



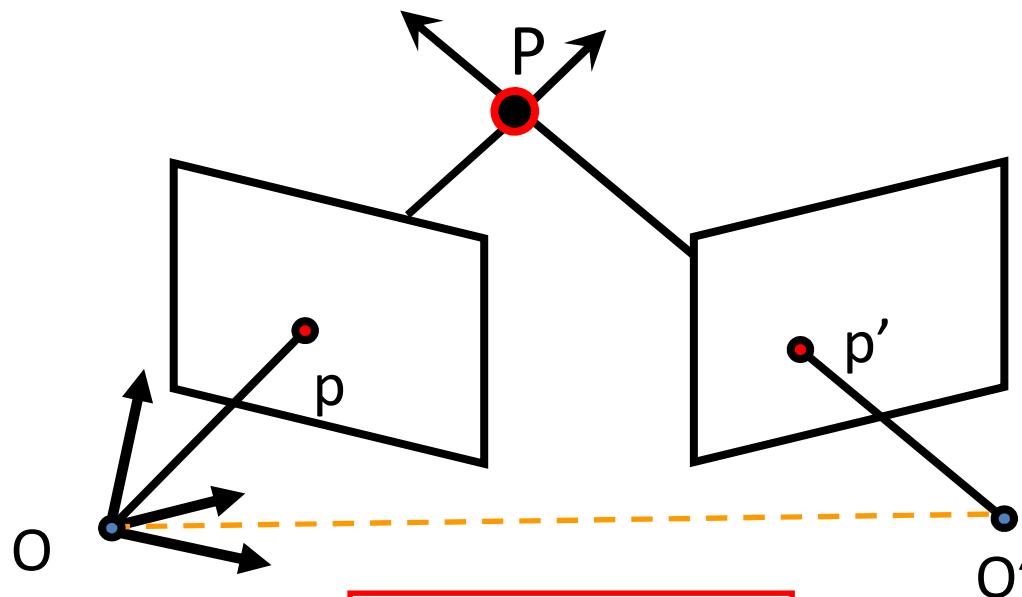
Lecture 10: Multi-view geometry

Professor Fei-Fei Li

Stanford Vision Lab

First, a reminder of the Fundamental Matrix: The Fundamental Matrix Song

<http://danielwedge.com/fmatrix/>



$$p^T F p' = 0$$

F = Fundamental Matrix
(Faugeras and Luong, 1992)

What we will learn today?

- Rectification for stereo vision
- Correspondence problem (**Problem Set 2 (Q3)**)
- Active stereo vision systems
- Structure from motion

Reading:

[HZ] Chapters: 4, 9, 11
[FP] Chapters: 10

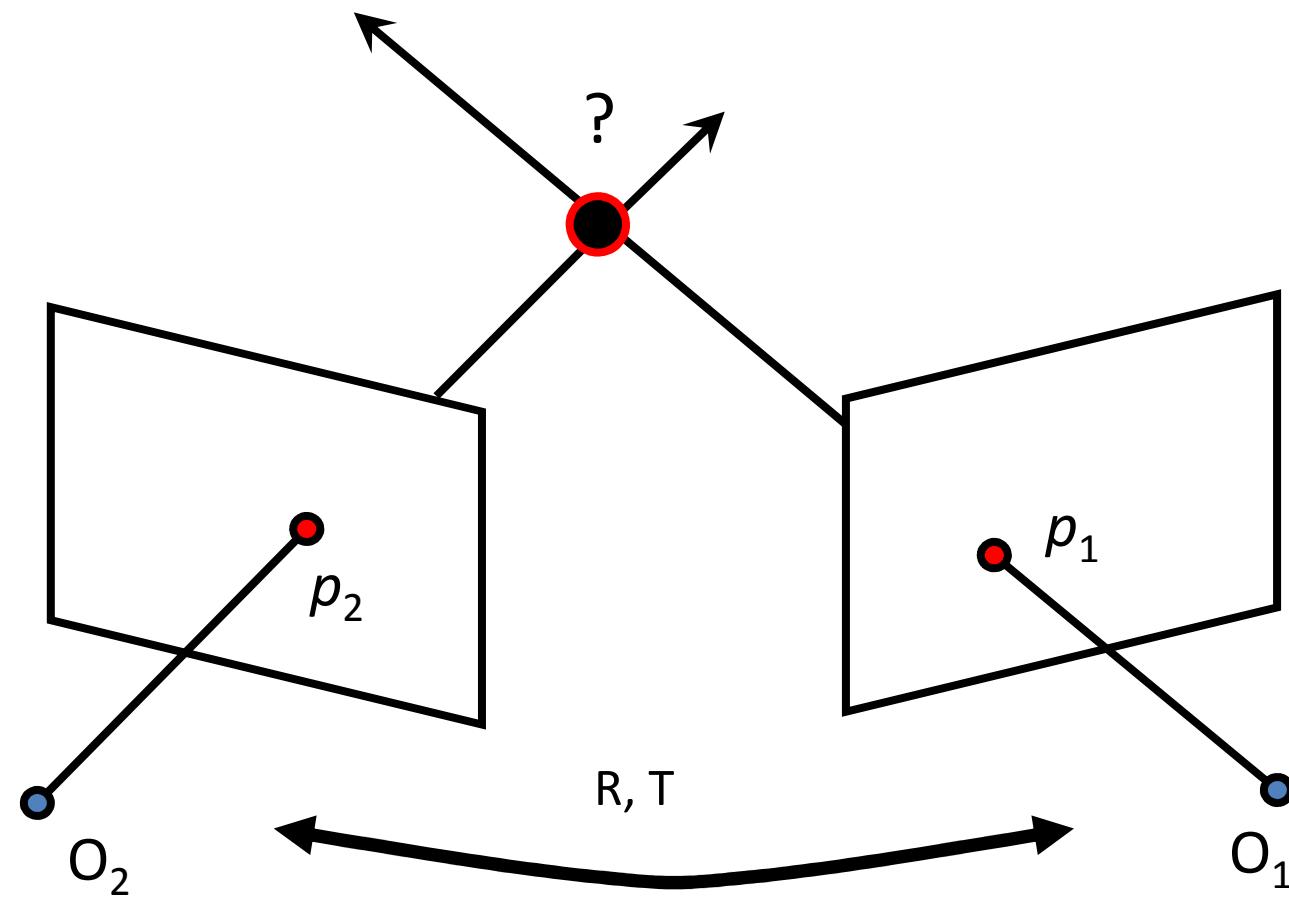
What we will learn today?

- Rectification for stereo vision
- Correspondence problem (Problem Set 2 (Q3))
- Active stereo vision systems
- Structure from motion

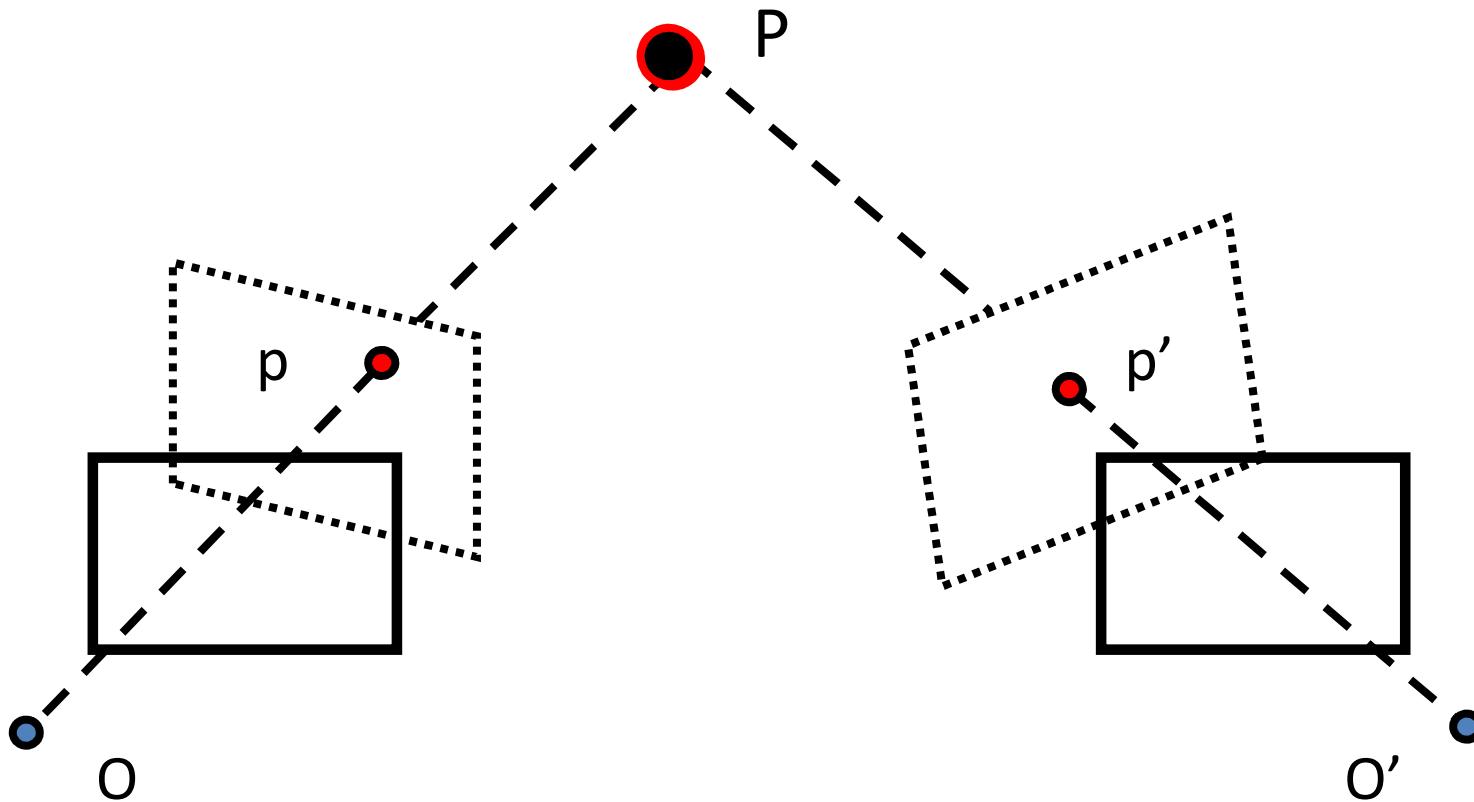
Reading:

[HZ] Chapters: 4, 9, 11
[FP] Chapters: 10

Two eyes help!

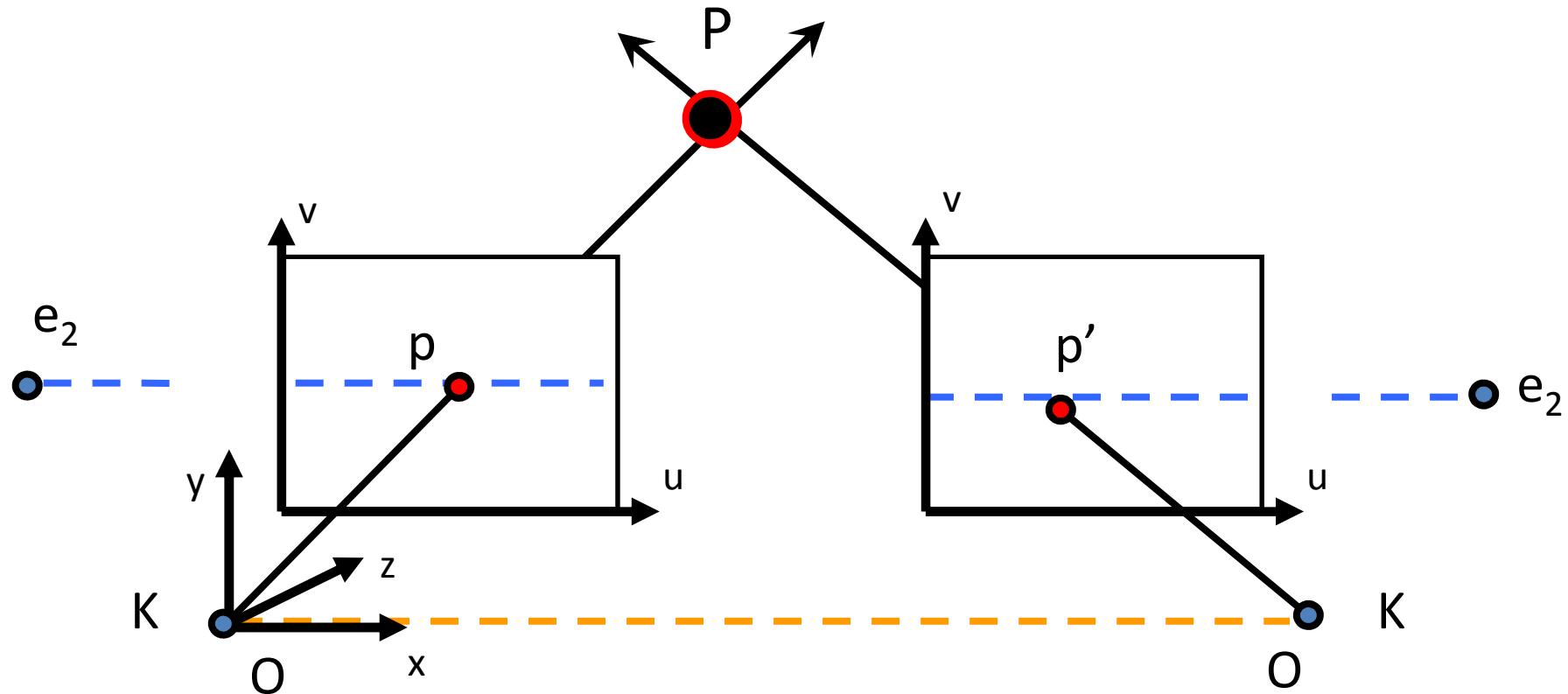


Rectification



- Make two camera images “parallel”
 - Correspondence problem becomes easier

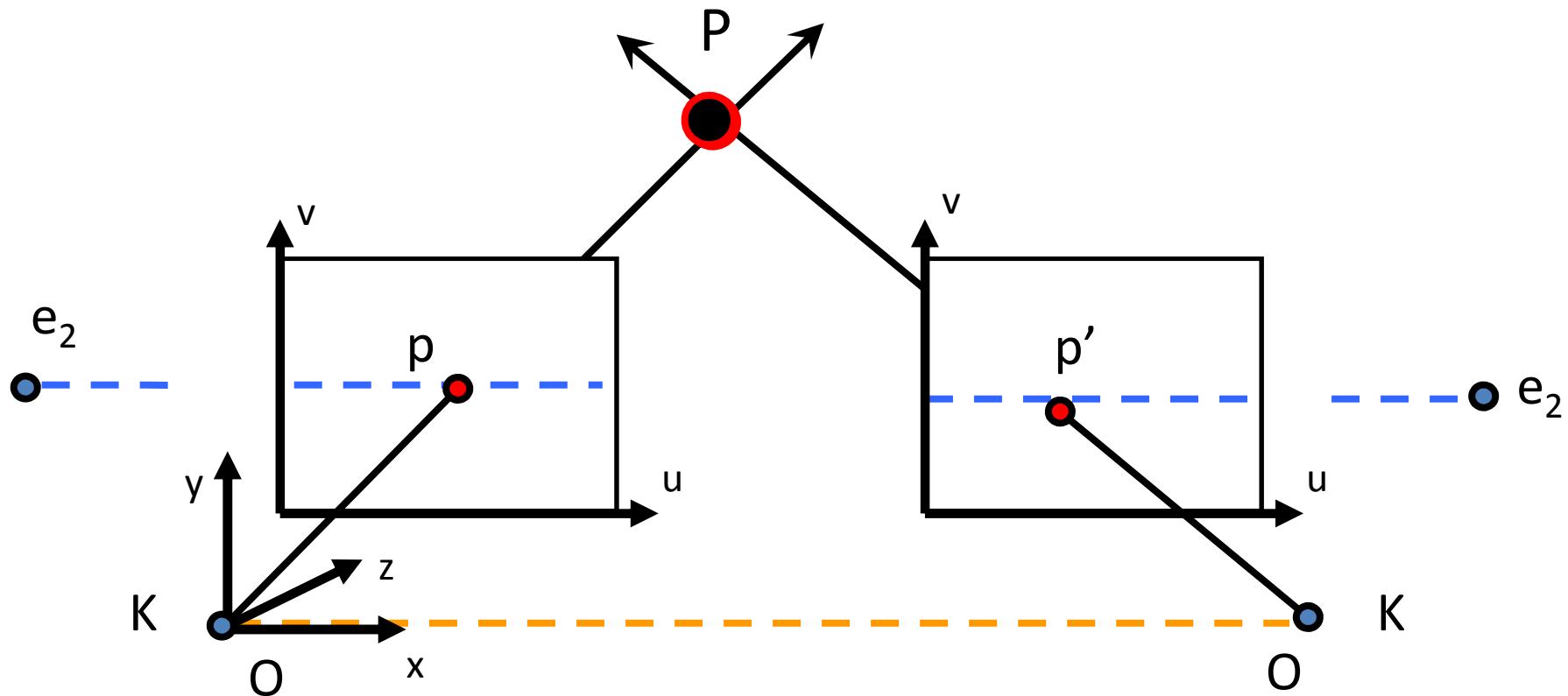
Rectification



- Parallel epipolar lines
- Epipoles at infinity
- $v = v'$

Let's see why....

Rectification



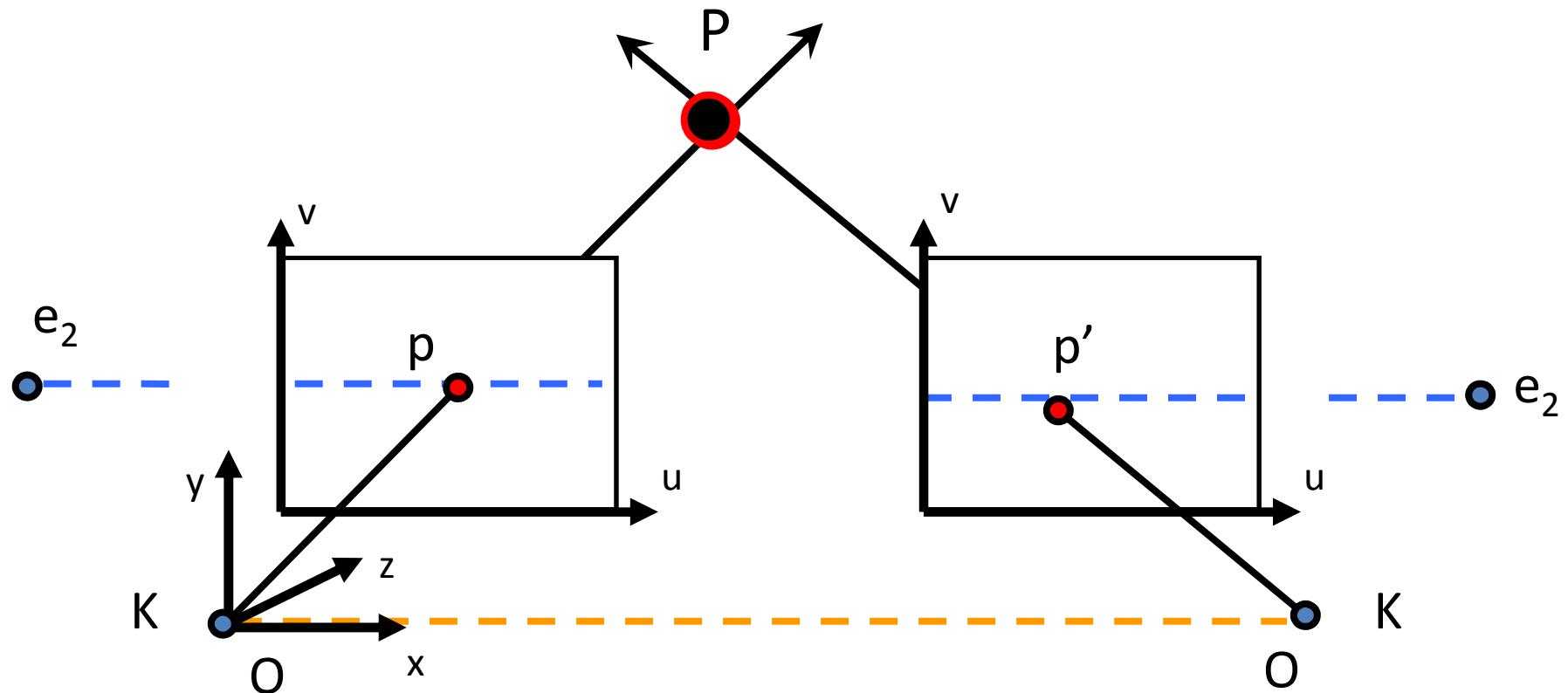
$K_1 = K_2 = \text{known}$
x parallel to O_1O_2

$$E = [t_x]R$$

Cross product as matrix multiplication

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = [\mathbf{a}_\times] \mathbf{b}$$

