Lecture 10: Multi-view geometry

Professor Fei-Fei Li

Stanford Vision Lab

What we will learn today?

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

Reading:

[HZ] Chapters: 4, 9, 11

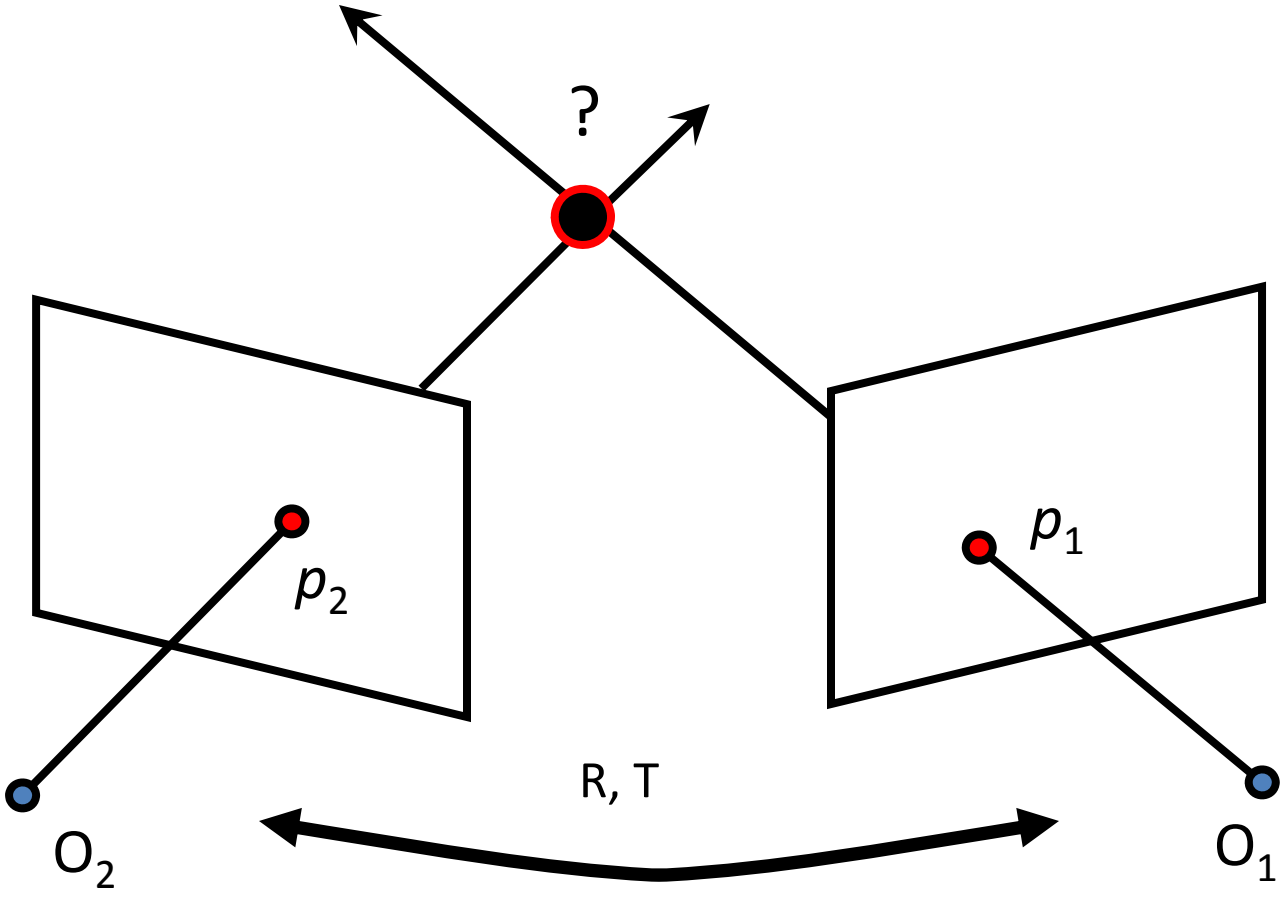
[FP] Chapters: 10

What we will learn today?

- 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!



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

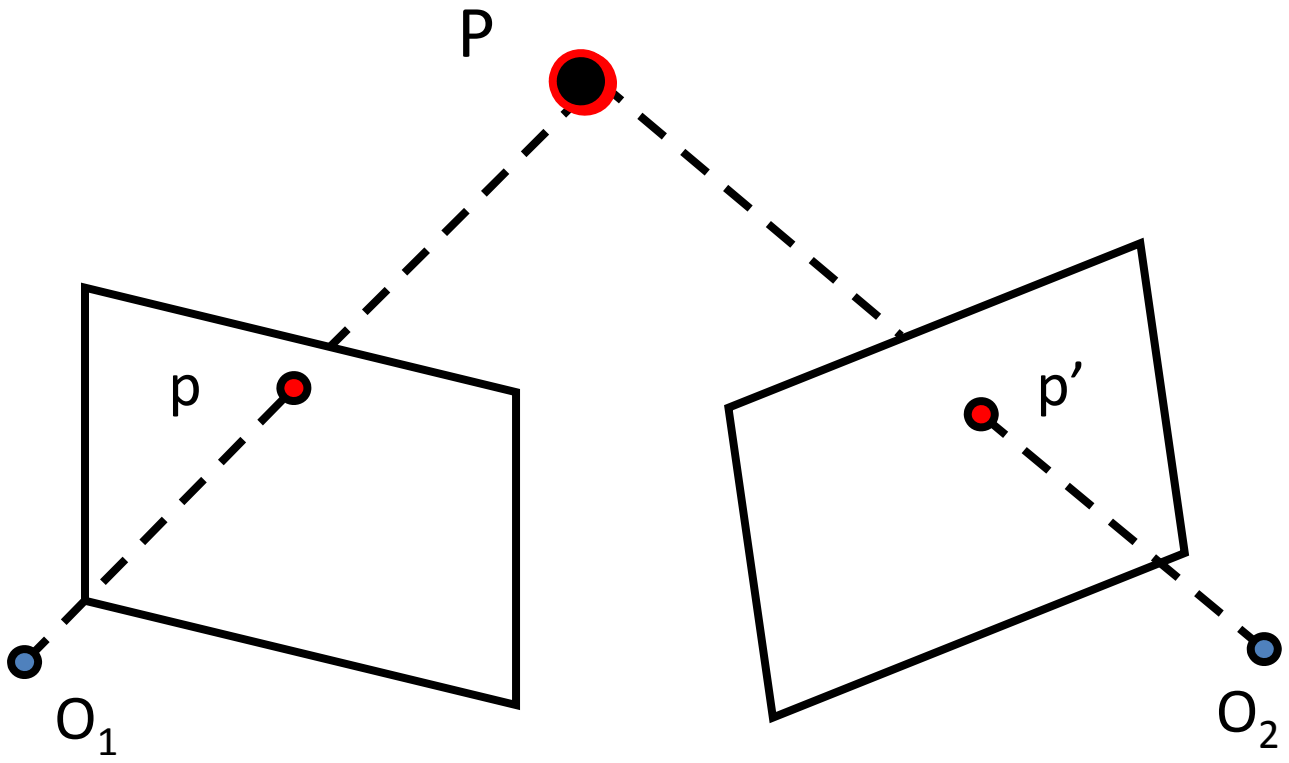
This lecture (#10)

- **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.

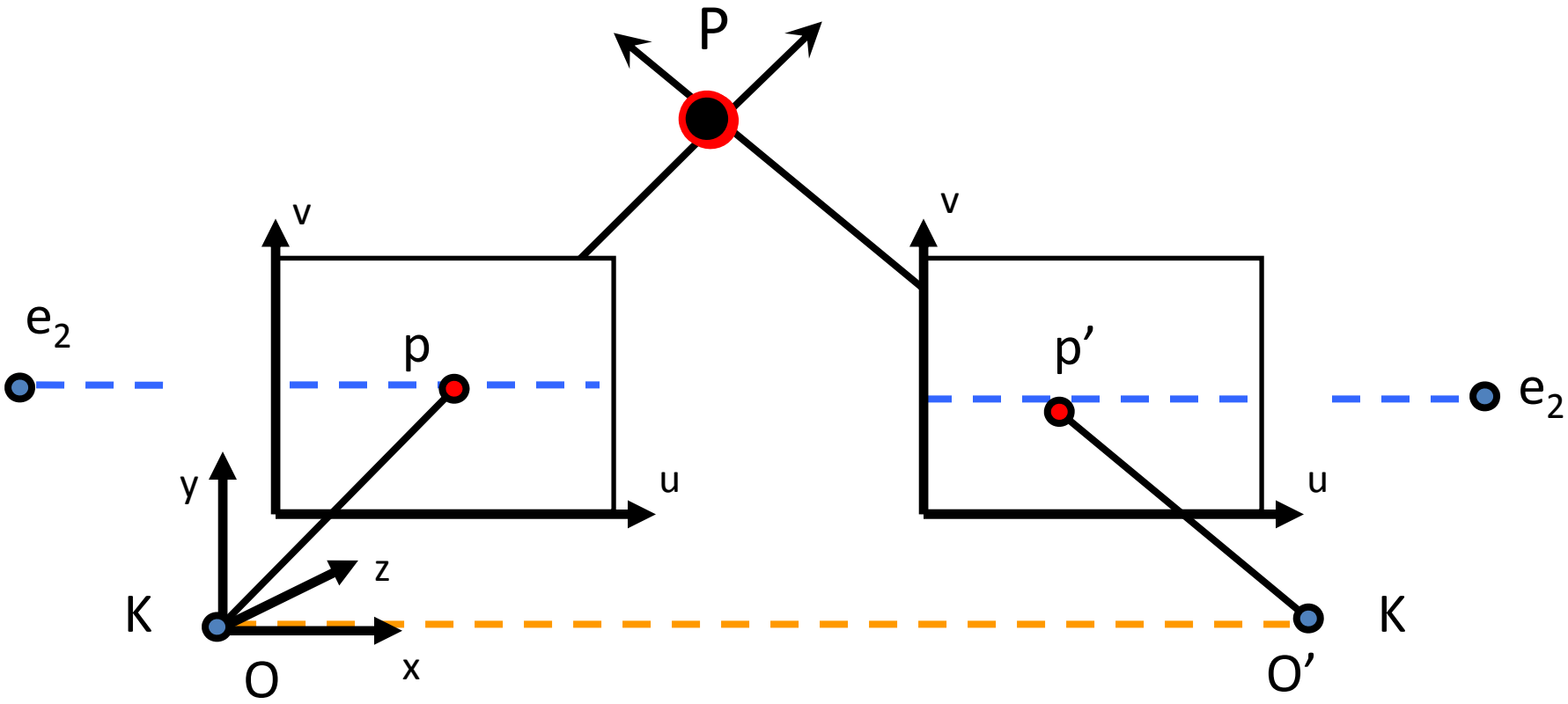
Stereo-view geometry



Subgoals:

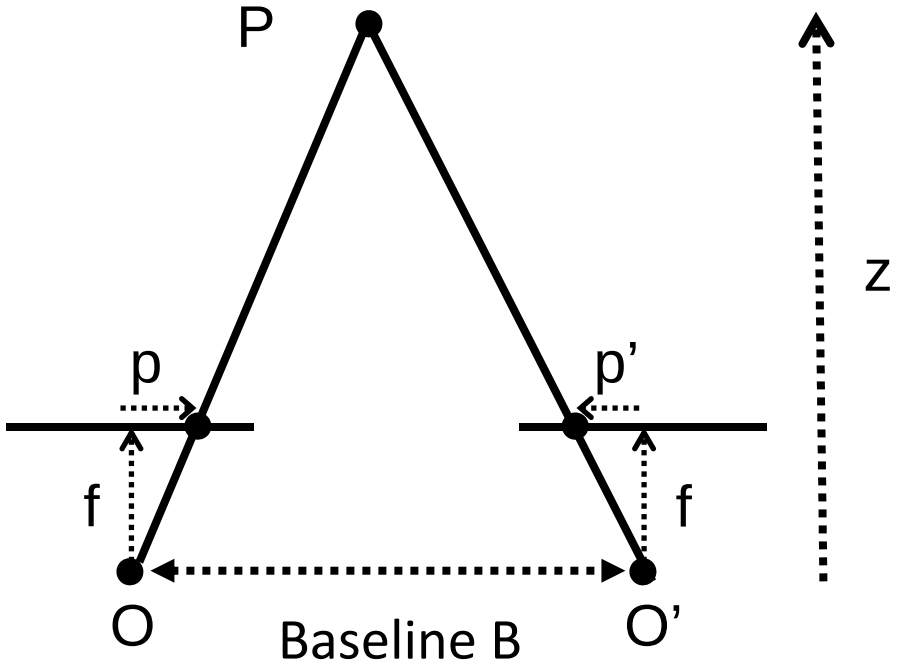
- Solve the correspondence problem
- Use corresponding observations to triangulate (depth estimation)

Depth estimation



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

Depth estimation



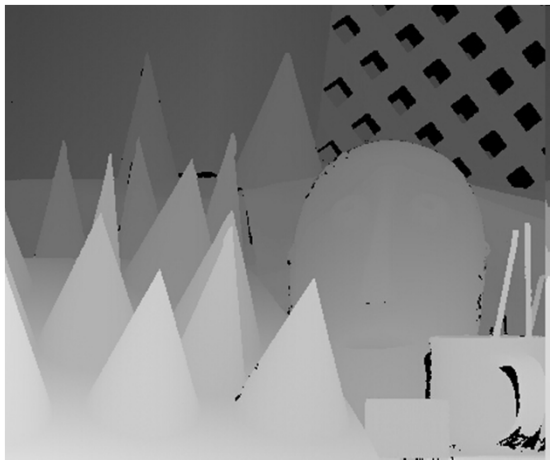
$$p - p' = \frac{B \cdot f}{z}$$ = disparity

