Computational Vision

Stereopsis

Szeliski: 11.1 – 11.5

Trucco, chapter 8 (177-198)

Stereopsis

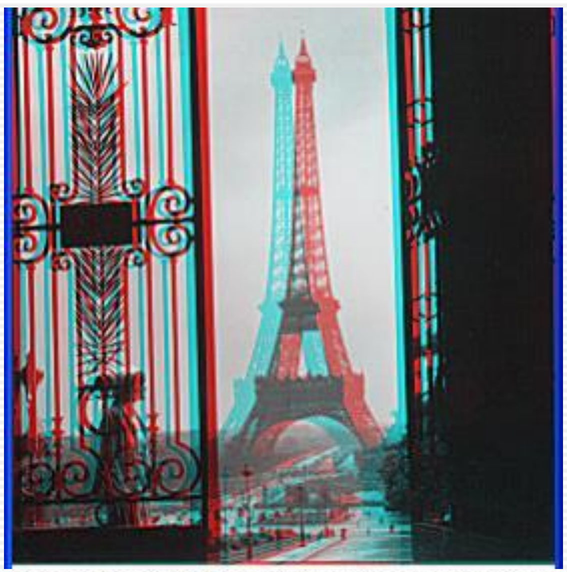
- Recovering 3D information (depth) from two images.
  - The correspondence problem.
  - The reconstruction problem.
  - Epipolar constraint.
  - The 8-point algorithm.

Stereo photography and stereo viewers

Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

Invented by Sir Charles Wheatstone, 1838

Image from fisher-price.com



http://www.johnsonshawmuseum.org

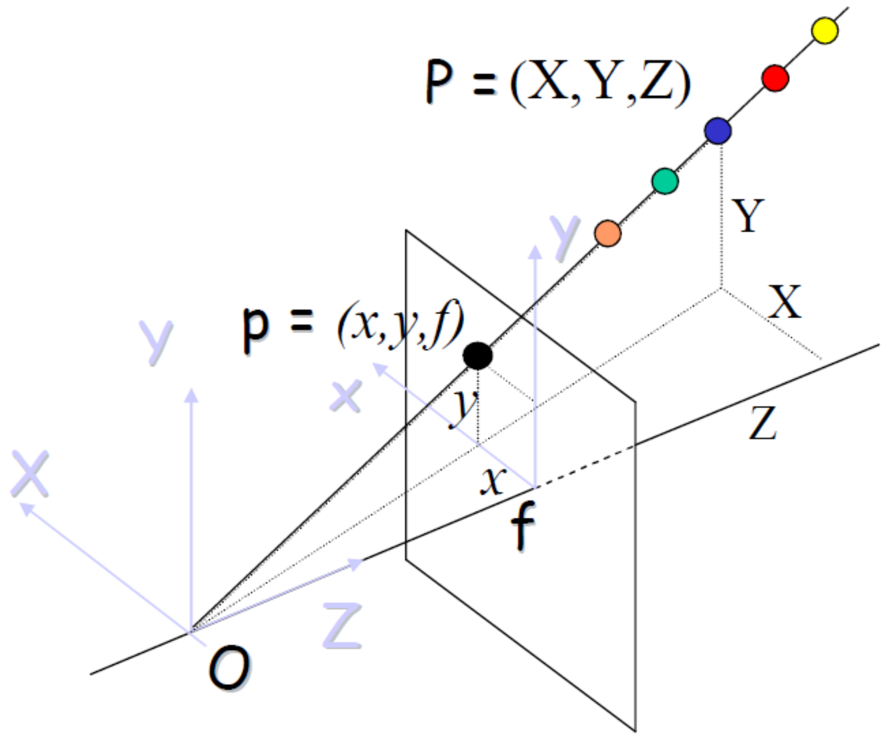
Check this out!

http://www.well.com/user/jimg/stereo/stereo_list.html

VR headsets and 360 degree videos

- Oculus Facebook
- Google VR Jump [discontinued]
- Microsoft Hololens
- Meta glasses

Why Stereo Vision?



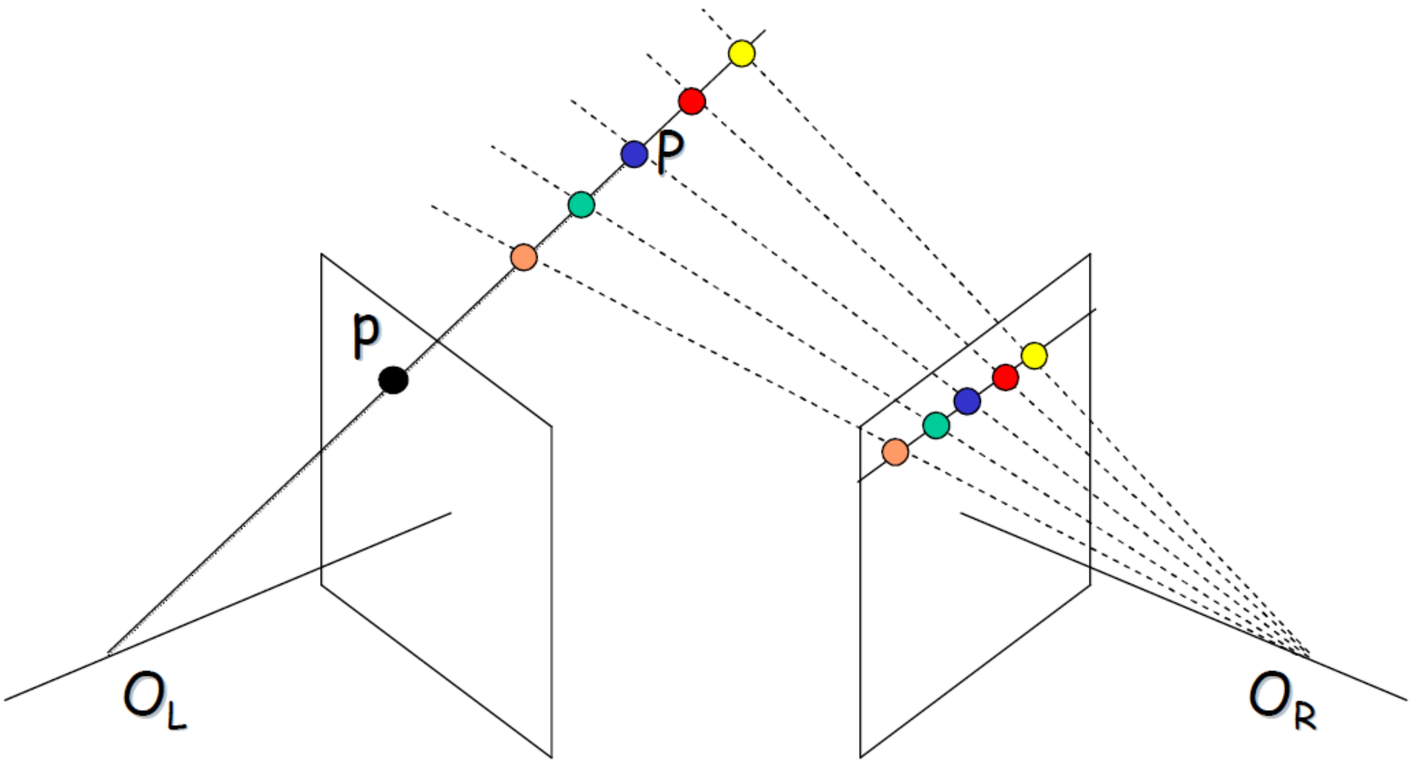
$$x = f\frac{X}{Z} = f\frac{kX}{kZ}$$

$$y = f\frac{Y}{Z} = f\frac{kY}{kZ}$$

Fundamental Ambiguity:

Any point on the ray OP has image p

Why Stereo Vision?



A second camera can resolve the ambiguity, enabling measurement of depth via triangulation.

The 2 problems of Stereo

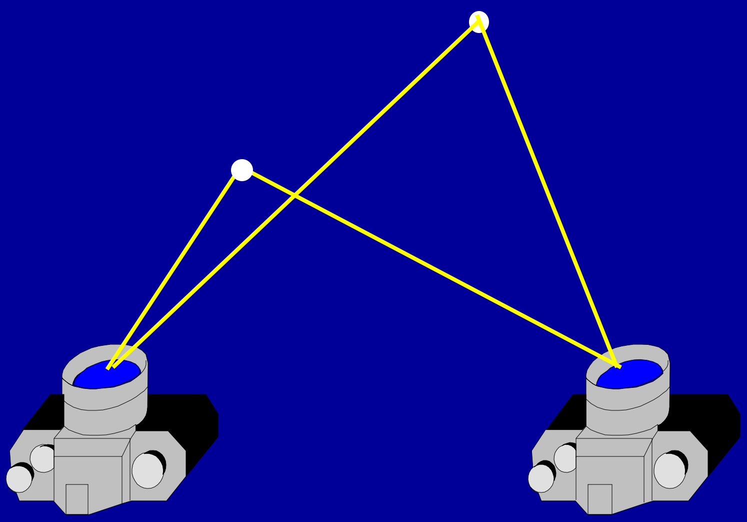
The setting: Simultaneous acquisition of 2 images (left, right) of a static scene.

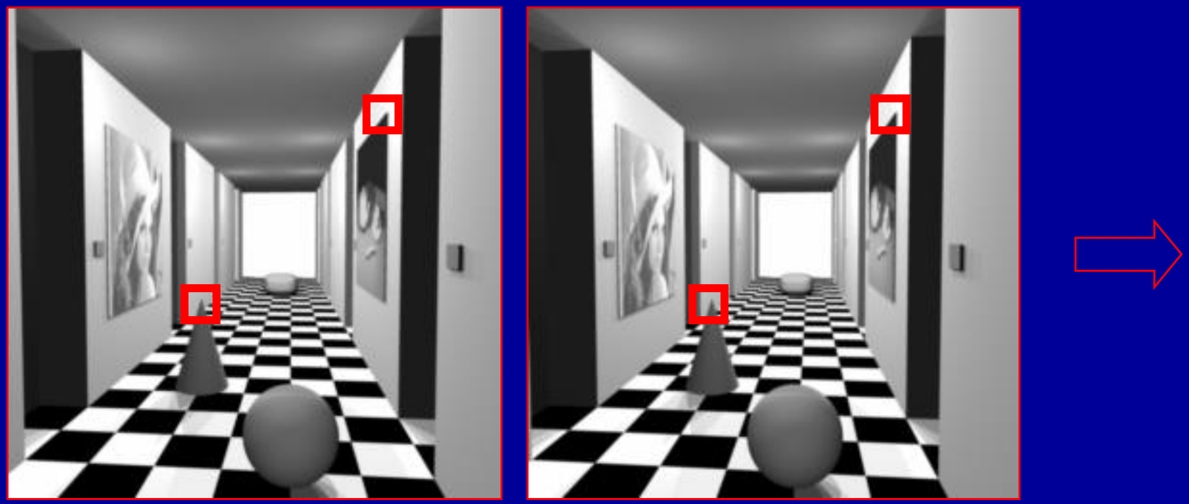
- Correspondence: *Which parts of the left and right images are projections of the same scene element?*
- Reconstruction: *Given:*
  - *A number of corresponding points between the left and right image,*
  - *Information on the geometry of the stereo system,*

Find:

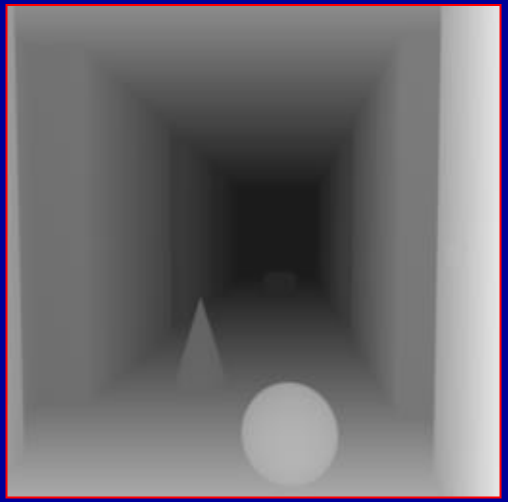
3-D structure of observed objects.

Stereo Vision

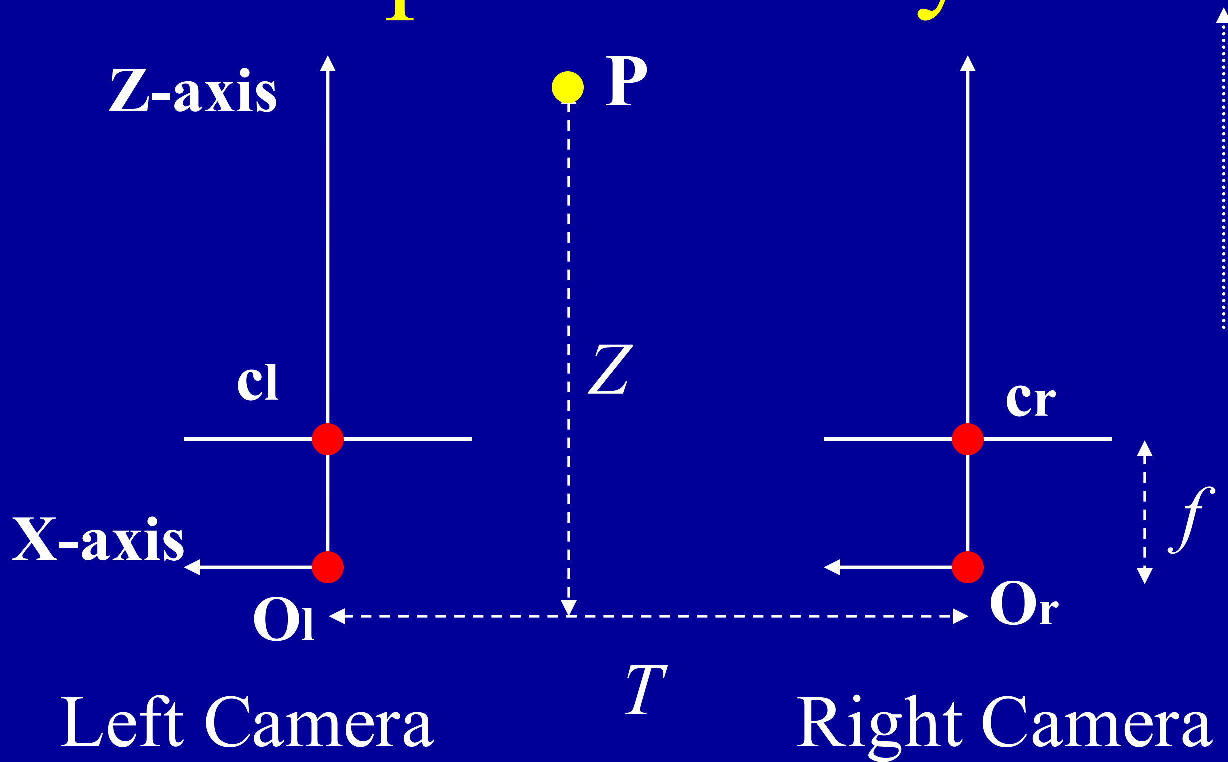




depth map

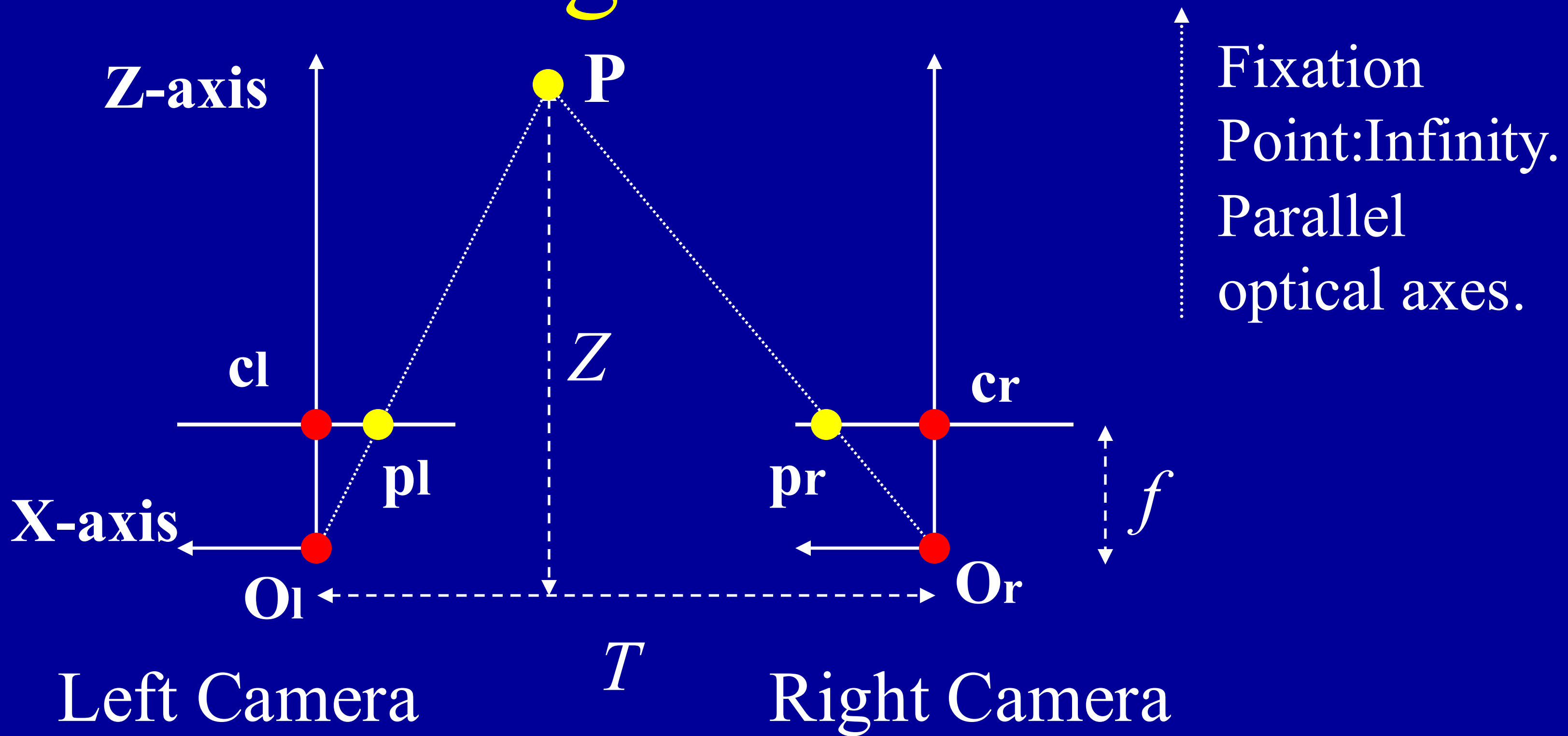


A simple stereo system



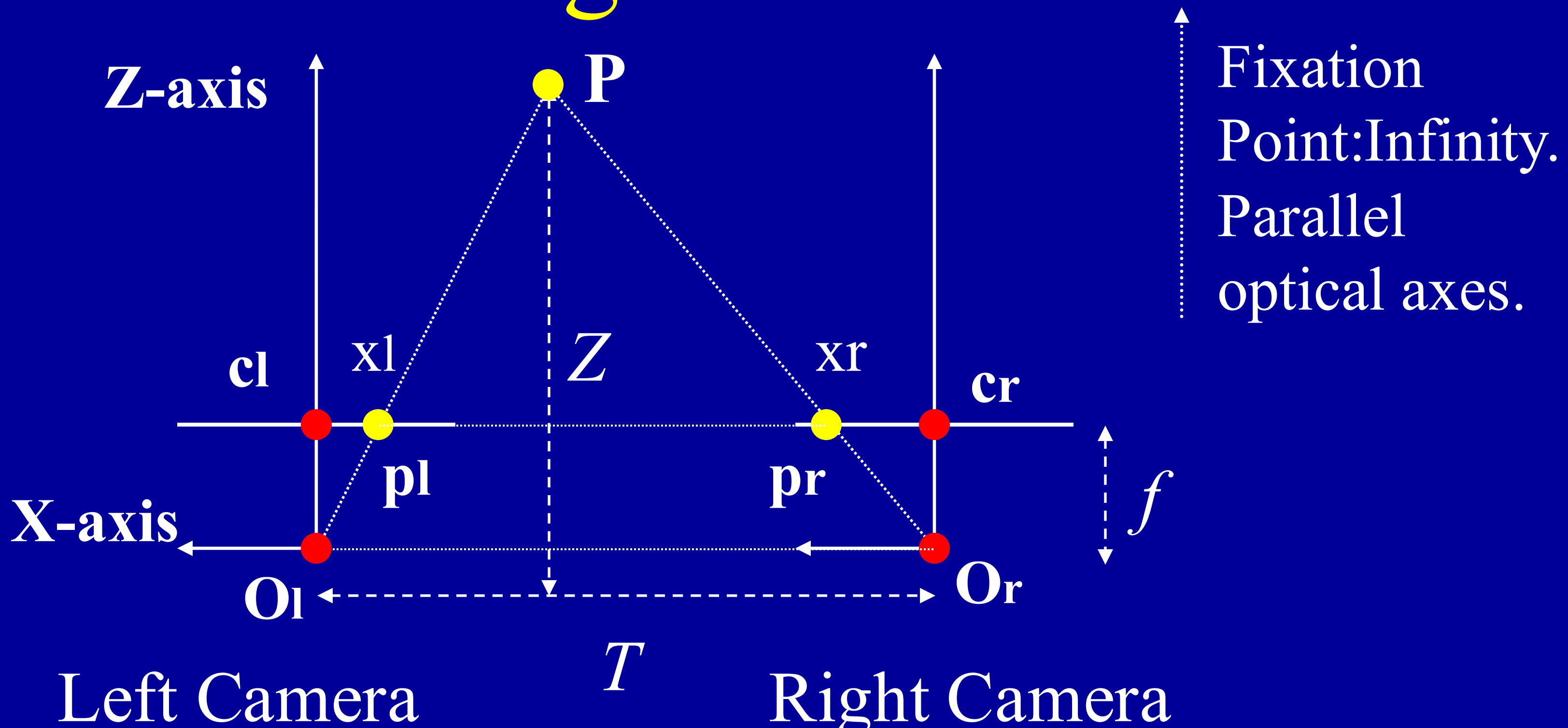
Fixation Point:Infinity.
Parallel optical axes.

