https://tinyurl.com/4wwbdbe2

Computer Vision 1

HC6b

Shape from X, Stereo, Structure from Motion

Dr. Martin Oswald, Dr. Dimitris Tzionas, Dr. Arun Mukundan,
(m.r.oswald, d.tzionas, a.mukundan)@uva.nl

University of Amsterdam

computer vision lab

Week 7 - Guest Lectures

dr. Vincent Leroy
Research Scientist
Naver Labs Europe

Tuesday

dr. Sezer Karaoglu
3DUniversum
Co-Founder & CTO

Tuesday

dr. Javier Romero
Research Scientist
Meta Reality Labs

Thursday

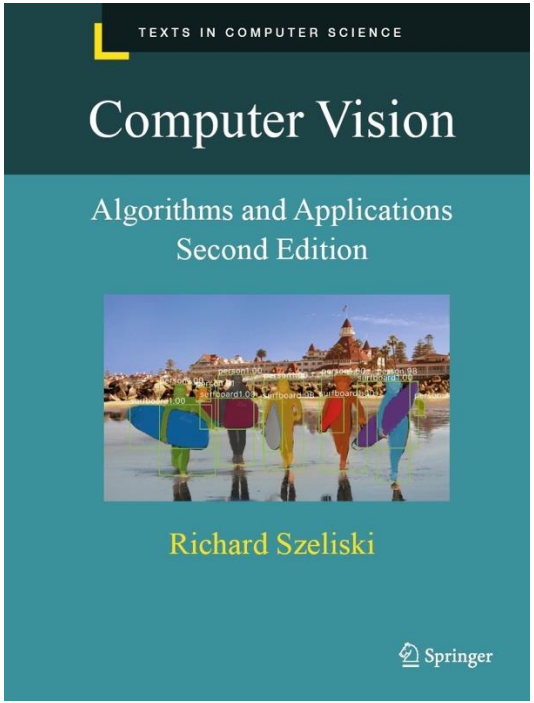
Very easy questions from each of the speakers talk will be in the exam, each lecture worth 1 pt

Today's Agenda

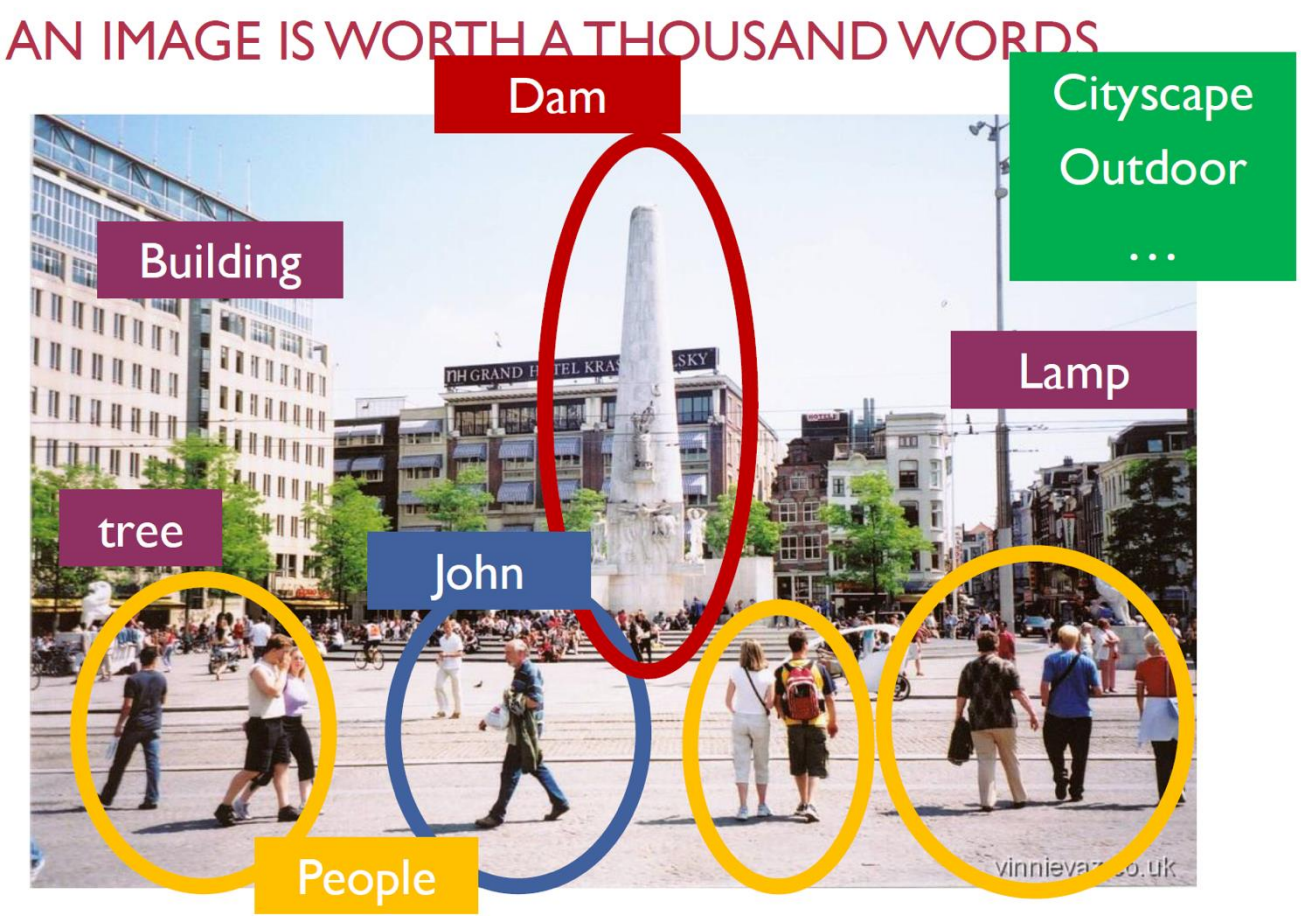
- Shape from X
- Camera Calibration
- 2-view Case:
  - Epipolar Geometry
  - Stereo Vision
  - Image Rectification
  - Stereo Reconstruction
- Multi-view Case: Structure from Motion

Textbook

- Chapter 12, 13, 14
- 3D Vision is mostly covered in CV2, a brief introduction in CV1



Motivation: Image Understanding



Motivation: Scene Understanding

So far mostly 2D Computer Vision:

- What is in the image?
- Object categories
- Object counts
- 2D object locations (bbox/mask)

3D Computer Vision:

- Where are objects in the scene?
- Distances
- Relative 3D positions
- Spatial object relations
- Object sizes
- Geometric shape of environment
- Scene Representations
- Vision-language models
- etc.

CV2

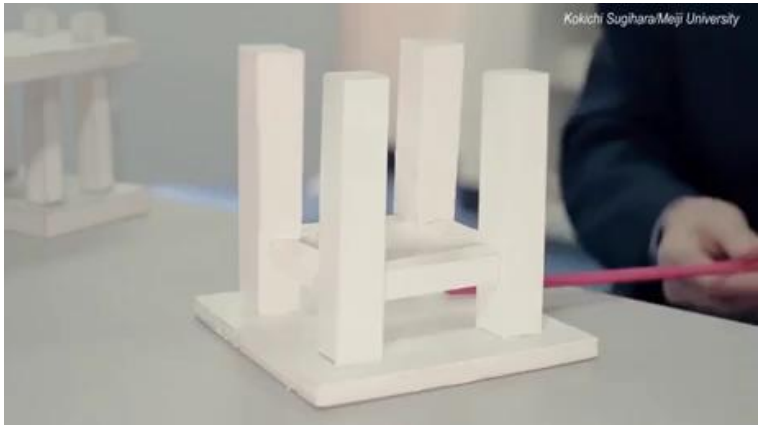
Challenges in 3D Reconstruction

Julian Beever with his anamorphic Coke bottle

- Output of monocular depth estimator
- or: How to stop an autonomous car?

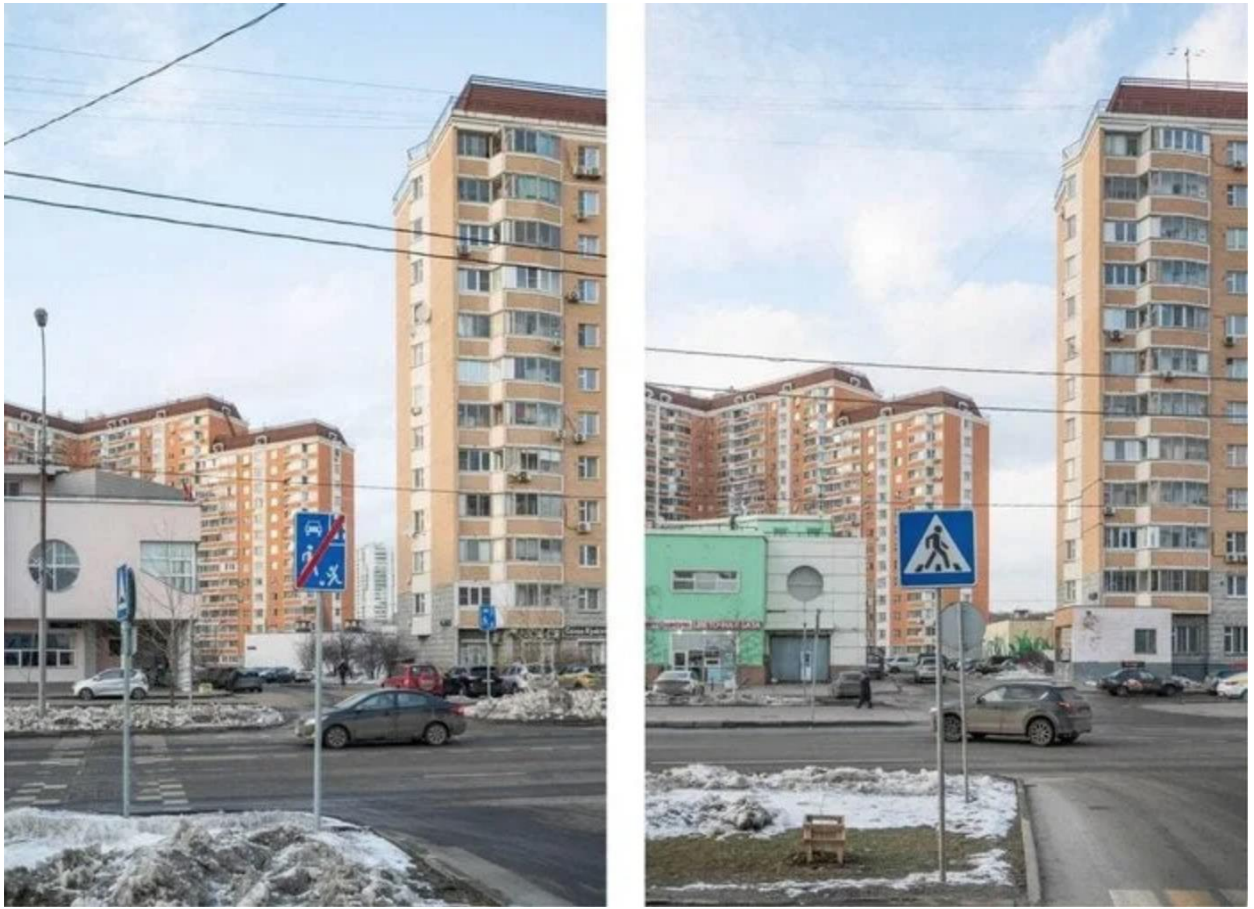
Challenges: View Ambiguity

Challenges: View Ambiguity

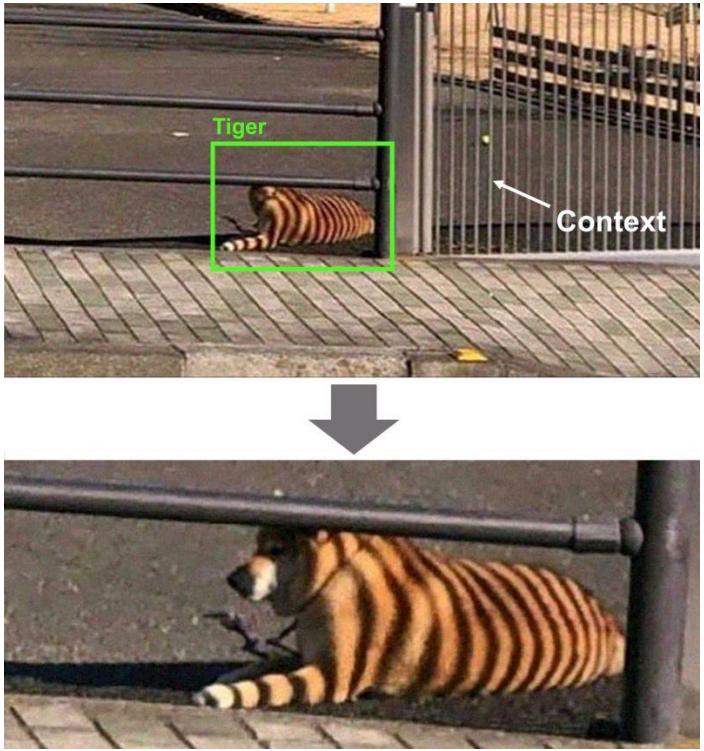


Challenges: Texture vs. Shape

Challenges in Localization



Challenges in Scene Understanding



Today's Agenda

- Shape from X
- Camera Calibration
- 2-view Case:
  - Epipolar Geometry
  - Stereo Vision
  - Image Rectification
  - Stereo Reconstruction
- Multi-view Case: Structure from Motion

Shape-from-X

X ∈ [

- Shading
- Texture
- Focus / Defocus
- Specularities
- Shadows
- Silhouettes
- Motion (/ Structure-from-Motion)
- Multiple Light Sources (photometric stereo)
- …

]

Shape-from-Shading

- Goal is to reconstruct a 3D profile from a single image
- Typical assumptions:
  - Diffuse material with constant albedo
  - Known light source at infinity
  - Known camera at infinity (orthography)
- Exploits relationship between shading and surface normal direction

