



CSc I6716
Spring 2011



Topic 3 of Part II Stereo Vision

Zhigang Zhu, City College of New York zhu@cs.cny.cuny.edu



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

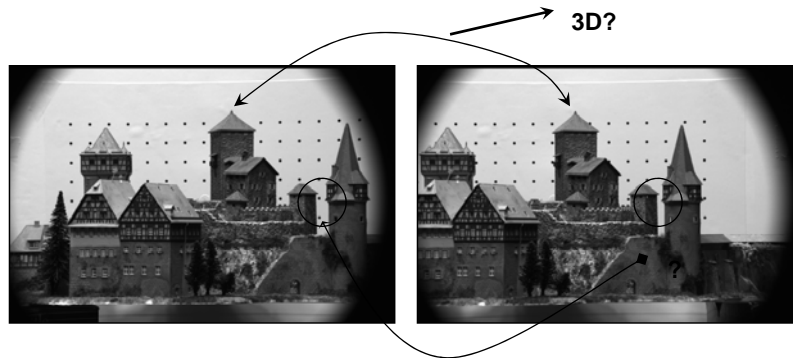
- Two primary Sub-problems
 - Correspondence problem (stereo match) -> disparity map
 - “Similar” instead of “Same”
 - Occlusion problem: some parts of the scene are visible only in one eye
 - Reconstruction problem -> 3D
 - What we need to know about the cameras' parameters
 - Often a stereo calibration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