Triangulation



Calibrated Cameras

Triangulation



Calibrated Cameras

Similar triangles:

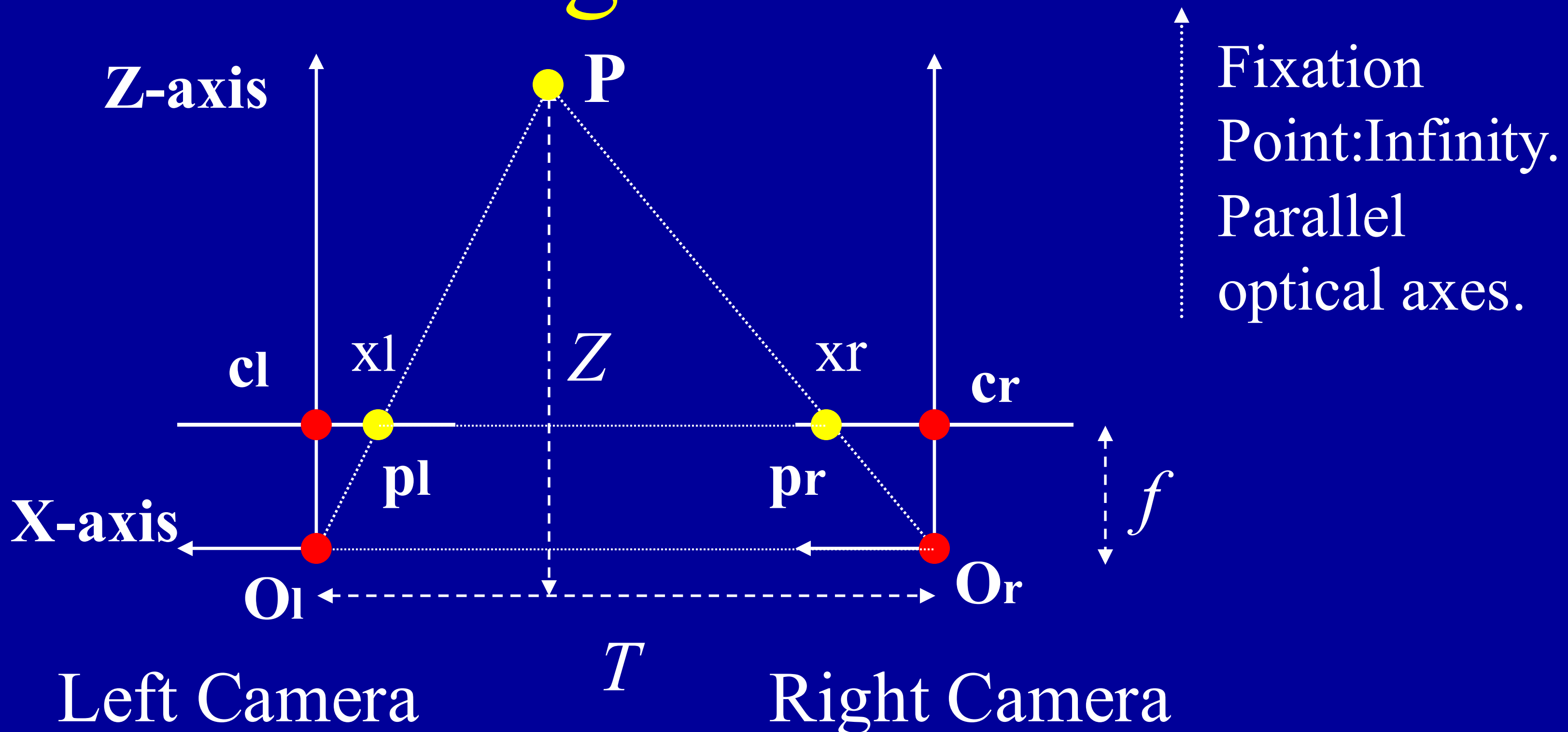
$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z} \Rightarrow Z = f\frac{T}{d}, d = x_l - x_r$$

d:disparity (difference in retinal positions).

T:baseline.

Depth (Z) is inversely proportional to d (fixation at infinity)

Triangulation



Calibrated Cameras

Similar triangles:

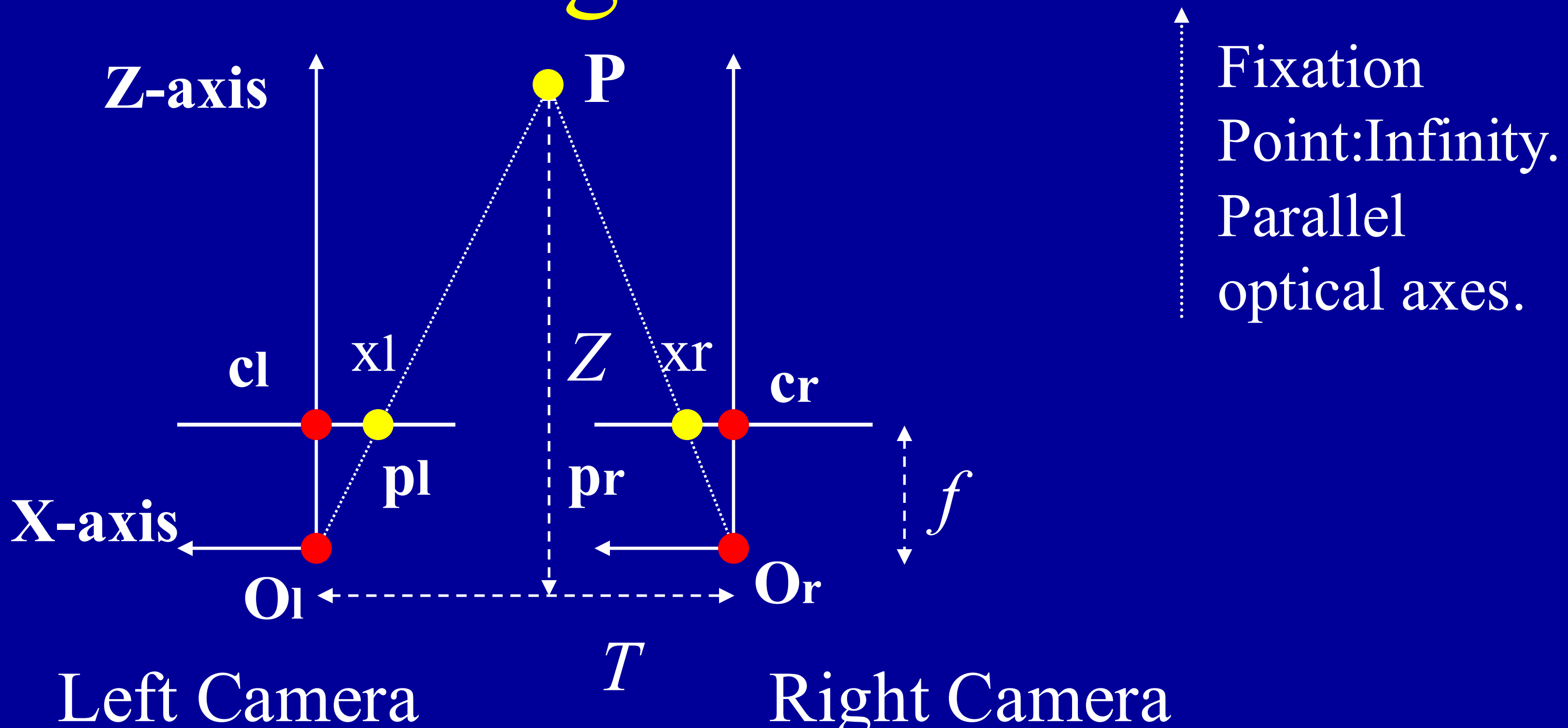
$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z} \Rightarrow Z = f\frac{T}{d}, d = x_l - x_r$$

d:disparity (difference in retinal positions).

T:baseline.

Baseline T: accuracy/robustness of depth calculation.

Triangulation



Calibrated Cameras

Similar triangles:

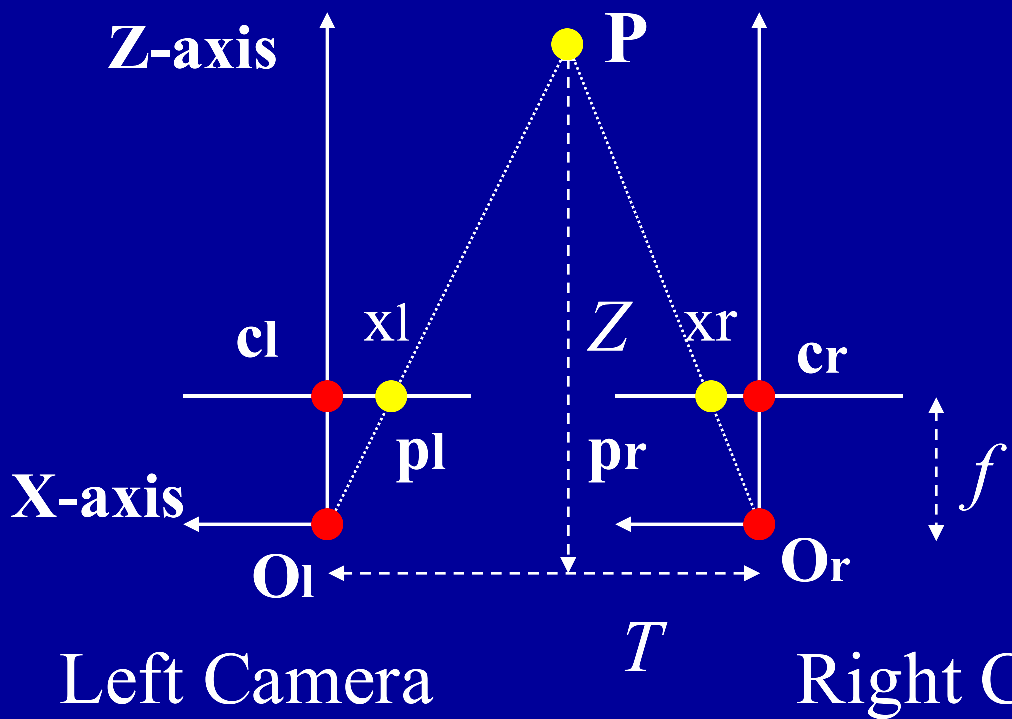
$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z} \Rightarrow Z = f\frac{T}{d}, d = x_l - x_r$$

d:disparity (difference in retinal positions).

T:baseline.

Small baselines: less accurate measurements.

Triangulation



Fixation Point:Infinity. Parallel optical axes.

Calibrated Cameras

Similar triangles:

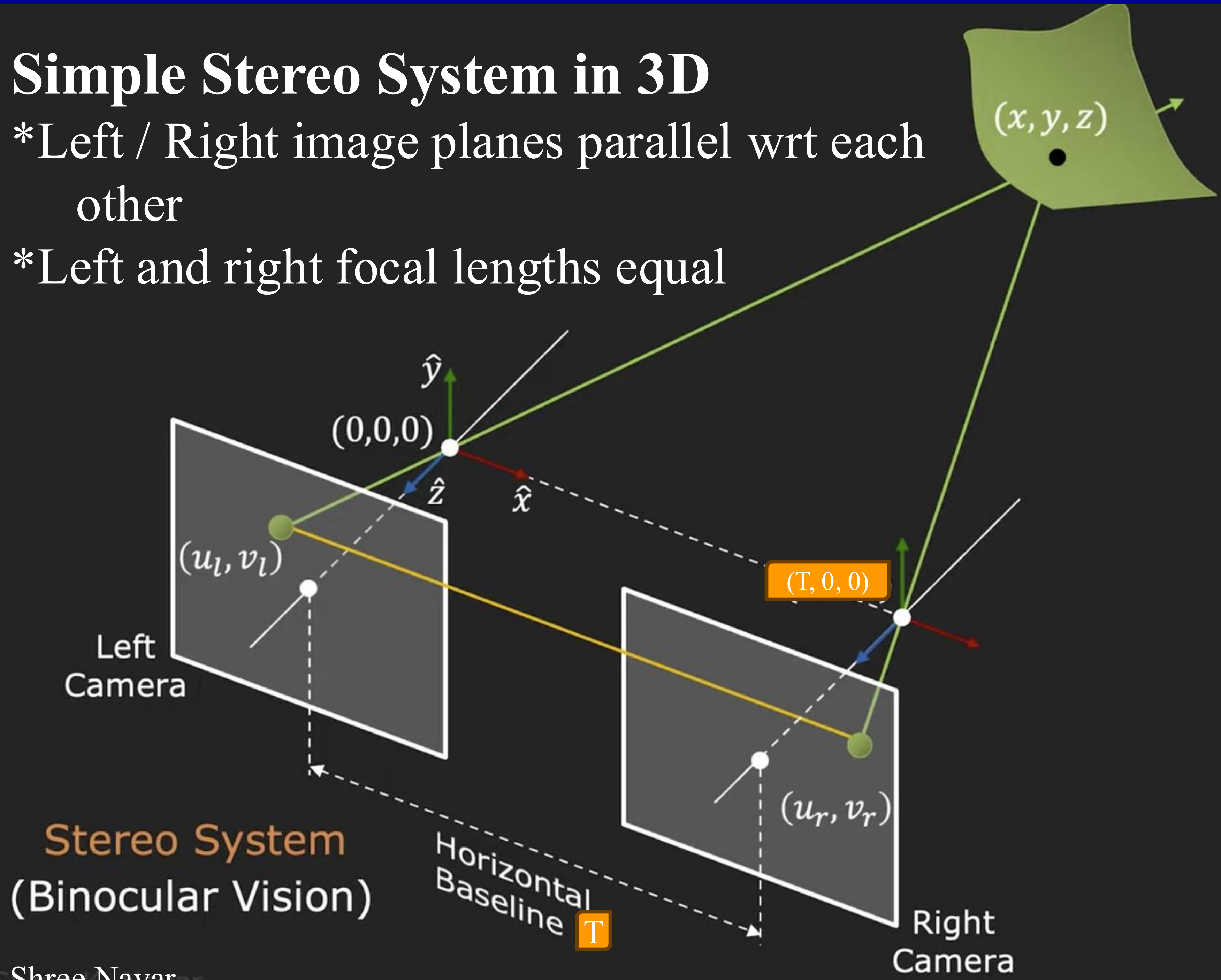
$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z} \Rightarrow Z = f\frac{T}{d}, d = x_l - x_r$$

d:disparity (difference in retinal positions).
T:baseline.
Large baselines: occlusions/foreshortening.

Simple Stereo System in 3D

*Left / Right image planes parallel wrt each other

*Left and right focal lengths equal



Shree Nayar

From perspective projection:

$$(u_l, v_l) = \left(f_x \frac{x}{z} + o_x, f_y \frac{y}{z} + o_y\right) \qquad (u_r, v_r) = \left(f_x \frac{x - T}{z} + o_x, f_y \frac{y}{z} + o_y\right)$$

Solving for (x, y, z):

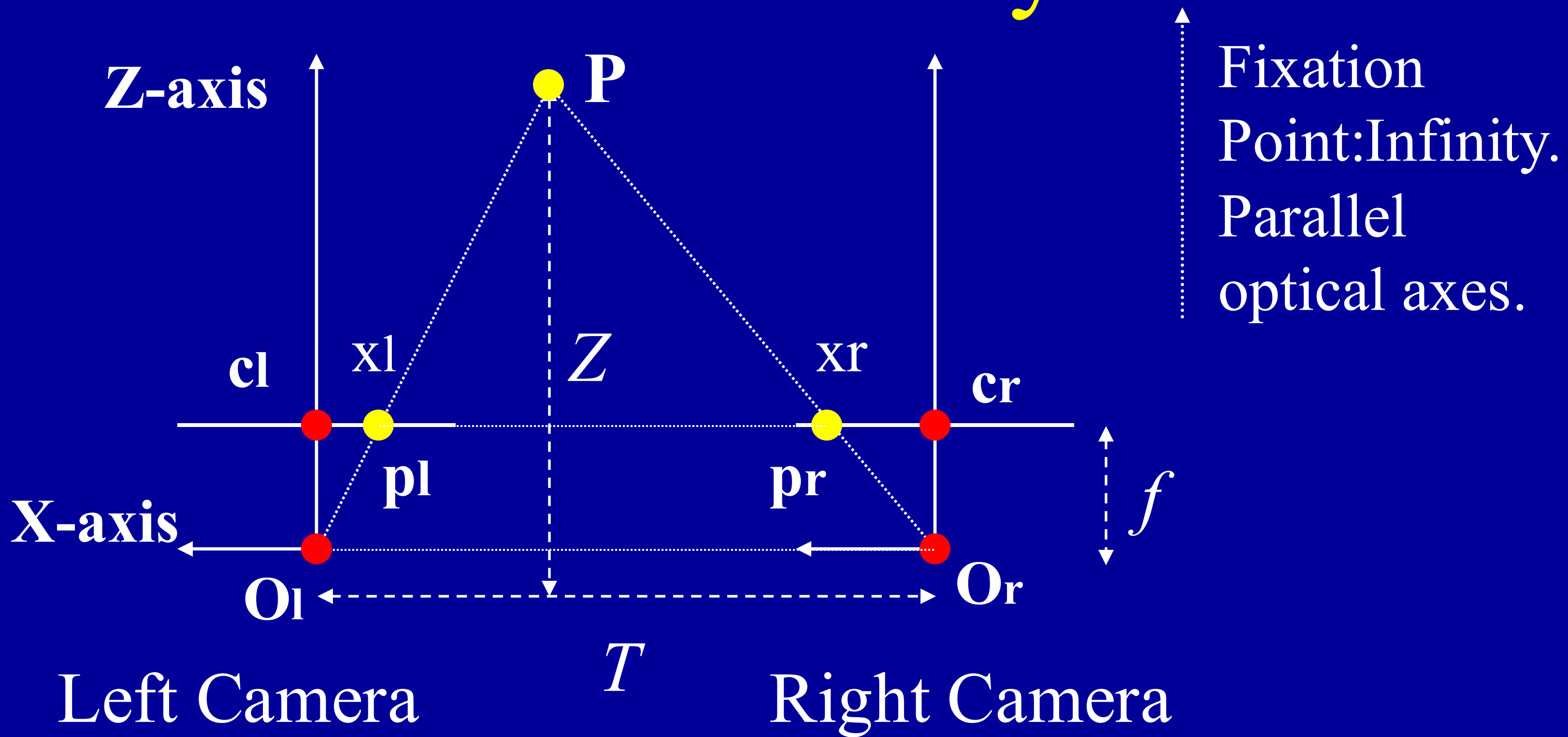
$$x = \frac{T(u_l - o_x)}{(u_l - u_r)} \qquad y = \frac{T f_x (v_l - o_y)}{f_y (u_l - u_r)} \qquad \boxed{z = \frac{T f_x}{(u_l - u_r)}}$$

where $(u_l - u_r)$ is called Disparity.

Depth z is inversely proportional to Disparity.

Shree Nayar

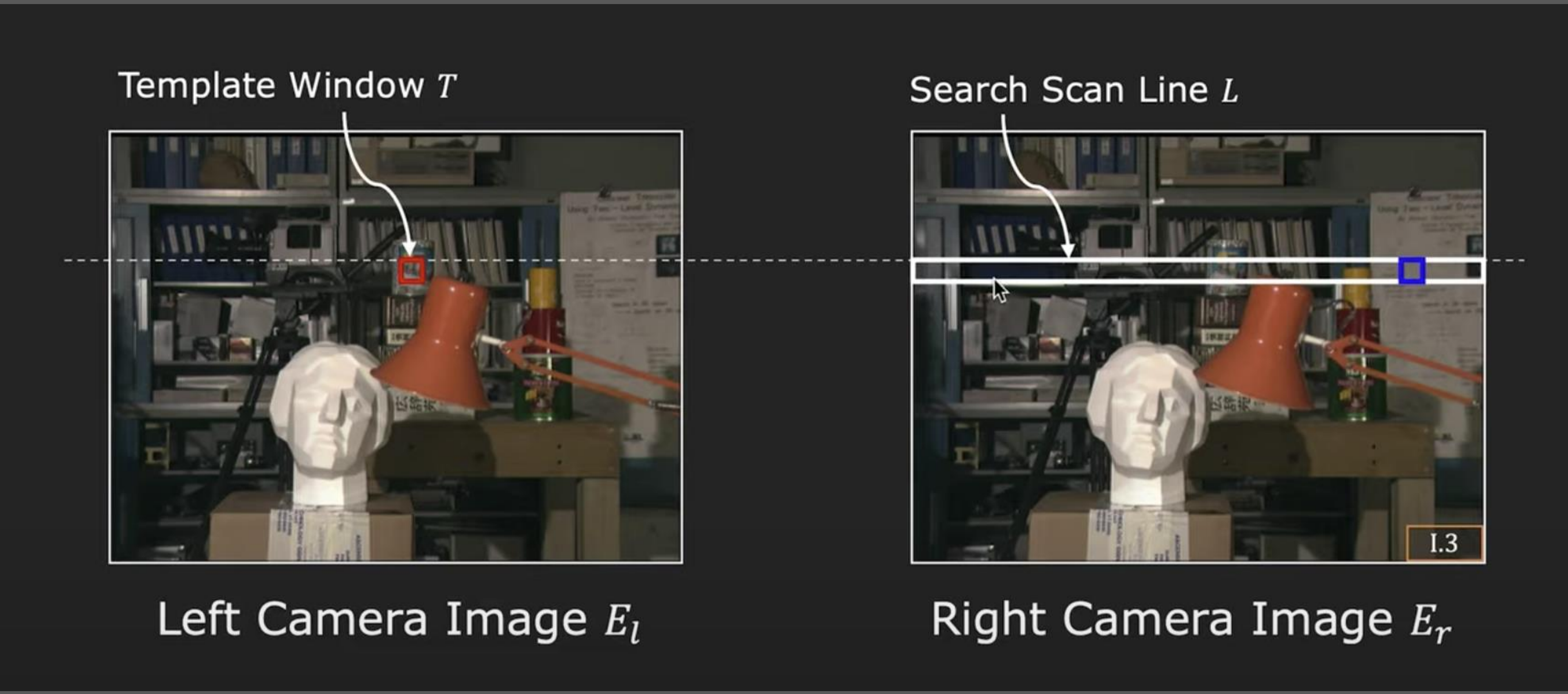
Parameters of Stereo System



1) Intrinsic parameters (i.e. f, cl, cr)
2) Extrinsic parameters: relative position and orientation of the 2 cameras.