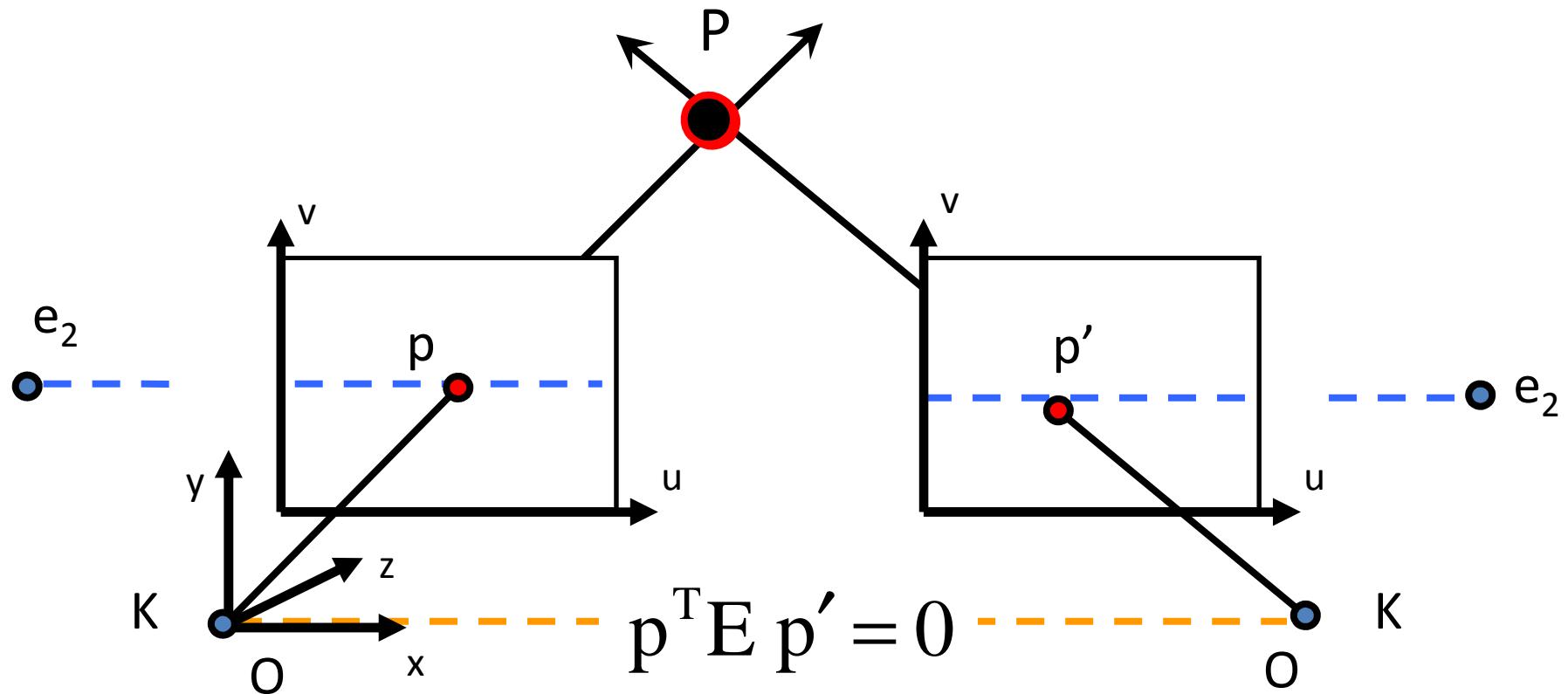
Rectification



$K_1 = K_2 = \text{known}$
 $x \text{ parallel to } O_1 O_2$

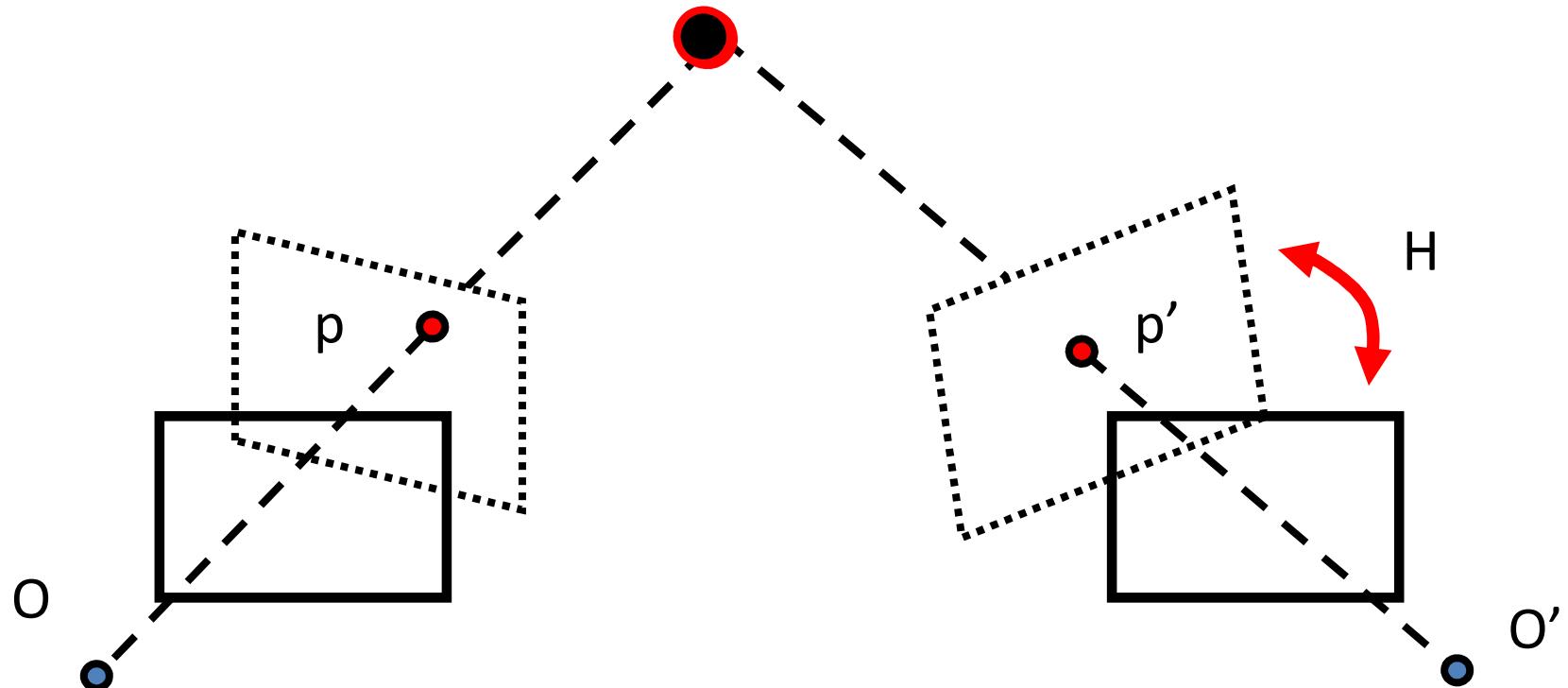
$$E = [t_x]R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \rightarrow v = v'?$$

Rectification



$$(u \quad v \quad 1) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0 \quad (u \quad v \quad 1) \begin{pmatrix} 0 \\ -T \\ Tv' \end{pmatrix} = 0 \quad Tv = Tv' \rightarrow v = v'$$

Rectification



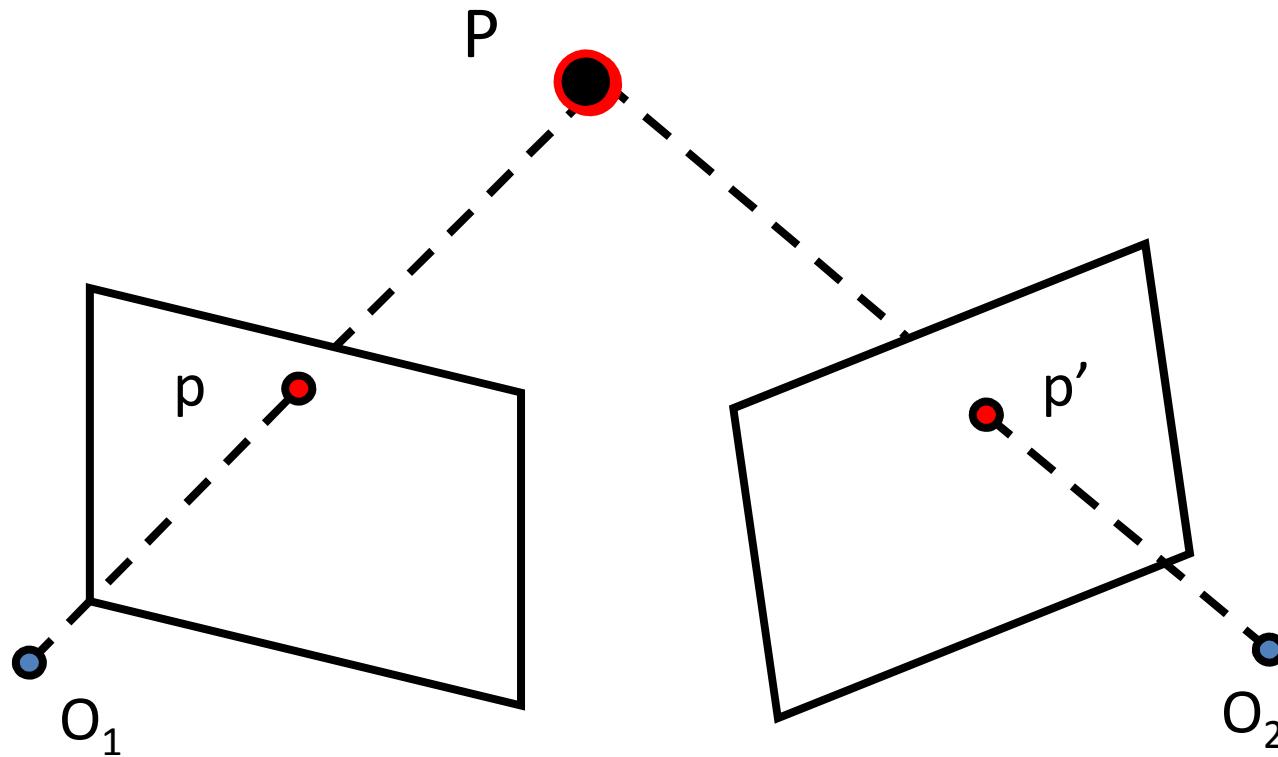
GOAL of rectification : Estimate a perspective transformation H that makes images parallel

Impose $v' = v$

- This leaves degrees of freedom for determining H
- If not appropriate H is chosen, severe projective distortions on image take place
- We impose a number of restriction while computing H

[HZ] Chapters: 11 (sec. 11.12)

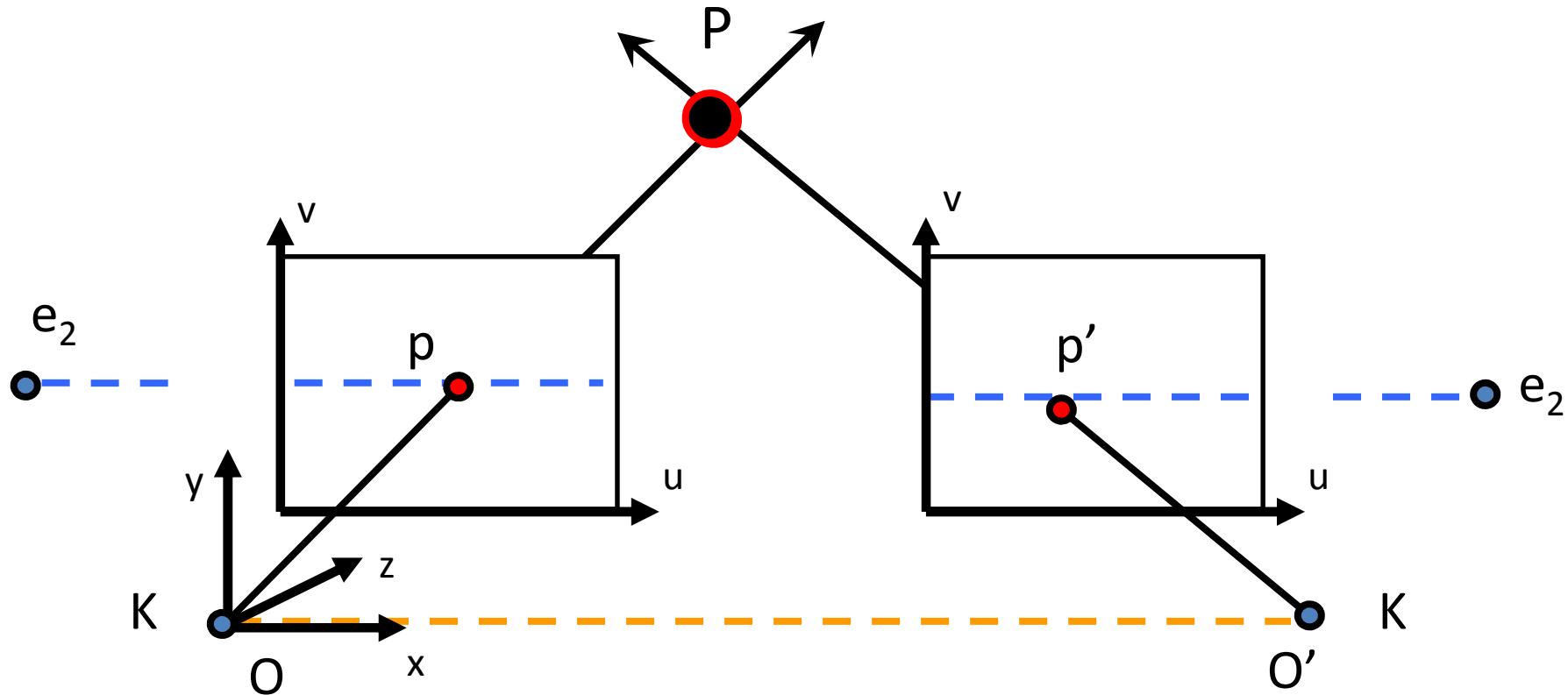
Stereo-view geometry



Subgoals:

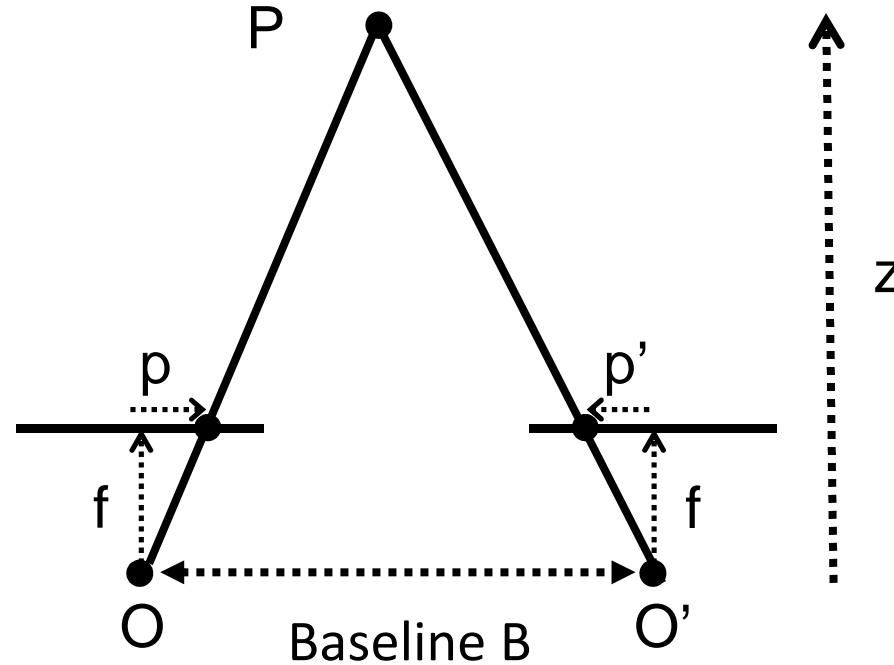
- Solve the correspondence problem
- Use corresponding observations to triangulate (depth estimation)

Depth estimation



- Assume cameras are calibrated and rectified
- Assume correspondence problem is solved
- **Goal: estimate depth (triangulate)**

Depth estimation

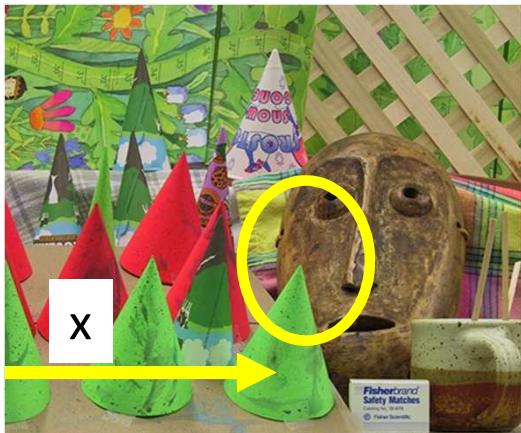


$$p - p' = \frac{B \cdot f}{z} = \text{disparity}$$

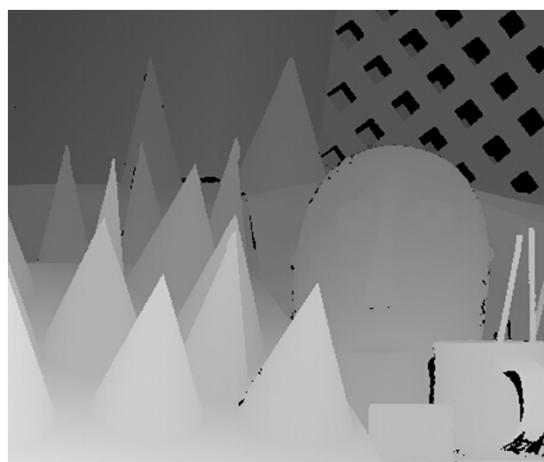
Note: Disparity is inversely proportional to depth