Note: Disparity is inversely proportional to depth

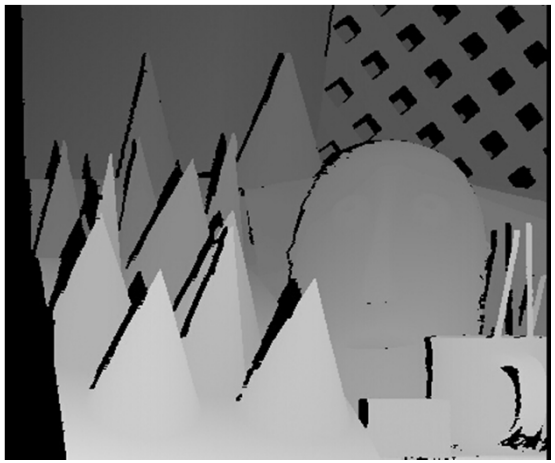
Depth estimation

Stereo pair

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



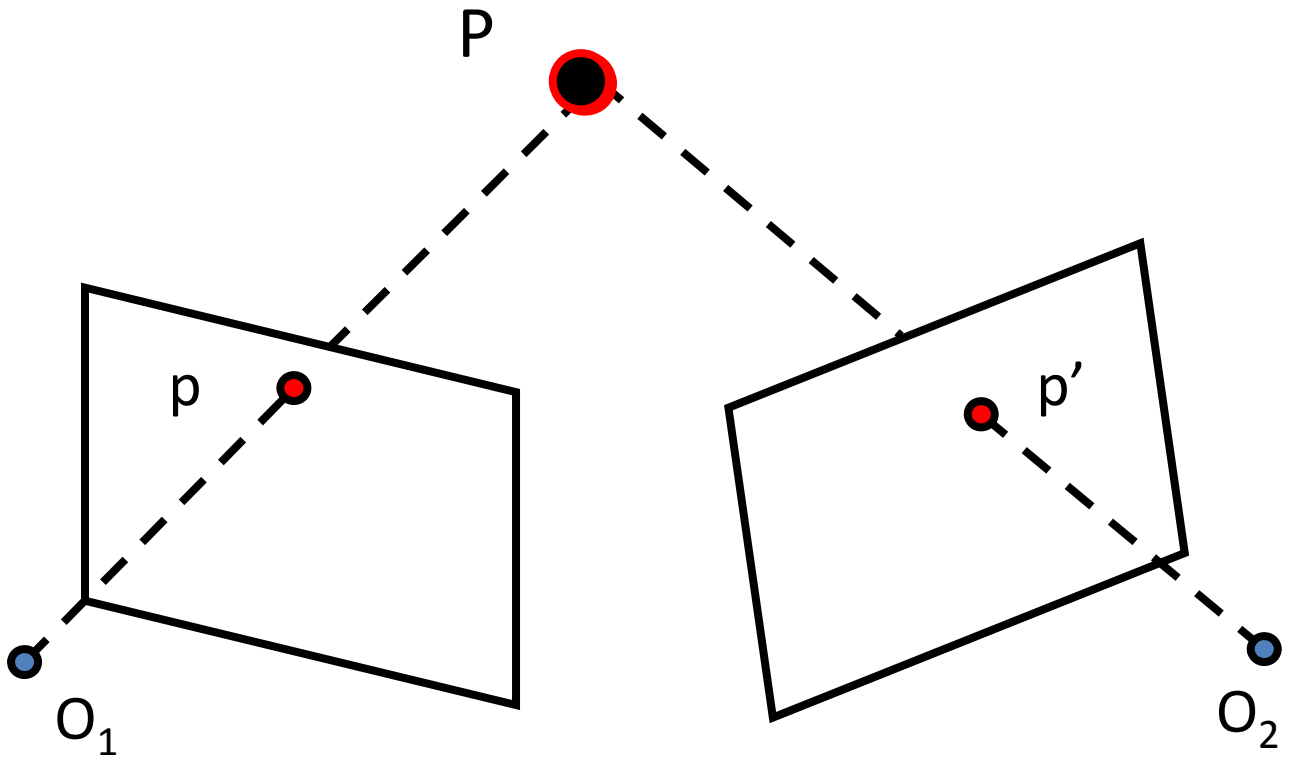
Disparity map / depth map



Disparity map with occlusions

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

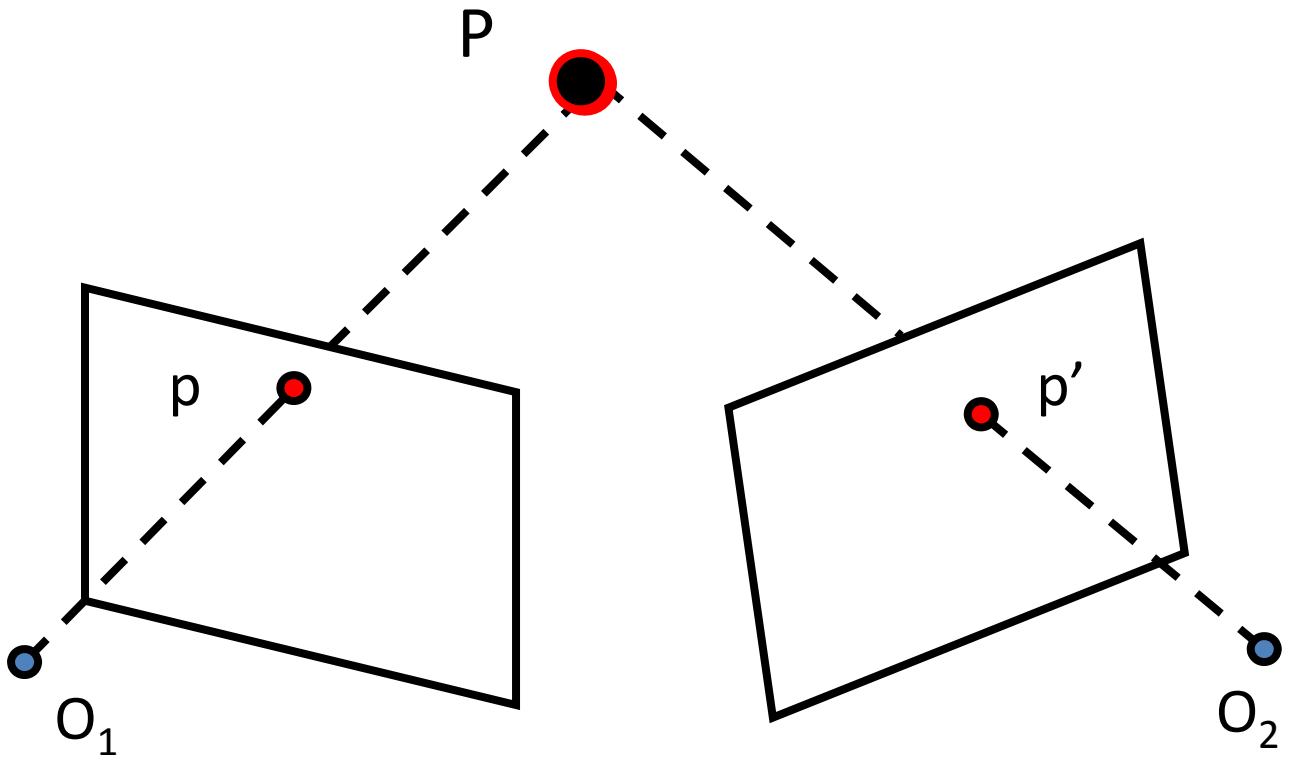
Depth estimation



Subgoals:

- Solve the correspondence problem
- Use corresponding observations to triangulate

Correspondence problem



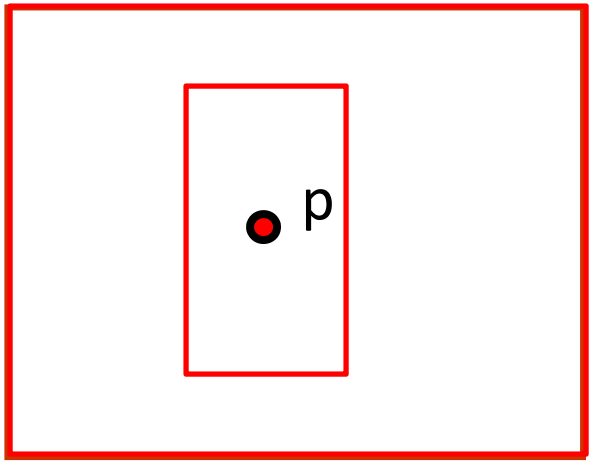
Given a point in 3D, discover corresponding observations
in left and right images [also called binocular fusion problem]

Correspondence problem

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

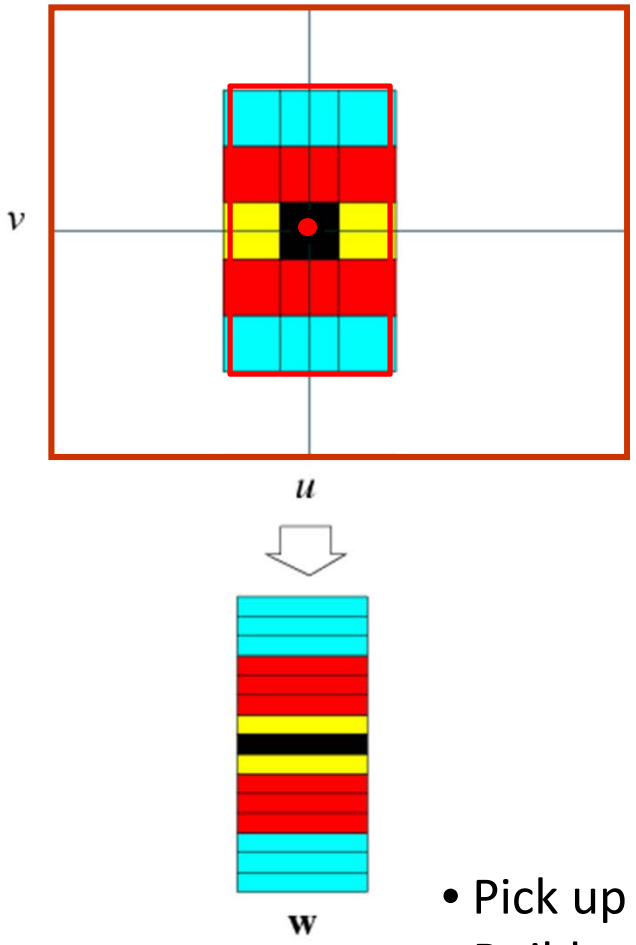
[FP] Chapters: 11

Correlation Methods



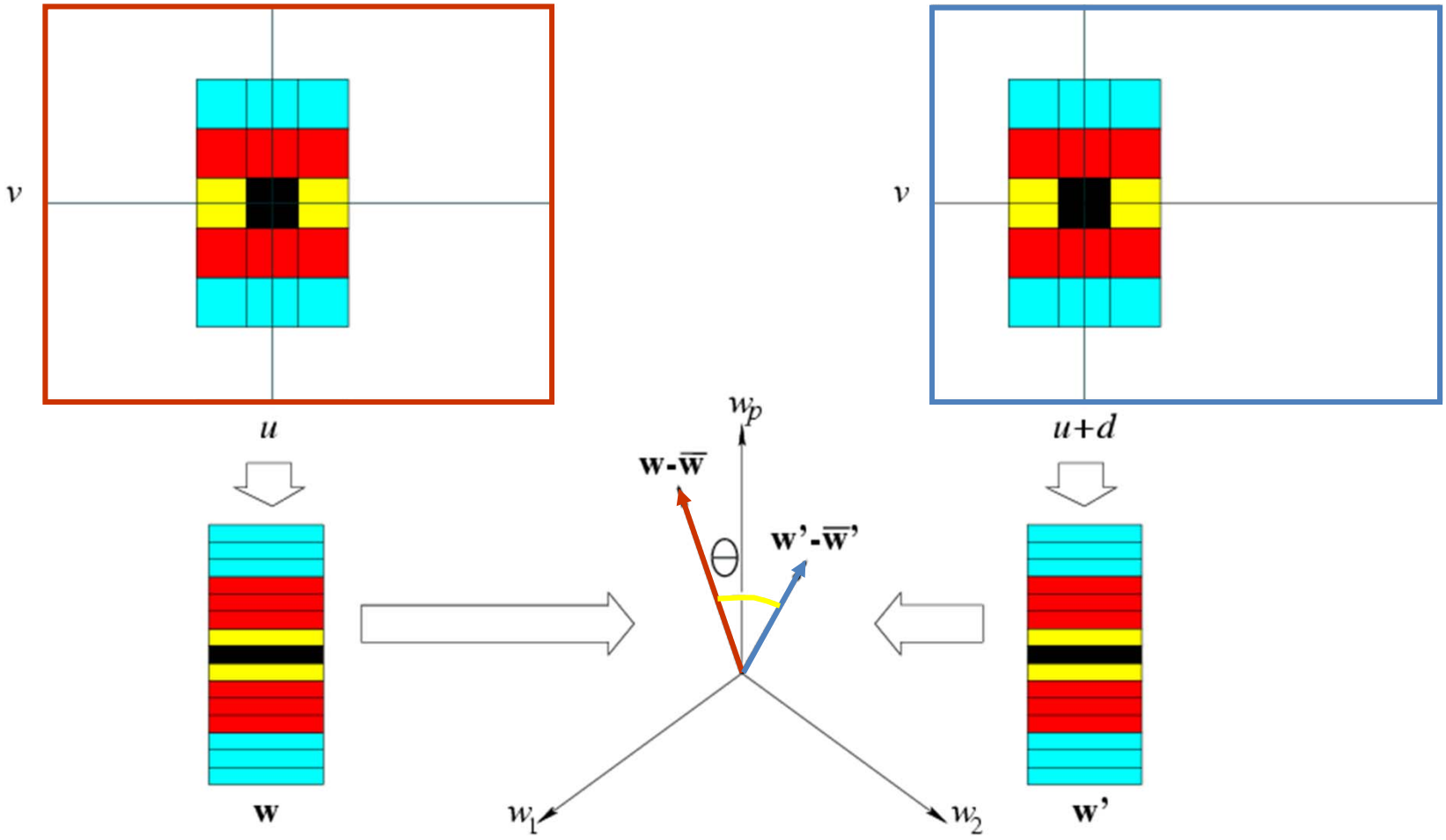
• Pick up a window around p(u,v)

Correlation Methods



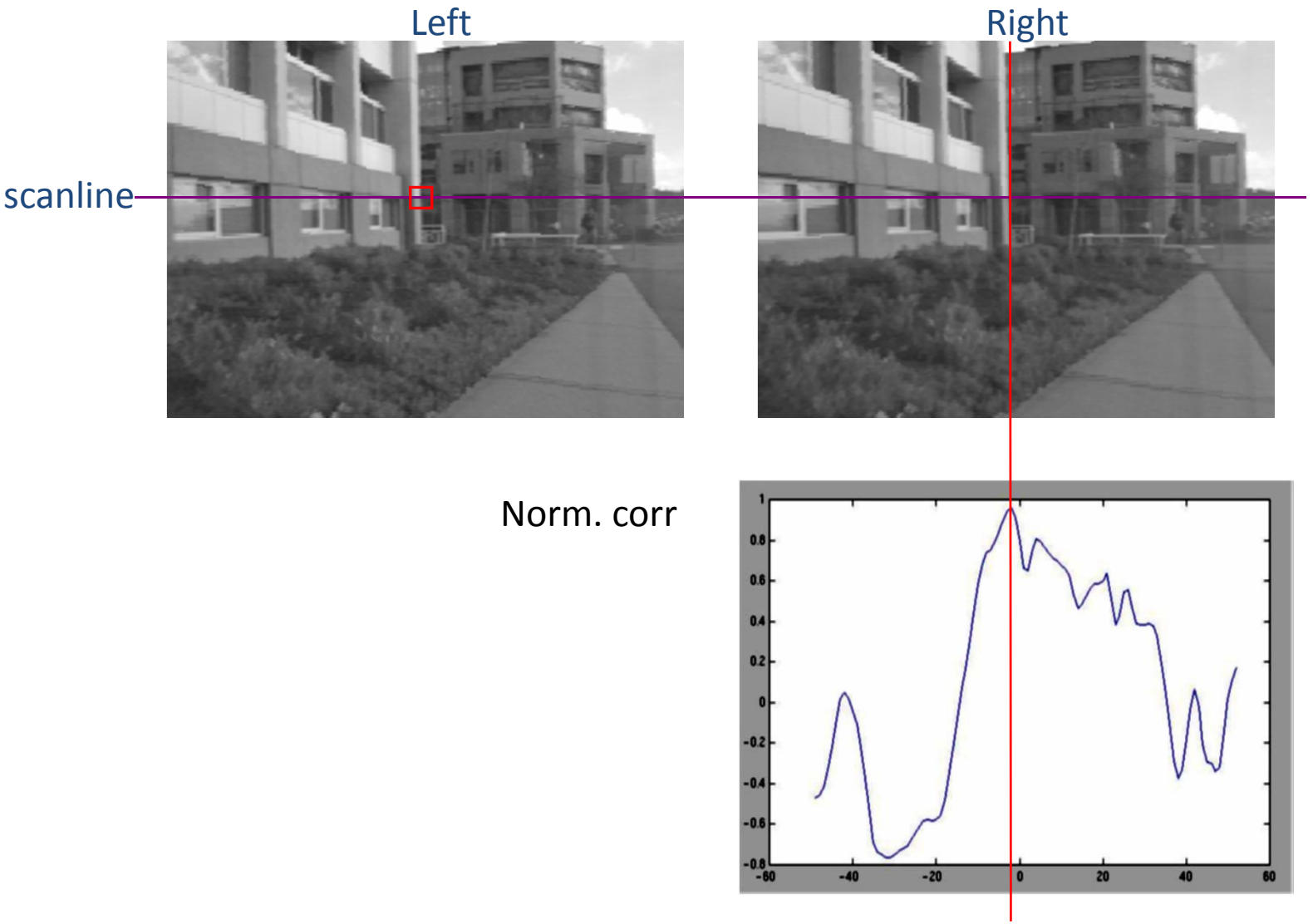
- Pick up a window around p(u,v)
- Build vector W
- Slide the window along v line in image 2 and compute w'
- Keep sliding until w·w' is maximized.

