# 1. Stereo Vision

- **Problem**
  - Infer 3D structure of a scene from two or more images taken from different viewpoints

- **Two primary Sub-problems**
  - Correspondence problem (stereo match) -> disparity map
    - Similarity instead of identity
    - Occlusion problem: some parts of the scene are visible only in one eye
  - Reconstruction problem -> 3D
    - What we need to know about the cameras' parameters
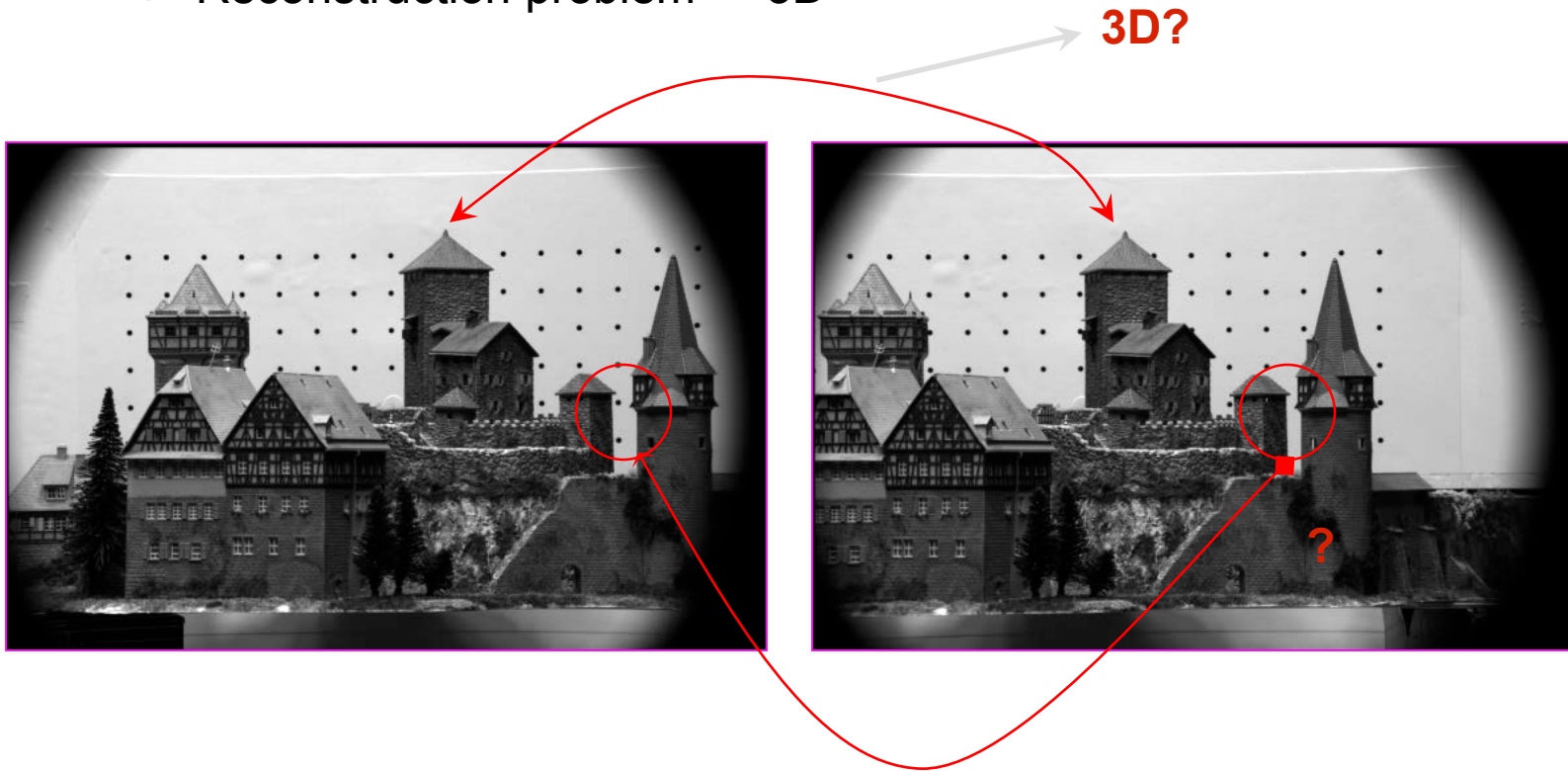    - Often a stereo calibration problem

- **Lectures on Stereo Vision**
  - Stereo Geometry – Epipolar Geometry (*)
  - Correspondence Problem (*) – Two classes of approaches
  - 3D Reconstruction Problems – Three approaches

■ **Problems**
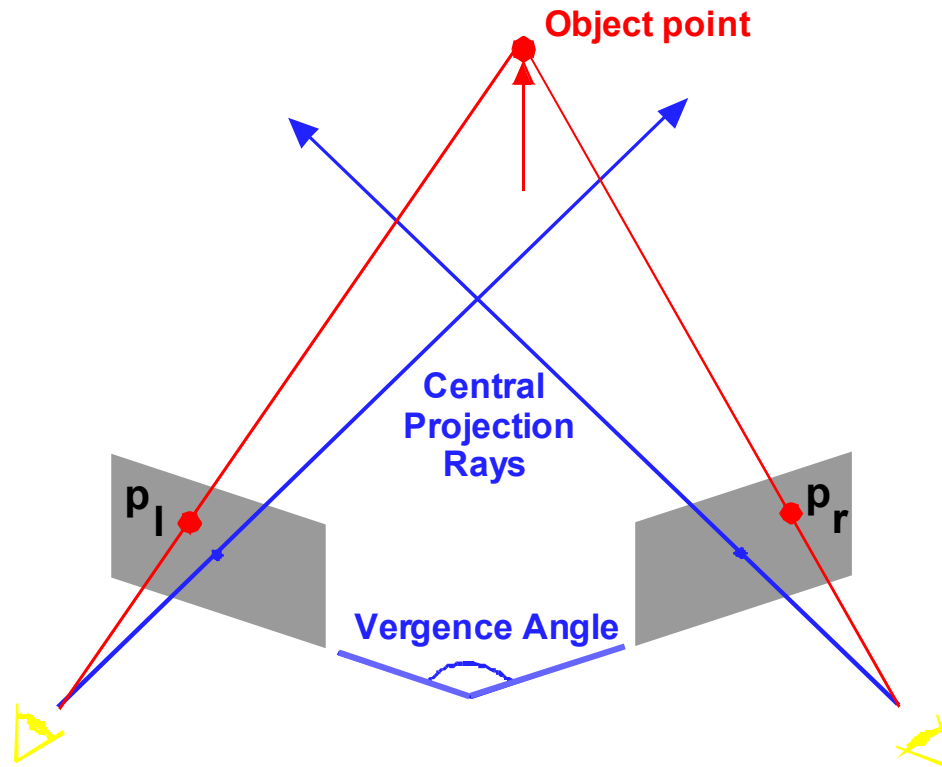- Correspondence problem (stereo match) -> disparity map
- Reconstruction problem -> 3D

**3D?**



**?**

CMU CIL Stereo Dataset : Castle sequence
http://www-2.cs.cmu.edu/afs/cs/project/cil/ftp/html/cil-ster.html

# Stereo Geometry

**Object point**
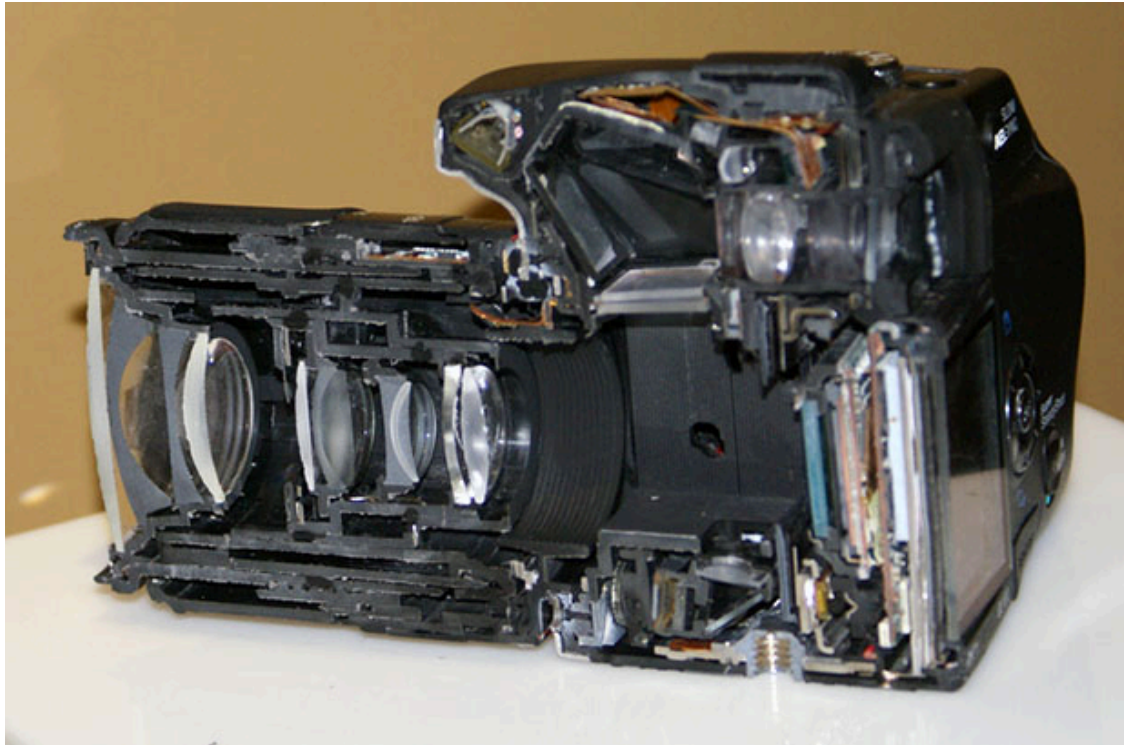
**Central Projection Rays**
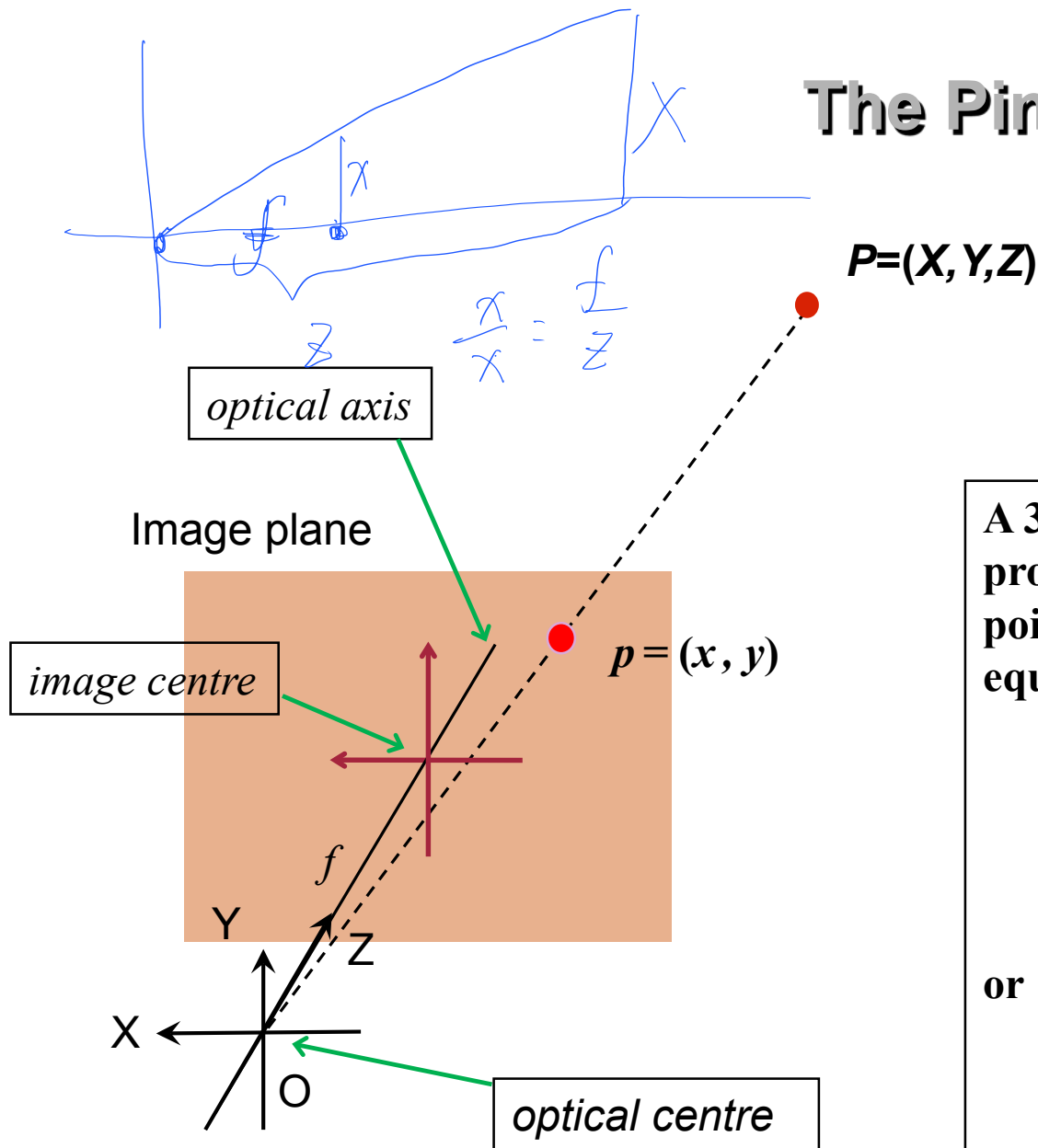
$p_l$

$p_r$

**Vergence Angle**

- Converging Axes – Usual setup of human eyes
- Depth obtained by triangulation
- Correspondence problem: $p_l$ and $p_r$ correspond to the left and right projections of P, respectively.