Depth estimation

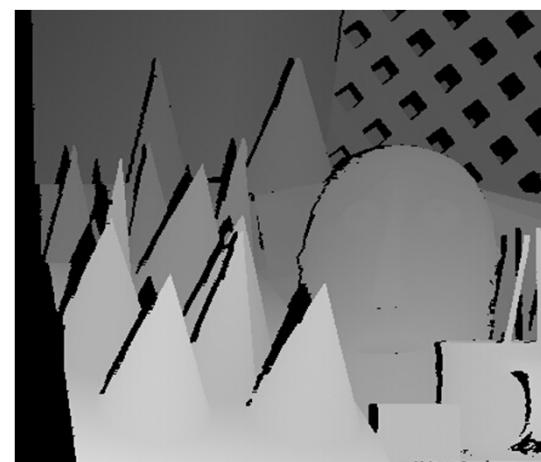
Stereo pair



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



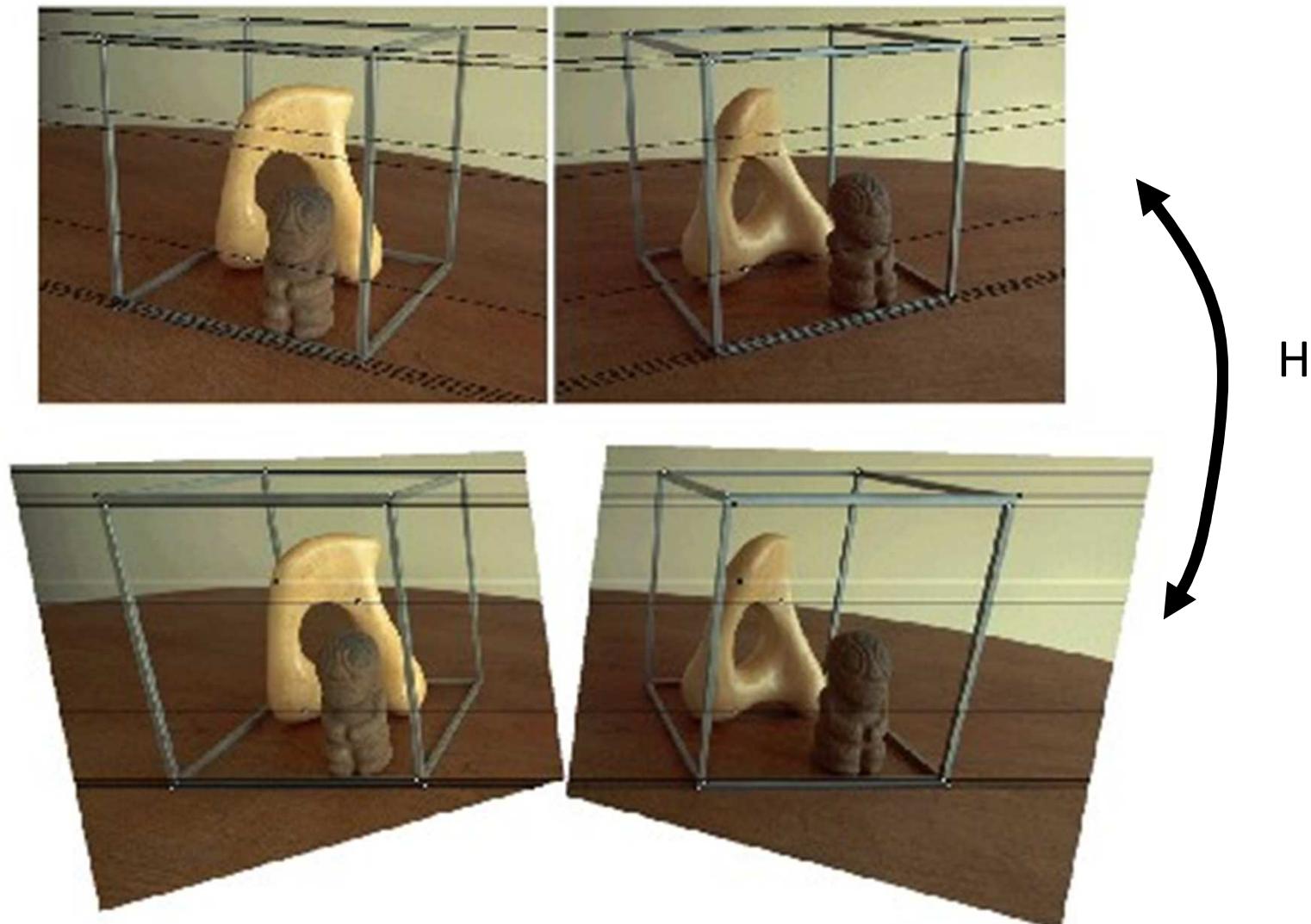
Disparity map / depth map



Disparity map with occlusions

<http://vision.middlebury.edu/stereo/>

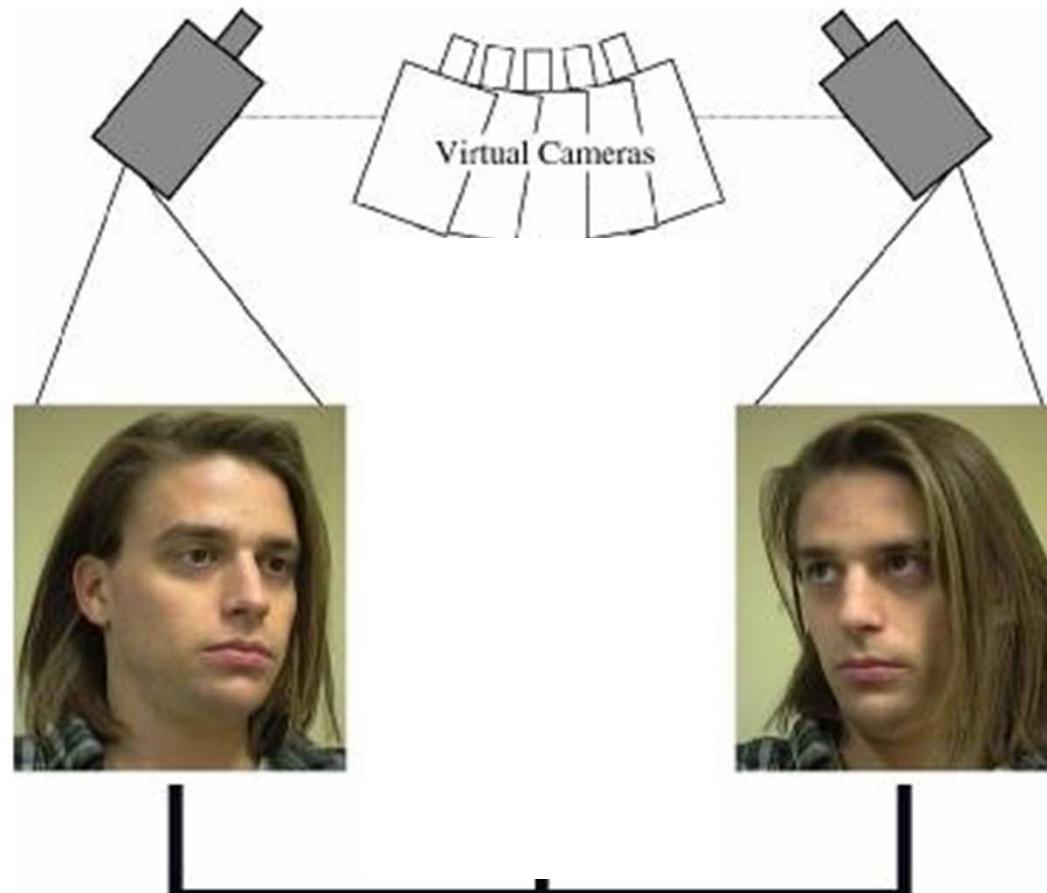
Rectification



Courtesy figure S. Lazebnik

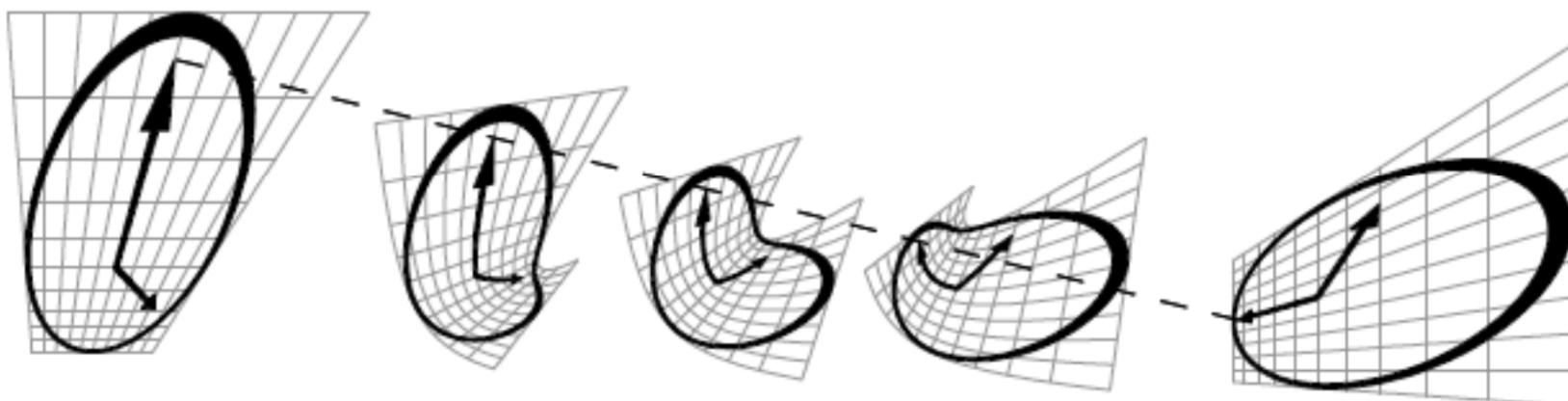
Application: view morphing

S. M. Seitz and C. R. Dyer, *Proc. SIGGRAPH 96*, 1996, 21-30

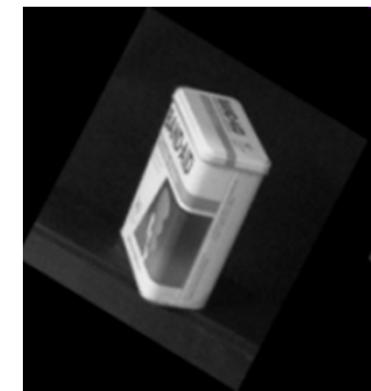


Application: view morphing

If rectification is not applied, the morphing procedure does not generate geometrically correct interpolations



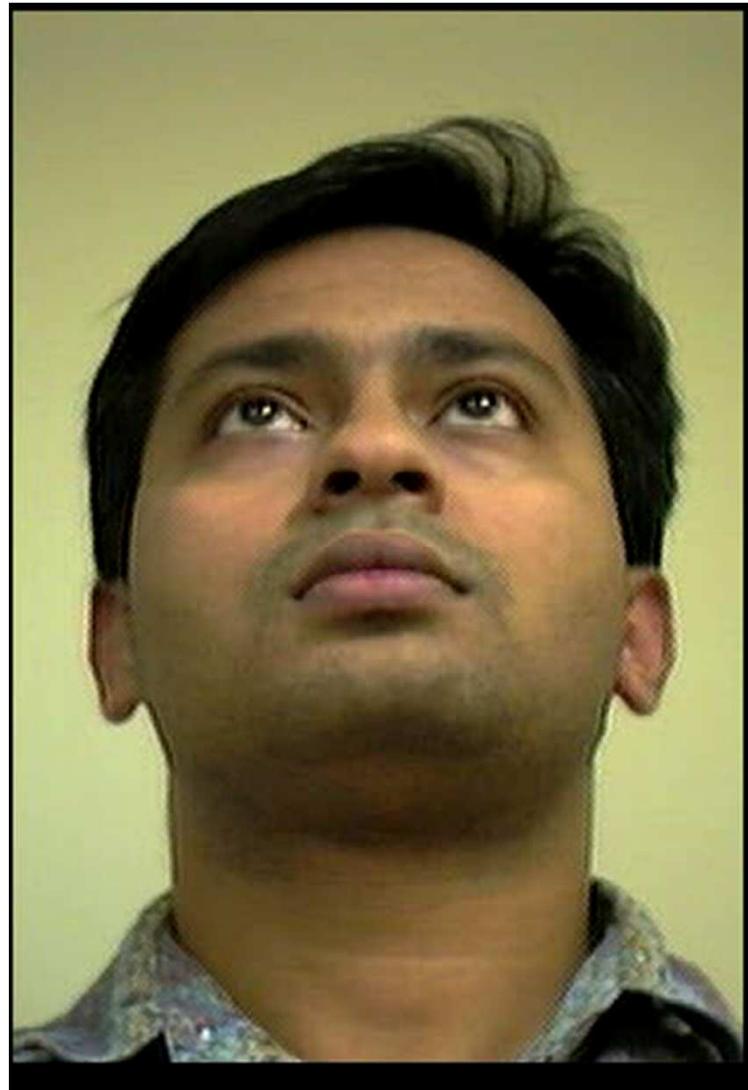
Application: view morphing



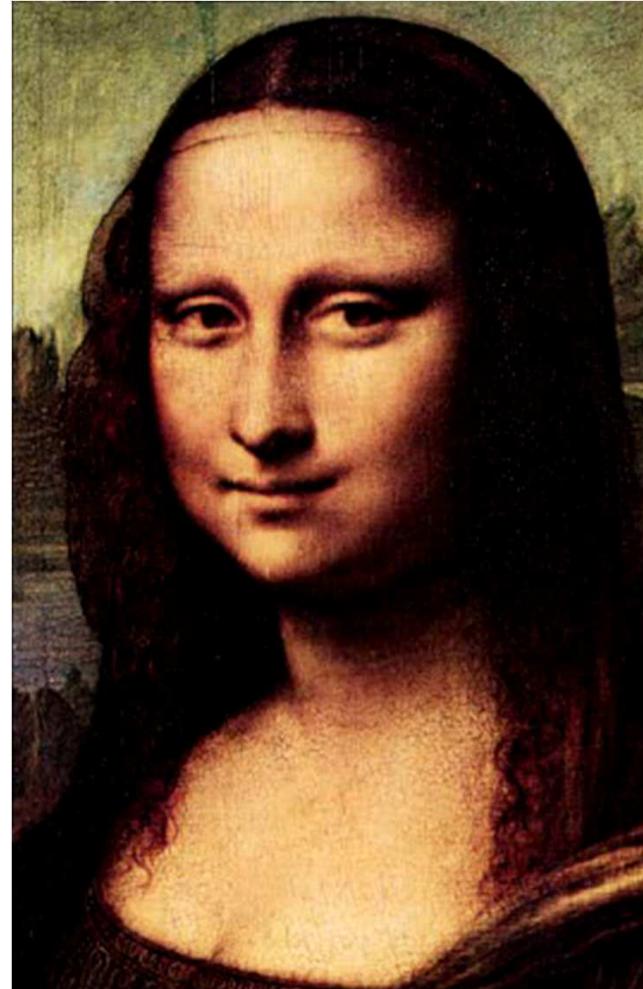
Application: view morphing



Application: view morphing



Application: view morphing



What we will learn today?

- Rectification for stereo vision
- Correspondence problem (**Problem Set 2 (Q3)**)
- Active stereo vision systems
- Structure from motion

Reading:

[HZ] Chapters: 4, 9, 11
[FP] Chapters: 10

Stereo-view geometry

- **Correspondence:** Given a point in one image, how can I find the corresponding point x' in another one?
- **Camera geometry:** Given corresponding points in two images, find camera matrices, position and pose.
- **Scene geometry:** Find coordinates of 3D point from its projection into 2 or multiple images.

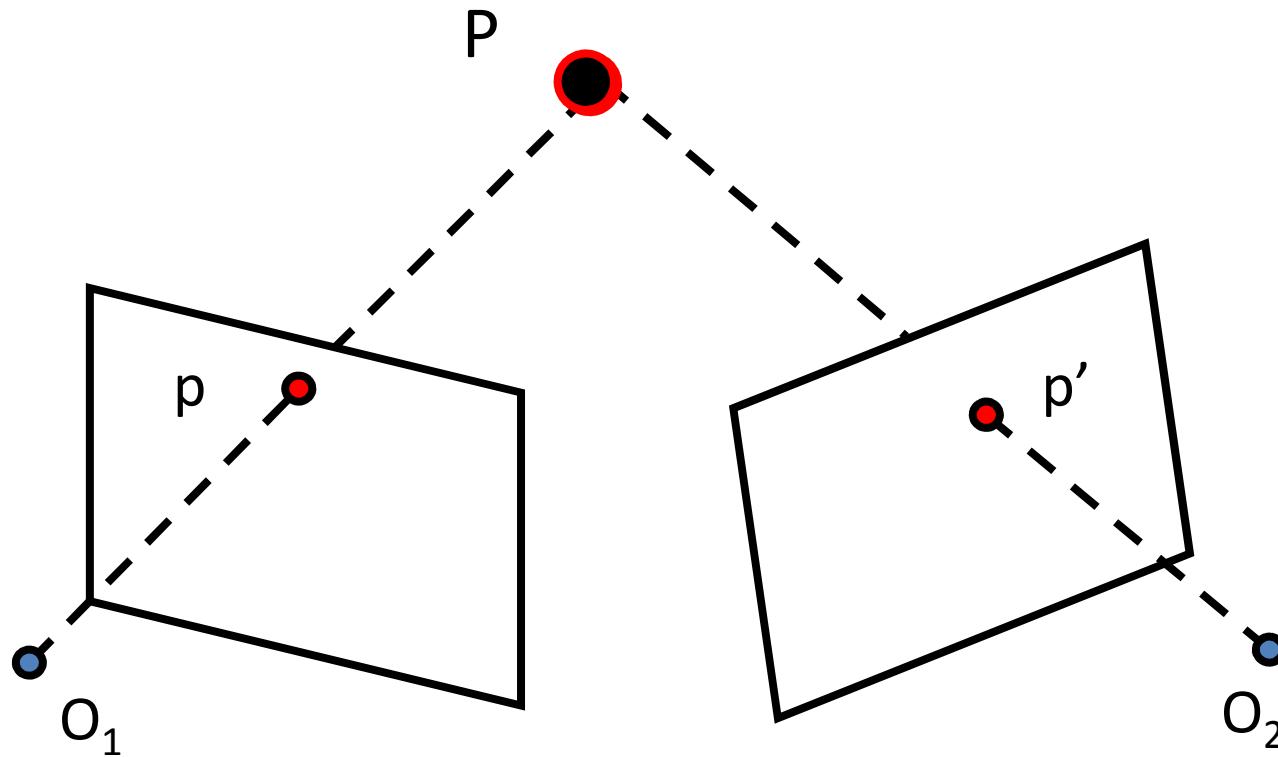
Previous lecture (#9)

Stereo-view geometry

This lecture (#10)

- **Correspondence:** Given a point in one image, how can I find the corresponding point x' in another one?
- **Camera geometry:** Given corresponding points in two images, find camera matrices, position and pose.
- **Scene geometry:** Find coordinates of 3D point from its projection into 2 or multiple images.

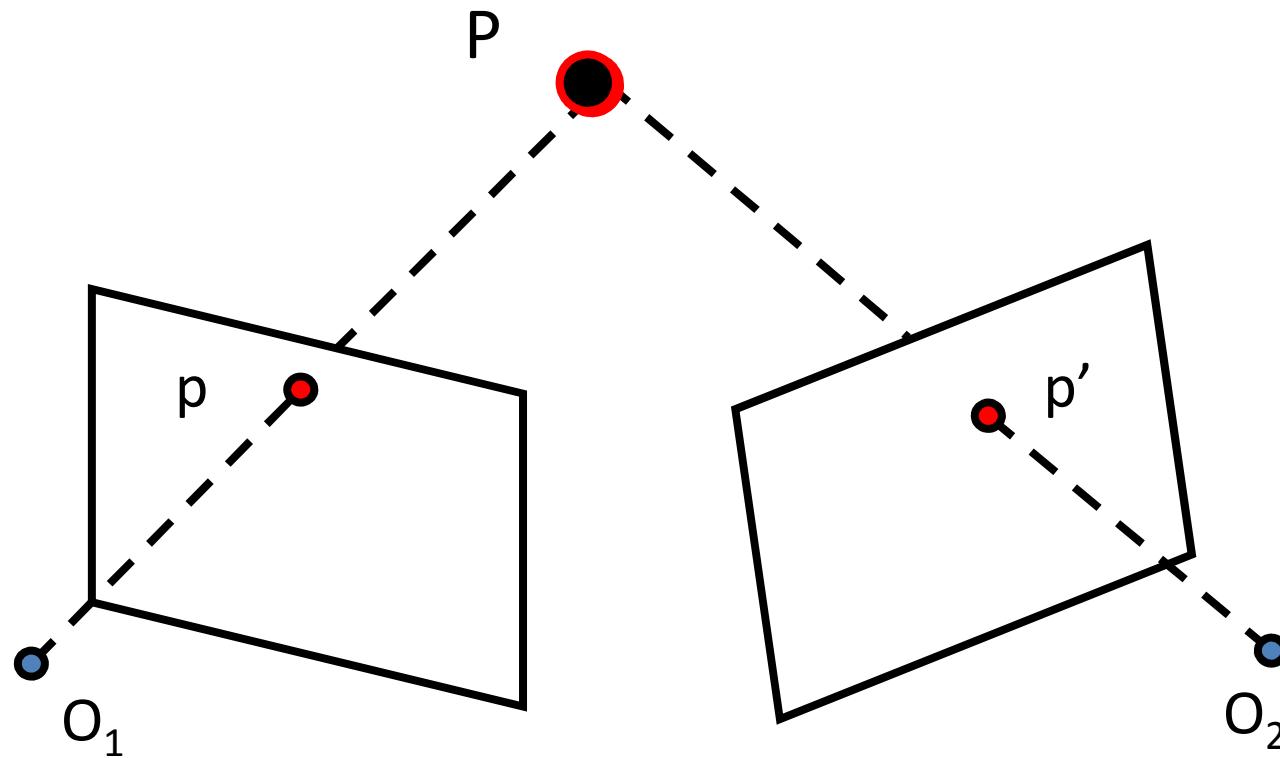
Depth estimation



Subgoals:

- Solve the correspondence problem
- Use corresponding observations to triangulate

Correspondence problem



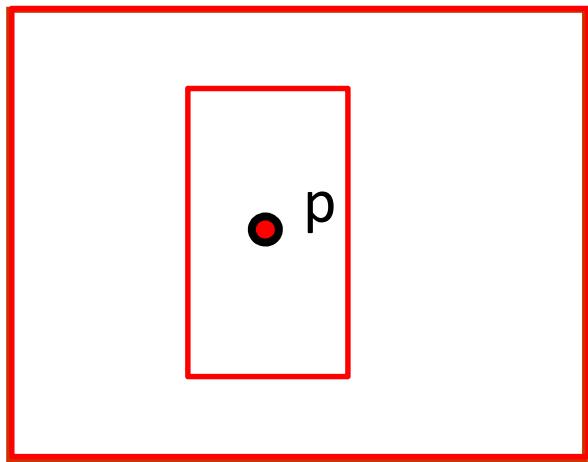
Given a point in 3D, discover corresponding observations
in left and right images [also called binocular fusion problem]

Correspondence problem

- A Cooperative Model (Marr and Poggio, 1976)
- Correlation Methods (1970--)
- Multi-Scale Edge Matching (Marr, Poggio and Grimson, 1979-81)