Correlation Methods



Normalized Correlation; minimize: $\dfrac{(\mathrm{w}-\overline{\mathrm{w}})(\mathrm{w}'-\overline{\mathrm{w}}')}{\left\|(\mathrm{w}-\overline{\mathrm{w}})(\mathrm{w}'-\overline{\mathrm{w}}')\right\|}$

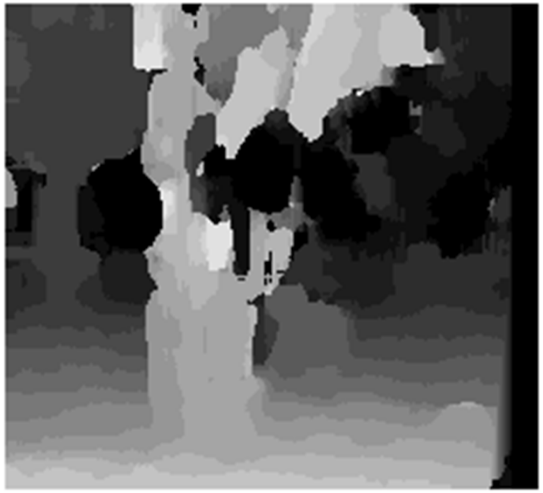
Correlation Methods



Credit slide S. Lazebnik

Correlation Methods

Window size = 3



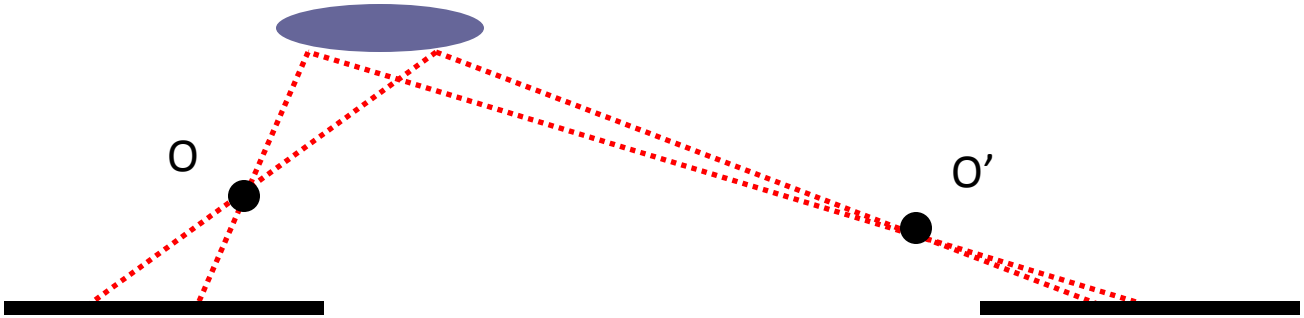
Window size = 20

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

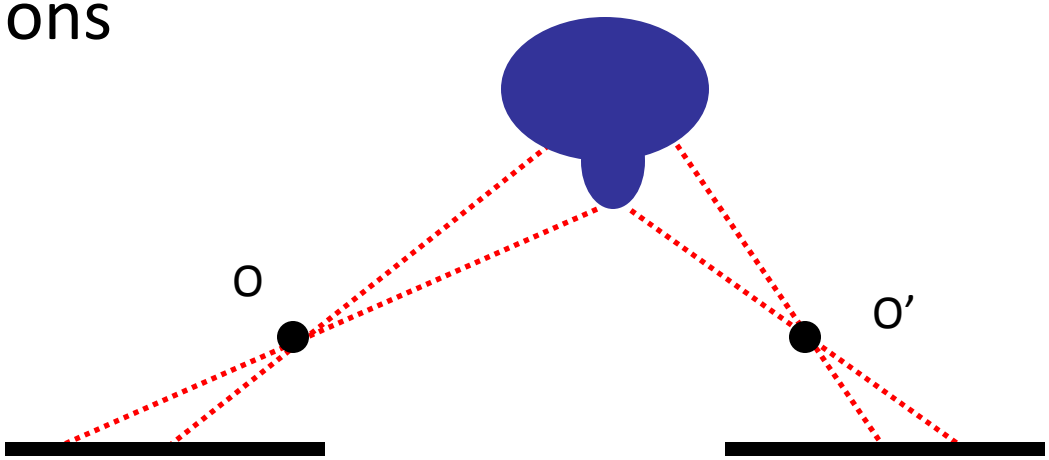
Credit slide S. Lazebnik

Issues with Correlation Methods

- Fore shortening effect



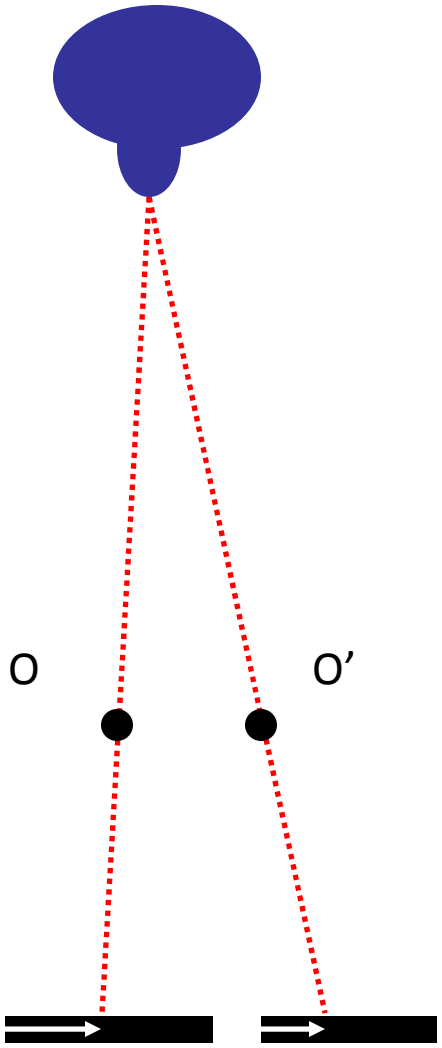
- Occlusions



It is desirable to have small B/z ratio!

Issues with Correlation Methods

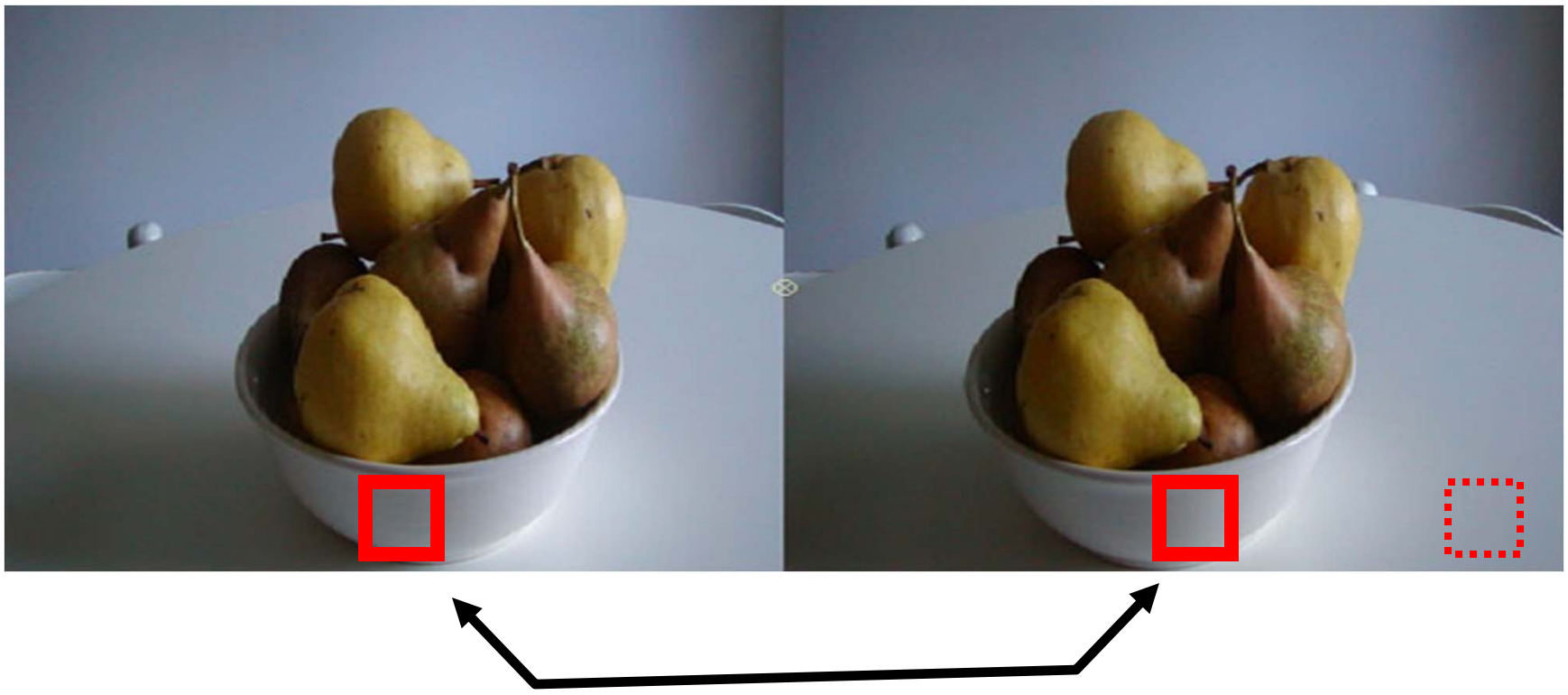
- small B/z ratio



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

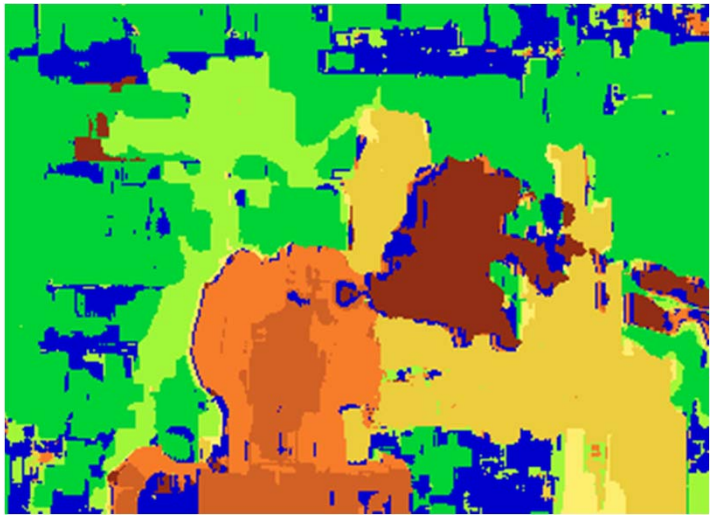
- Repetitive patterns

Results with window search

Data

Ground truth

Window-based matching



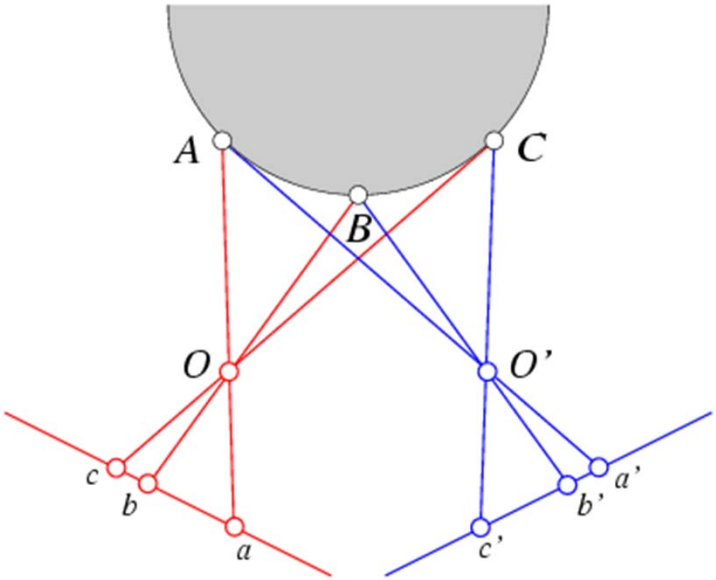
Credit slide S. Lazebnik

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

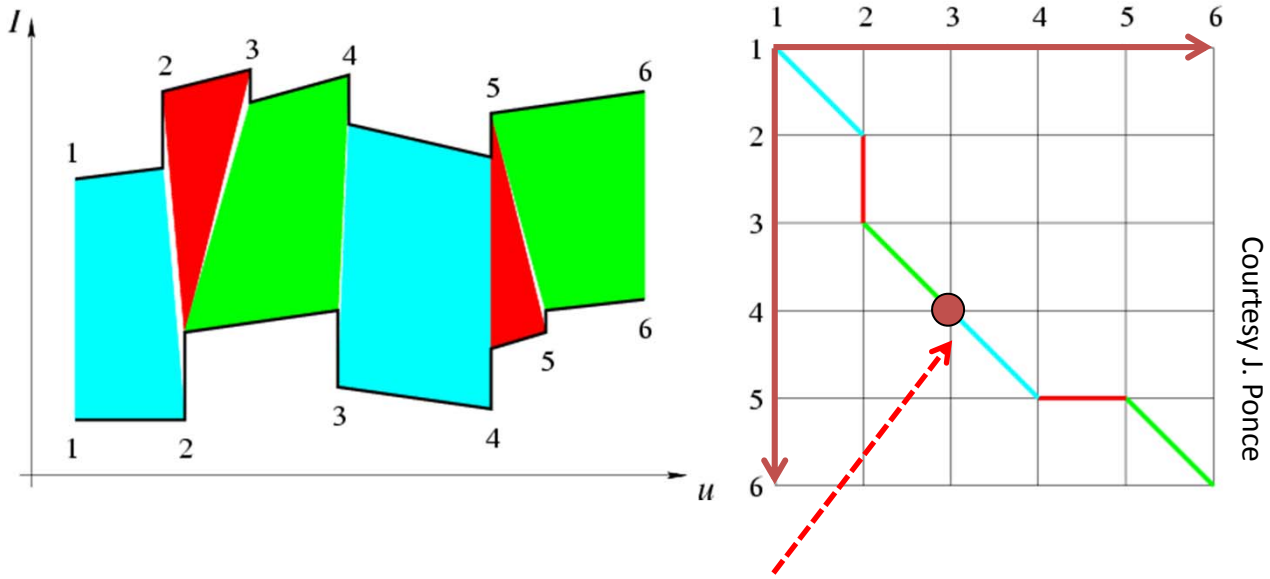


Courtesy J. Ponce

Dynamic Programming

[Uses ordering constraint]

(Baker and Binford, 1981)



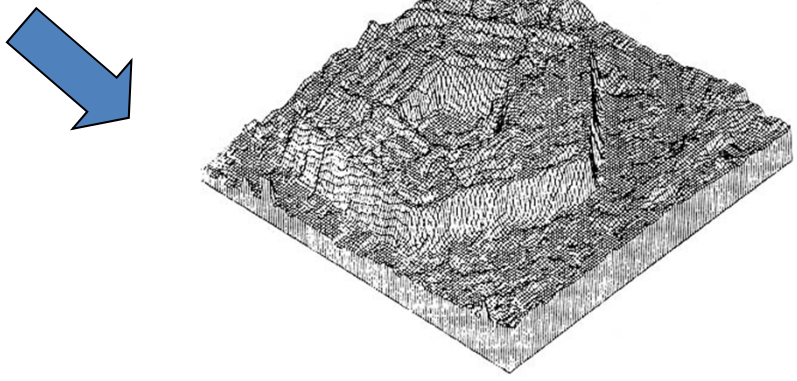
- Nodes = matched feature points (e.g., edge points).
- Arcs = matched intervals along the epipolar lines.
- Arc cost = discrepancy between intervals.