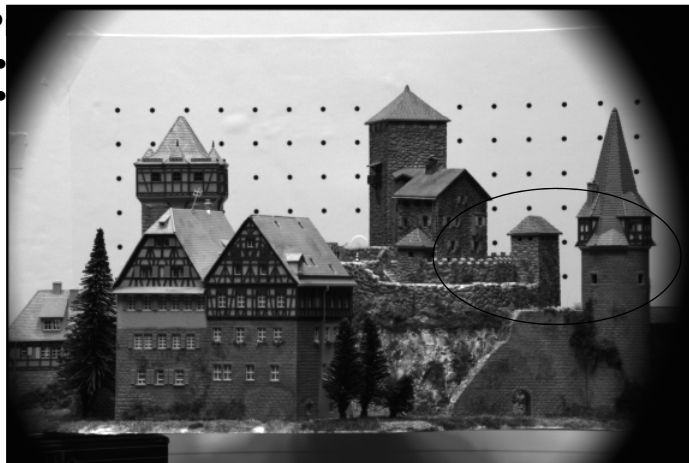


CMU CIL Stereo Dataset : Castle sequence

<http://www-2.cs.cmu.edu/afs/cs/project/cil/ftp/html/cil-ster.html>



■ P

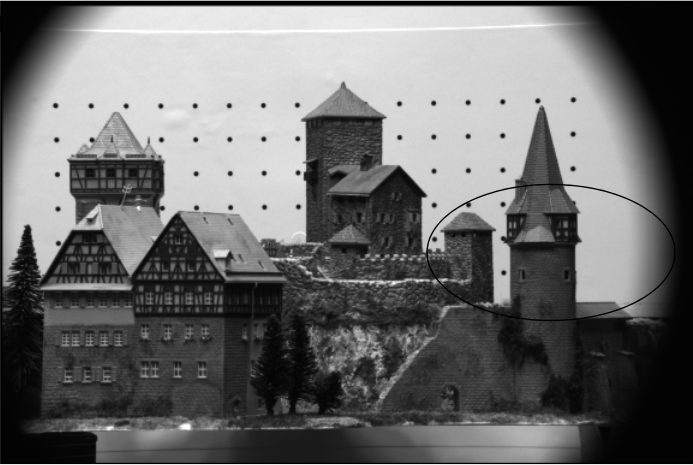




■ P

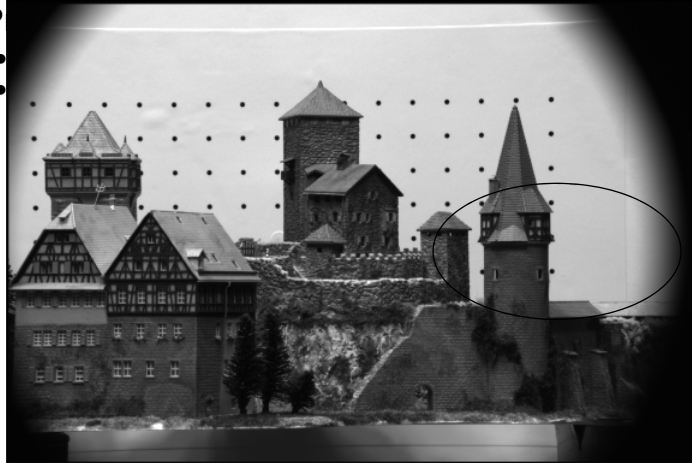


■ P

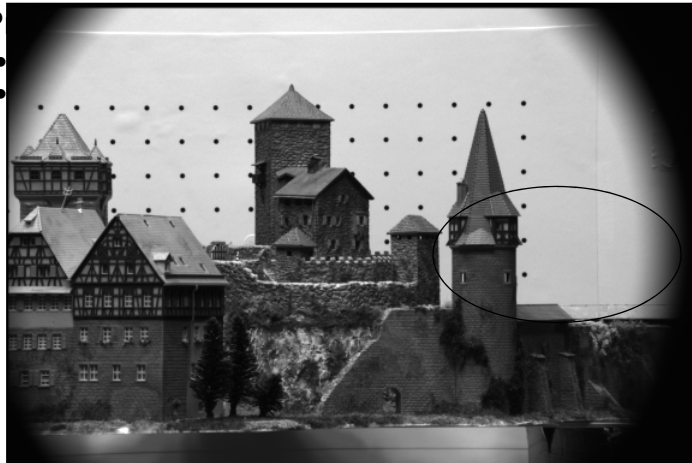




■ P

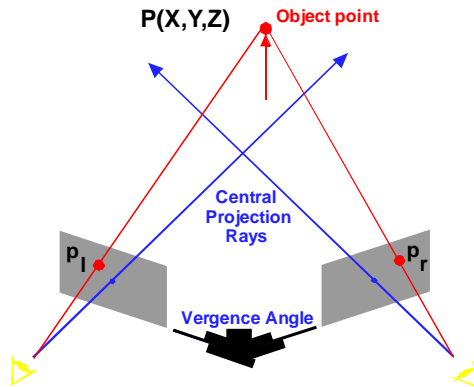


■ P



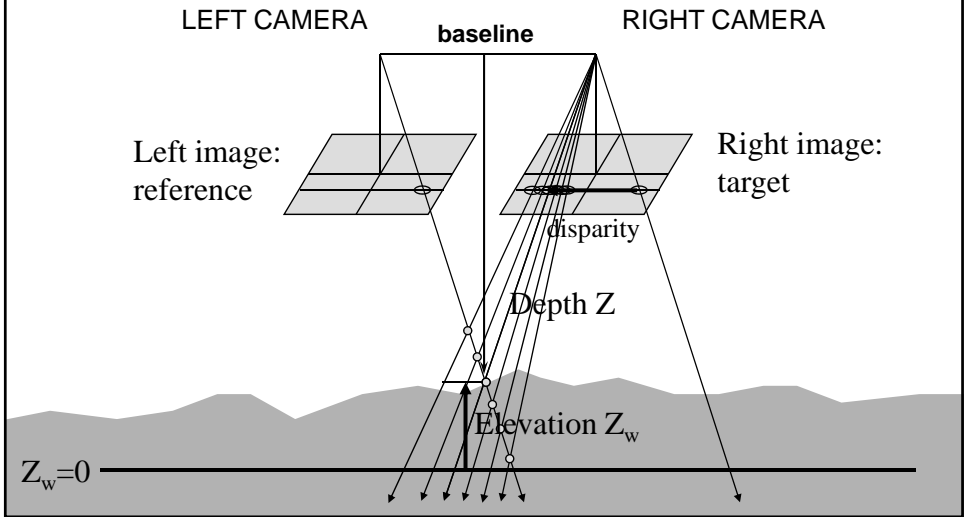
- A Simple Stereo Vision System
 - Disparity Equation
 - Depth Resolution
 - Fixated Stereo System
 - Zero-disparity Horopter

- Epipolar Geometry
 - Epipolar lines – Where to search correspondences
 - Epipolar Plane, Epipolar Lines and Epipoles
 - <http://www.ai.sri.com/~luong/research/Meta3DViewer/EpipolarGeo.html>
 - Essential Matrix and Fundamental Matrix
 - Computing E & F by the Eight-Point Algorithm
 - Computing the Epipoles
- Stereo Rectification

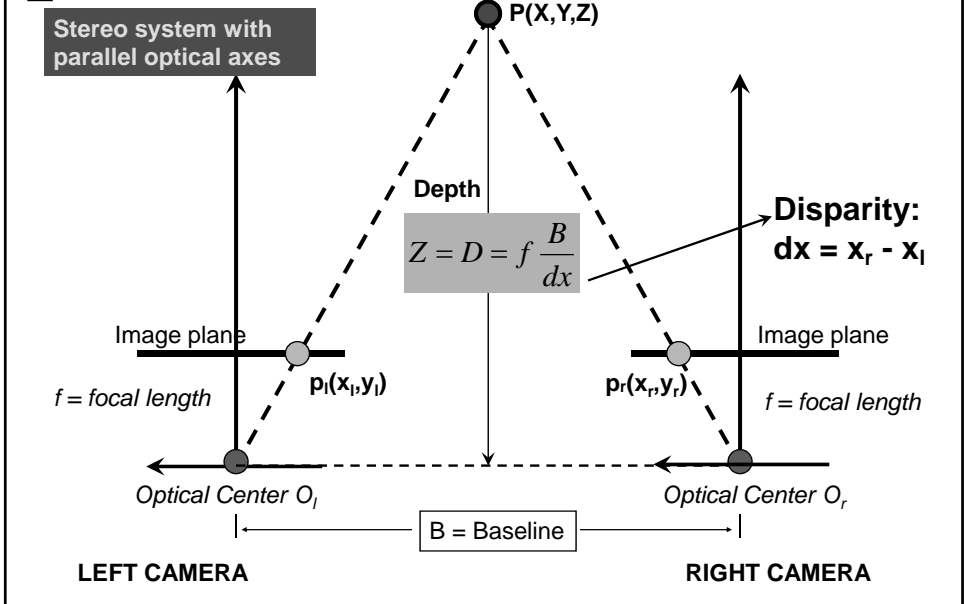


- Converging Axes – Usual setup of human eyes
- Depth obtained by triangulation
- Correspondence problem: p_l and p_r correspond to the left and right projections of P , respectively.