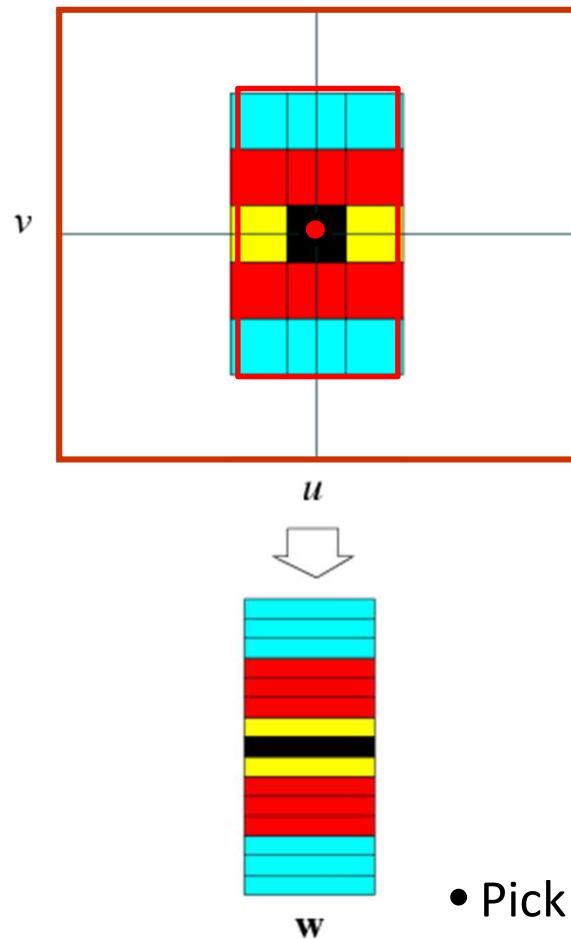
[FP] Chapters: 11

Correlation Methods



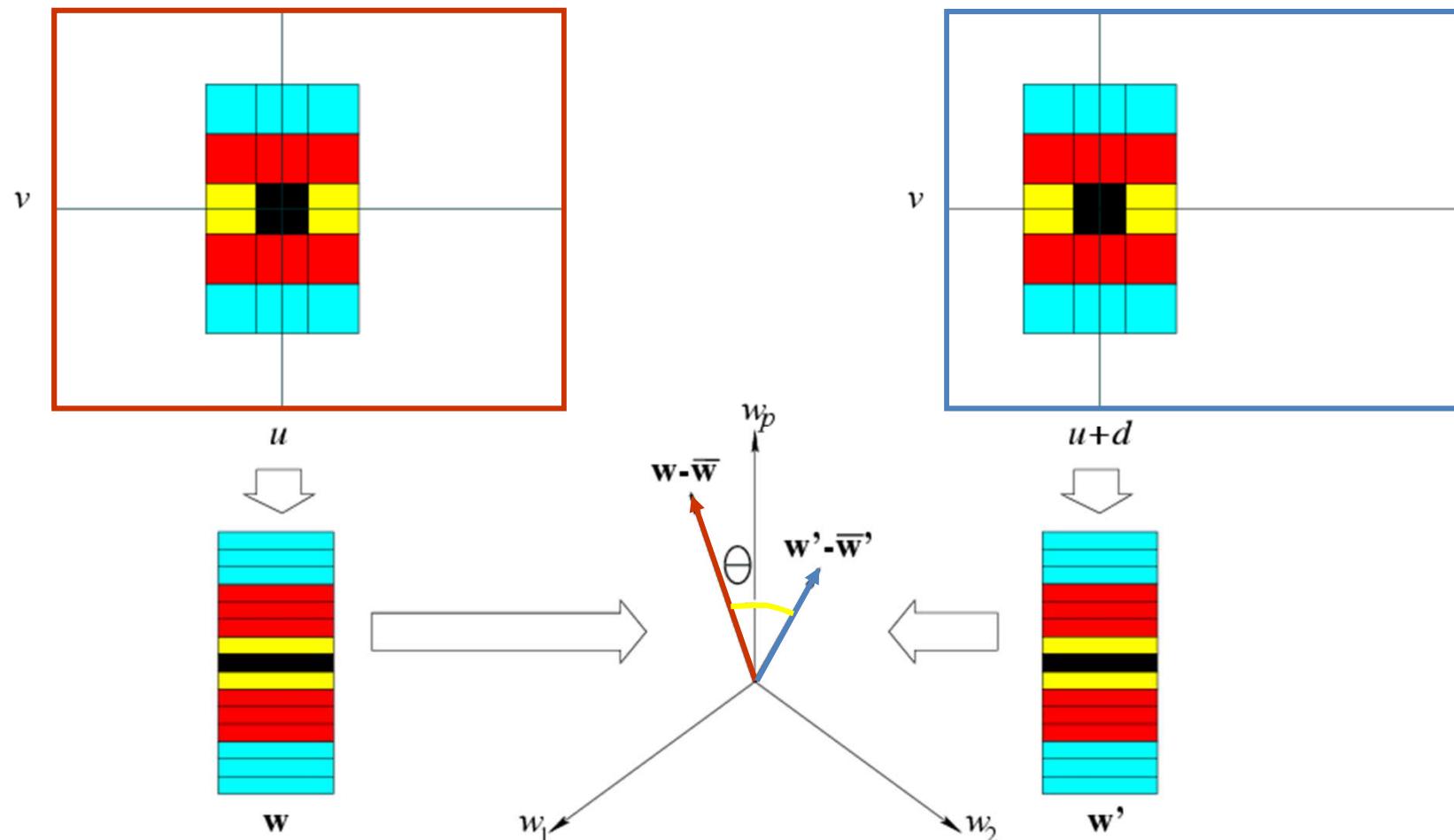
- Pick up a window around $p(u,v)$

Correlation Methods



- Pick up a window around $p(u,v)$
- Build vector W
- Slide the window along v line in image 2 and compute w'
- Keep sliding until $w \cdot w'$ is maximized.

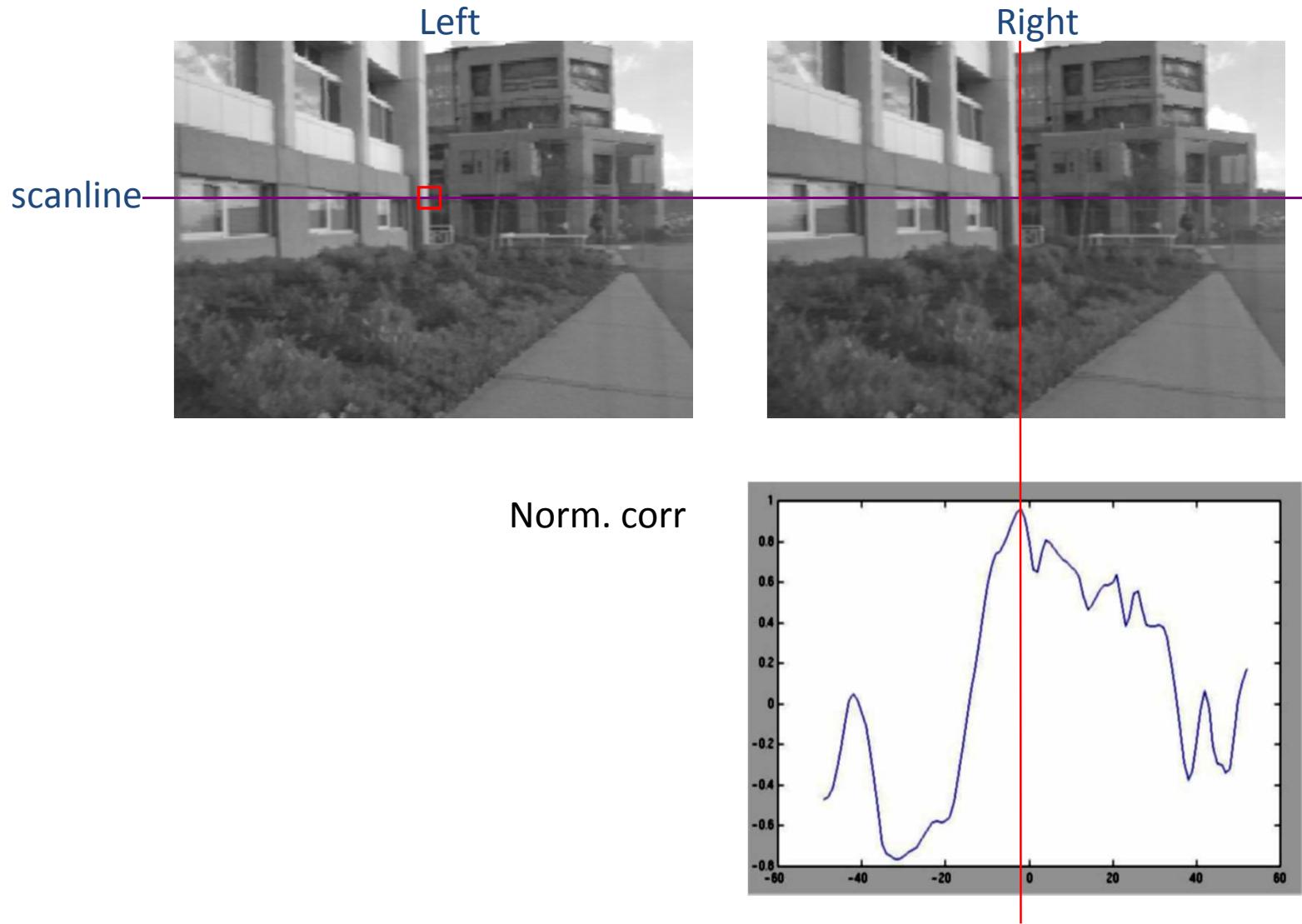
Correlation Methods



Normalized Correlation; minimize:

$$\frac{(w - \bar{w})(w' - \bar{w}')}{\|(w - \bar{w})(w' - \bar{w}')\|}$$

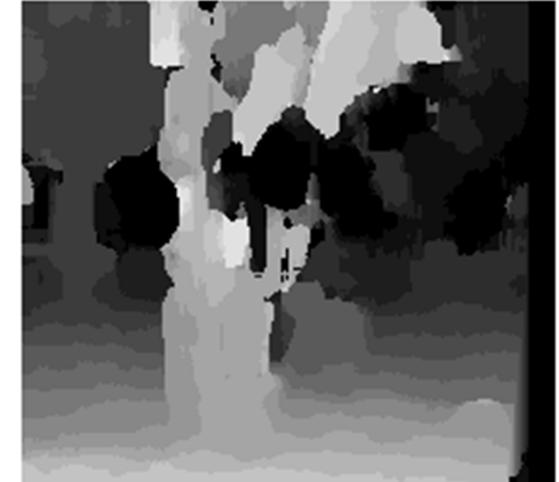
Correlation Methods



Correlation Methods



Window size = 3



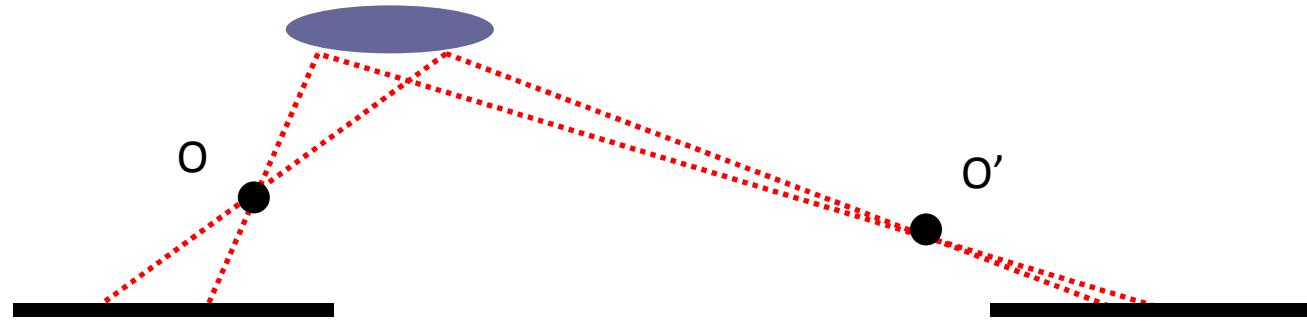
Window size = 20

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

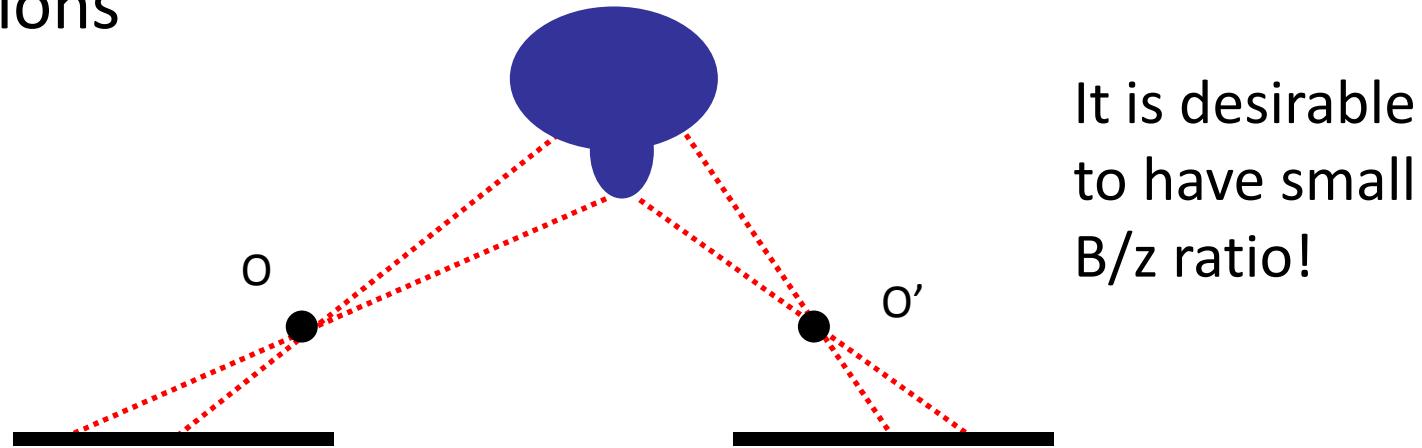
Credit slide S. Lazebnik

Issues with Correlation Methods

- Fore shortening effect

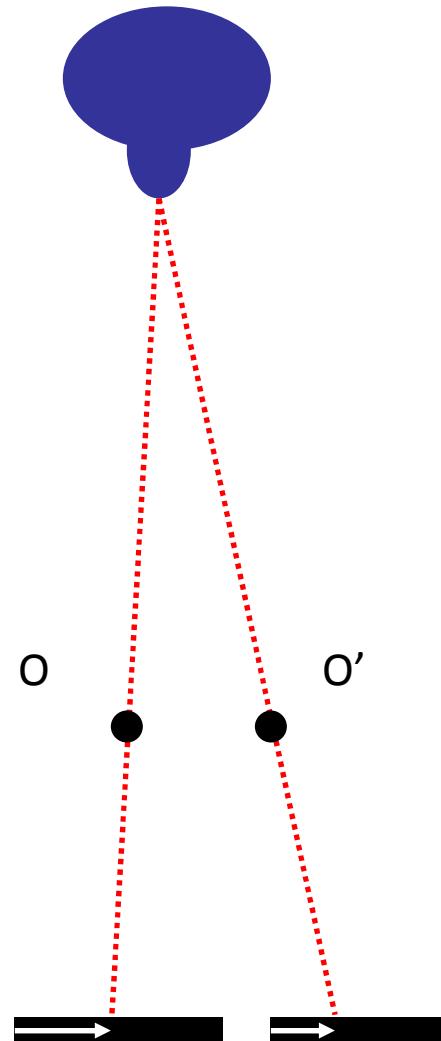


- Occlusions



Issues with Correlation Methods

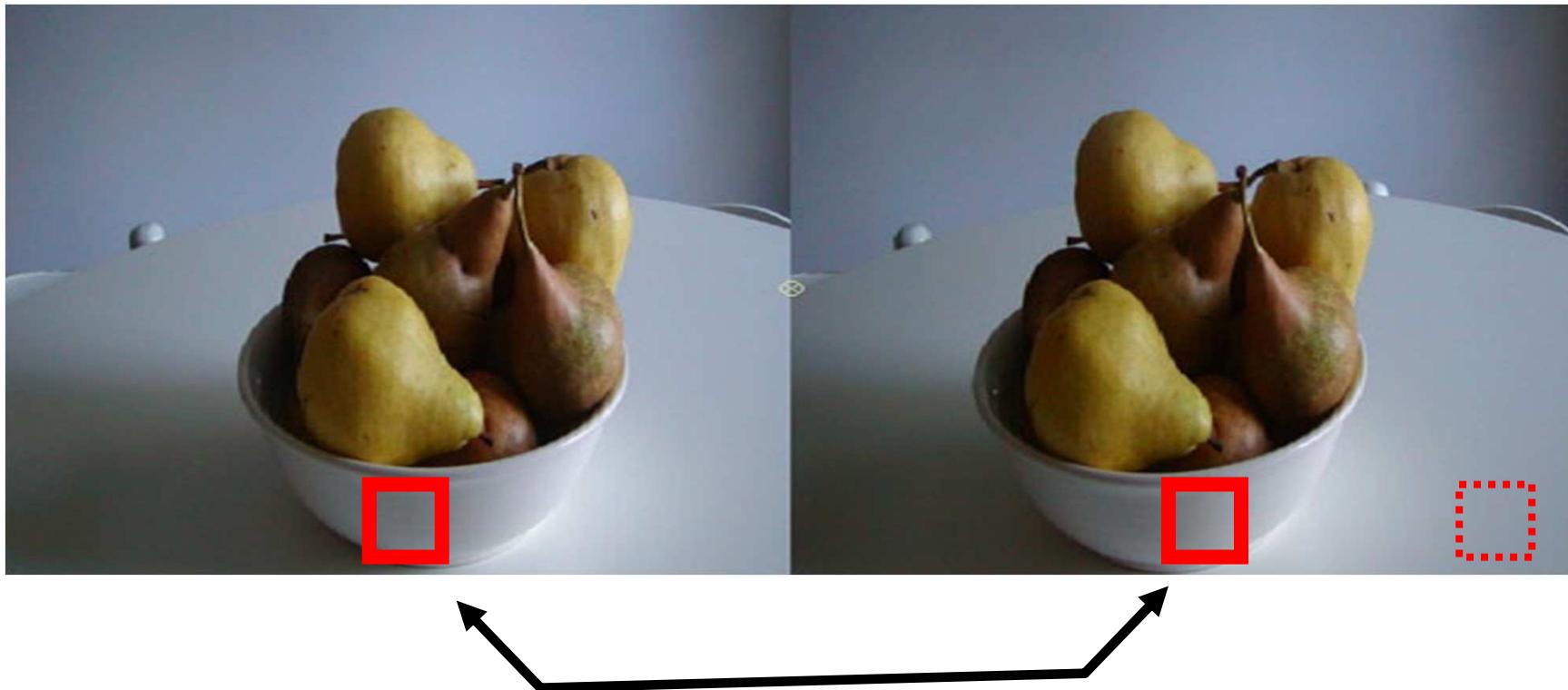
- small B/z ratio



Small error in measurements implies large error in estimating depth

Issues with Correlation Methods

- Homogeneous regions



Hard to match pixels in these regions

Issues with Correlation Methods

- Repetitive patterns



Results with window search

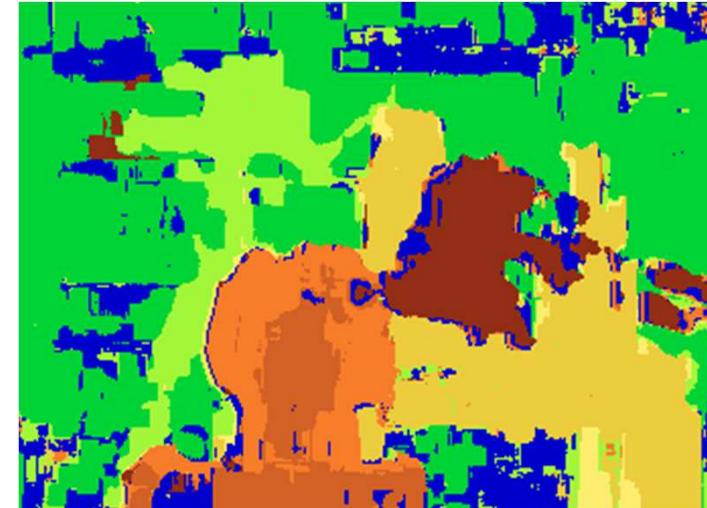
Data



Ground truth



Window-based matching



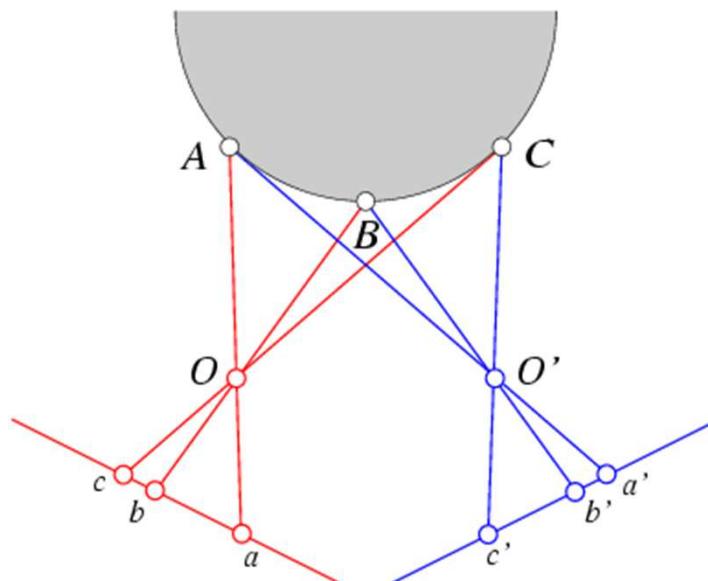
Credit slide S. Lazebnik

Improving correspondence: Non-local constraints

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

Improving correspondence: Non-local constraints

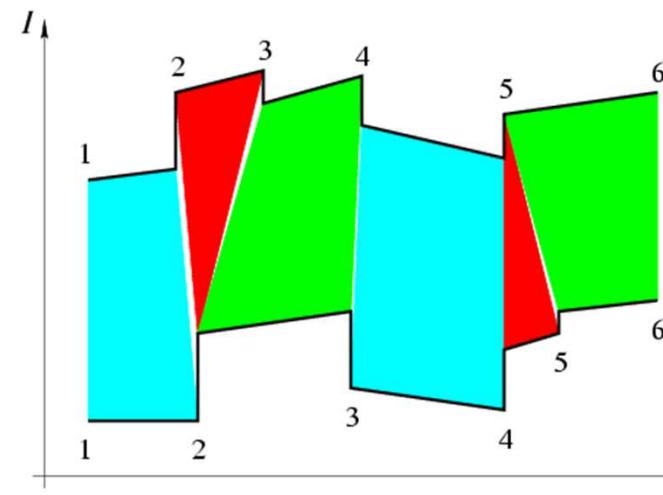
- Uniqueness
 - For any point in one image, there should be at most one matching point in the other image
- Ordering
 - Corresponding points should be in the same order in both views



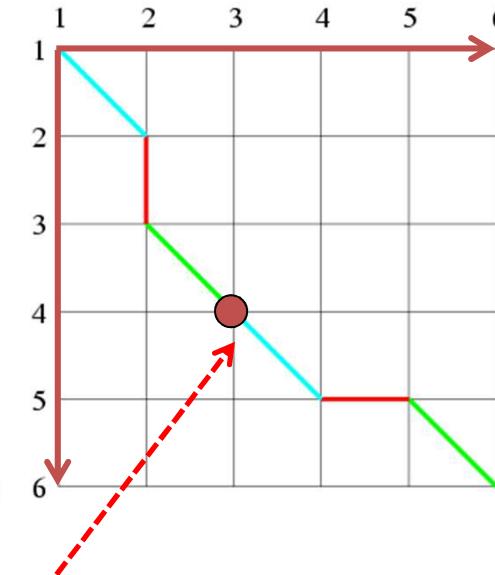
Courtesy J. Ponce

Dynamic Programming

[Uses ordering constraint]



(Baker and Binford, 1981)



Courtesy J. Ponce

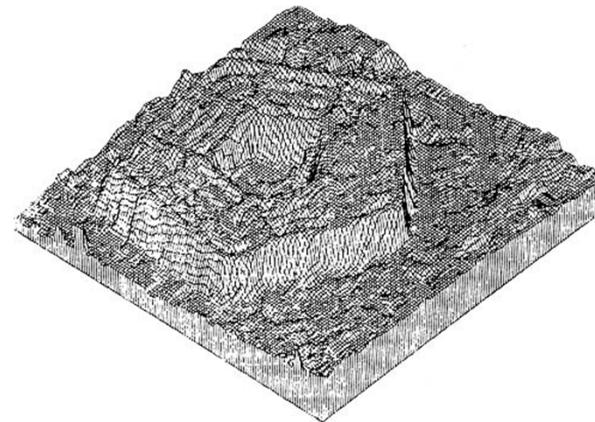
- Nodes = matched feature points (e.g., edge points).
- Arcs = matched intervals along the epipolar lines.
- Arc cost = discrepancy between intervals.