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)



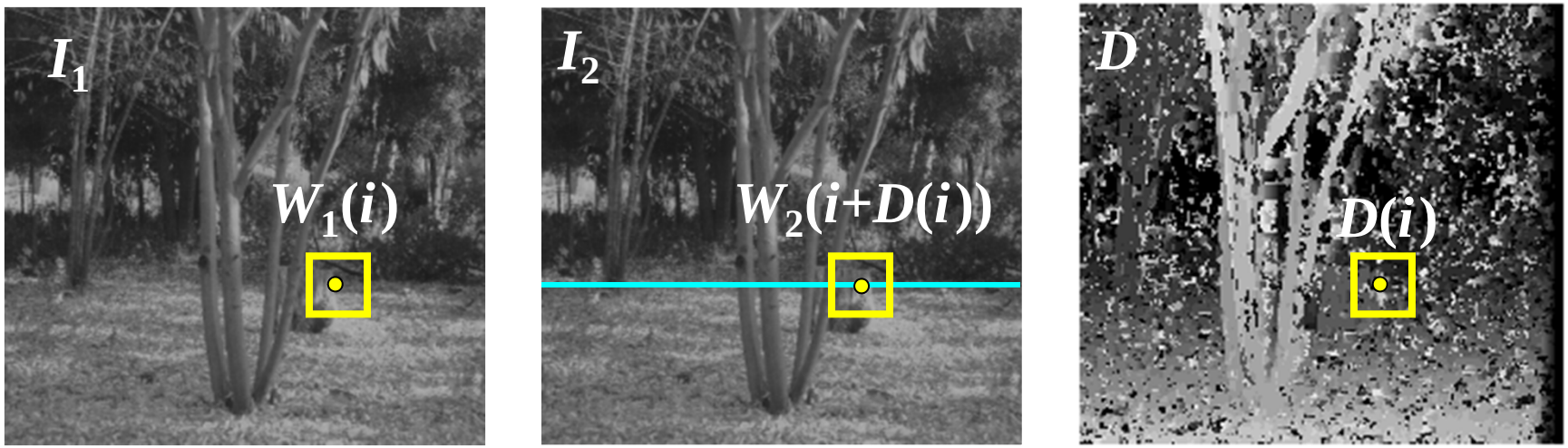
Improving correspondence: Non-local constraints

- 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



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

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

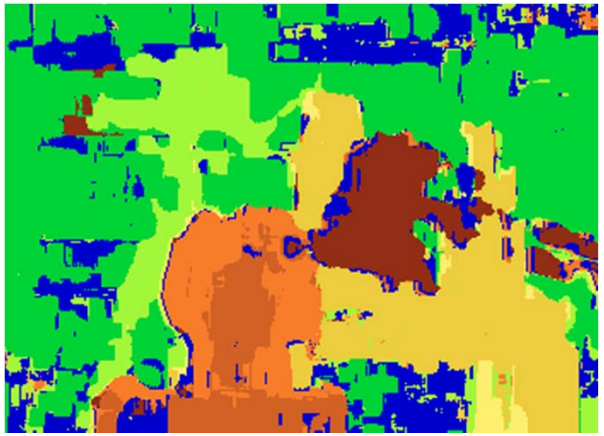
$$E_{\text{smooth}} = \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)$$

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

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

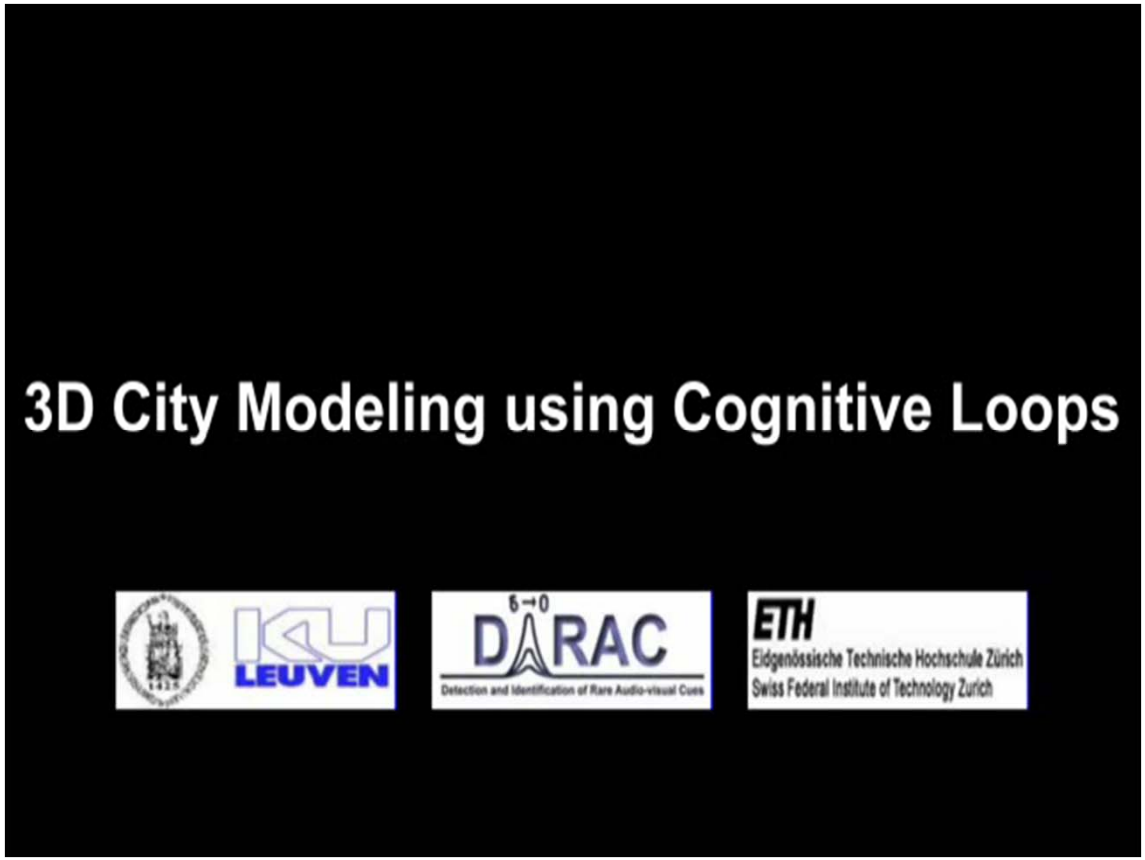
Foreground/background segmentation using stereo

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.



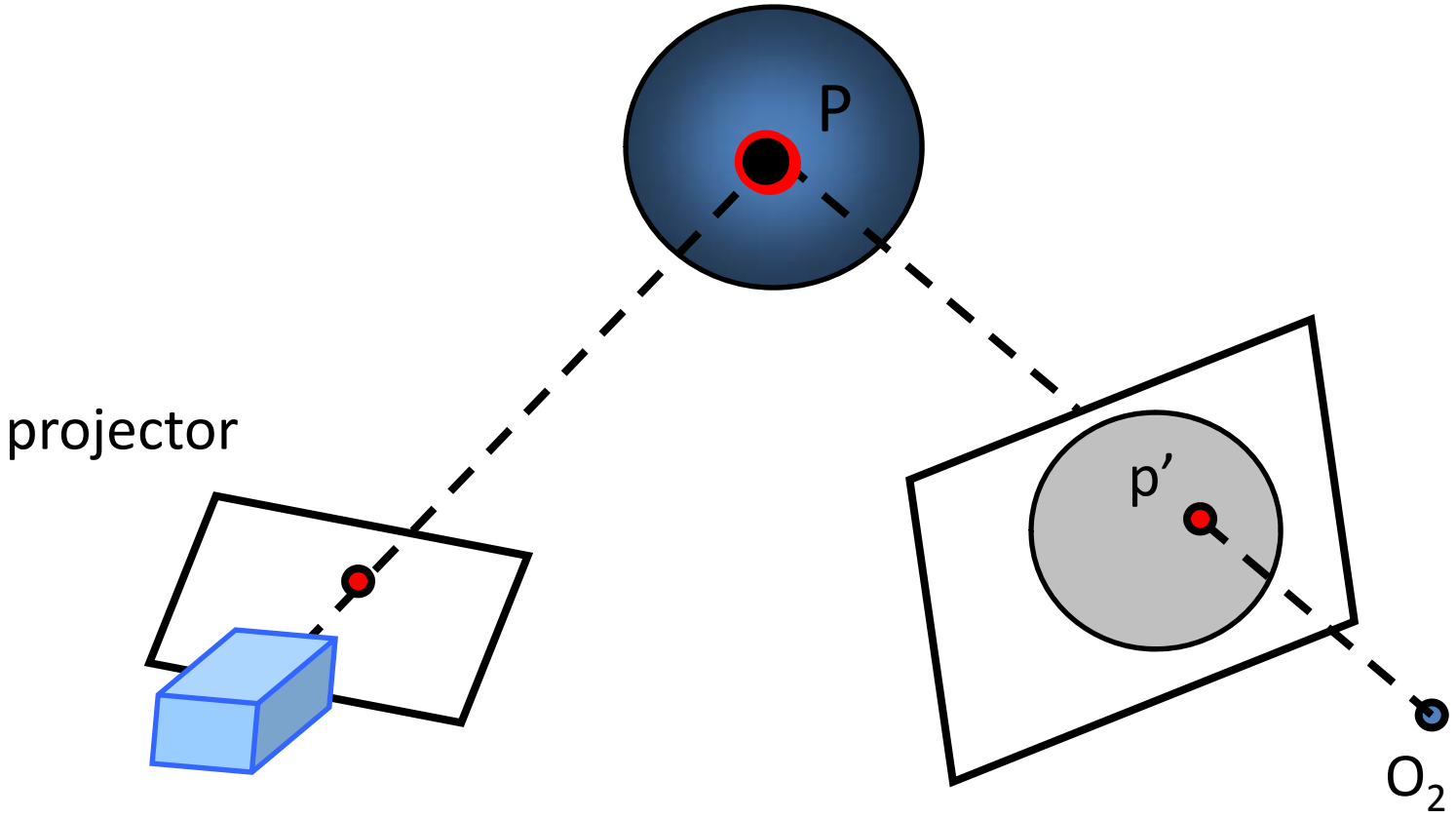
http://www.vision.ee.ethz.ch/showroom/index.en.html#

What we will learn today?

- Stereo vision
- Correspondence problem
- Active stereo vision systems
- Structure from motion

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

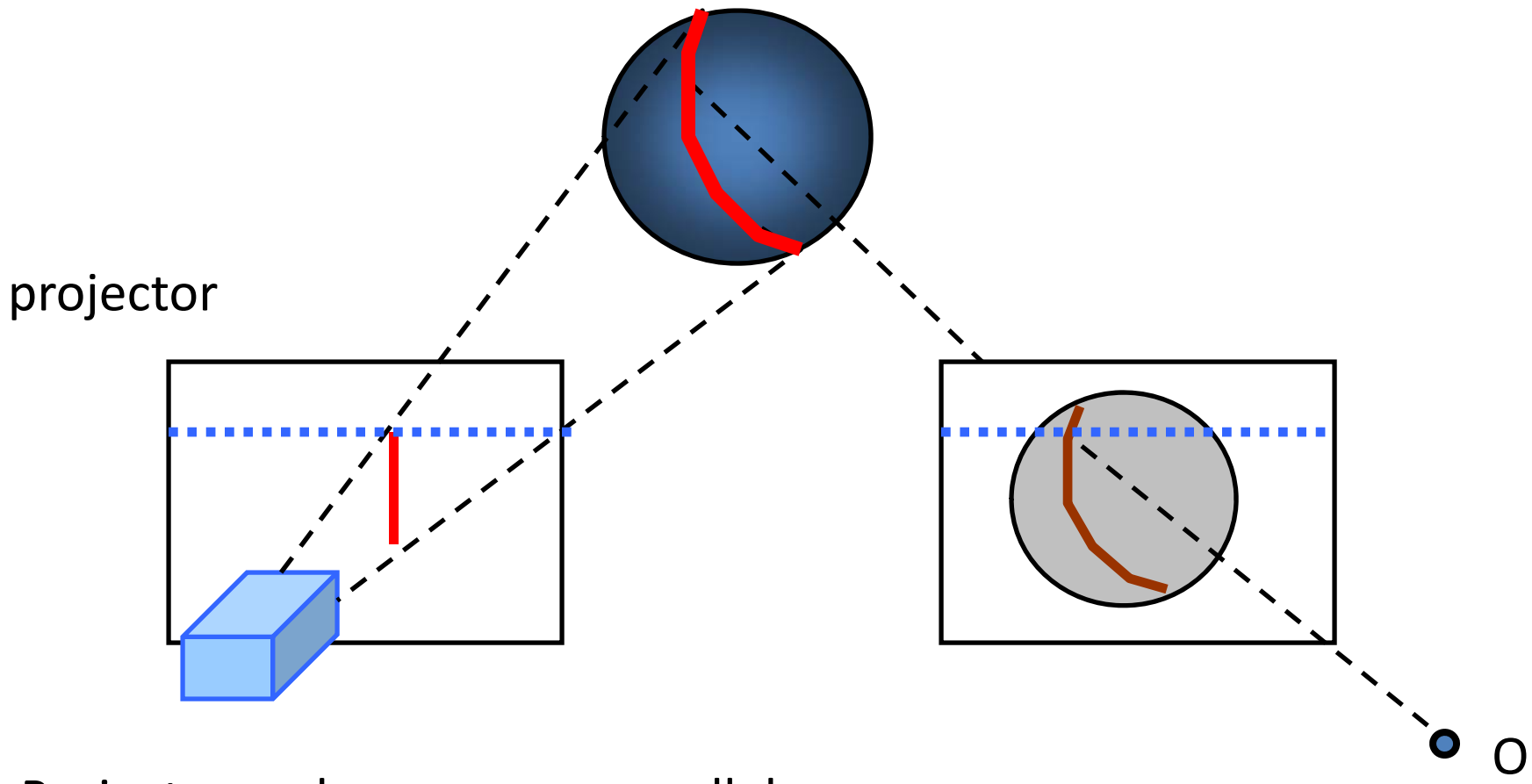
Active stereo (point)



Replace one of the two cameras by a projector

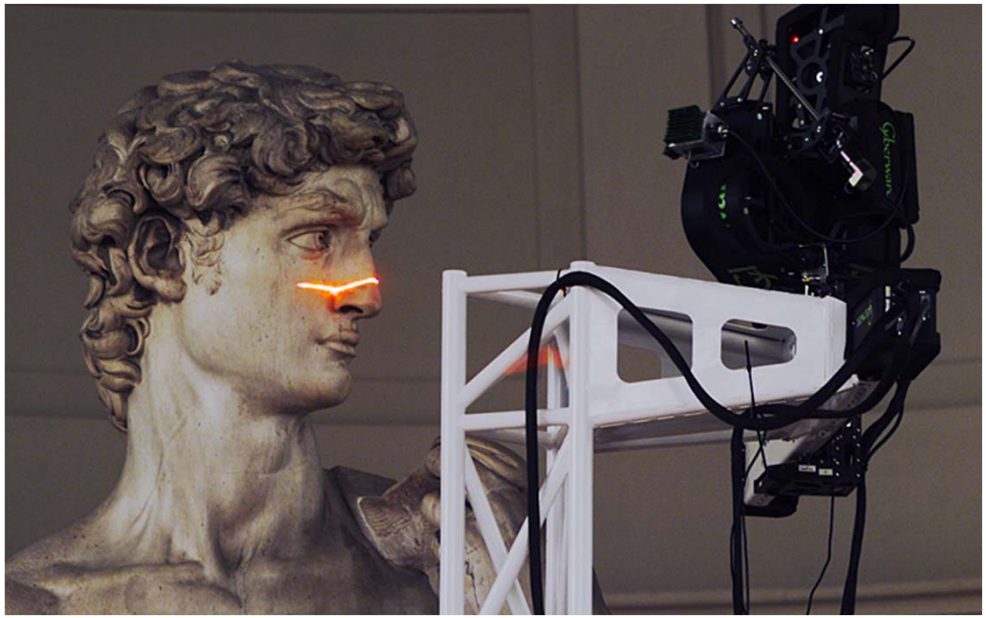
- Single camera
- Projector geometry calibrated
- What's the advantage of having the projector? Correspondence problem solved!