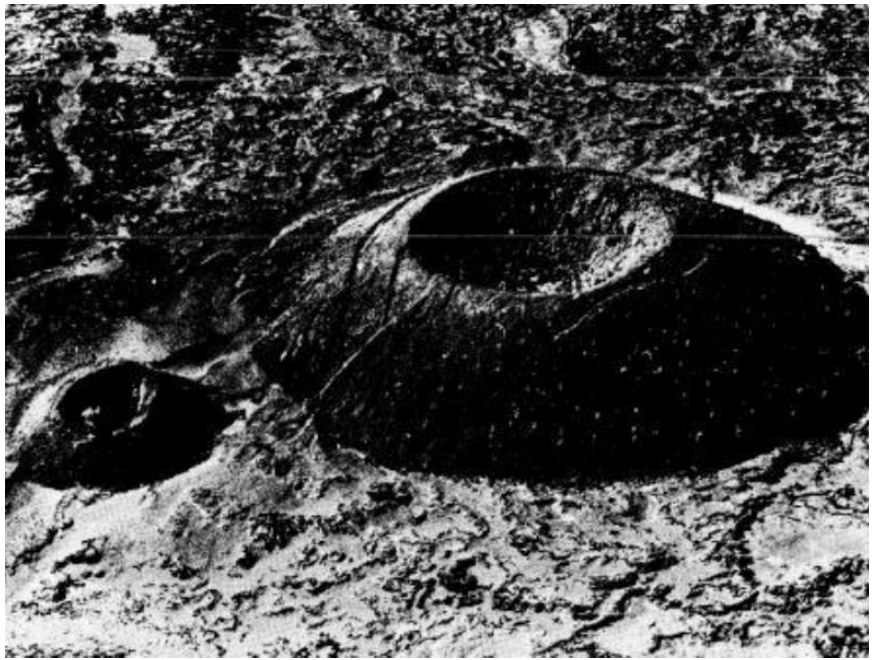
Challenges / Problems:

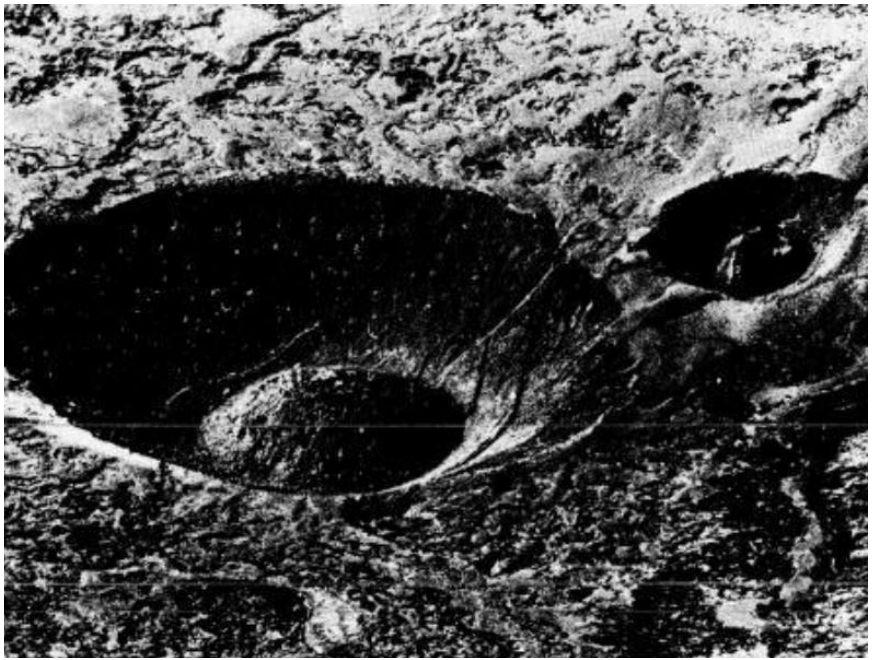
- Light direction ambiguities
- Non-Lambertion surfaces
- Texture vs. Shading differentiation
  -> works best with textureless, homogeneously colored objects

Image of the surface of planat Mars. Source: NASA

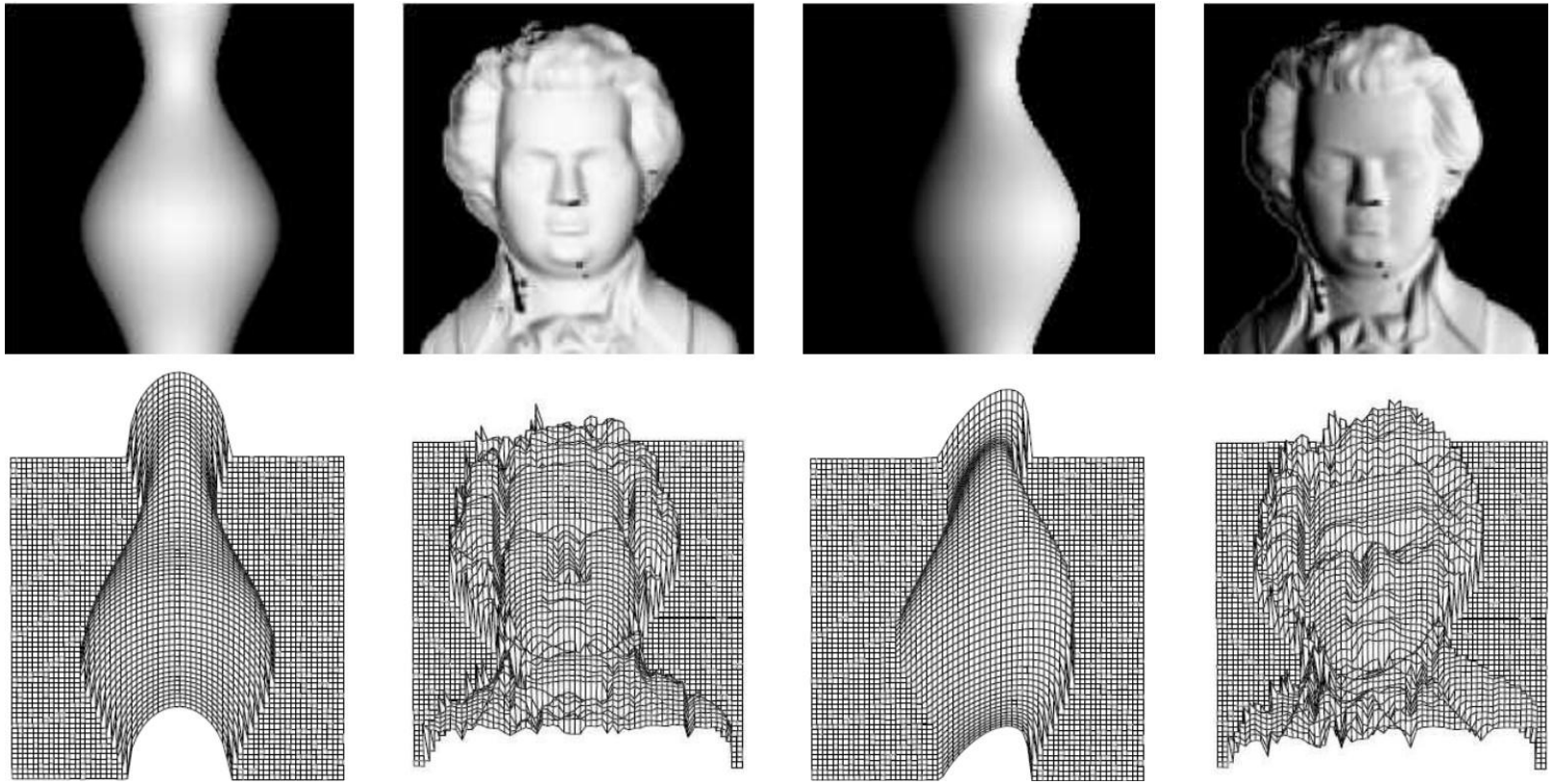
Shape-from-Shading: Ambiguities

Human vision resolves light direction ambiguities by assuming light is coming from above.





Shape-from-Shading – Early Results

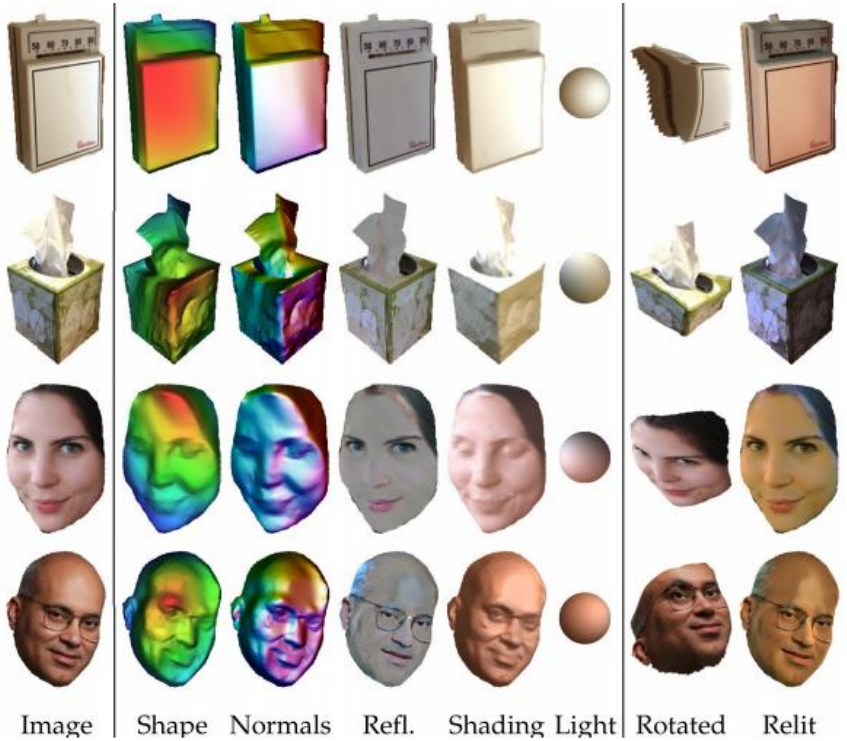


[Frankot and Chellappa: A Method for enforcing integrability in shape from shading algorithms. TPAMI, 1988]

Shape-from-Shading – Results

SIRFS: Shape, Illumination, and Reflectance from Shading

- Modern version of SfS
- Input: Single masked RGB image
- Estimates depth, normals, reflectance, shading, lighting
- Does not assume known point light
- Does not assume constant color
- Much harder problem than SfS

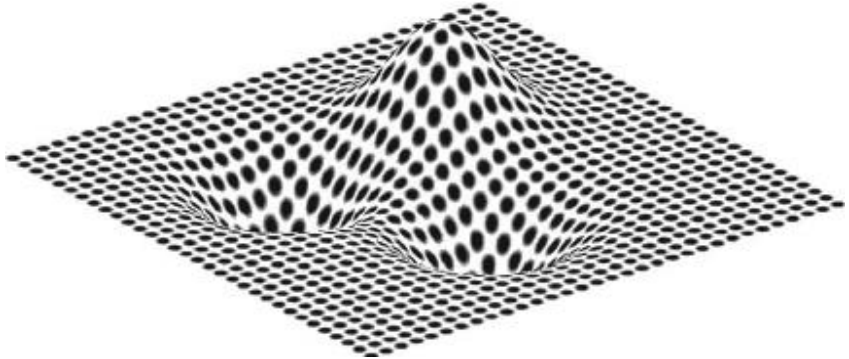


[Slide credit: Andreas Geiger]

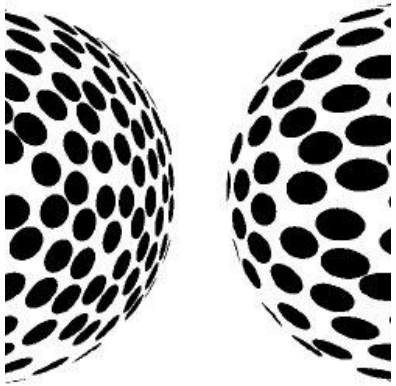
[Barron and Malik: Shape, Illumination, and Reflectance from Shading. TPAMI, 2015]

Shape-from-Texture

- Goal is to reconstruct a 3D profile from a single image



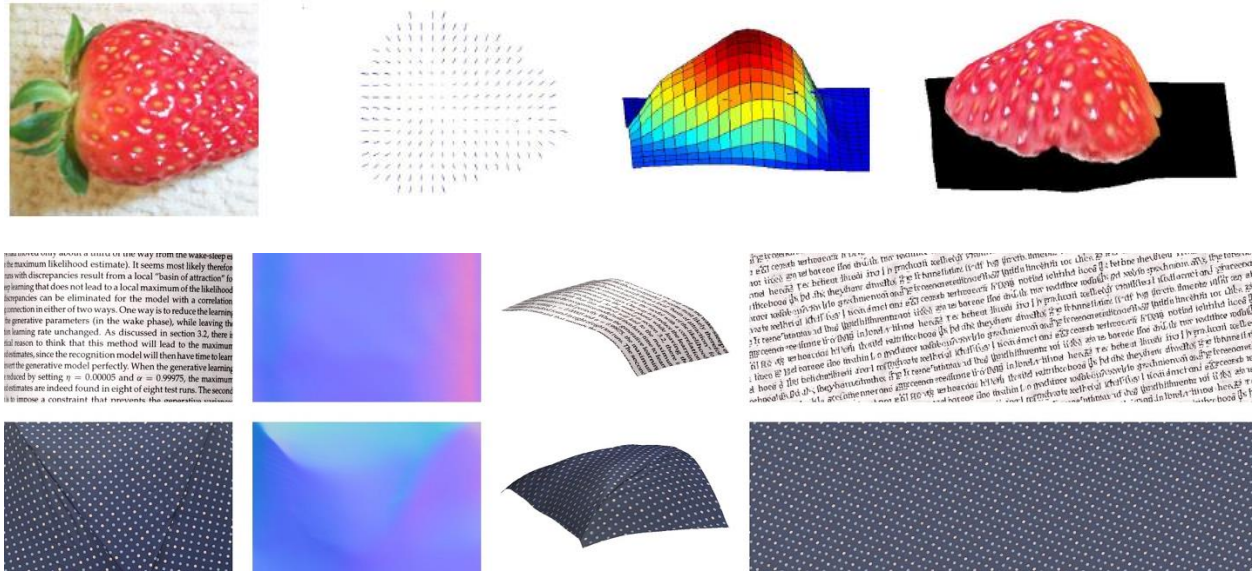
[Source: Gumsey et al., Journal of Vision, 2006]



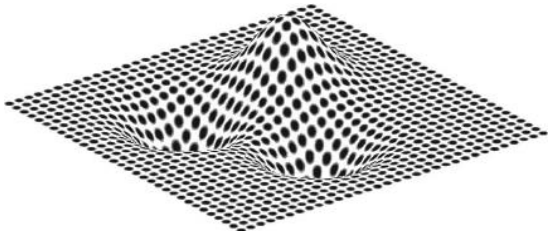
[Source: Verbin, Zickler, CVPR 2020, http://vision.seas.harvard.edu/sft/]

Shape-from-Texture

- Goal is to reconstruct a 3D profile from a single image assuming a repetitive texture pattern