<u>STEREO CALIBRATION PROBLEM</u>

Depth from disparity



Depth from disparity

image I(x,y)

Disparity map D(x,y)

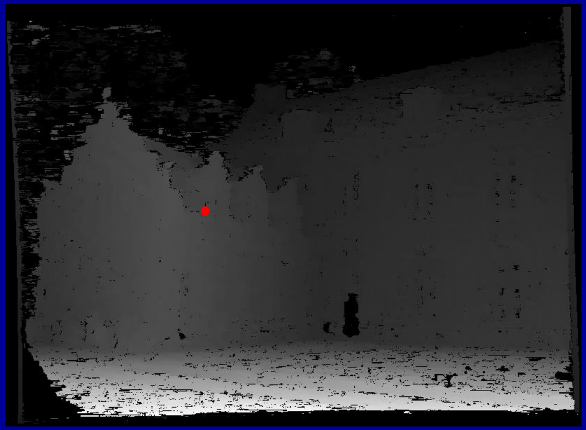
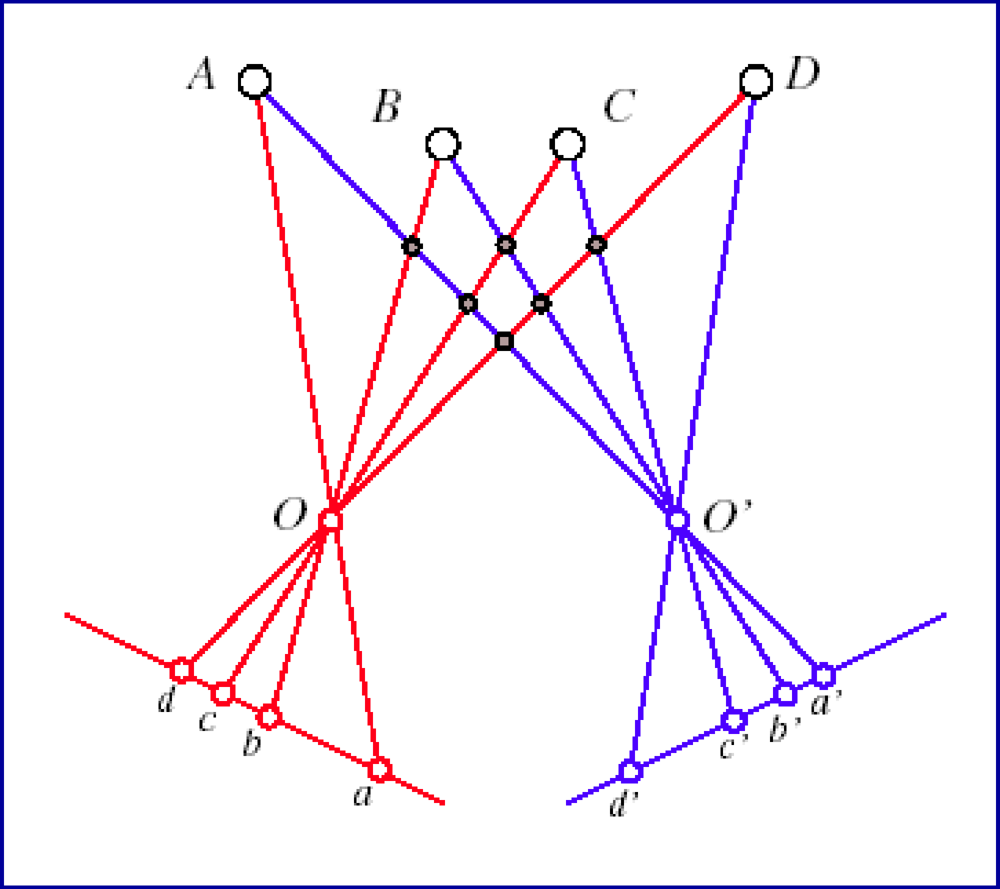


image I´(x´,y´)

(x´,y´)=(x+D(x,y), y)

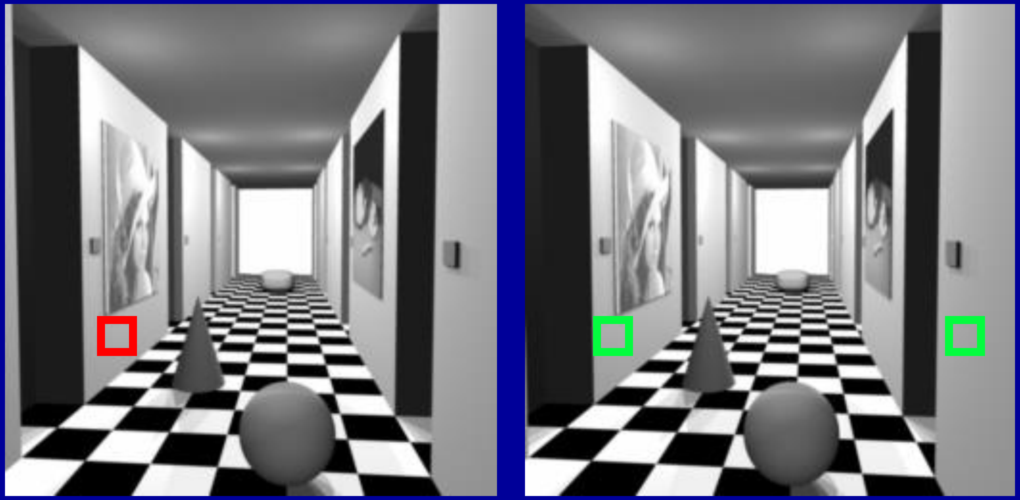
So if we could find the **corresponding points** in two images, we could **estimate relative depth**…

Correspondence Is Difficult



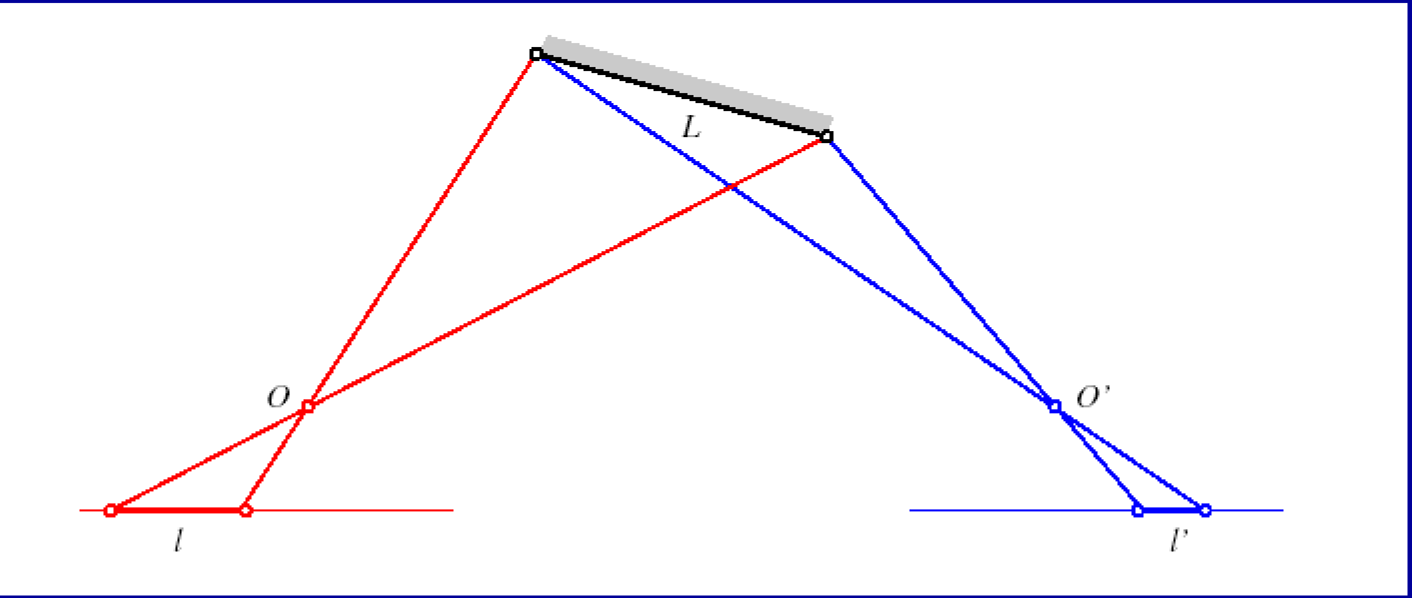
Ambiguity: there may be many possible 3D reconstructions.

Correspondence Is Difficult



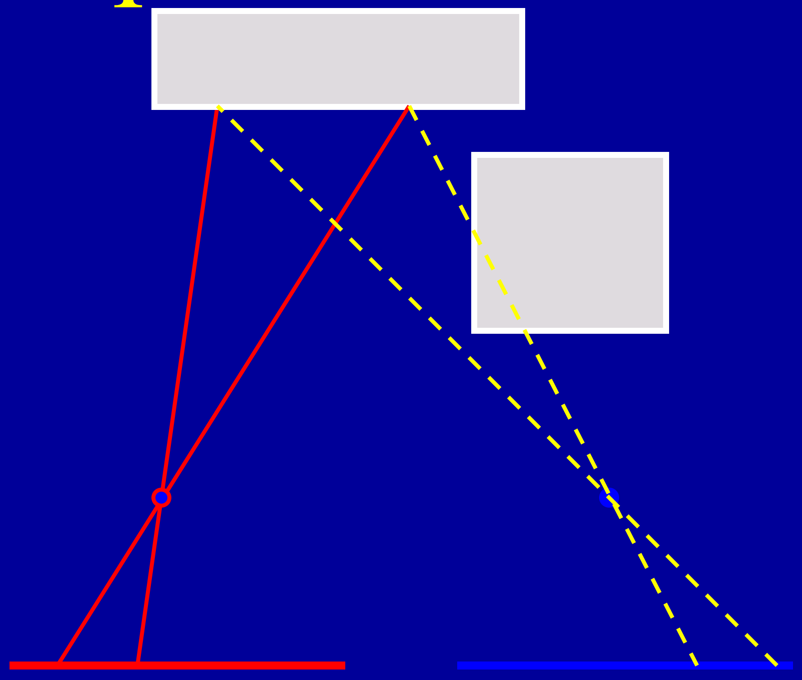
No texture: difficult to find a unique match.

Correspondence Is Difficult



Foreshortening: the projection in each image is different.

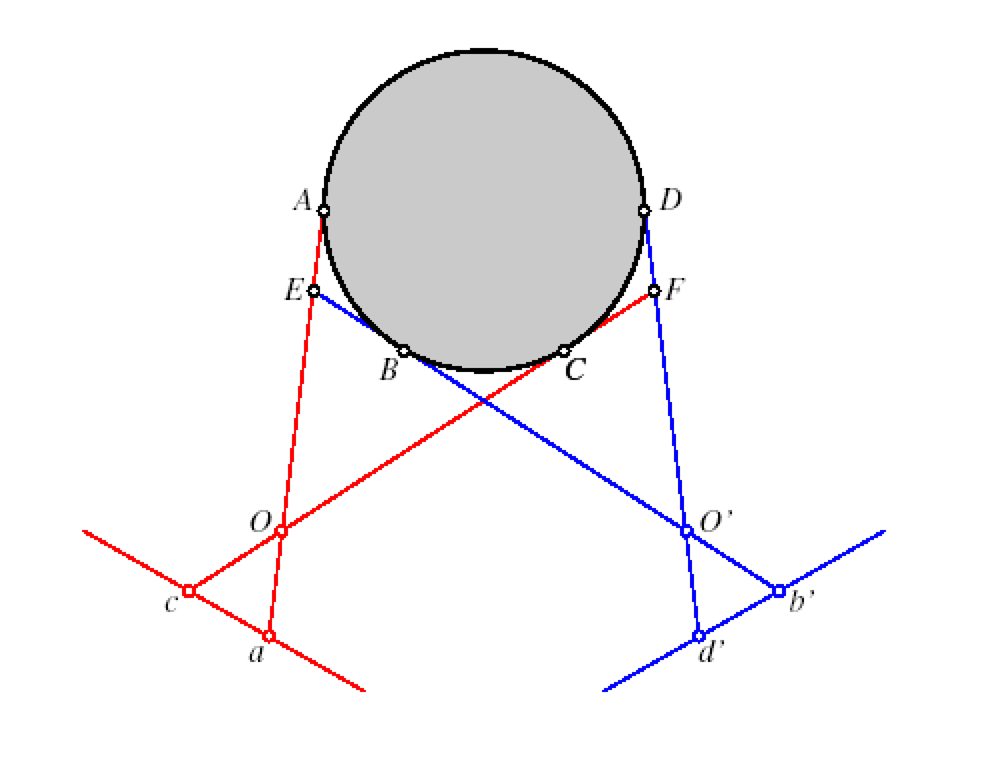
Correspondence Is Difficult



Occlusions: there may not be a correspondence.

Assumptions: 1) Most scene points are visible from both views.
2) Corresponding image regions are similar.

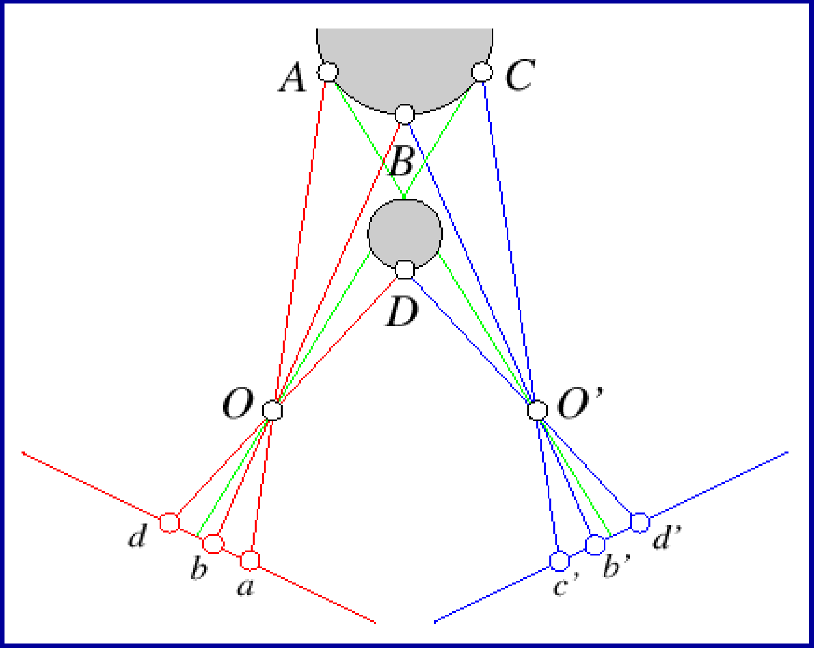
Correspondence Is Difficult



Curved surfaces: triangulation produces incorrect position.

Ordering Constraint

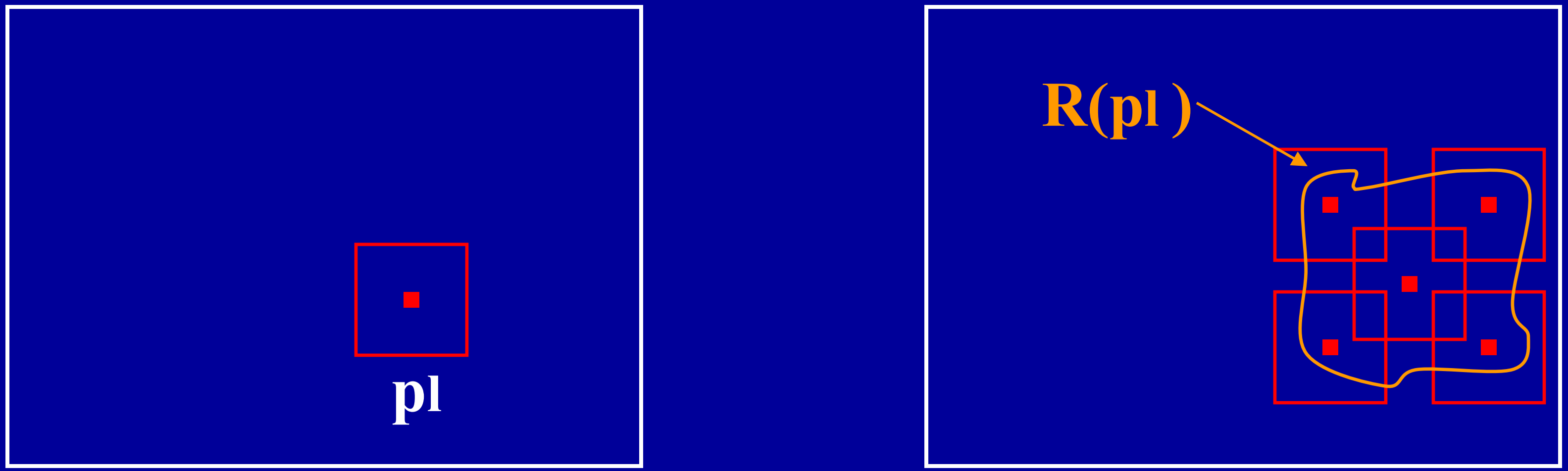
- Order of matching features usually the same in both images
- But not always: occlusion



Methods For Correspondence

- Correlation based (dense correspondences).
- Feature based (such as edges/lines/corners).

Correlation-Based Methods

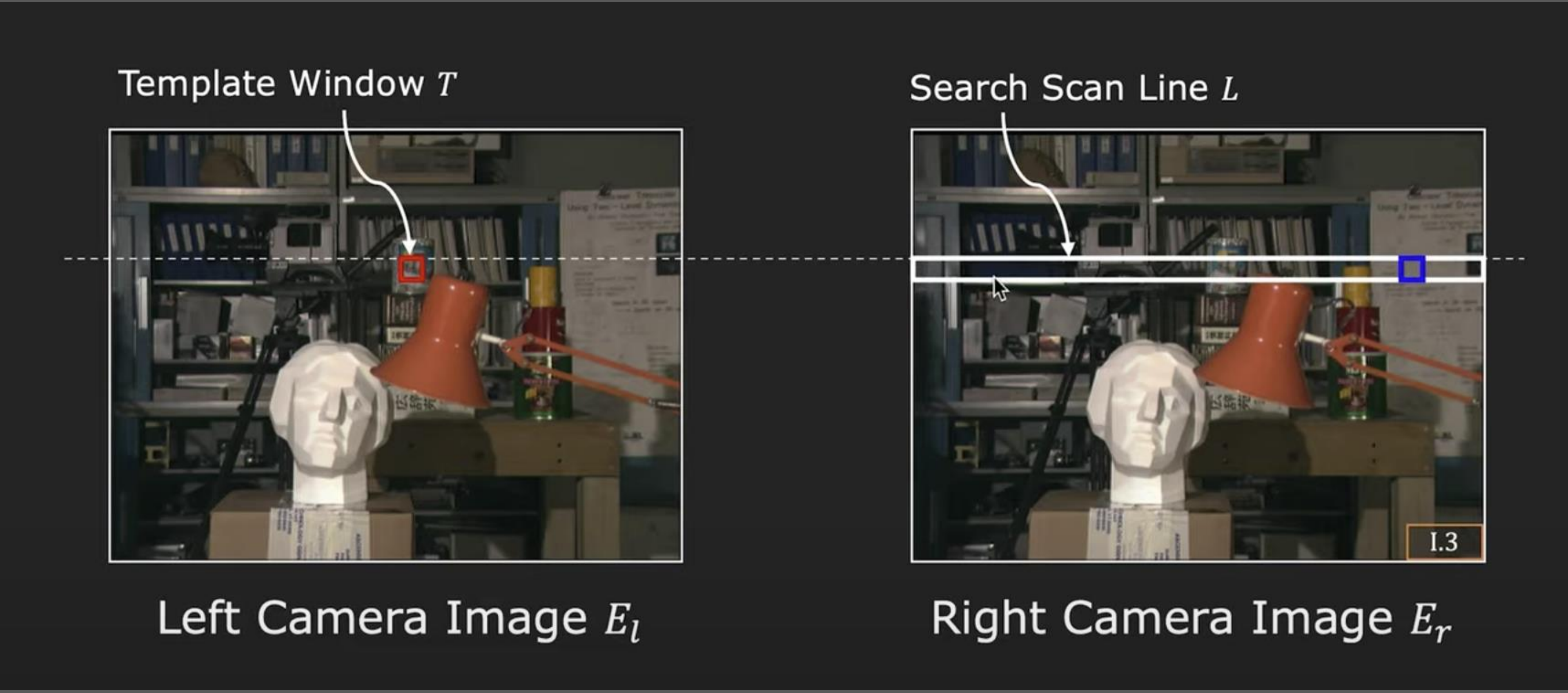


Left Image Right Image

1)For each pixel pl in the left image
search in a region R(pl) in the right image for corresponding pixel pr.
2)Use image windows of size (2W+1)x(2W+1).
3)Select the pixel pr that maximizes a correlation function.

HAVE TO SPECIFY: Region R, size W, and correlation function ψ.