Active stereo (stripe)



-Projector and camera are parallel
- Correspondence problem solved!

Active stereo (stripe)



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

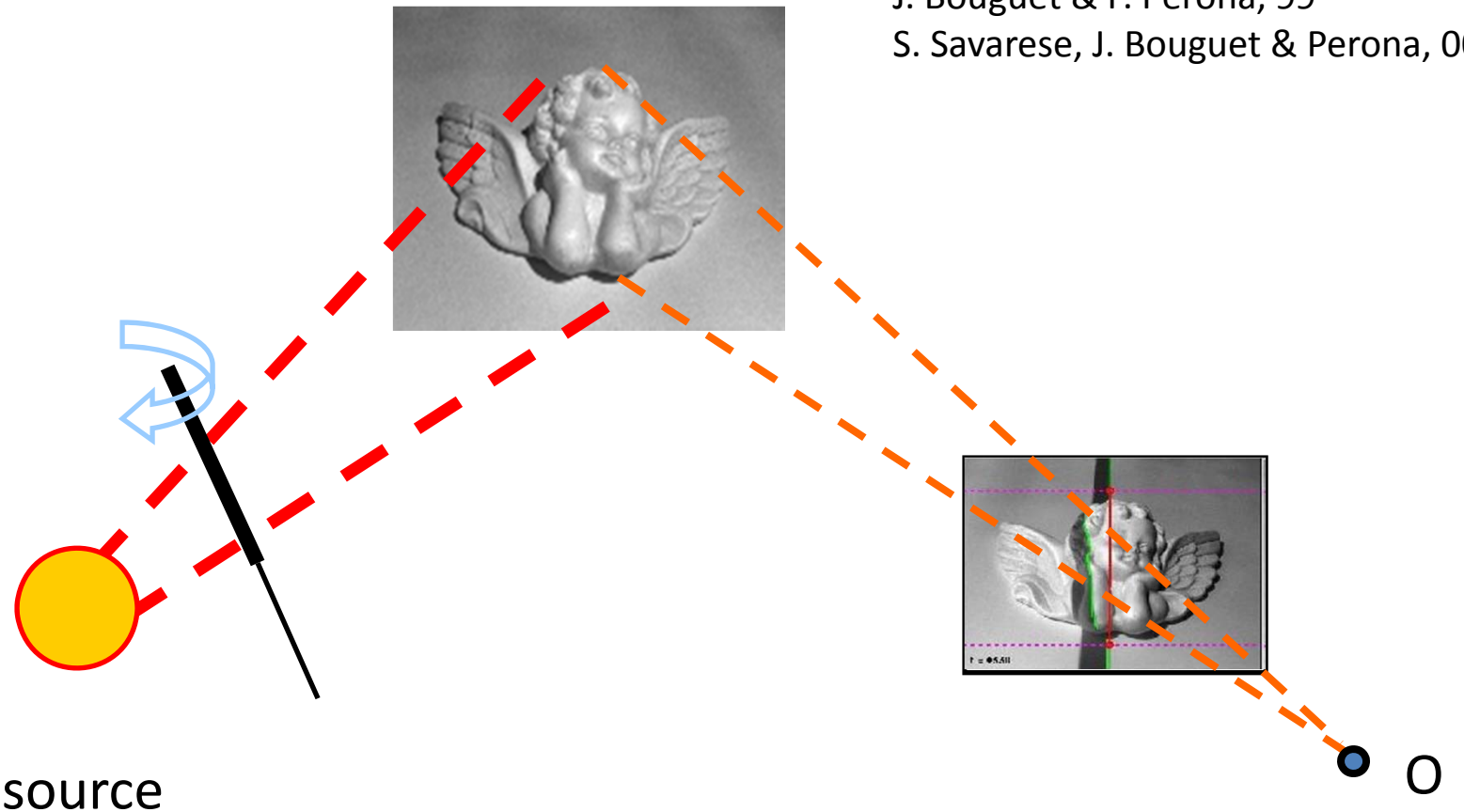
- Optical triangulation
  - Project a single stripe of laser light
  - Scan it across the surface of the object
  - This is a very precise version of structured light scanning

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

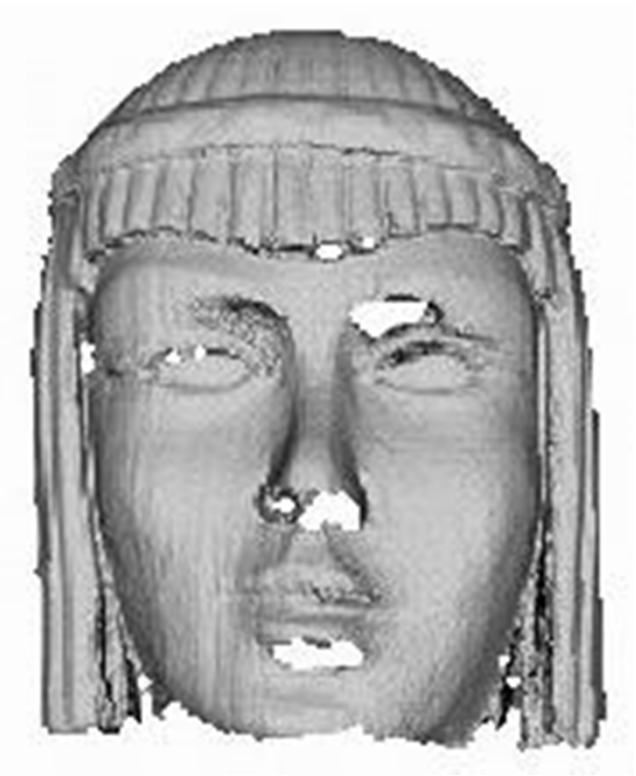


Light source

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

Active stereo (shadows)

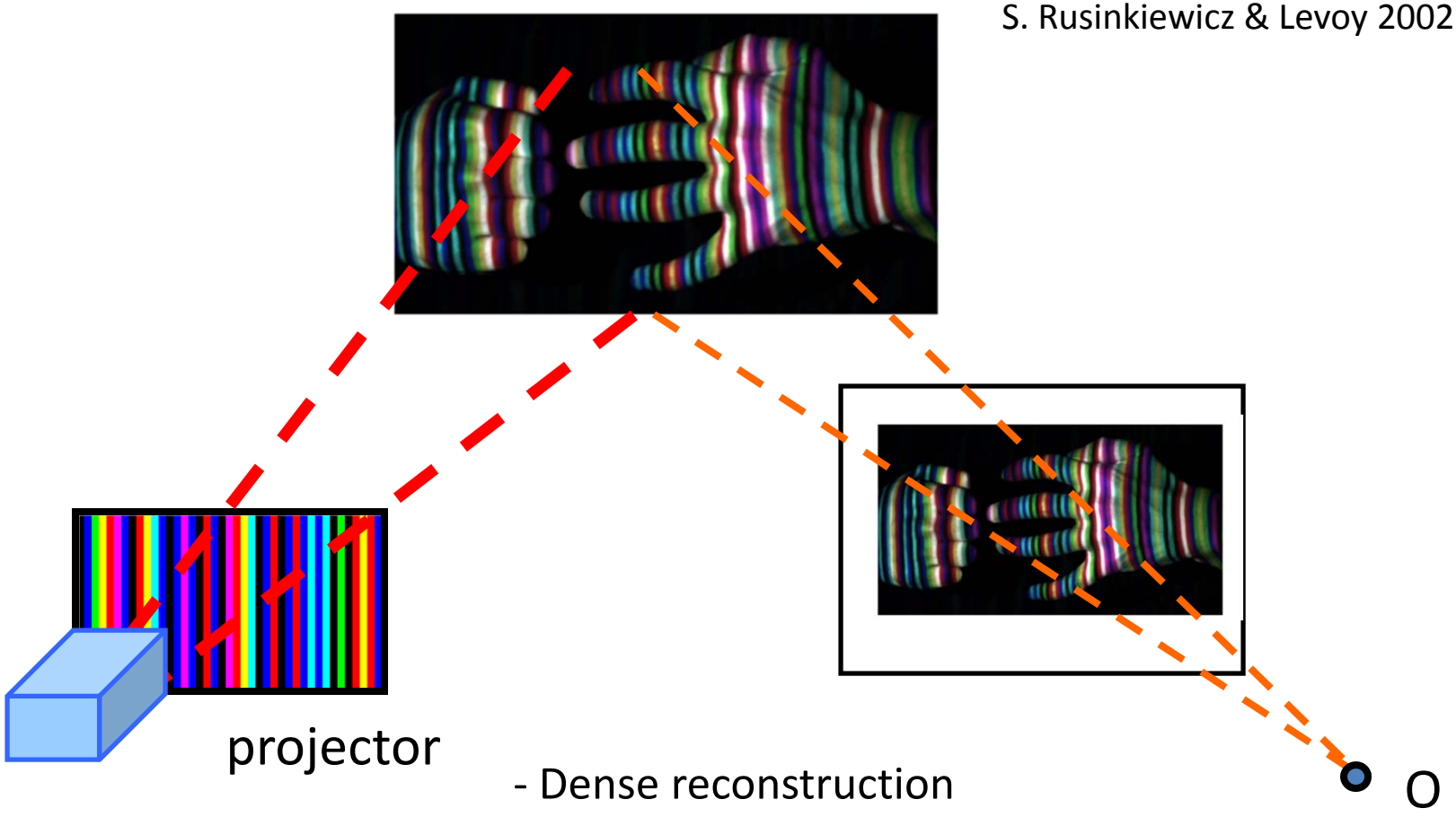
J. Bouguet & P. Perona, 99
S. Savarese, J. Bouguet & Perona, 00



Active stereo (color-coded stripes)

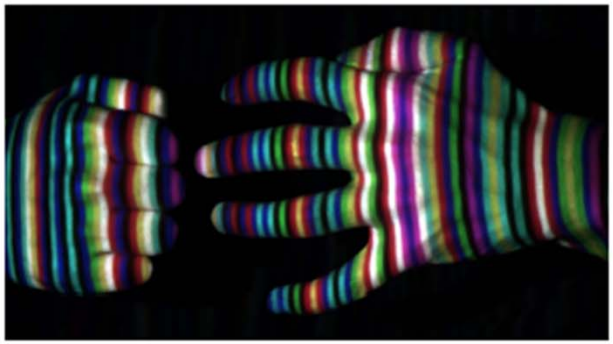
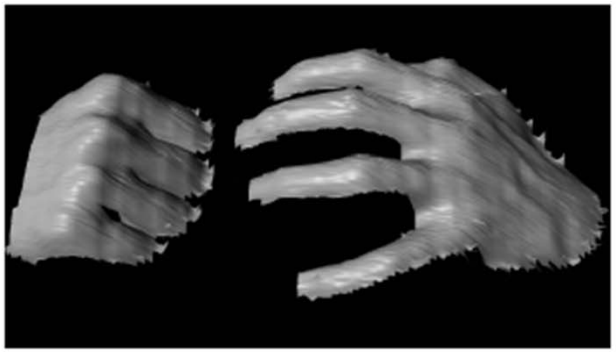
L. Zhang, B. Curless, and S. M. Seitz 2002

S. Rusinkiewicz & Levoy 2002



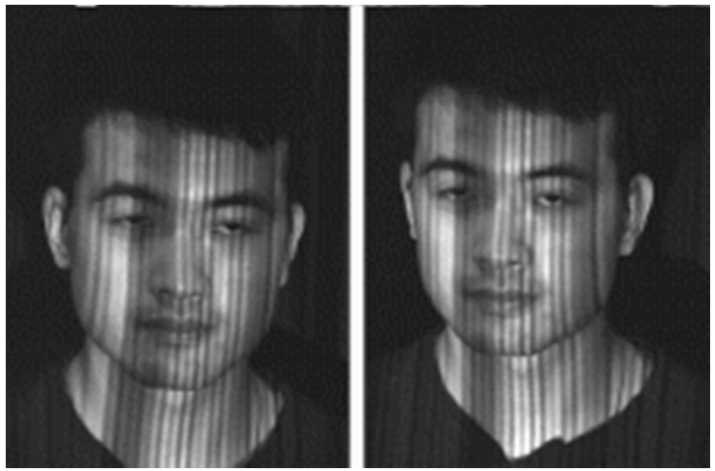
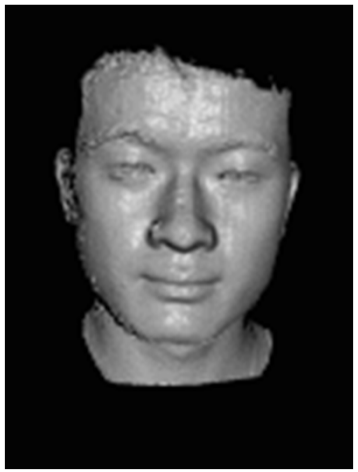
- Dense reconstruction
- Correspondence problem again
- Get around it by using color codes

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

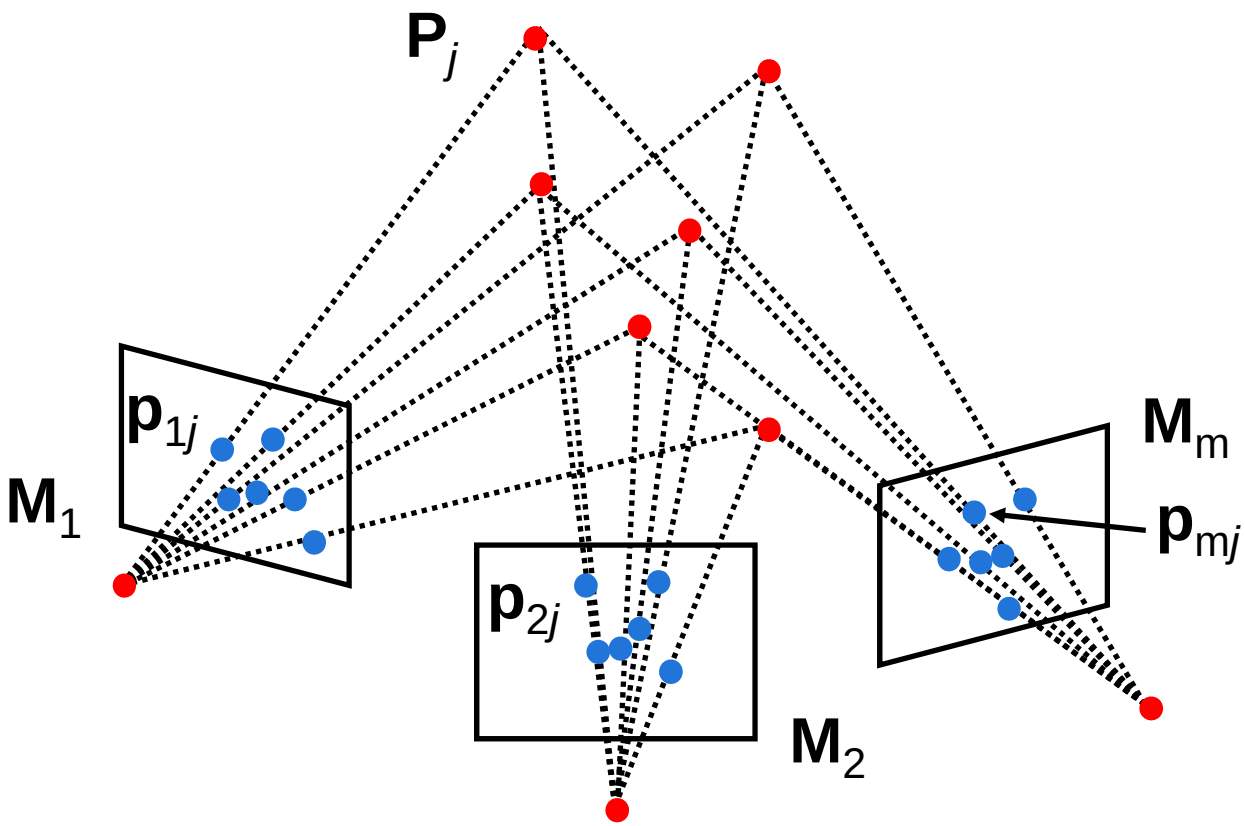



What we will learn today?