Find the minimum-cost path going monotonically down and right from the top-left corner of the graph to its bottom-right corner.

Dynamic Programming

(Baker and Binford, 1981)

(Ohta and Kanade, 1985)



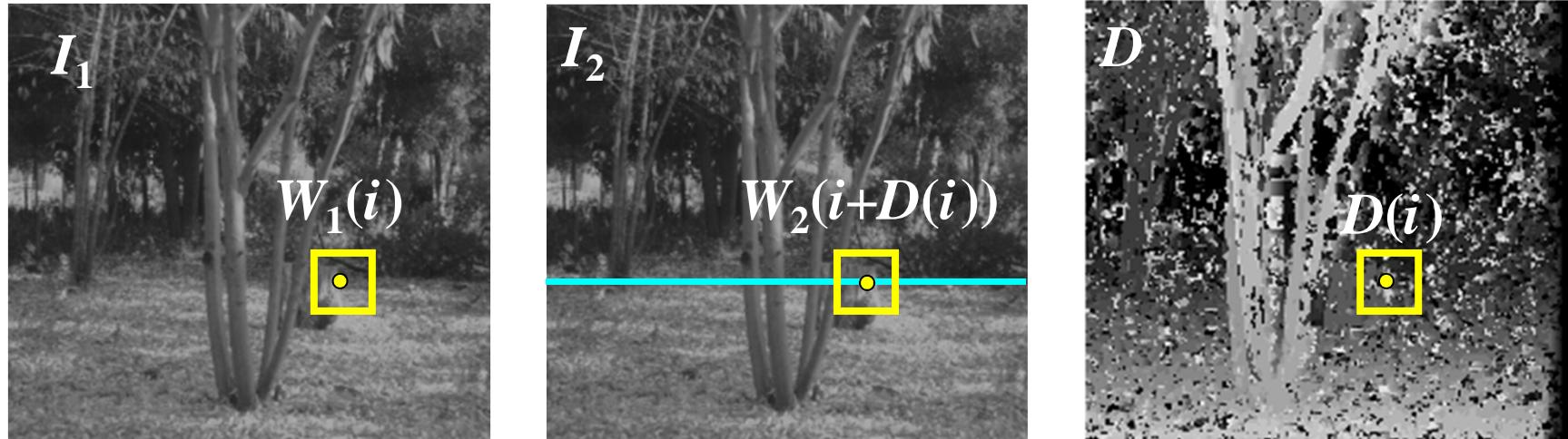
Improving correspondence: Non-local constraints

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

Disparity is typically a smooth function of x (except in occluding boundaries)

Stereo matching as energy minimization

Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 01



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

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

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

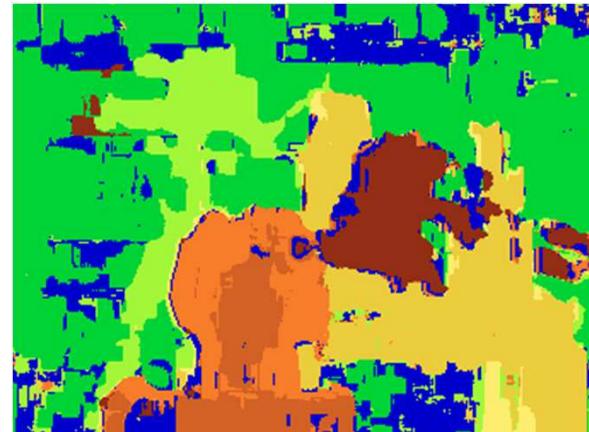
- Energy functions of this form can be minimized using *graph cuts*

Stereo matching as energy minimization

Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 01



Ground truth



Window-based
matching



Energy minimization

Two-frame stereo correspondence algorithms

[Click here](#)

<http://www.middlebury.edu/stereo/>

Stereo SDK vision software development kit

A. Criminisi, A. Blake and D. Robertson



Foreground/background segmentation using stereo

V. Kolmogorov, A. Criminisi, A. Blake, G. Cross and C. Rother.

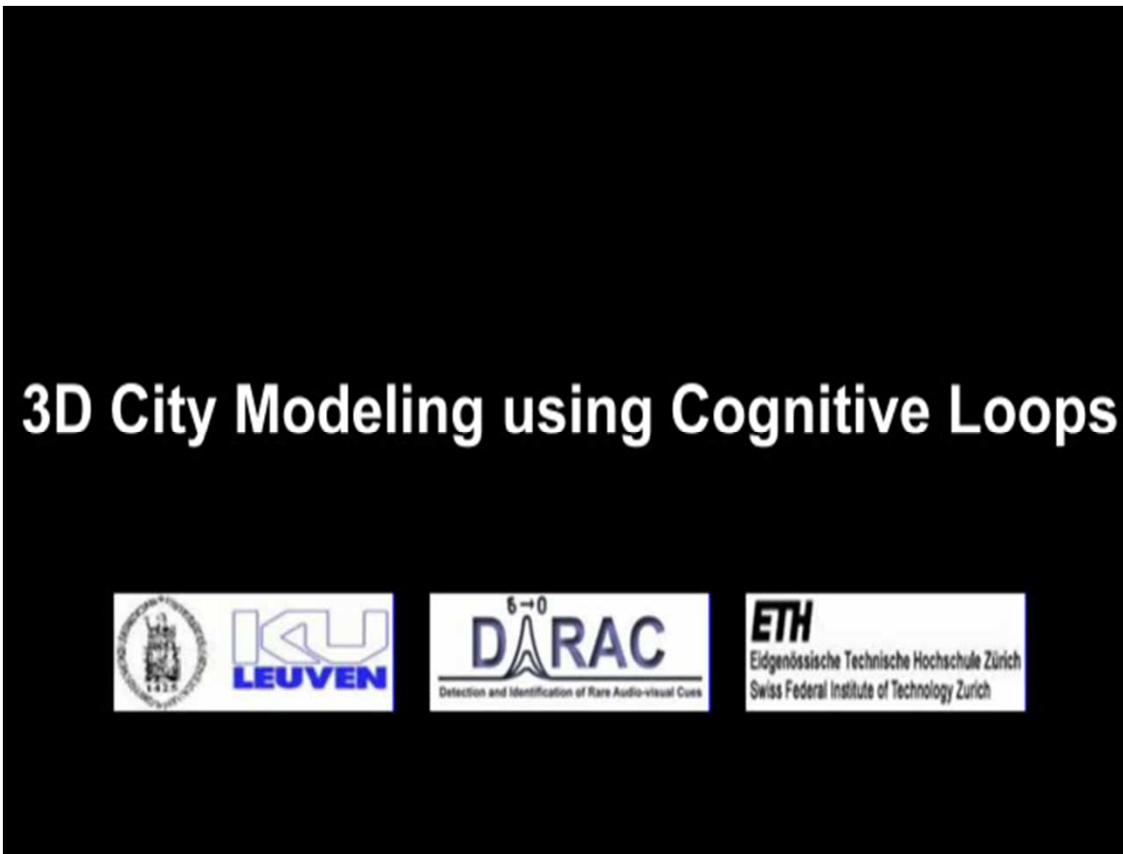
Bi-layer segmentation of binocular stereo video CVPR 2005



http://research.microsoft.com/~antcrim/demos/ACriminisi_Recognition_CowDemo.wmv

3D Urban Scene Modeling using a stereo system

3D Urban Scene Modeling Integrating Recognition and Reconstruction,
N. Cornelis, B. Leibe, K. Cornelis, L. Van Gool, IJCV 08.



<http://www.vision.ee.ethz.ch/showroom/index.en.html#>

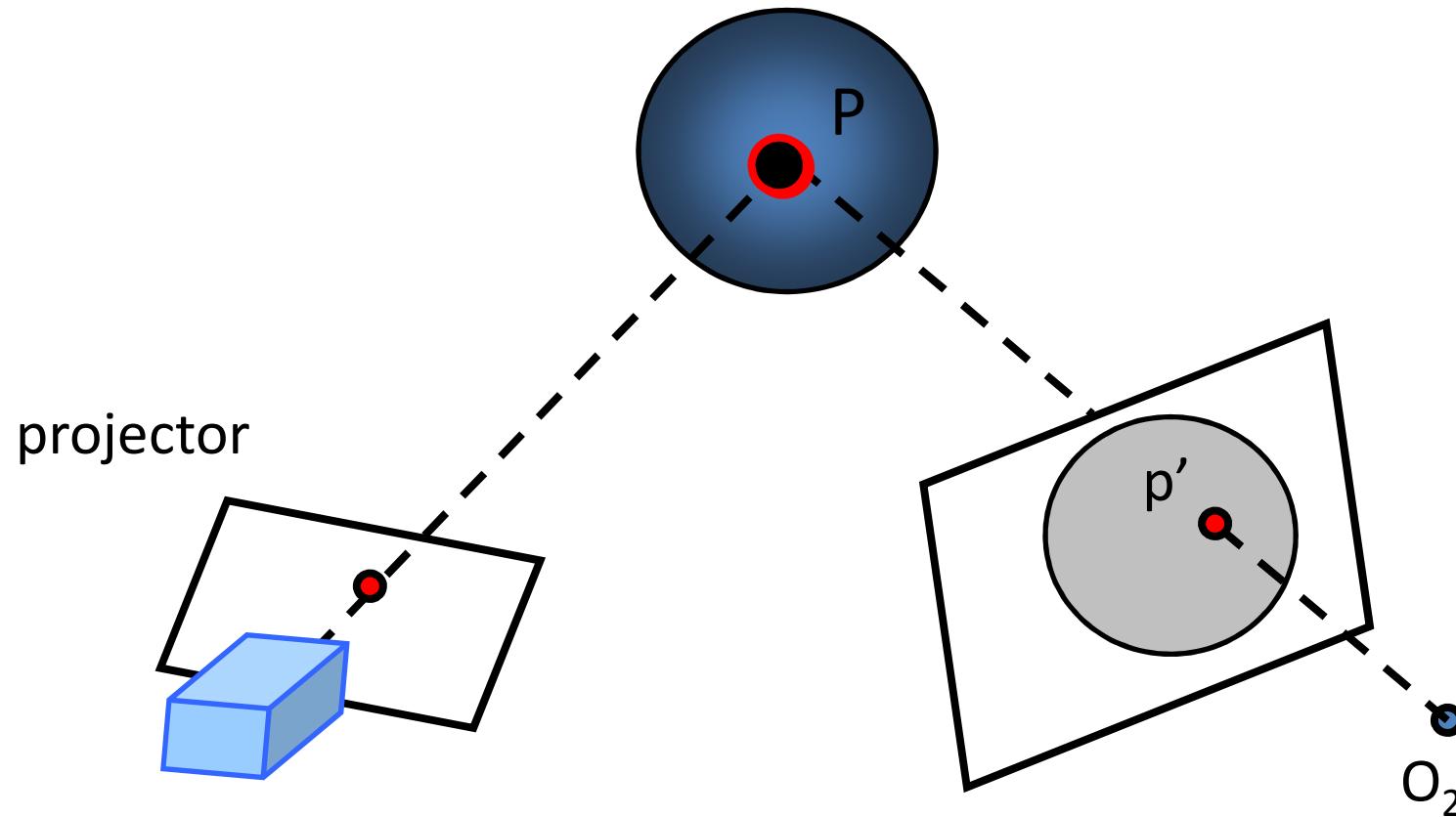
What we will learn today?

- Rectification for stereo vision
- Correspondence problem (Problem Set 2 (Q3))
- Active stereo vision systems (see Friday Oct 28 TA Section)
- Structure from motion

Reading:

[HZ] Chapters: 4, 9, 11
[FP] Chapters: 10

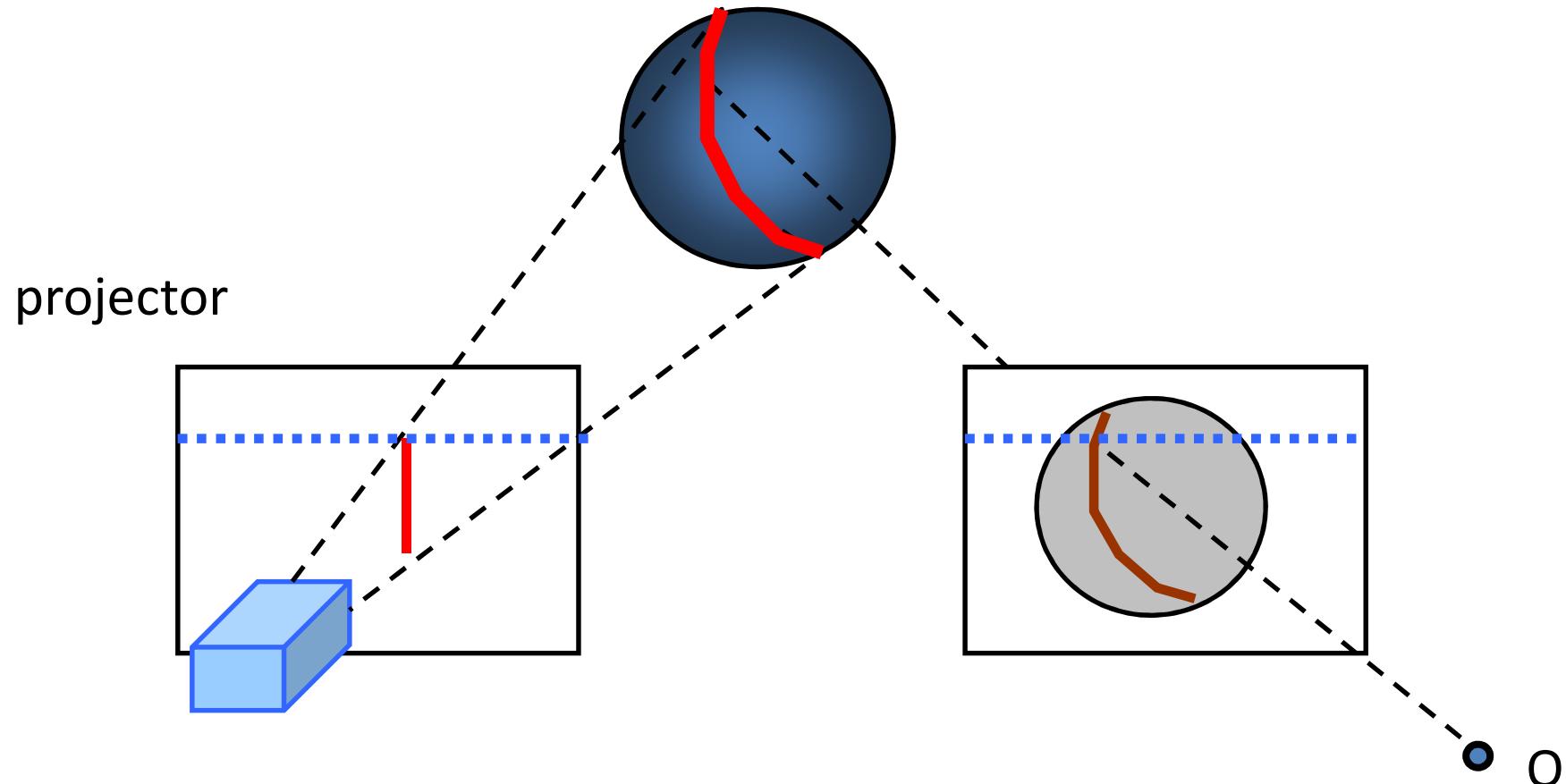
Active stereo (point)



Replace one of the two cameras by a projector

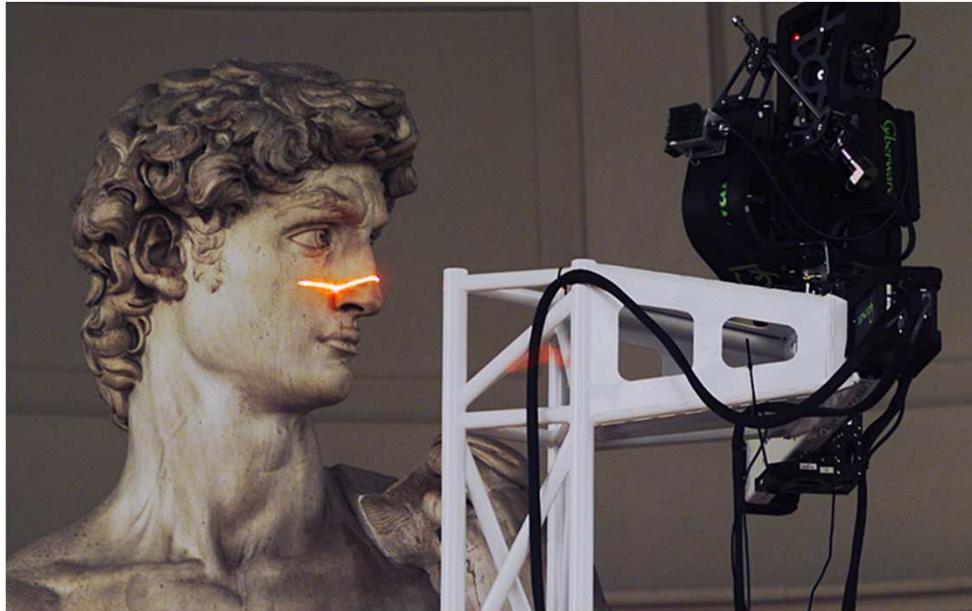
- Single camera
- Projector geometry calibrated
- What's the advantage of having the projector? Correspondence problem solved!

Active stereo (stripe)



- Projector and camera are parallel
- Correspondence problem solved!