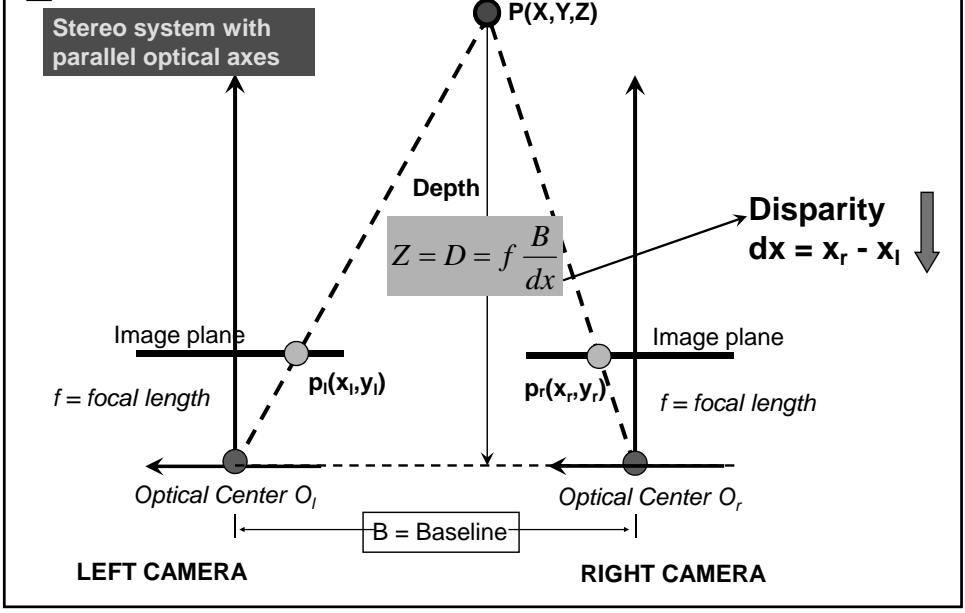
3D Computer Vision and Video Computing **A Simple Stereo System**



3D Computer Vision and Video Computing **Disparity Equation**

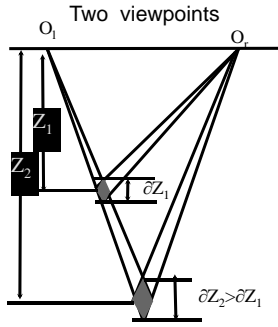


3D Computer Vision and Video Computing **Disparity vs. Baseline**



3D Computer Vision and Video Computing **Depth Accuracy**

- 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) $\propto 1/\text{Depth}$
 - Depth Error $\propto \text{Depth}^2$
 - Nearer the point, better the depth estimation
- An Example
 - $f = 16 \times 512/8$ pixels, $B = 0.5$ m
 - Depth error vs. depth



Absolute error

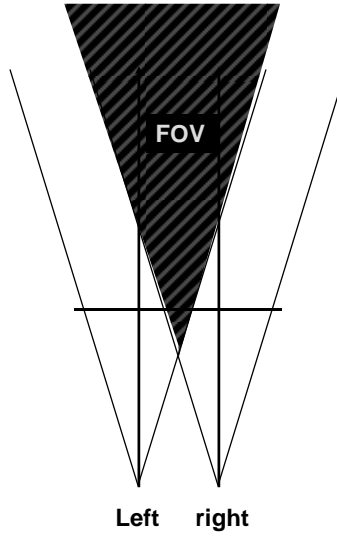
$$\delta Z = \frac{Z^2}{fB} \delta(dx)$$

Relative error

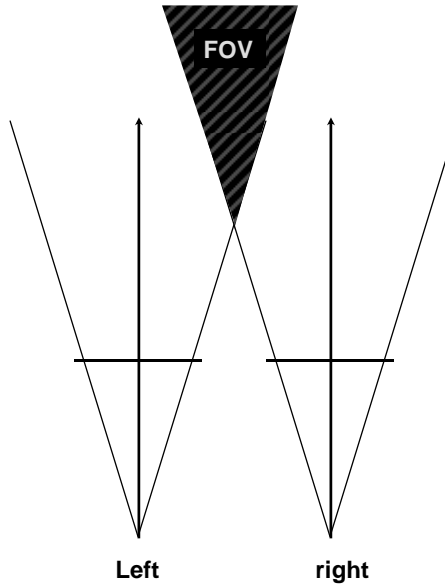
$$\frac{\delta Z}{Z} = \frac{Z}{fB} \delta(dx)$$



- 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 θ
 - The common FOV Increases

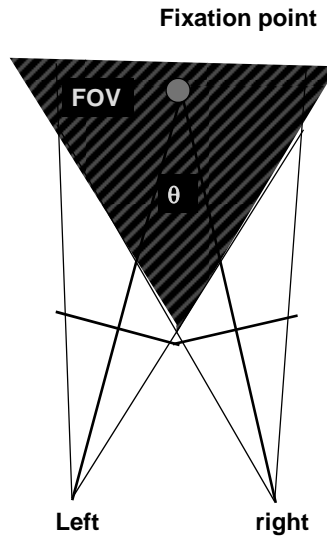


- Stereo with Parallel Axes
 - Short baseline
 - large common FOV
 - large depth error
 - Long baseline
 - small depth error
 - small common FOV
 - More occlusion problems
- Two optical axes intersect at the Fixation Point
 - converging angle θ
 - The common FOV Increases



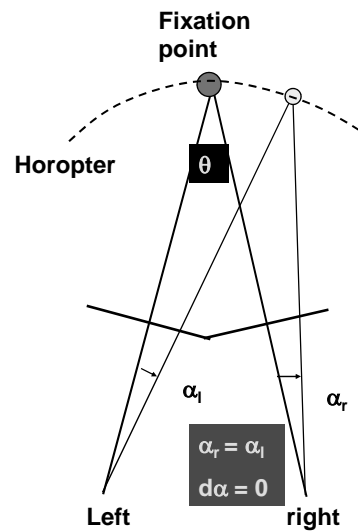
3D Computer Vision and Video Computing **Stereo with Converging Cameras**

- Two optical axes intersect at the Fixation Point
 - converging angle θ
 - 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 Depth²
 - Nearer the point, better the depth estimation