# The Pinhole Camera

$P=(X,Y,Z)$

optical axis

Image plane

image centre

$p = (x, y)$

$f$

Y

Z

X

O

optical centre

A 3D point *P* is projected on the image point *p*. The projection equation is
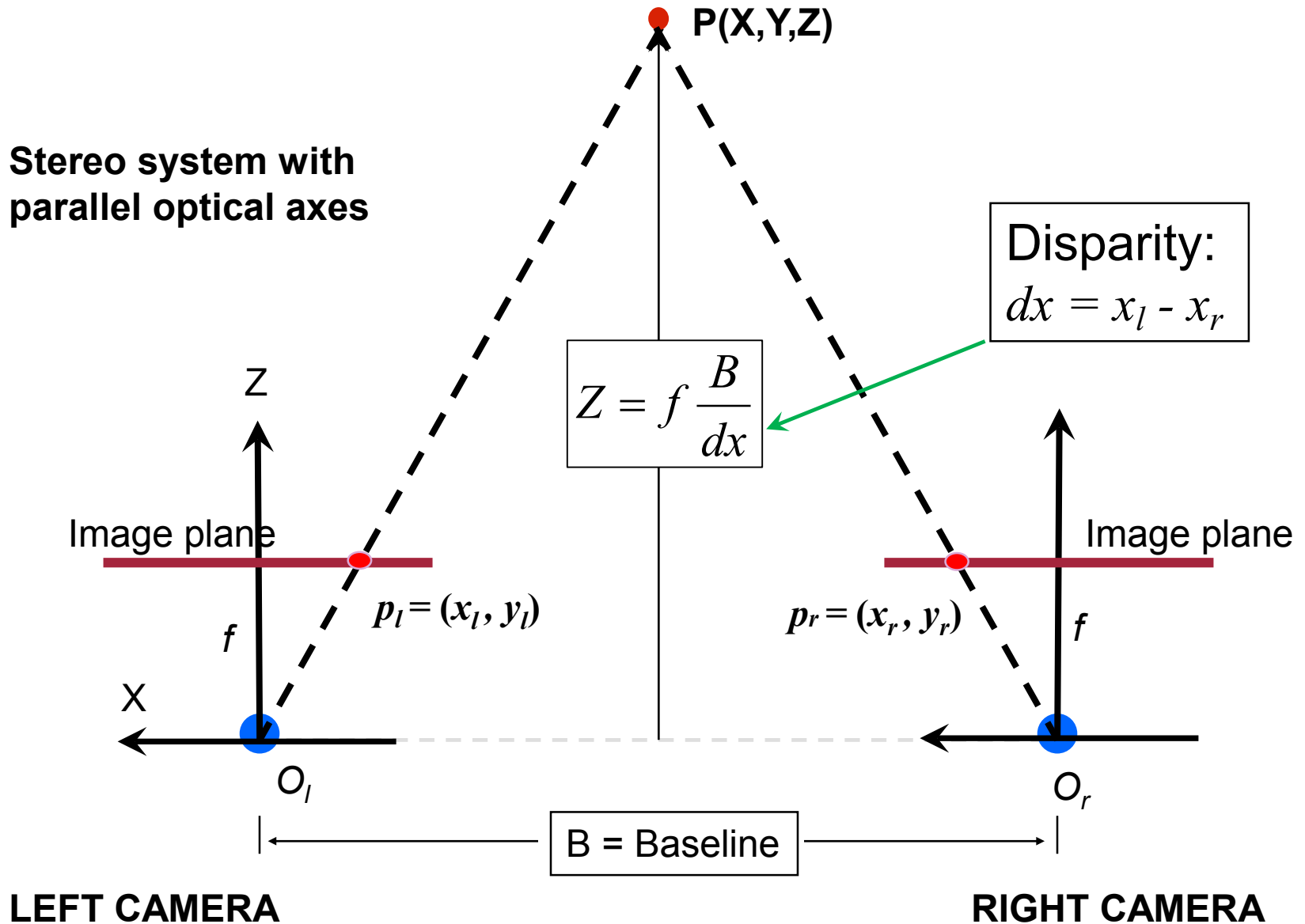
$$x = \frac{f}{Z}X$$

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

or

$$\begin{pmatrix} x \\ y \\ f \end{pmatrix} = \frac{f}{Z}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

# Disparity Equation

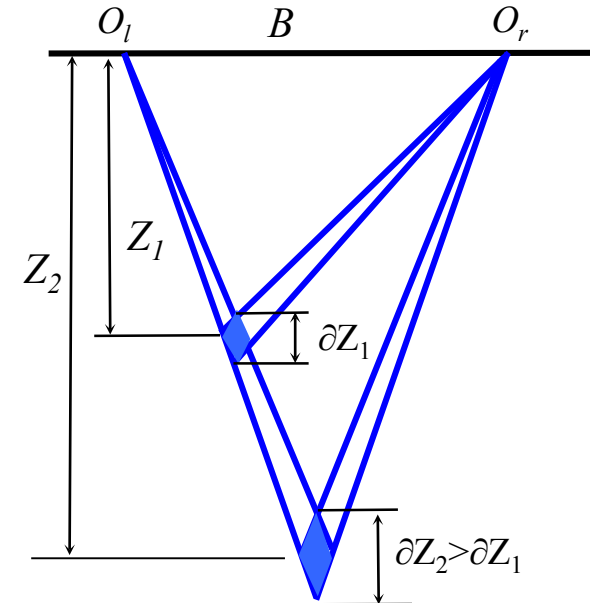**Stereo system with parallel optical axes**

P(X,Y,Z)

Disparity:
$$dx = x_l - x_r$$

$$Z = f\,\frac{B}{dx}$$

Z

Image plane

$p_l = (x_l, y_l)$

$p_r = (x_r, y_r)$

Image plane

$f$

$f$

X

$O_l$

$O_r$

B = Baseline

**LEFT CAMERA**

**RIGHT CAMERA**

- Given the same image localization error
  - Angle of cones in the figure
- Depth Accuracy (Depth Resolution) vs. Baseline
  - Depth Error $\propto$ 1/B (Baseline length)
  - PROS of Longer baseline,
    - better depth estimation
  - CONS
    - smaller common FOV
    - Correspondence harder due to occlusion
- Depth Accuracy (Depth Resolution) vs. Depth
  - Disparity (>0) $\propto$ 1/ Depth
  - Depth Error $\propto$ Depth$^2$
  - Nearer the point, better the depth estimation

$$Z = \frac{fB}{dx} \qquad \frac{\partial z}{\partial(dx)} = -\frac{fB}{(dx)^2} = -\frac{Z^2}{fB}$$

$O_l$    $B$    $O_r$

$Z_2$   $Z_1$

$\partial Z_1$

$\partial Z_2 > \partial Z_1$

$B \uparrow \text{error} \downarrow$

**Absolute error**

$$\partial Z = \frac{Z^2}{fB} \partial(dx)$$

**Relative error**

$$\frac{\partial Z}{Z} = \frac{Z}{fB} \partial(dx)$$

- For the ST Electronics project, we have obtained the following parameters.
- $Z$ = 500m
- $B$ = 3m
- $f$ = 2700 pixel (obtained from calibration)
- We set $\partial(dx) = 1$ pixel
- The absolute error

$$\partial Z = \frac{Z^2}{fB} \partial(dx) = 30m$$

- If we can obtain sub-pixel accuracy, for example, $\partial(dx) = 0.25$ pixel, then, the error is reduced to

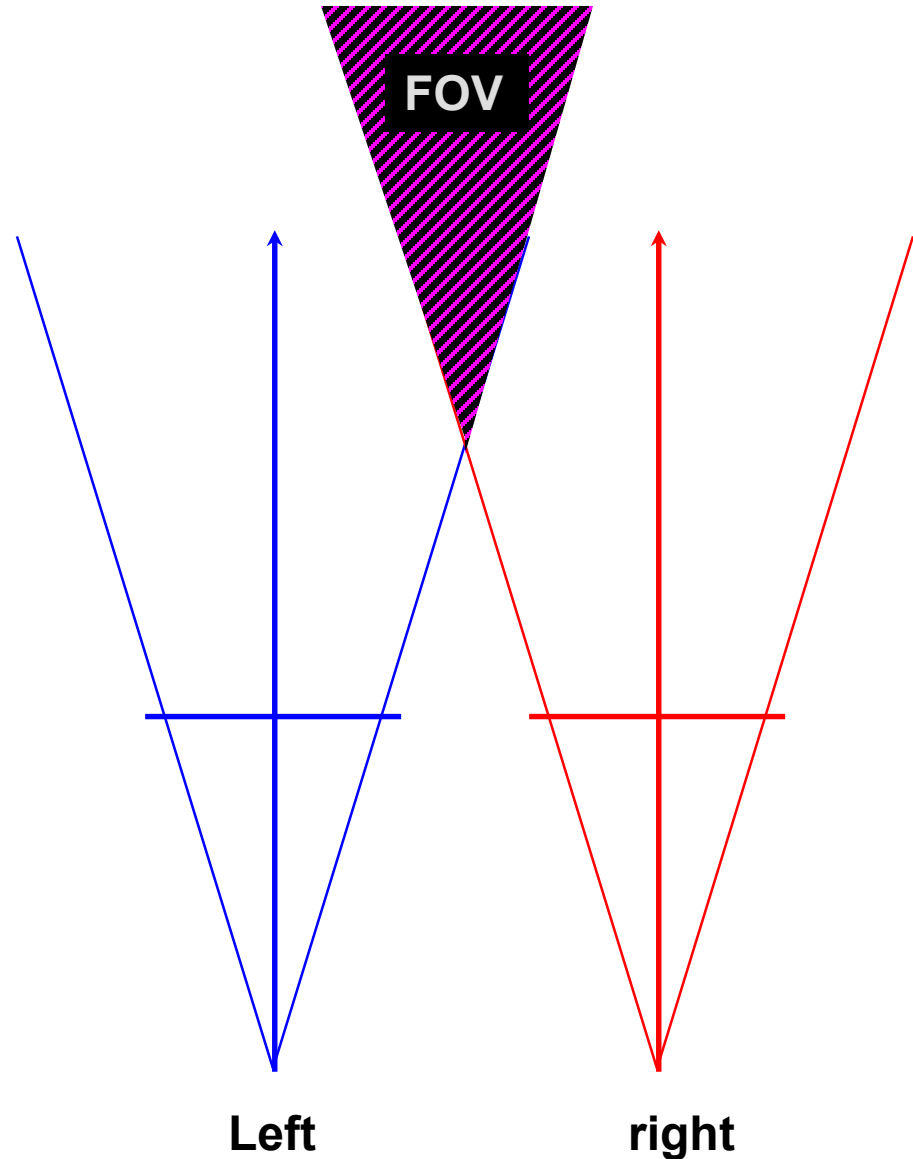$$\partial Z = \frac{Z^2}{fB} \partial(dx) = 7.5m$$

# Stereo with Parallel Cameras

- **Stereo with Parallel Axes**
  - Short baseline
    - large common FOV (Field of View)
    - large depth error
  - Long baseline
    - small depth error
    - small common FOV
    - More occlusion problems

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

**FOV**

**Left    right**
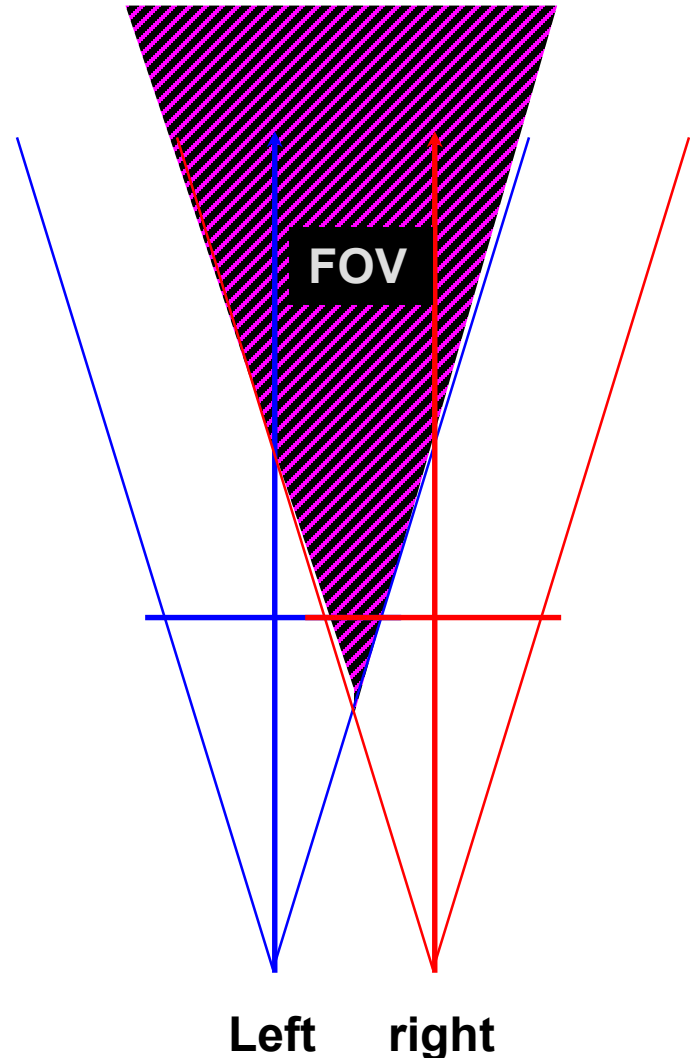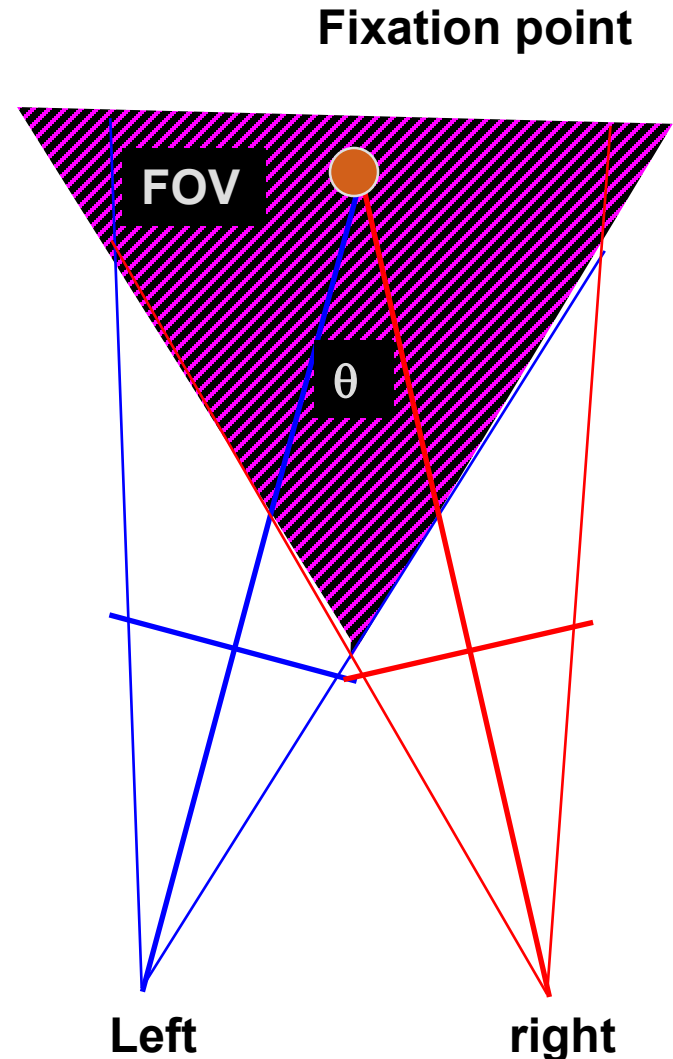
# Stereo with Parallel Cameras