Active stereo (stripe)



Digital Michelangelo Project
<http://graphics.stanford.edu/projects/mich/>

- Optical triangulation
 - Project a single stripe of laser light
 - Scan it across the surface of the object
 - This is a very precise version of structured light scanning

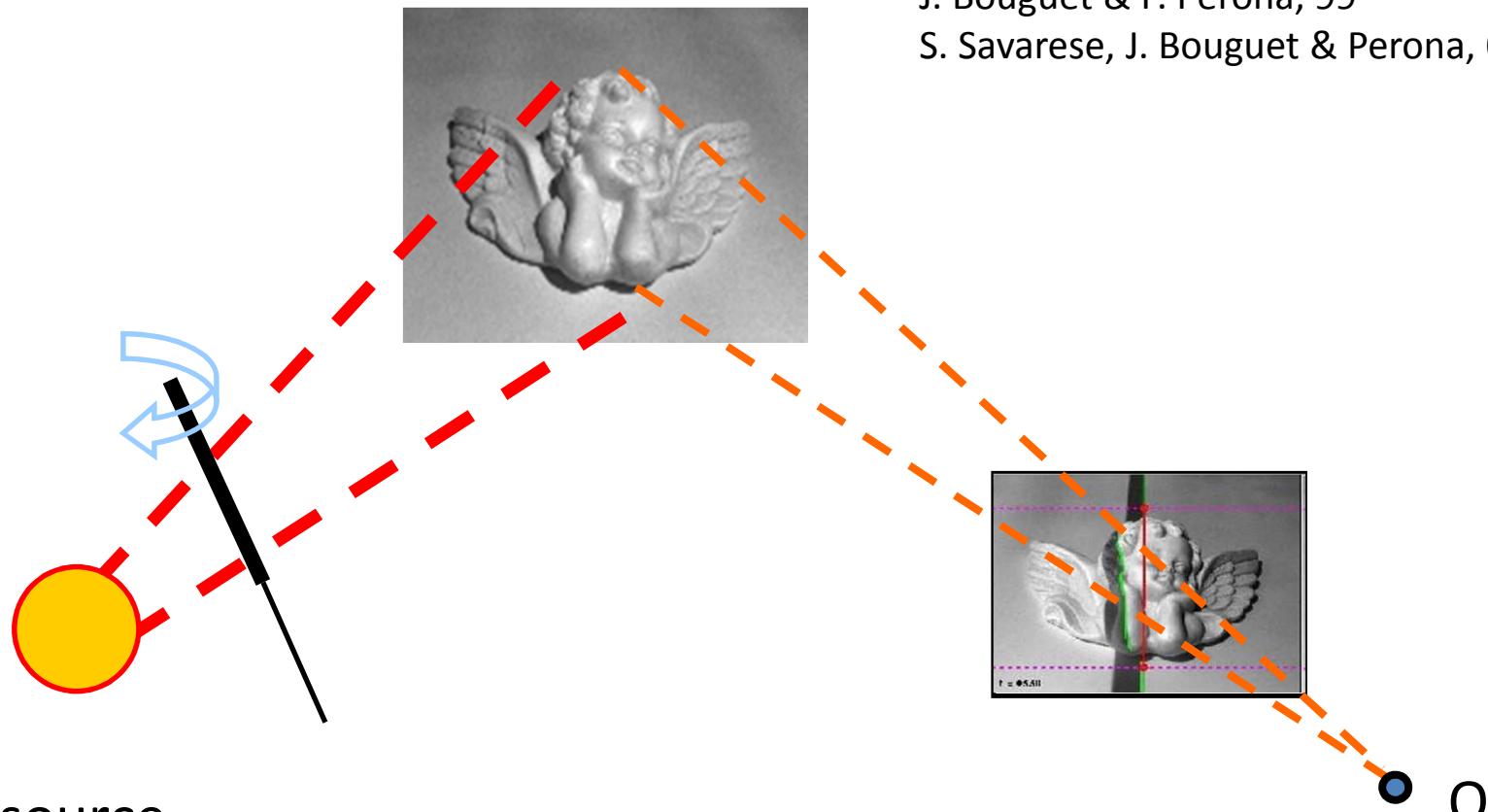
Active stereo (stripe)



Digital Michelangelo Project
<http://graphics.stanford.edu/projects/mich/>

Active stereo (shadows)

J. Bouguet & P. Perona, 99
S. Savarese, J. Bouguet & Perona, 00

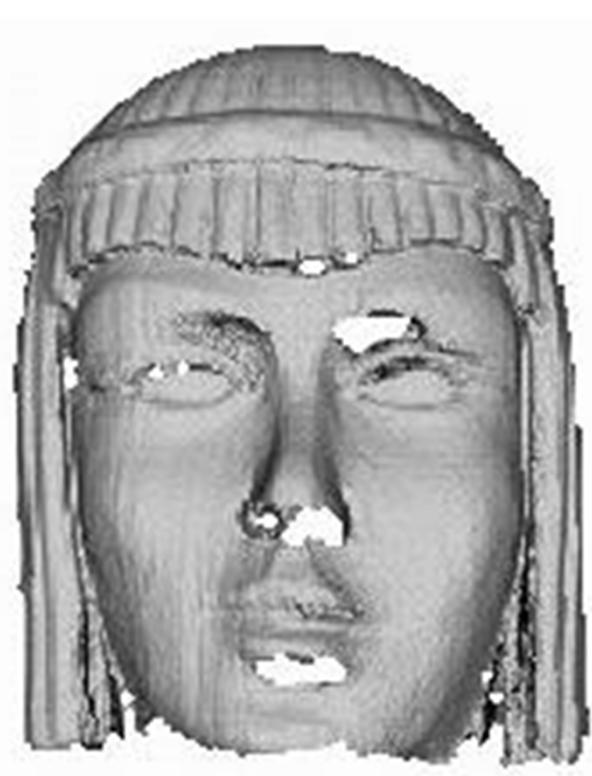


Light source

- 1 camera, 1 light source
- very cheap setup
- calibrated light source

Active stereo (shadows)

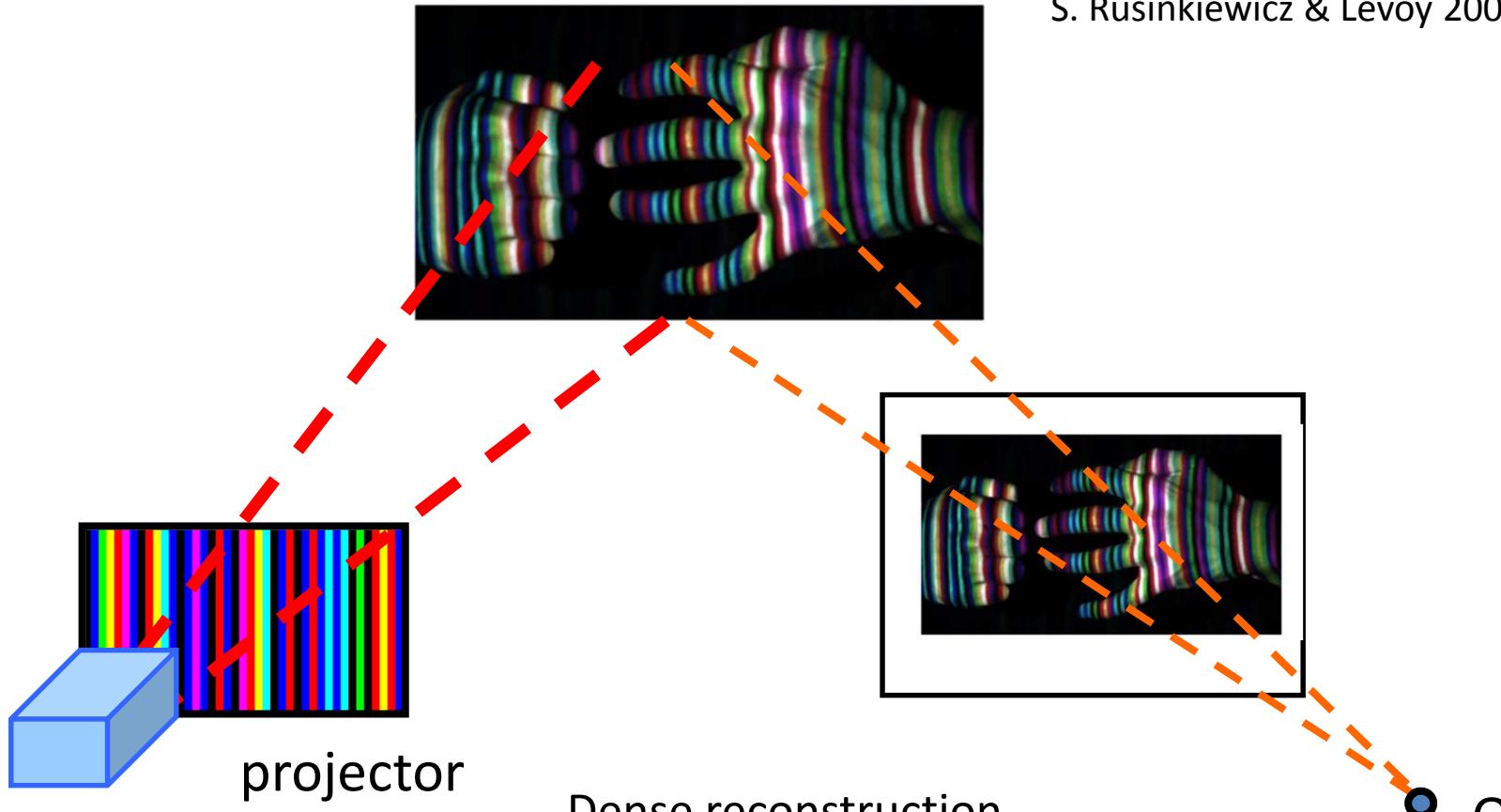
J. Bouguet & P. Perona, 99
S. Savarese, J. Bouguet & Perona, 00



Active stereo (color-coded stripes)

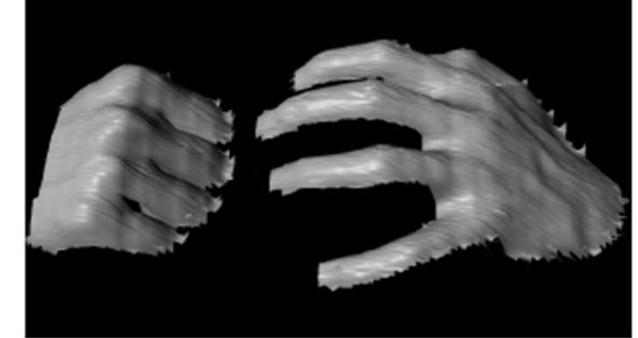
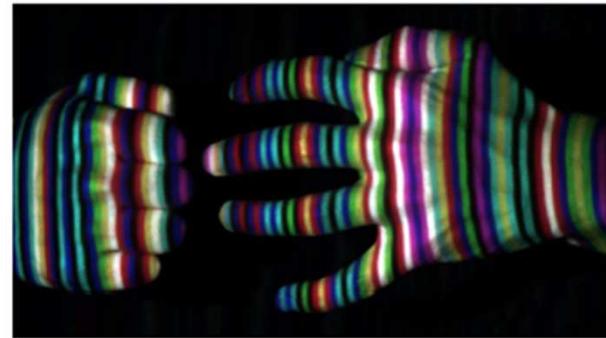
L. Zhang, B. Curless, and S. M. Seitz 2002

S. Rusinkiewicz & Levoy 2002



- Dense reconstruction
- Correspondence problem again
- Get around it by using color codes

Active stereo (color-coded stripes)



Rapid shape acquisition: Projector + stereo cameras

L. Zhang, B. Curless, and S. M. Seitz. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. *3DPVT* 2002



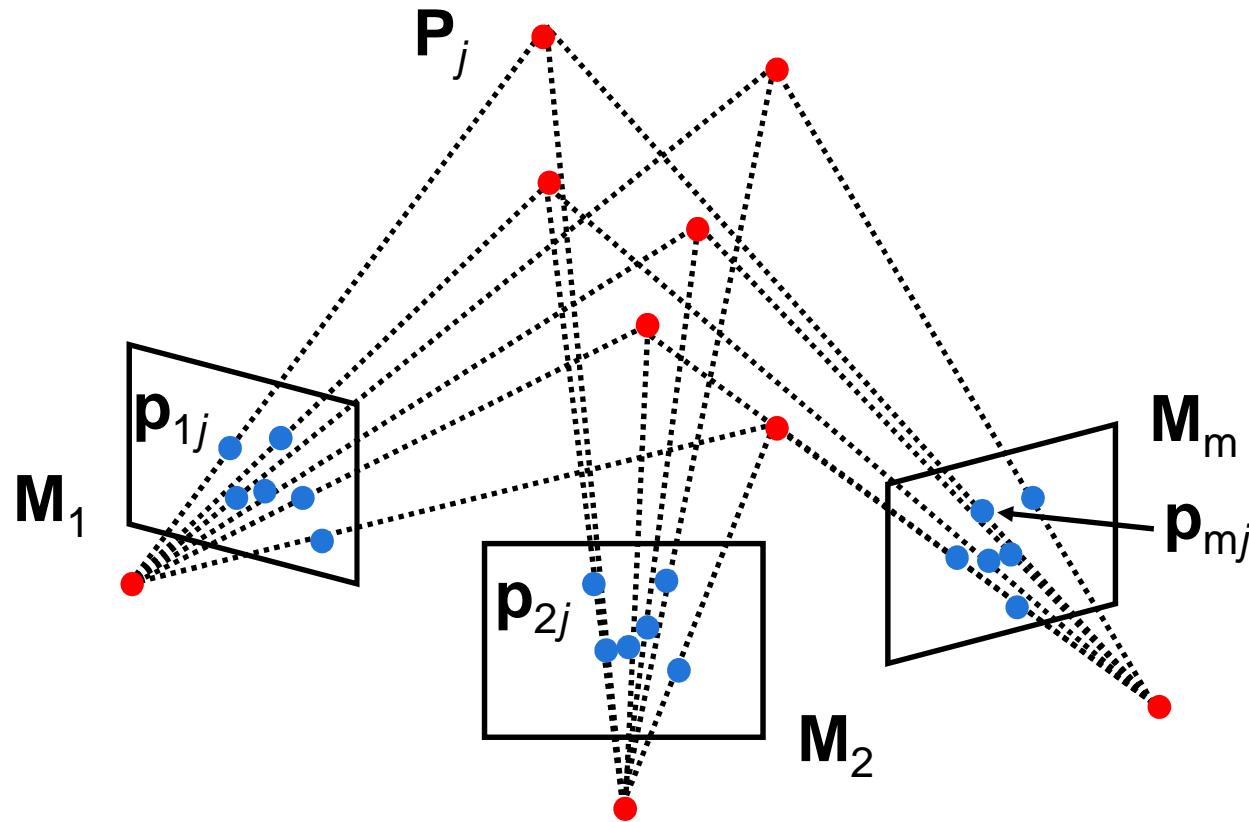
What we will learn today?

- Rectification for stereo vision
- Correspondence problem (Problem Set 2 (Q3))
- Active stereo vision systems
- Structure from motion

Reading:

[HZ] Chapters: 4, 9, 11
[FP] Chapters: 10

Multiple view geometry - the structure from motion problem

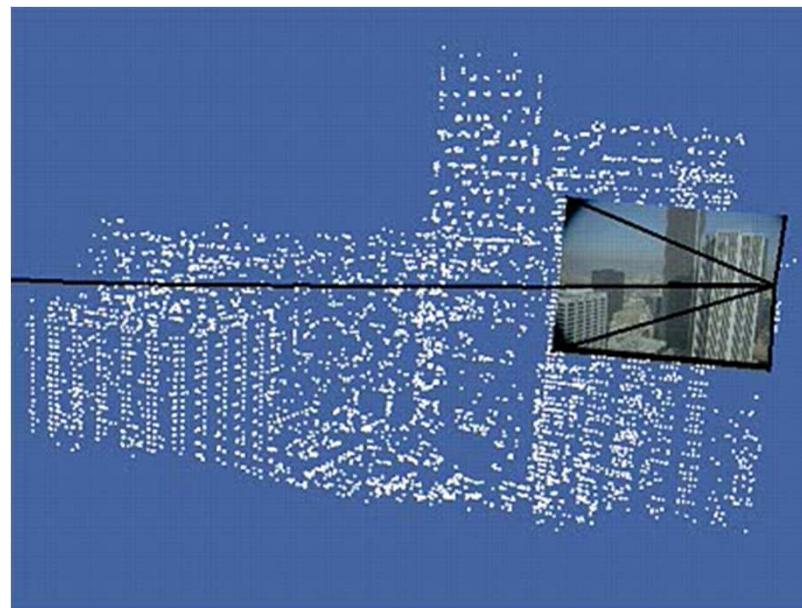


Given m images of n fixed 3D points;

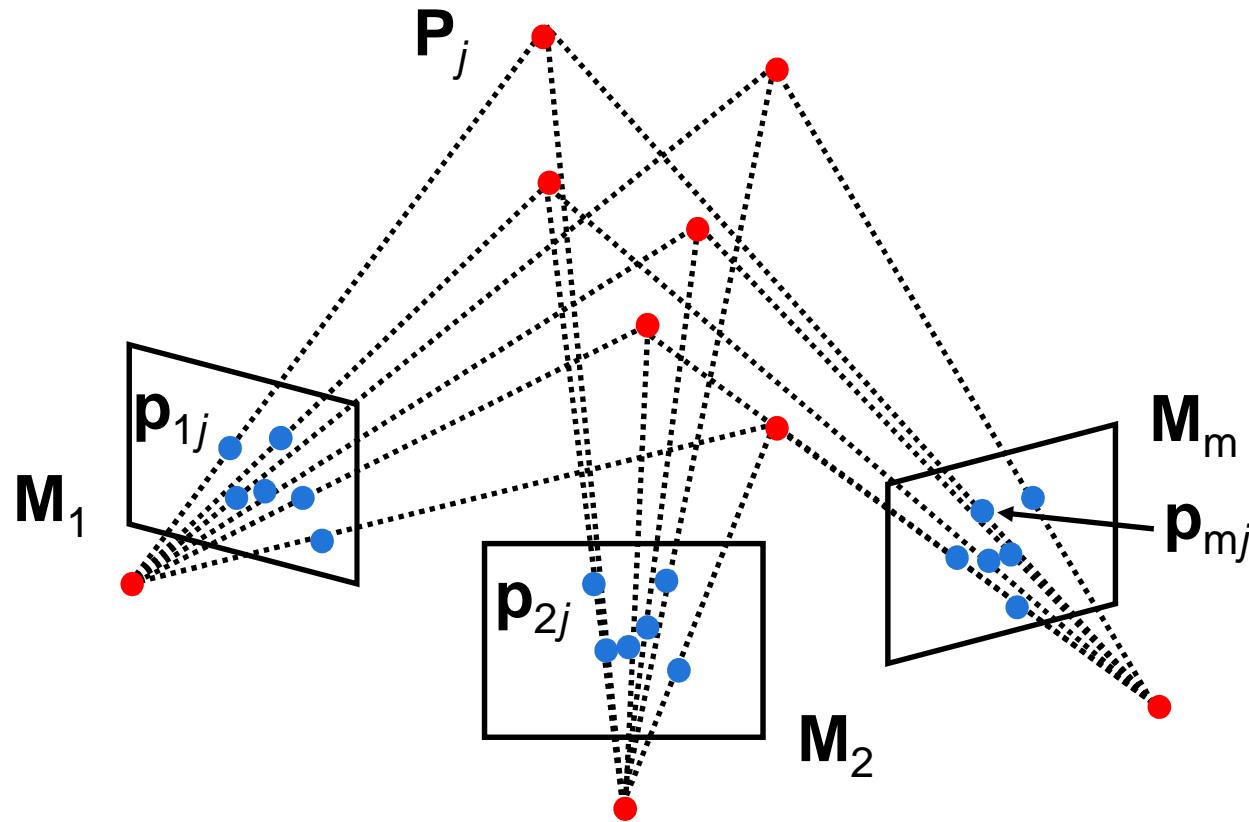
$$\mathbf{p}_{ij} = \mathbf{M}_i \mathbf{P}_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

Multiple view geometry - the structure from motion problem

Courtesy of Oxford **Visual Geometry Group**



Multiple view geometry - the structure from motion problem



From the $m \times n$ correspondences p_{ij} , estimate:

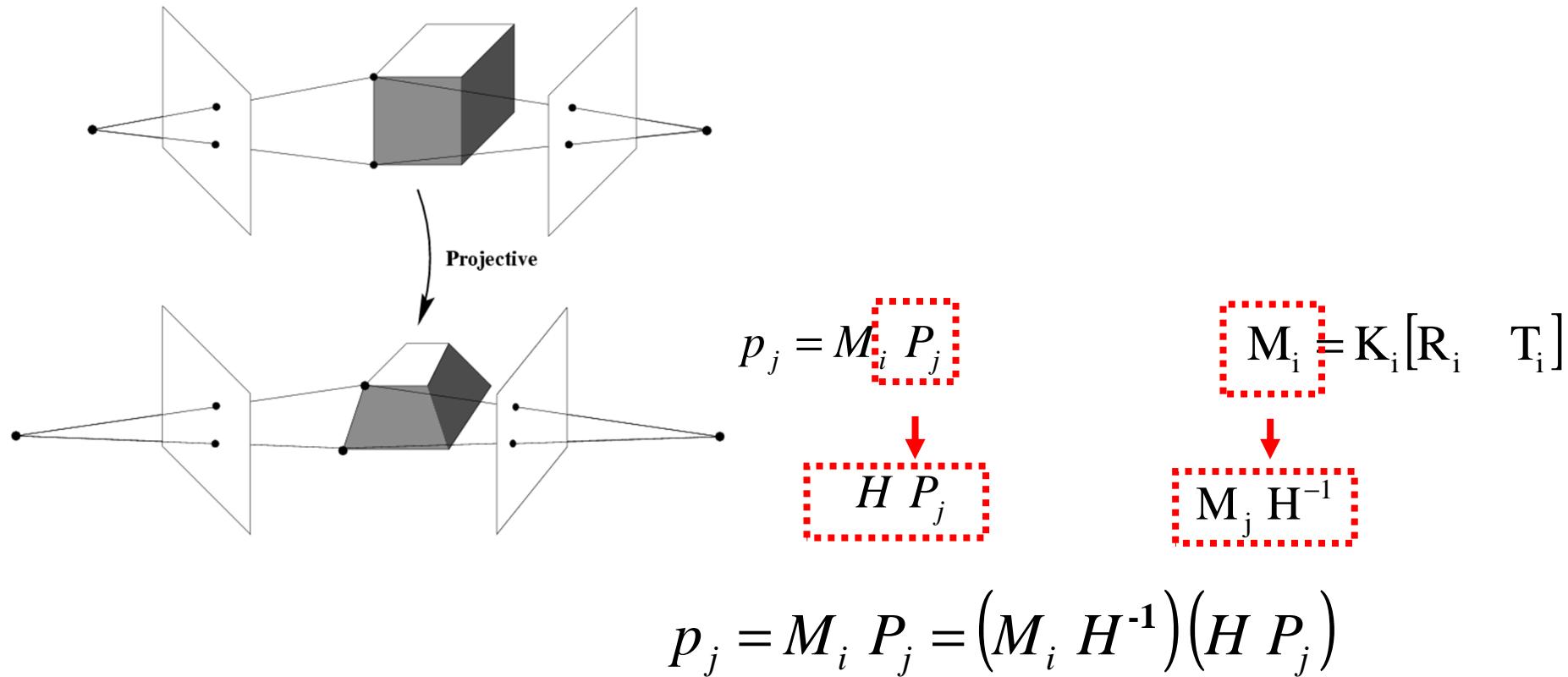
- m projection matrices M_i
- n 3D points P_j

motion

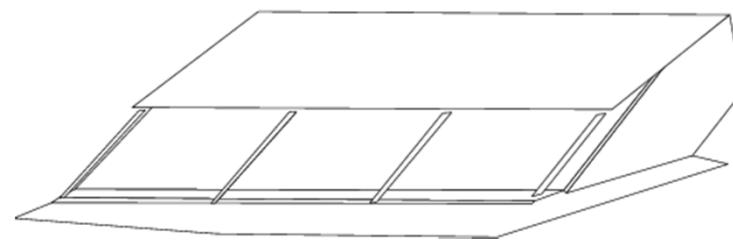
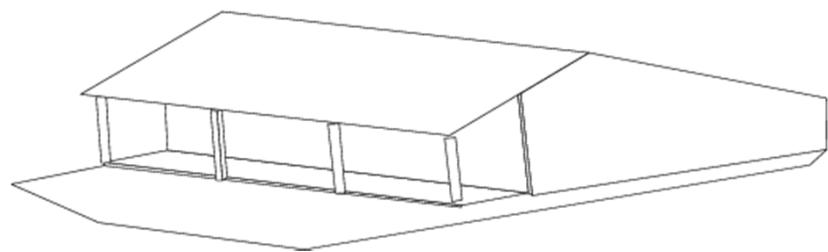
structure

Structure from motion ambiguity

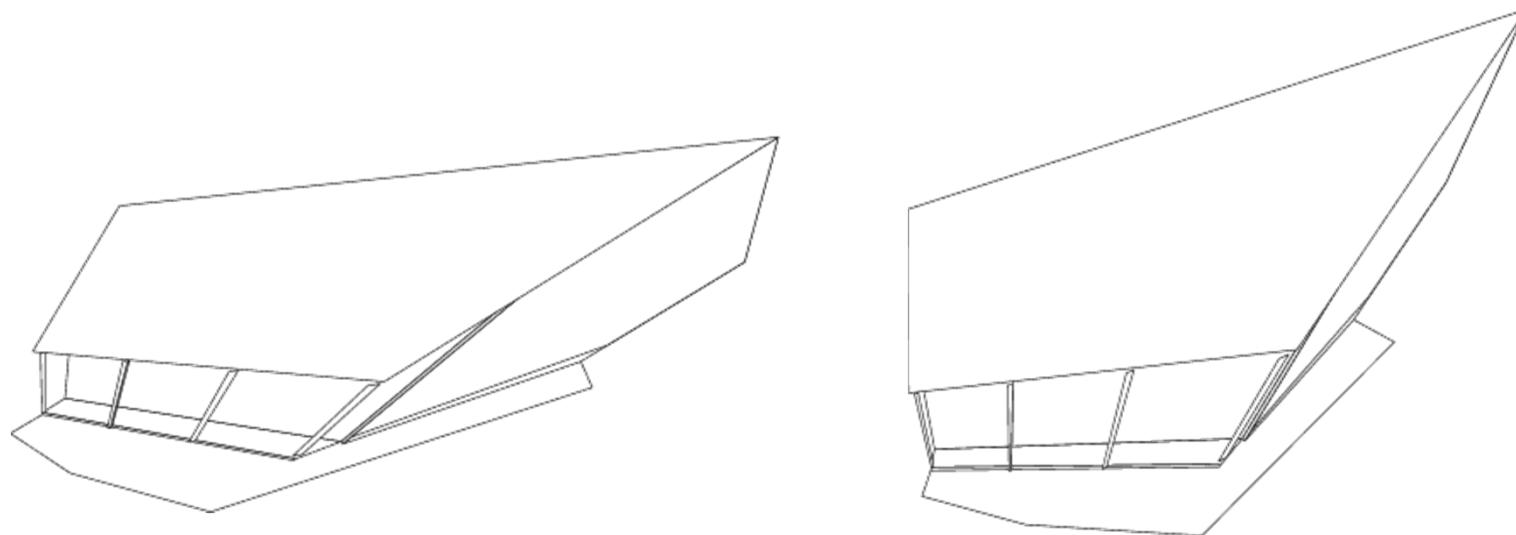
- SFM can be solved up to a N-degree of freedom ambiguity
- In the general case (nothing is known) the ambiguity is expressed by an arbitrary **affine** or **projective transformation**



Affine ambiguity



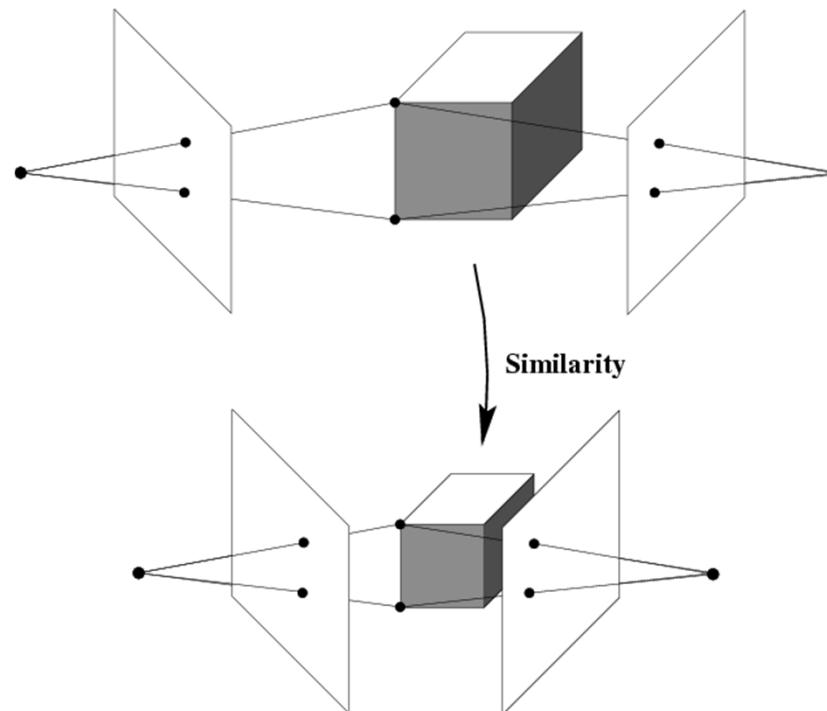
Prospective ambiguity



Structure from motion ambiguity

- A scale (similarity) ambiguity exists even for calibrated cameras
- For calibrated cameras, the similarity ambiguity is the **only** ambiguity
- A reconstruction up to scale is called **metric**

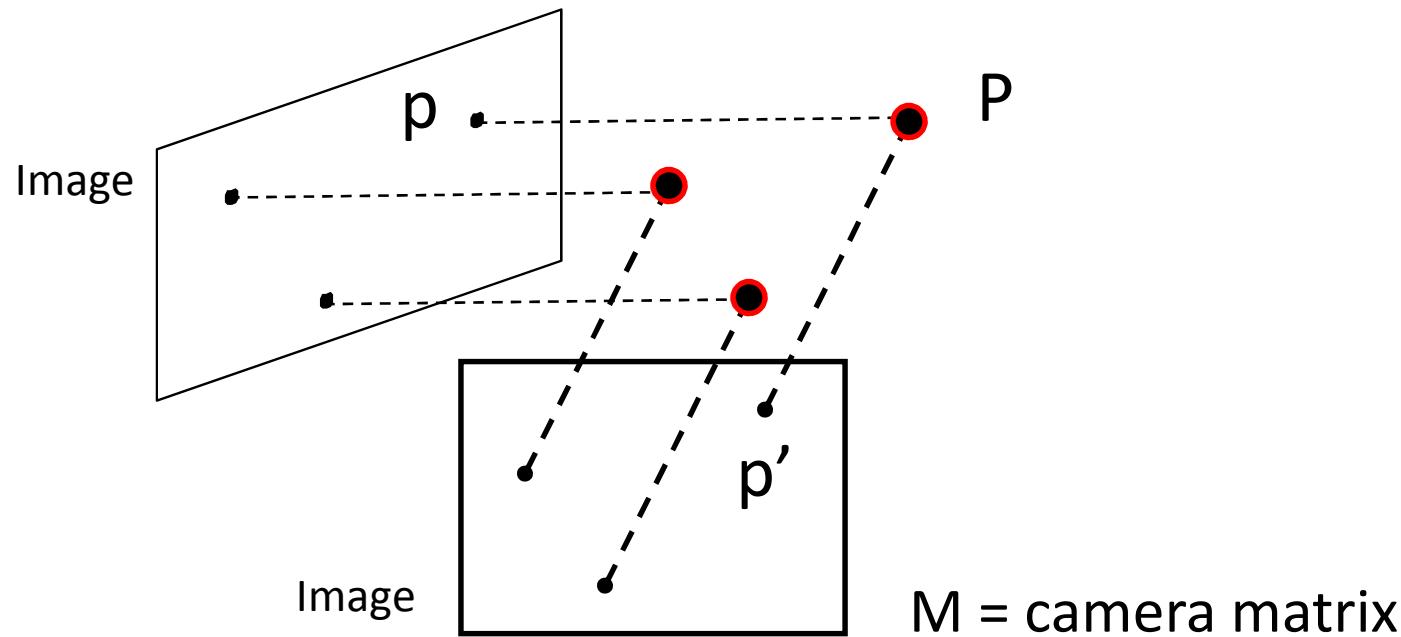
[Longuet-Higgins '81]



Multiple view geometry - overview

- Recover camera and geometry up to ambiguity
 - Use affine approximation if possible (**affine SFM**)
 - Algebraic methods
 - Factorization methods
- Metric upgrade to obtain solution up to scale
(remove perspective or affine ambiguity)
 - Self-calibration
- Bundle adjustment **(optimize solution across all observations)**

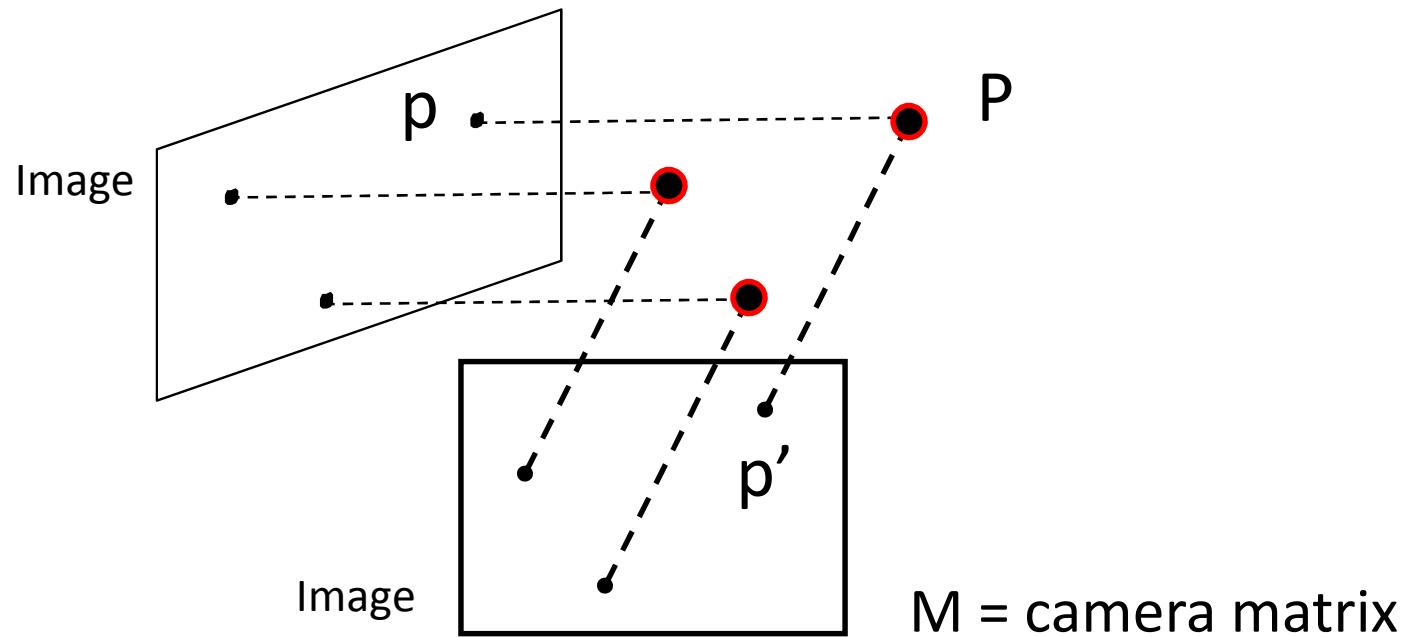
Affine structure from motion - a simpler problem



From the $m \times n$ correspondences \mathbf{p}_{ij} , estimate:

- m projection matrices \mathbf{M}_i (affine cameras)
- n 3D points \mathbf{P}_j

Affine structure from motion - a simpler problem



$$\mathbf{p} = \begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{AP} + \mathbf{b} = M \begin{bmatrix} \mathbf{P} \\ 1 \end{bmatrix}; \quad \mathbf{M} = [\mathbf{A} \quad \mathbf{b}]$$

Affine structure from motion - a simpler problem

Given m images of n fixed points P_j ($=X_i$) we can write

$$\mathbf{p}_{ij} = \mathcal{M}_i \begin{pmatrix} \mathbf{P}_j \\ 1 \end{pmatrix} = \mathcal{A}_i \mathbf{P}_j + \mathbf{b}_i \quad \text{for } i = 1, \dots, \boxed{m} \quad \text{and } j = 1, \dots, \boxed{n}.$$

N of cameras N of points

Problem: estimate the m 2×4 matrices \mathcal{M}_i and
the n positions \mathbf{P}_j from the $m \times n$ correspondences \mathbf{p}_{ij} .

How many equations and how many unknown?

$2m \times n$ equations in $8m+3n$ unknowns

Two approaches:

- Algebraic approach (affine epipolar geometry; estimate F ; cameras; points)
- Factorization method

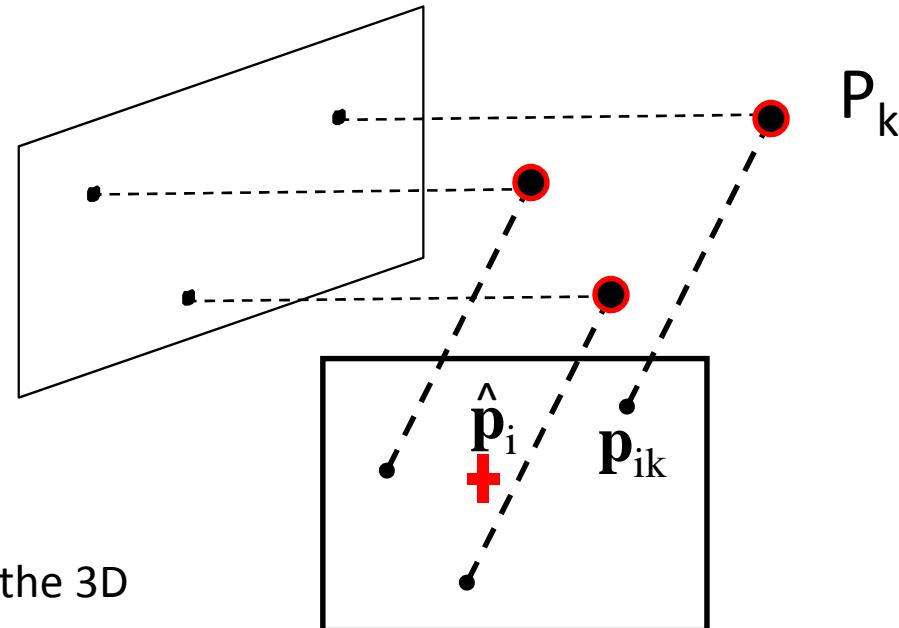
Factorization method

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method, 1992.

Two steps:

- Centering the data
- Factorization

- Centering: subtract the centroid of the image points
- Assume that the origin of the world coordinate system is at the centroid of the 3D points



$$\begin{aligned}\hat{p}_{ij} &= p_{ij} - \frac{1}{n} \sum_{k=1}^n p_{ik} = \mathbf{A}_i P_j + \mathbf{b}_i - \frac{1}{n} \sum_{k=1}^n (\mathbf{A}_i P_k + \mathbf{b}_i) \\ &= \mathbf{A}_i \left(P_j - \frac{1}{n} \sum_{k=1}^n P_k \right) = \mathbf{A}_i \hat{P}_j\end{aligned}$$

Factorization method

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method, 1992.

Two steps:

- Centering the data
- Factorization

After centering, each normalized point \mathbf{p}_{ij} is related to the 3D point \mathbf{P}_i by

$$\hat{\mathbf{p}}_{ij} = \mathbf{A}_i \mathbf{P}_j$$

Factorization method

- Let's create a $2m \times n$ data (measurement) matrix:

$$\mathbf{D} = \begin{bmatrix} \hat{p}_{11} & \hat{p}_{12} & \cdots & \hat{p}_{1n} \\ \hat{p}_{21} & \hat{p}_{22} & \cdots & \hat{p}_{2n} \\ & & \ddots & \\ \hat{p}_{m1} & \hat{p}_{m2} & \cdots & \hat{p}_{mn} \end{bmatrix}$$

cameras
($2m$)

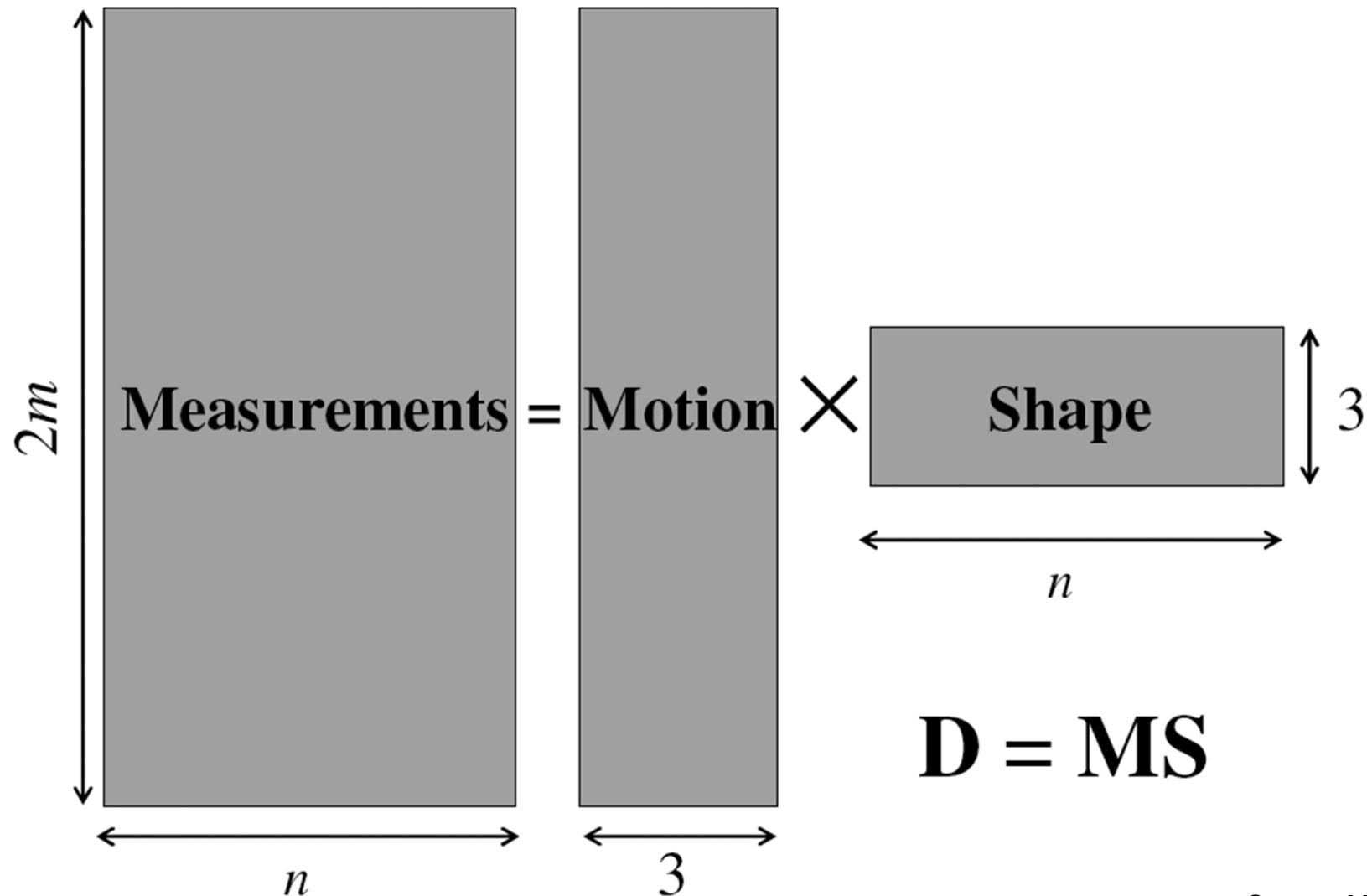
points (n)