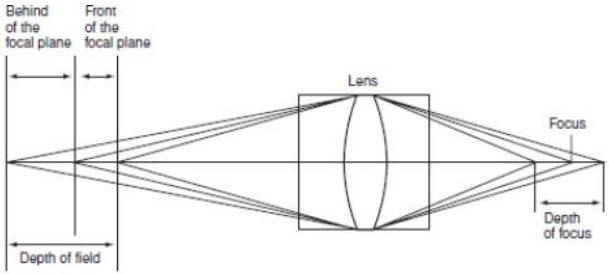
[Source: Verbin, Zickler, CVPR 2020, http://vision.seas.harvard.edu/sft/]

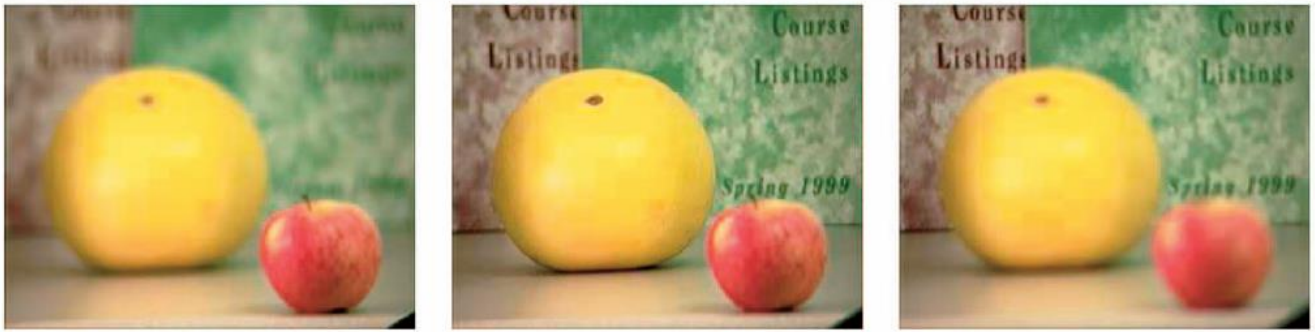


[Source: Gumsey et al., Journal of Vision, 2006]

Shape-from-Focus

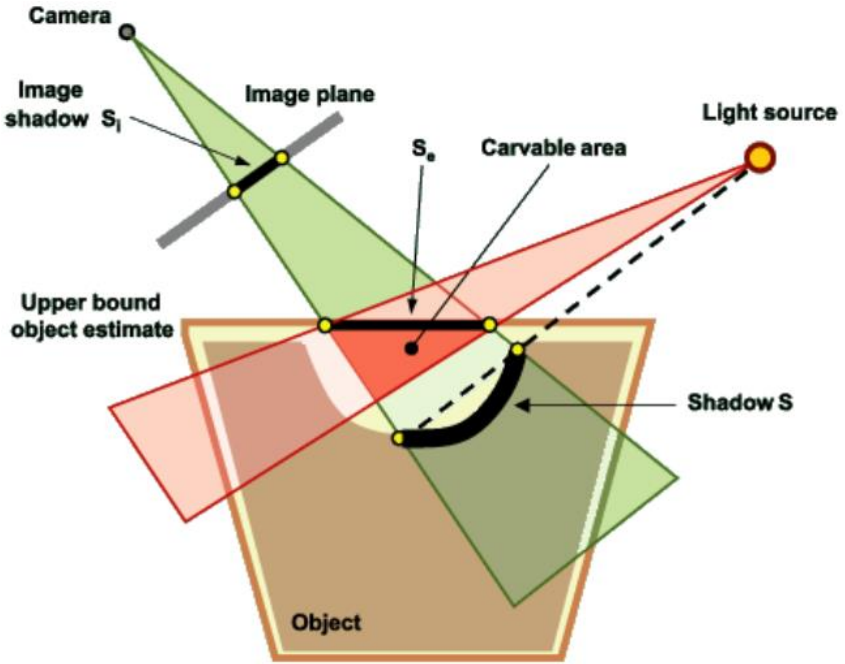
- Sweep through focus settings “most sharp” pixels correlate with the depth (most high frequencies)

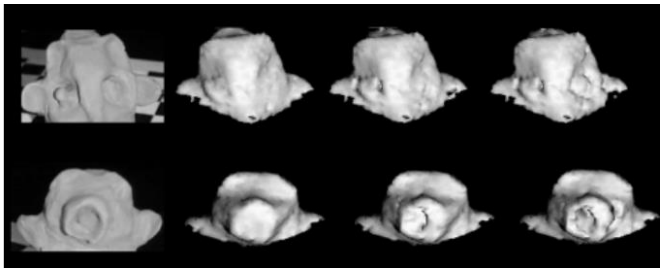




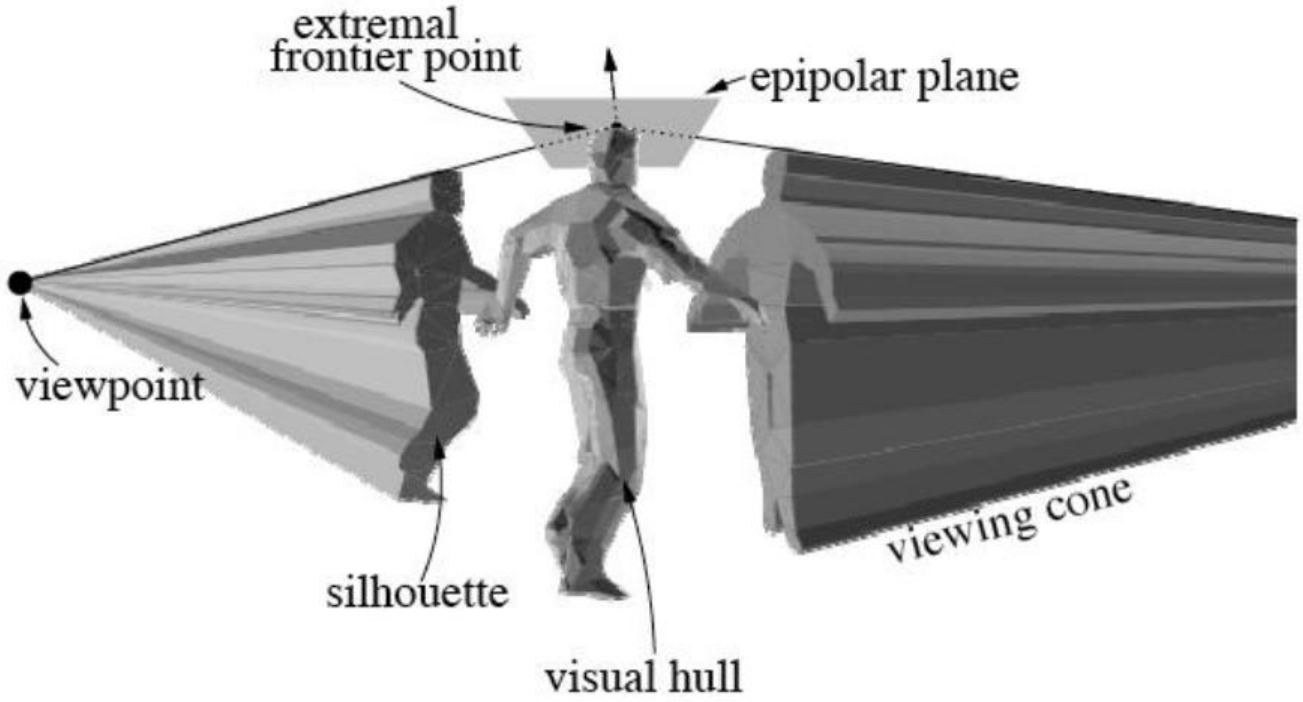
Shape-from-Shadows

- Detect shadows and use them to recover shape cues





Shape-from-Silhouettes

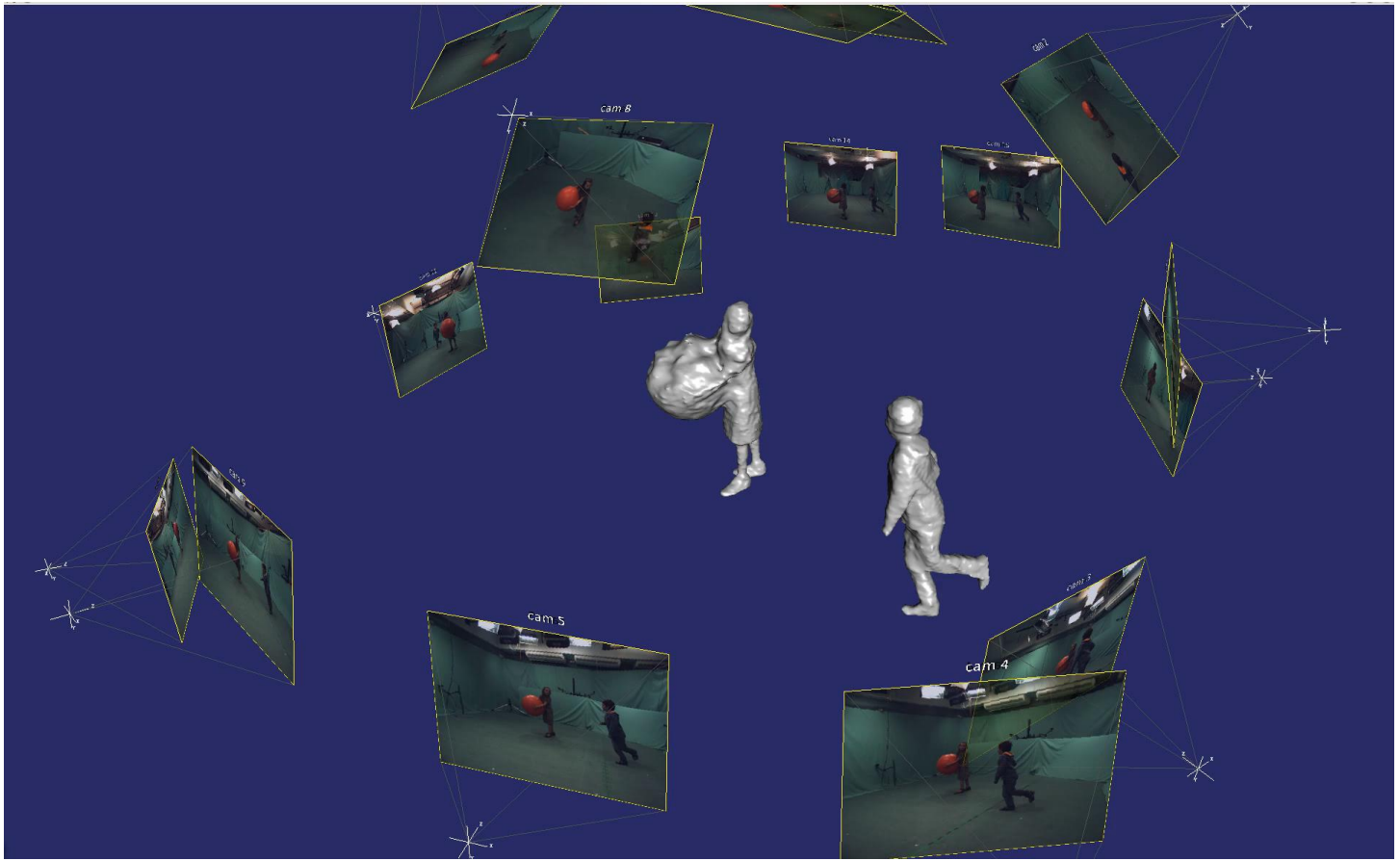


Visual Hull: Intersection of all silhouette-based viewing cones (with known cameras)

Shape-from-Silhouettes

[Miller, PhD Thesis 2007, https://www.youtube.com/watch?v=21YUA-SalO0]

Shape-from-Silhouettes



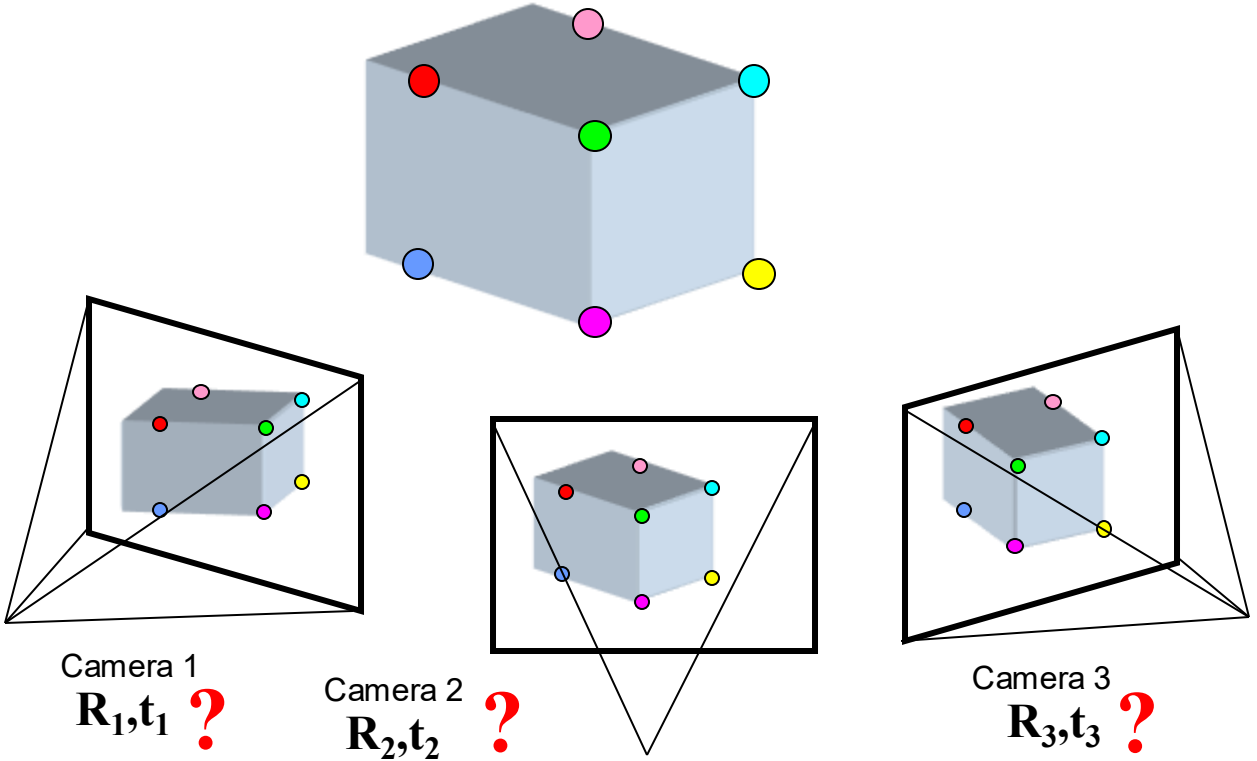
[Oswald, Cremers, BMVC 2014]