- Stereo with Parallel Axes
  - Short baseline
    - large common FOV
    - large depth error
  - Long baseline
    - small depth error
    - small common FOV
    - More occlusion problems

Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases
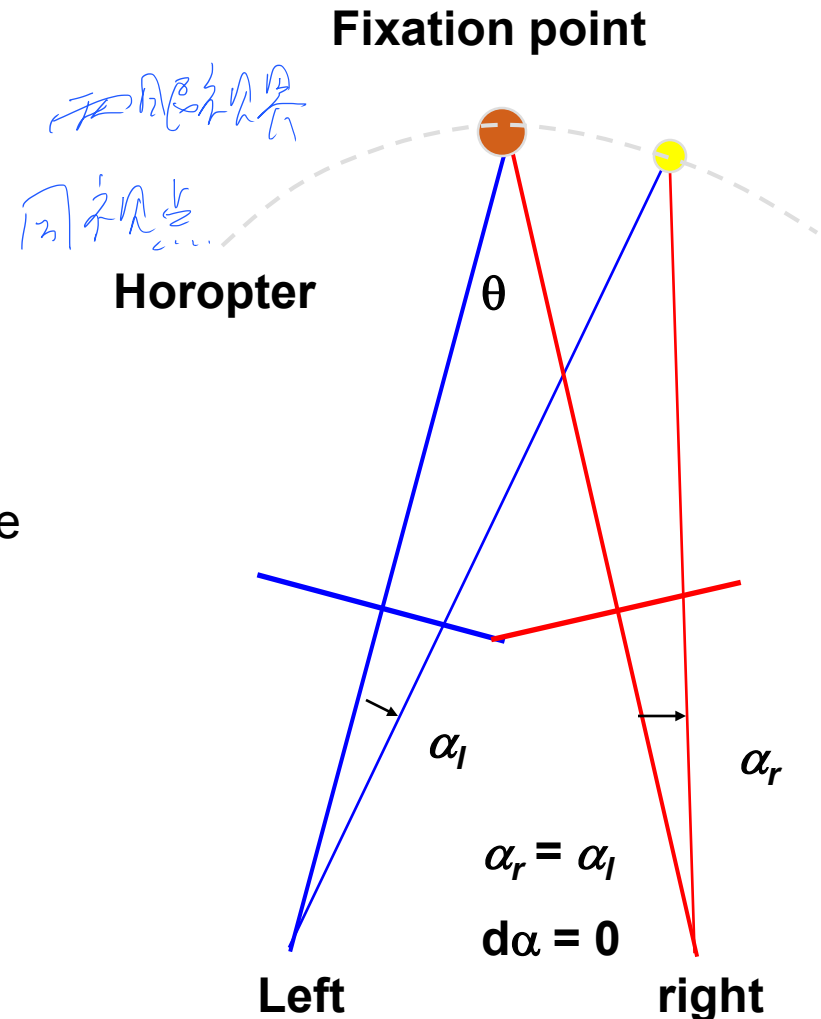
**FOV**

**Left**          **right**

# Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - >0 : outside of the horopter
    - <0 : inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
  - Nearer the point, better the depth estimation

**Fixation point**

FOV

$\theta$

**Left**          **right**

# Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - >0 : outside of the horopter
    - <0 : inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
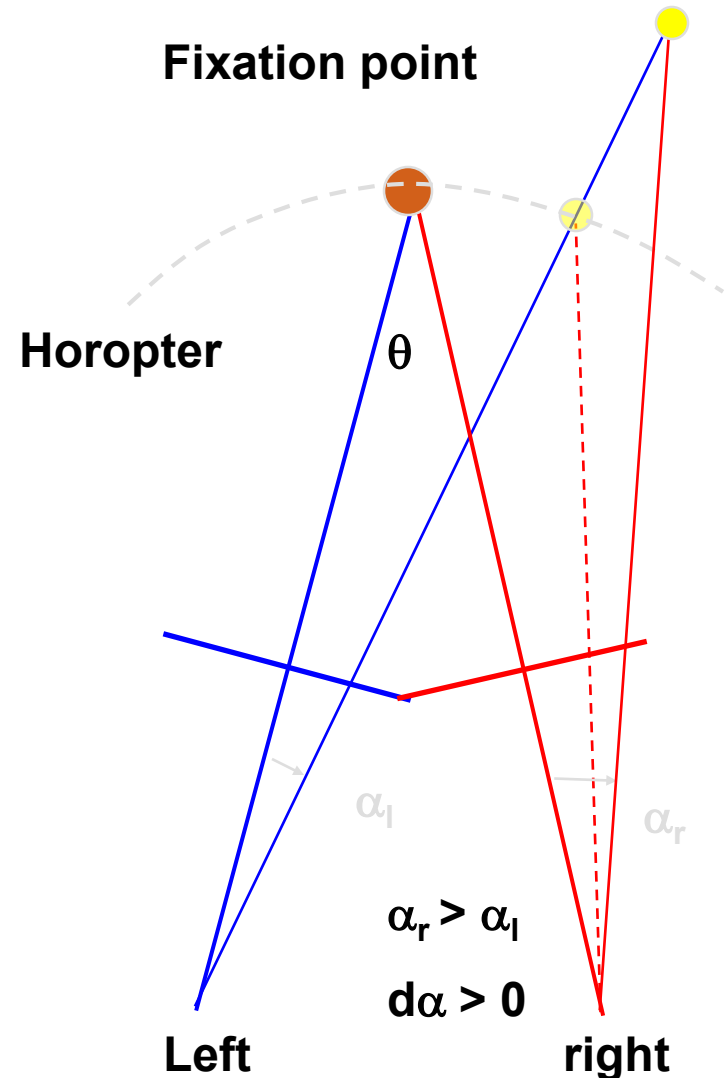  - Nearer the point, better the depth estimation

**Fixation point**

**Horopter**

$\theta$

$\alpha_l$

$\alpha_r$

$\alpha_r = \alpha_l$

$d\alpha = 0$

**Left**

**right**

# Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - >0 : outside of the horopter
    - <0 : inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
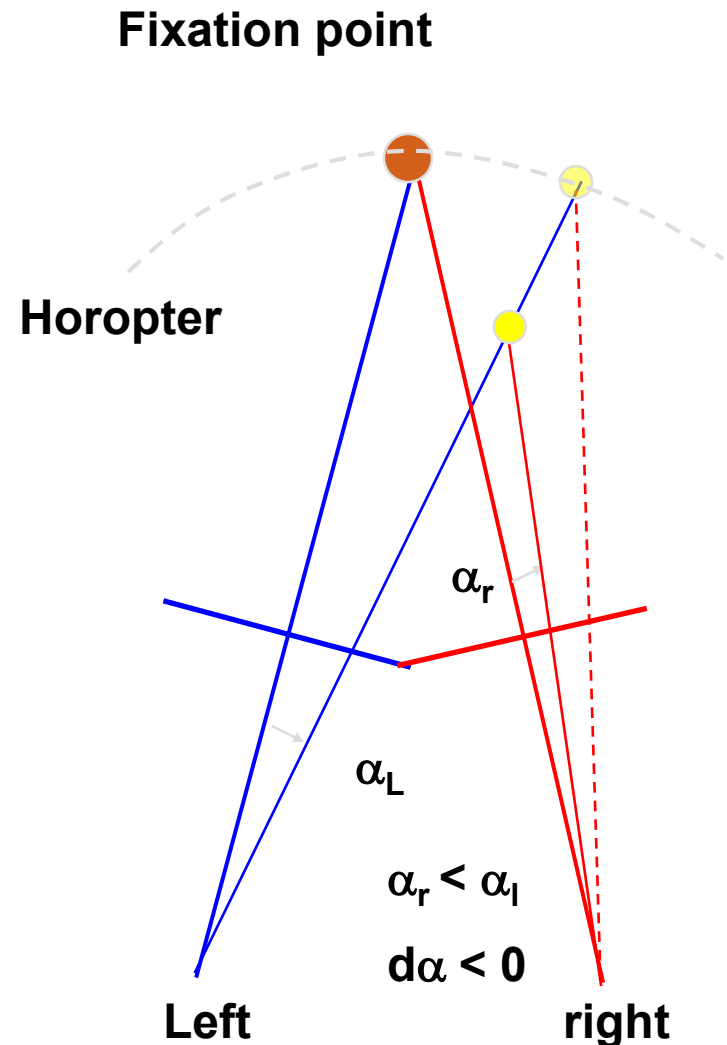  - Nearer the point, better the depth estimation

**Fixation point**

**Horopter**

$\theta$

$\alpha_l$

$\alpha_r$

$\alpha_r > \alpha_l$

$d\alpha > 0$

**Left**

**right**

# Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - >0 : outside of the horopter
    - <0 : inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
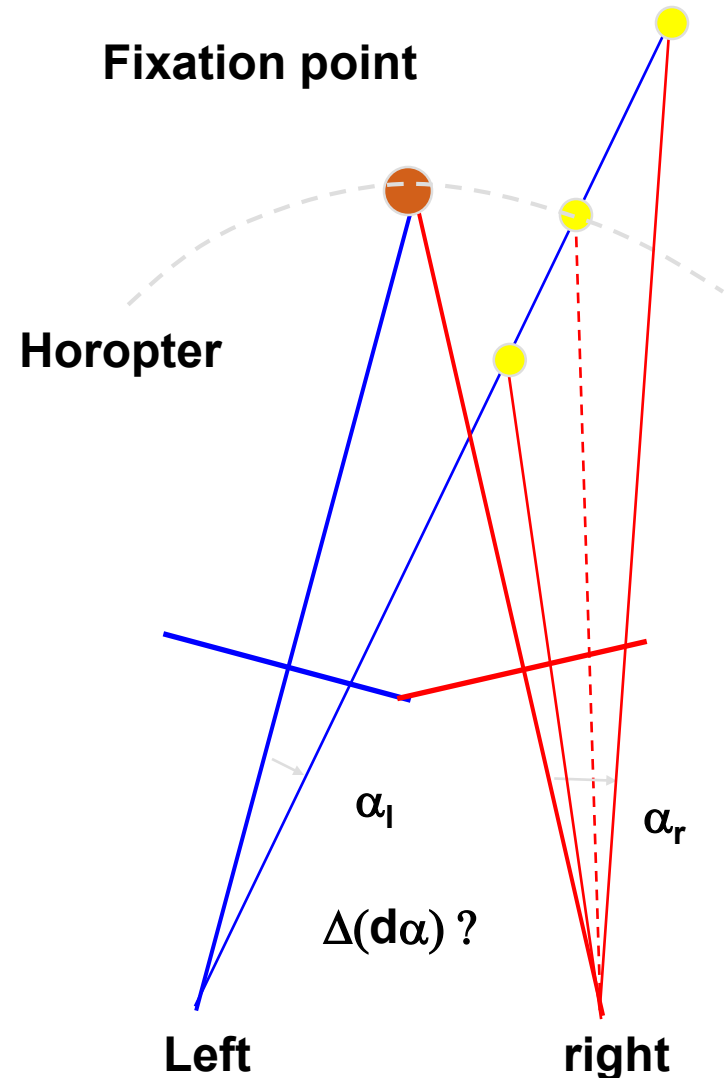  - Nearer the point, better the depth estimation

**Fixation point**

**Horopter**

$\alpha_r$

$\alpha_L$

$\alpha_r < \alpha_l$

$d\alpha < 0$

**Left**

**right**

# Stereo with Converging Cameras

- Two optical axes intersect at the Fixation Point
  - converging angle $\theta$
  - The common FOV Increases

- Disparity properties
  - Disparity uses angle instead of distance
  - Zero disparity at fixation point
    - and the Zero-disparity horopter
  - Disparity increases with the distance of objects from the fixation points
    - >0 : outside of the horopter
    - <0 : inside the horopter

- Depth Accuracy vs. Depth
  - Depth Error $\propto$ Depth$^2$
  - Nearer the point, better the depth estimation

**Fixation point**

**Horopter**

$\alpha_l$

$\alpha_r$

$\Delta(\mathbf{d}\alpha)$ ?