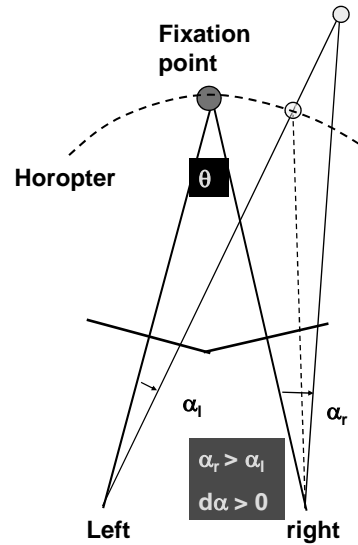
3D Computer Vision and Video Computing **Stereo with Converging Cameras**

- Two optical axes intersect at the Fixation Point
 - converging angle θ
 - 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 Depth²
 - Nearer the point, better the depth estimation



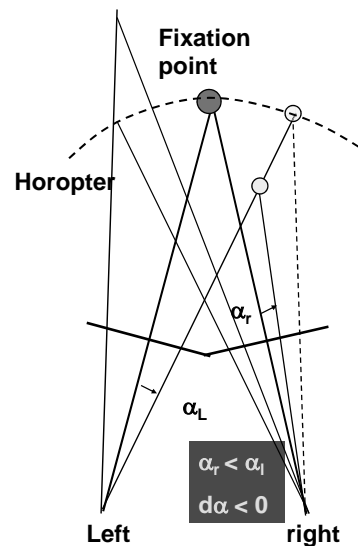
3D Computer Vision and Video Computing **Stereo with Converging Cameras**

- Two optical axes intersect at the Fixation Point
 - converging angle θ
 - 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 Depth²
 - Nearer the point, better the depth estimation

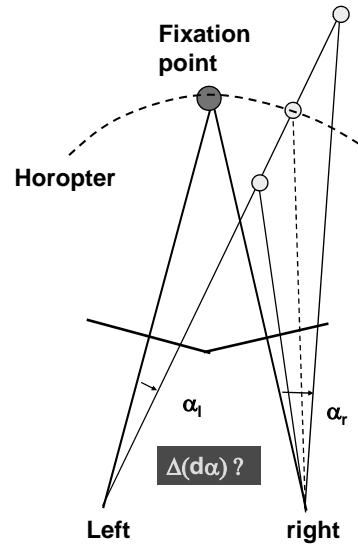


3D Computer Vision and Video Computing **Stereo with Converging Cameras**

- Two optical axes intersect at the Fixation Point
 - converging angle θ
 - 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 Depth²
 - Nearer the point, better the depth estimation

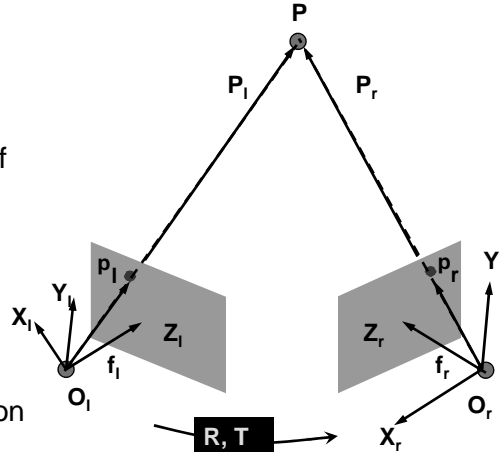


- Two optical axes intersect at the Fixation Point
 - converging angle θ
 - The common FOV Increases
- Disparity properties
 - Disparity uses angle instead of distance
 - and the Zero-disparity horopter
 - Zero disparity at fixation point
 - >0 : outside of the horopter
 - <0 : inside the horopter
 - 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 Depth²
 - Nearer the point, better the depth estimation

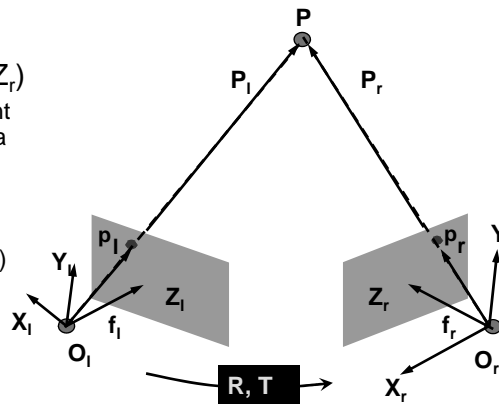


- Homework #4 online, due on May 03 before class

- Intrinsic Parameters
 - Characterize the transformation from camera to pixel coordinate systems of each camera
 - Focal length, image center, aspect ratio
- Extrinsic parameters
 - Describe the relative position and orientation of the two cameras
 - Rotation matrix R and translation vector T



- Notations
 - $P_l = (X_l, Y_l, Z_l)$, $P_r = (X_r, Y_r, Z_r)$
 - Vectors of the same 3-D point P , in the left and right camera coordinate systems respectively
 - Extrinsic Parameters
 - Translation Vector $T = (O_r - O_l)$
 - Rotation Matrix R
- $p_l = (x_l, y_l, z_l)$, $p_r = (x_r, y_r, z_r)$
 - Projections of P on the left and right image plane respectively
 - For all image points, we have $z_l = f_l$, $z_r = f_r$

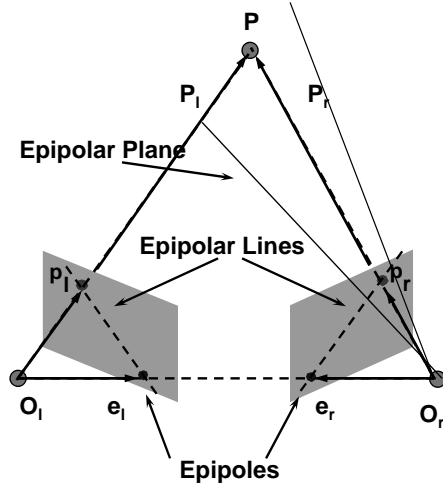


$$P_r = R(P_l - T)$$

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

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

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



- Equation of the epipolar plane
 - Co-planarity condition of vectors \$P_1\$, \$T\$ and \$P_1-T\$
- Essential Matrix \$E = RS\$
 - 3x3 matrix constructed from \$R\$ and \$T\$ (extrinsic only)
 - Rank (\$E\$) = 2, two equal nonzero singular values