- Stereo vision
- Correspondence problem
- Active stereo vision systems
- Structure from motion

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

Multiple view geometry - the structure from motion problem

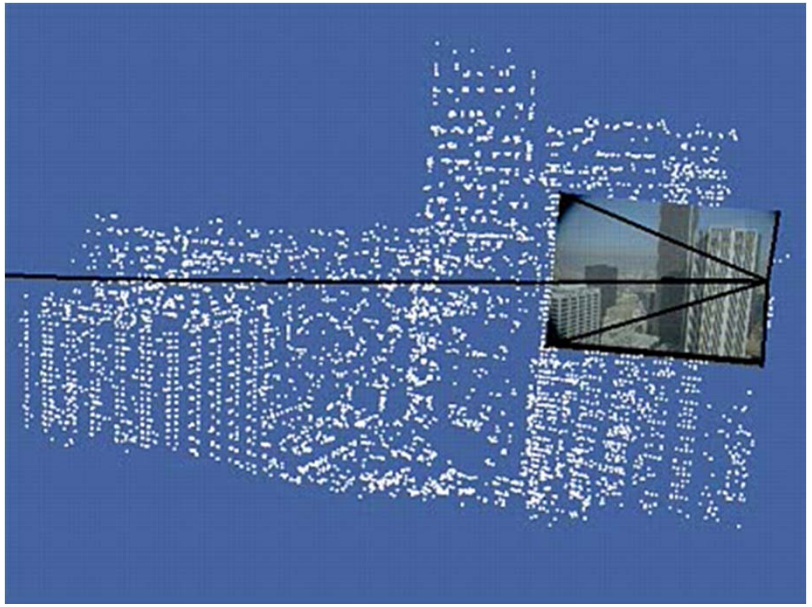


Given *m* images of *n* fixed 3D points;

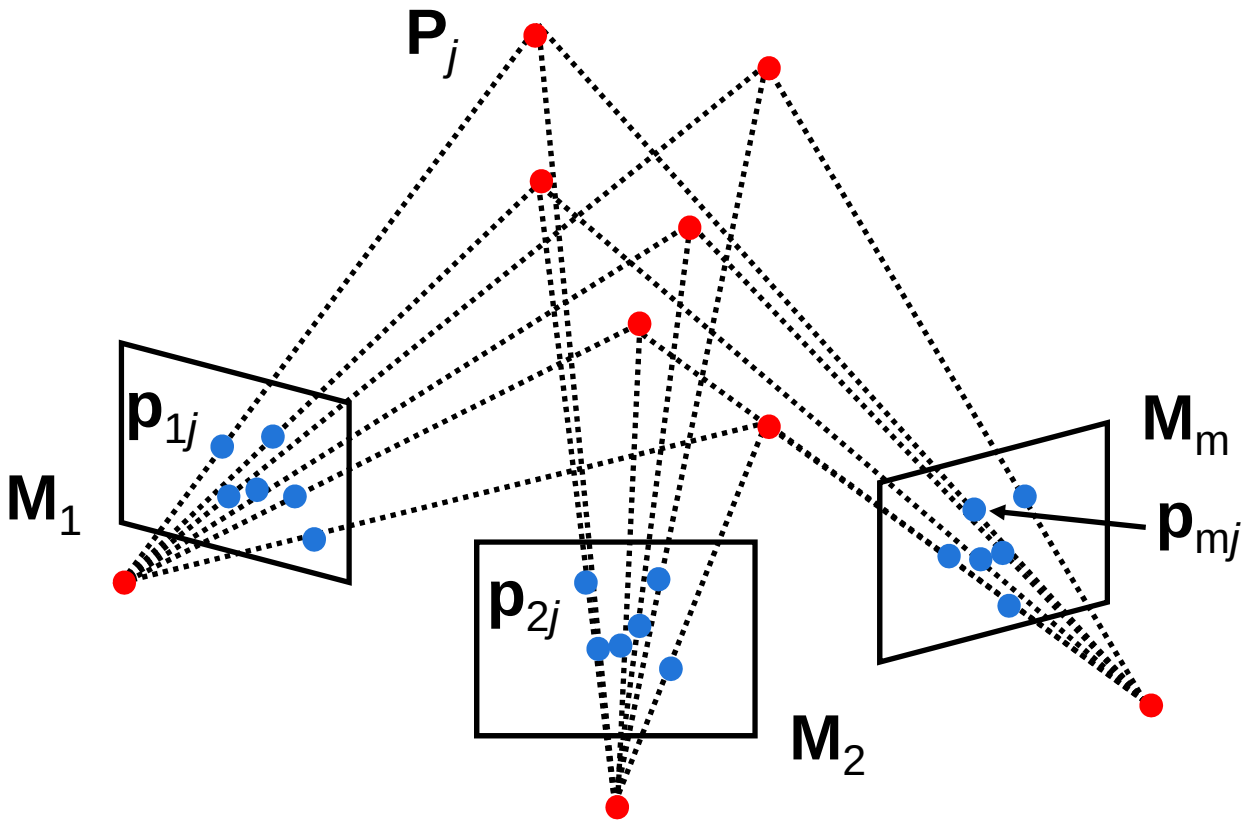
$$\mathbf{p}_{ij} = \mathbf{M}_i \mathbf{P}_j, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$$

Multiple view geometry - the structure from motion problem

Courtesy of Oxford **Visual Geometry Group**



Multiple view geometry - the structure from motion problem

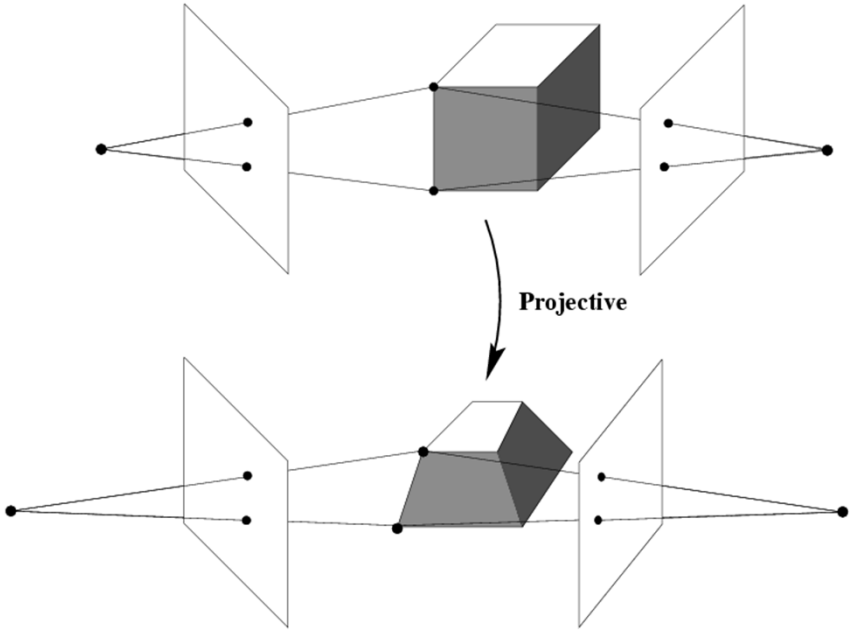


From the mxn correspondences $\mathbf{p}_{ij}$, estimate:

•*m* projection matrices $\mathbf{M}_i$	motion
•*n* 3D points $\mathbf{P}_j$	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**



$$p_j = M_i \, P_j$$

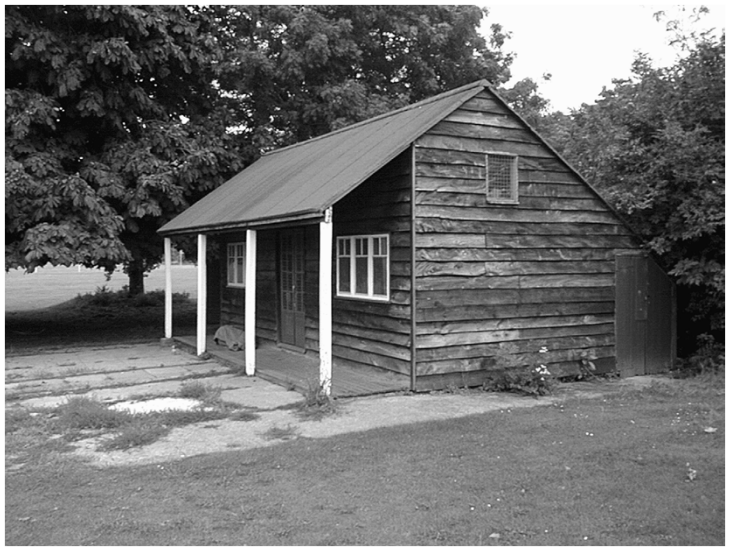
$$M_i \rightarrow H \, P_j$$

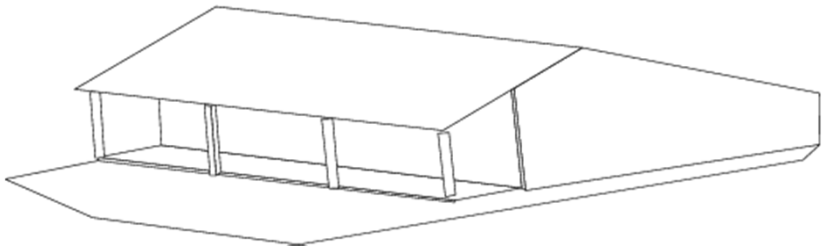
$$M_i = K_i \begin{bmatrix} R_i & T_i \end{bmatrix}$$

