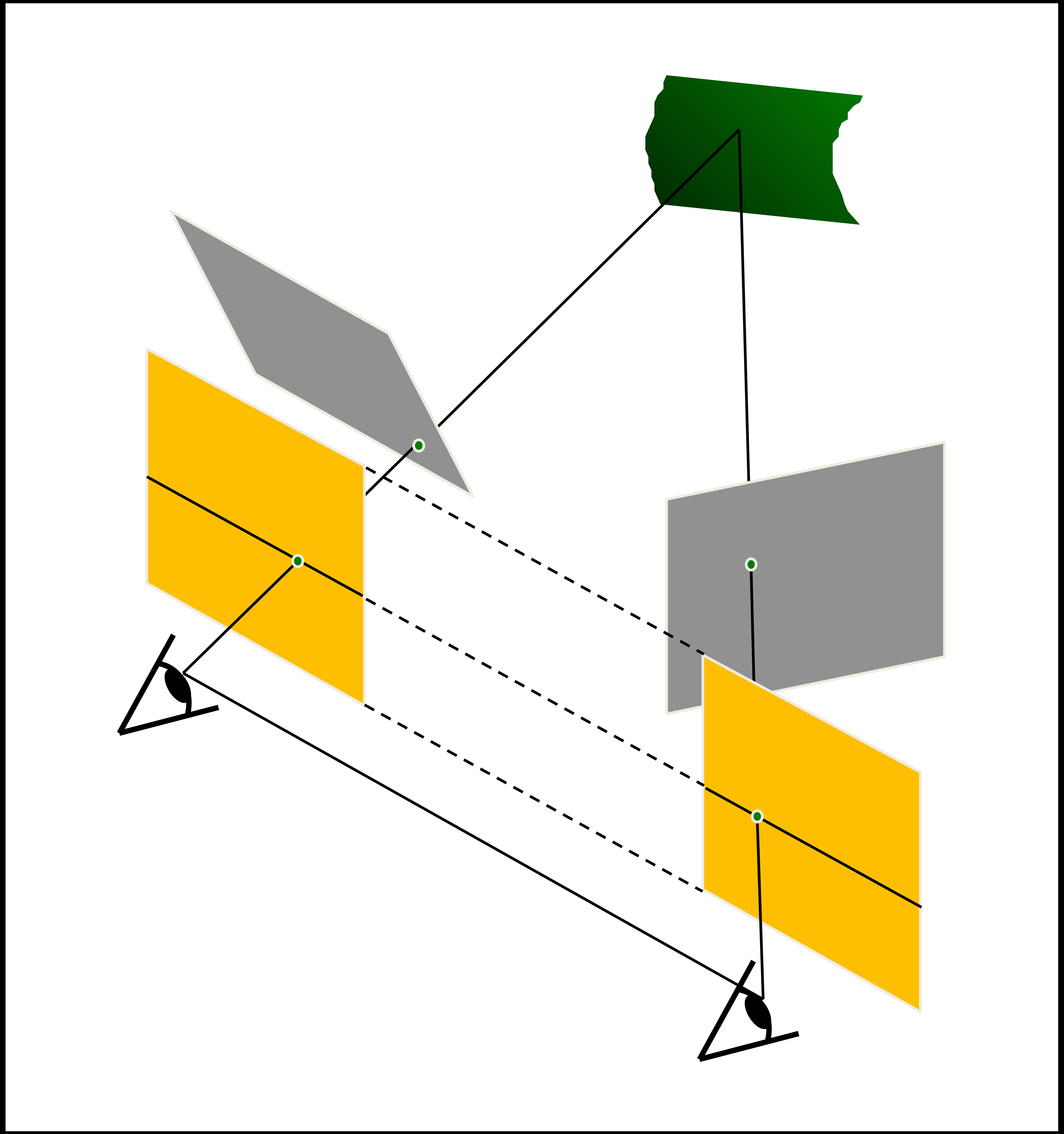
$$disparity = x - x' = \frac{B \cdot f}{z}$$

STEREO IMAGE RECTIFICATION

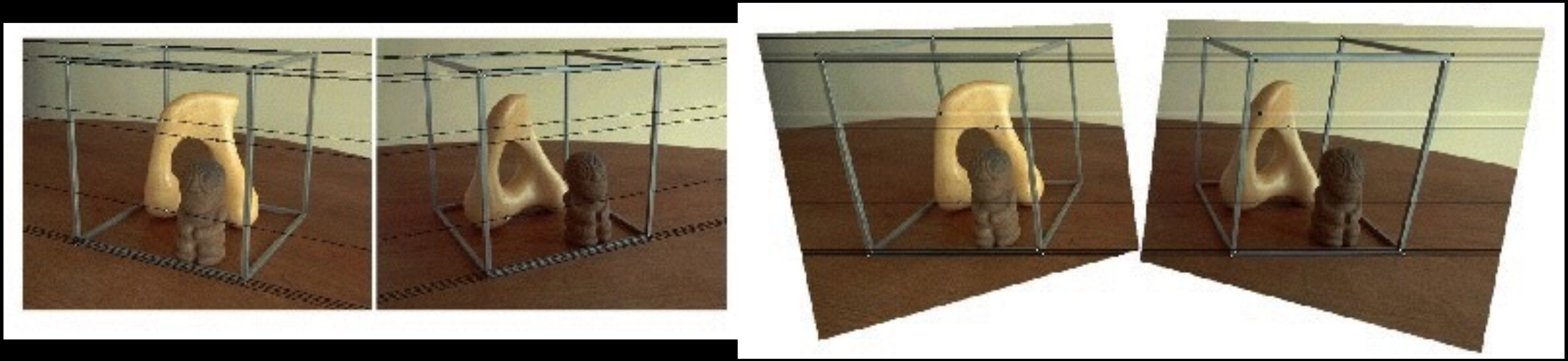


STEREO IMAGE RECTIFICATION

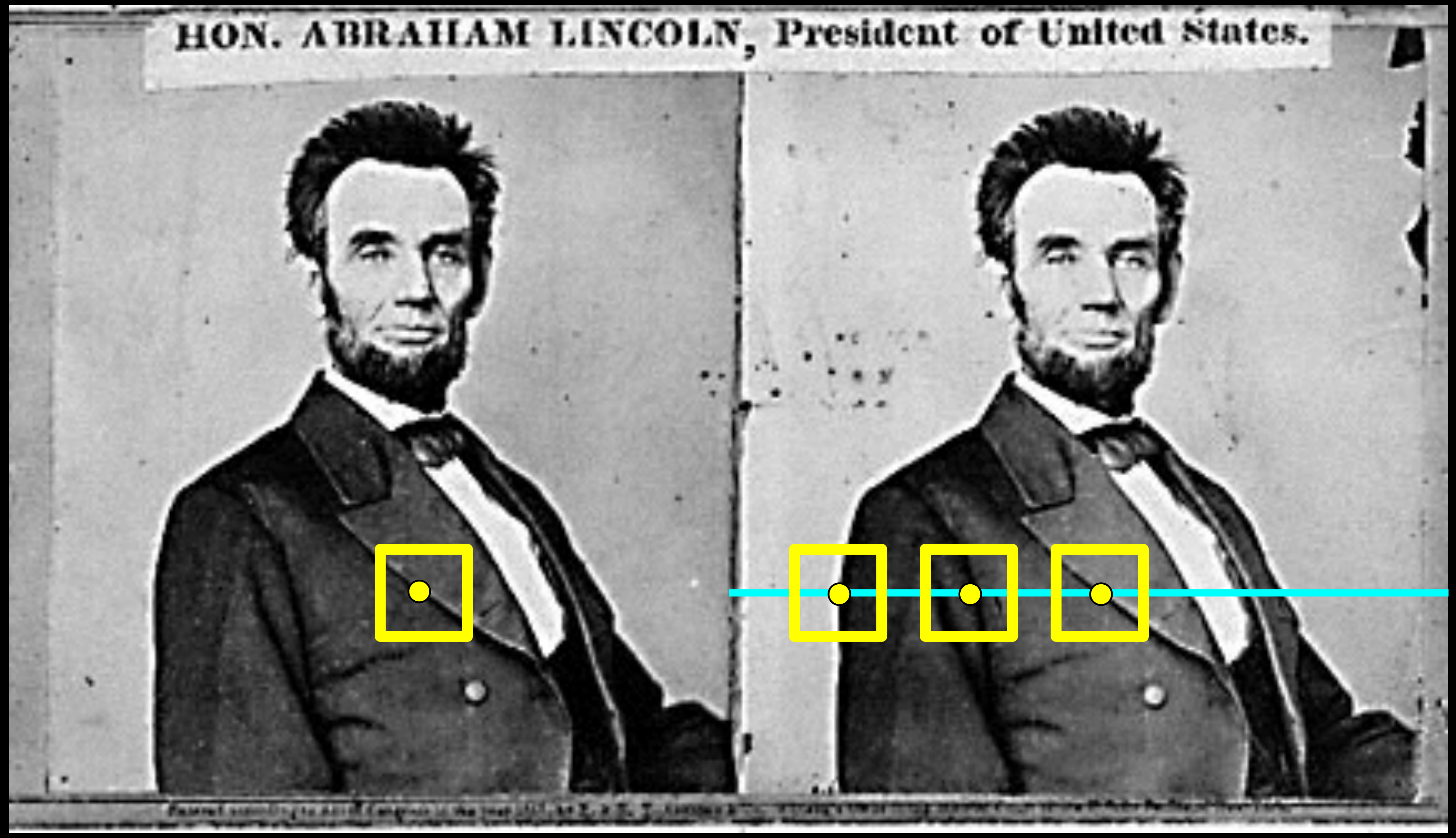
- Reproject image planes onto a common plane parallel to the line between optical centers
- Pixel motion is horizontal after this transformation
- Two homographies (3x3 transform), one for each input image reprojection
- C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition, 1999.



RECTIFICATION EXAMPLE



BASIC STEREO MATCHING ALGORITHM

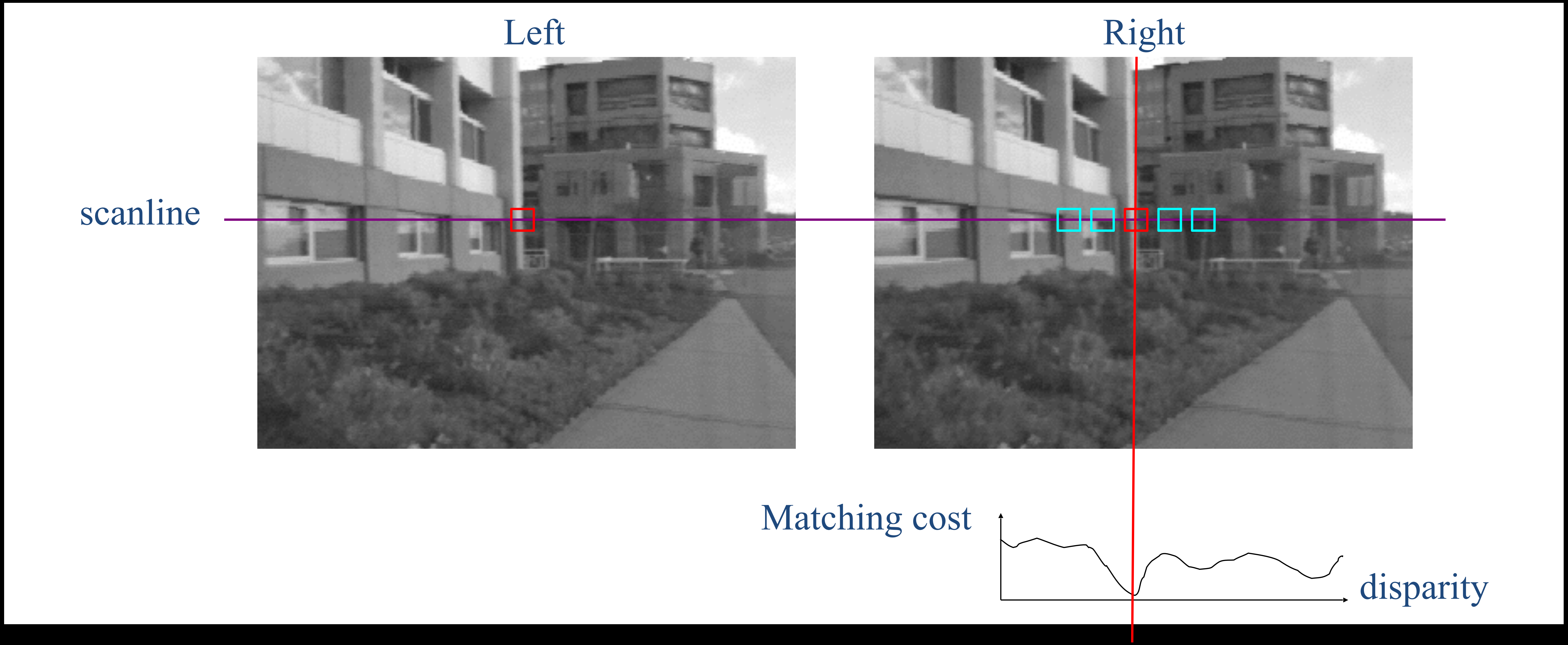


- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x′
  - Compute disparity x-x′ and set depth(x) = 1/(x-x′)

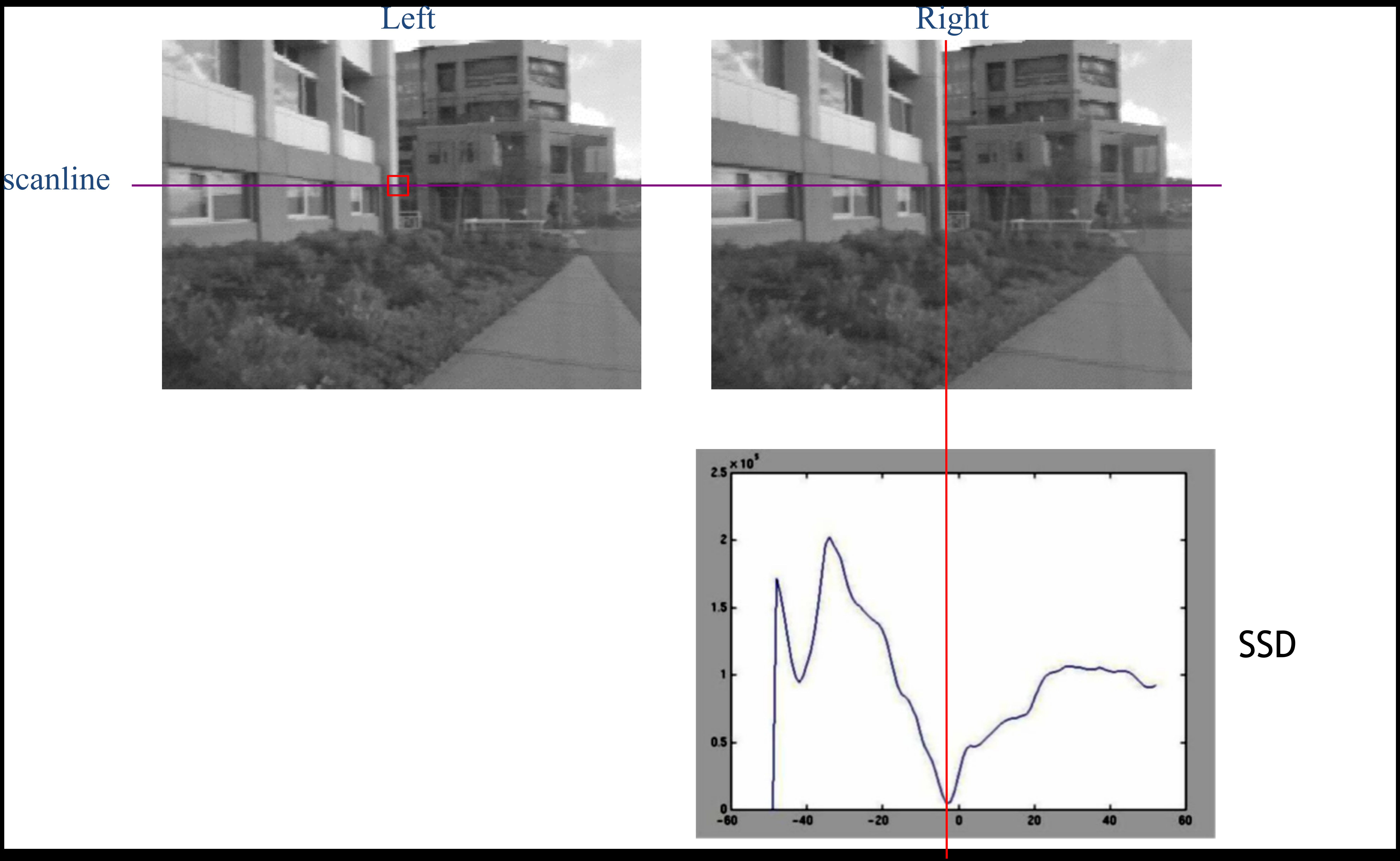
CORRESPONDENCE SEARCH

- Slide a window along the right scanline and compare contents of that window with the reference window in the left image