Today's Agenda

- Shape from X
- Camera Calibration
- 2-view Case:
  - Epipolar Geometry
  - Stereo Vision
  - Image Rectification
  - Stereo Reconstruction
- Multi-view Case: Structure from Motion

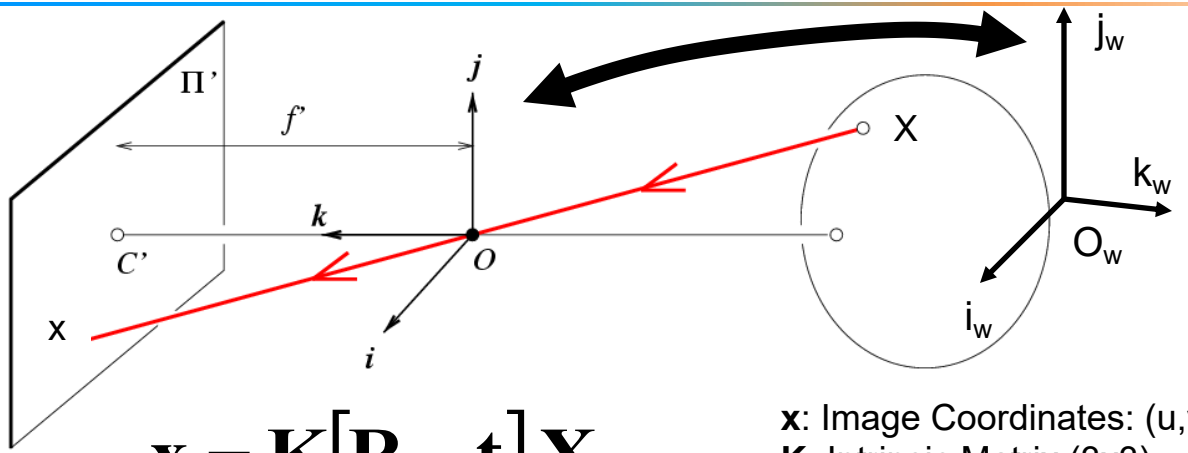
Camera Calibration

Camera 'Motion': Given a set of corresponding 2D/3D points in two or more images, compute the camera parameters.



[Slide credit: Noah Snavely]

Recap: Camera Model



$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\mathbf{X}$$

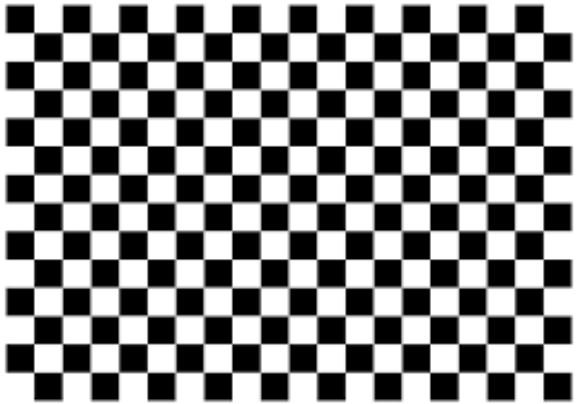
Extrinsic Matrix

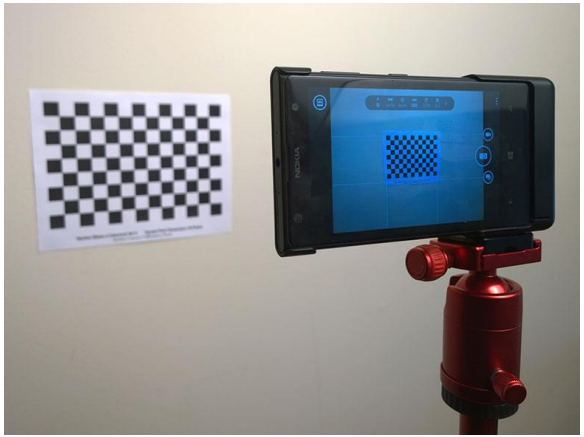
x: Image Coordinates: (u,v,1)
K: Intrinsic Matrix (3x3)
R: Rotation (3x3)
t: Translation (3x1)
X: World Coordinates: (X,Y,Z,1)

$$w\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}f_x & s & u_0\\0 & f_y & v_0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}r_{11} & r_{12} & r_{13} & t_x\\r_{21} & r_{22} & r_{23} & t_y\\r_{31} & r_{32} & r_{33} & t_z\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix}$$

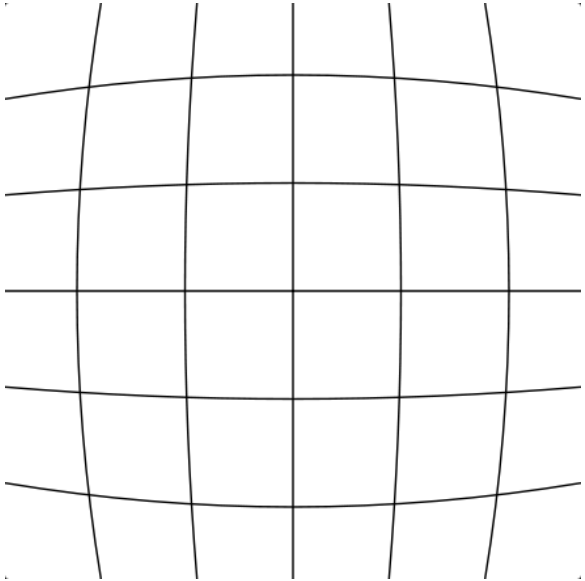
Intrinsic Calibration

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

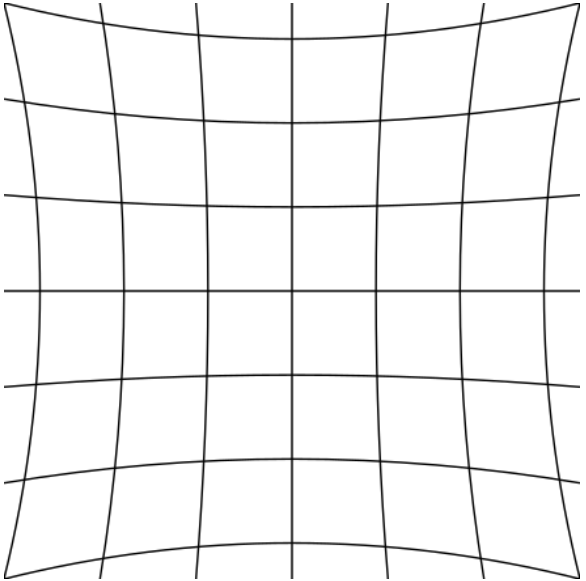




Intrinsic Calibration: Non-linear Distortion

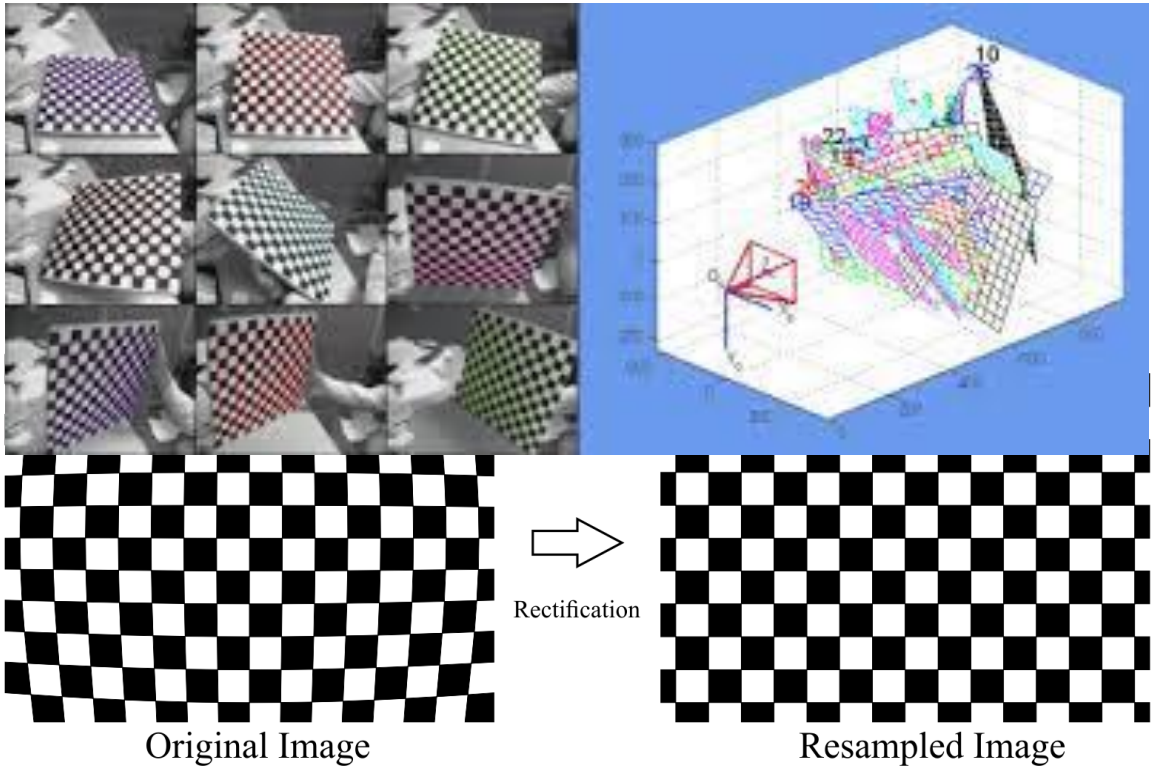


Barrel distortion



Pincushion distortion

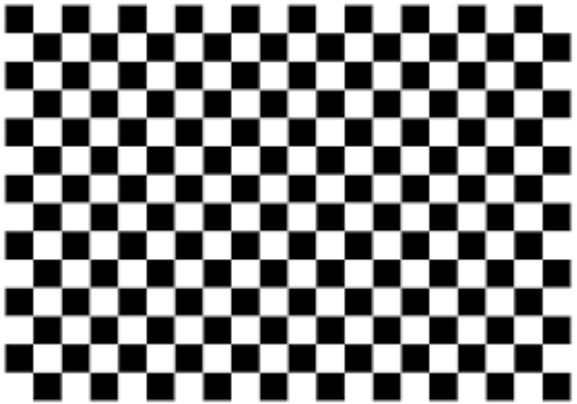
Intrinsic Calibration: Non-linear Distortion



- Typically, one preprocesses all images to remove non-linear distortion before estimating computing dense correspondences and stereo.

Extrinsic Calibration

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

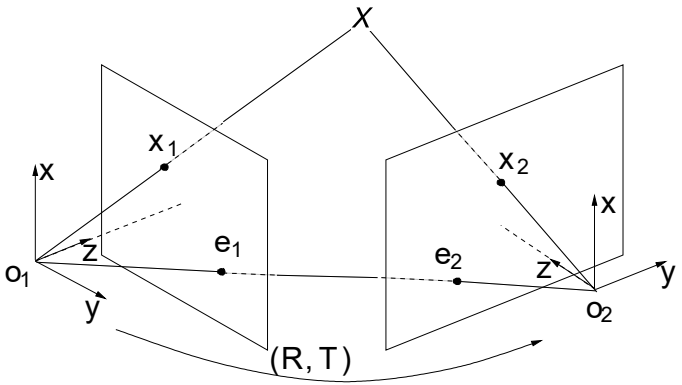


- With sufficiently many point correspondences within two views, one can solve for all unknowns!

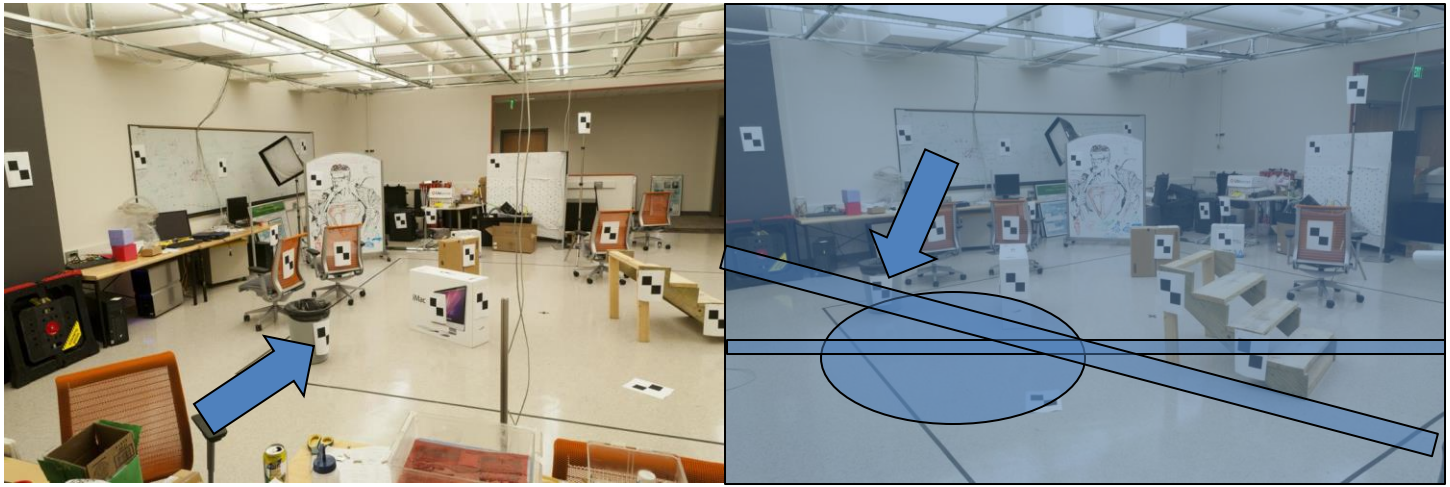
Today's Agenda

- Shape from X
- Camera Calibration
- 2-view Case:
  - Epipolar Geometry
  - Stereo Vision
  - Image Rectification
  - Stereo Reconstruction
- Multi-view Case: Structure from Motion

Two-View Geometry

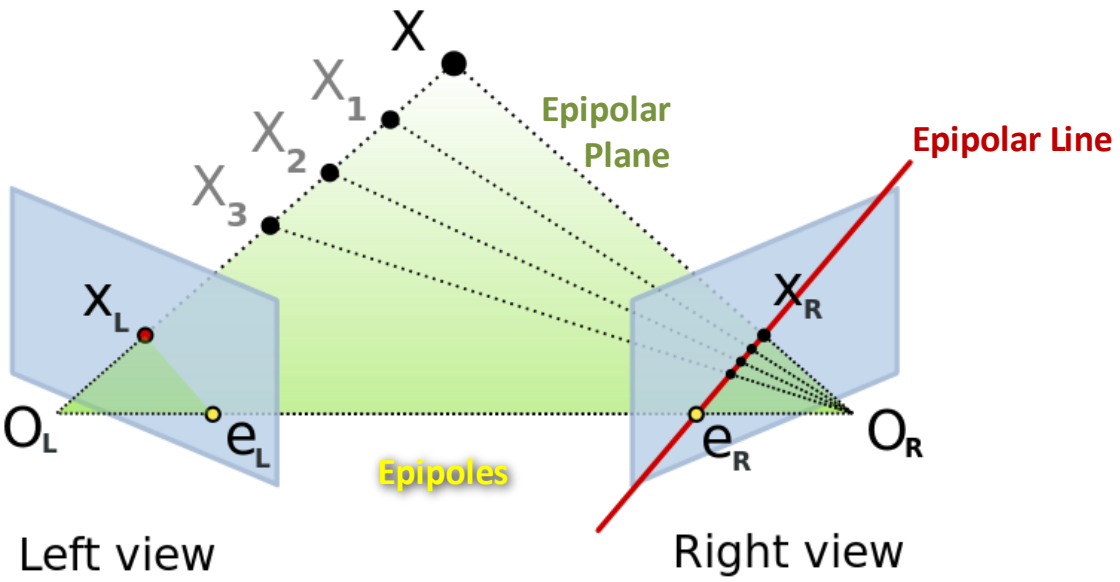


Motivation: Epipolar Geometry



Epipolar geometry contraints 2D search to 1D!