Correlation-based



Special camera setup (image planes parallel to each other)

Similarity Metrics

Find pixel $(k,l) \in L$ with Minimum Sum of Absolute Differences:

$$SAD(k,l) = \sum_{(i,j)\in T} |E_l(i,j) - E_r(i+k, j+l)|$$

Find pixel $(k,l) \in L$ with Minimum Sum of Squared Differences:

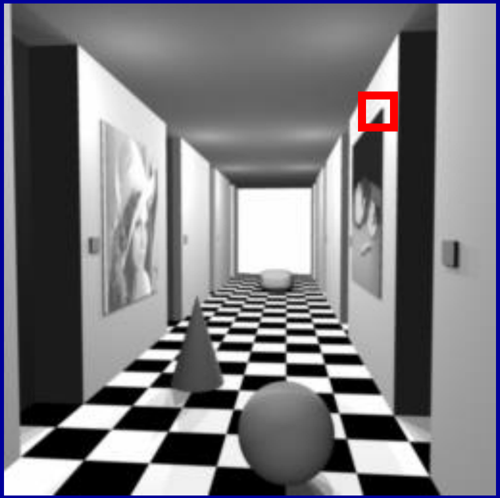
$$SSD(k,l) = \sum_{(i,j)\in T} |E_l(i,j) - E_r(i+k, j+l)|^2$$

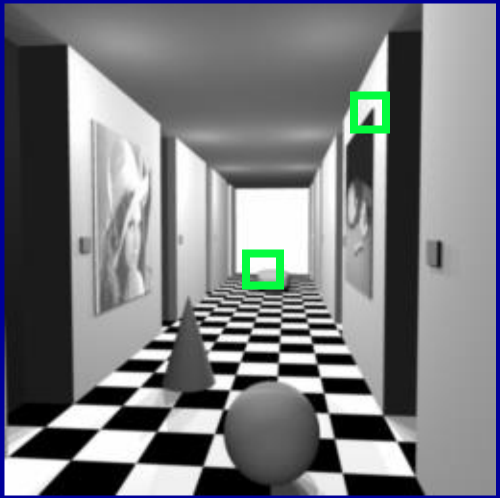
Find pixel $(k,l) \in L$ with Maximum Normalized Cross-Correlation

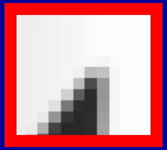
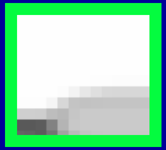
$$NCC(k,l) = \frac{\sum_{(i,j)\in T} E_l(i,j)\, E_r(i+k, j+l)}{\sqrt{\sum_{(i,j)\in T} E_l(i,j)^2 \sum_{(i,j)\in T} E_r(i+k, j+l)^2}}$$

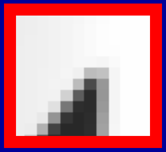
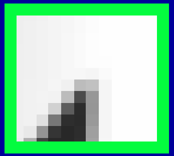
Shree Nayar

Correspondence

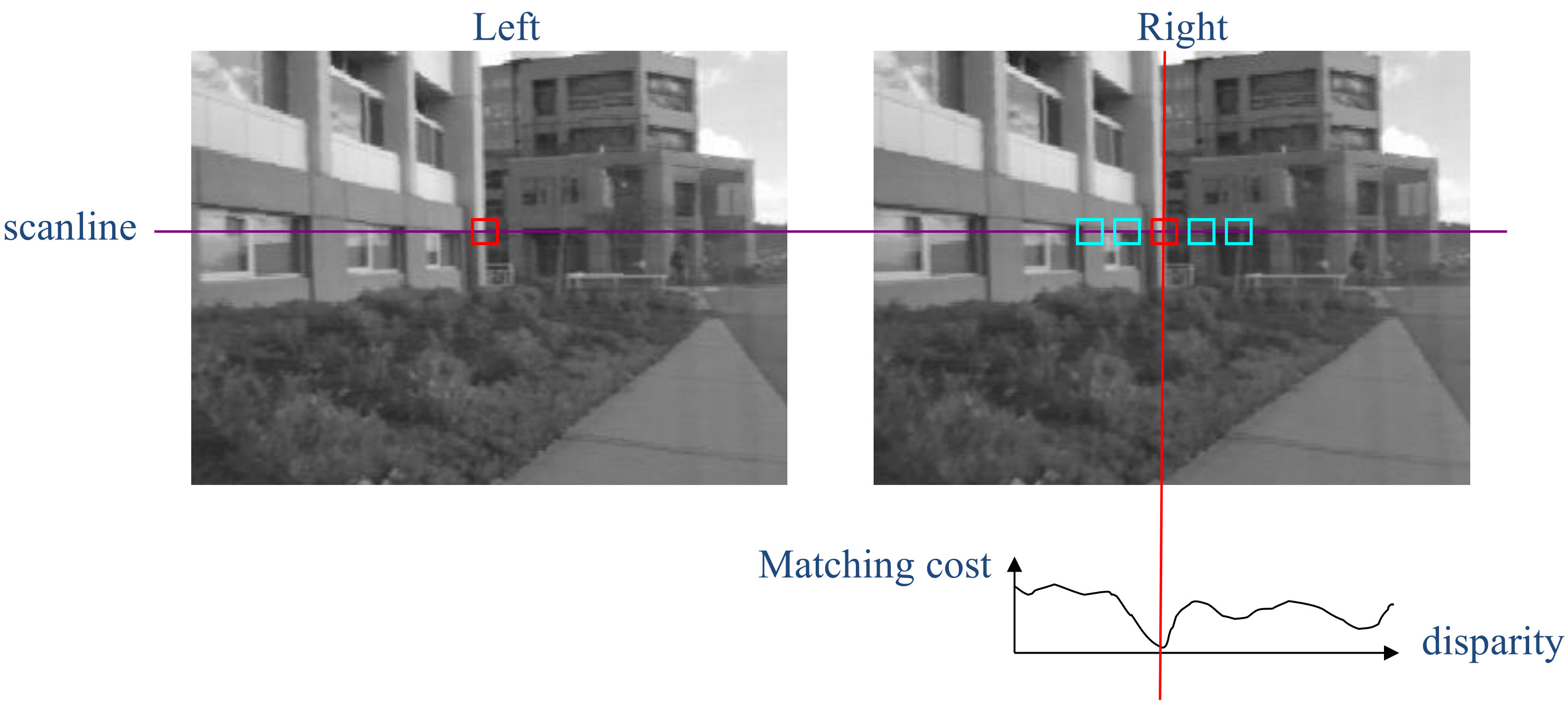




$\sum$ ( -  $)^2$ = ssd

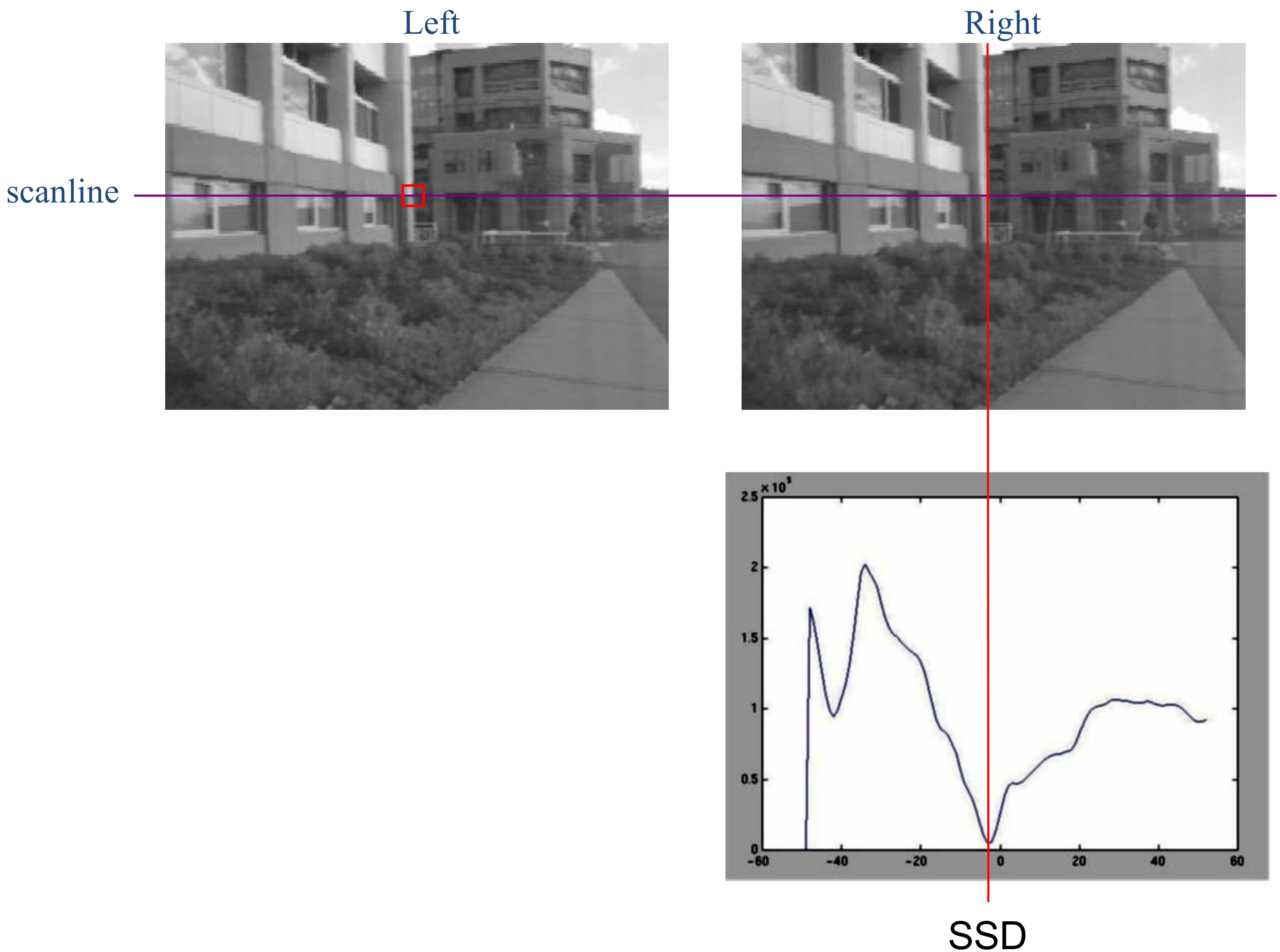
$\sum$ ( -  $)^2$ = ssd

Correspondence search

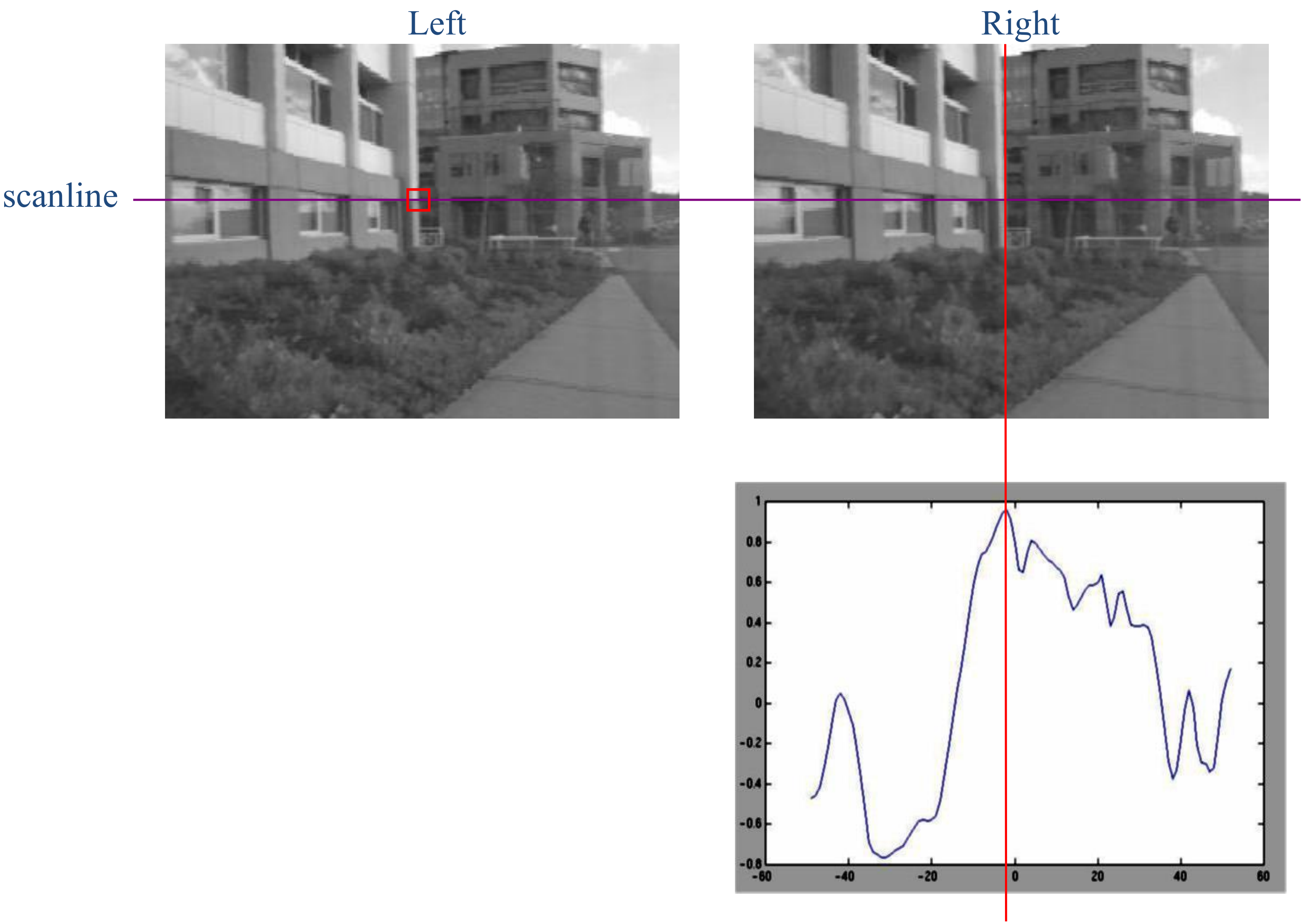


- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search



Correspondence search

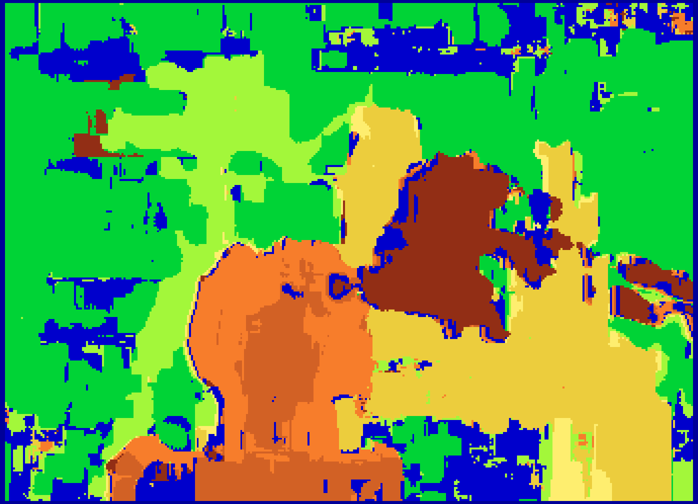


Norm. corr

Results with window search

Data

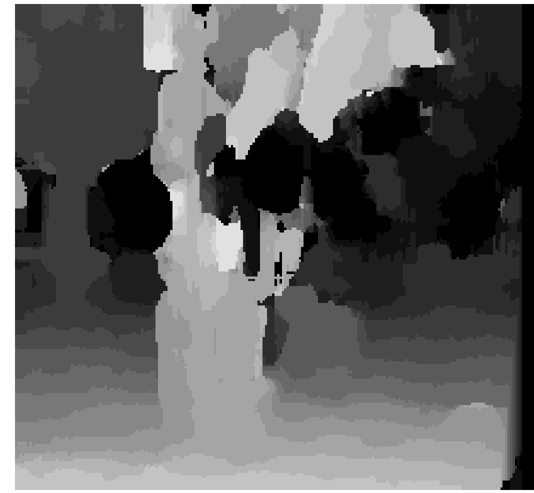
Window-based matching



Ground truth

Effect of window size

W = 3

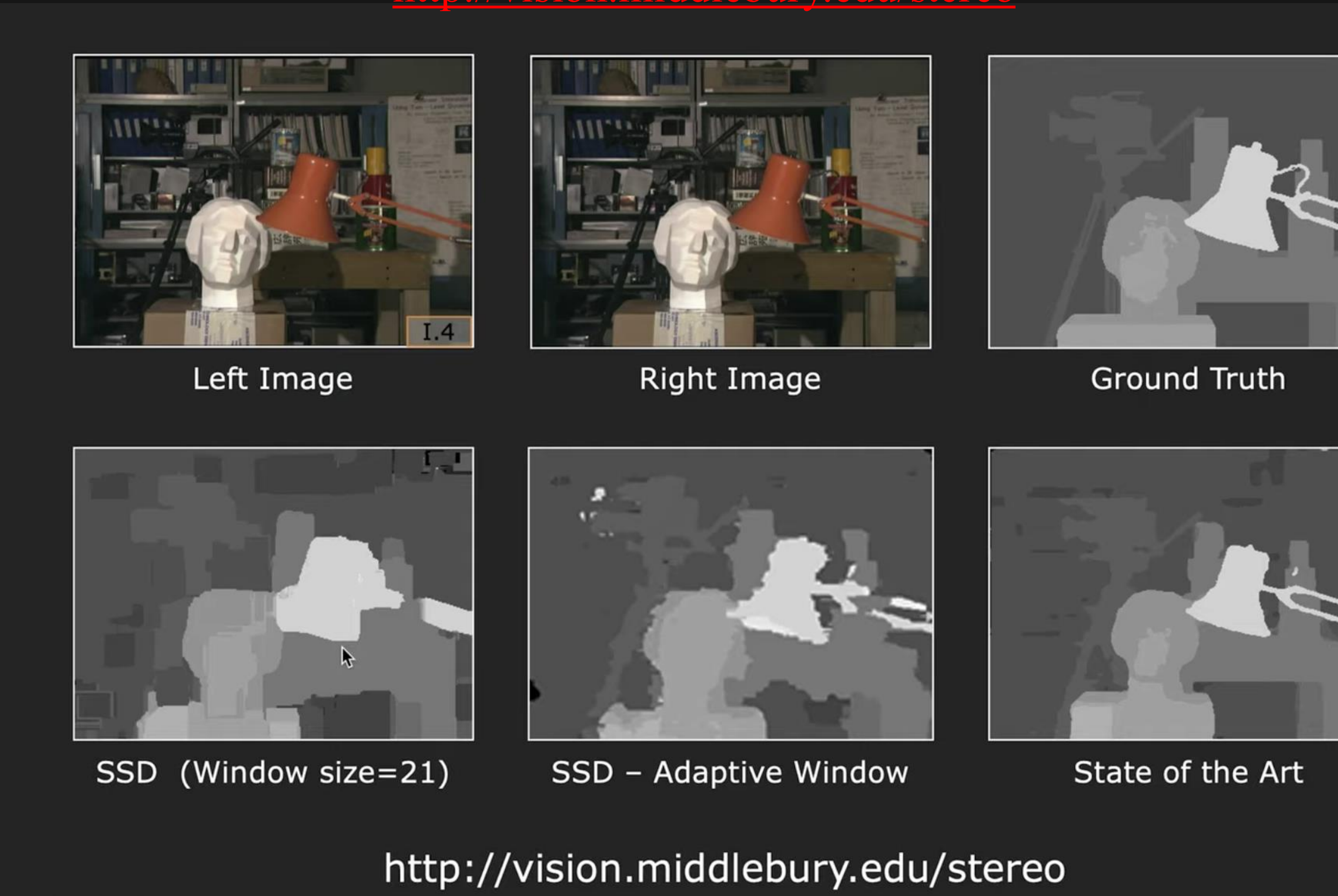


W = 20

- Smaller window
  - \+ More detail
  - – More noise

- Larger window
  - \+ Smoother disparity maps
  - – Less detail

http://vision.middlebury.edu/stereo

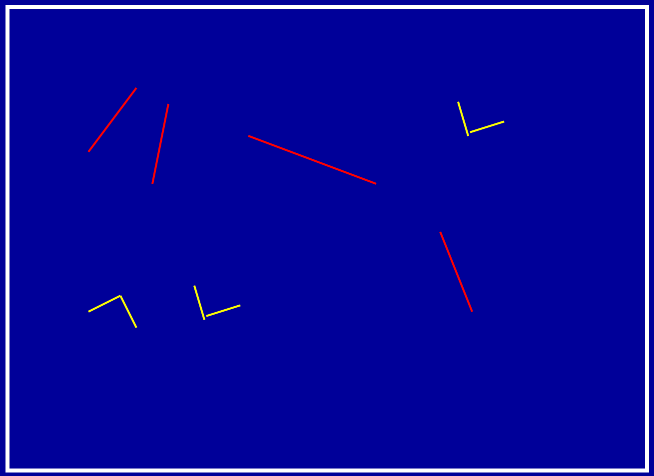


More recent methods

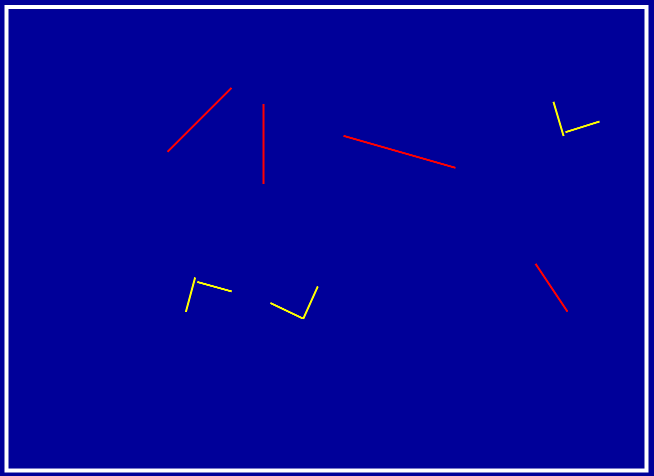
- Learn correspondences:

Paper link:
https://arxiv.org/pdf/1512.02134.pdf

Feature-Based Methods



Left Image



Right Image

Match **sparse** sets of extracted features.

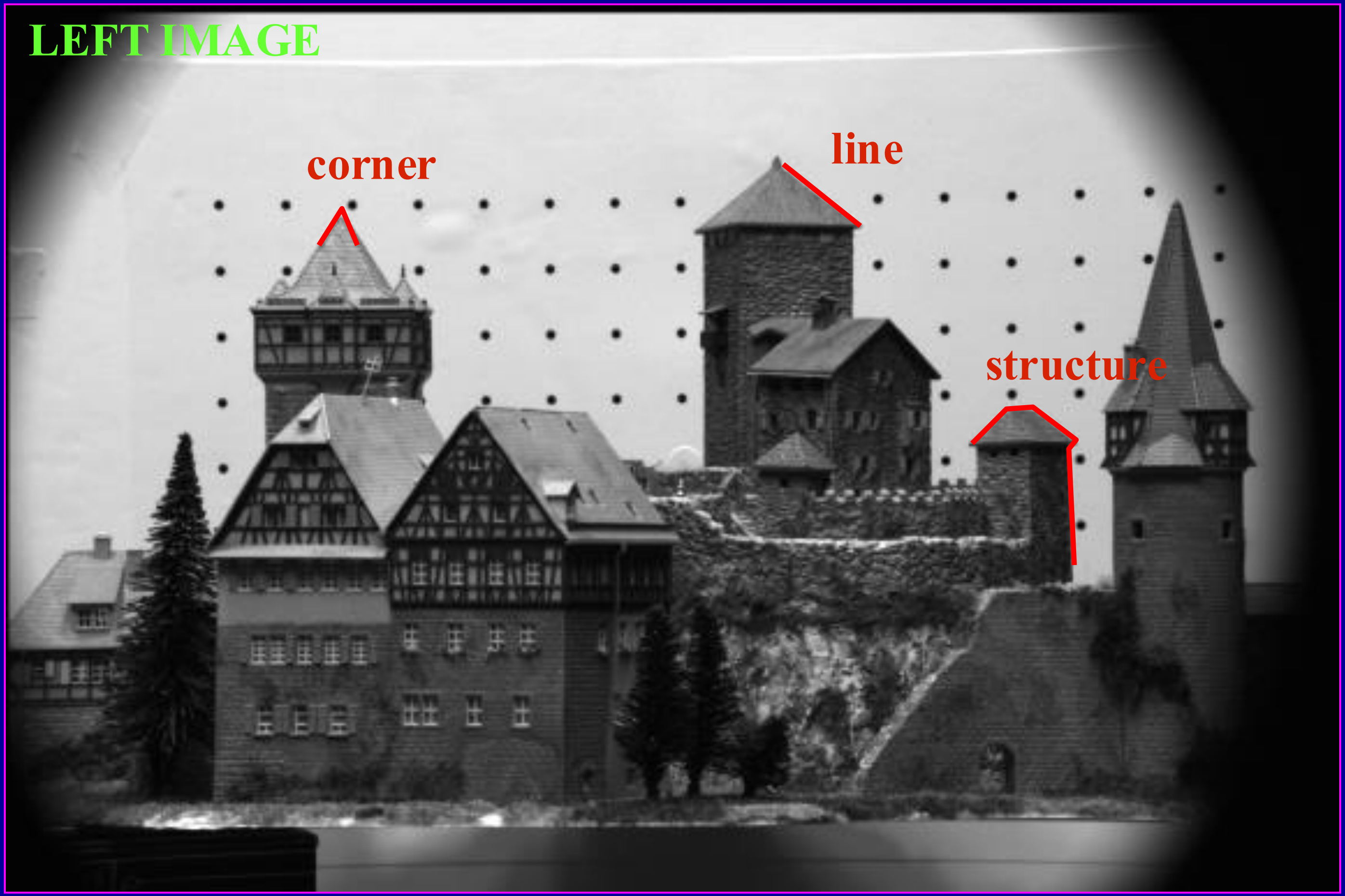
A feature descriptor for a line could contain:

length **l**, orientation **o**, midpoint **(x,y)**, average contrast **c**

An example similarity measure (w's are weights):