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

Rank (\$R\$) = 3

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

Rank (\$S\$) = 2

$$P_r^T E P_1 = 0$$

$$P_r = R(P_1 - T)$$

$$p_i = \frac{f_i}{Z_i} P_i \quad \Bigg| \quad p_r = \frac{f_r}{Z_r} P_r$$

$$p_r^T E p_i = 0$$

- Essential Matrix $E = RS$ $\mathbf{p}_r^T E \mathbf{p}_l = 0$
 - A natural link between the stereo point pair and the extrinsic parameters of the stereo system
 - One correspondence \rightarrow a linear equation of 9 entries
 - Given 8 pairs of $(p_l, p_r) \rightarrow E$
 - Mapping between points and epipolar lines we are looking for
 - Given $p_l, E \rightarrow p_r$ on the projective line in the right plane
 - Equation represents the epipolar line of 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

- 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

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

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

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

$$\begin{array}{c} \mathbf{p}_r^T E \mathbf{p}_l = 0 \\ \mathbf{p}_l = \mathbf{M}_l^{-1} \bar{\mathbf{p}}_l \quad \mathbf{p}_r = \mathbf{M}_r^{-1} \bar{\mathbf{p}}_r \\ \downarrow \\ \bar{\mathbf{p}}_r^T \mathbf{F} \bar{\mathbf{p}}_l = 0 \\ \mathbf{F} = \mathbf{M}_r^{-T} \mathbf{E} \mathbf{M}_l^{-1} \end{array}$$

■ Fundamental Matrix

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

- Rank (F) = 2
- 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

$$\bar{\mathbf{p}}_r^T \mathbf{F} \bar{\mathbf{p}}_l = 0 \Rightarrow \begin{pmatrix} x_{im}^{(r)} & y_{im}^{(r)} & 1 \end{pmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{pmatrix} x_{im}^{(l)} \\ y_{im}^{(l)} \\ 1 \end{pmatrix} = 0$$

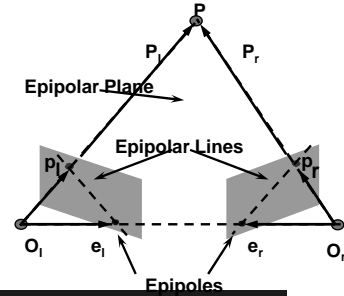
■ Input: n point correspondences (n >= 8)

- Construct homogeneous system $\mathbf{Ax} = 0$ from $\bar{\mathbf{p}}_r^T \mathbf{F} \bar{\mathbf{p}}_l = 0$
 - $\mathbf{x} = (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})$: entries in F
 - Each correspondence give one equation
 - A is a nx9 matrix
- Obtain estimate \mathbf{F}^\wedge by SVD of A $\mathbf{A} = \mathbf{UDV}^T$
 - 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 \mathbf{F}^\wedge $\hat{\mathbf{F}} = \mathbf{UDV}^T$
 - Set the smallest singular value to 0: $\mathbf{D} \rightarrow \mathbf{D}'$
 - Correct estimate of F : $\mathbf{F}' = \mathbf{UD}'\mathbf{V}^T$

■ Output: the estimate of the fundamental matrix, F'

■ Similarly we can compute E given intrinsic parameters

$\bar{p}_r^T F \bar{p}_l = 0$ e_l lies on all the epipolar lines of the left image
 \Downarrow
 $\bar{p}_r^T F \bar{e}_l = 0$ For every p_r
 \Downarrow F is not identically zero
 $F \bar{e}_l = 0$

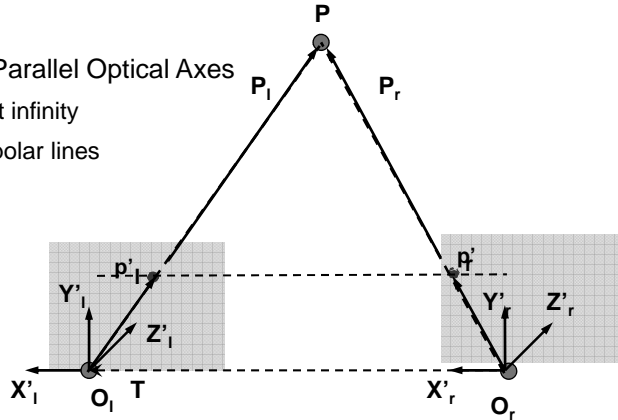


- Input: Fundamental Matrix F
 - Find the SVD of F $F = UDV^T$
 - The epipole e_l is the column of V corresponding to the null singular value (as shown above)
 - The epipole e_r is the column of U corresponding to the null singular value
- Output: Epipole e_l and e_r

- Homework #4 online, due on May 03 before class

■ Stereo System with Parallel Optical Axes

- Epipoles are at infinity
- Horizontal epipolar lines

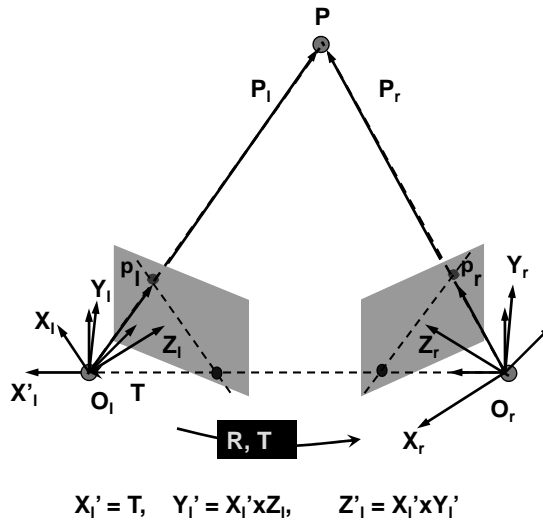


■ Rectification

- Given a stereo pair, the intrinsic and extrinsic parameters, find the image transformation to achieve a stereo system of horizontal epipolar lines
- A simple algorithm: Assuming calibrated stereo cameras