Epipolar Geometry



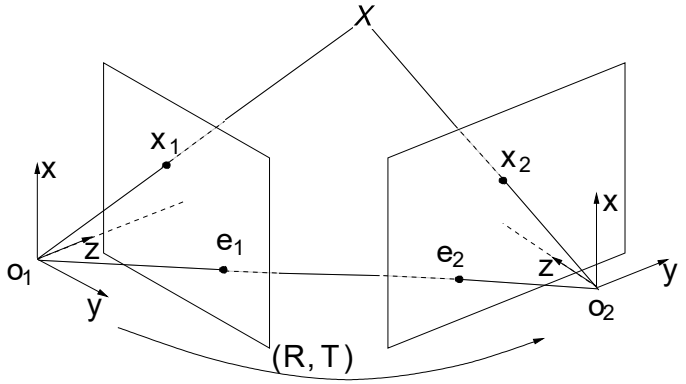
Epipolar Geometry

[Noah Snavely, Fundamental matrix demo, 2022.
https://www.cs.cornell.edu/courses/cs5670/2022sp/demos/FundamentalMatrix/?demo=demo1]

Essential Matrix

We assume **known** (/identity) intrinsic camera parameters.

Point x_1 is the projection of 3D point X with unknown depth λ_1.



$$\lambda_1 x_1 = X \qquad \lambda_2 x_2 = RX + T$$

$$\lambda_2 x_2 = R(\lambda_1 x_1) + T$$

$\widehat{T}$

$(\widehat{T}v \equiv T \times v)$

$\widehat{T}$ *is* a skew-symmetric matrix replicating the cross product in 3D

$$\lambda_2 \widehat{T} x_2 = \lambda_1 \widehat{T} R x_1$$

$$x_2^\top \lambda_2 \widehat{T} x_2 = 0$$

$$x_2^\top \widehat{T} R x_1 = 0$$

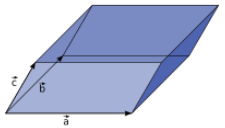
epipolar constraint

$$E = \widehat{T} R \quad \in \mathbb{R}^{3\times 3}$$

essential matrix

Geometrically, the epipolar constraint states that the vectors $\overrightarrow{o_1 X}$, $\overrightarrow{o_2 o_1}$, $\overrightarrow{o_2 X}$ form a plane, that is, their triple product (measuring the volume of the parallelepiped) is zero:

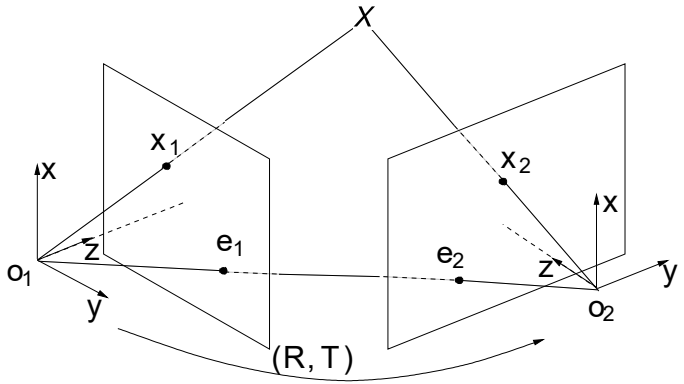
$$volume = x_2^\top (T \times R x_1) = 0$$



Fundamental Matrix

When intrinsic camera parameters are **unknown**:

The fundamental matrix is directly linked to the essential matrix via multiplication of the (inverse) calibration matrix.



$$\boldsymbol{x}_2^\top \hat{T} R \boldsymbol{x}_1 = 0 \quad \Leftrightarrow \quad \boldsymbol{x}_2'^\top K^{-\top} \hat{T} R K^{-1} \boldsymbol{x}_1' = 0$$

$$\boldsymbol{x}_2'^\top F \boldsymbol{x}_1' = 0$$

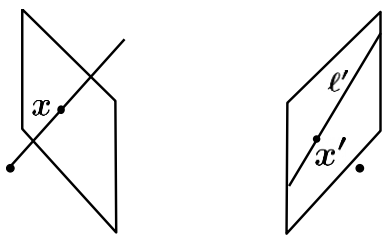
epipolar constraint

$$F \equiv K^{-\top} \hat{T} R K^{-1} = K^{-\top} E K^{-1}$$

fundamental matrix

- Algebraic representation of epipolar geometry
- unique 3x3 rank 2 matrix (7 DoF)
- Maps points to epipolar lines

$$\boldsymbol{\ell}' = F\boldsymbol{x}$$
$$\boldsymbol{\ell} = F^T\boldsymbol{x}'$$



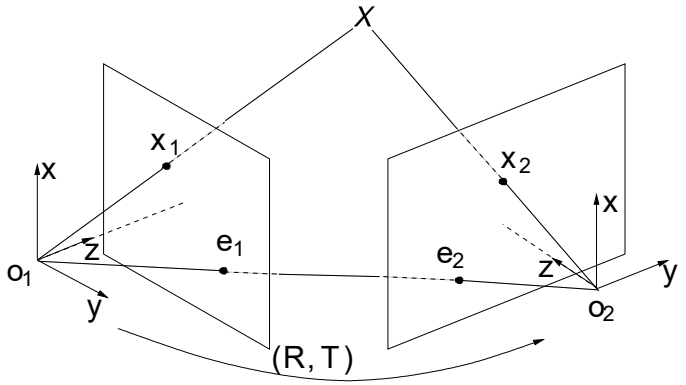
- Epipolar constraint

$$\boldsymbol{x}'^T F \boldsymbol{x} = 0$$

(Fundamental Matrix Song: http://danielwedge.com/fmatrix/)

Overview: Essential & Fundamental Matrix

- Describe geometric relations between pairs of images
- Contains all information about relative orientation from corresponding points
- 3x3 matrix
- Rank deficiency: $\mathrm{rank}(E) = \mathrm{rank}(F) = 2$ due to scale ambiguity
- Coplanarity constraint $x_2{}^T E x_1 = 0$ / $x_2{}^T F x_1 = 0$ of observed point pairs leads to $Af = 0$ → solve via SVD



Essential Matrix

- Known intrinsics
- $x_2{}^T E x_1 = 0$
- 5 DOF
- 5-point algorithm

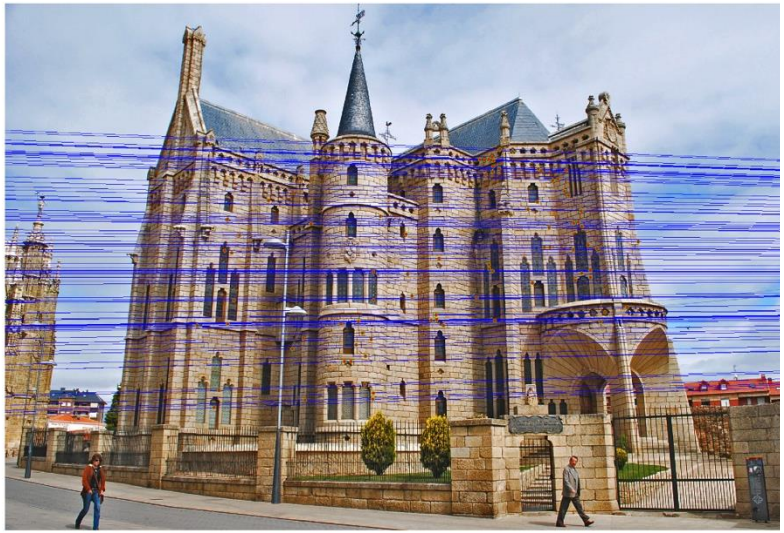
Fundamental Matrix

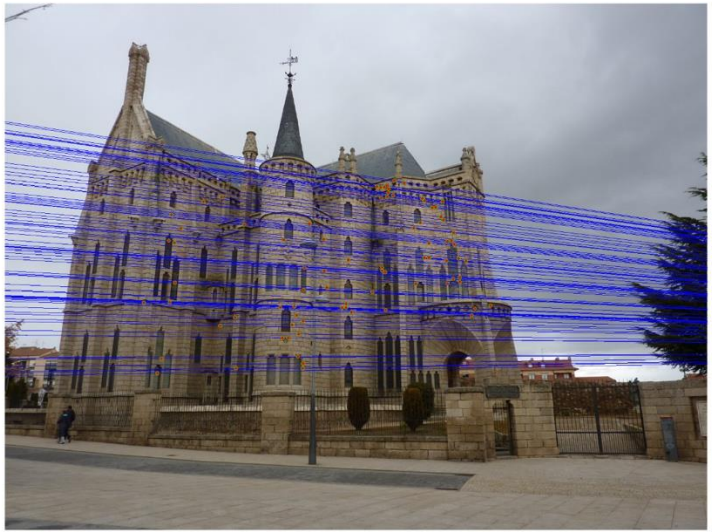
- Unknown intrinsics
- $x_2{}^T F x_1 = 0$
- 7 DOF
- 8-point algorithm

$$F \equiv K^{-\top} \hat{T} R K^{-1} = K^{-\top} E K^{-1}$$

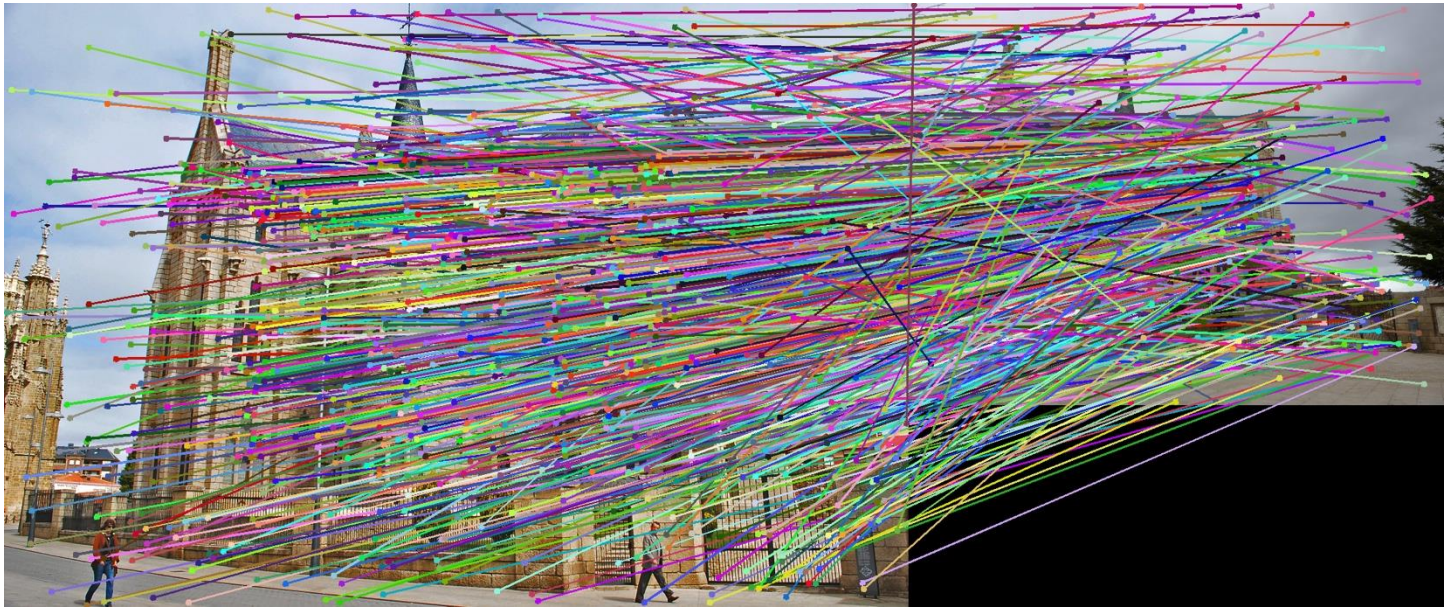
(Fundamental Matrix Song: http://danielwedge.com/fmatrix/)