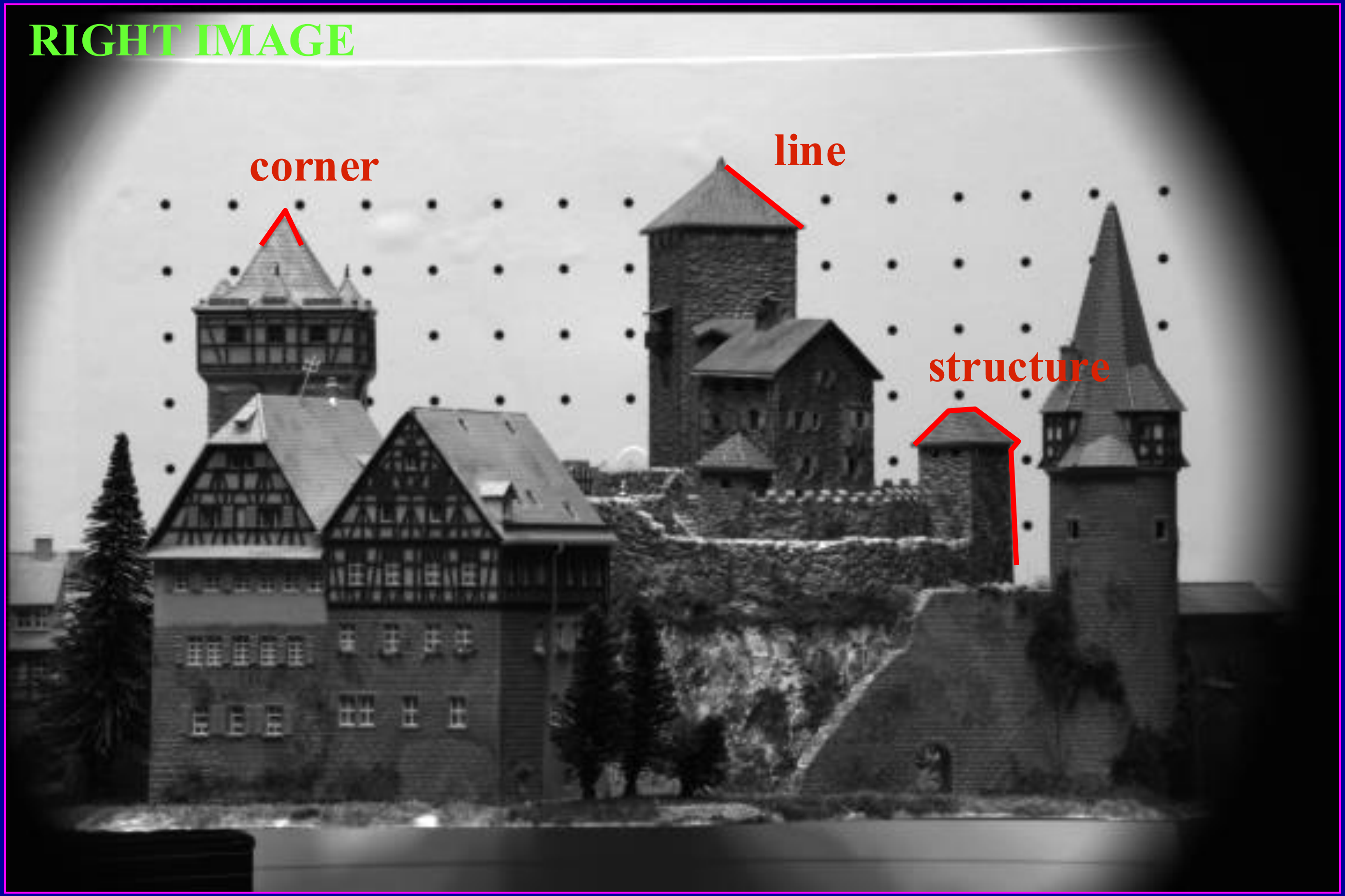
$$S = \frac{1}{w_0(l_l - l_r)^2 + w_1(\theta_1 - \theta_2)^2 + w_2(m_l - m_r)^2 + w_3(c_l - c_r)^2}$$

Correspondence By Features



From Sebastian Thrun/Jana Kosecka

Correspondence By Features



- Search in the right image… the disparity (dx, dy) is the displacement when the similarity measure is maximum

From Sebastian Thrun/Jana Kosecka

Comparison

Correlation-Based

- Dense depth maps.
- Need textured images
- Sensitive to foreshorening/illumination changes
- Need close views

Feature-Based

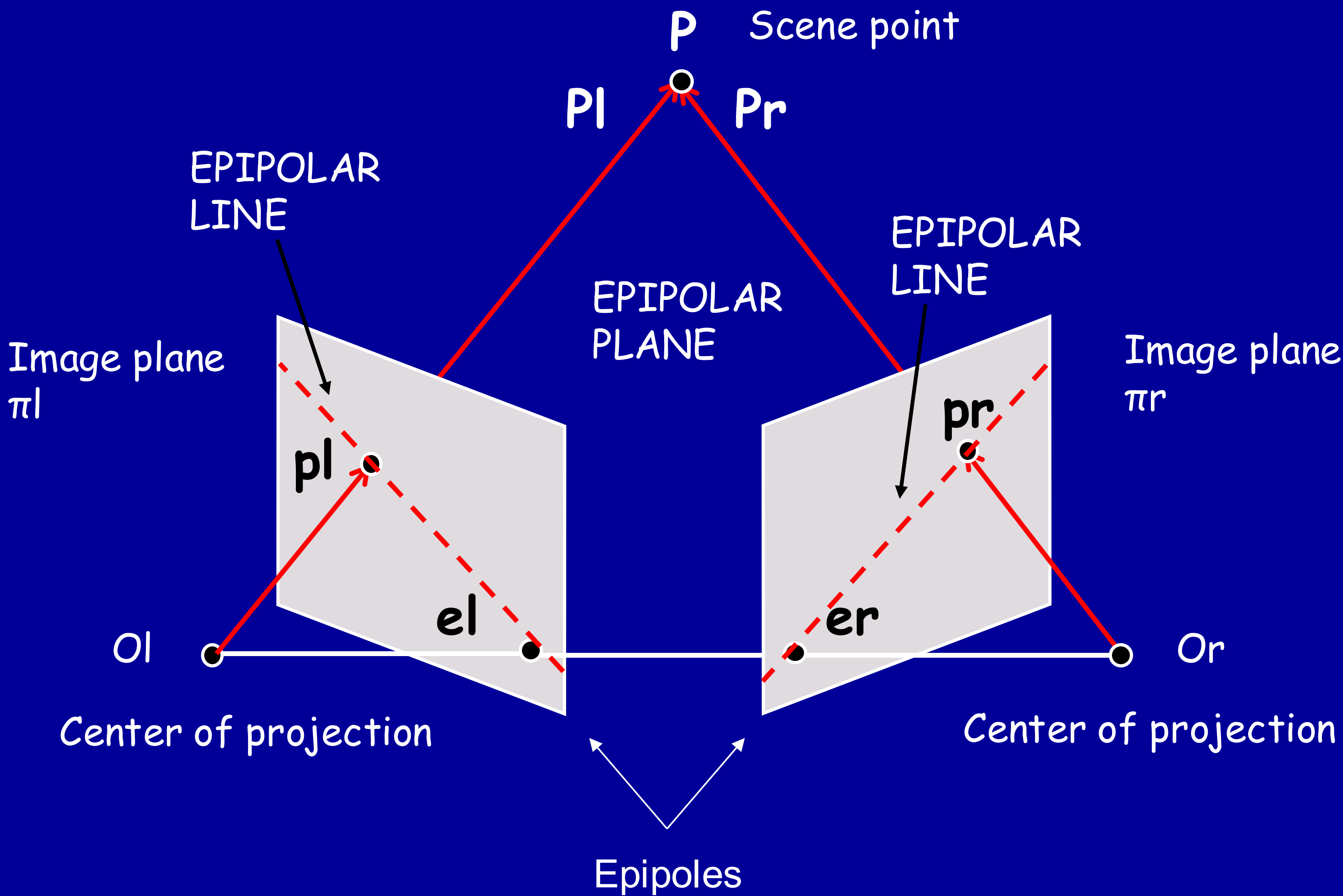
- Sparse depth maps.
- Insensitive to illumination changes.
- A-priori info used.
- Faster.

Problems: *occlusions/spurious matches:*
=>Introduce constraints in matching
(i.e. left-right consistency constraint)

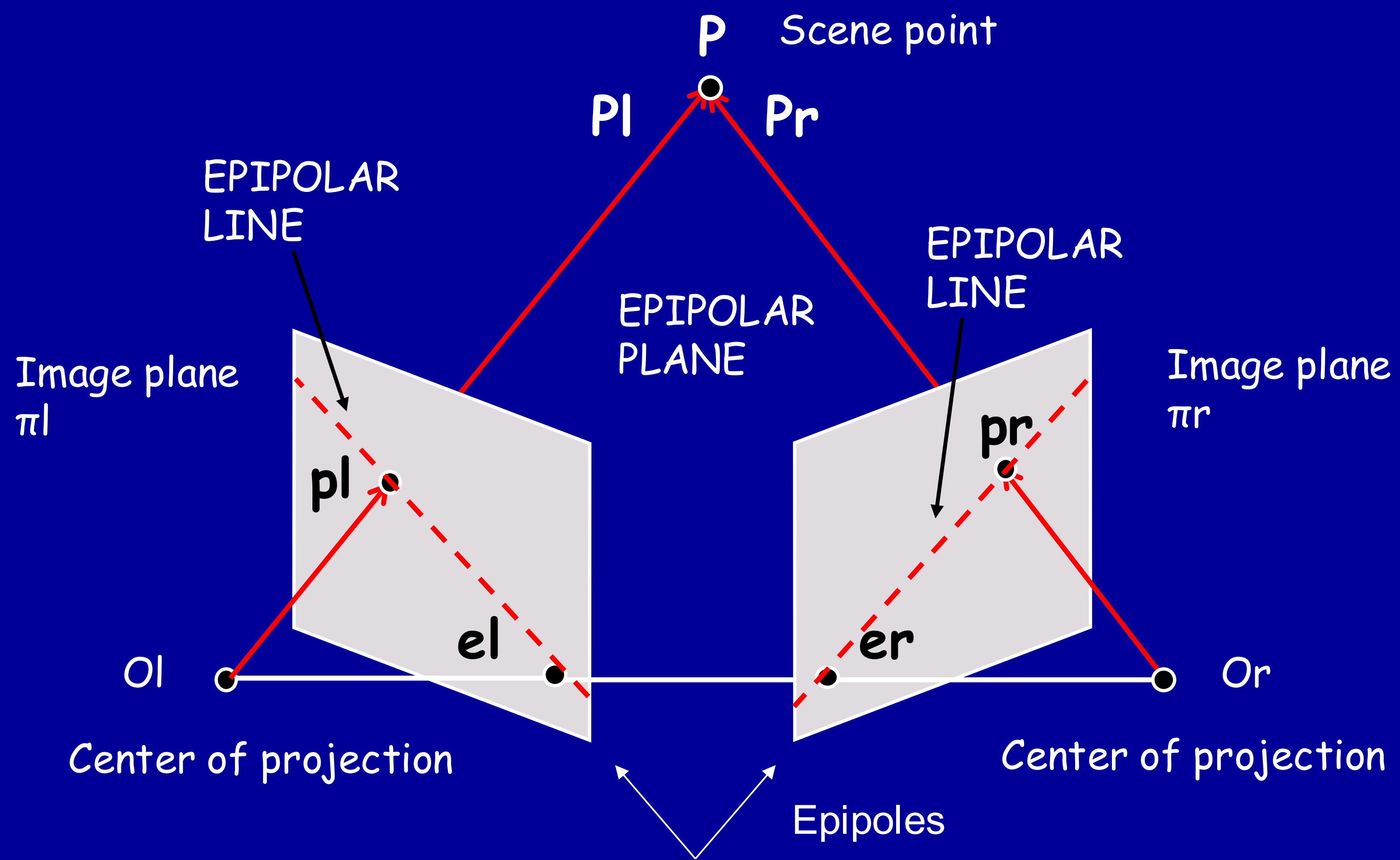
Evaluation Benchmarks

- Middlebury stereo dataset:
  - http://vision.middlebury.edu/stereo/
- KITTI
  - http://www.cvlibs.net/datasets/kitti/

Epipolar Constraint



Epipolar Constraint

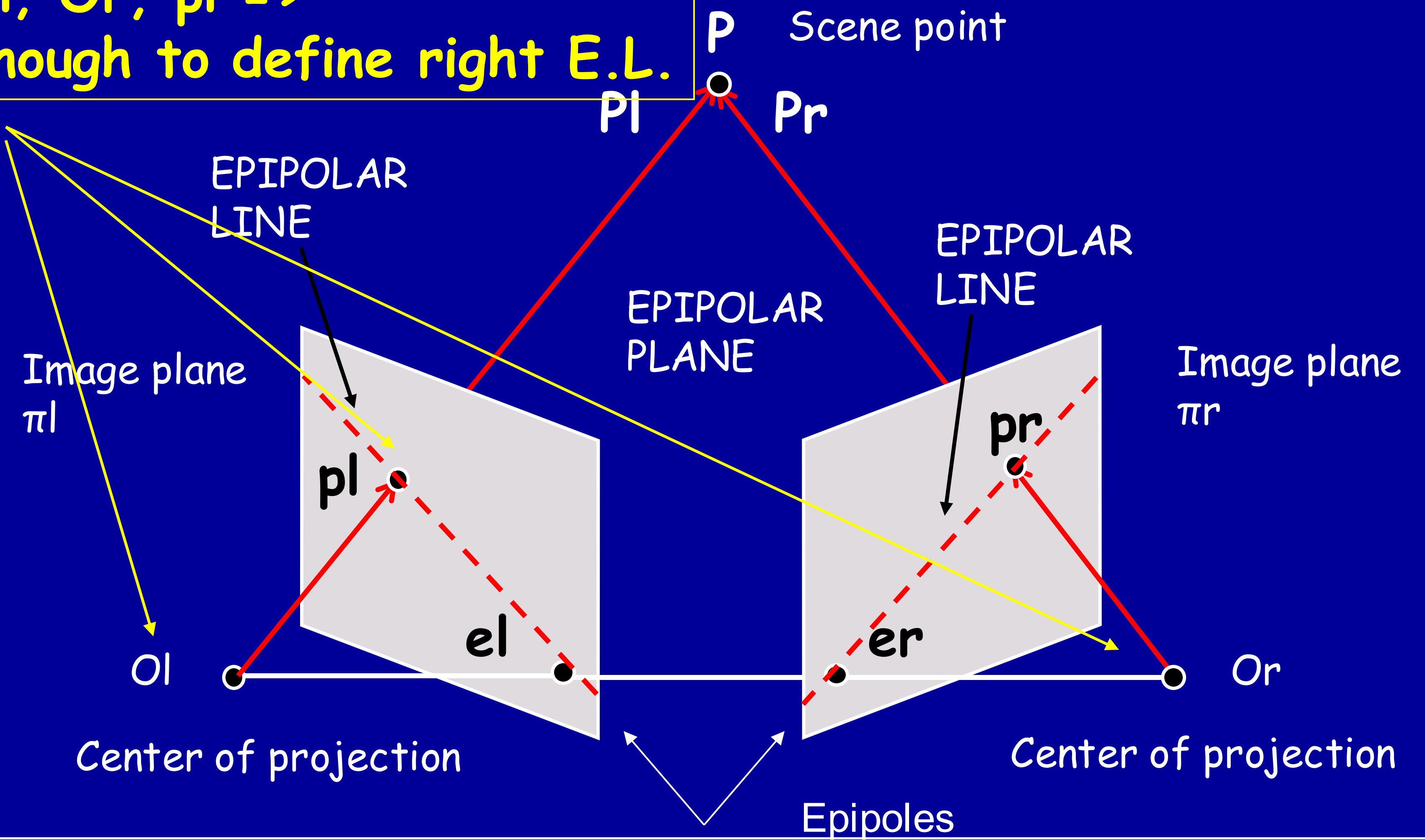


Extrinsic parameters: Left/Right Camera Frames:

Pr=R(Pl-T), T=Or-Ol (1)

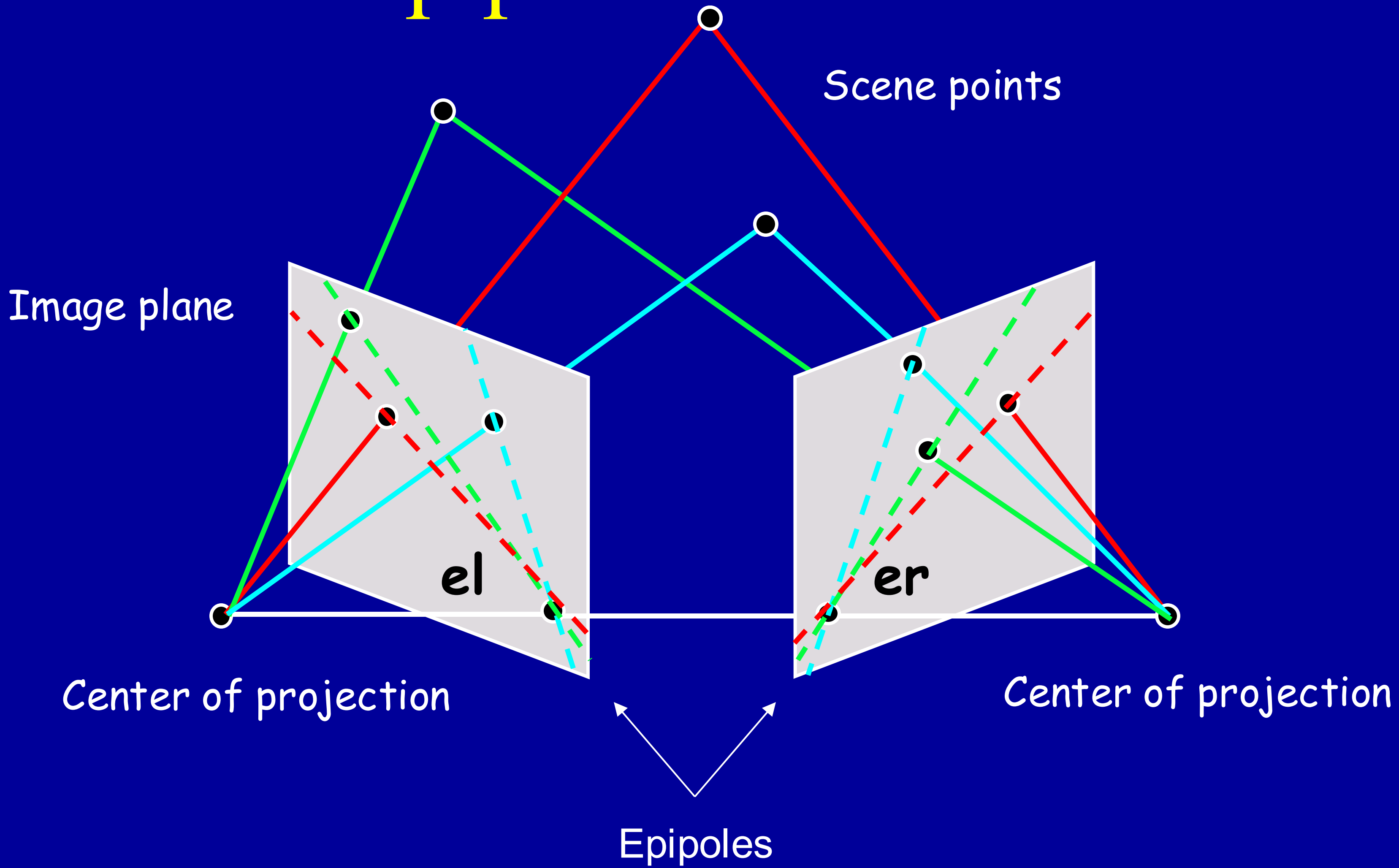
Epipolar Constraint

Ol, Or, pl =>
Enough to define right E.L.



Given pl, pr is constrained to lie on the Epipolar Line (E.L.).
For each left pixel pl, find the corresponding right E.L.
Searching for pr reduces to a 1-D problem.

Epipolar Constraint

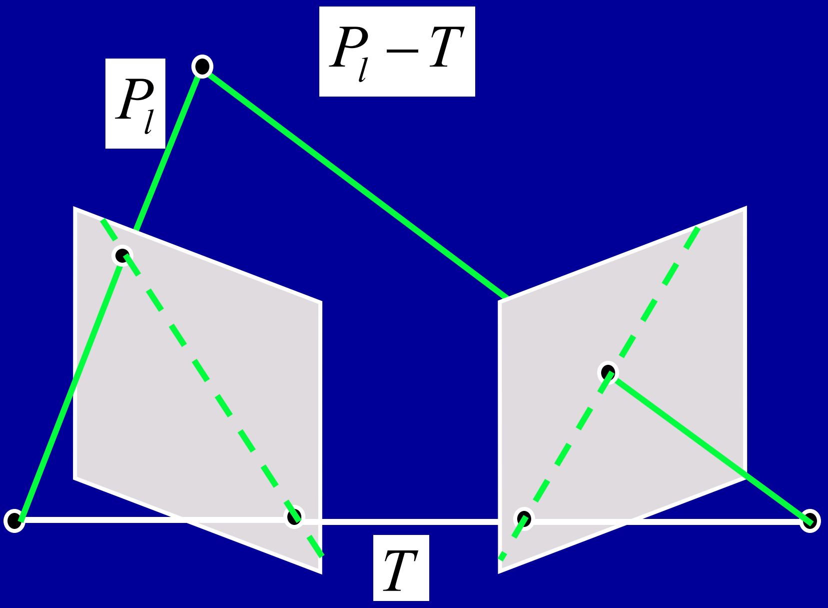


All E.L.s go through epipoles.
Parallel image planes => epipoles at infinity.

Essential Matrix

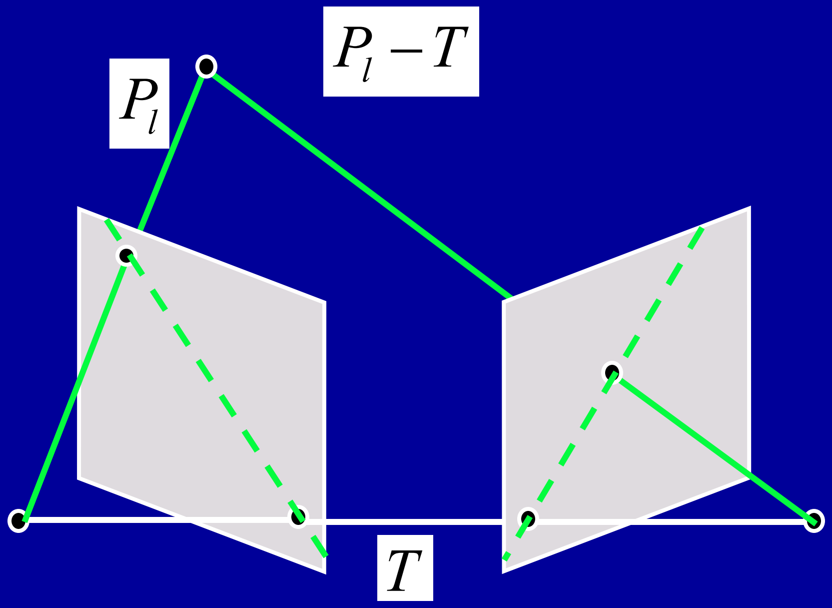
Estimate the epipolar geometry: correspondence between points and E.L.s.