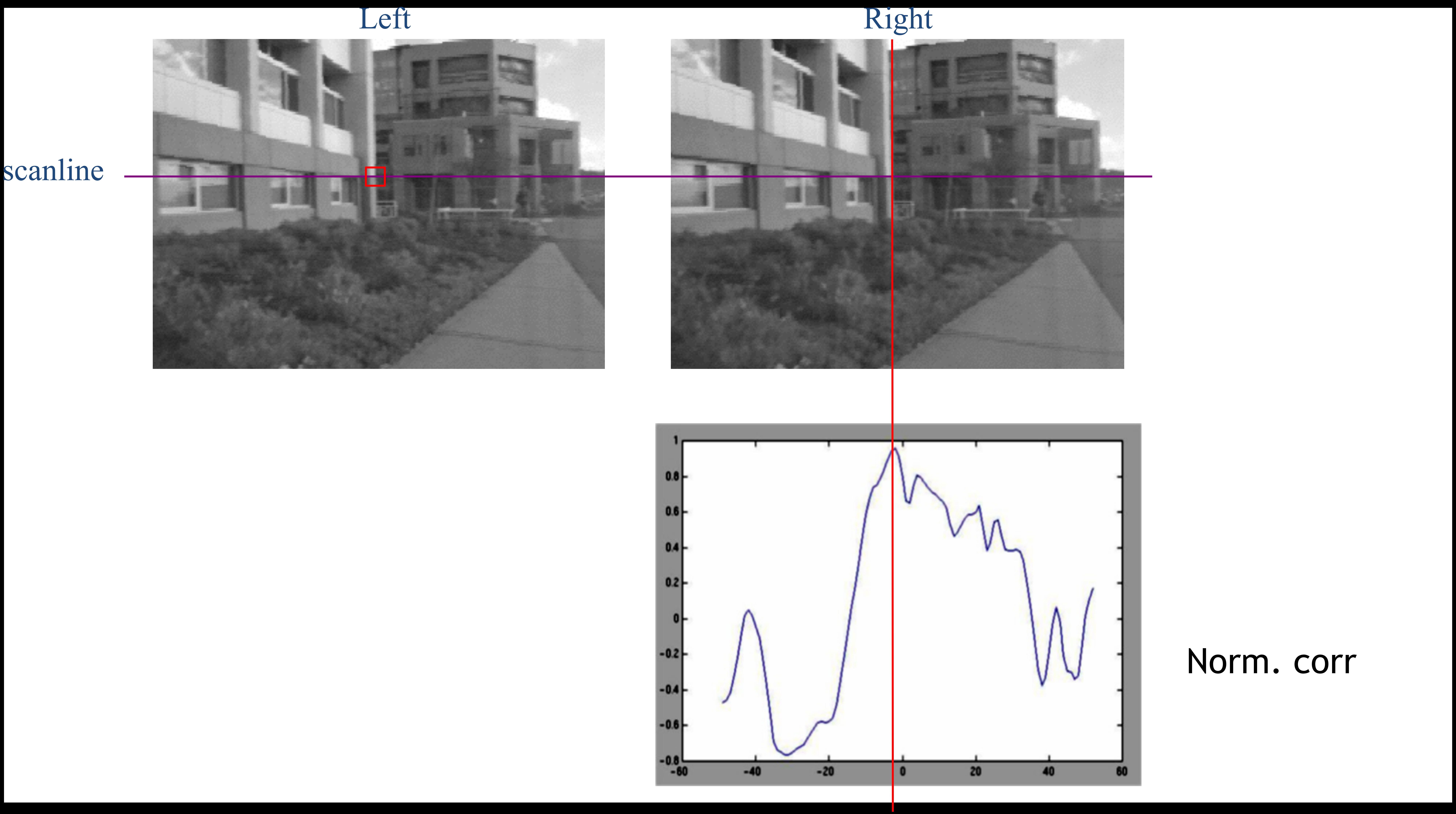
- Matching cost: SSD or normalized correlation



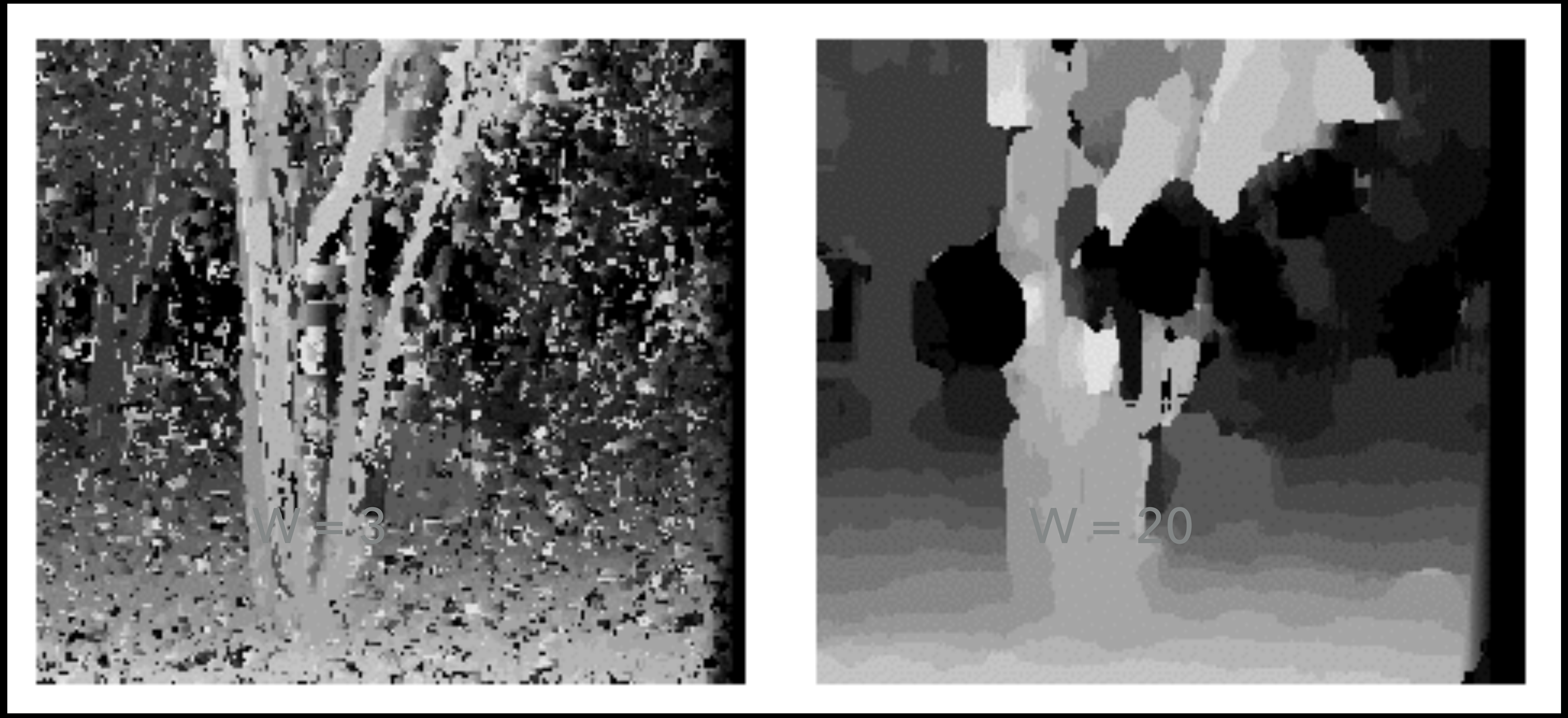
CORRESPONDENCE SEARCH



CORRESPONDENCE SEARCH



EFFECT OF WINDOW SIZE



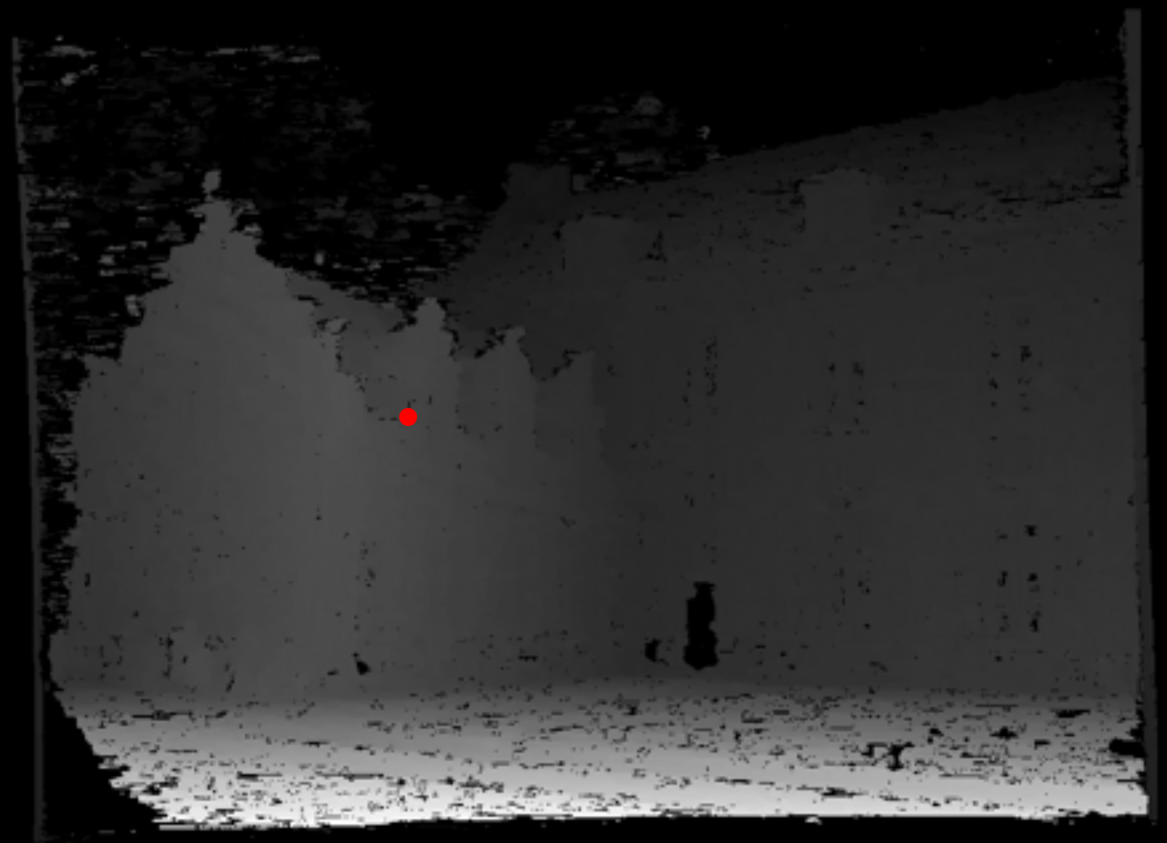
- Smaller window
  - More detail
  - More noise
- Larger window
  - Smoother disparity maps
  - Less detail

DISPARITY MAP

IMAGE I(X,Y)

DISPARITY MAP D(X,Y)

IMAGE I´(X´,Y´)

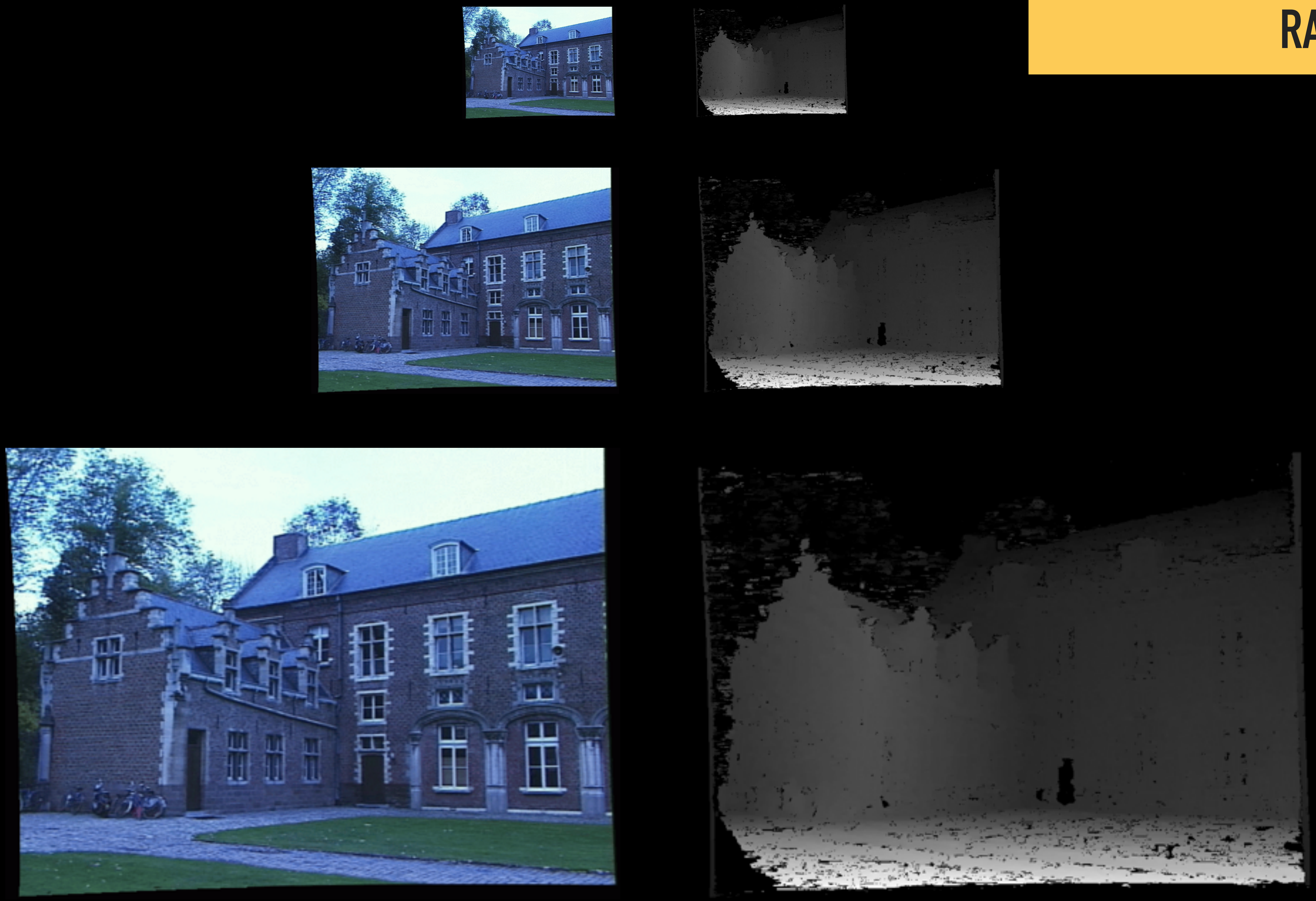


$$(X',Y') = (X + D(X,Y), Y)$$

HIERARCHICAL STEREO MATCHING

ALLOWS FASTER COMPUTATION
DEALS WITH LARGE DISPARITY RANGES

DOWNSAMPLING
(GAUSSIAN PYRAMID)