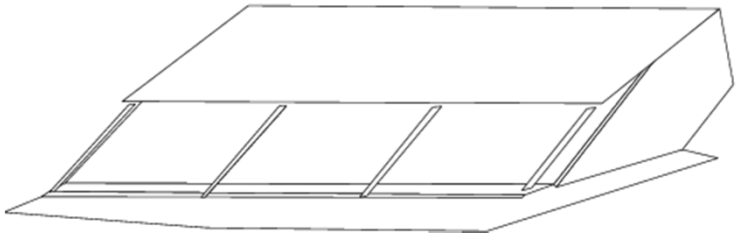
$$M_j \, H^{-1}$$

$$p_j = M_i \, P_j = \left(M_i \, H^{-1}\right)\left(H \, P_j\right)$$

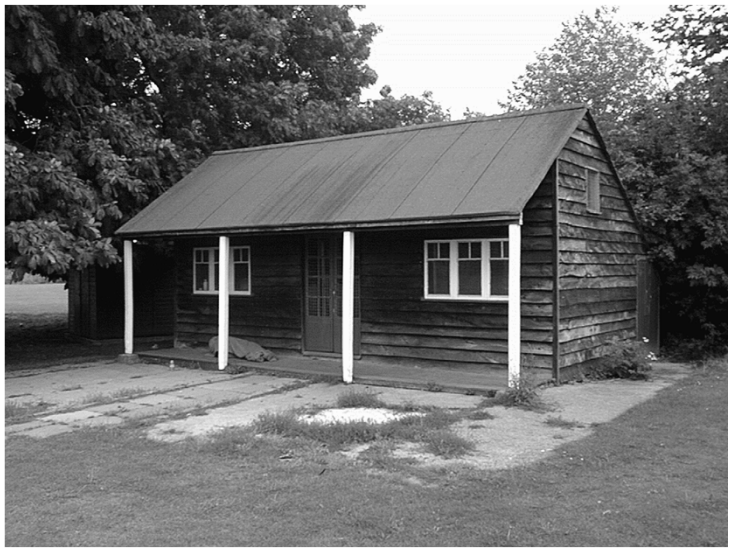
Affine ambiguity

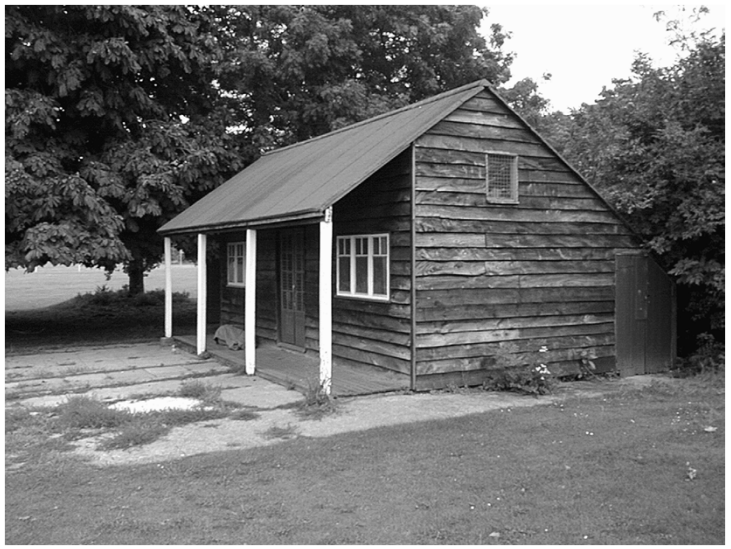


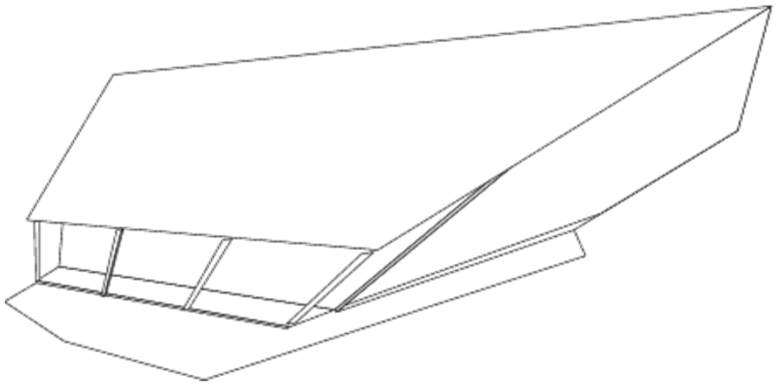


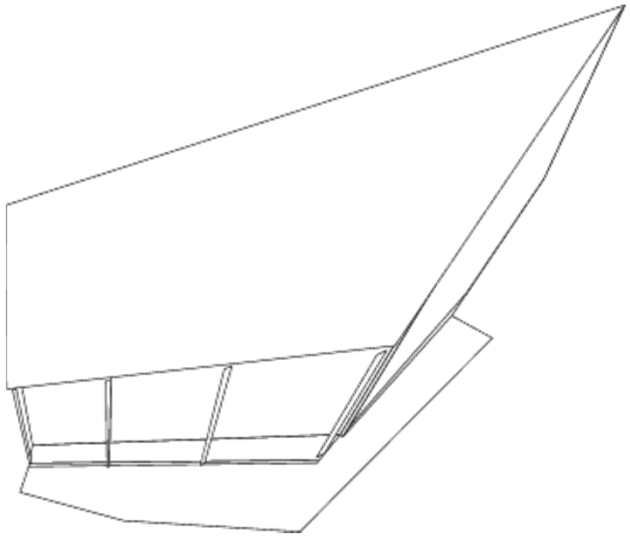


Prospective ambiguity







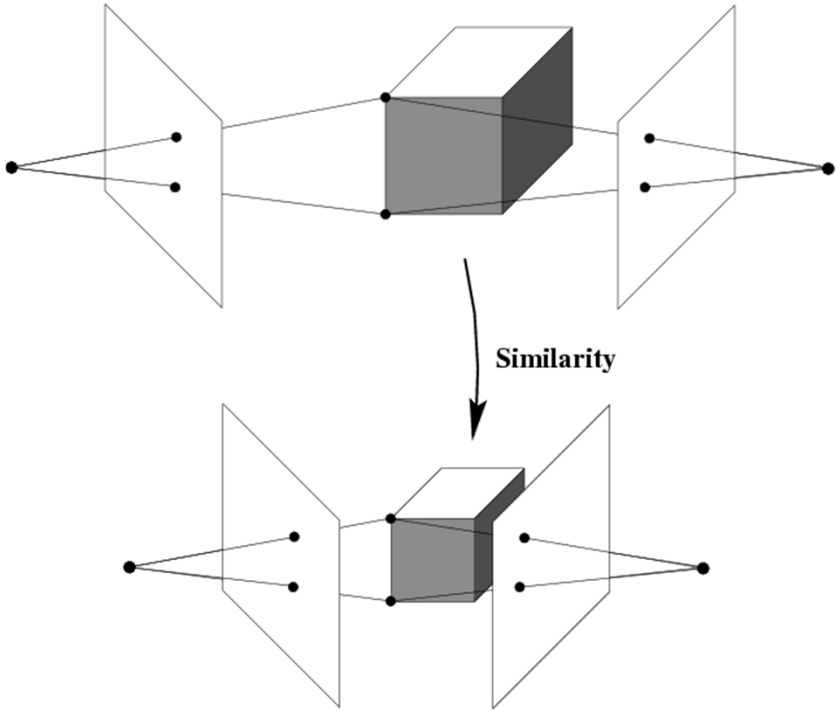


Structure from motion ambiguity

- A **scale (similarity) ambiguity** exists even for calibrated cameras
- For calibrated cameras, the similarity ambiguity is the only ambiguity

[Longuet-Higgins '81]

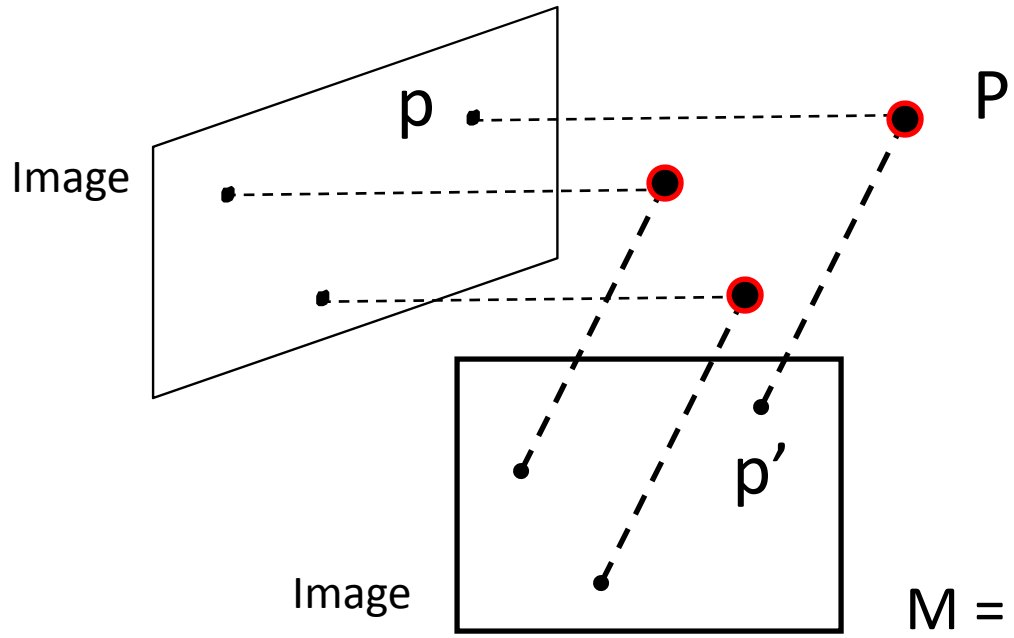
- A reconstruction up to scale is called **metric**



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

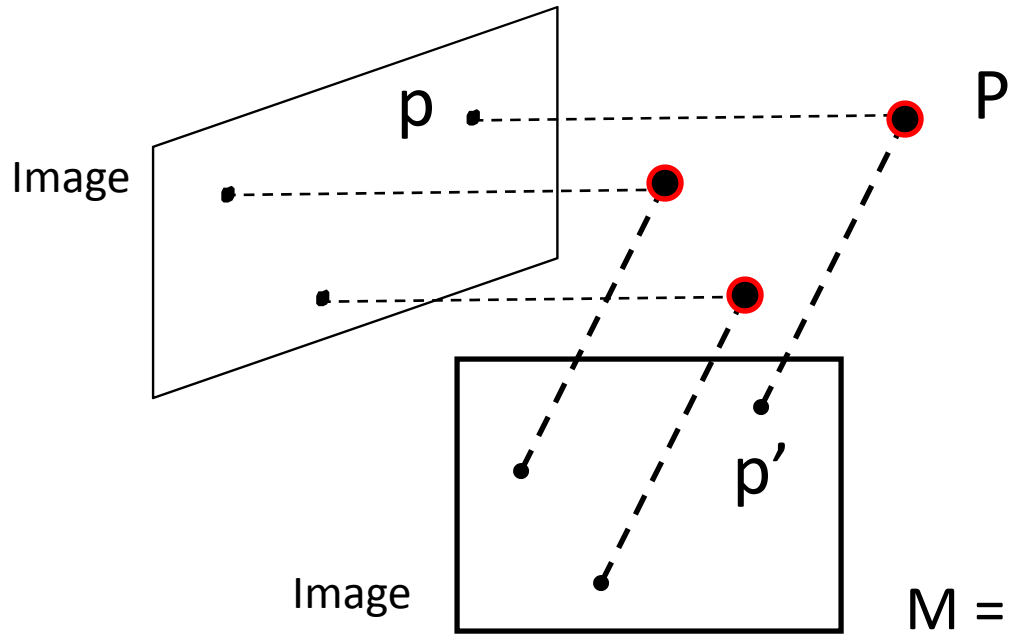


M = camera matrix

From the mxn 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



M = camera matrix

$$\mathbf{p} = \begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{AP} + \mathbf{b} = M \begin{bmatrix} \mathbf{P} \\ 1 \end{bmatrix}; \qquad \mathrm{M} = \begin{bmatrix} \mathrm{A} & \mathrm{b} \end{bmatrix}$$

Affine structure from motion - a simpler problem

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

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

N of cameras N of points

Problem: estimate the m 2×4 matrices M_i and
the n positions P_j from the m×n correspondences $\mathbf{p}_{ij}$.

How many equations and how many unknown?

2m × n equations in 8m+3n unknowns

Two approaches:

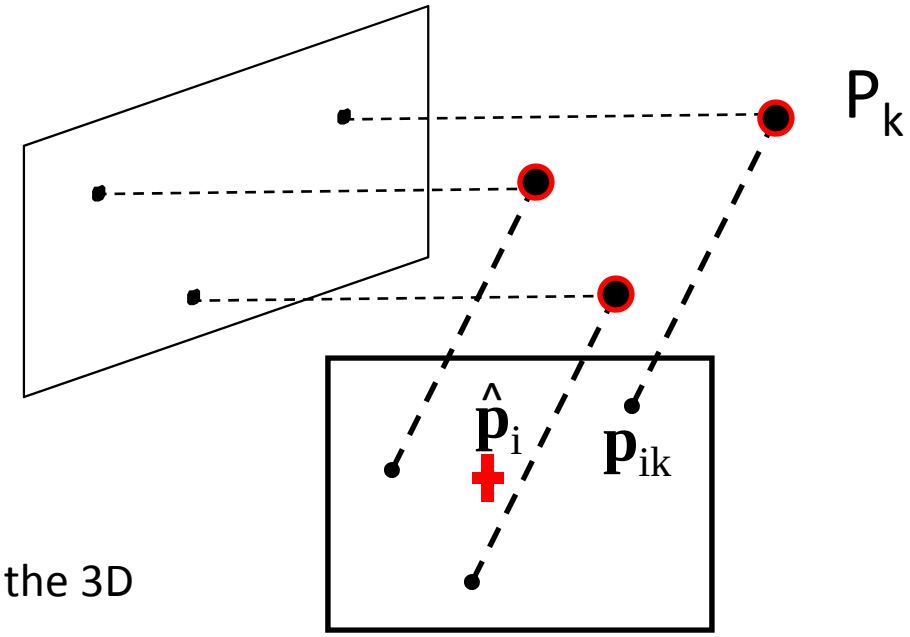
- Algebraic approach (affine epipolar geometry; estimate F; cameras; points)
- Factorization method

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

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

$$= \mathbf{A}_i\left(P_j - \frac{1}{n}\sum_{k=1}^{n} P_k\right) = \mathbf{A}_i \hat{P}_j$$

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

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

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 × n data (measurement) matrix:

$$\mathbf{D} = \begin{bmatrix} \hat{p}_{11} & \hat{p}_{12} & \cdots & \hat{p}_{1n} \\ \hat{p}_{21} & \hat{p}_{22} & \cdots & \hat{p}_{2n} \\ & & \ddots & \\ \hat{p}_{m1} & \hat{p}_{m2} & \cdots & \hat{p}_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_m \end{bmatrix}_{\mathrm{M}} \overbrace{\begin{bmatrix} P_1 & P_2 & \cdots & P_n \end{bmatrix}}^{\mathrm{S}}$$

(2*m* × n)

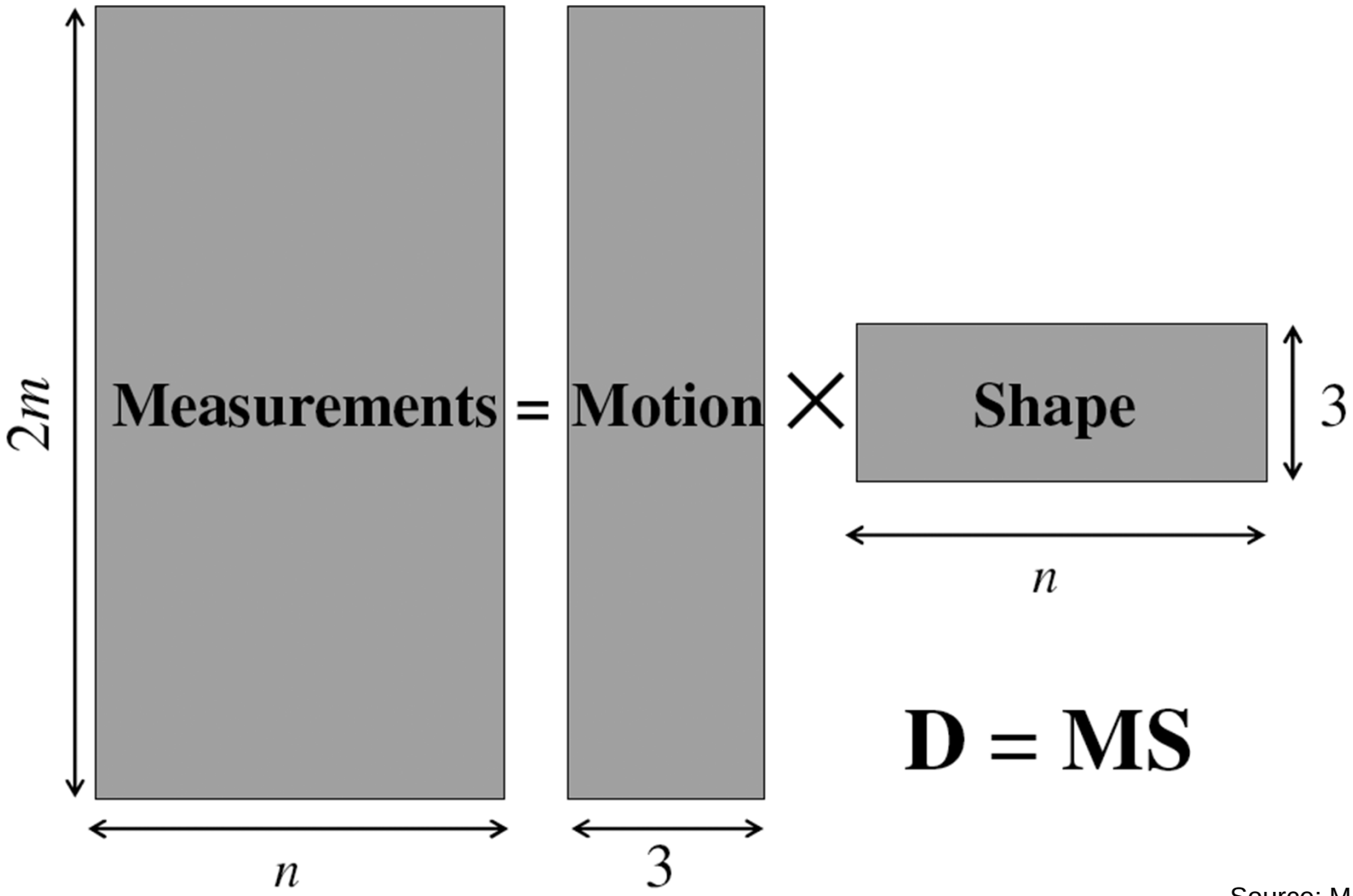
cameras (2*m* × 3)

points (3 × *n*)

The measurement matrix **D = M S** has rank 3

(it's a product of a 2mx3 matrix and 3xn matrix)