P_l, P_l-T and T are coplanar



Essential Matrix

Estimate the epipolar geometry: correspondence between points and E.L.s.



P_l, P_l-T and T are coplanar

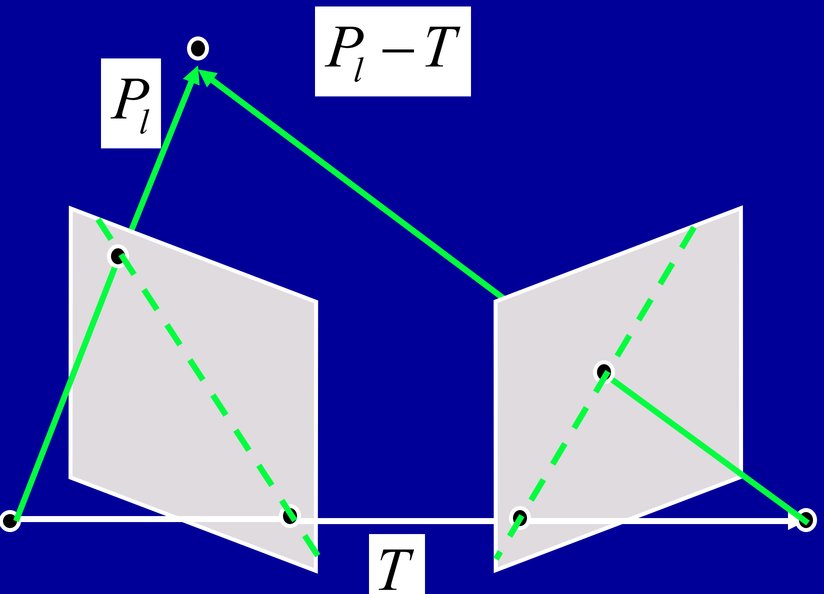
$$(P_l - T)^T T \times P_l = 0$$

(1)

$$(R^T P_r)^T \left(T \times P_l\right) = 0$$

Essential Matrix

Estimate the epipolar geometry: correspondence between points and E.L.s.



P_l, P_l-T and T are coplanar

$$(P_l - T)^T T \times P_l = 0$$

(1)

$$(R^T P_r)^T (T \times P_l) = 0$$

$$P_r^T R (T \times P_l) = 0$$

$$T \times P_l = S P_l$$

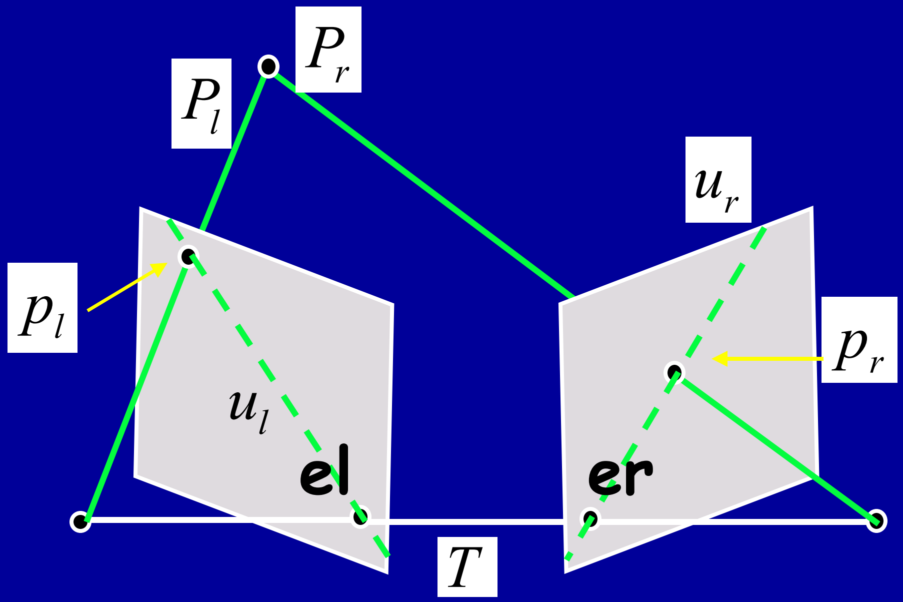
$$S = \begin{bmatrix} 0 & -T_Z & T_Y \\ T_Z & 0 & -T_X \\ -T_Y & T_X & 0 \end{bmatrix}$$

$$P_r^T (RS) P_l = 0$$

$$P_r^T E P_l = 0$$

Link bw/ epipolar constraint and extrinsic parameters of stereo system.

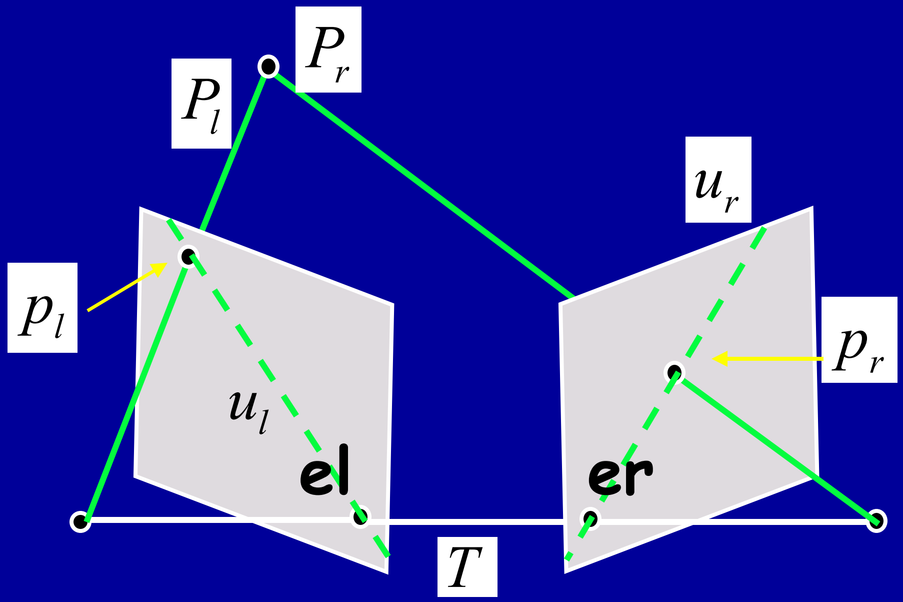
Essential Matrix (E)



$$P_r^{\,T} E P_l = 0$$

Essential Matrix (E)

$$P_r^T E P_l = 0$$

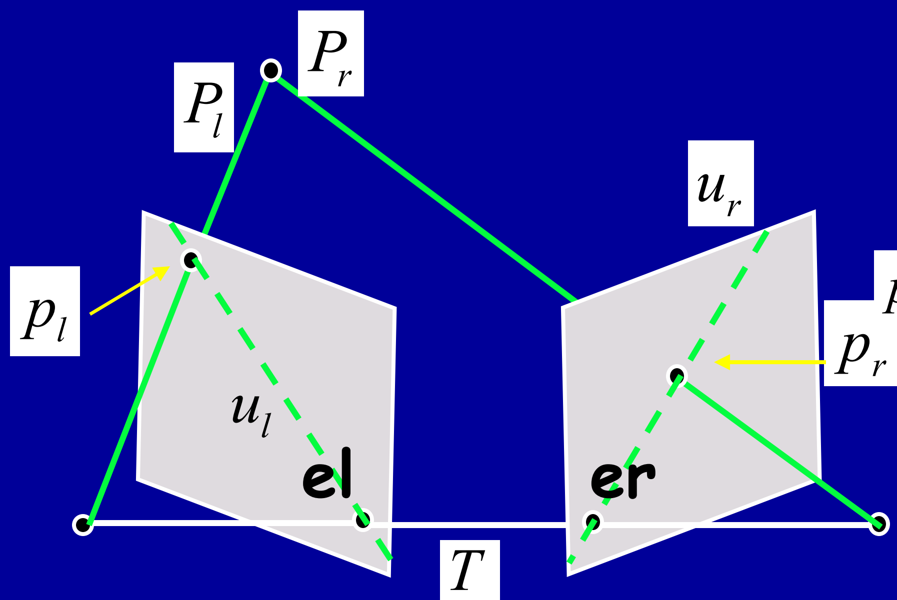


Perspective Projection:

$$p_r = \frac{f_r}{Z_r} P_r$$

$$p_l = \frac{f_l}{Z_l} P_l$$

Essential Matrix (E)



$$P_r^T E P_l = 0$$

Perspective Projection

$$p_r^T E p_l = 0$$

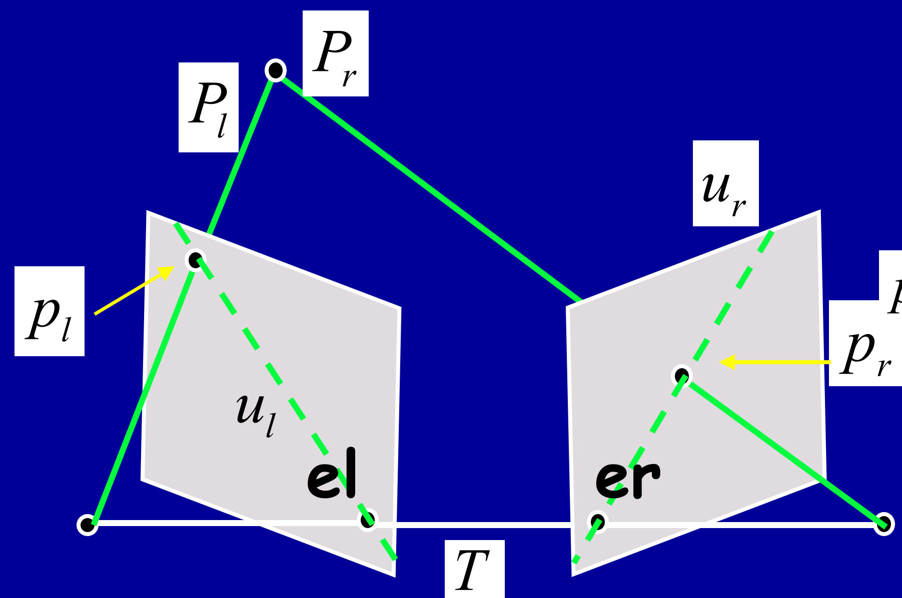
p_l and p_r are points in camera coordinates

$$p_r^T \underbrace{R\begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}}_{E} p_l = 0$$

Essential matrix

Rank 2

Essential Matrix (E)



$$p_r^{\,T} E p_l = 0$$

p_l **and** p_r **are points in camera coordinates**

$$\underbrace{p_r^{\,T} R \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}}_{E} p_l = 0$$

<u>Essential matrix</u>

Rank 2

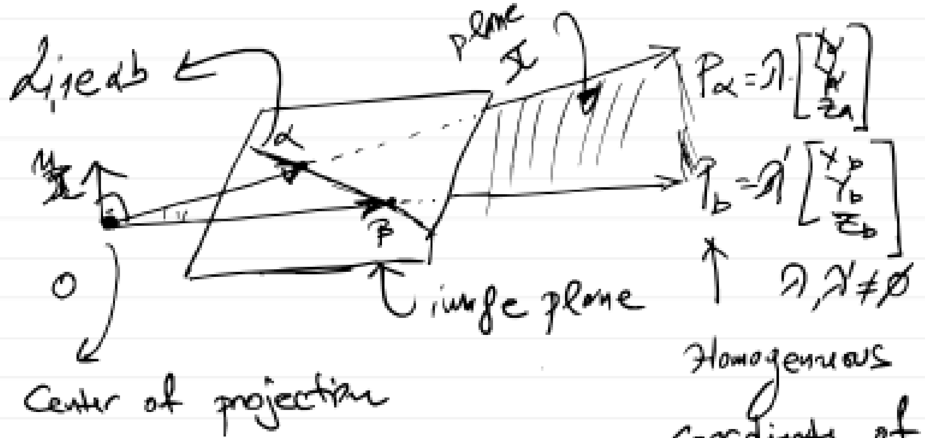
Epipolar lines are found by

$$u_r = E p_l$$

$$u_l = E^T p_r$$

Note:

Lines in Projective space



So α & β can be represented by 3-vectors P_α & P_b with any $\lambda, \lambda' \neq 0$

The image coordinates of α for example are:

$$\left(\frac{\lambda x_\alpha}{\lambda z_\alpha}, \frac{\lambda y_\alpha}{\lambda z_\alpha} \right) = \left(\frac{x_\alpha}{z_\alpha}, \frac{y_\alpha}{z_\alpha} \right)$$

Line αb is the intersection b/w image plane and plane $OP_\alpha P_b$ (π) with normal $n_\pi = \pm \text{Normalize}(P_\alpha \times P_b)$

Therefore a line can be represented by a 3-vector n_π in projective space!

Note:

Lines in Projective space

Every point p on line ab satisfies equation

DOT PRODUCT $\rightarrow$ $u_\pi^T \cdot p = 0$.

So for example if $u_\pi = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$

and p is a point $\begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix}$ in

projective space, then:

$$u_\pi^T p = 0 \Rightarrow \underline{u_x \cdot p_x + u_2 \cdot p_y + u_3 = 0}$$

Camera Models (linear versions)

$$M_{int} = \begin{pmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{pmatrix}$$

Elegant decomposition.
No distortion!

$$M_{ext} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \mathrm{Tx} \\ r_{21} & r_{22} & r_{23} & \mathrm{Ty} \\ r_{31} & r_{32} & r_{33} & \mathrm{Tz} \end{pmatrix},$$

The Linear Matrix Equation of Perspective Projections

Homogeneous
Coordinates

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = M_{int} M_{ext} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}.$$

Measured Pixel
$(x_{im}=x_1/x_3,\ y_{im}=x_2/x_3)$

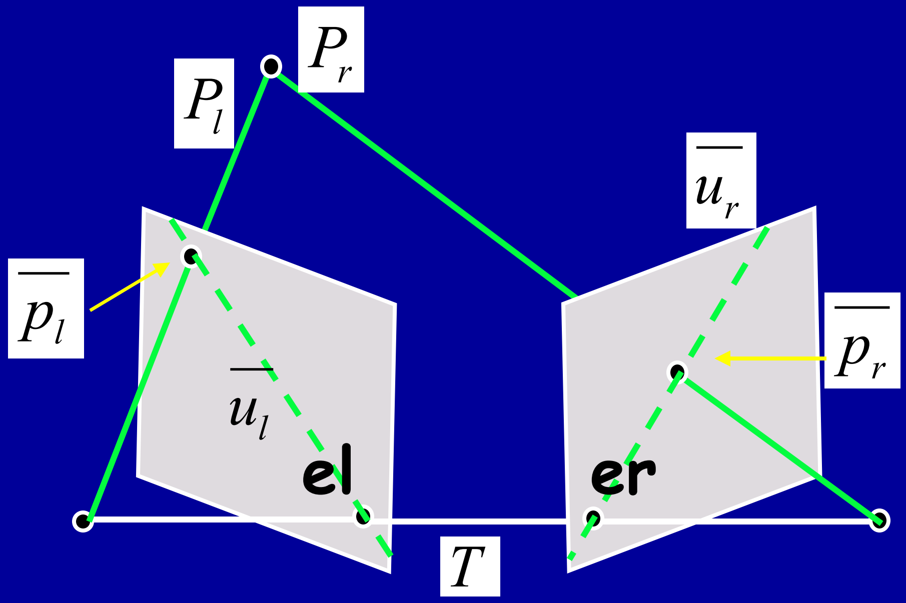
World Point
(Xw, Yw,Zw)

?

Fundamental Matrix (F)

Ml (Mr) matrix of intrinsic parameters for left (right) camera.

$$p_r^T E p_l = 0$$



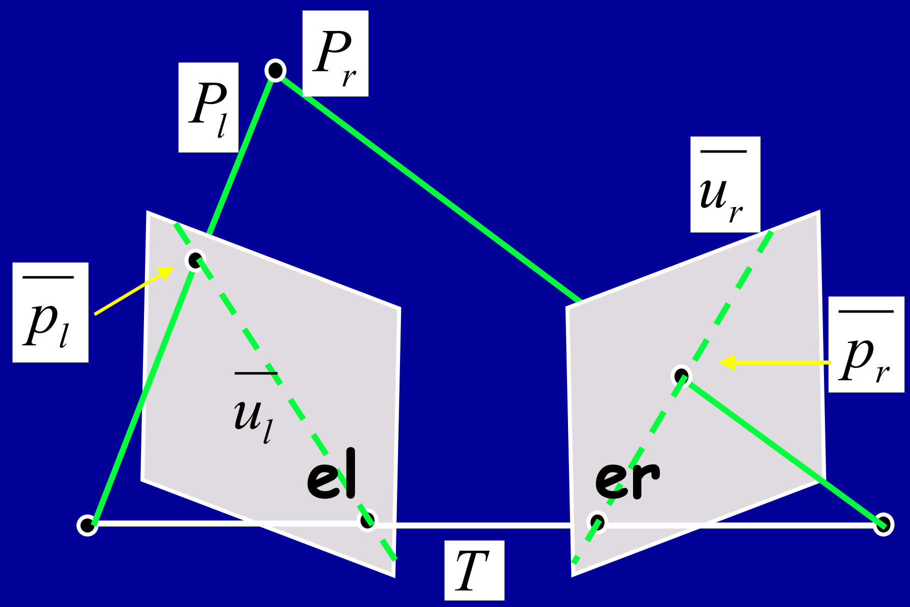
Converting to pixel coordinates

$$\overline{p}_l = M_l p_l$$

$$\overline{p}_r = M_r p_r$$

Fundamental Matrix (F)

Ml (Mr) matrix of intrinsic parameters for left (right) camera.



Fundamental matrix F

$$\overline{p}_r^{\,T} F \overline{p}_l = 0$$

$$F = M_r^{-T} E M_l^{-1}$$

F: pixel coordinates !

E: camera coordinates !

Epipolar lines:

$$\overline{u_r} = F\overline{p}_l$$

$$\overline{u_l} = F^T \overline{p}_r$$

Conclusions

Essential Matrix

- Encodes information on extrinsic parameters.
- Has rank 2.
- Its 2 non-zero singular values are equal.

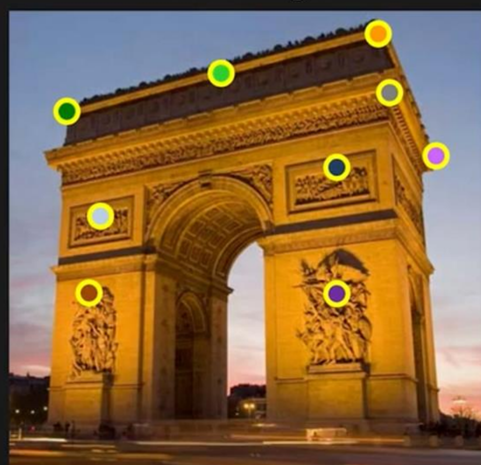
Fundamental Matrix

- Encodes information on both the extrinsic and intrinsic parameters.
- Has rank 2.

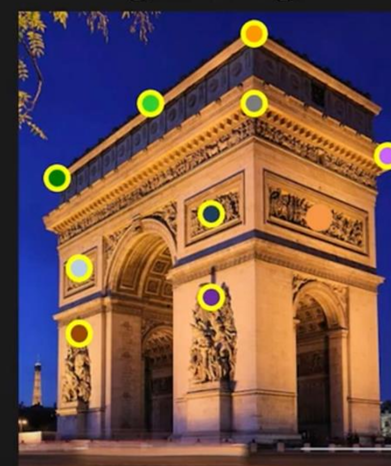
Stereo Calibration
Finding the epipolar geometry

Find a set of corresponding features in left and right images
(e.g. using SIFT or hand-picked)

Left image



Right image



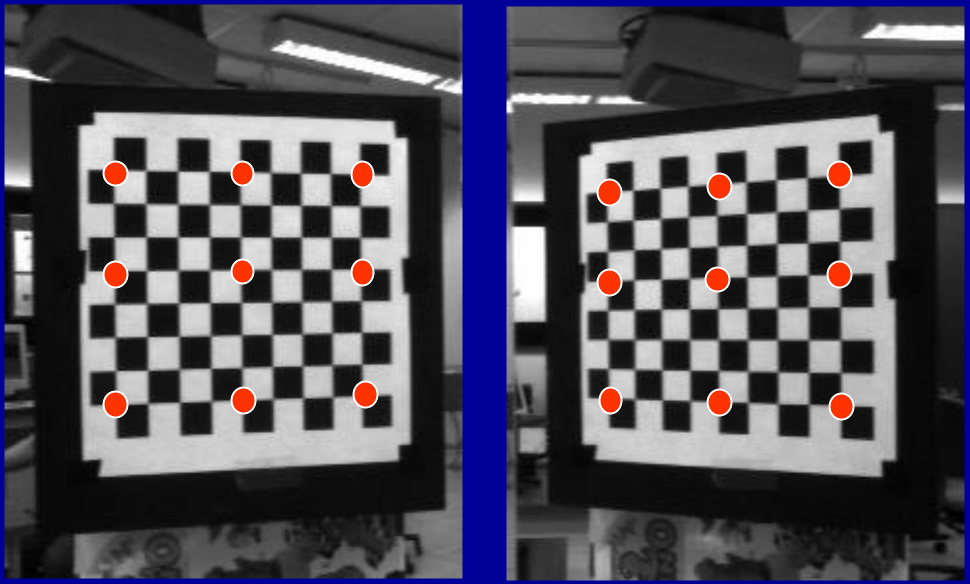
$(u_l^{(1)}, v_l^{(1)})$ $\quad$ $(u_r^{(1)}, v_r^{(1)})$