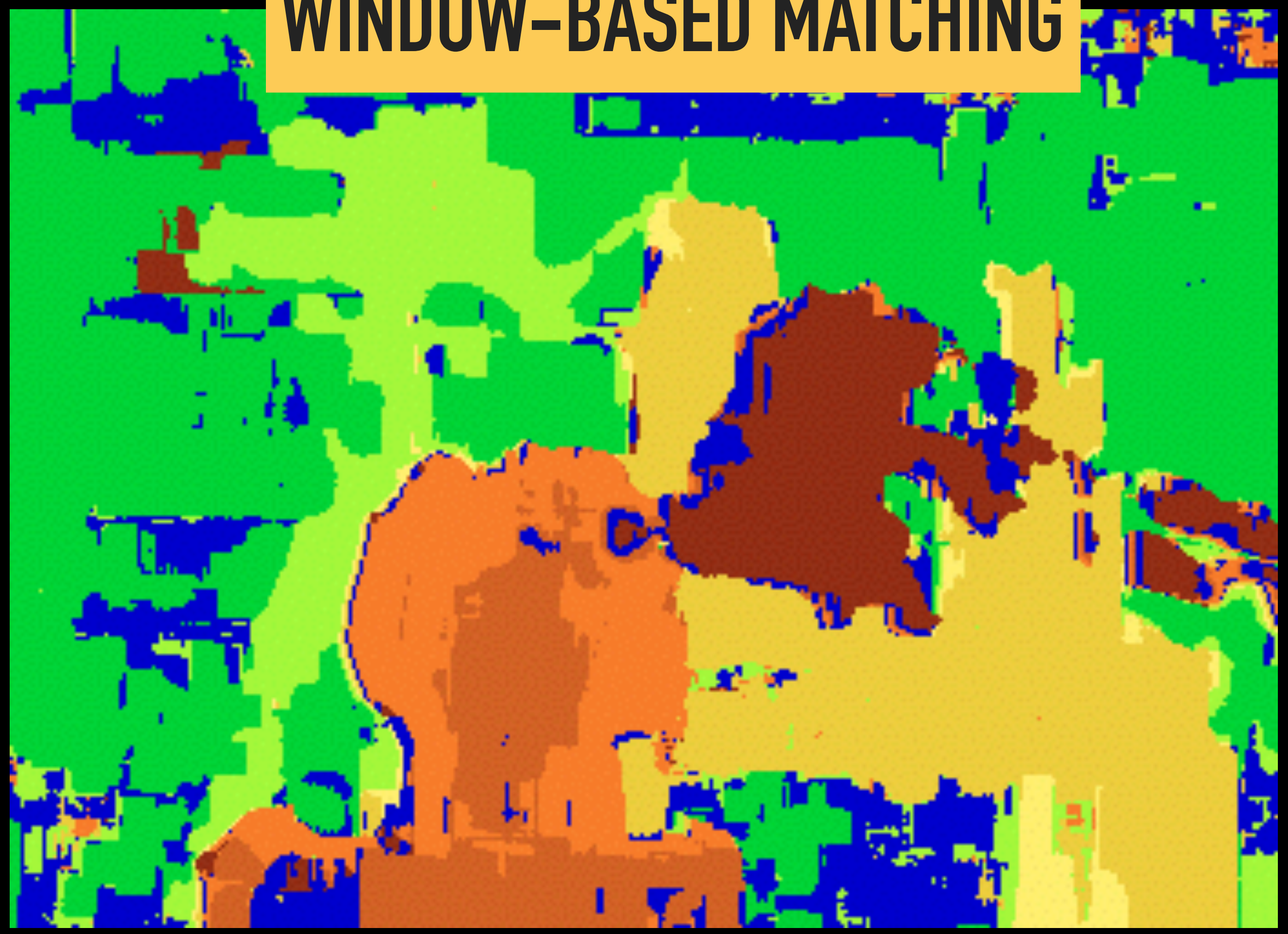
DISPARITY PROPAGATION

(FALKENHAGEN´97;VAN MEERBERGEN,VERGAUWEN,POLLEFEYS,VANGOOL IJCV‘02)

RESULTS WITH WINDOW SEARCH

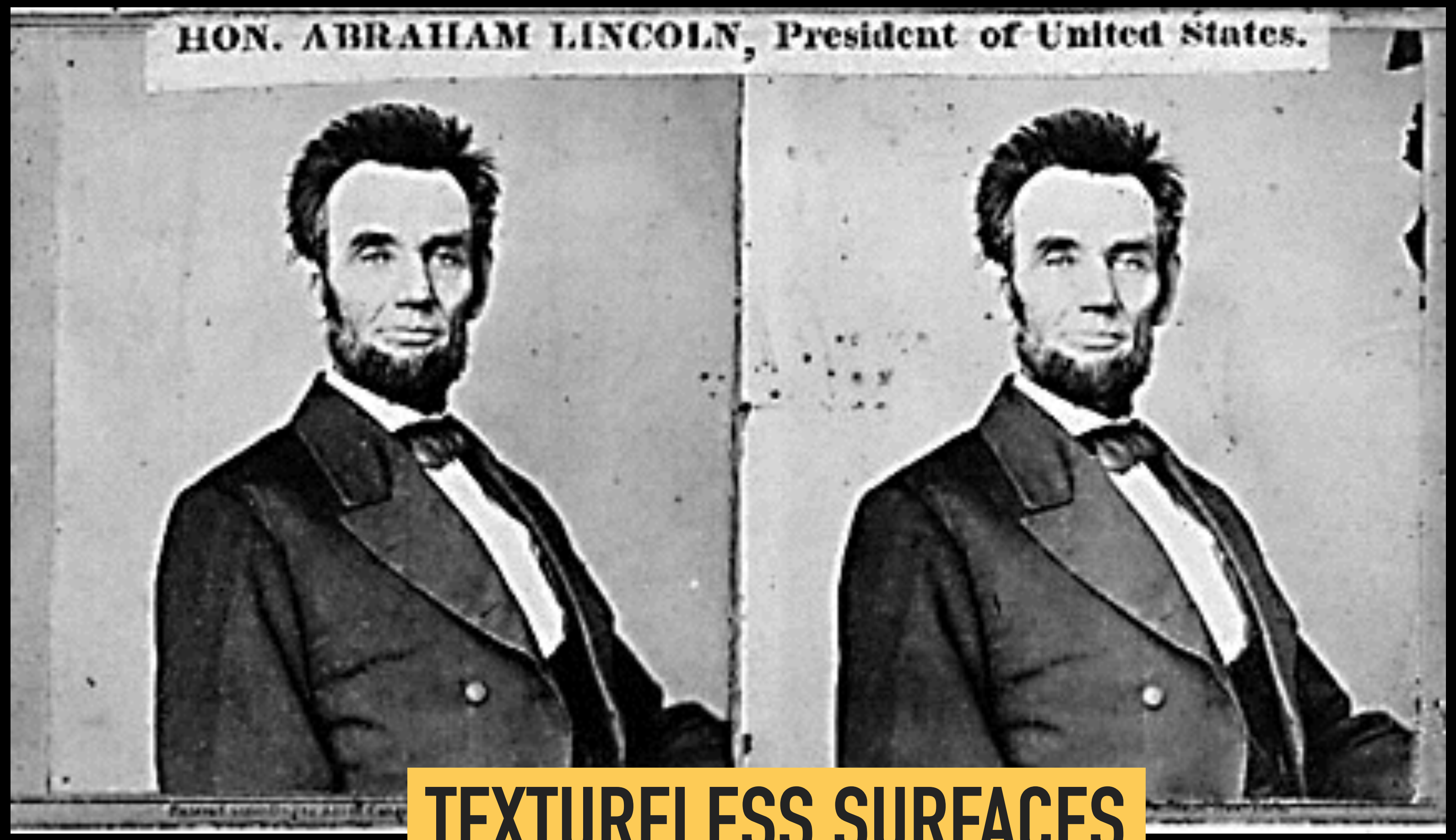
DATA

WINDOW-BASED MATCHING



GROUND TRUTH

FAILURES OF CORRESPONDENCE SEARCH



TEXTURELESS SURFACES

OCCLUSIONS, REPETITION

NON-LAMBERTIAN SURFACES, SPECULARITIES

HOW CAN WE IMPROVE WINDOW-BASED MATCHING?

- So far, matches are independent for each point

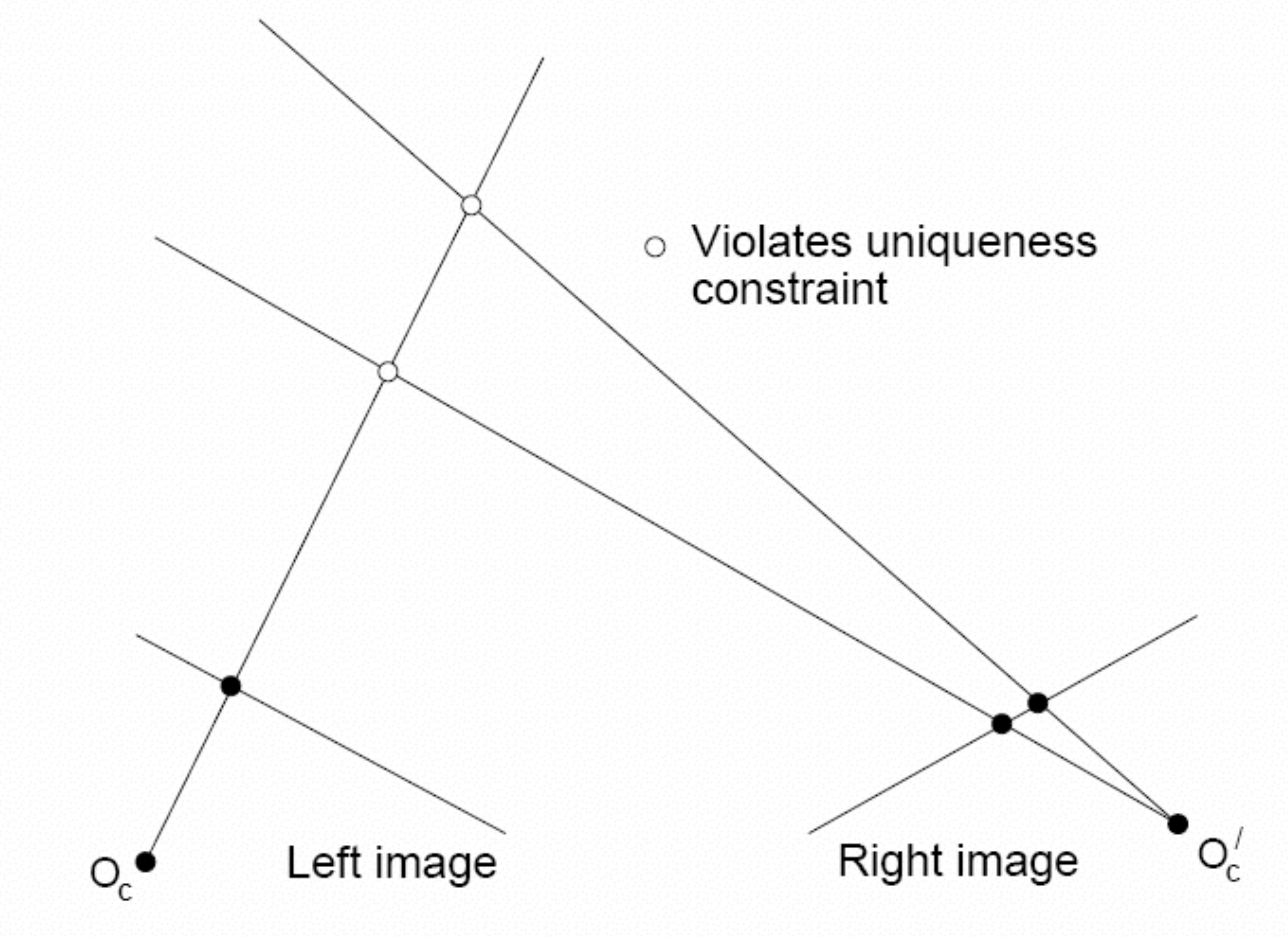
- What constraints or priors can we add?

HOW CAN WE IMPROVE WINDOW-BASED MATCHING?

- The similarity constraint is local (each reference window is matched independently)
- Need to enforce non-local correspondence constraints or priors