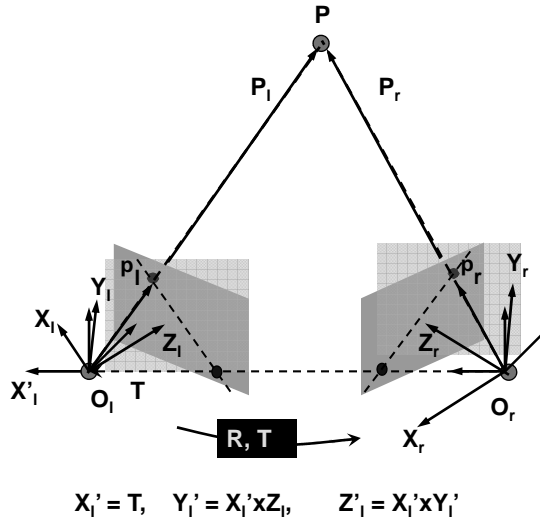
■ Algorithm

- Rotate both left and right camera so that they share the same X axis : $O_r - O_l = T$
- Define a rotation matrix R_{rect} for the left camera
- Rotation Matrix for the right camera is $R_{rect}R^T$
- Rotation can be implemented by image transformation



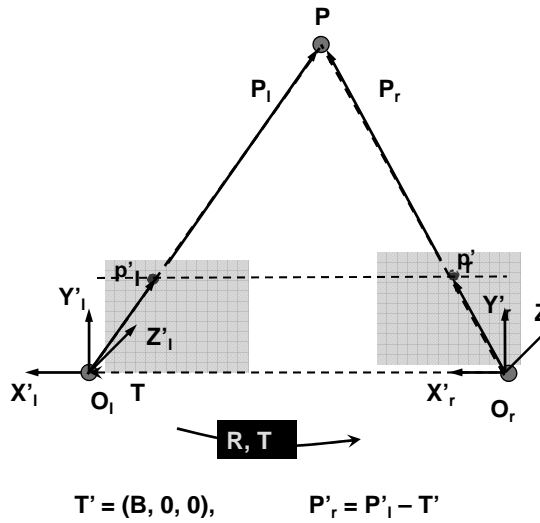
Algorithm

- Rotate both left and right camera so that they share the same X axis : $O_r - O_l = T$
- Define a rotation matrix R_{rect} for the left camera
- Rotation Matrix for the right camera is $R_{rect}R^T$
- Rotation can be implemented by image transformation



Algorithm

- Rotate both left and right camera so that they share the same X axis : $O_r - O_l = T$
- Define a rotation matrix R_{rect} for the left camera
- Rotation Matrix for the right camera is $R_{rect}R^T$
- Rotation can be implemented by image transformation



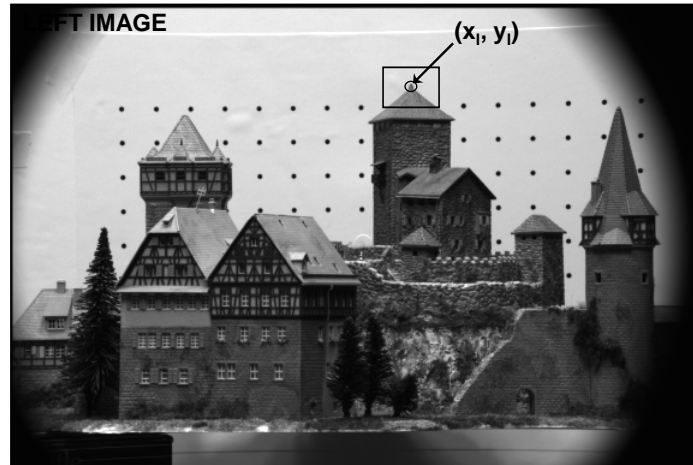
- Purpose
 - where to search correspondences
- Epipolar plane, epipolar lines, and epipoles
 - known intrinsic (f) and extrinsic (R, T)
 - co-planarity equation
 - known intrinsic but unknown extrinsic
 - essential matrix
 - unknown intrinsic and extrinsic
 - fundamental matrix
- Rectification
 - Generate stereo pair (by software) with parallel optical axis and thus horizontal epipolar lines

$$\mathbf{P}_r^T \mathbf{R}^T \mathbf{T} \times \mathbf{P}_l = 0$$

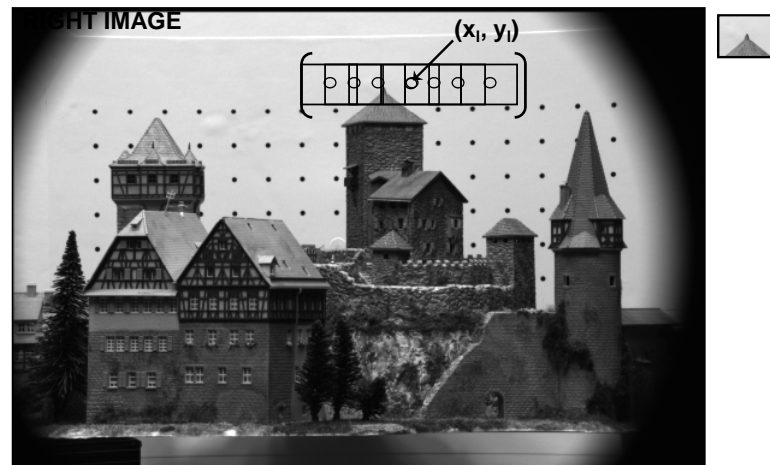
$$\mathbf{p}_r^T \mathbf{E} \mathbf{p}_l = 0$$