Factorization method

- Let's create a $2m \times n$ data (measurement) matrix:

The measurement matrix $\mathbf{D} = \mathbf{M} \mathbf{S}$ has rank 3
(it's a product of a 2×3 matrix and $3 \times n$ matrix)

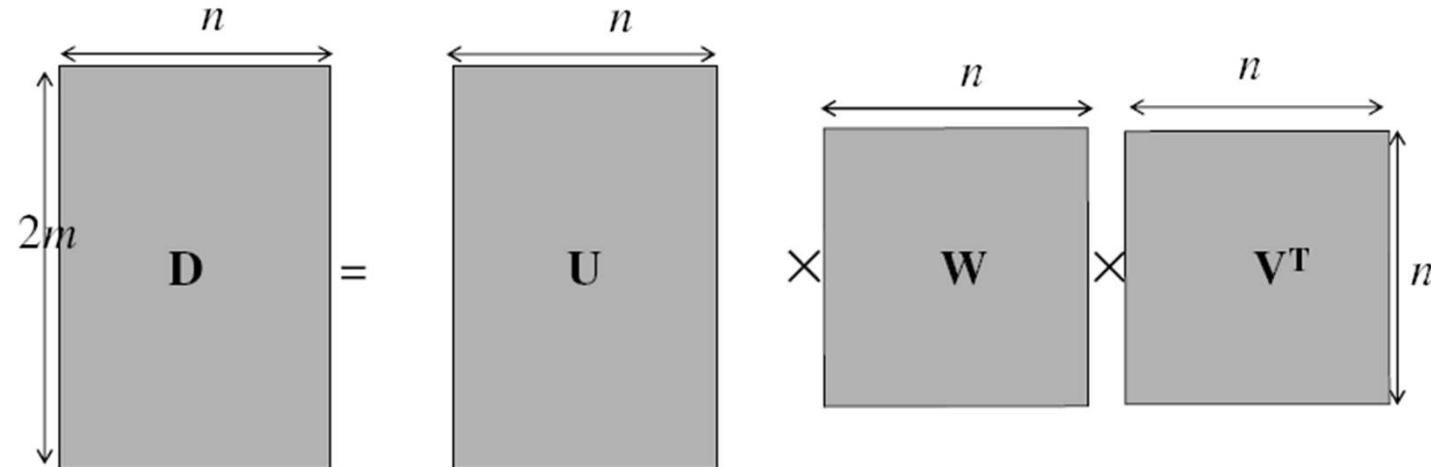
Factorization method



Source: M. Hebert

Factorization method

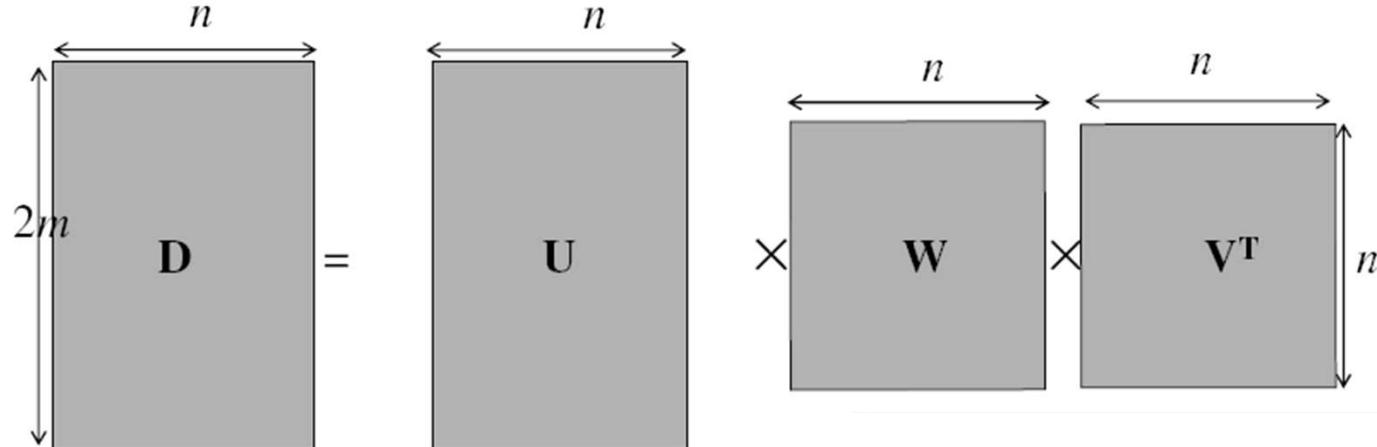
- Singular value decomposition of D:



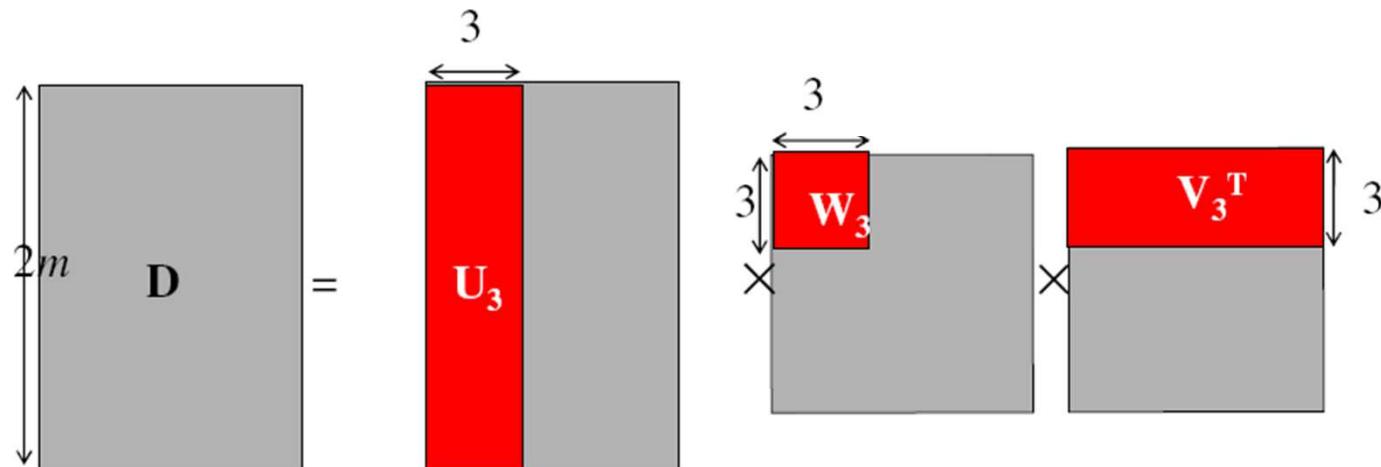
Source: M. Hebert

Factorization method

- Singular value decomposition of D:



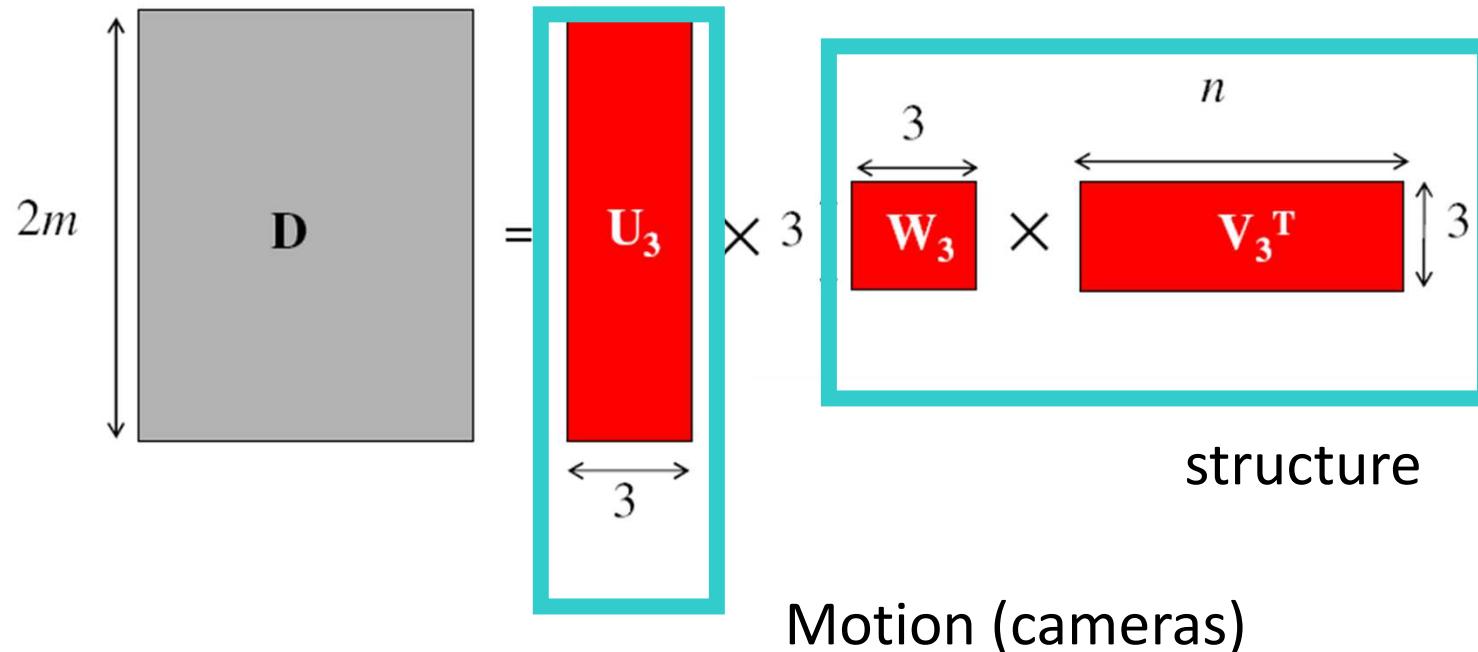
Since rank (D)=3, there are only 3 non-zero singular values



Source: M. Hebert

Factorization method

- Obtaining a factorization from SVD:



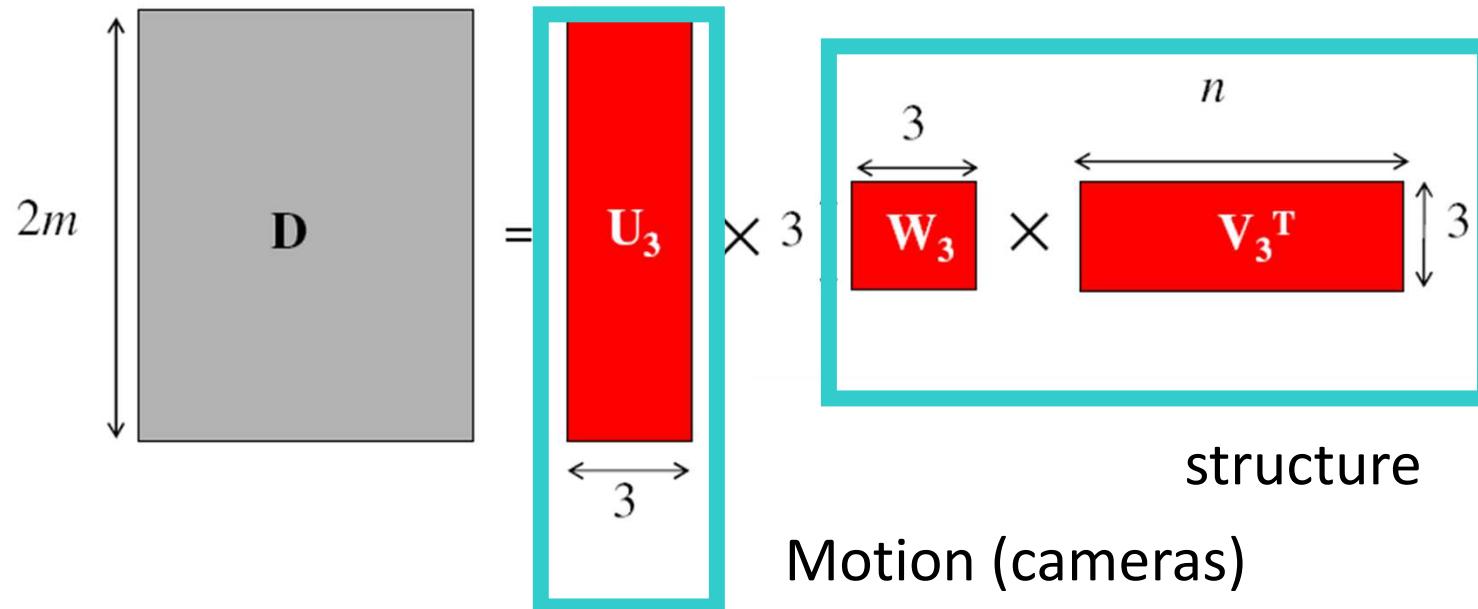
What is the issue here?

D has rank > 3 because of
- measurement noise
- affine approximation

Source: M. Hebert

Factorization method

- Obtaining a factorization from SVD:



Theorem: When \mathcal{A} has a rank greater than p , $\mathcal{U}_p \mathcal{W}_p \mathcal{V}_p^T$ is the best possible rank- p approximation of \mathcal{A} in the sense of the Frobenius norm.

$$\mathcal{D} = \mathcal{U}_3 \mathcal{W}_3 \mathcal{V}_3^T$$

$$\begin{cases} \mathcal{A}_0 = \mathcal{U}_3 \\ \mathcal{P}_0 = \mathcal{W}_3 \mathcal{V}_3^T \end{cases}$$

Factorization method

1. Given: m images and n features \mathbf{p}_{ij}
2. For each image i , center the feature coordinates
3. Construct a $2m \times n$ measurement matrix \mathbf{D} :
 - Column j contains the projection of point j in all views
 - Row i contains one coordinate of the projections of all the n points in image i
4. Factorize \mathbf{D} :
 - Compute SVD: $\mathbf{D} = \mathbf{U} \mathbf{W} \mathbf{V}^T$
 - Create \mathbf{U}_3 by taking the first 3 columns of \mathbf{U}
 - Create \mathbf{V}_3 by taking the first 3 columns of \mathbf{V}
 - Create \mathbf{W}_3 by taking the upper left 3×3 block of \mathbf{W}
5. Create the motion and shape matrices:
 - $\mathbf{M} = \mathbf{M} = \mathbf{U}_3$ and $\mathbf{S} = \mathbf{W}_3 \mathbf{V}_3^T$ (or $\mathbf{U}_3 \mathbf{W}_3^{1/2}$ and $\mathbf{S} = \mathbf{W}_3^{1/2} \mathbf{V}_3^T$)

Source: M. Hebert

Multiple view geometry - overview

- Recover camera and geometry up to ambiguity
 - Use affine approximation if possible (affine SFM)
 - Algebraic methods
 - Factorization methods
- Metric upgrade to obtain solution up to scale
(remove perspective or affine ambiguity)
 - Self-calibration
- Bundle adjustment (optimize solution across all observations)

Self-calibration

[HZ] Chapters: 18

- Prior knowledge on cameras or scene can be used to add constraints and remove ambiguities
- Obtain metric reconstruction (up to scale)

Condition	N. Views
• Constant internal parameters	3
• Aspect ratio and skew known • Focal length and offset vary	4*
• Aspect ratio and skew known • Focal length and offset vary	5*
• skew =0, all other parameters vary	8*

Multiple view geometry - overview

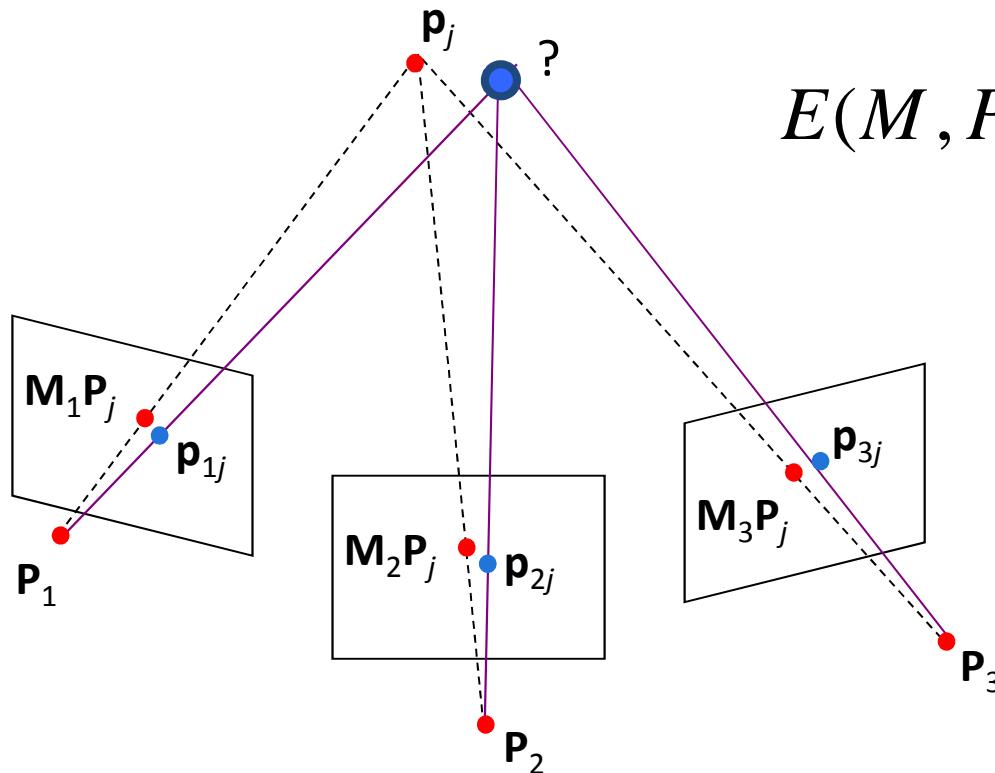
- Recover camera and geometry up to ambiguity
 - Use affine approximation if possible (affine SFM)
 - Algebraic methods
 - Factorization methods
- Metric upgrade to obtain solution up to scale (remove ambiguity)
 - Self-calibration
- **Bundle adjustment (optimize solution across all observations)**

Bundle adjustment

[HZ] Chapters: 18

- Non-linear method for refining structure and motion
- Minimizing re-projection error
- It can be used before or after metric upgrade

$$E(M, P) = \sum_{i=1}^m \sum_{j=1}^n D(p_{ij}, M_i P_j)^2$$



Bundle adjustment

[HZ] Chapters: 18

- Non-linear method for refining structure and motion
- Minimizing re-projection error
- It can be used before or after metric upgrade

$$E(M, P) = \sum_{i=1}^m \sum_{j=1}^n D(p_{ij}, M_i P_j)^2$$

- **Advantages**

- Handle large number of views
- Handle missing data

- **Limitations**

- Large minimization problem (parameters grow with number of views)
- Requires good initial condition

Photo synth

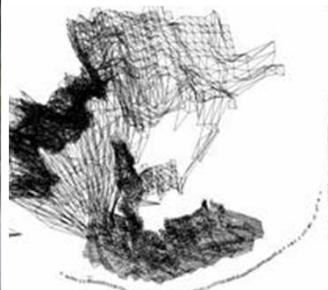
Noah Snavely, Steven M. Seitz, Richard Szeliski, "[Photo tourism: Exploring photo collections in 3D](#)," ACM Transactions on Graphics (SIGGRAPH Proceedings), 2006,

<http://phototour.cs.washington.edu/bundler/>



Bundler is a structure-from-motion (SfM) system for unordered image collections (for instance, images from the Internet) written in C and C++.

Additional references



D. Nistér, PhD thesis '01
See also *5-point algorithm*



M. Pollefeys et al 98---



M. Brown and D. G. Lowe, 05

What we have learned today

- Rectification for stereo vision
- Correspondence problem (**Problem Set 2 (Q3)**)
- Active stereo vision systems (see Friday Oct 28 TA Section)
- Structure from motion

Reading:

[HZ] Chapters: 4, 9, 11
[FP] Chapters: 10