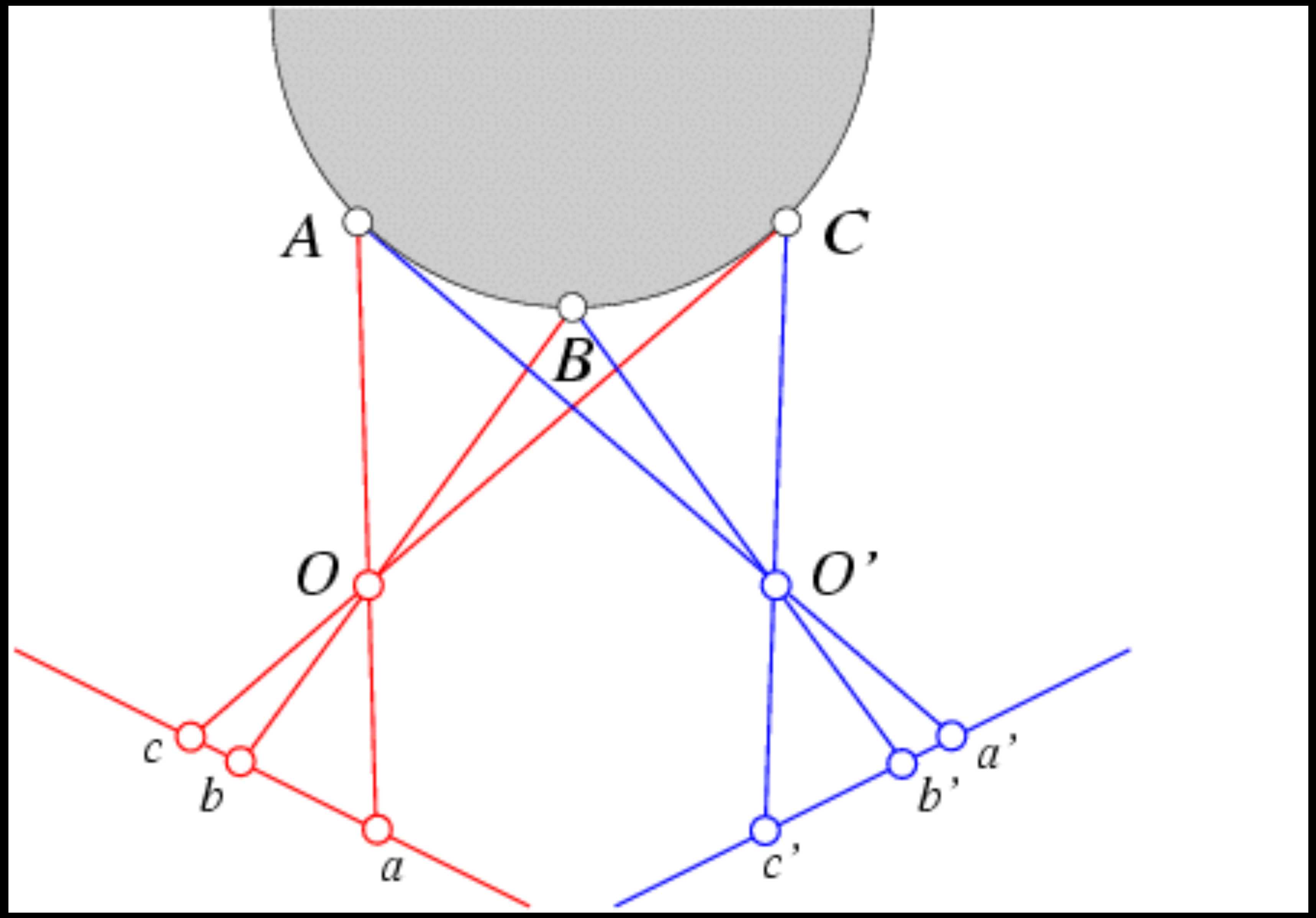
STEREO CONSTRAINTS/PRIORS

- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image



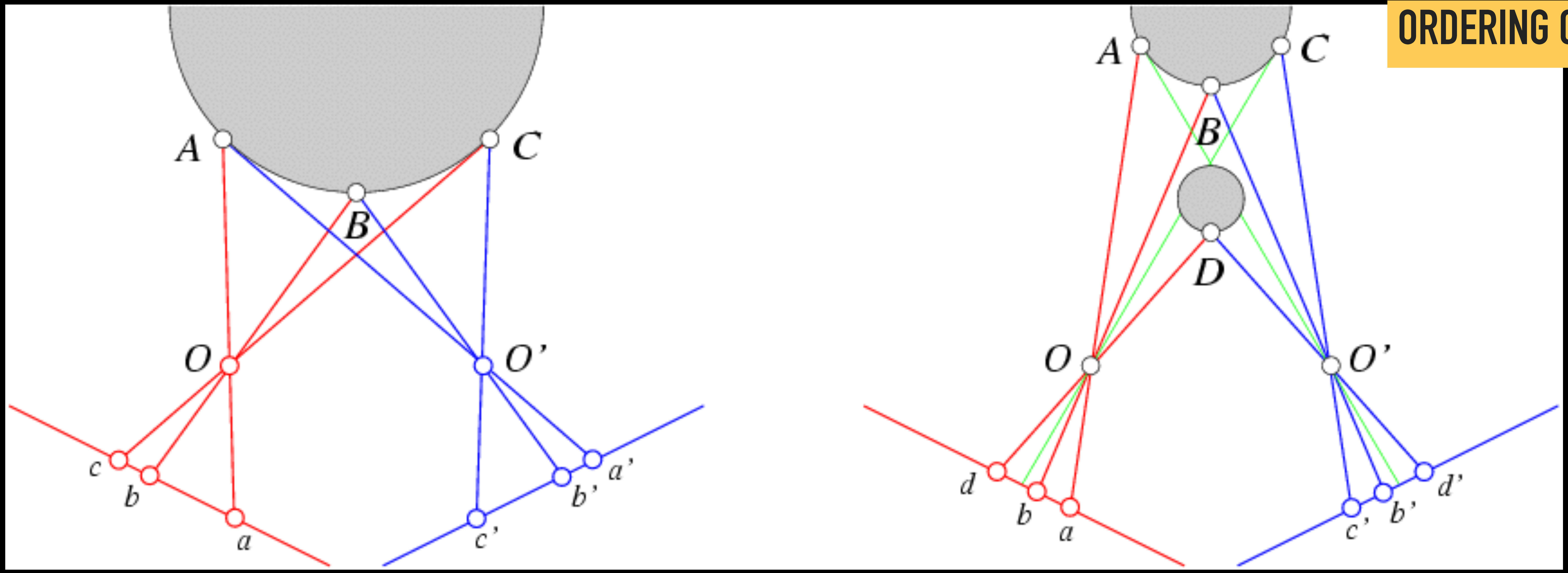
STEREO CONSTRAINTS/PRIORS

- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views



STEREO CONSTRAINTS/PRIORS

- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views

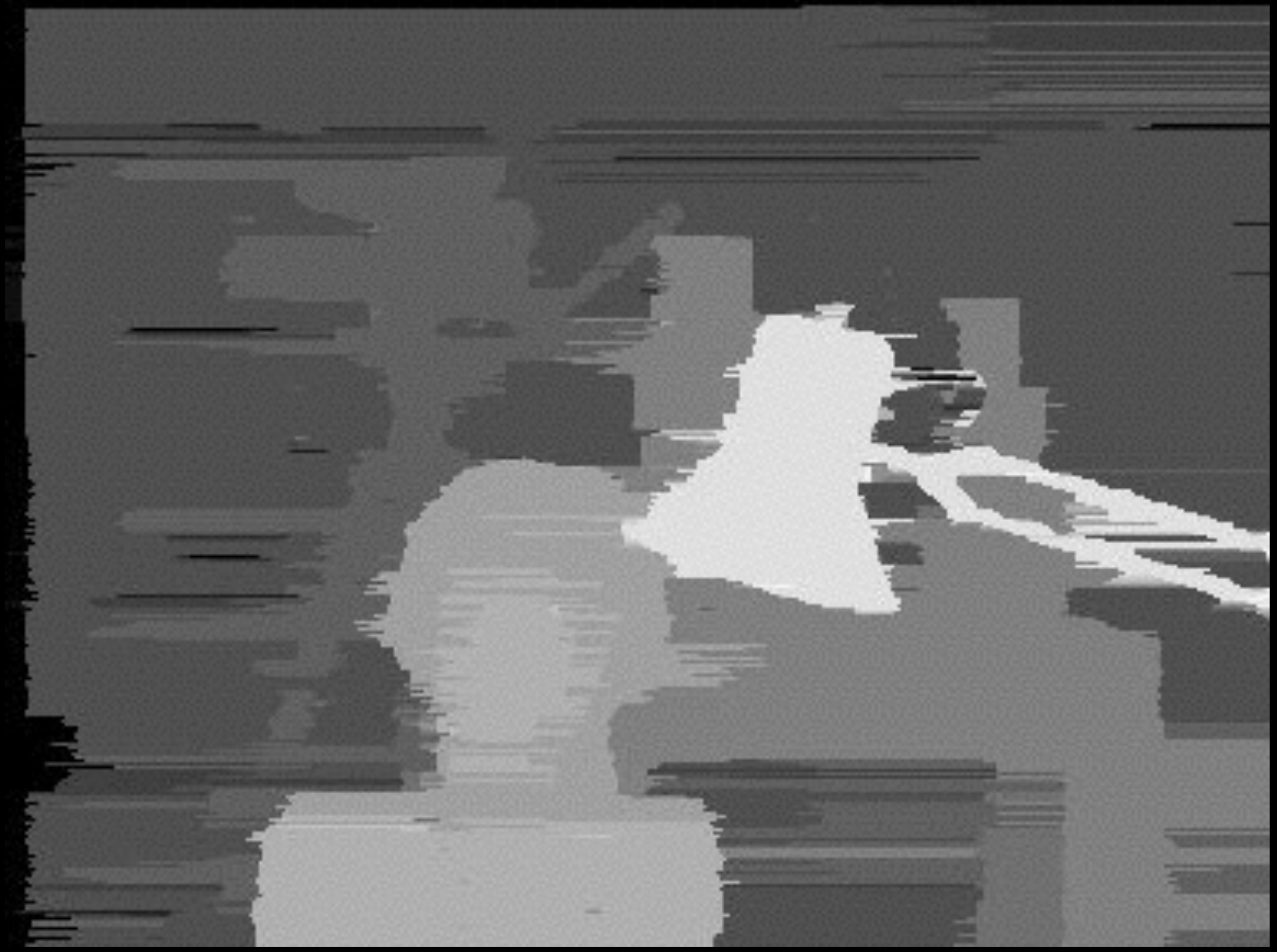


ORDERING CONSTRAINT DOESN'T HOLD

NON-LOCAL CONSTRAINTS

- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views
- Smoothness
  - We expect disparity values to change slowly (for the most part)