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

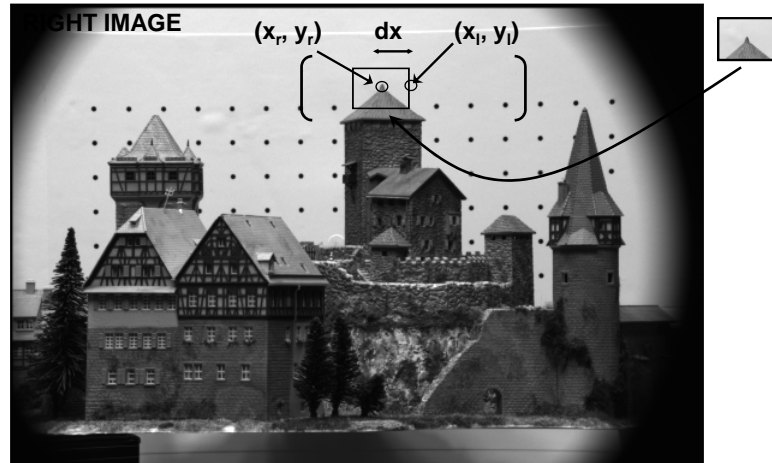
- Three Questions
 - What to match?
 - Features: point, line, area, structure?
 - Where to search correspondence?
 - Epipolar line?
 - How to measure similarity?
 - Depends on features
- Approaches
 - Correlation-based approach
 - Feature-based approach
- Advanced Topics
 - Image filtering to handle 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



- For Each point (x_i, y_i) in the left image, define a window centered at the point



- ... search its corresponding point within a search region in the right image



- ... the disparity (dx , dy) is the displacement when the correlation is maximum



- 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



■ 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 $\bar{\mathbf{d}} = (\bar{d}_x, \bar{d}_y) = \arg \max_{\mathbf{d} \in R} \{c(dx, dy)\}$

■ Similarity criterion

• Cross-Correlation $\Psi(u, v) = uv$

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

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



■ PROS

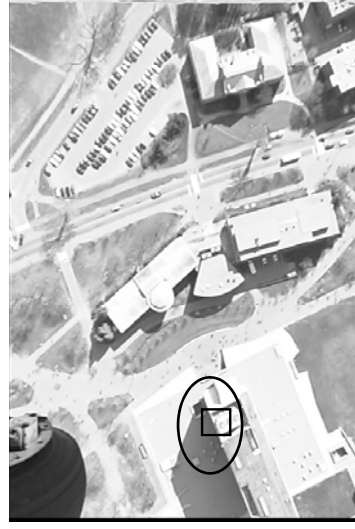
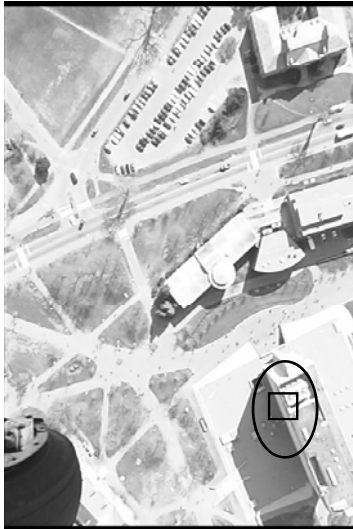
- Easy to implement
- Produces dense disparity map
- Maybe slow

■ CONS

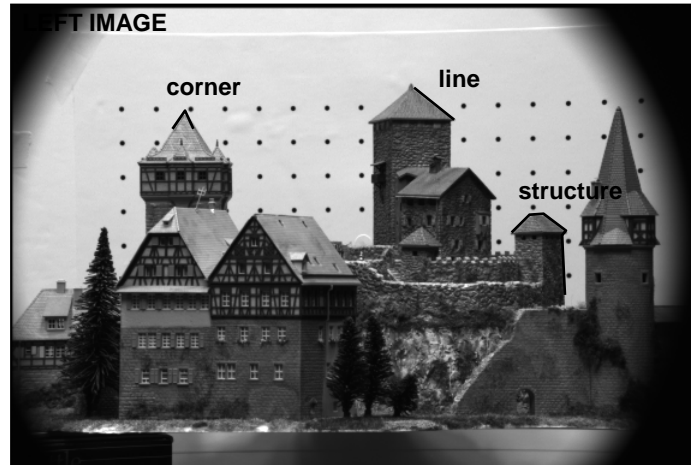
- Needs textured images to work well
- Inadequate for matching image pairs from very different viewpoints due to illumination changes
- Window may cover points with quite different disparities
- Inaccurate disparities on the occluding boundaries



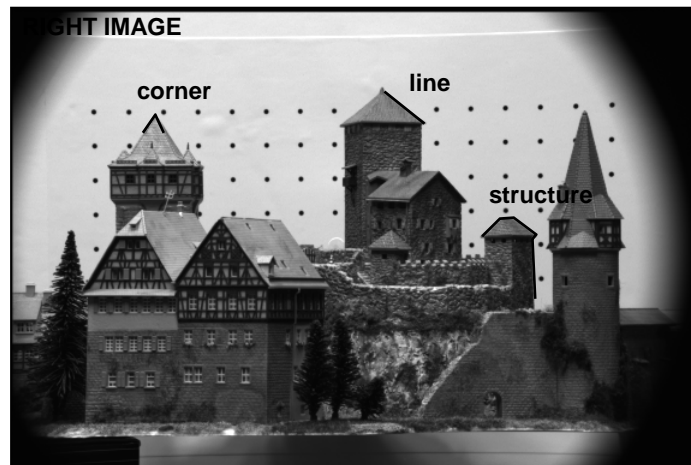
- A Stereo Pair of UMass Campus – texture, boundaries and occlusion



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



- 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?