**Left**

**right**

# Parameters of a Stereo System
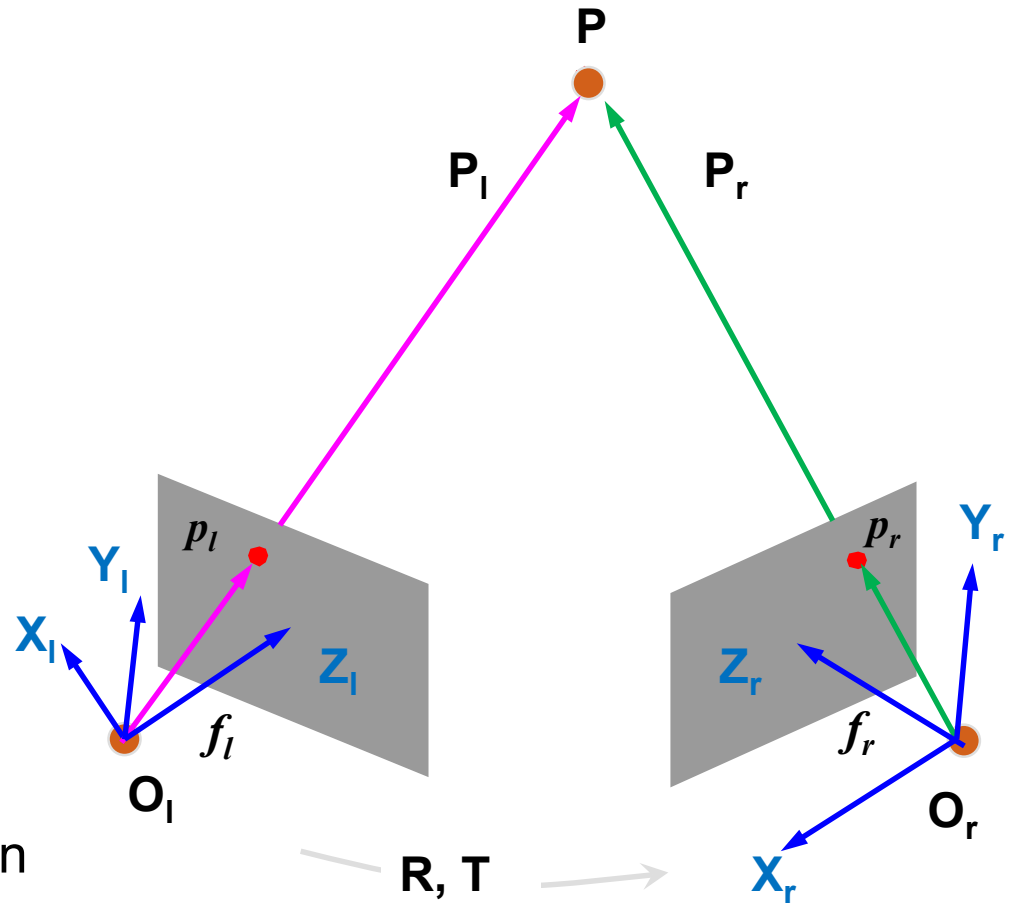
**Intrinsic Parameters**

Characterize the transformation from camera to pixel coordinate systems of each camera:

- focal length,
- image center,
- aspect ratio,
- distortion

**Extrinsic parameters**

Describe the relative position and orientation of the two cameras:

- Rotation matrix R
- translation vector T

- Notations
  - $P_l = (X_l, Y_l, Z_l)^T$, $P_r = (X_r, Y_r, Z_r)^T$
    - Vectors of the same 3-D point $P$, in the left and right camera coordinate systems respectively
  - Projection Equation

$$\mathbf{p}_l = \frac{f_l}{Z_l}\mathbf{P_l} \qquad \mathbf{p}_r = \frac{f_r}{Z_r}\mathbf{P_r}$$
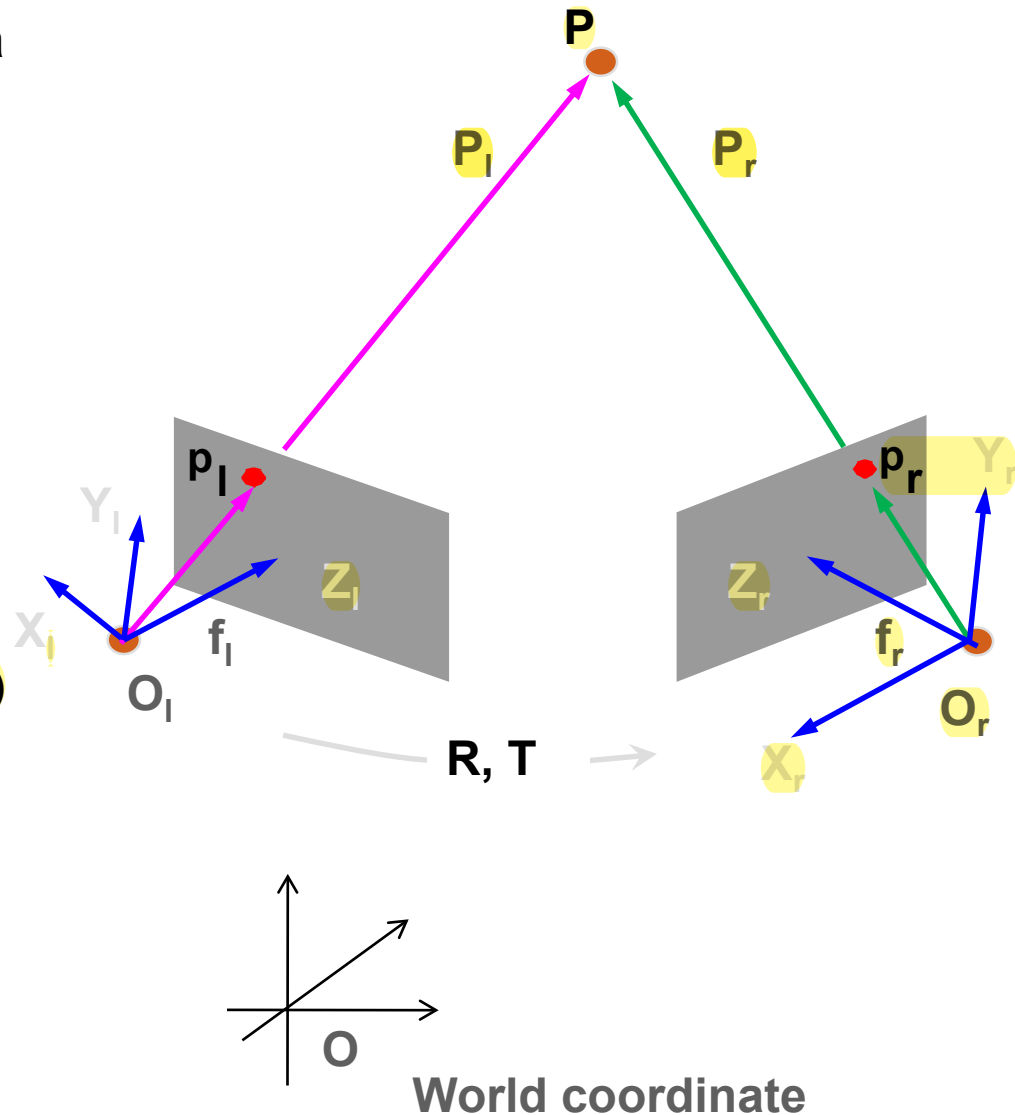
  - $p_l = (x_l, y_l, f_l)$, $p_r = (x_r, y_r, f_r)$
  - Extrinsic Parameters
    - Translation Vector T = $(O_r - O_l)$
    - Rotation Matrix R and

$$\mathbf{P_r} = \mathbf{R}^T(\mathbf{P_l} - \mathbf{T})$$

  - **P**, **O**$_l$ and **O**$_r$ are in the world coordinates.

# Epipolar Geometry

视线

P

$P_l$

$P_r$

$p_l$

$p_r$

$Y_l$

$Z_l$

$X_l$

$f_l$

$O_l$

$Y_r$

$Z_r$

$f_r$

$O_r$

$X_r$

**R, T**

O

**World coordinate**

- Motivation: where to search correspondences?
  - Epipolar Plane
    - A plane going through point P and the centers of projection (COPs) of the two cameras
  - Conjugated Epipolar Lines
    - Lines where epipolar plane intersects the image planes
  - Epipoles
    - The image in one camera of the COP of the other
- Epipolar Constraint
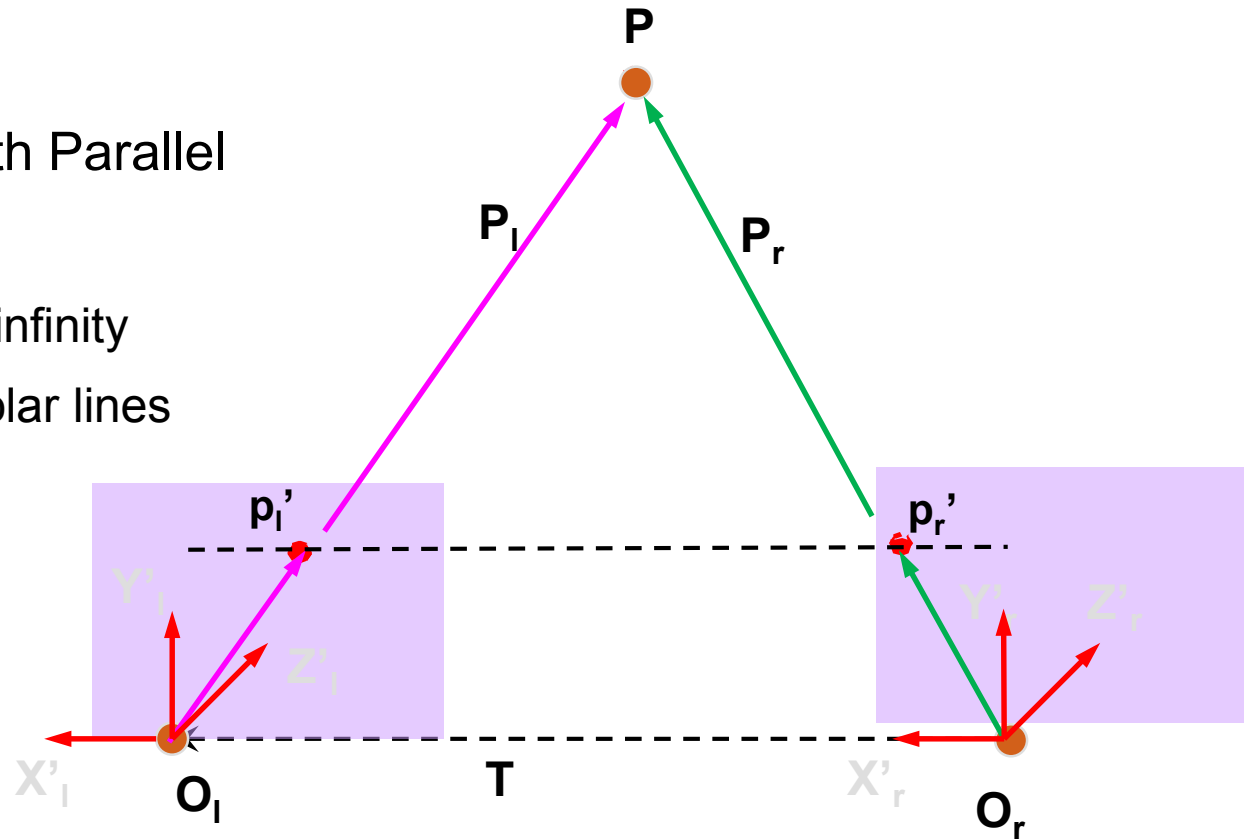  - Corresponding points must lie on conjugated epipolar lines
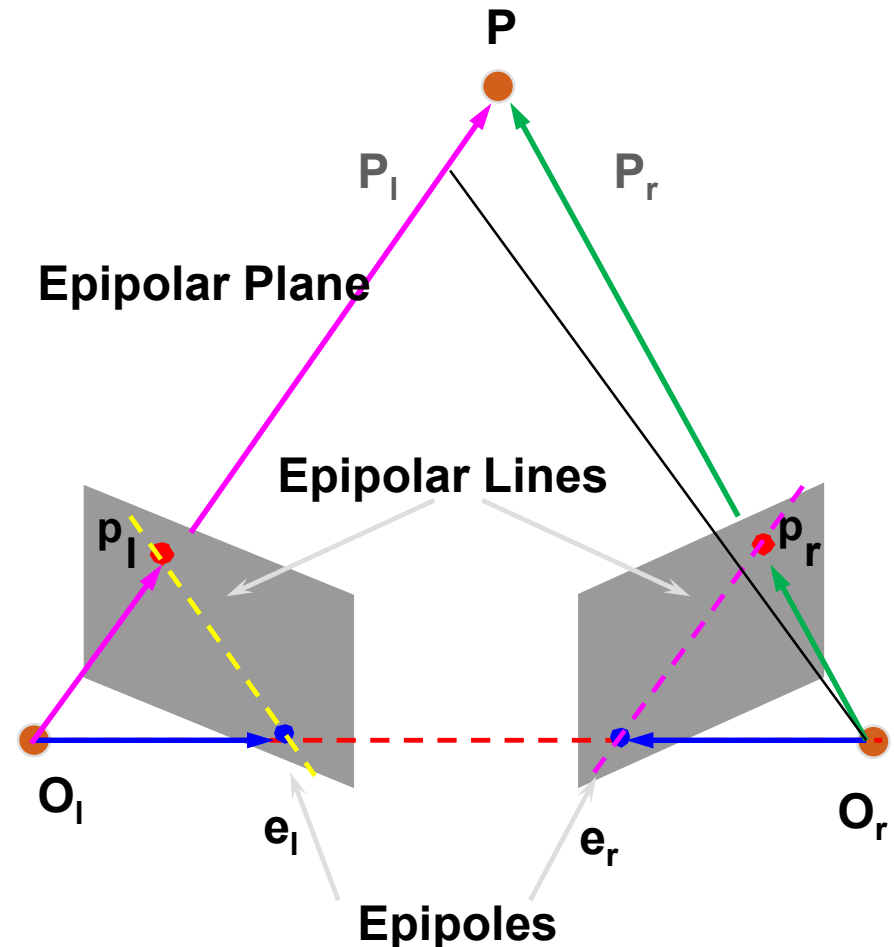
**P**