Epipolar Lines

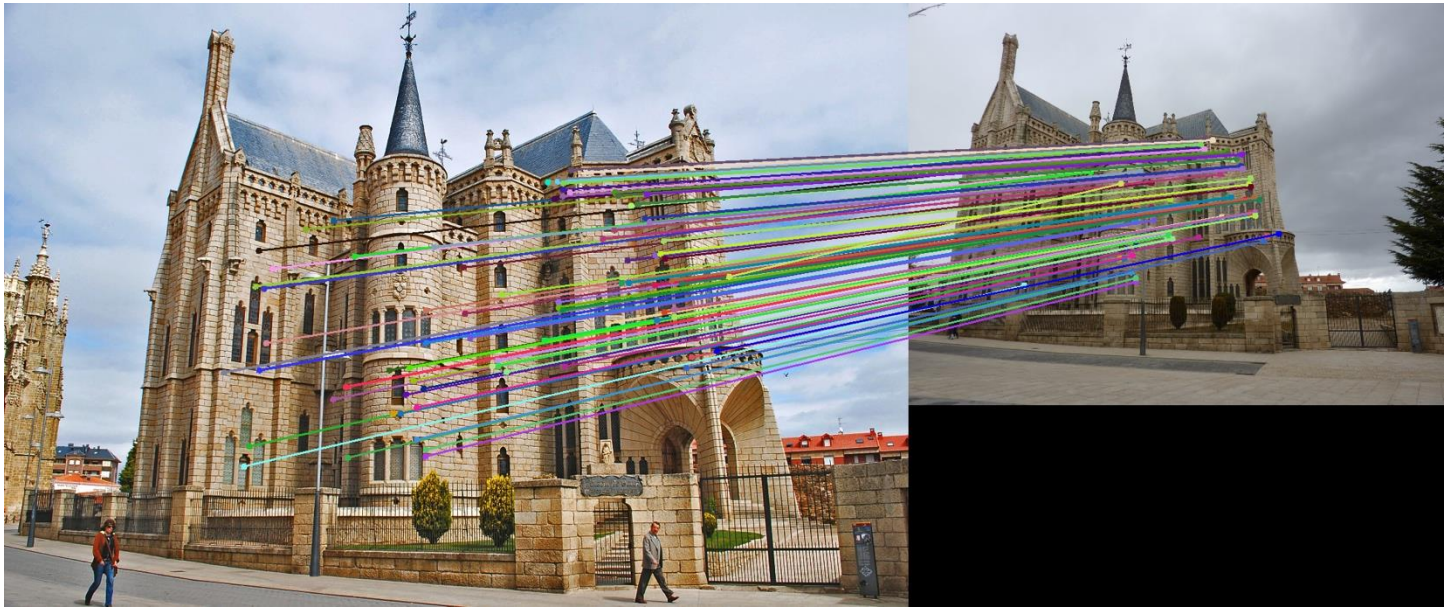




Feature Correspondences



Feature Correspondences

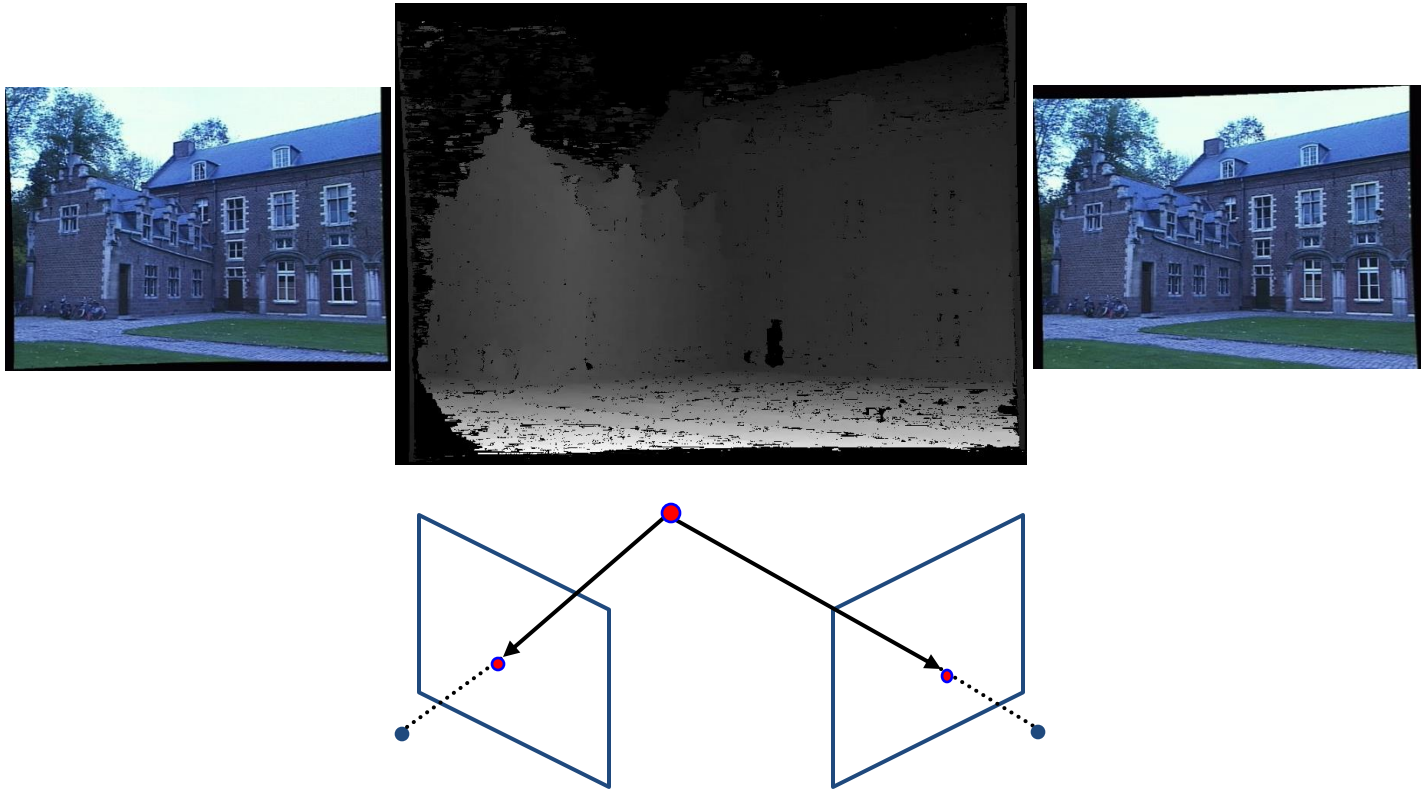


Keep only the matches at are “inliers” with respect to the epilolar geometry (via RANSAC).

Today's Agenda

- Shape from X
- Camera Calibration
- 2-view Case:
  - Epipolar Geometry
  - Stereo Vision
  - Image Rectification
  - Stereo Reconstruction
- Multi-view Case: Structure from Motion

Stereo Vision



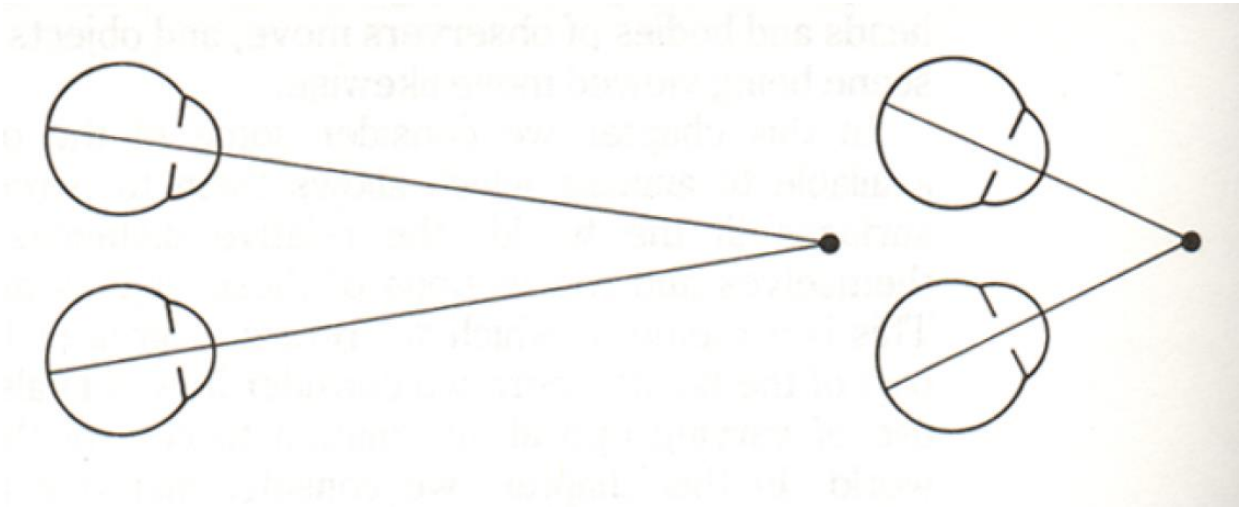
[Slide credit: James Hays and Kristen Grauman]

Stereo Vision

If stereo were critical for depth perception, navigation, recognition, etc., then rabbits would never have evolved.

Human stereopsis

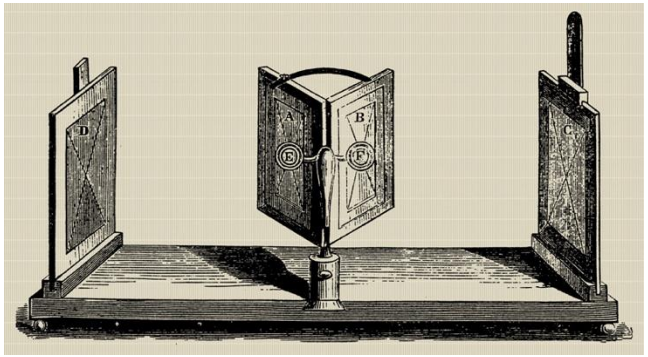
Human eyes **fixate** on point in space – rotate so that corresponding images form in centers of fovea.



[From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology]

Stereo photography and stereo viewers

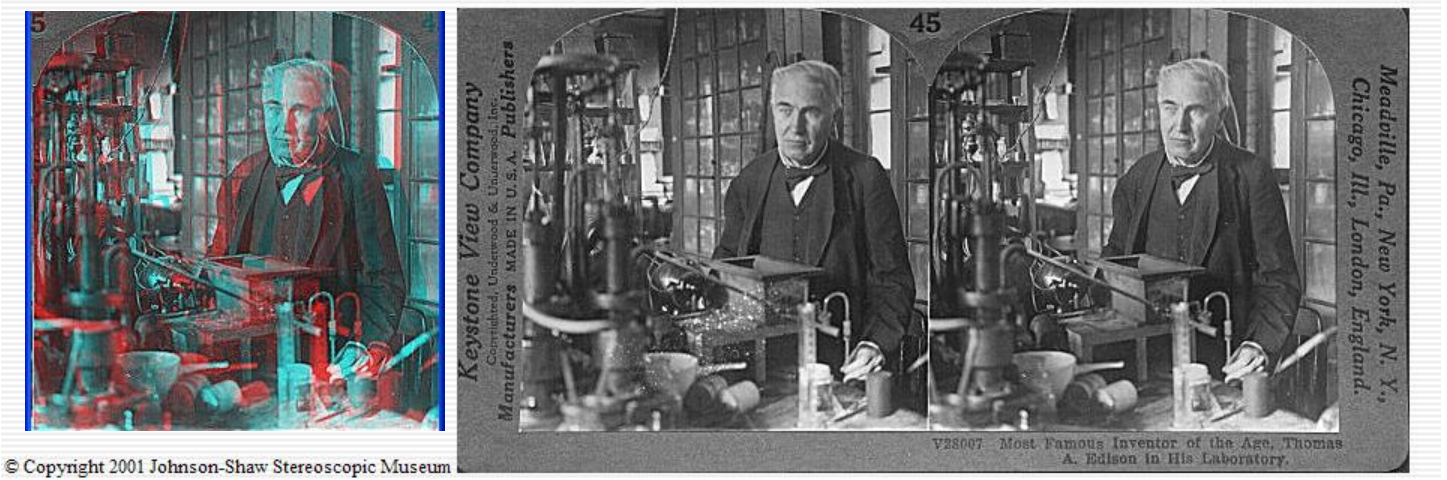
Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.



Invented by Sir Charles Wheatstone, 1838

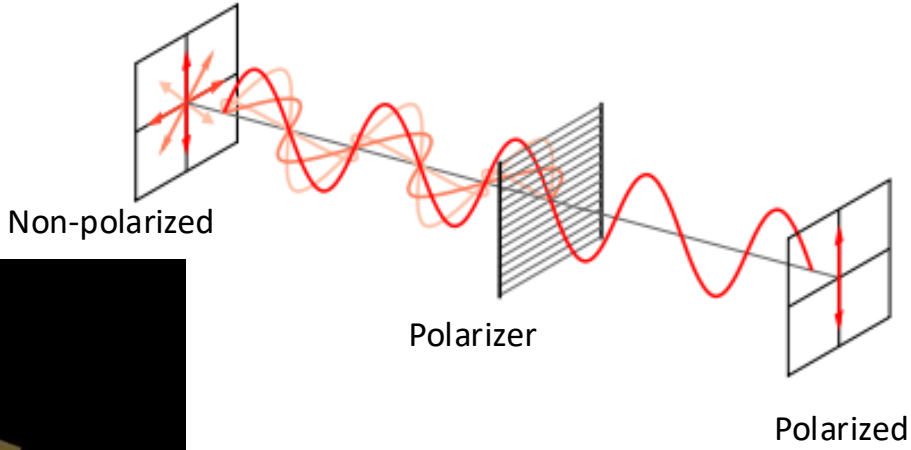
[Image from fisher-price.com]