Factorization method

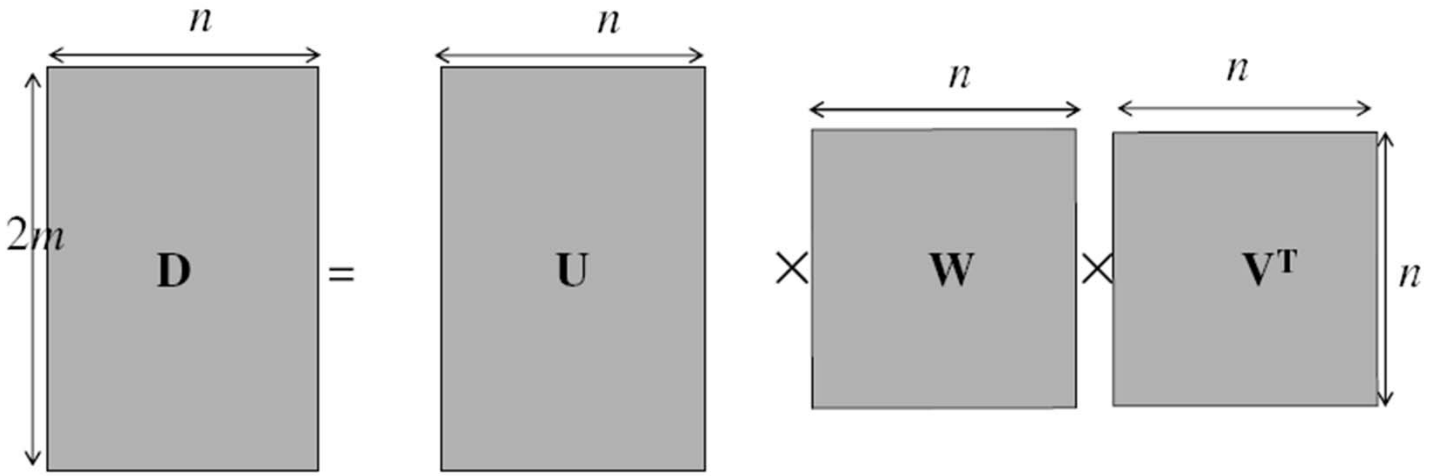


$$\mathbf{D} = \mathbf{MS}$$

Source: M. Hebert

Factorization method

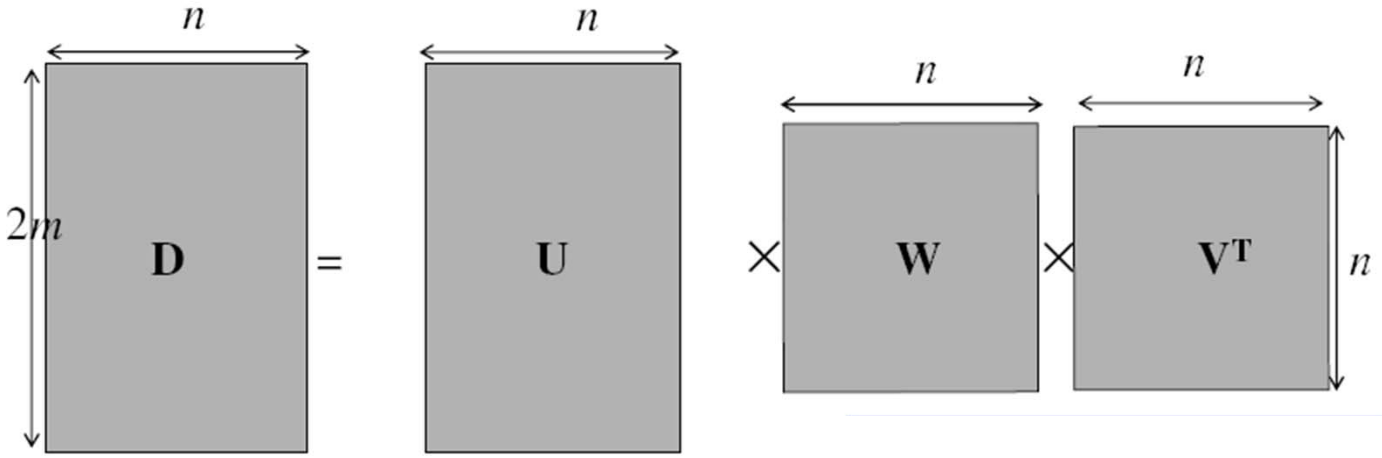
- Singular value decomposition of D:



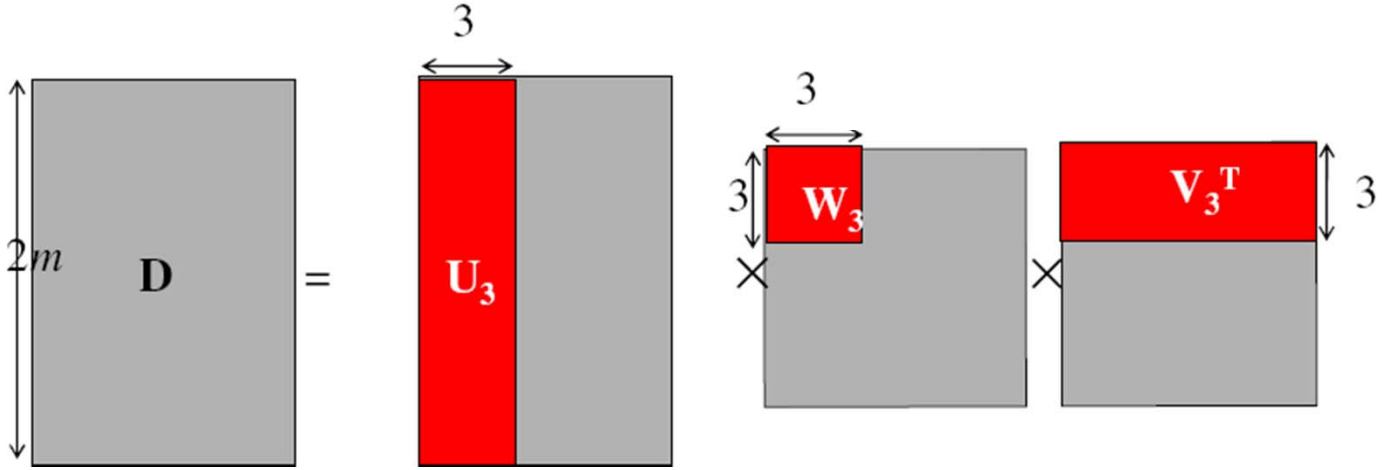
Source: M. Hebert

Factorization method

- Singular value decomposition of D:



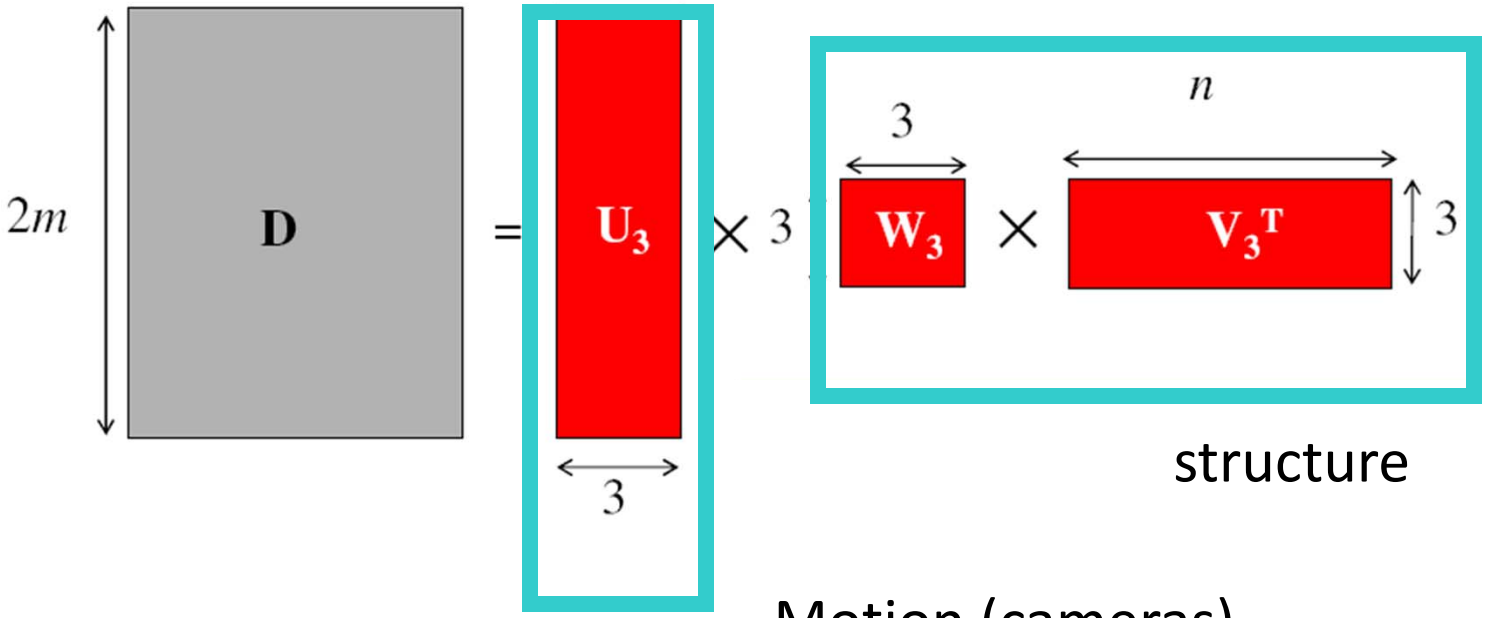
Since rank (D)=3, there are only 3 non-zero singular values



Source: M. Hebert

Factorization method

- Obtaining a factorization from SVD:



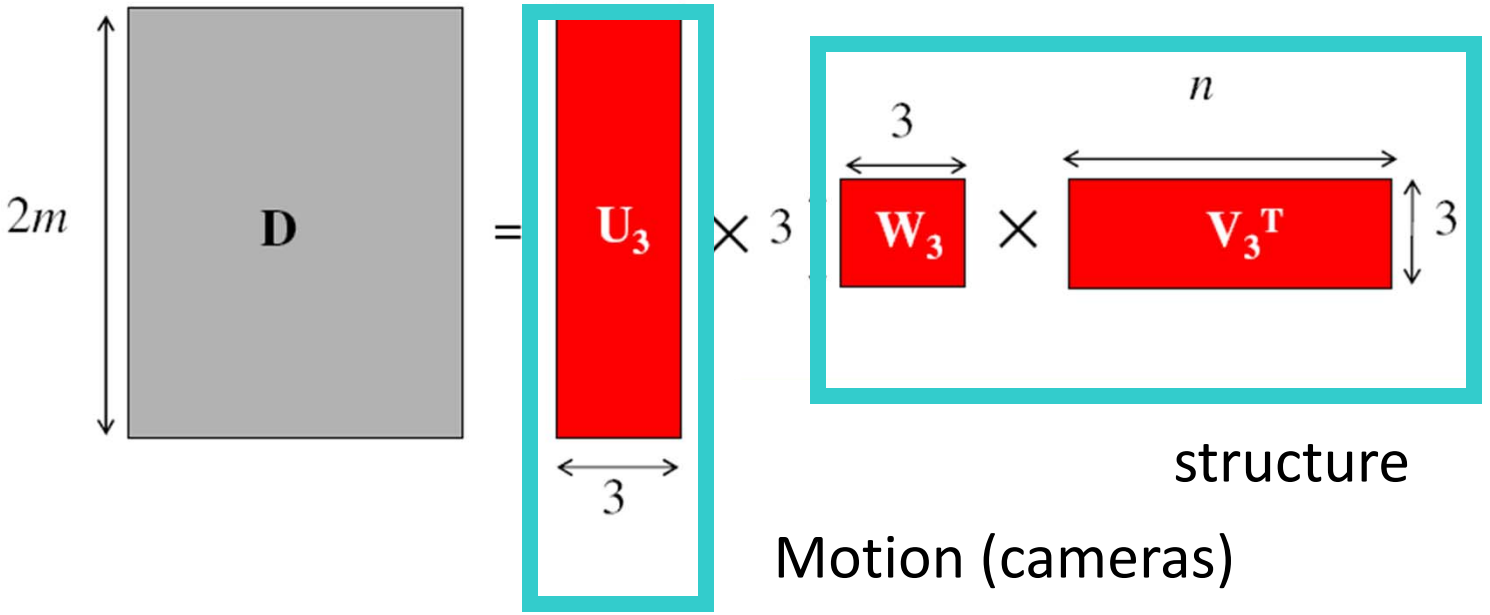
What is the issue here?

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, c*enter the feature coordinates
3. Construct a $2m \times n$ measurement matrix **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 **D**:
   - Compute SVD: $\mathbf{D} = \mathbf{U}\,\mathbf{W}\,\mathbf{V}^{\mathbf{T}}$
   - Create $\mathbf{U}_3$ by taking the first 3 columns of **U**
   - Create $\mathbf{V}_3$ by taking the first 3 columns of **V**
   - Create $\mathbf{W}_3$ by taking the upper left 3×3 block of **W**
5. Create the motion and shape matrices:
   - $\mathbf{M} = \mathbf{M} = \mathbf{U}_3$ and $\mathbf{S} = \mathbf{W}_3\mathbf{V}_3^{\mathrm{T}}$ (or $\mathbf{U}_3\mathbf{W}_3^{½}$ and $\mathbf{S} = \mathbf{W}_3^{½}\,\mathbf{V}_3^{\mathrm{T}}$)

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*

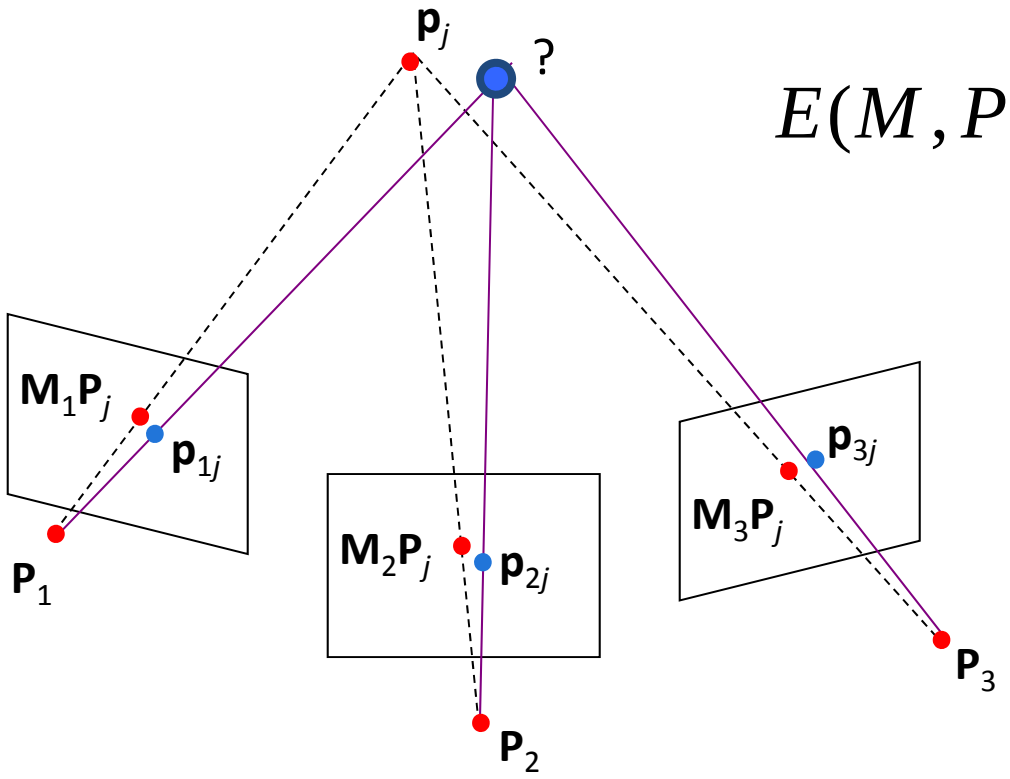
Multiple view geometry - overview

- 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



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

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

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

- 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?

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

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