$P_l$   $P_r$

**Epipolar Plane**

**Epipolar Lines**

$p_l$   $p_r$

$O_l$   $e_l$   $e_r$   $O_r$

**Epipoles**
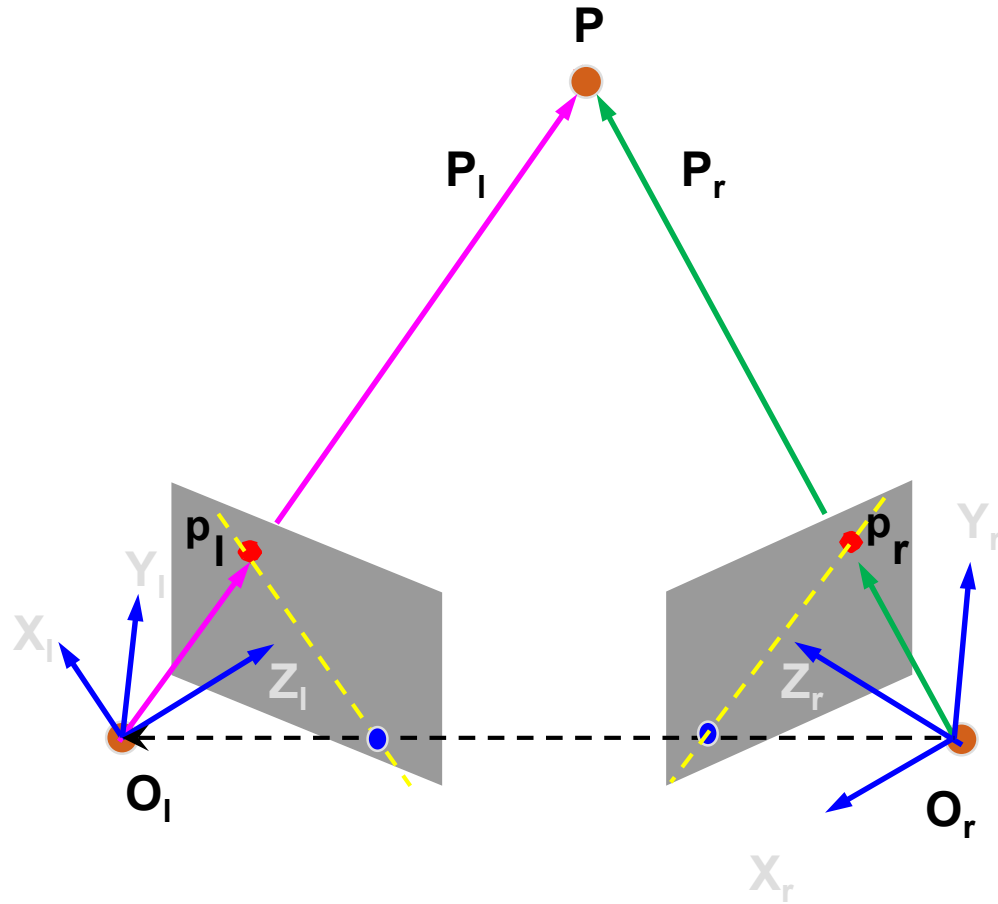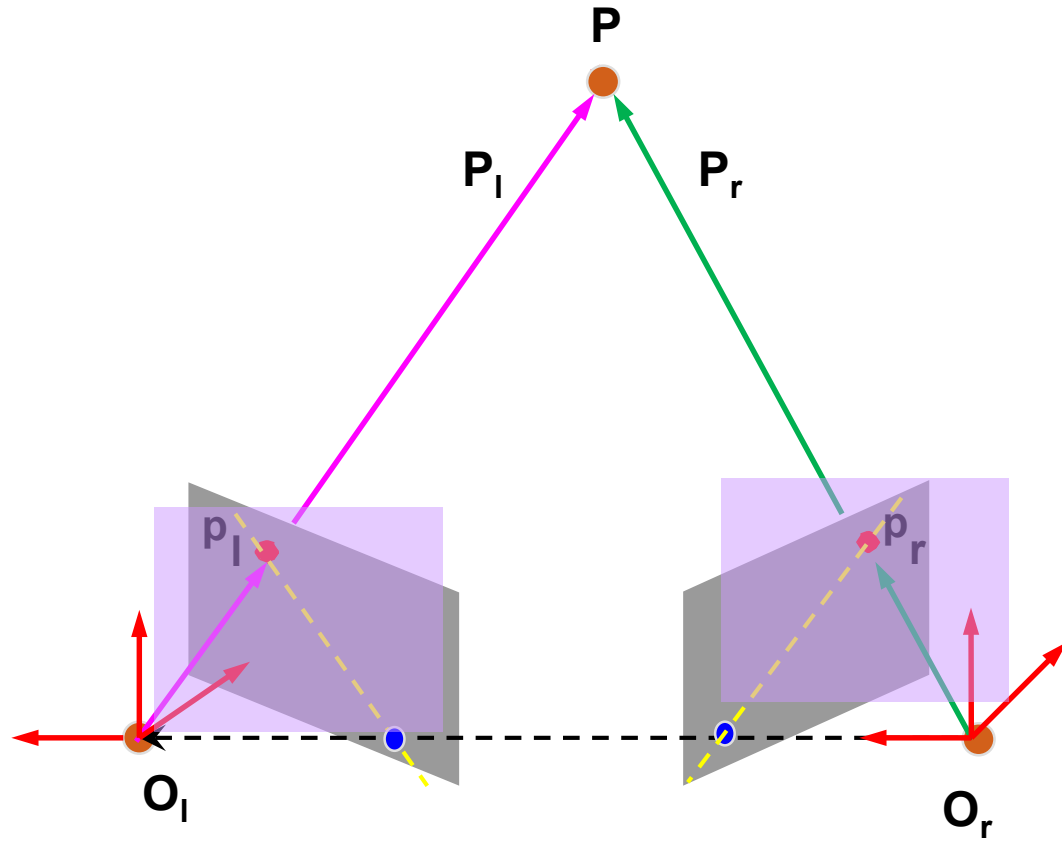
## Algorithm(optional)

- Rotate both left and right camera so that they share the same X axis : $O_r\text{-}O_l = T$
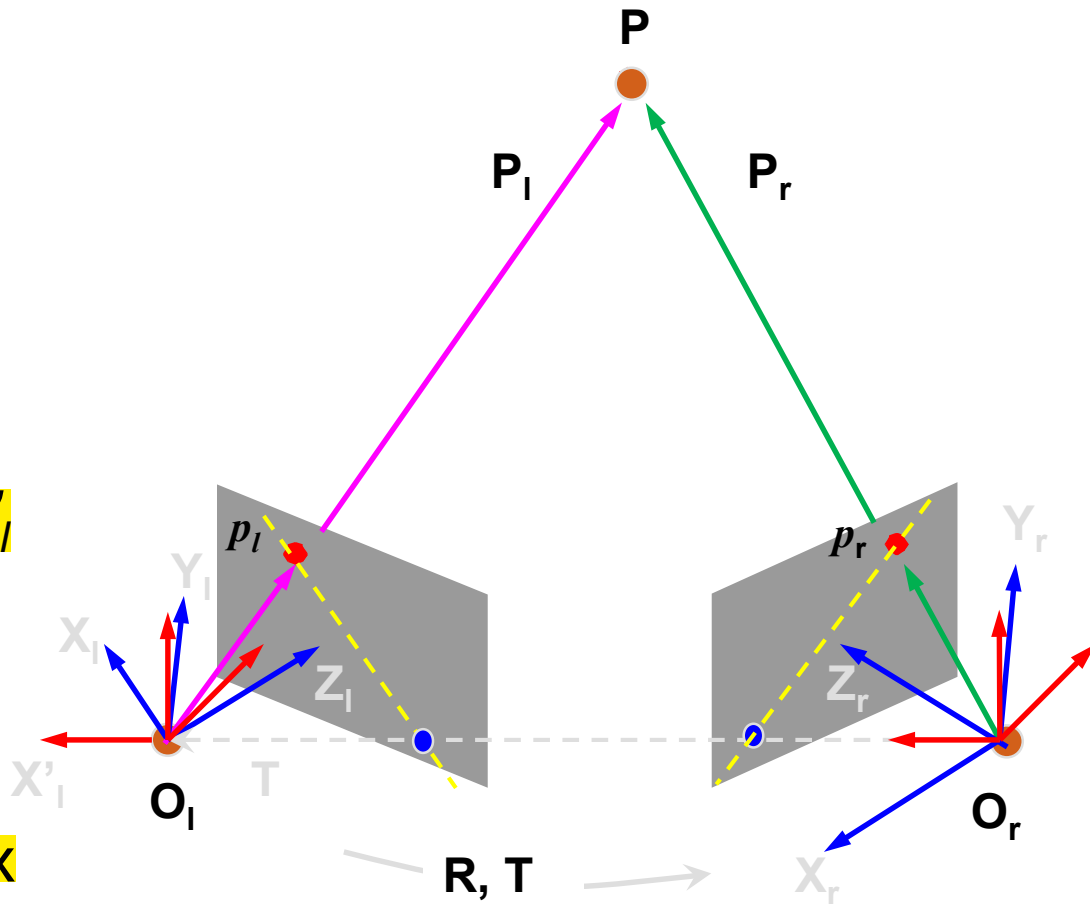
- Find a new virtual camera pose: $O_l'X_l'Y_l'Z_l'$

$$X_l' = \frac{-T}{\|T\|} \qquad Y_l' = -X_l' \times Z_l$$

$$Z_l' = X_l' \times Y_l'$$

- Form the rotation matrix $R_{rect}$ for the left camera from $O_lX_lY_lZ_l$ to $O_l'X_l'Y_l'Z_l'$

- Rotation Matrix for the right camera is $R_{rect}R^T$



**P**

**P$_l$**      **P$_r$**

$p_l$      $p_r$

$Y_l$      $Y_r$

$X_l$      $Z_l$      $Z_r$

$X'_l$      **T**      $O_r$

**O$_l$**

**R, T**      $X_r$

How to find **R$_{rect}$**?

The left image will be rotated to:

$R_{rect}(x_l, y_l, f_l)^T$
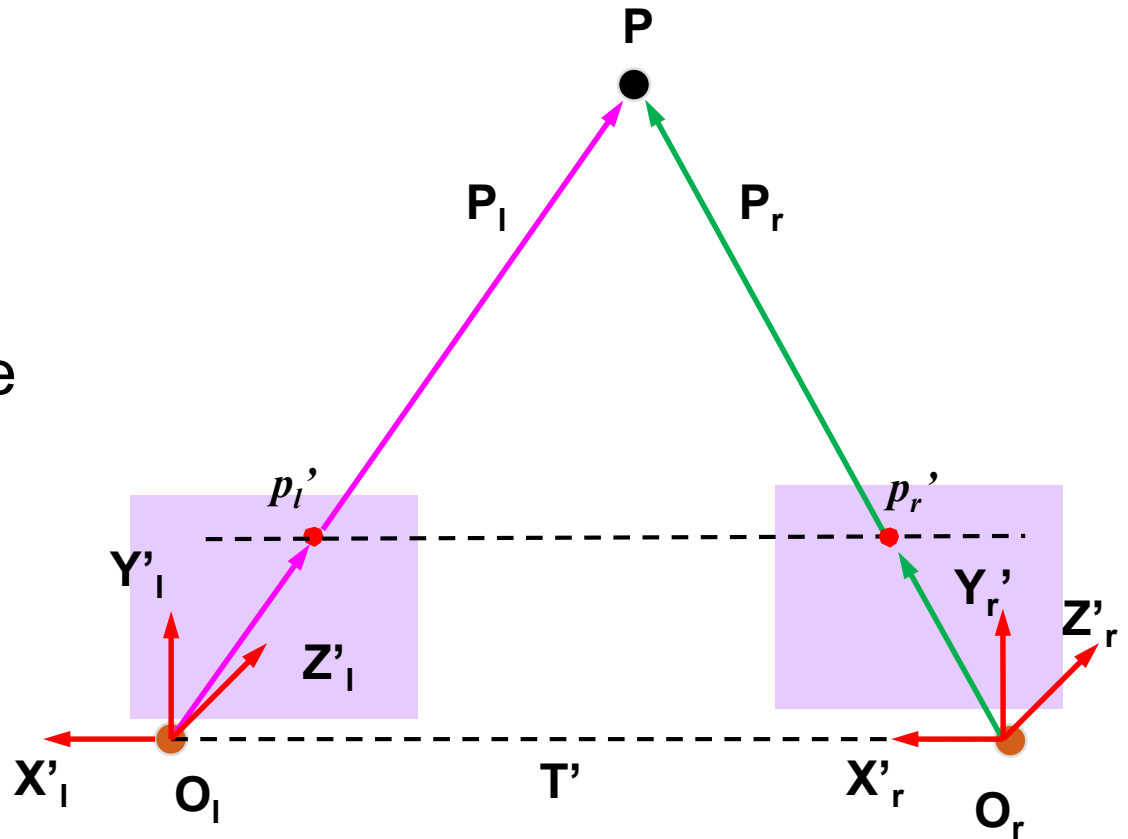
The right image will be rotated to:
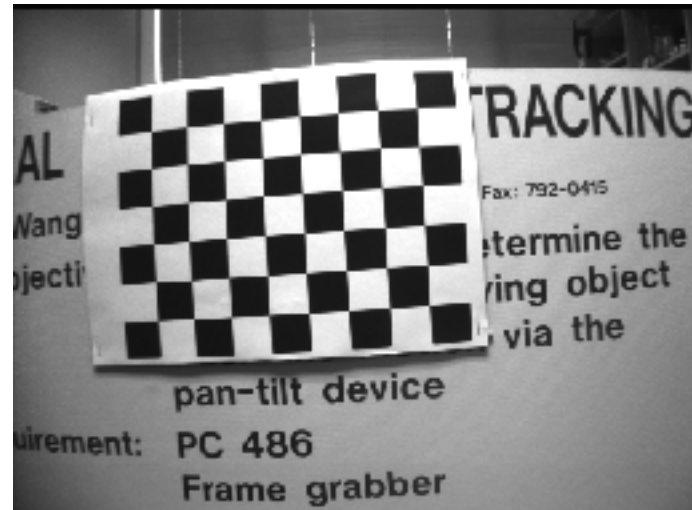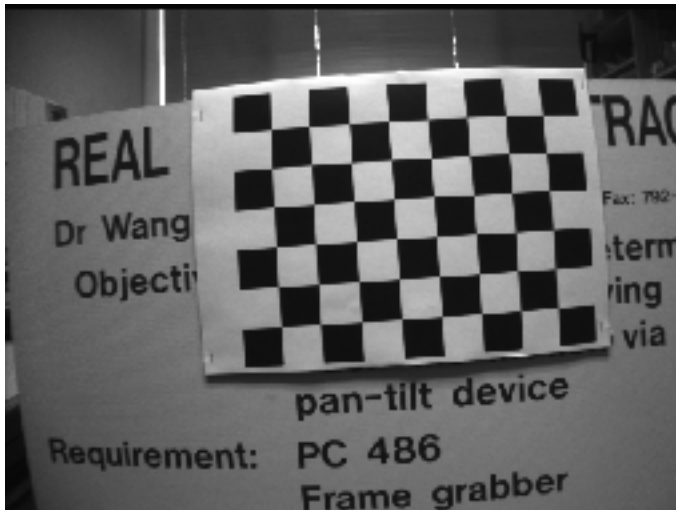
$R_{rect} R^T (x_r, y_r, f_r)^T$