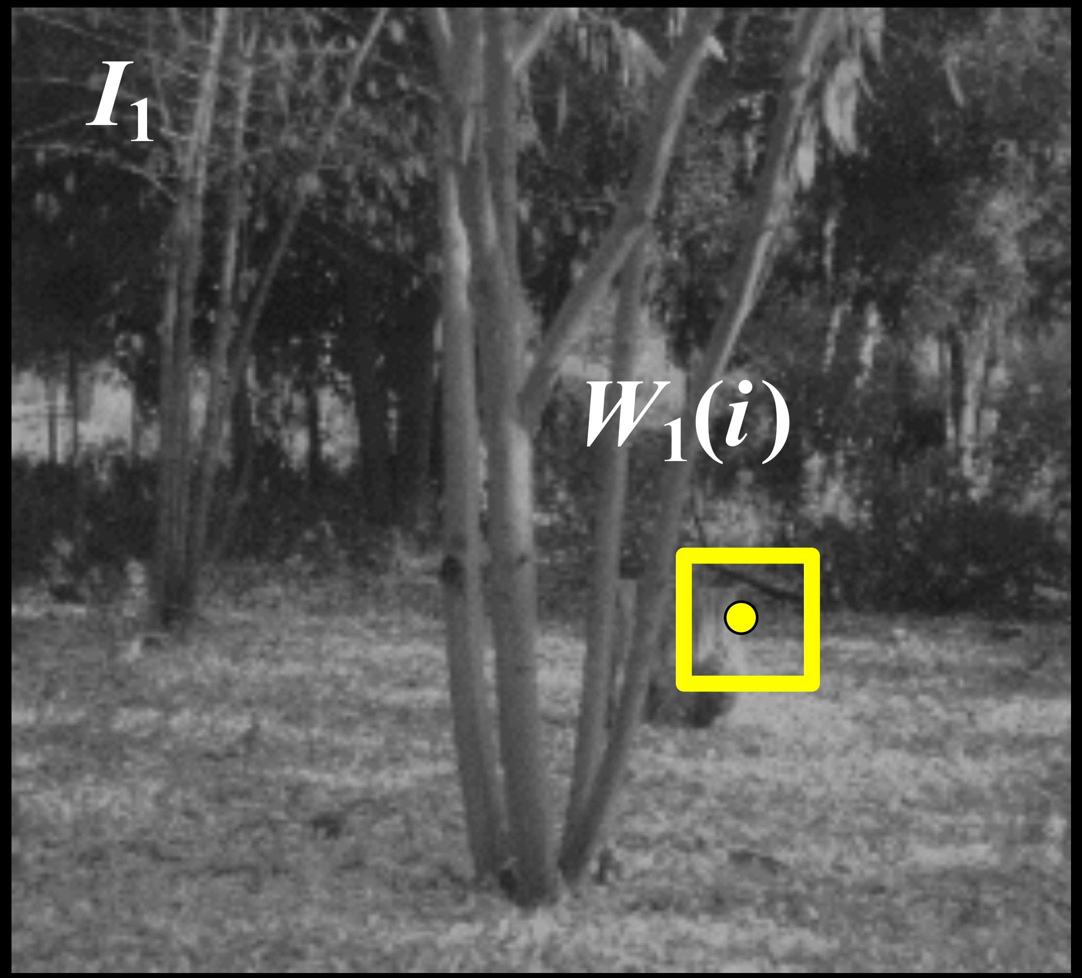
ONLY TAKING INTO ACCOUNT SCAN LINES

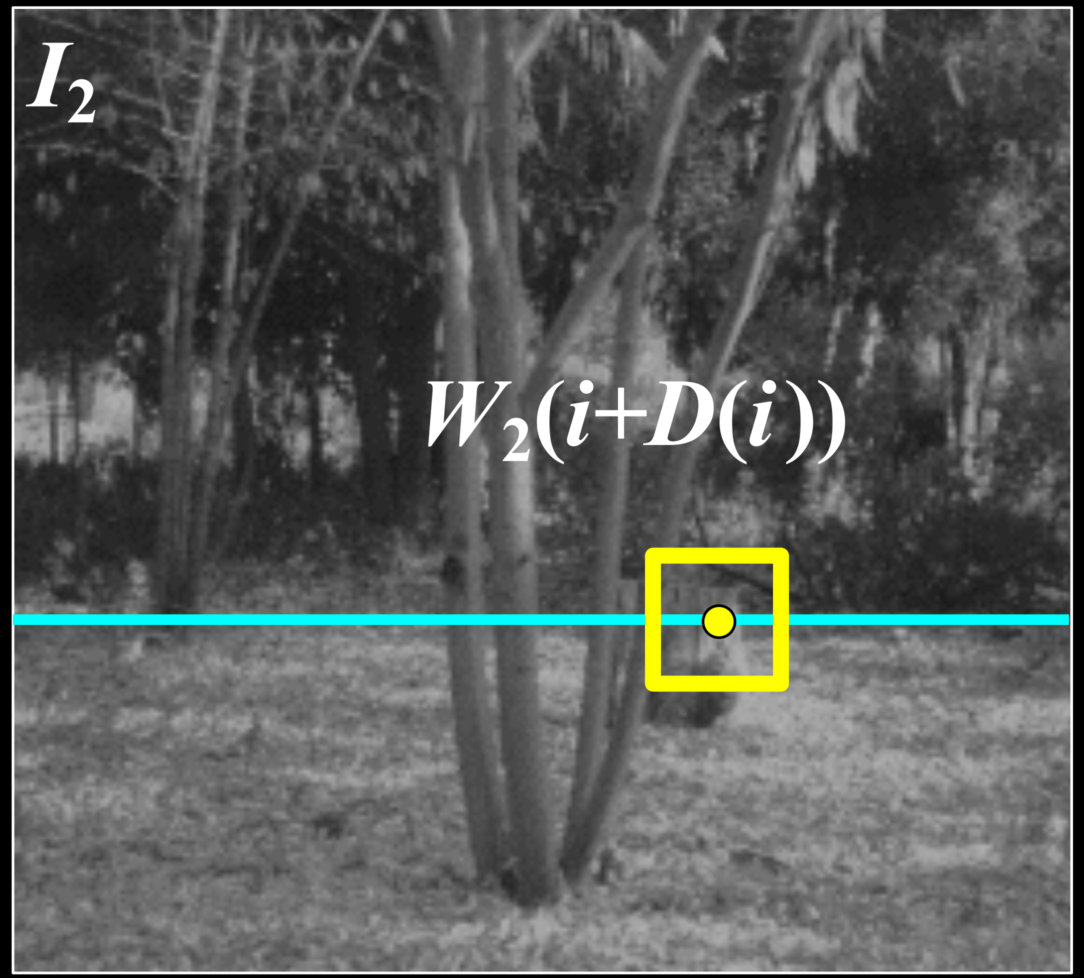


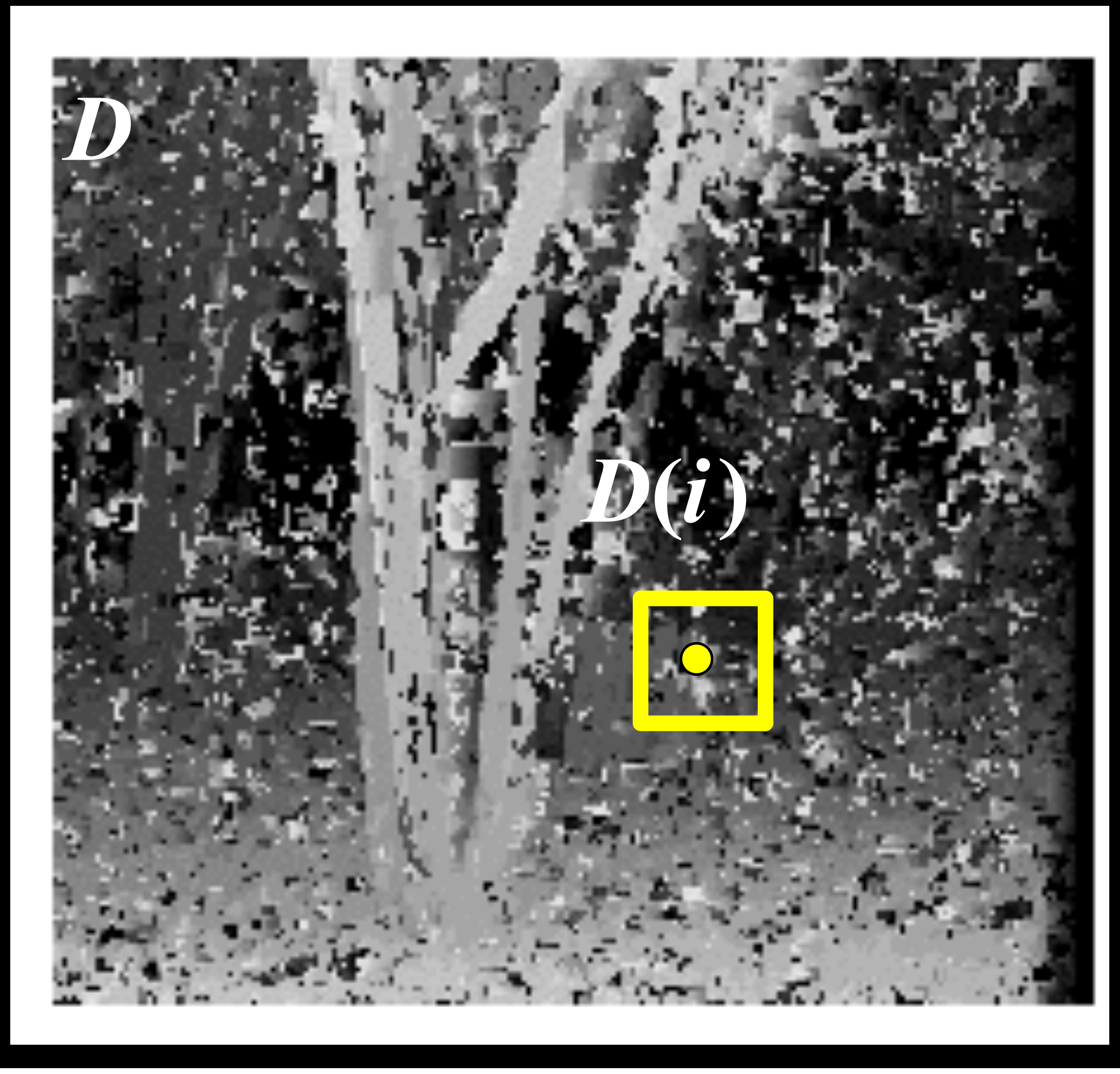
STEREO MATCHING AS ENERGY MINIMIZATION

- Energy functions on Markov Random Fields
- can be minimized using graph cuts, believe propagation, …
- http://vision.middlebury.edu/MRF/results/tsukuba/index.html

TEXT







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

$$E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2 \qquad E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)$$

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

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

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

MANY OF THESE CONSTRAINTS CAN BE ENCODED IN AN ENERGY FUNCTION AND SOLVED USING GRAPH CUTS

GRAPH CUTS

GROUND TRUTH

- http://vision.middlebury.edu/MRF/results/tsukuba/index.html

FOR THE LATEST AND GREATEST: HTTP://WWW.MIDDLEBURY.EDU/STEREO/

TAXONOMY OF STEREO ALGORITHMS

- matching cost computation;
- cost (support) aggregation;
- disparity computation and optimization; and
- disparity refinement.

MATCHING COSTS FUNCTIONS

- The most common pixel-based matching costs include sums of squared intensity differ- ences (SSD)
- Sum absolute intensity differences (SAD)
- Mean-squared error (MSE)
- mean absolute difference (MAD)
- normalized cross-correlation
- [Szeliski, R. and Scharstein, D. (2004)]. *Sampling the disparity space image.*

STEREO WITH SEGMENTATION

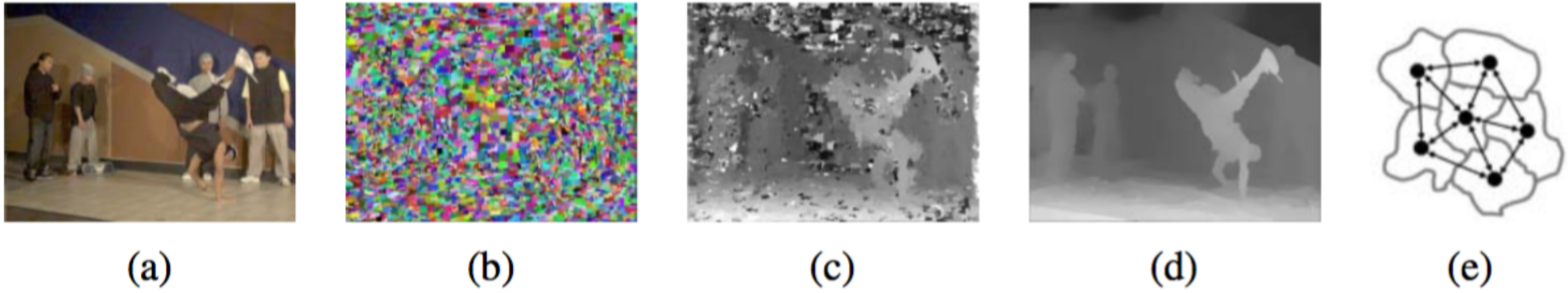


Figure 11.12 Segmentation-based stereo matching (Zitnick, Kang, Uyttendaele *et al.* 2004) © 2004 ACM: (a) input color image; (b) color-based segmentation; (c) initial disparity estimates; (d) final piecewise-smoothed disparities; (e) MRF neighborhood defined over the segments in the disparity space distribution (Zitnick and Kang 2007) © 2007 Springer.

PATCH MATCH STEREO

PATCHMATCH ALGORITHM

- Very fast to compute aproximate closest neighbour in images
- Saves displacements
- Random start
- Propagates
- Linear to the number of iterations

PATCHMATCH ALGORITHM

- Initialize with random displacements
- Compare with left and up (or right and down)
- Some random samples in the rest of the image
- Iterate

PATCHMATCH STEREO

Supplementary Material for the Paper:
PatchMatch Stereo - Stereo Matching with
Slanted Support Windows

PATCHMATCH STEREO

PatchMatch Based Joint View Selection and Depthmap Estimation

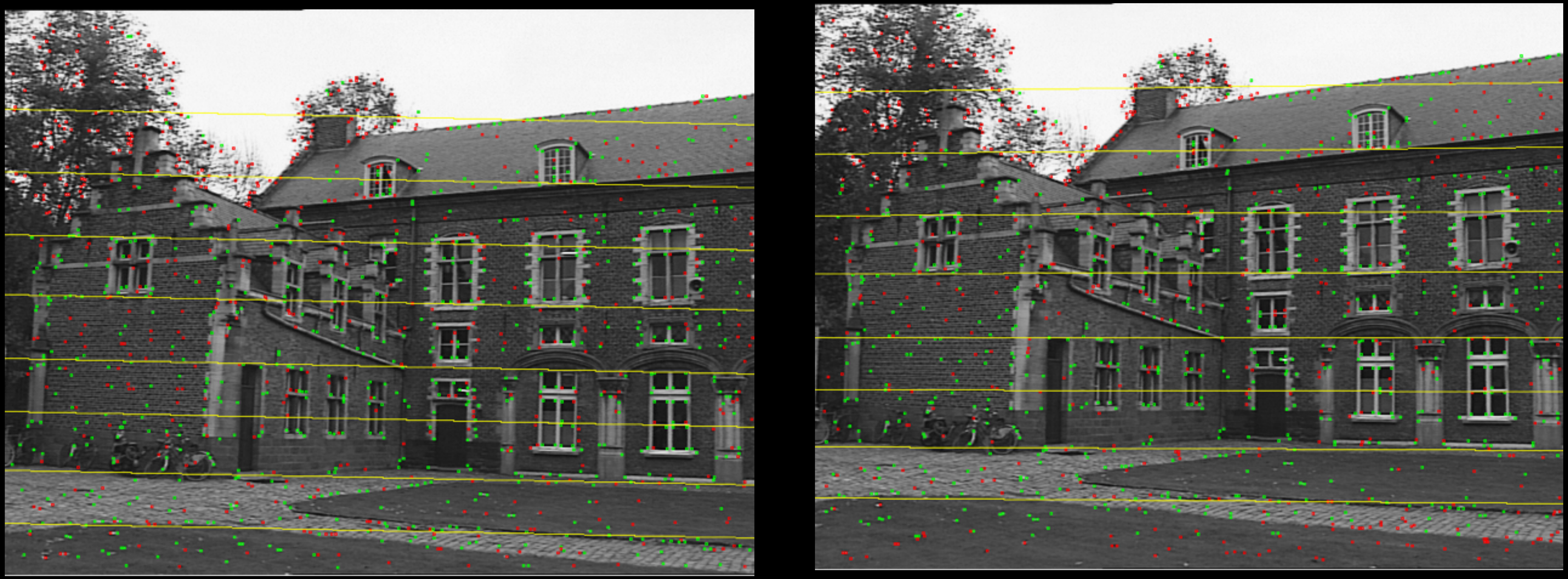
Enliang Zheng, Enrique Dunn, Vladimir Jojic, Jan-Michael Frahm
{ezheng, dunn, vjojic,jmf}@cs.unc.edu

SUMMARY

- Depth from disparity is trivial
- Compute Disparity is complicated
  - Correspondence problem - metrics are imperfect
  - Constraint for a particular problem
  - Global optimizations
  - …
  - Lot's of solutions

STEREO IMAGE RECTIFICATION

TWO-VIEW GEOMETRY



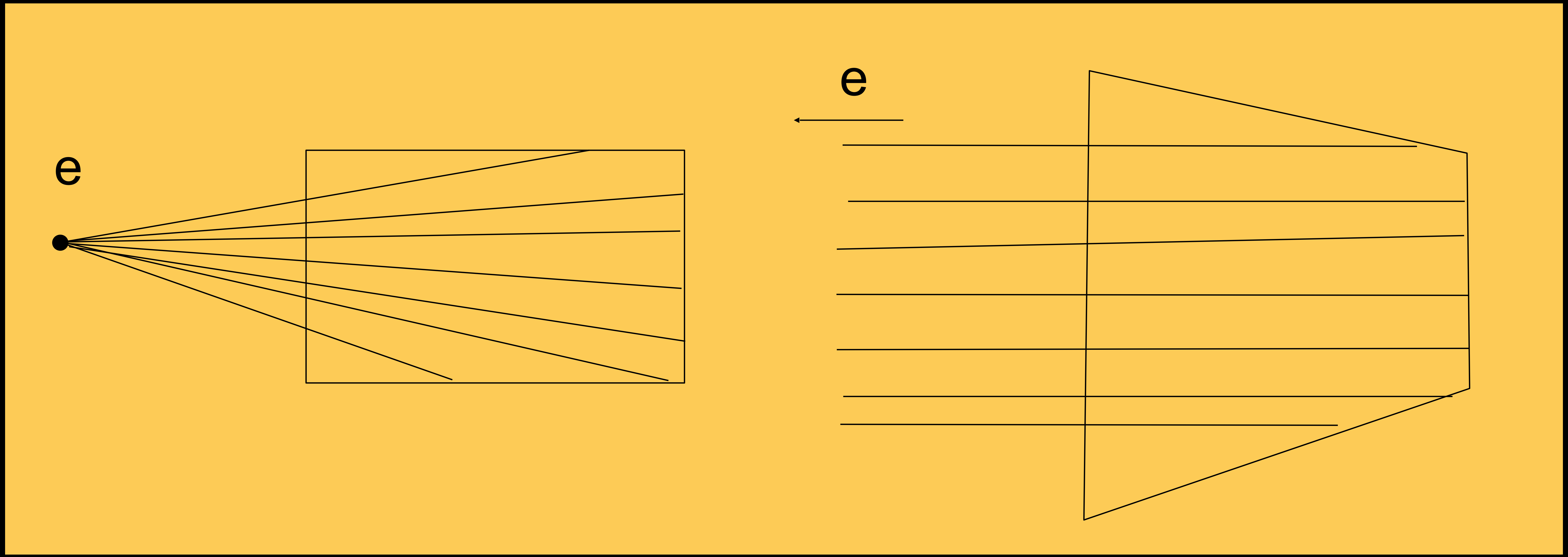
geometric relations between two views is fully

described by recovered 3x3 matrix F

IMAGE PAIR RECTIFICATION

Apply projective transformation so that epipolar lines correspond to horizontal scanlines

simplify stereo matching by warping the images



map epipole e to (1,0,0)

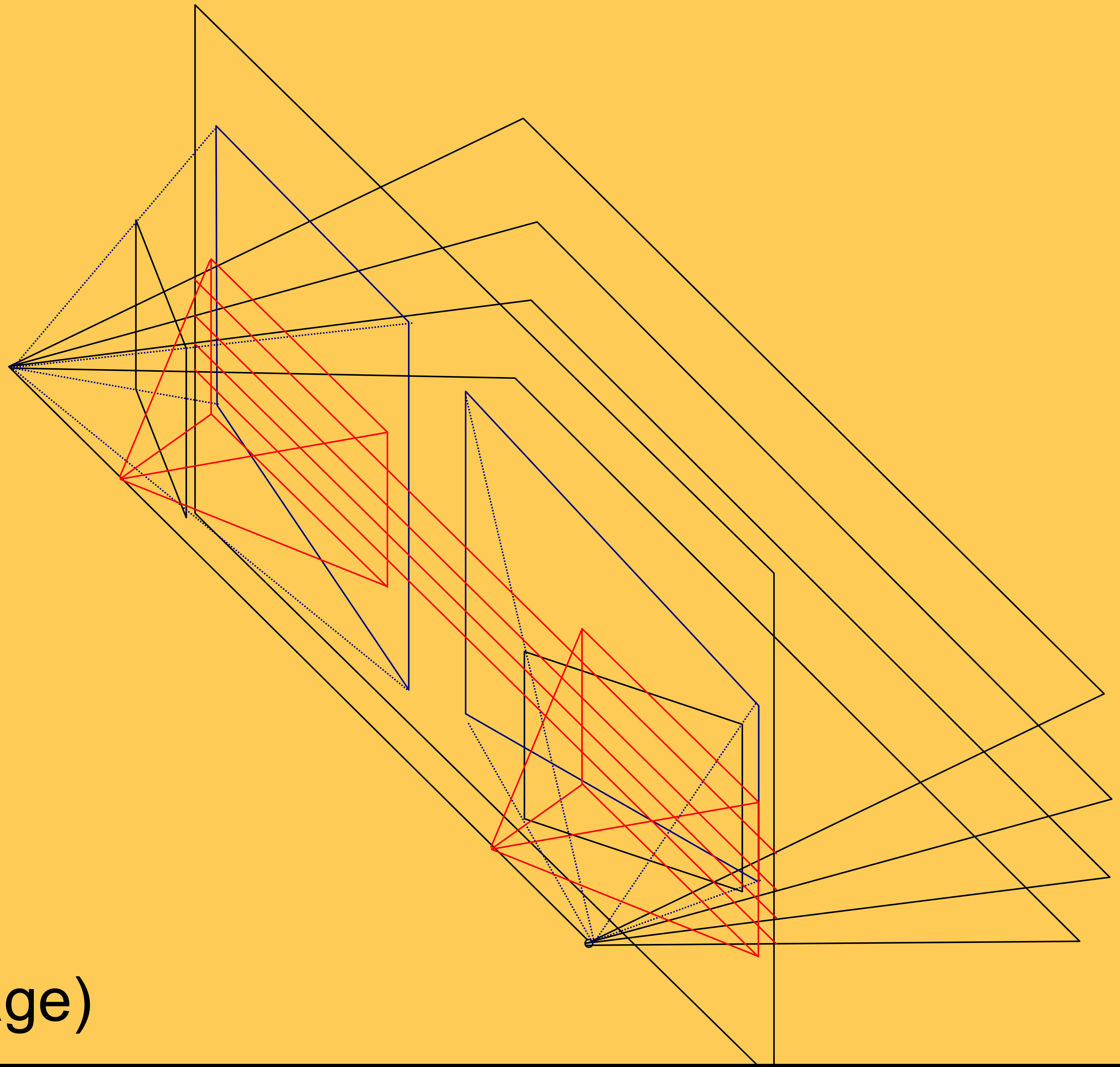
try to minimize image distortion

problem when epipole in (or close to) the image

IMAGE RECTIFICATION

- Planar Rectification
- Polar Rectification

PLANAR RECTIFICATION



Bring two views
to standard stereo setup
(moves epipole to ∞)
(not possible when in/close to image)