- Homework #4 online, due on May 03 before class



- 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

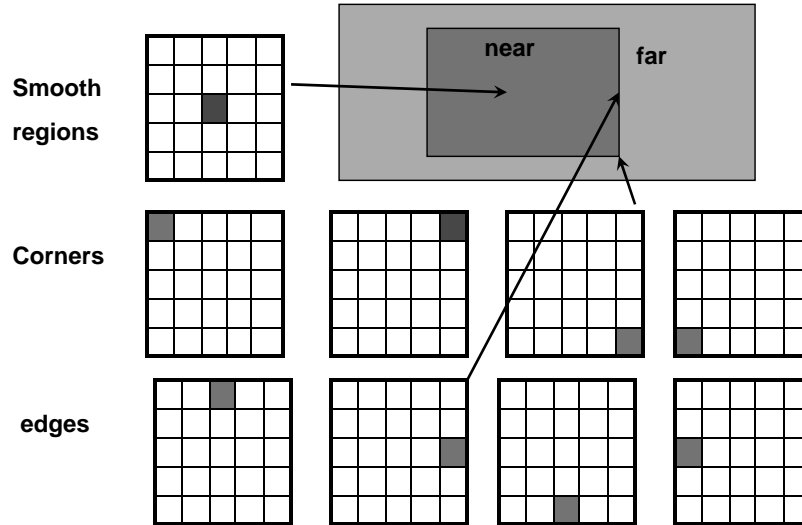
 - Laplacian filtering (2nd order derivative)
 - Use derivative rather than intensity (or original color)



- Adaptive windows to deal with multiple disparities
 - Adaptive Window Approach (Kanade and Okutomi)
 - statistically adaptive technique which selects at each pixel the window size that minimizes the uncertainty in disparity estimates
 - A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment, *T. Kanade and M. Okutomi, Proc. 1991 IEEE International Conference on Robotics and Automation*, Vol. 2, April, 1991, pp. 1088-1095

 - Multiple window algorithm (Fusiello, et al)
 - Use 9 windows instead of just one to compute the SSD measure
 - The point with the smallest SSD error amongst the 9 windows and various search locations is chosen as the best estimate for the given points
 - A Fusiello, V. Roberto and E. Trucco, Efficient stereo with multiple windowing, IEEE CVPR pp858-863, 1997

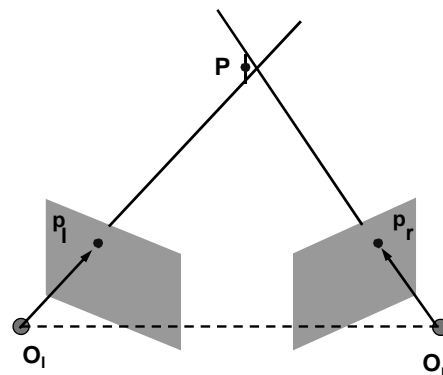
■ Multiple windows to deal with multiple disparities



- Sub-pixel matching to improve accuracy
 - Find the peak in the correlation curves
- Self-consistency to reduce false matches esp. for occlusions
 - Check the consistency of matches from L to R and from R to L
- Multiple Resolution Approach
 - From coarse to fine for efficiency in searching correspondences
- Local warping to account for perspective distortion
 - Warp from one view to the other for a small patch given an initial estimation of the (planar) surface normal
- Multi-baseline Stereo
 - Improves both correspondences and 3D estimation by using more than two cameras (images)

- What we have done
 - Correspondences using either correlation or feature based approaches
 - Epipolar Geometry from at least 8 point correspondences
- Three cases of 3D reconstruction depending on the amount of a priori knowledge on the stereo system
 - Both intrinsic and extrinsic known -> can solve the reconstruction problem unambiguously by triangulation
 - Only intrinsic known -> recovery structure and extrinsic up to an unknown scaling factor
 - Only correspondences -> reconstruction only up to an unknown, global projective transformation (*)

- Assumption and Problem
 - Under the assumption that both intrinsic and extrinsic parameters are known
 - Compute the 3-D location from their projections, p_l and p_r
- Solution
 - Triangulation: Two rays are known and the intersection can be computed
 - Problem: Two rays will not actually intersect in space due to errors in calibration and correspondences, and pixelization
 - Solution: find a point in space with minimum distance from both rays



- Assumption and Problem Statement
 - Under the assumption that only intrinsic parameters and more than 8 point correspondences are given
 - Compute the 3-D location from their projections, p_l and p_r , as well as the extrinsic parameters
- Solution
 - Compute the essential matrix E from at least 8 correspondences
 - Estimate T (up to a scale and a sign) from $E (=RS)$ using the orthogonal constraint of R , and then R
 - End up with four different estimates of the pair (T, R)
 - Reconstruct the depth of each point, and pick up the correct sign of R and T .
 - Results: reconstructed 3D points (up to a common scale);
 - The scale can be determined if distance of two points (in space) are known

(* not required for this course; needs advanced knowledge of projective geometry)

- Assumption and Problem Statement
 - Under the assumption that only n (≥ 8) point correspondences are given
 - Compute the 3-D location from their projections, p_l and p_r
- Solution
 - Compute the Fundamental matrix F from at least 8 correspondences, and the two epipoles
 - Determine the projection matrices
 - Select five points (from correspondence pairs) as the projective basis
 - Compute the projective reconstruction
 - Unique up to the unknown projective transformation fixed by the choice of the five points



- 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 (*)



- Understanding 3D structure and events from motion

Motion

- Homework #4 online, due on May 03 before class