After rectification,

T' = (B, 0, 0),
P'$_r$ = P'$_l$ − T'

Another procedure is required here: to correct the radial distortion. We skip its derivation.
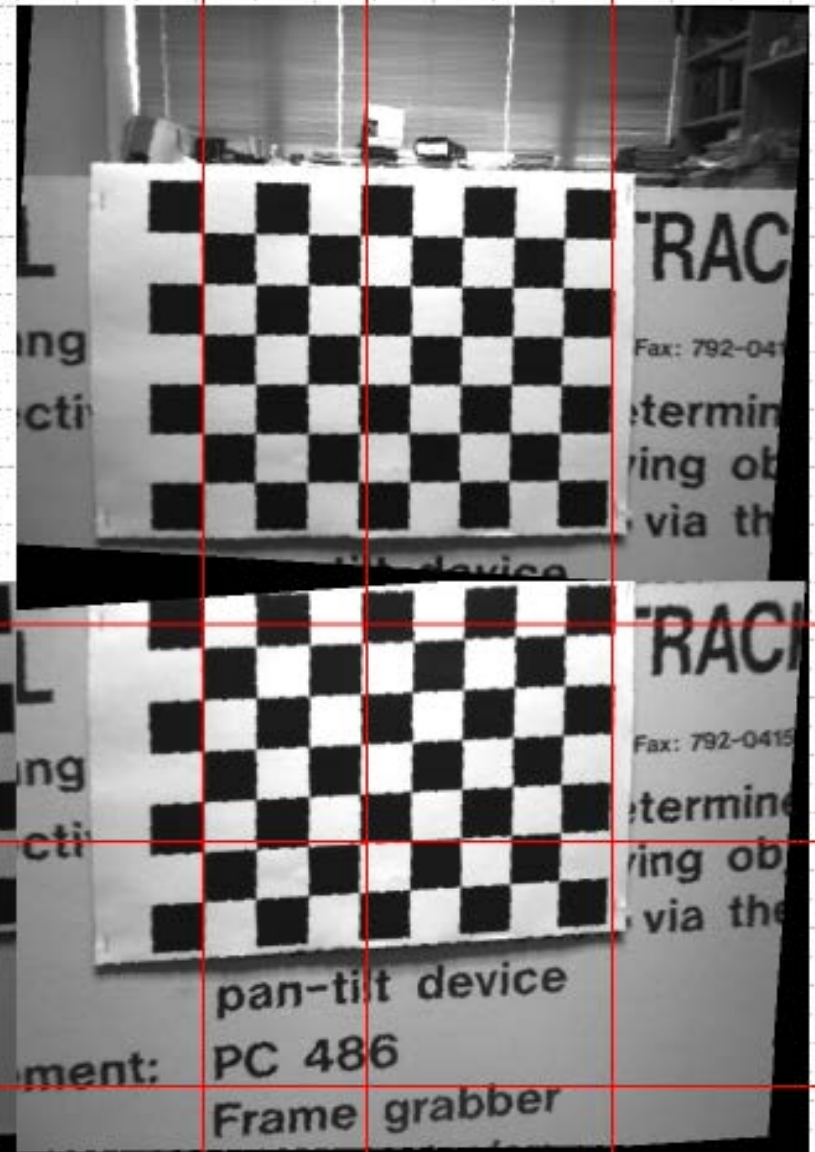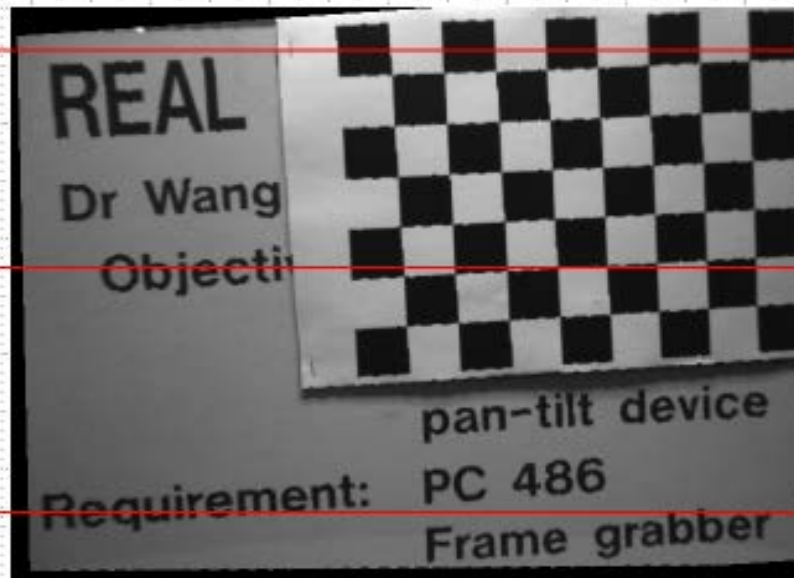
**Radial Distortion**

$$x_\mathrm{u} = x_\mathrm{c} + \frac{x_\mathrm{d} - x_\mathrm{c}}{1 + K_1 r^2 + K_2 r^4 + \cdots}$$

$$y_\mathrm{u} = y_\mathrm{c} + \frac{y_\mathrm{d} - y_\mathrm{c}}{1 + K_1 r^2 + K_2 r^4 + \cdots}$$

$(x_d, y_d)$ = distorted image point as projected on image plane using specified lens,

$(x_u, y_u)$ = undistorted image point as projected by an ideal pinhole camera,

$(x_c, y_c)$ = distortion center (assumed to be the principal point),

$K_n$ = n[th] radial distortion coefficient (can be found by collecting many images at            different distance, pose, …, and apply an optimisation equation.)

$$r = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2}$$

- Equation of the epipolar plane
  - Co-planarity condition of vectors $P_l$, T and $P_l$-T

$$(\mathbf{P_l} - \mathbf{T})^T \mathbf{T} \times \mathbf{P_l} = 0$$

$$\mathbf{P_r} = \mathbf{R}^T (\mathbf{P_l} - \mathbf{T})$$

- Essential Matrix    E = RS
  - 3x3 matrix constructed from R and T (extrinsic only)
    - Rank (E) = 2, two equal nonzero singular values

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

$$S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

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

$$\mathbf{p}_l = \frac{f_l}{Z_l} \mathbf{P_l}$$

$$\mathbf{p}_r = \frac{f_r}{Z_r} \mathbf{P_r}$$

**Rank (R) =3**

**Rank (S) =2**

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

- **Essential Matrix  E = RS**

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

  - A natural link between the stereo point pair and the extrinsic parameters of the stereo system
    - One correspondence -> a linear equation of 9 entries
    - Given 8 pairs of $(p_l, p_r)$  -> E

  - Mapping between points and epipolar lines we are looking for
    - Given $p_l$, E -> $p_r$ on the projective line in the right plane
    - Equation represents the epipolar line of either $p_r$ (or $p_l$) in the right (or left) image

- Note:
  - **$p_l$, $p_r$** are in the camera coordinate system, not pixel coordinates that we can measure(eg. unit focal length).

# Fundamental Matrix

- **Mapping between points and epipolar lines in the pixel coordinate systems**
  - With no prior knowledge on the stereo system
- **From Camera to Pixels:** Matrices of intrinsic parameters

$$\mathbf{M}_{\text{int}} = \begin{bmatrix} -f_x & 0 & o_x \\ 0 & -f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

**Rank ($\mathbf{M}_{\text{int}}$) =3**

- Questions:
  - What are $f_x$, $f_y$, $o_x$, $o_y$ ?
  - How to measure $\overline{p}_l$ in images?

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

$$\mathbf{p}_r = \mathbf{M}_r^{-1} \overline{\mathbf{p}}_r \qquad \mathbf{p_l} = \mathbf{M_l}^{-1} \overline{\mathbf{p}_l}$$

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

$$\mathbf{F} = \mathbf{M}_r^{-\mathbf{T}} \mathbf{E} \mathbf{M}_l^{-1}$$

- **Fundamental Matrix**
  - Rank (F) = 2

$$\mathbf{F} = \mathbf{M}_r^{-\mathbf{T}} \mathbf{E} \mathbf{M}_l^{-1}$$

  - Encodes info on both intrinsic and extrinsic parameters

  - Enables full reconstruction of the epipolar geometry
  - In pixel coordinate systems without any knowledge of the intrinsic and extrinsic parameters
  - Linear equation of the 9 entries of F

$$\overline{\mathbf{p}}_\mathbf{r}^{\mathbf{T}} \mathbf{F} \overline{\mathbf{p}}_\mathbf{l} = 0 \implies (x_{im}^{(l)} \quad y_{im}^{(l)} \quad 1) \begin{bmatrix} f11 & f12 & f13 \\ f21 & f22 & f23 \\ f31 & f32 & f33 \end{bmatrix} \begin{pmatrix} x_{im}^{(r)} \\ y_{im}^{(r)} \\ 1 \end{pmatrix} = 0$$

# Computing F: The Eight-point Algorithm

- Input: n point correspondences ( n >= 8)
  - Construct homogeneous system Ax= 0 from $\overline{\mathbf{p}}_{\mathbf{r}}^{\mathbf{T}} \mathbf{F} \overline{\mathbf{p}}_{\mathbf{l}} = 0$
    - x = ($f_{11}, f_{12}, ,f_{13}, f_{21}, f_{22}, f_{23} f_{31}, f_{32}, f_{33}$) : entries in F
    - Each correspondence give one equation $x_l x_r, x_l y_r, x_l,$ $y_l x_r, y_l y_r, y_r$ $x_r, y_r, 1$
    - A is a nx9 matrix
  - Obtain estimate F^ by SVD of A $\quad \mathbf{A} = \mathbf{UDV}^{T}$
    - x (up to a scale) is column of V corresponding to the least singular value
  - Enforce singularity constraint: since Rank (F) = 2
    - Compute SVD of F^ $\quad \hat{\mathbf{F}} = \mathbf{UDV}^{T}$
    - Set the smallest singular value to 0:  D -> D'
    - Correct estimate of F : $\quad \mathbf{F'} = \mathbf{UD'V}^{T}$
- Output:  the estimate of the fundamental matrix, F'
- Similarly we can compute E given intrinsic parameters

# Locating the Epipoles from F

$$\overline{p_r}^T F \overline{p_l} = 0$$

$e_l$ lies on all the epipolar lines of the left image

$$\overline{p_r}^T F \overline{e_l} = 0$$

**For every $p_r$**

F is not identically zero

$$F \overline{e_l} = 0$$
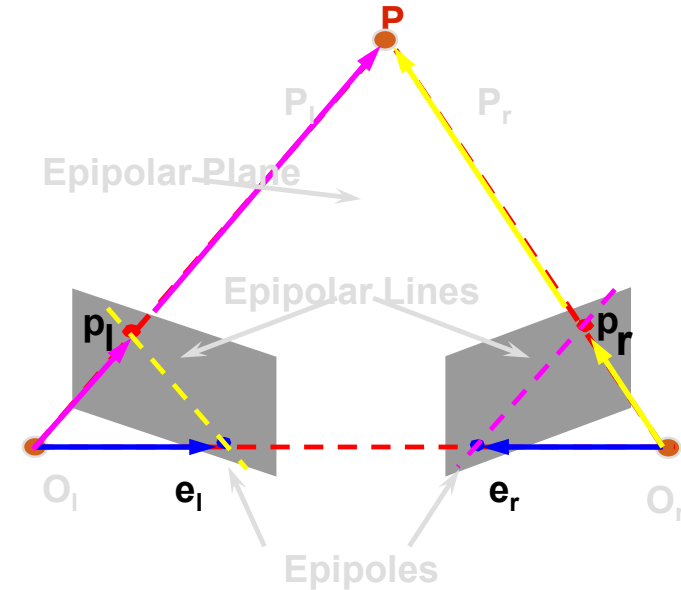
- **Input: Fundamental Matrix F**

  $$F = UDV^T$$

  - Find the SVD of F
  - The epipole $e_l$ is the column of V corresponding to the null singular value (as shown above)
  - The epipole $e_r$ is the column of U corresponding to the null singular value
- Output:  Epipole $e_l$ and $e_r$