PLANAR RECTIFICATION

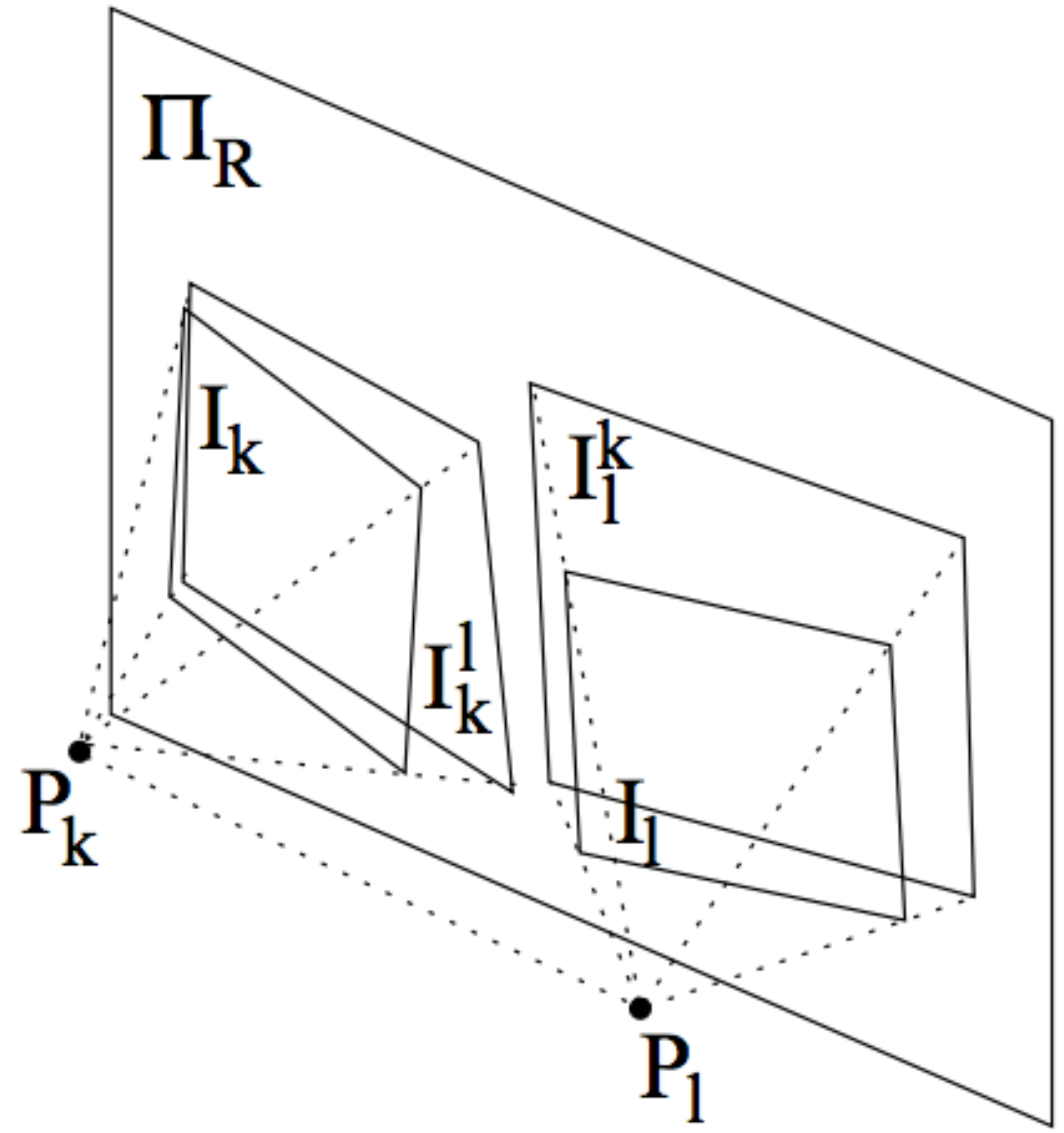


Figure 7.1: Planar rectification: $(\mathbf{I}_k^l, \mathbf{I}_l^k)$ are the rectified images for the pair $(\mathbf{I}_k, \mathbf{I}_l)$ (the plane Π_R should be parallel to the baseline $(\mathbf{P}_k, \mathbf{P}_l)$).

PLANAR RECTIFICATION

- Selecting a plane parallel with the baseline.
- The two image are then reprojected into this plane.
- These new images satisfy the standard stereo setup.
- The different methods for rectification mainly differ in how the remaining degrees of freedom are chosen.

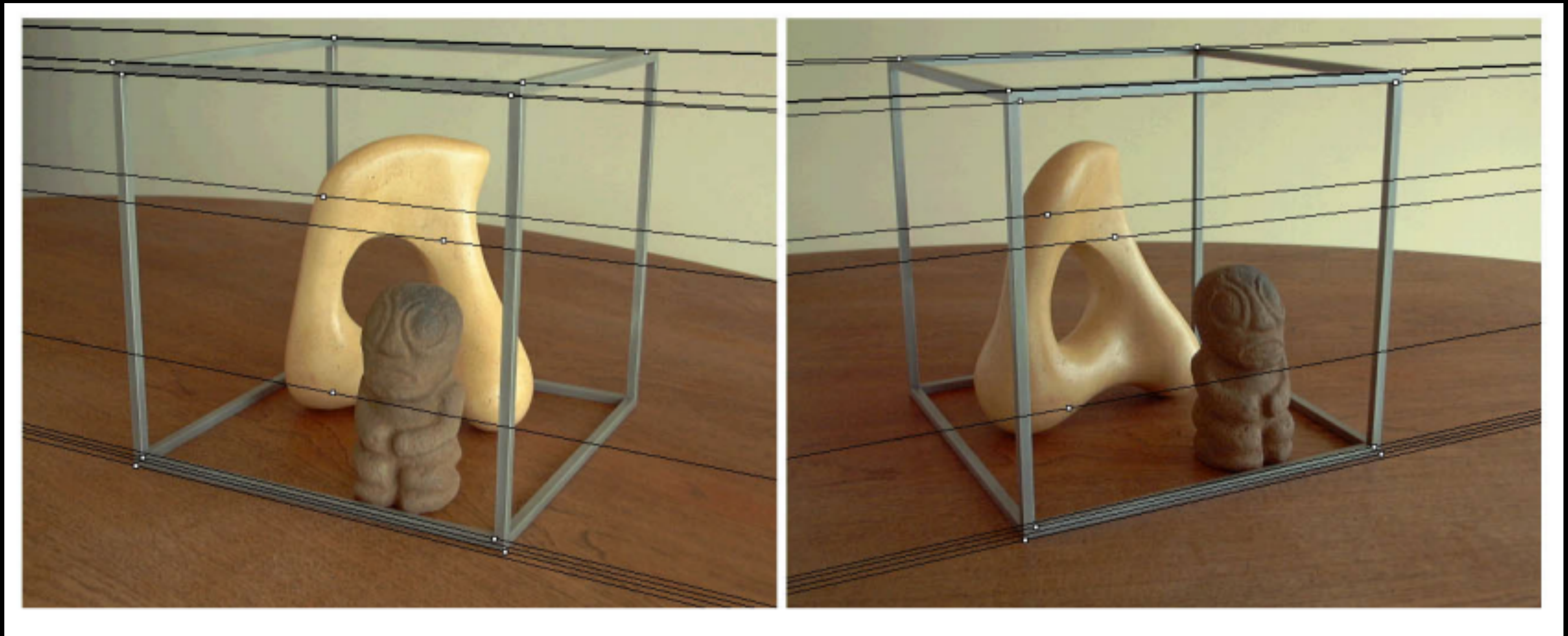
PLANAR RECTIFICATION

- Find an homography that sets epipoles to the infinity
- This is not unique:
  - Minimizes distortion
- [Loop&Zhang] Transformation closer to an affine transformation
- [Fusiello] Rotate cameras around the origin
  - Requires camera matrices

PLANAR RECTIFICATION

- Find an homography that sets epipoles to the infinity
- This is not unique:
  - Minimizes distortion
- [Loop&Zhang] Transformation closer to an affine transformation
  - http://research.microsoft.com/en-us/um/people/zhang/Papers/TR99-21.pdf
- [Fusiello] Rotate cameras around the origin
  - Requires camera matrices
  - http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/FUSIELLO2/rectif_cvol.html
  - http://www.diegm.uniud.it/fusiello/demo/rect/

PLANAR RECTIFICATION: [LOOP 1999]

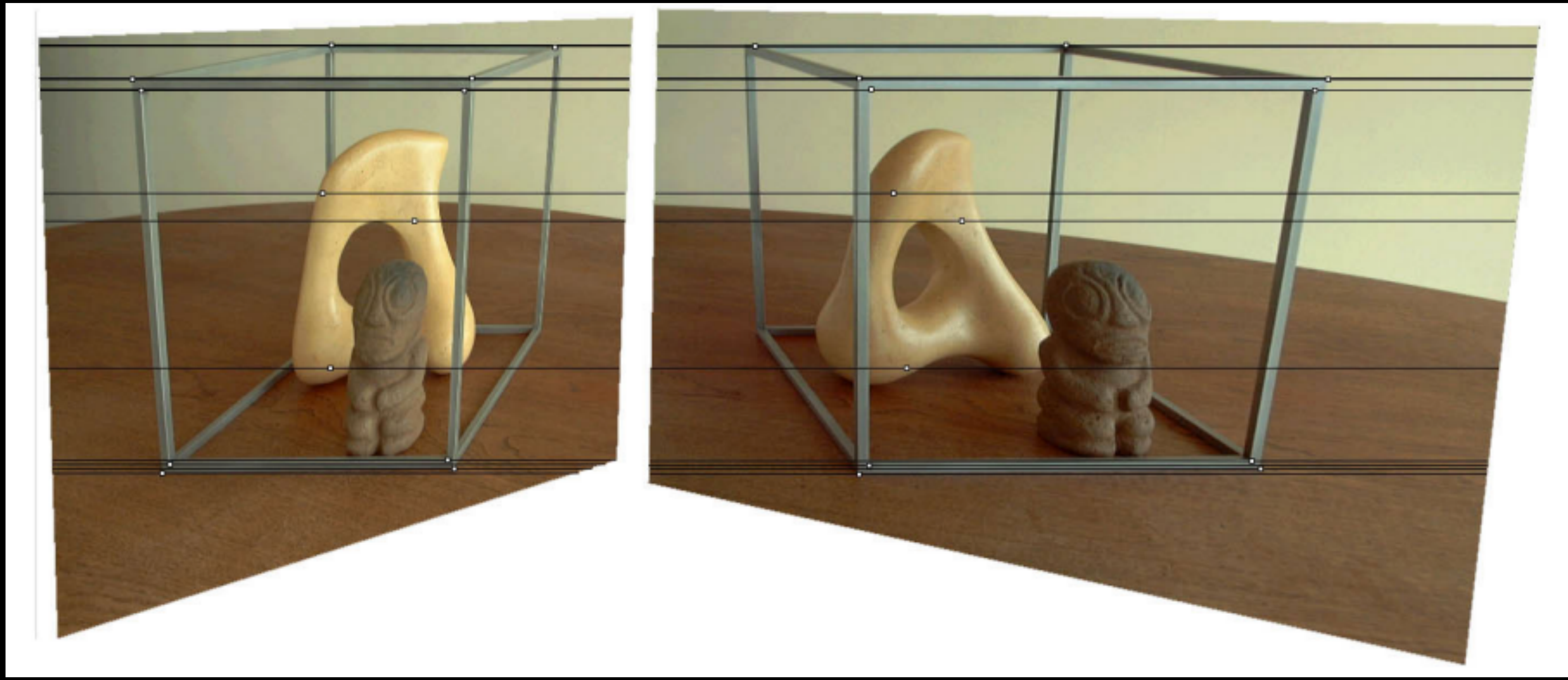


- Original image pair overlayed with several epipolar lines.
- Find homography H. Decompose it into Projective, Similarity and Shearing

PLANAR RECTIFICATION: [LOOP 1999]

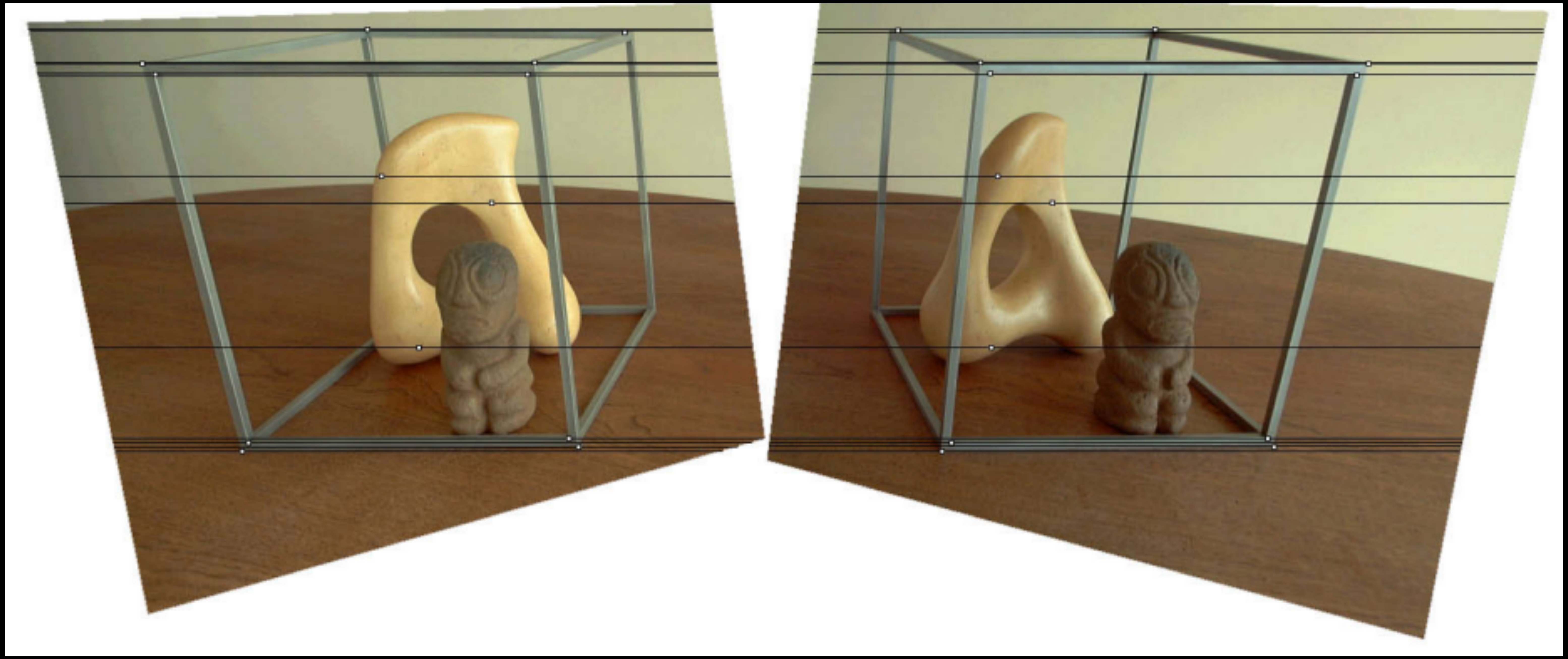
- Image pair transformed by the specialized projective mapping