$\vdots$ $\quad$ $\vdots$

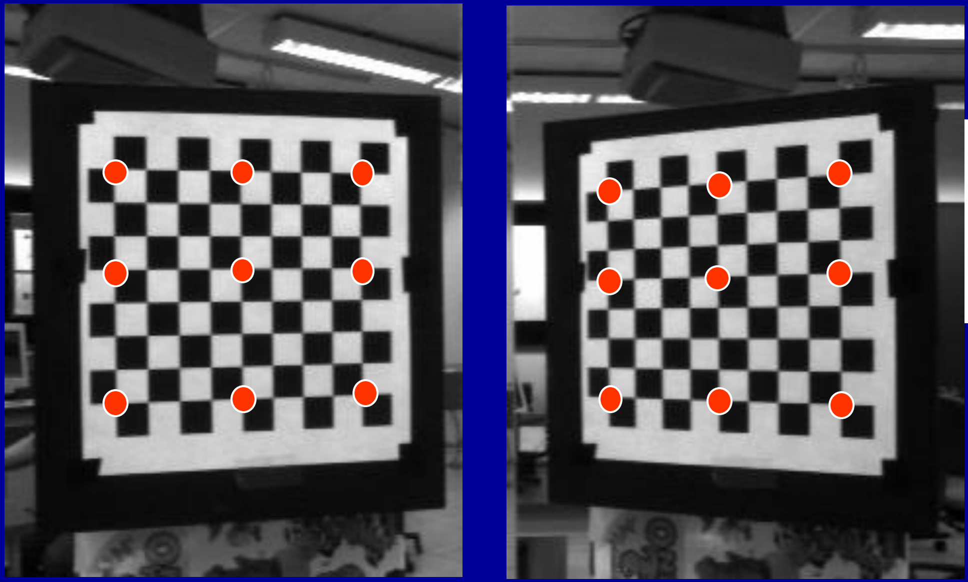
$(u_l^{(m)}, v_l^{(m)})$ $\quad$ $(u_r^{(m)}, v_r^{(m)})$

Stereo Calibration

Finding the epipolar geometry



Stereo Calibration
Finding the epipolar geometry



$$\overline{p}^{i\,T}_{r} F \, \overline{p}^{i}_{l} = 0$$

Problem: Find the fundamental matrix from a set of image correspondences

$$\left\{\left(\overline{p}^{i}_{l}, \overline{p}^{i}_{r}\right)\right\}, i = 1 \ldots N$$

$$\bar{p}_e^{\,i} = \begin{bmatrix} u_e^{(i)} \\ v_e^{(i)} \\ 1 \end{bmatrix} \qquad \bar{p}_r^{\,i} = \begin{bmatrix} u_r^{(i)} \\ v_r^{(i)} \\ 1 \end{bmatrix}$$

$$\bar{p}_r^{\,i\,T} \cdot F \cdot \bar{p}_e^{\,i} = \emptyset \Rightarrow$$

$$\begin{bmatrix} u_r^{(i)} & v_r^{(i)} & 1 \end{bmatrix} \cdot \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \cdot \begin{bmatrix} u_e^{(i)} \\ v_e^{(i)} \\ 1 \end{bmatrix} = \emptyset$$

$$\Rightarrow$$

$$\begin{bmatrix} u_r^{(i)} & v_r^{(i)} & 1 \end{bmatrix} \cdot \begin{bmatrix} f_{11} u_e^{(i)} + f_{12} v_e^{(i)} + f_{13} \\ f_{21} u_e^{(i)} + f_{22} v_e^{(i)} + f_{23} \\ f_{31} u_e^{(i)} + f_{32} v_e^{(i)} + f_{33} \end{bmatrix} = \emptyset$$

$$\Rightarrow$$

$$(u_r^{(i)} u_e^{(i)}) f_{11} + (u_r^{(i)} v_e^{(i)}) f_{12} + u_r^{(i)} f_{13} +$$

$$(v_r^{(i)} u_e^{(i)}) f_{21} + (v_r^{(i)} v_e^{(i)}) f_{22} + v_r^{(i)} f_{23} +$$

$$u_e^{(i)} f_{31} + v_e^{(i)} f_{32} + f_{33} = \emptyset$$

$$\bar{p}_e^{\,i} = \begin{bmatrix} u_e^{(i)} \\ v_e^{(i)} \\ 1 \end{bmatrix} \qquad \bar{p}_r^{\,i} = \begin{bmatrix} u_r^{(i)} \\ v_r^{(i)} \\ 1 \end{bmatrix}$$

Known

$$\bar{p}_r^{\,i\,T} \cdot F \cdot \bar{p}_e^{\,i} = \emptyset \Longrightarrow$$

Uknown

$$\begin{bmatrix} u_r^{(i)} & v_r^{(i)} & 1 \end{bmatrix} \cdot \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \cdot \begin{bmatrix} u_e^{(i)} \\ v_e^{(i)} \\ 1 \end{bmatrix} = \emptyset$$

$$\Longrightarrow$$

$$\begin{bmatrix} u_r^{(i)} & v_r^{(i)} & 1 \end{bmatrix} \cdot \begin{bmatrix} f_{11}u_e^{(i)} + f_{12}v_e^{(i)} + f_{13} \\ f_{21}u_e^{(i)} + f_{22}v_e^{(i)} + f_{23} \\ f_{31}u_e^{(i)} + f_{32}v_e^{(i)} + f_{33} \end{bmatrix} = \emptyset$$

$$\Longrightarrow$$

$$(u_r^{(i)}u_e^{(i)})f_{11} + (u_r^{(i)}v_e^{(i)})f_{12} + u_r^{(i)}f_{13} +$$
$$(v_r^{(i)}u_e^{(i)})f_{21} + (v_r^{(i)}v_e^{(i)})f_{22} + v_r^{(i)}f_{23} +$$
$$u_e^{(i)}f_{31} + v_e^{(i)}f_{32} + f_{33} = \emptyset$$

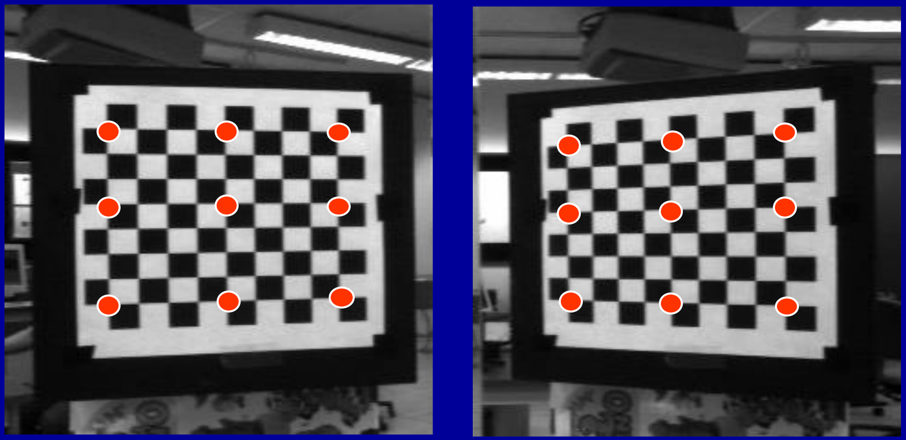
Stereo Calibration
Finding the epipolar geometry

$$
\begin{array}{c}
\\
\\
i \rightarrow \\
\\
\\
\end{array}
\begin{bmatrix}
u_r^{(1)} u_e^{(1)} & u_r^{(1)} v_e^{(1)} & u_r^{(1)} & v_r^{(1)} u_e^{(1)} & v_r^{(1)} v_e^{(1)} & v_r^{(1)} & u_e^{(1)} & v_e^{(1)} & 1 \\
 & & & \vdots & & & & & \\
u_r^{(i)} u_e^{(i)} & u_r^{(i)} v_e^{(i)} & u_r^{(i)} & v_r^{(i)} u_e^{(i)} & v_r^{(i)} v_e^{(i)} & v_r^{(i)} & u_e^{(i)} & v_e^{(i)} & 1 \\
 & & & \vdots & & & & & \\
u_r^{(n)} u_e^{(n)} & u_r^{(n)} v_e^{(n)} & u_r^{(n)} & v_r^{(n)} u_e^{(n)} & v_r^{(n)} v_e^{(n)} & v_r^{(n)} & u_e^{(n)} & v_e^{(n)} & 1
\end{bmatrix}_{n\times 9}
\begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix}_{9\times 1}
=
\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1}
$$

Homogeneous system: $\longrightarrow$ $\boxed{A_{n\times 9} \cdot v_{9\times 1} = \mathbf{0}_{n\times 1}}$

n equations, 9 unknowns

The 8-point algorithm



n>=8 correspondences

$$\overline{p}^{i}{}_{r}{}^{T} F \, \overline{p}^{i}{}_{l} = 0 \quad \Rightarrow \quad A\mathbf{v} = \mathbf{0}$$

v: the 9 elements of F (**uknown**)

A: n x 9 measurement matrix (**known**).

Solve using **SVD** (solution up to **a scale factor**).

Enforce rank(F)=2 (SVD on the computed F).
Be careful: numerical instabilities.

The 8-point algorithm

n>=8 correspondences

$$\overline{p}'^{i}_{r}{}^{T} F \, \overline{p}'^{i}_{l} = 0 \quad \Rightarrow \quad A\mathbf{v} = \mathbf{0}$$

1) SVD: $A = UDV^T$ => v is column V: smaller singular value

2) Construct F using values of v

3) Apply SVD to F (again!): $F = U'D'V'^T$

4) Enforce rank(F) = 2 by setting D' smaller sing. value to 0.

5) So $F' = U'D''V'^T$ is the estimate of the fundamental matrix (D'' is D' with smaller element set to 0).

Find R, T from Fundamental Matrix

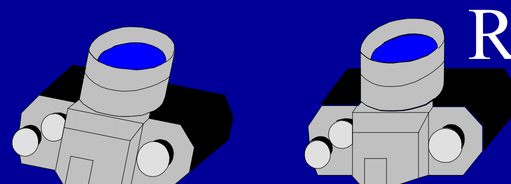
- If cameras are calibrated (internals known)

$$F = M_r^{-T} E M_l^{-1}$$ → $E = M_r^T F M_l$

- But, $E = R\ S_{[T]}$ → Solve for R, T (SVD)

Epipolar Lines – Example

Left Image



Right Image

Epipolar Lines – Example

Left Epipolar Lines

Right E. L.

Example

Finding Epipolar Lines: Example

Given the Fundamental matrix,

$$F = \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix}$$

Left Image

Right Image

Finding Epipolar Lines: Example

Given the Fundamental matrix,

$$F = \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix}$$

and the left image point

$$\tilde{\boldsymbol{u}}_l = \begin{bmatrix} 343 \\ 221 \\ 1 \end{bmatrix}$$

Left Image

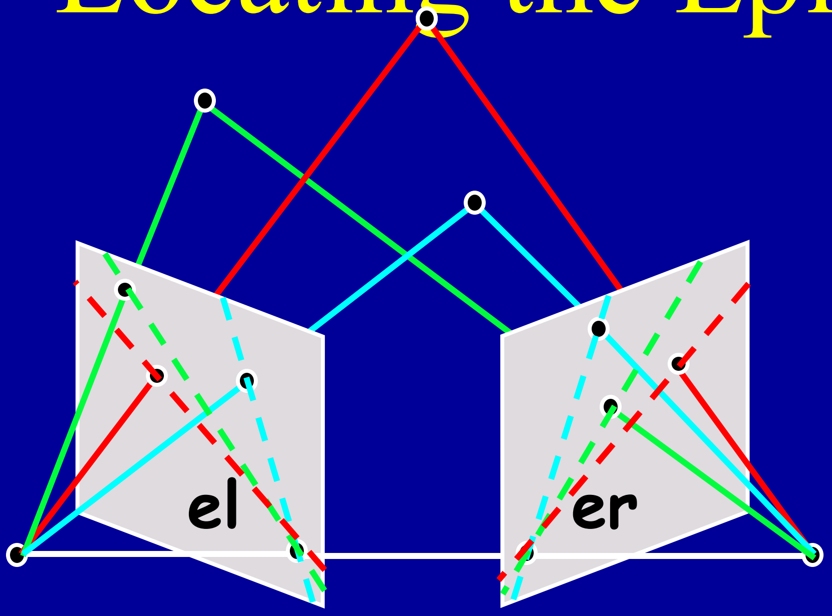
Right Image

Epipolar Line

The equation for the epipolar line in the right image is

$$.03u_r + .99v_r - 265 = 0$$

Locating the Epipoles from E & F



Accurate epipole localization:

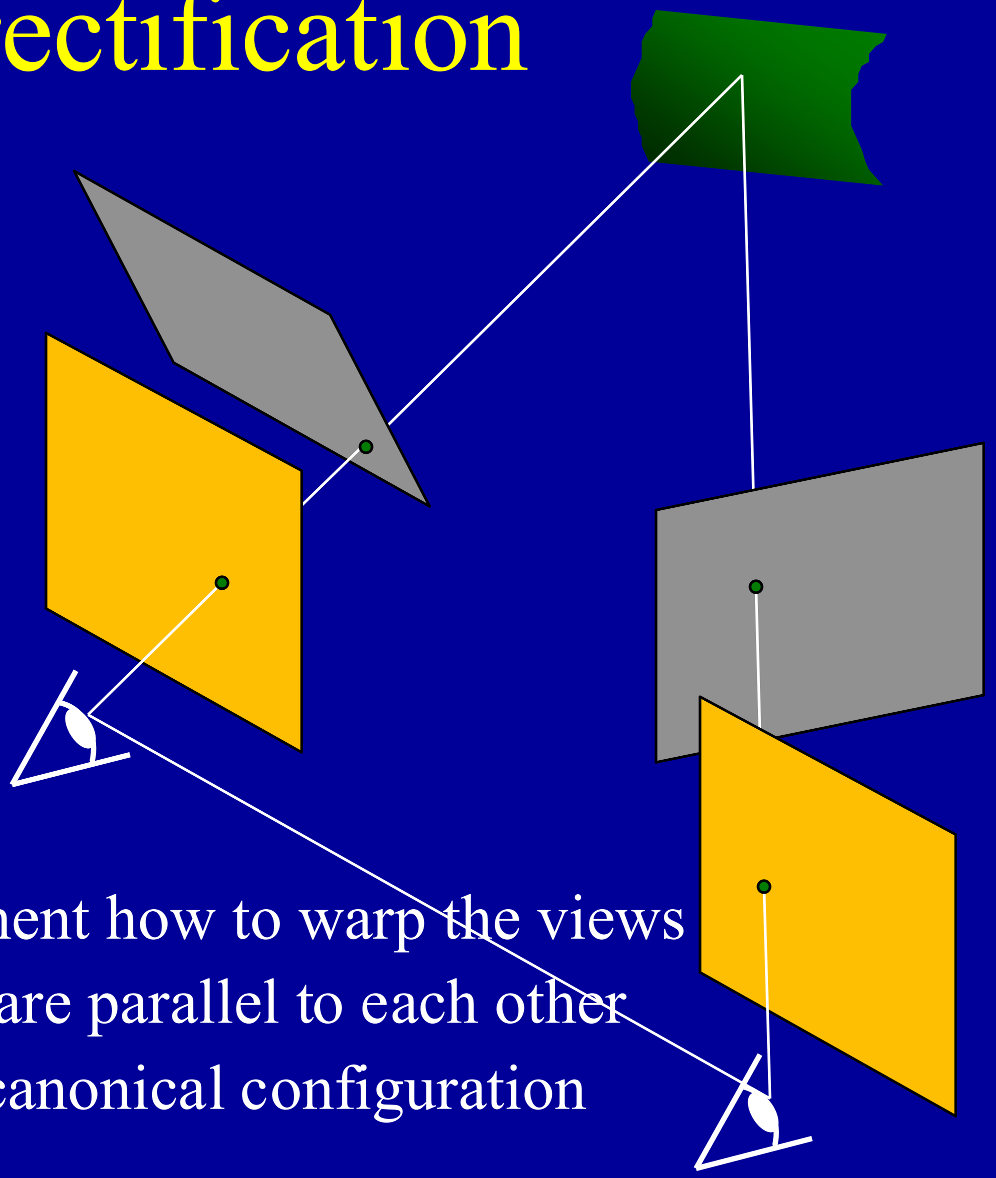
1) Refining epipolar lines.
2) Checking for consistency.
3) Uncalibrated stereo.

F => el, er in pixel coordinates.
E => el, er in camera coordinates.

Fact: All epipolar lines pass through epipoles.
So

$$\bar{p}_r^{\,T} F \bar{e}_l = 0, \forall \bar{p}_r$$

Image rectification



- Given general displacement how to warp the views
- Such that epipolar lines are parallel to each other
- How to warp it back to canonical configuration

(Seitz)

Epipolar rectification

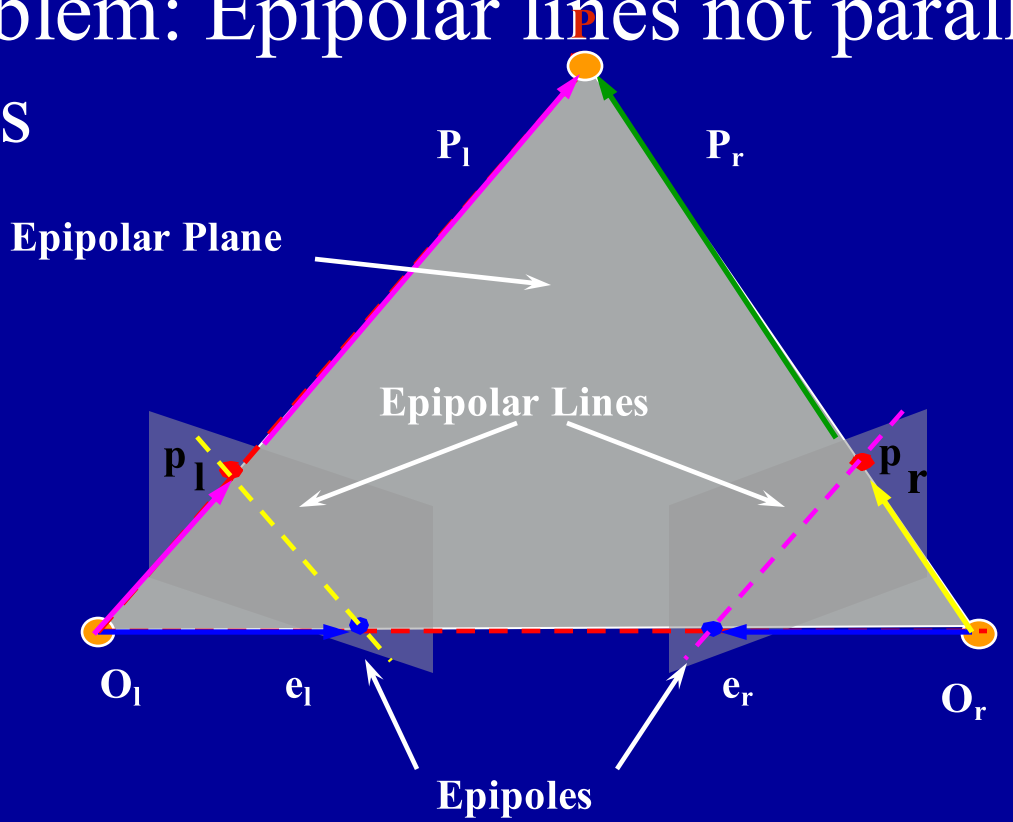
- Rectified Image Pair
- Corresponding epipolar lines are aligned with the scan-lines
- Search for dense correspondence is a 1D search

Epipolar rectification

Rectified Image Pair

Rectification

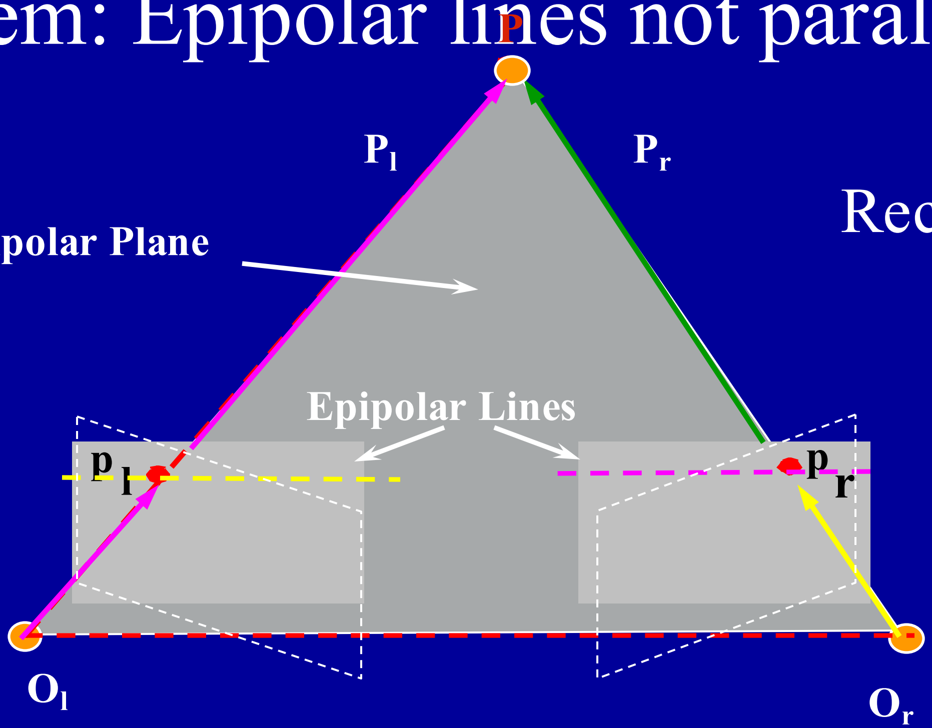
- Problem: Epipolar lines not parallel to scan lines



From Sebastian Thrun/Jana Kosecka

Rectification

- Problem: Epipolar lines not parallel to scan lines



Rectified Images

Epipoles at infinity

From Sebastian Thrun/Jana Kosecka

Image Rectification (cont.)

- Perform by rotating the cameras
- *Not* equivalent to rotating the images
- The lines through the centers become parallel to each other, and the epipoles move to infinity

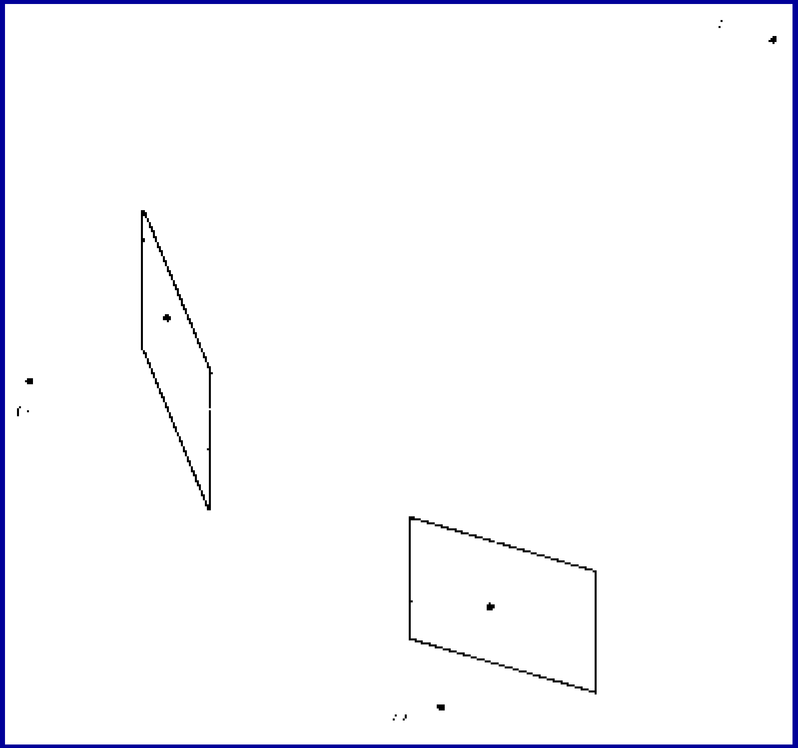
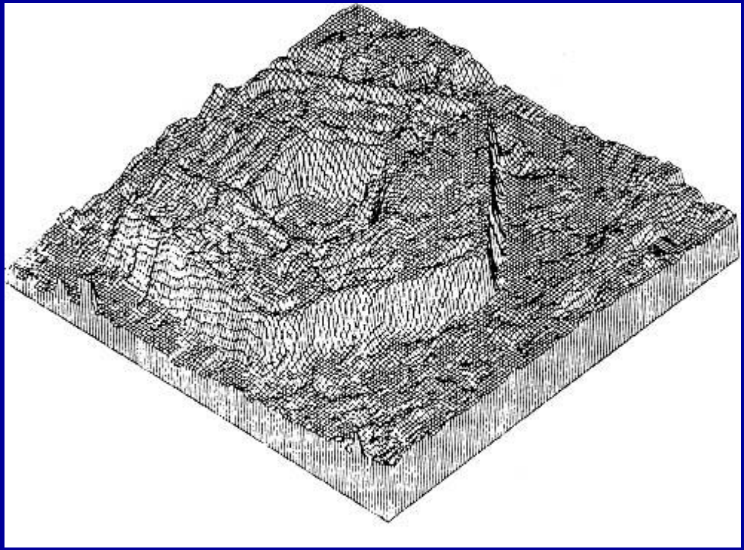


Image Rectification (cont.)