- **Purpose**
  - where to search correspondences

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

- **Epipolar plane, epipolar lines, and epipoles**
  - known intrinsic (f) and extrinsic (R, T)
    - co-planarity equation
  - known intrinsic but unknown extrinsic

    $$p_r{}^T E p_l = 0$$
    - essential matrix
  - unknown intrinsic and extrinsic
    - fundamental matrix

    $$\overline{p_r}{}^T F \overline{p_l} = 0$$
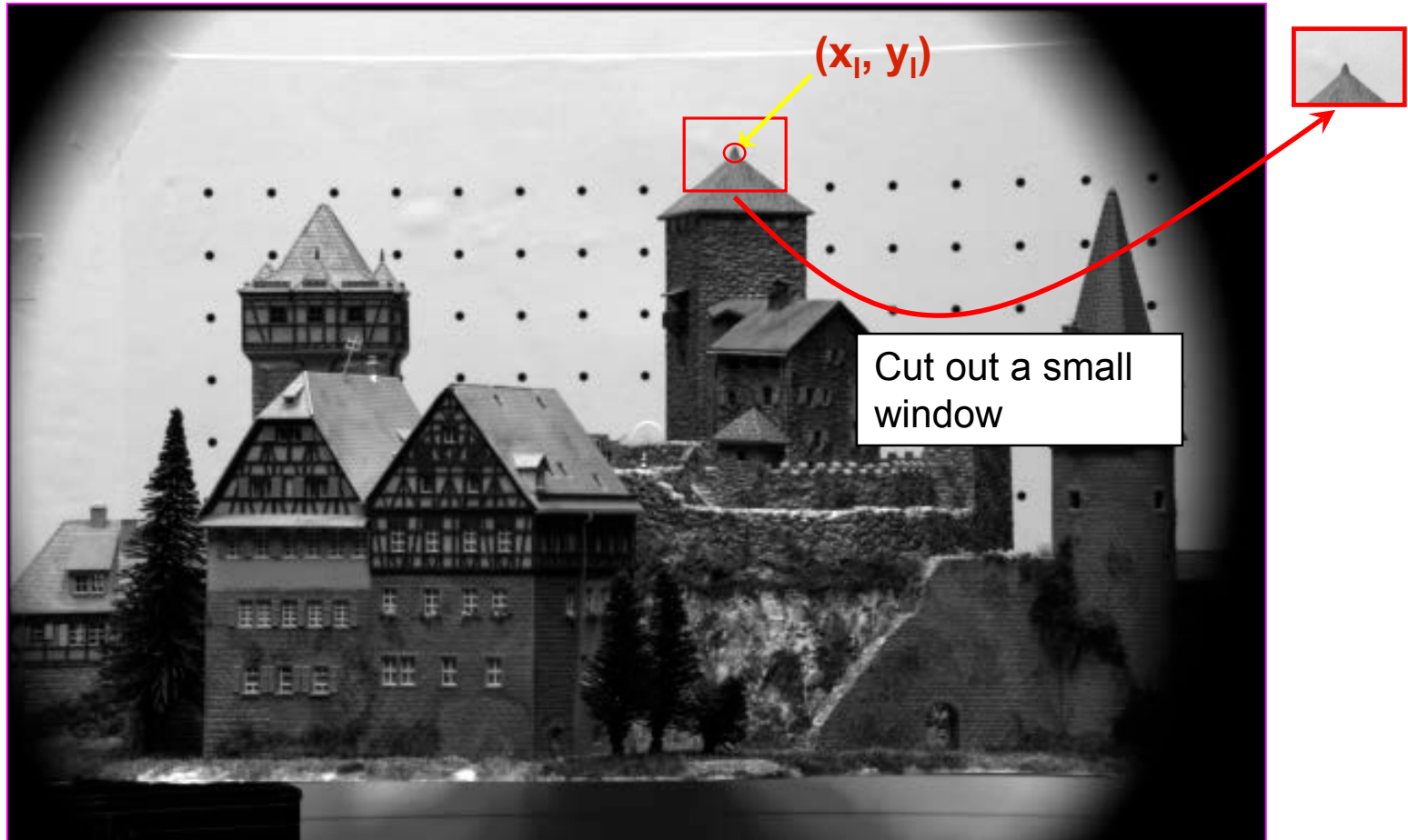
- **Rectification**
  - Generate stereo pair (by software) with parallel optical axis and thus horizontal epipolar lines

# Correspondence problem

- Three Questions
  - What to match?
    - Features: point, line, area, structure?
  - Where to search correspondence?
    - Epipolar line?
  - How to measure similarity?
    - Depends on features
- Approaches
  - Correlation-based approach
  - Feature-based approach
- Advanced Topics
  - Image filtering to handle illumination changes
  - Adaptive windows to deal with multiple disparities
  - Local warping to account for perspective distortion
  - Sub-pixel matching to improve accuracy
  - Self-consistency to reduce false matches
  - Multi-baseline stereo

**LEFT IMAGE**



$(x_l, y_l)$

Cut out a small window

- For Each point $(x_l, y_l)$ in the left image, define a window centered at the point

**RIGHT IMAGE**



$(x_r, y_r)$  **dx**  $(x_l, y_l)$

The epipolar line

Move this window on the right image along the epipolar line to find the same pattern; (pixel by pixel)
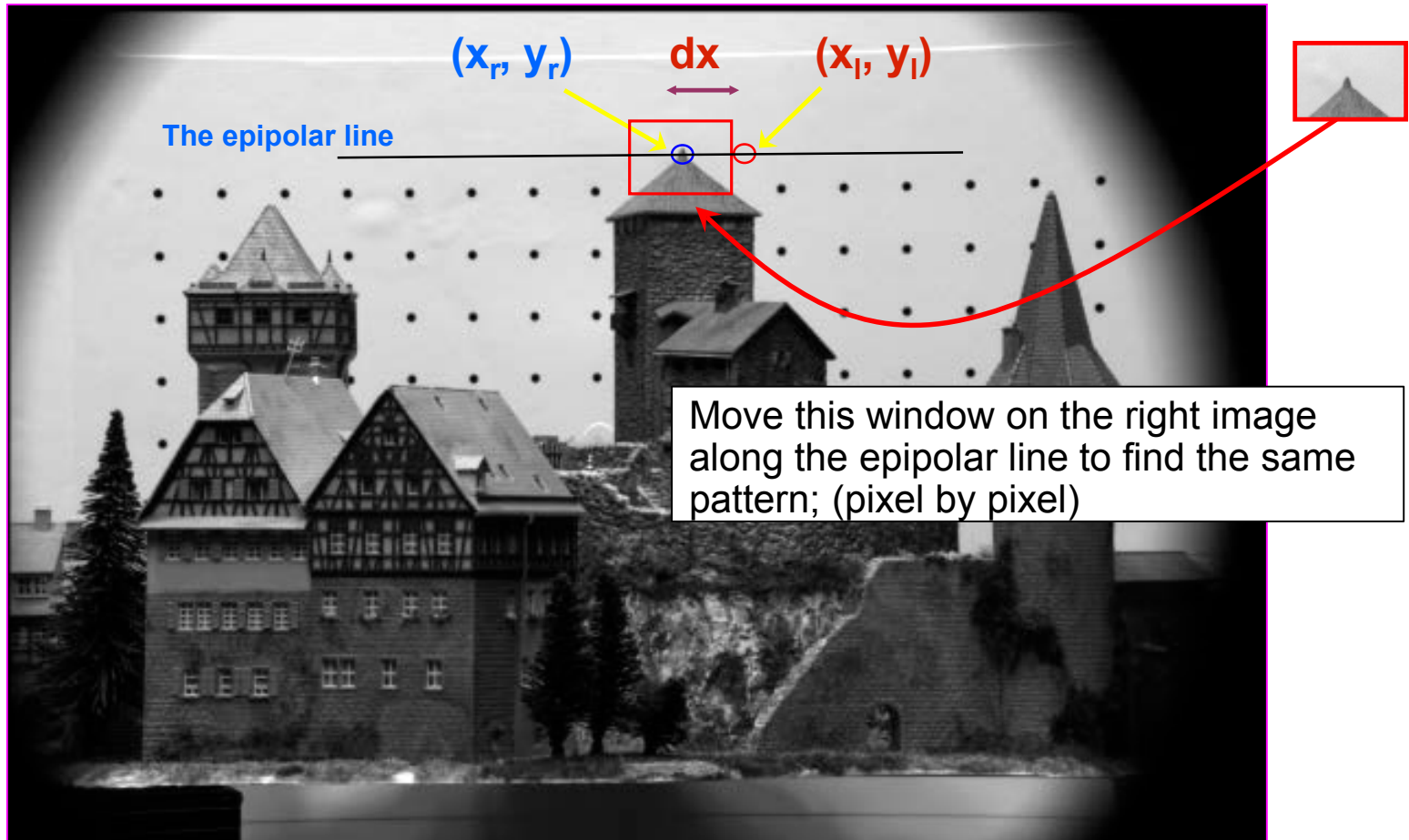
- The epipolar line is perfectly horizontal, because we have parallel optical axes after rectification. It means reduced effort in searching for matched pattern, and less ambiguity.

# Correlation Approach

- **Elements to be matched**
  - Image window of fixed size centered at each pixel in the left image
- **Similarity criterion**
  - A measure of similarity between windows in the two images
  - The corresponding element is given by window that maximizes the similarity criterion within a search region
- **Search regions**
  - Theoretically, search region can be reduced to a 1-D segment, along the epipolar line, and within the disparity range.
  - In practice, search a slightly larger region due to errors in calibration

# Correlation Approach

- **Equations**

$$c(dx, dy) = \sum_{k=-W}^{W} \sum_{l=-W}^{W} \psi(I_l(x_l + k, y_l + l), I_r(x_l + dx + k, y_l + dy + l))$$

- **disparity**

$$\overline{\mathbf{d}} = (\overline{dx}, \overline{dy}) = \arg \max_{\mathbf{d} \in R} \{c(dx, dy)\}$$

- **Similarity criterion**
  - Cross-Correlation
  
  $$\Psi(u, v) = uv$$

  - Sum of Square Difference (SSD)
  
  $$\Psi(u, v) = -(u - v)^2$$

  - Sum of Absolute Difference(SAD)
  
  $$\Psi(u, v) = -|u - v|$$

# Correlation Approach

- **PROS**
  - Easy to implement
  - Produces dense disparity map
  - Maybe slow

- **CONS**
  - Needs textured images to work well
  - Inadequate for matching image pairs from very different viewpoints due to illumination changes
  - Window may cover points with quite different disparities
  - Inaccurate disparities on the occluding boundaries

# Template matching of Lena's hat

Scan the template over each pixel of the image of Lena and compute $\gamma$ for each pixel.

We obtain the 2D $\gamma$-map. If we represent higher values of $\gamma$ as brighter intensity, we get the $\gamma$ image. The max value of $\gamma$ is deemed the best match.
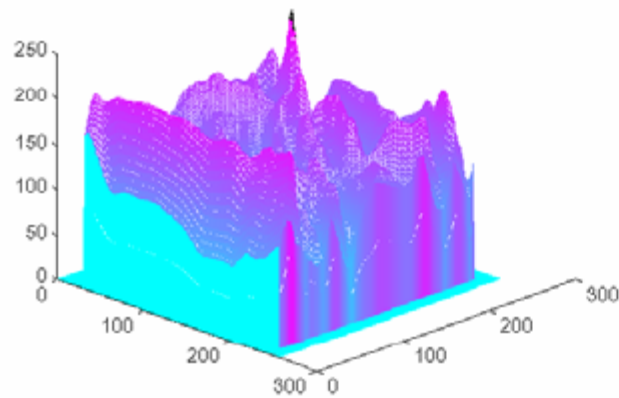


Original image of Lena



template



$\gamma$ - map



Max $\gamma$
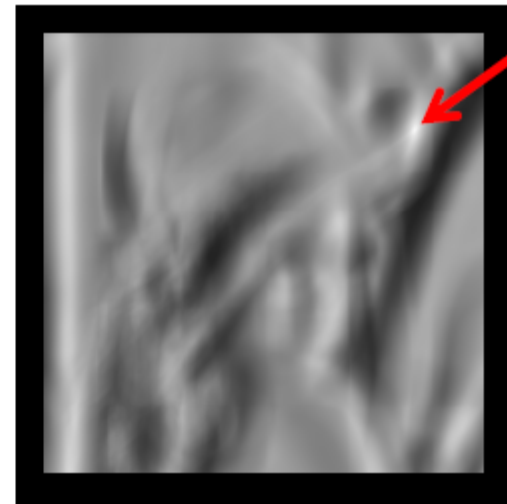
Result of template matching

① Vectorize the template, $t = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{vmatrix}$

$\vec{t} = (1, 1, 1, 1, 0, 1, 0, 0, 0)$

find the mean

$$t_{mean} = \frac{5}{9}$$

Adjust $\vec{t}$, so that it is ⟨zero-meaned⟩.

$\vec{t} = (1 - t_{mean}, \cdots )$

$$= (\frac{4}{9}, \frac{4}{9}, \frac{4}{9}, \frac{4}{9}, -\frac{5}{9}, \frac{4}{9}, -\frac{5}{9}, -\frac{5}{9}, -\frac{5}{9})$$

Make it into unit vector;

$$\|\vec{t}\| = \sqrt{(\frac{4}{9})^2 + (\frac{4}{9})^2 + \cdots} = \frac{\sqrt{180}}{9}$$

$$\vec{T} = \frac{\vec{t}}{\|\vec{t}\|} = \frac{9}{\sqrt{180}} (\frac{4}{9}, \frac{4}{9}, \cdots)$$

$$= \frac{1}{\sqrt{180}} (4, 4, 4, 4, -5, \cdots)$$

Note: $\|\vec{T}\| = 1$

**Copy out each pixel, form a 1x9 vector. If the template is 15x15, the vector is 1x225.**

not subject to

lighting

② vectorize the image $\quad i = \begin{bmatrix} 100 & 100 & 100 \\ 100 & 10 & 100 \\ 10 & 10 & 10 \end{bmatrix}$

$\therefore \vec{i} = (100, 100, 100, 100, 10, \cdots)$

find the mean

$\quad i_{mean} = 60$

adjust $\vec{i}$, it is zero-meaned

$\quad \vec{i} = (100-60, \ 100-60, \cdots)$

$\quad = (40, \ 40, \ 40, \ 40, \ -50, \cdots)$

make it into unit vector

$\quad \|\vec{i}\| = \sqrt{40^2 + 40^2 + \cdots} = \sqrt{18000}$

$\therefore \quad \hat{I} = \dfrac{\vec{i}}{\|\vec{i}\|} = \dfrac{1}{\sqrt{18000}} (40, \ 40, \cdots)$

$\quad = \dfrac{1}{\sqrt{180}} (4, \ 4, \ 4, \ 4, \ -5, \cdots)$

This vectorised template is then used to match against data. For example, a 3x3 window extracted from the right image on the epipolar line. The left is a sample. The triangle pattern is there, but this image is brighter.