PLANAR RECTIFICATION: [LOOP 1999]



- Image pair transformed by the similarity. Epipolar lines are aligned

PLANAR RECTIFICATION: [LOOP 1999]



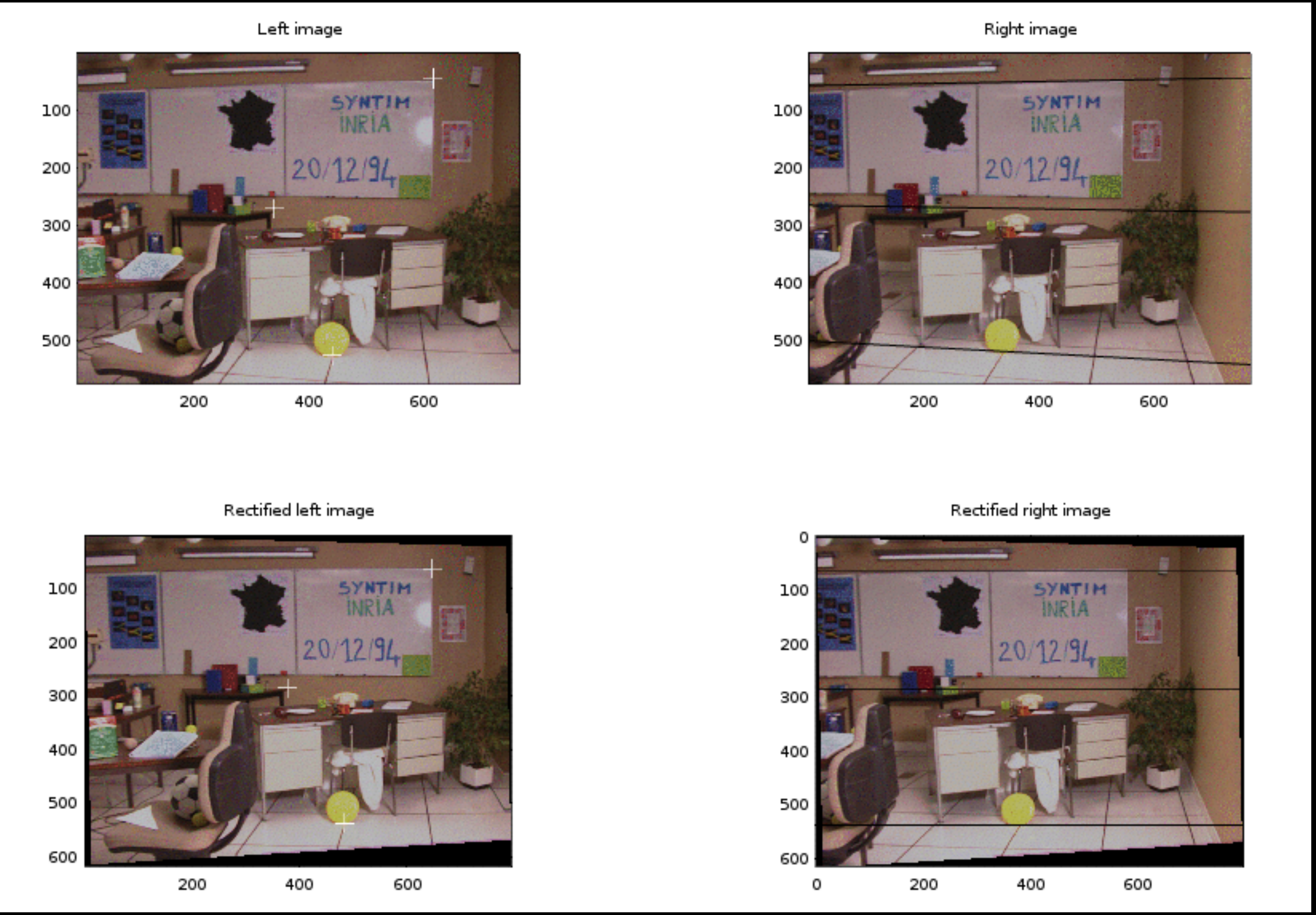
- Final image rectification after shearing transform

PLANAR RECTIFICATION: [FUSIELLO 2000]

http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/FUSIELLO2/rectif_cvol.html

- From a P1 and P2
  - Rotate both cameras so X axis is parallel to the baseline
  - Create new Pn1 and Pn2
    - Decompose P1 and P2
    - Set new R, the same for both cameras, and create new Pn1 and Pn2
    - Calculate the Homography H that transforms the images for the new Pn1 and Pn2

PLANAR RECTIFICATION: [FUSIELLO 2000]



LIMITATIONS PLANAR RECTIFICATION

- This approach fails when the epipoles are located in the images since this would have to results in infinitely large images. Even when this is not the case the image can still become very large (i.e. if the epipole is close to the image)

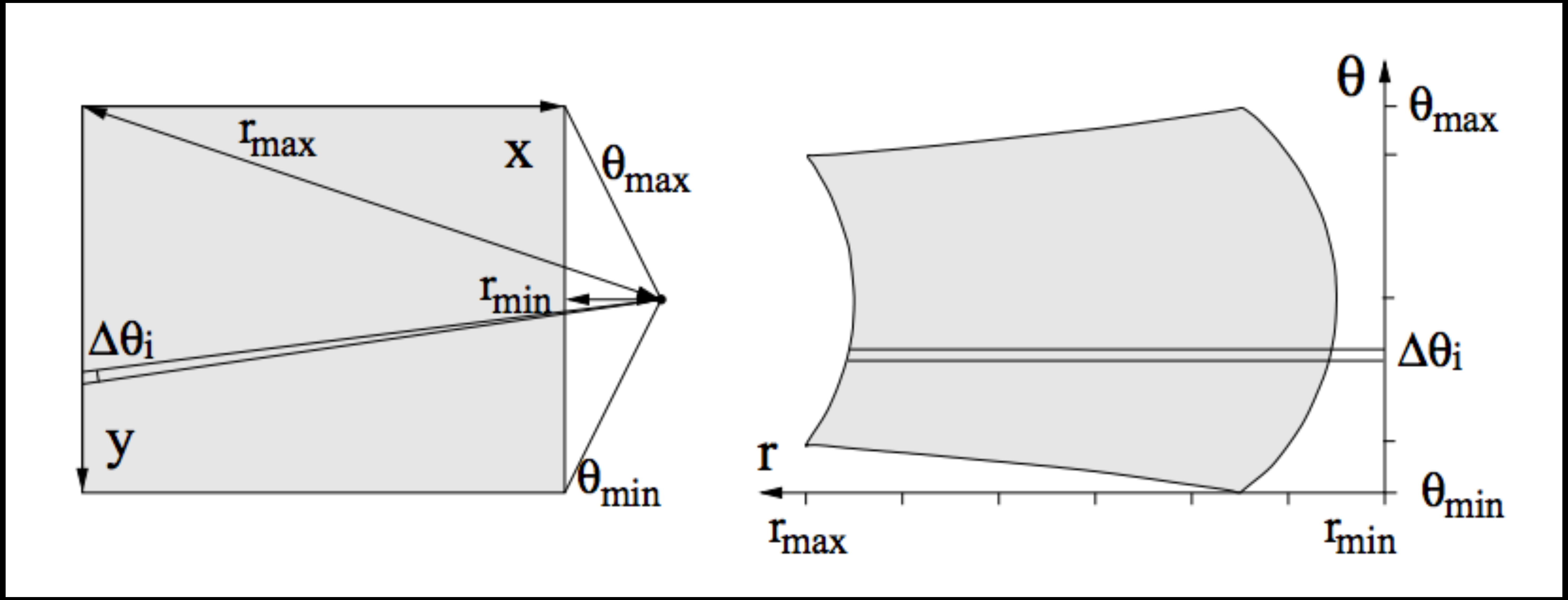
POLAR RECTIFICATION

- The image size is as small as can be achieved without compressing parts of the images
- https://www.inf.ethz.ch/personal/marc.pollefeys/pubs/PollefeysICCV99.pdf

POLAR RECTIFICATION

(Pollefeys et al. ICCV’99)

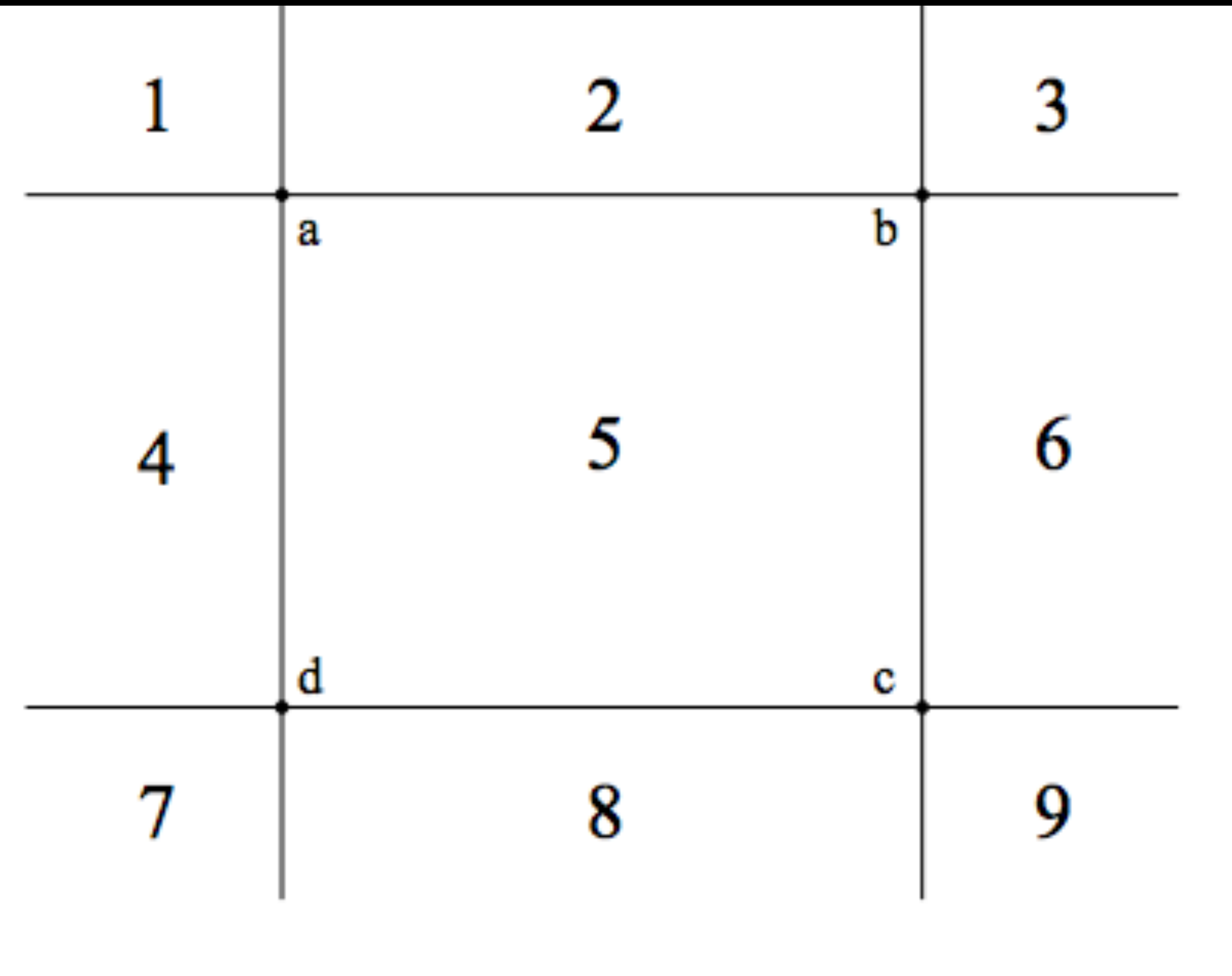
Polar re-parameterization around epipoles
Requires only (oriented) epipolar geometry
Preserve length of epipolar lines
Choose $\Delta\theta$ so that no pixels are compressed