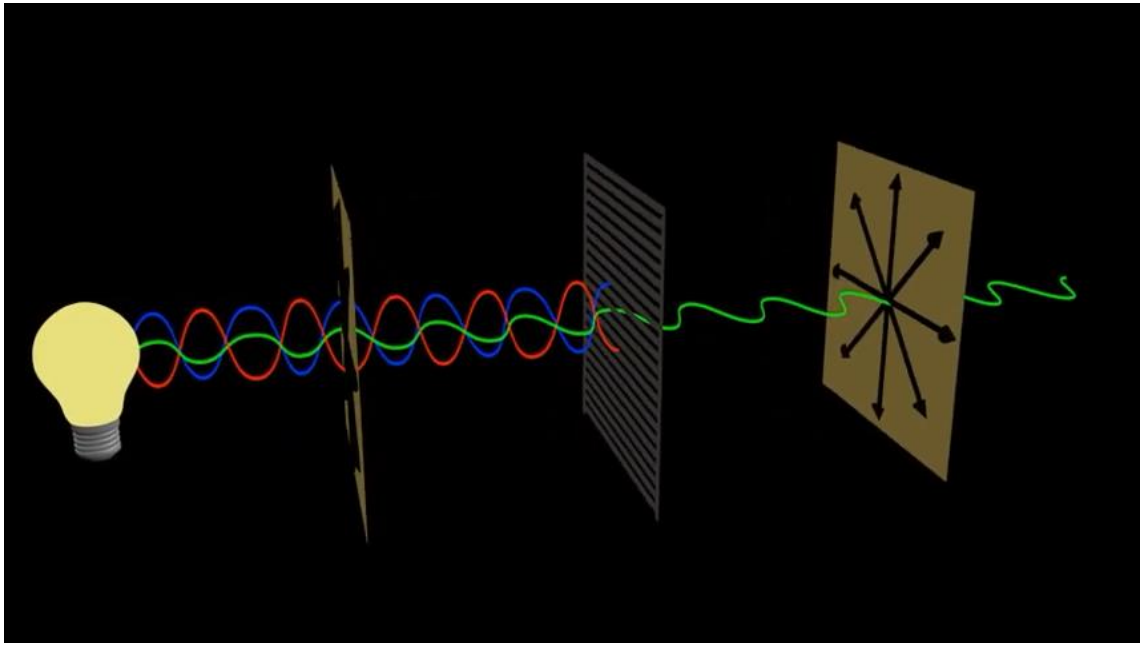
Anaglyph Stereo



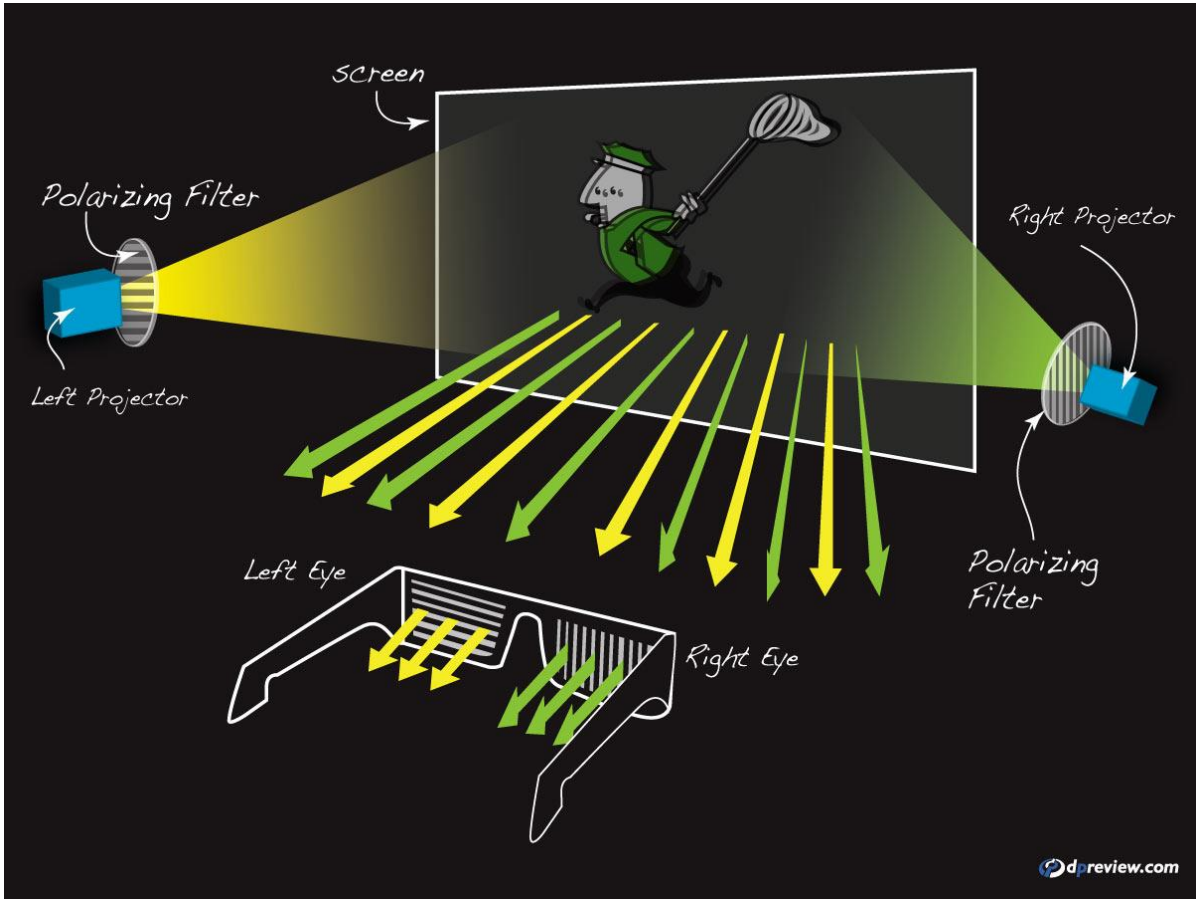
[http://www.johnsonshawmuseum.org]

3D Movies: View Separation via Polarization





3D Movies: View Separation via Polarization



Stereo Vision

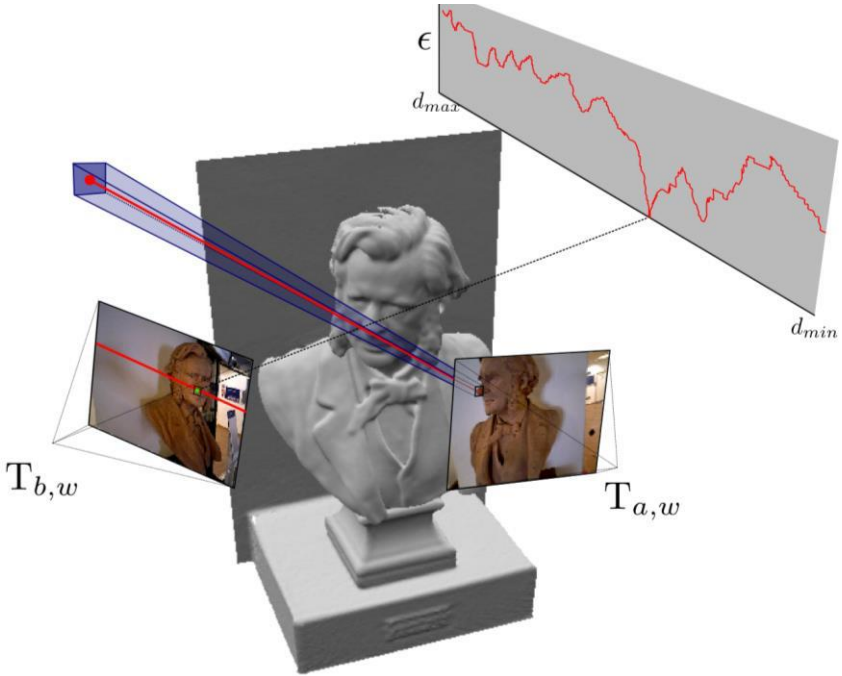
Two cameras, simultaneous views

Single moving camera and static scene

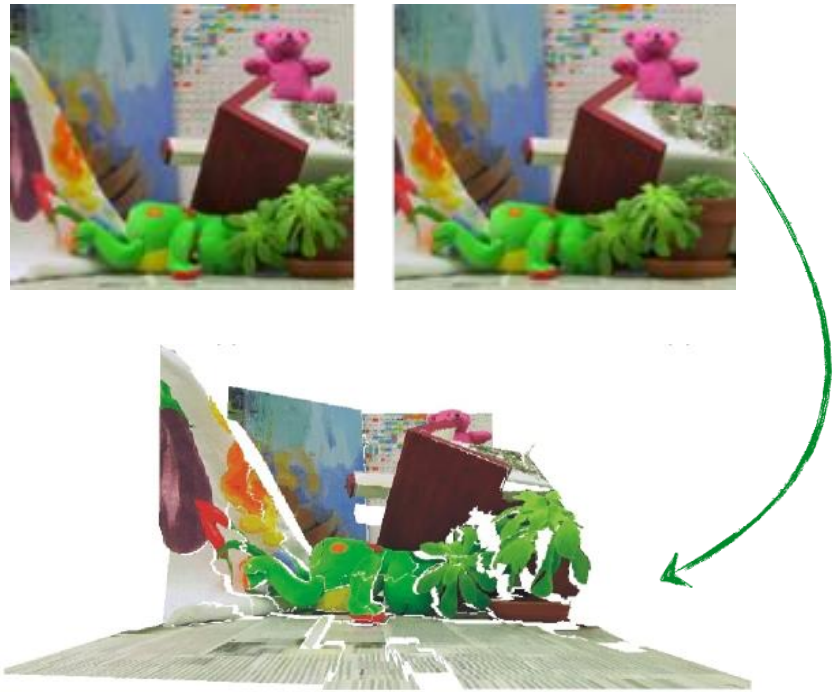
Why multiple views?

Structure and depth can be ambiguous from single views...

Multi-view Stereo



General Setting

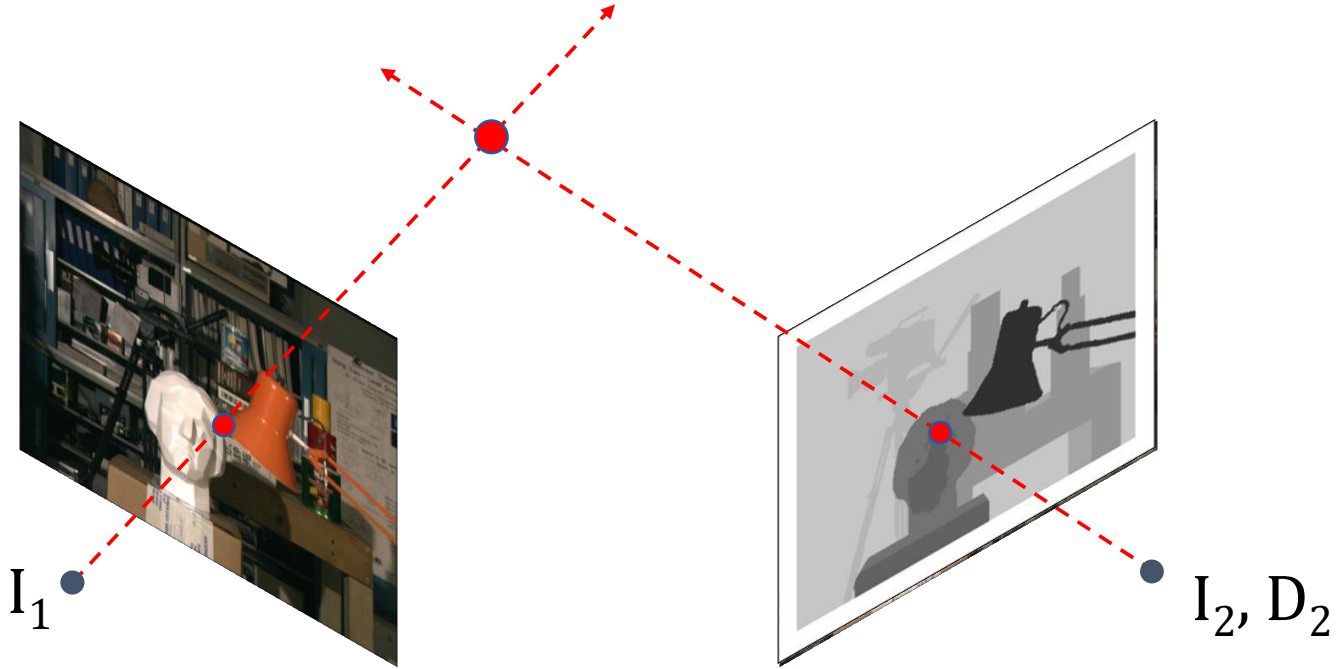


Rectificied Camera Setting

[Newcombe & Lovegrove, Geometric Reconstruction Lecture]

[Svetlana Lazebnik, Computer Vision Lecture]

Dense Scene Geometry General Model



$$I_2(x,y) = I_1(\pi(T_{12} K^{-1} D_2(x,y) [x,y]))$$

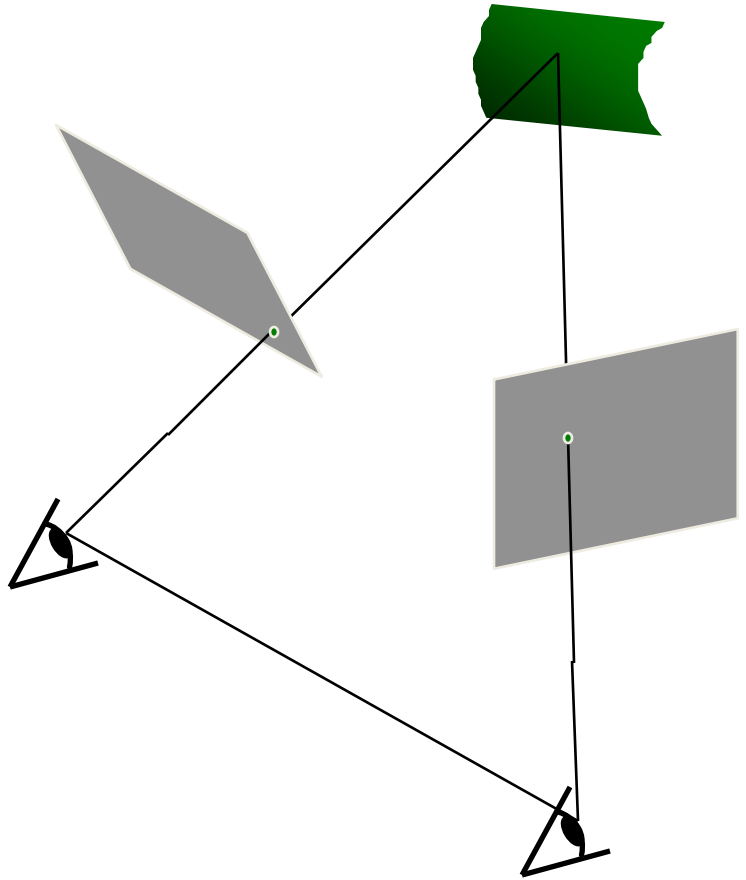
[Newcombe & Lovegrove, Geometric Reconstruction Lecture]

Today's Agenda

- Shape from X
- Camera Calibration
- 2-view Case:
  - Epipolar Geometry
  - Stereo Vision
  - Image Rectification
  - Stereo Reconstruction
- Multi-view Case: Structure from Motion