- Given extrinsic parameters T and R (relative position and orientation of the two cameras)
  - Rotate the left camera about the projection center so that the the epipolar lines become parallel to the horizontal axis
  - Apply the same rotation to the right camera
  - Rotate the right camera by R
  - Adjust the scale in both camera reference frames

3-D Reconstruction



3-D Reconstruction

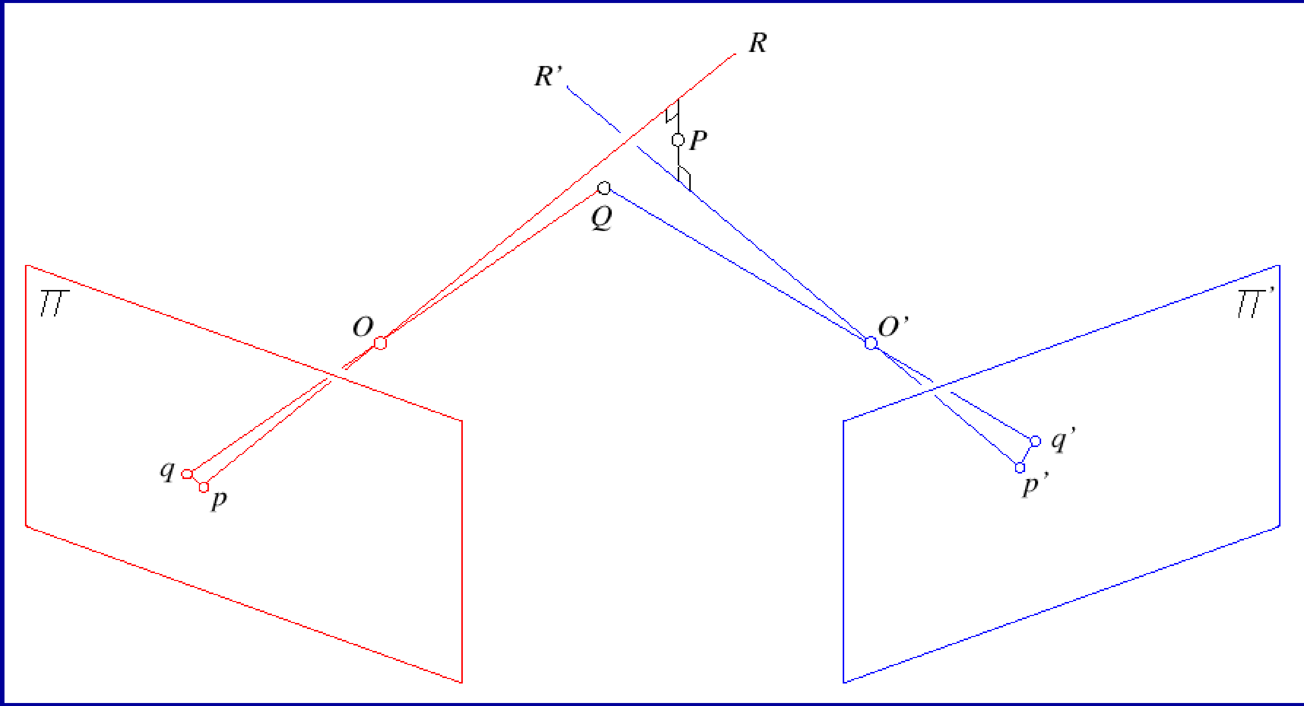
A Priori Knowledge	3-D Reconstruction from two views
Intrinsic and extrinsic	Unambiguous (triangulation)
Intrinsic only	Up to unknown scaling factor
No information	Up to unknown projective transformation

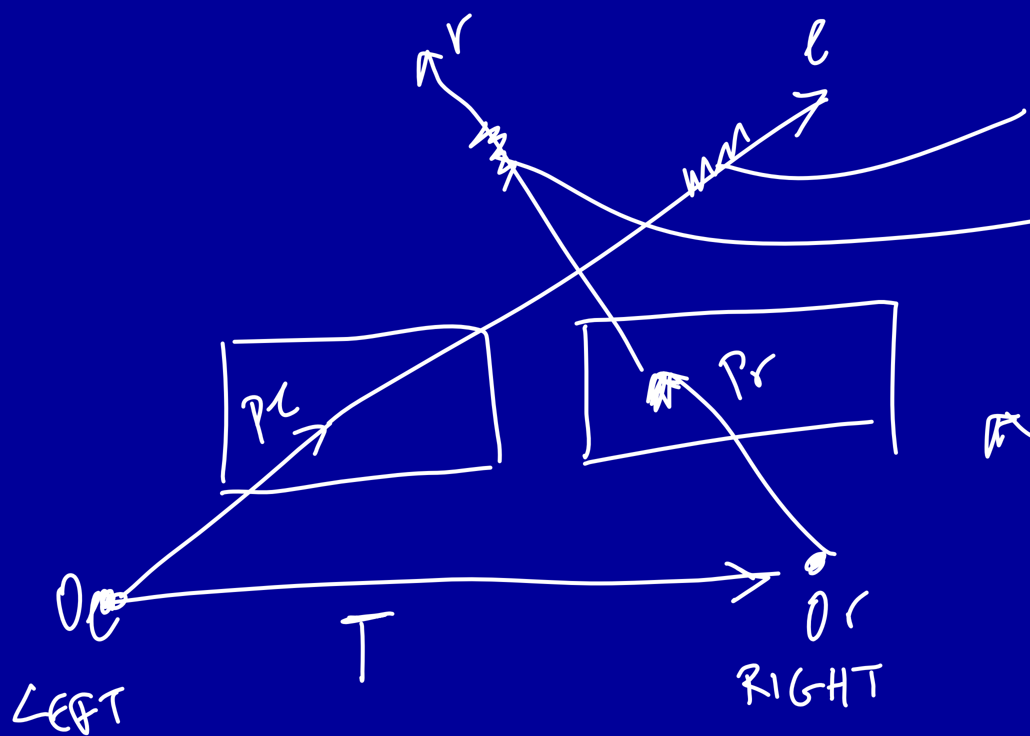
Reconstruction

- Given pair of image points p and p', and focal points O and O', find preimage P
- In theory: find P by intersecting the rays R=Op and R'=Op'
- In practice: R and R' won't actually intersect due to calibration and feature localization errors

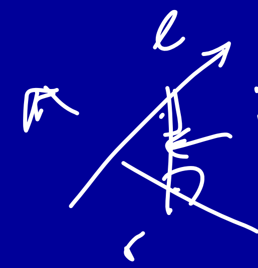
Reconstruction Approaches

- Geometric
  - Construct the line segment perpendicular to R and R' that intersects both rays and take its mid-point





$$\left.\begin{array}{l}\alpha \cdot p_l \\ \beta \cdot (R^T p_r) + T\end{array}\right\}$$ wrt left coord. system



is parallel to $p_l \times R^T p_r = w$

So w is $\perp$ to both lines l & r

$$\therefore\ \alpha \cdot p_l + c\,(p_l \times R^T p_r) = T + b\,(R^T p_r) \Longrightarrow$$

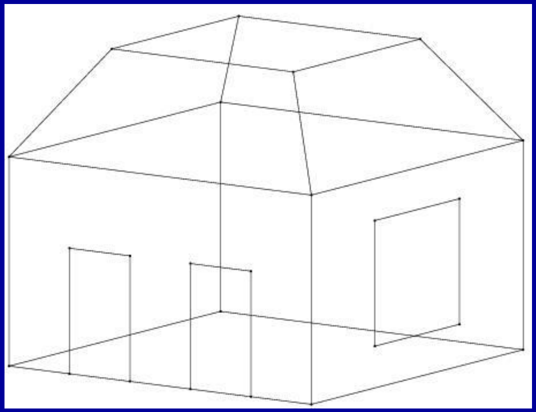
$$\Longrightarrow \alpha\, p_l - b\,(R^T p_r) + c\,(p_l \times R^T p_r) = T \Longrightarrow$$

3×3 system: compute α, b & c!

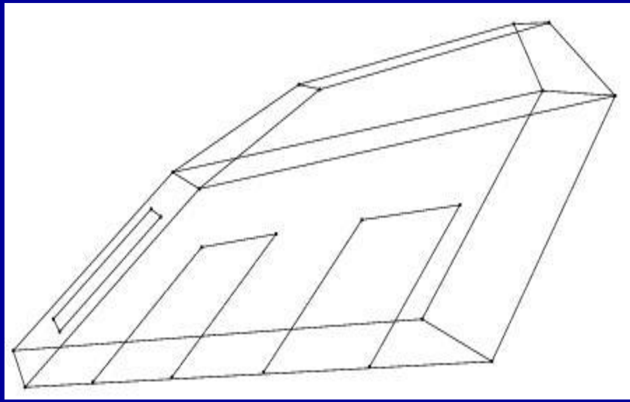
Check this out!

http://www.well.com/user/jimg/stereo/stereo_list.html

Projective Reconstruction



Euclidean reconstruction



Projective reconstruction

From Sebastian Thrun/Jana Kosecka

Euclidean vs Projective reconstruction

- Euclidean reconstruction – true metric properties of objects lengths (distances), angles, parallelism are preserved
- Unchanged under rigid body transformations
- => Euclidean Geometry – properties of rigid bodies under rigid body transformations, similarity transformation

- Projective reconstruction – lengths, angles, parallelism are NOT preserved – we get distorted images of objects – their distorted 3D counterparts --> 3D projective reconstruction
- => Projective Geometry

How can We Improve Stereo?

Space-time stereo scanner
uses unstructured light to aid
in correspondence

Result: Dense 3D mesh (noisy)

From Sebastian Thrun/Jana Kosecka

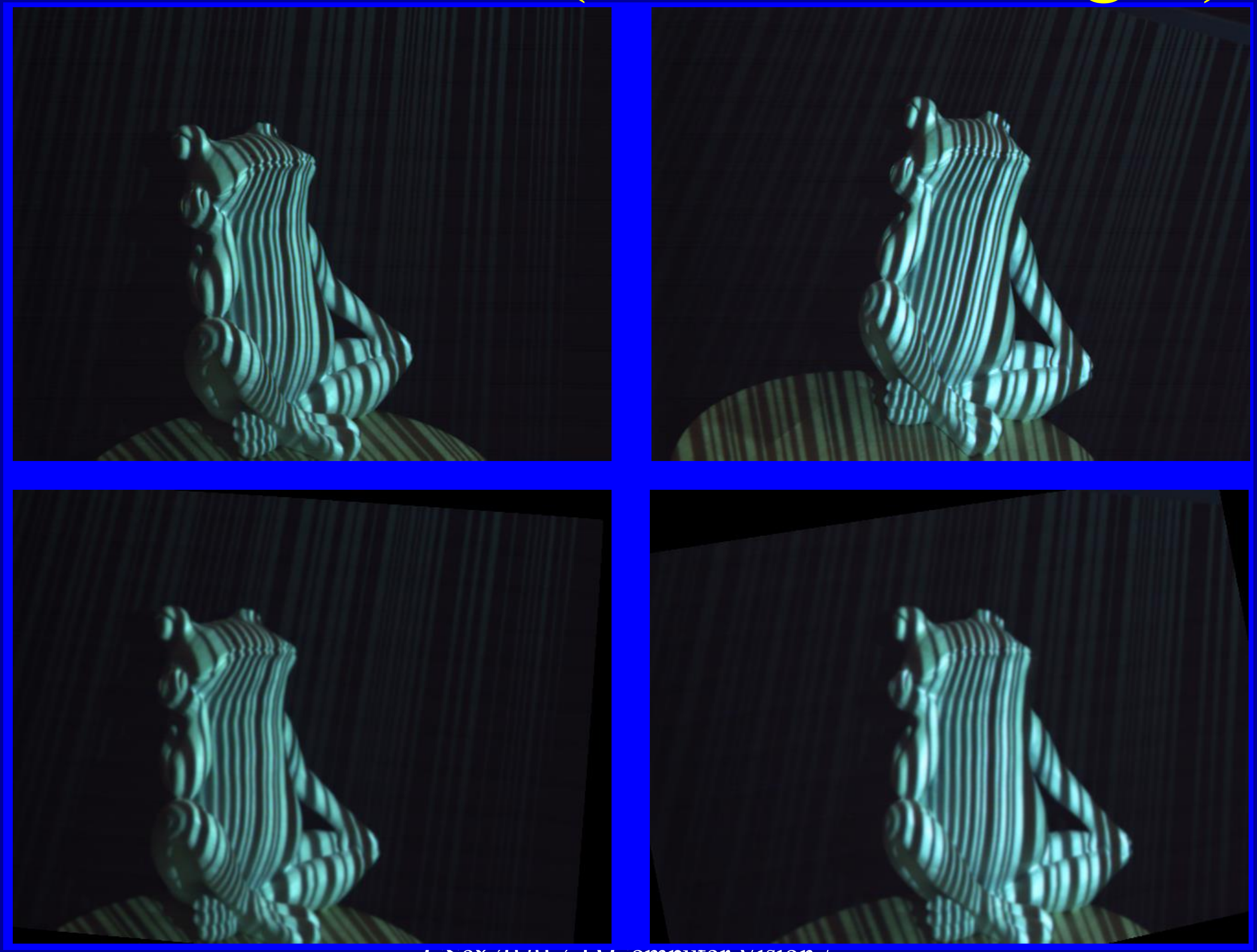
Active Stereo: Adding Texture to Scene

By James Davis,
Honda Research,
Now UCSC

From Sebastian Thrun/Jana Kosecka

Active Stereo (Structured Light)

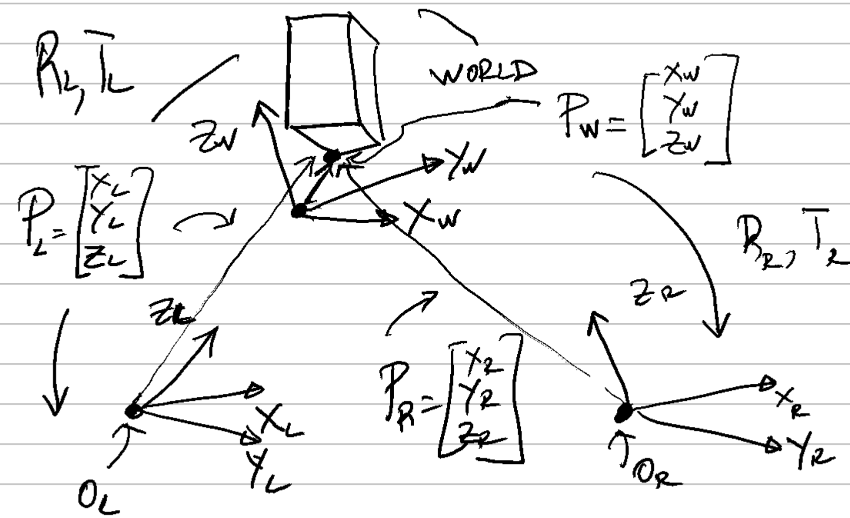
rectified



From Sebastian Thrun/Jana Kosecka

Older slides

Stereo Calibration Using Calibration object



FROM WORLD TO LEFT R_L, T_L:

$$P_L = R_L \cdot P_W + T_L \quad (1)$$

FROM WORLD TO RIGHT R_L, T_R:

$$P_R = R_R P_W + T_L \quad (2)$$

KNOWN FROM CALIBRATION

FROM RIGHT TO LEFT $R_{R\to L}, T_{R\to L}$:

$$\boxed{P_L = R_{R\to L} \cdot P_R + T_{R\to L}}$$

From (2): $R_R P_W = P_R - T_R \Rightarrow P_W = R_R^T (P_R - T_R)$ (3)

From (1) & (3): $P_L = R_L \cdot \left(R_R^T (P_R - T_R)\right) + T_L \Rightarrow$

$\Rightarrow$

Stereo Calibration Using Calibration object

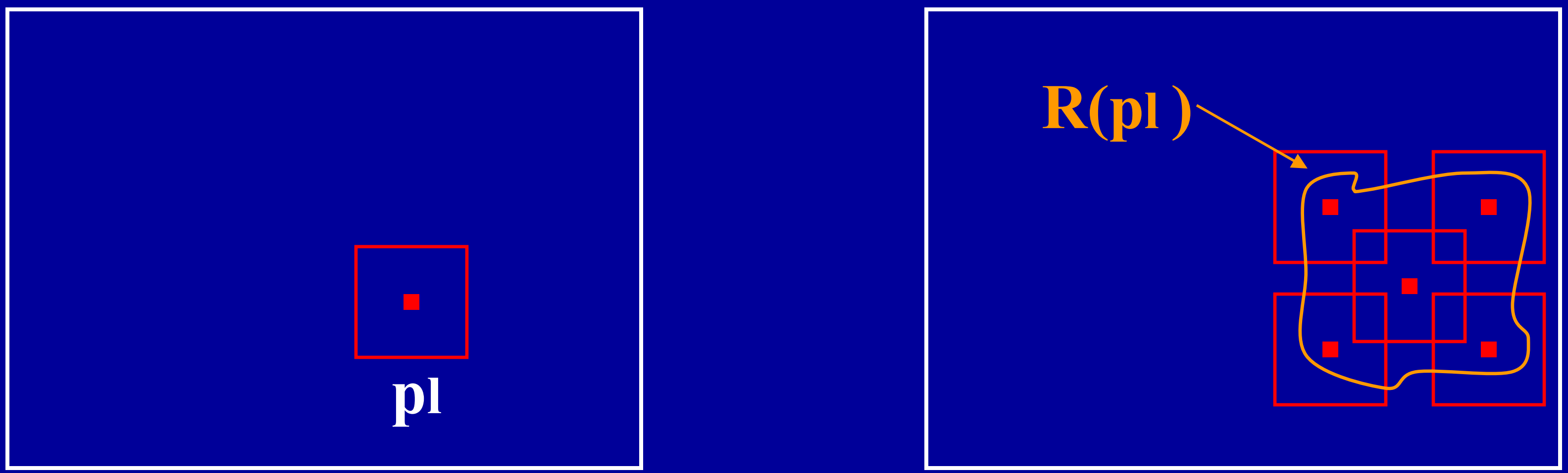
$$P_L = \underbrace{R_L \cdot R_R^T}_{R_{R\to L}} \cdot P_R \underbrace{- R_L \cdot R_R^T T_R + T_L}_{T_{R\to L}} \Longrightarrow$$

$$P_L = R_{R\to L} \cdot P_R + T_{R\to L} \quad (4)$$

$$\boxed{\begin{aligned} R_{R\to L} &= R_L \cdot R_R^T \\ T_{R\to L} &= T_L - R_{R\to L} \cdot T_R \end{aligned}}$$

Special case: if $R_{R\to L} = I$ then $T_{R\to L} = T_L - T_R$.

Correlation-Based Methods



Left Image Right Image

For each pixel pl=[i,j] in the left image

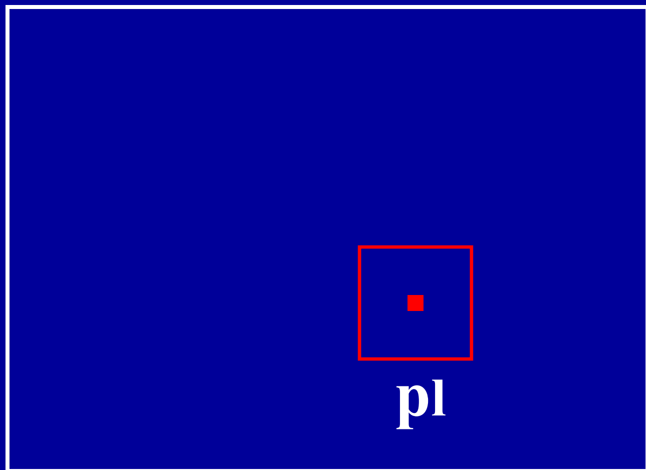
For each displacement **d**=[d1,d2] in R(pl)

Compute $$c(\mathbf{d}) = \sum_{k=-W}^{W} \sum_{l=-W}^{W} \psi(I_l(i+k, j+l), Ir(i+k-d_1, j+l-d_2))$$

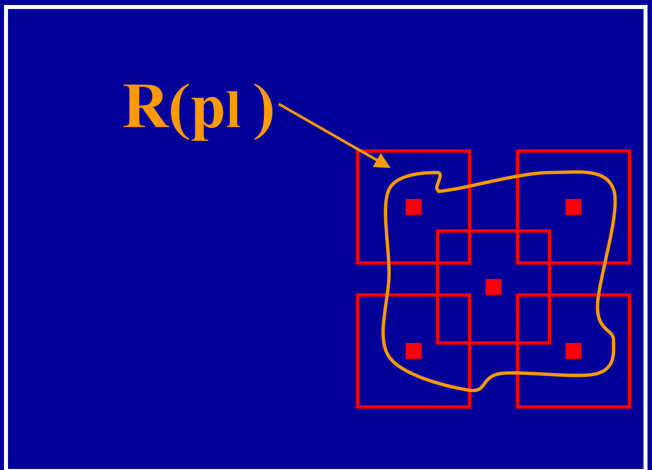
The disparity of pl is the **d** that maximizes c(**d**)

HAVE TO SPECIFY: Region R, size W, and correlation function ψ.

Correlation-Based Methods



Left Image



Right Image

$$\psi(u,v) = uv$$ ← CROSS-CORRELATION

$$\psi(u,v) = -(u-v)^2$$ ← SUM OF SQUARED DIFFERENCES **SSD**

SSD is usually preferred: handles different intensity scales.

Normalized cross-correlation is better (but is more expensive).