Repeat the same process to vectorise this data.

③ Inner product $\vec{T} \cdot \vec{I}$

note: $\vec{T} = \vec{I}$

∴ $\vec{T} \cdot \vec{I} = 1$

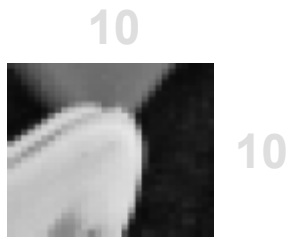④ You may try other pattern of

$$i = \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 3 & 2 & 1 \\ \hline 2 & 3 & 4 \\ \hline \end{array}$$

and observe the new

$$\vec{T} \cdot \vec{I} < 1$$

1 means the 2 patterns are identical. The worst case is -1 (when 2 vectors are in opposite direction).

Suppose the template has 10x10 pixels, we slice the template in raster order. Then form a list of 100 pixels. We treat this list as an 100 dimension vector, namely *w*. We search through the image. At each pixel location, we cut out a sample of 10x10, then form another 100-vector, call it *f*. The following method is better than SAD.

The objective is to measure the similarity between $f$ and $w$. Let

$f = (f_1, f_2, \ldots f_{100})$ and $w = (w_1, w_2, \ldots w_{100})$

We first find the mean of $f$ and $w$, and adjust each vector by subtracting the mean $f_{av}$ & $w_{av}$ and then normalise the two vectors into unit vectors.

Mean adjustment is necessary, as this reduces side effect of lighting changes between template and image.

As we understand that if two unit vectors are identical, the inner product of the two will produce 1. Any "dissimilarity" will make the product less than 1. The worst case is -1.

_Calculating ZNCC in vector form_

In the previous example, $f$ and $w$ were put into vector form thus :

$$f = [6\ 7\ 2\ 5\ 6\ 9\ 7\ 8\ 9\ ]^\mathrm{T}\ \text{and}\ w = [5\ 6\ 7\ 44\ 5\ 6\ 6\ 7\ 8\ ]^\mathrm{T}$$

In general, let $h = f - f_{ave}$ and $g = w - w_{ave}$, then

$$\gamma = \frac{h^T g}{\|h\|\ \|g\|}. \quad \textit{Since} \quad \frac{h}{\|h\|} \ \textit{is a unit vector and so is} \ \frac{g}{\|g\|},$$

therefore $\boxed{\gamma = \cos\theta}$ where $\theta$ is the angle between $h$ and $g$.

Recall : $a^\mathrm{T}b = \|a\|\ \|b\|\ \cos\theta$

**So ZNCC is simply the inner product between two unit vectors !!!**

Now you know why $\gamma$ is between $-1$ and $1$.

# This method is called ZNCC—zero centred normalised correlation coefficient

**Properties of** $\gamma = \mathbf{ZNCC}$ :

- $-1 \leq \gamma(s,t) \leq 1$    (why ?)

- summation is taken over the image region where $w$ and $f$ overlap.

- The ***maximum*** value of $\gamma(s,t)$ appears at the position where $w(x, y)$ best matches $f(x, y)$.

- Advantage : easy to implement.

- Disadvantages : sensitive to image rotation and image scale changes, etc.

# Correlation Approach (dense stereo example)
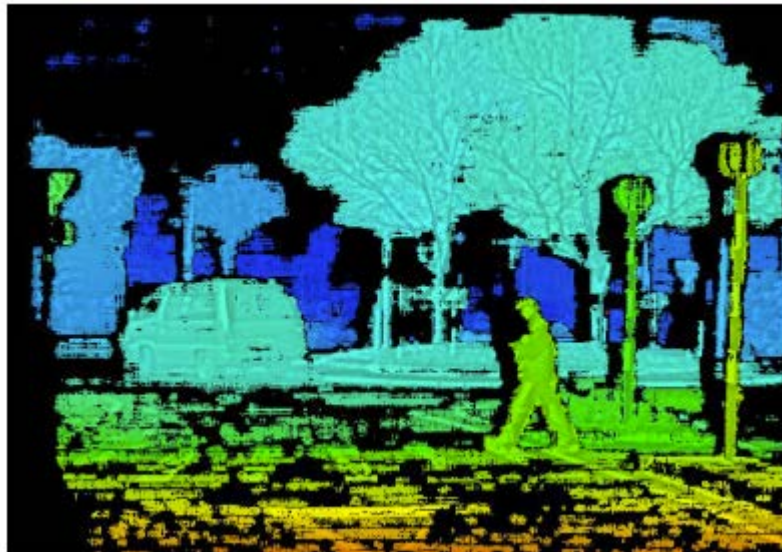
- A Stereo Pair of NTU Campus with corners superimposed



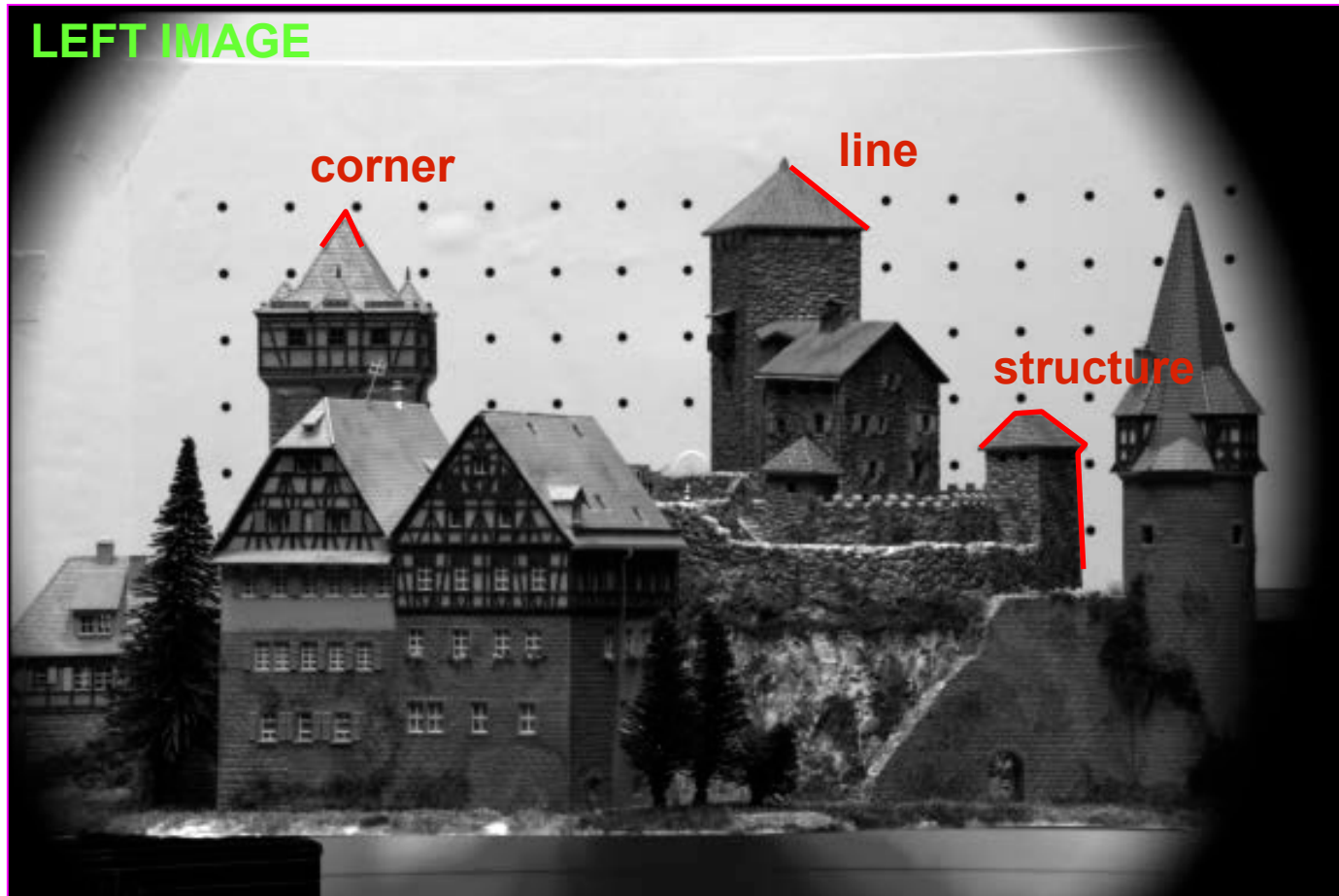**Actual epipolar line without rectification**

- **Features**
  - Edge points
  - Lines (length, orientation, average contrast)
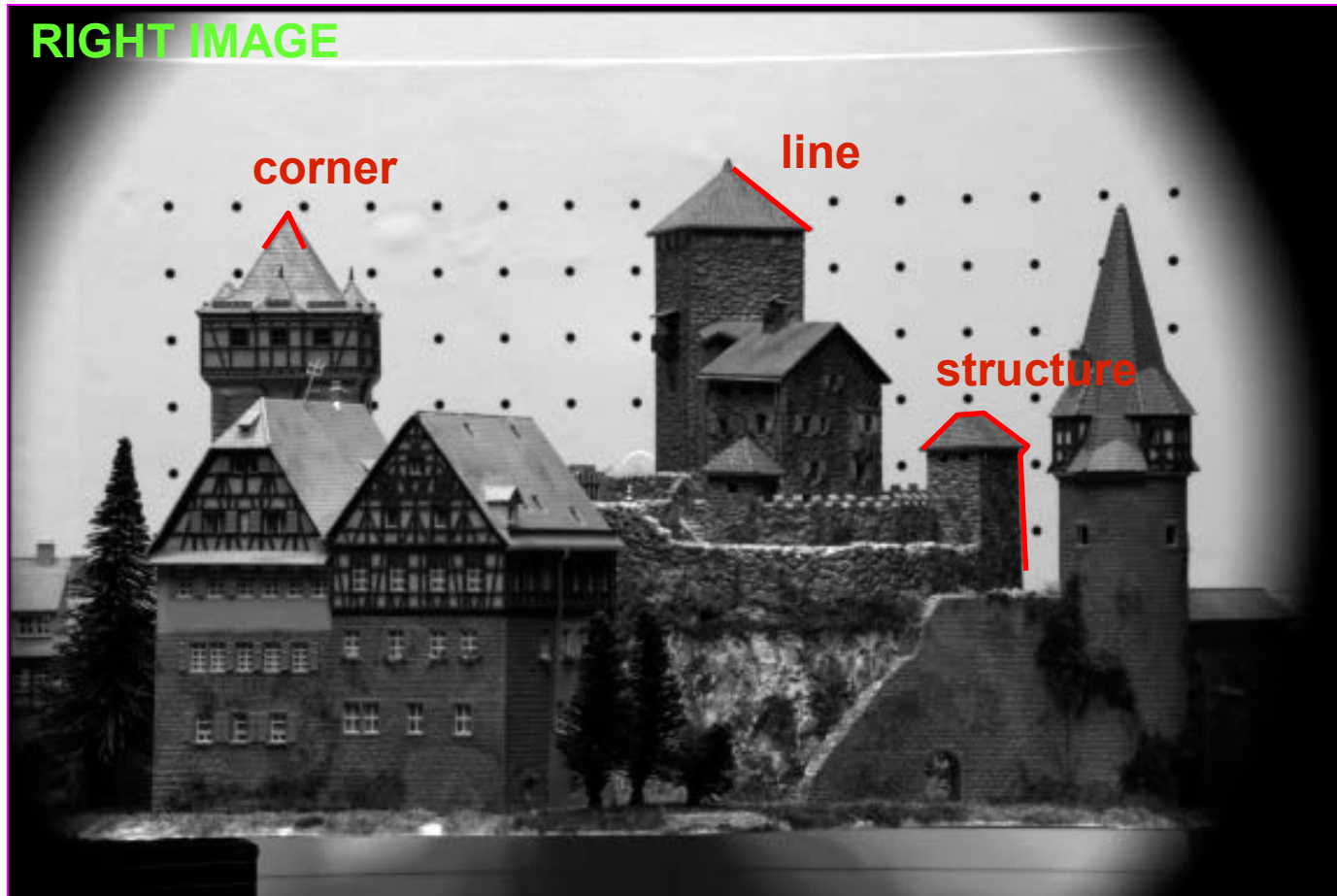  - Corners
  - SIFT

- **Matching algorithm**
  - Extract features in the stereo pair
  - Define similarity measure
  - Search correspondences using similarity measure and the epipolar geometry

- For each feature in the left image…

**RIGHT IMAGE**

corner

line

structure

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

- **PROS**
  - Relatively insensitive to illumination changes
  - Good for man-made scenes with strong lines but weak texture or textureless surfaces
  - Work well on the occluding boundaries (edges)
  - Could be faster than the correlation approach

- **CONS**
  - Only sparse depth map
  - Feature extraction may be tricky
    - Lines (Edges) might be partially extracted in one image
    - How to measure the similarity between two lines?

- Mainly used in correlation-based approach, but can be applied to feature-based match

- Image filtering to handle illumination changes

  - Image equalization
    - To make two images more similar in illumination

  - Laplacian filtering (2$^{nd}$ order derivative)
    - Use derivative rather than intensity (or original color)

- Fundamental concepts and problems of stereo
- Epipolar geometry and stereo rectification
- ~~Estimation of fundamental matrix from 8 point pairs~~
- Correspondence problem and two techniques: correlation and feature based matching
- Reconstruct 3-D structure from image correspondences given
  - Fully calibrated
  - Partially calibration
  - Uncalibrated stereo cameras (*)