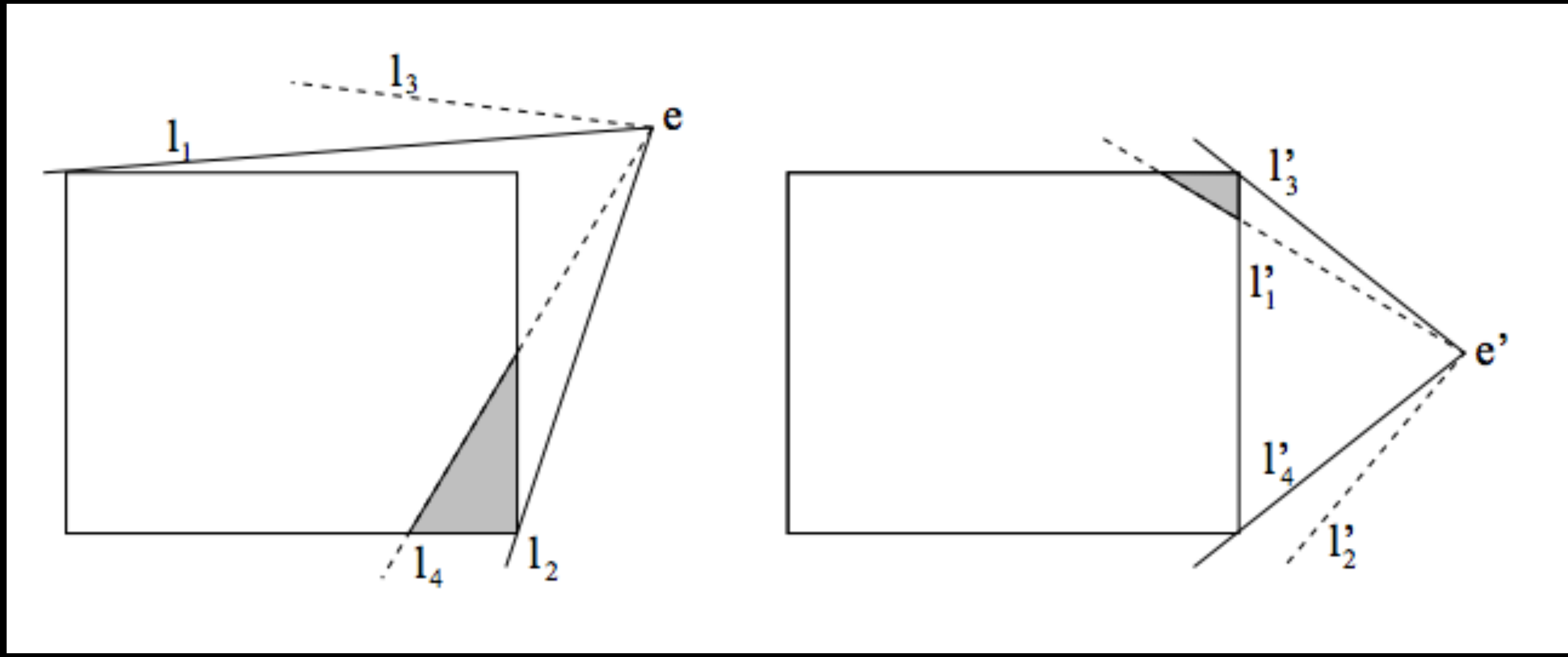


Works for all relative motions
Guarantees minimal image size

POLAR RECTIFICATION

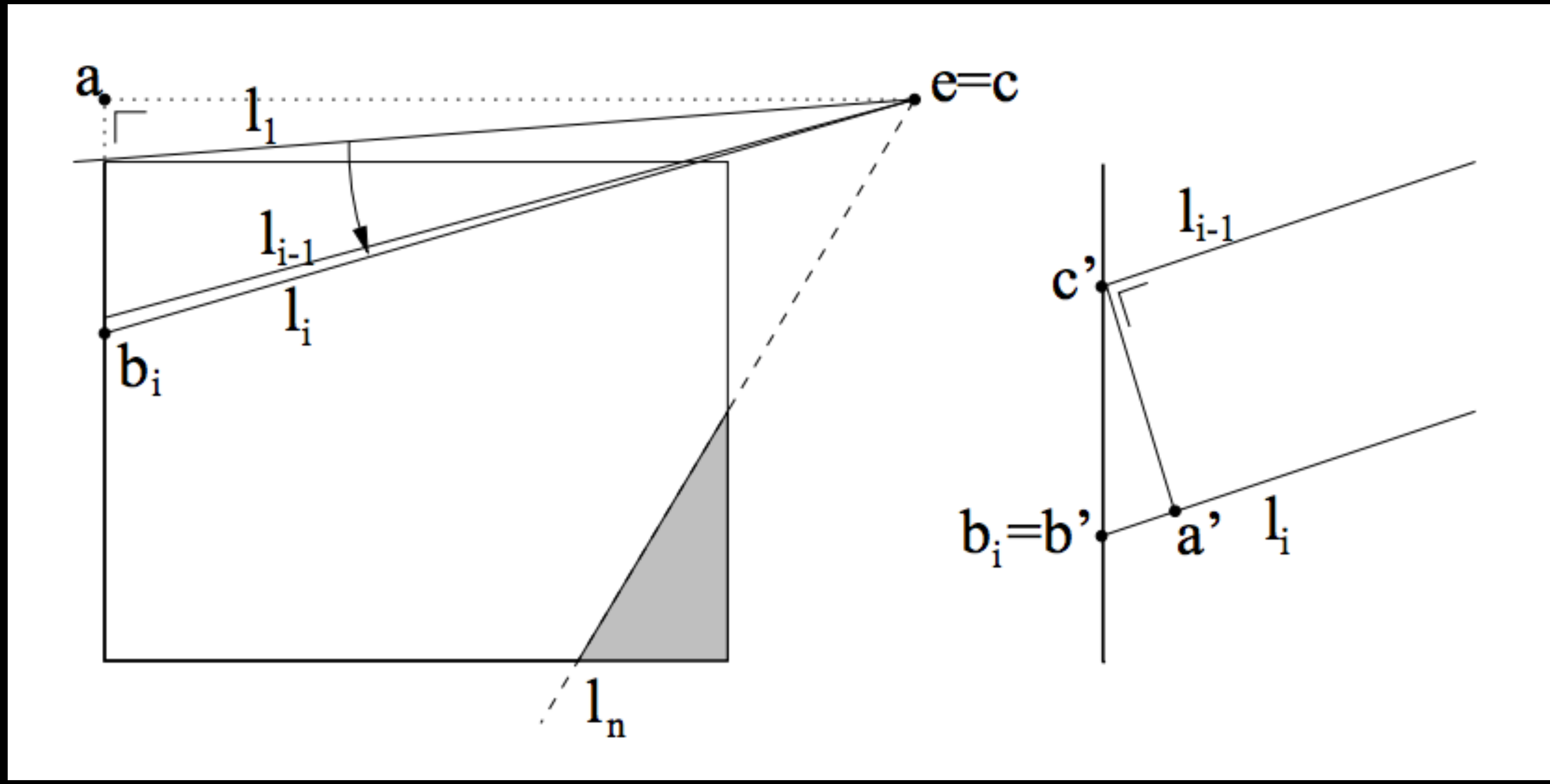
- Find where the epipole is.
- Find Extreme Epipolar lines
- Find Common Region



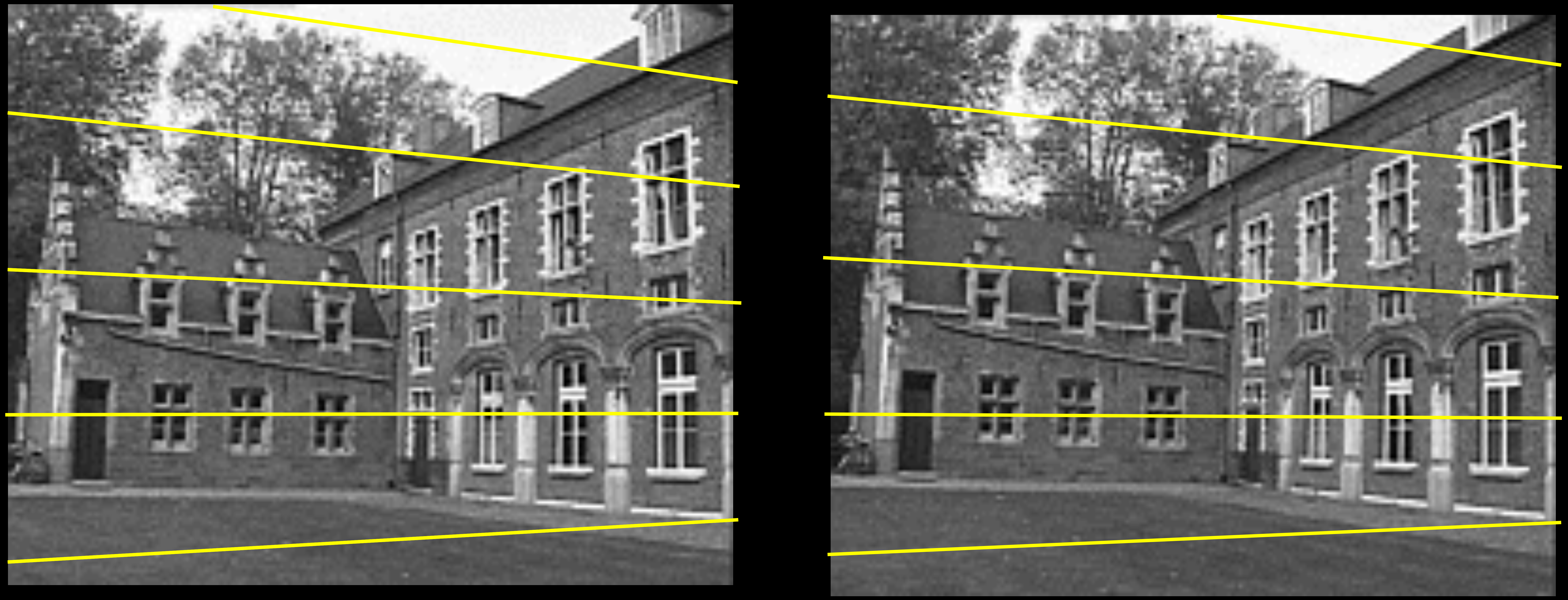


POLAR RECTIFICATION

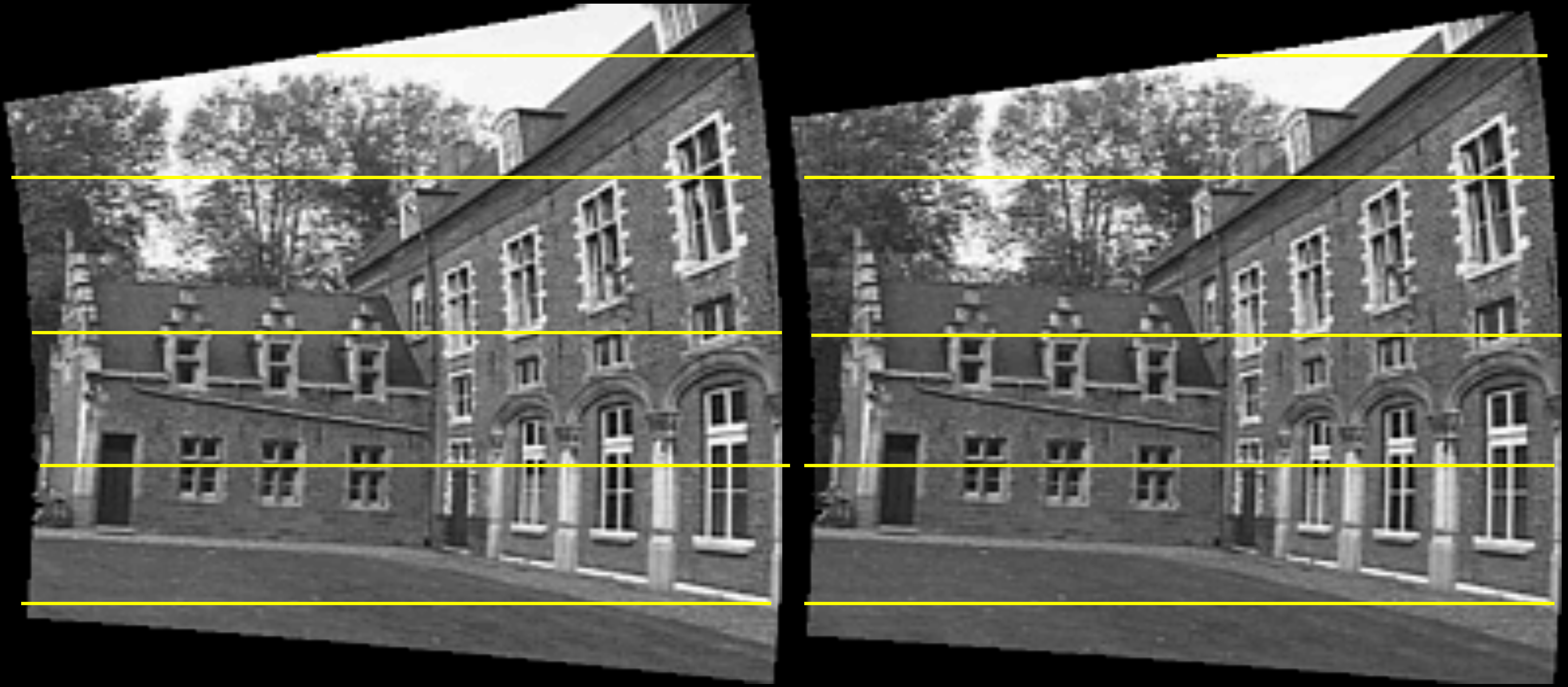
- To avoid losing pixel information the area of every pixel should be at least preserved when transformed to the rectified image. The worst case pixel is always located on the image border opposite to the epipole. A simple procedure to compute this step is depicted in Figure 7.6.
- The same procedure can be carried out in the other image. In this case the obtained epipolar line should be transferred back to the first image. The minimum of both displacements is carried out.



POLAR RECTIFICATION: EXAMPLE

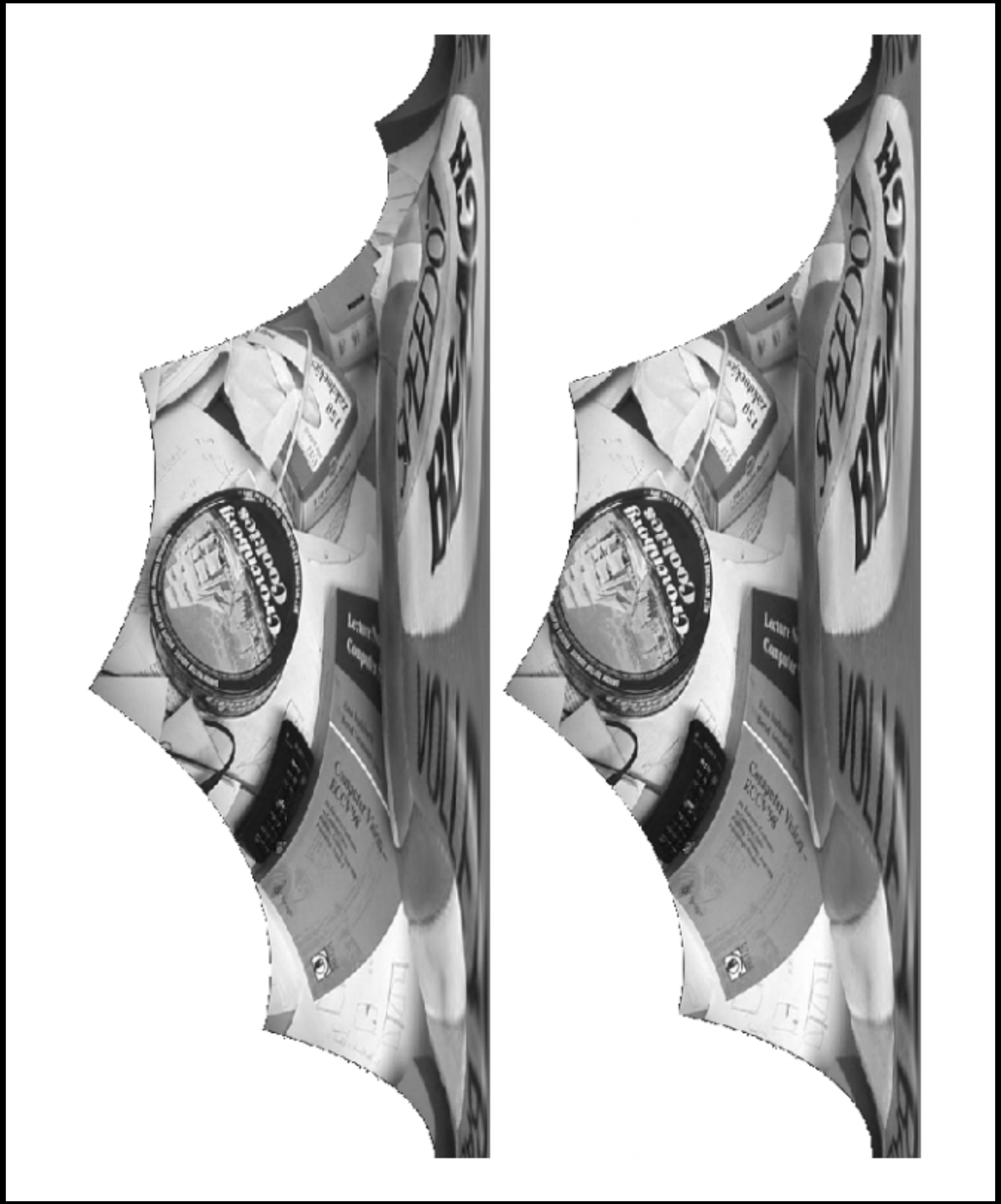


POLAR RECTIFICATION: EXAMPLE



POLAR RECTIFICATION: EXAMPLE

POLAR RECTIFICATION: EXAMPLE



NEXT WEEK
FEATURE DETECTION

FINDING 2D CORRESPONDENCES

- feature detection
- matching
- robust computation

MOVING A CLASS

- January 3rd
- Move to ?