Stereo Image Rectification

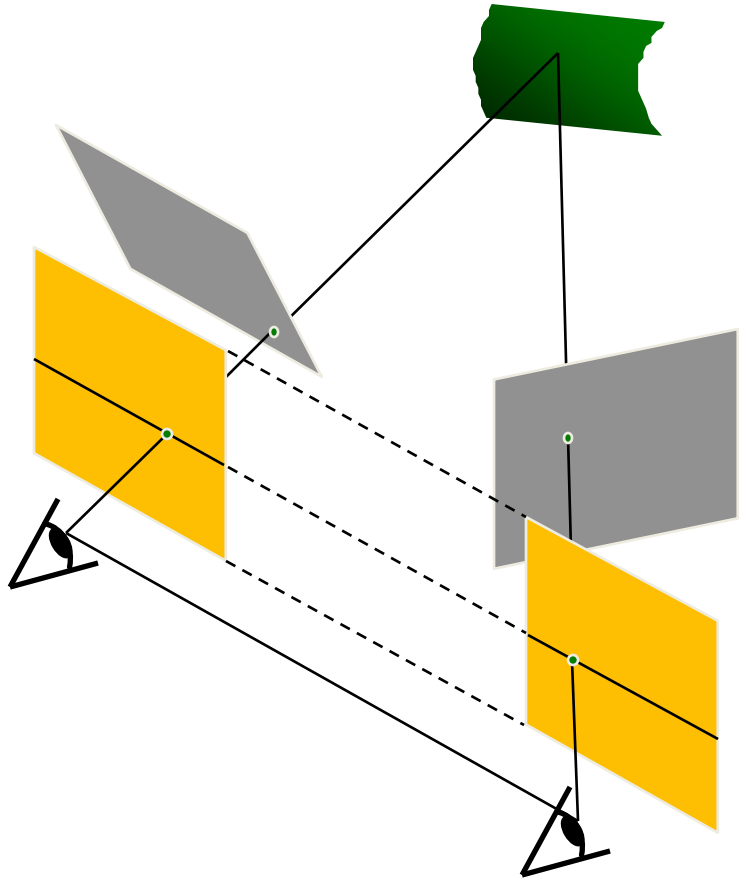
- Project each image onto a common plane parallel to the baseline using homographies



[Svetlana Lazebnik, Computer Vision Lecture]

Stereo Image Rectification

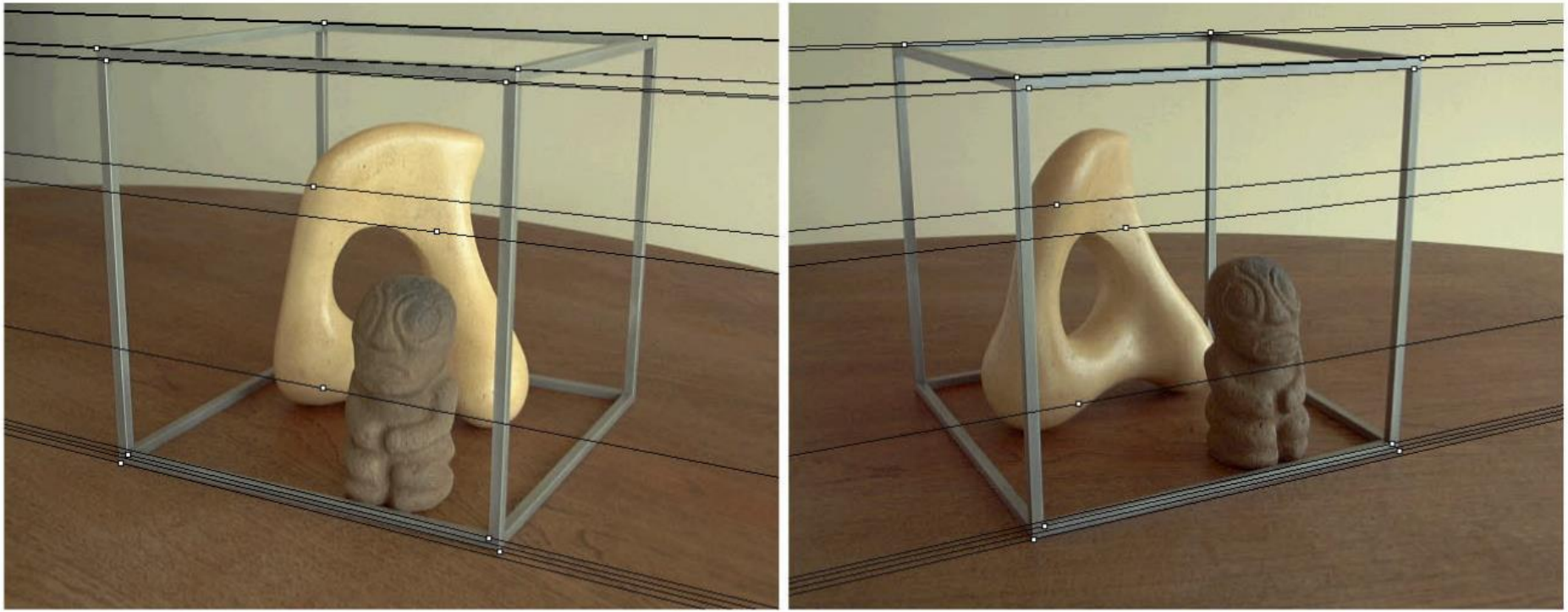
- Project each image onto a common plane parallel to the baseline using homographies



[Svetlana Lazebnik, Computer Vision Lecture]

Stereo Image Rectification

Before rectification



[Svetlana Lazebnik, Computer Vision Lecture]

Stereo Image Rectification

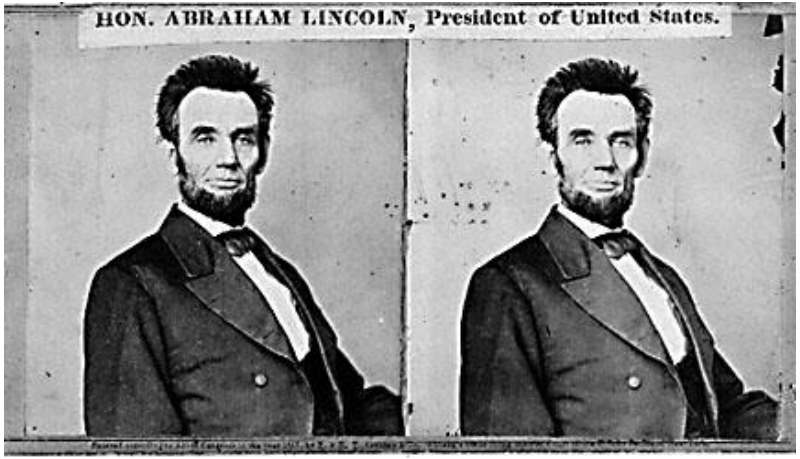
After rectification

[Svetlana Lazebnik, Computer Vision Lecture]

Today's Agenda

- Shape from X
- Camera Calibration
- 2-view Case:
  - Epipolar Geometry
  - Stereo Vision
  - Image Rectification
  - Stereo Reconstruction
- Multi-view Case: Structure from Motion

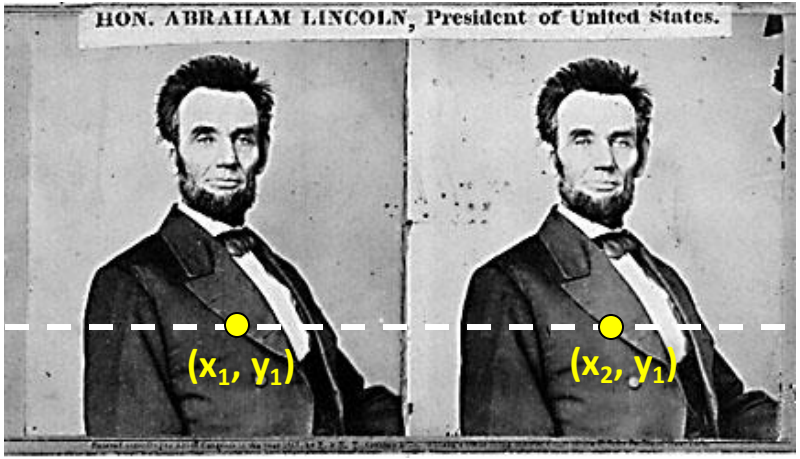
Stereo



- Given two images from different viewpoints
  - How can we compute the depth of each point in the image?
  - Based on *how much each pixel moves* between the two images

[Slide credit: Noah Snavely]

Stereo

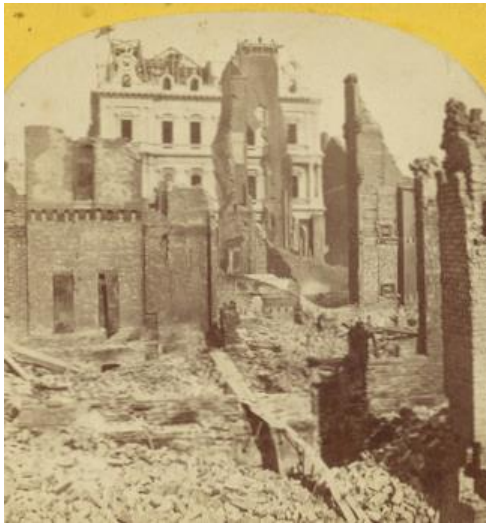


Two images captured by a purely horizontal translating camera
(*rectified* stereo pair)

$x_2 - x_1$ = the *disparity* of pixel (x_1, y_1)

[Slide credit: Noah Snavely]

Disparity = inverse depth



http://stereo.nypl.org/view/41729

(Or, hold a finger in front of your face and wink each eye in succession.)

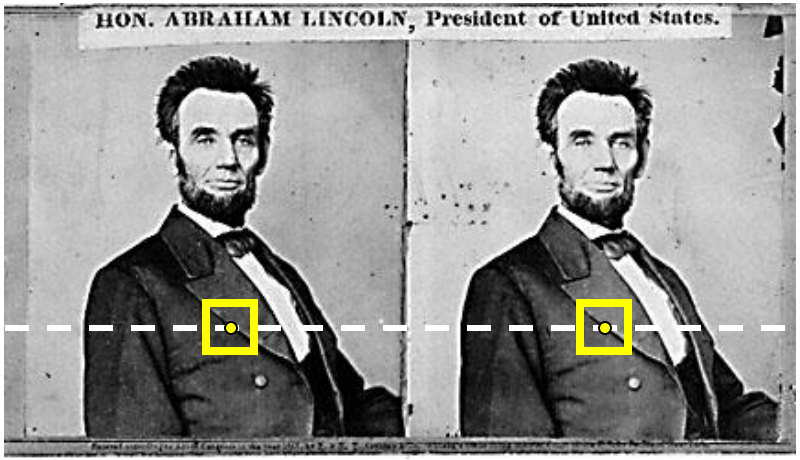
[Slide credit: Noah Snavely]

Your basic stereo matching algorithm

- **Match Pixels in Conjugate Epipolar Lines**
  - Assume brightness constancy
  - This is a challenging problem
  - Hundreds of approaches
    - A good survey and evaluation: http://www.middlebury.edu/stereo/

[Slide credit: Noah Snavely]

Your basic stereo matching algorithm



For each epipolar line

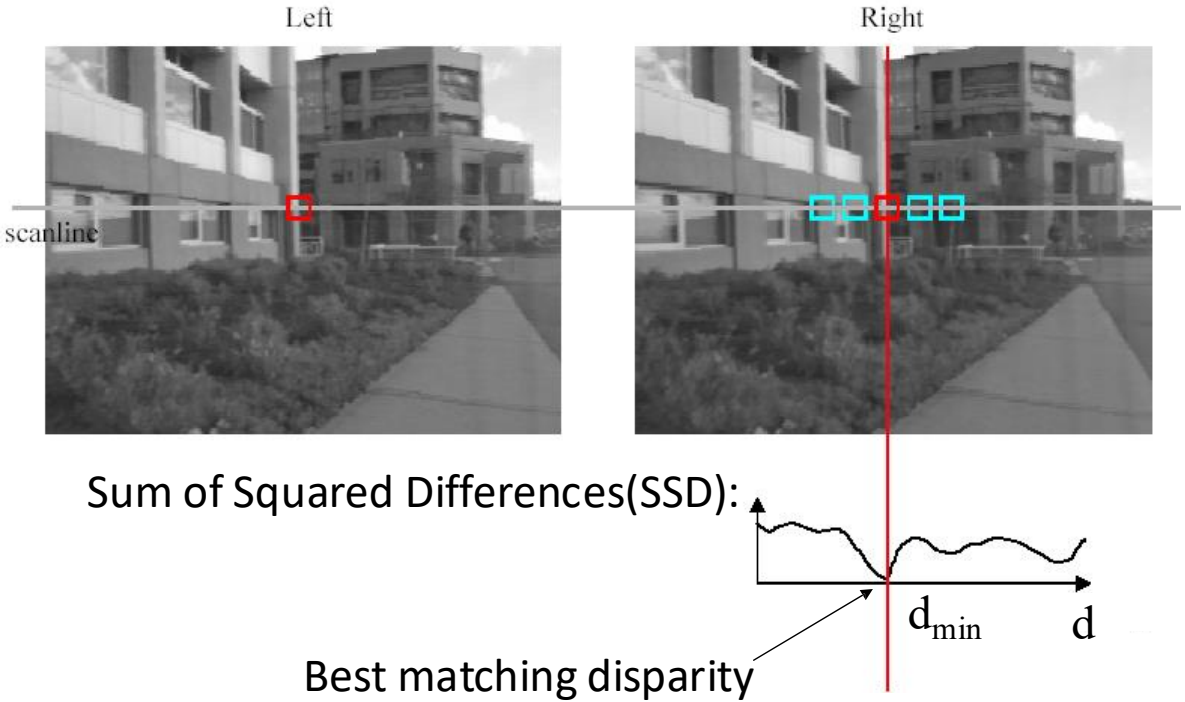
For each pixel in the left image

- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost

Improvement: match small **patches** instead of single **pixels**

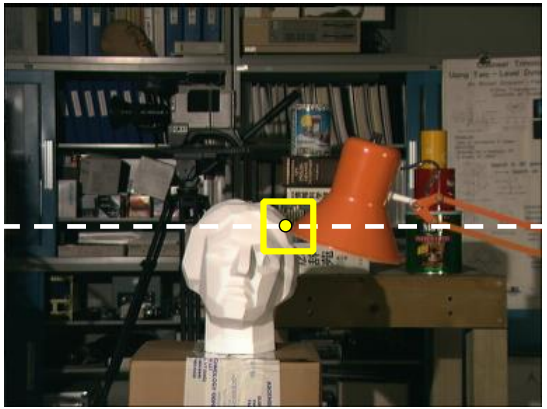
[Slide credit: Noah Snavely]

Rectified Stereo matching with SSD

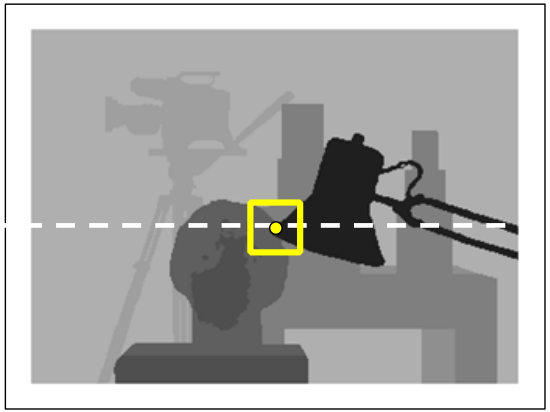


[Slide credit: Noah Snavely]

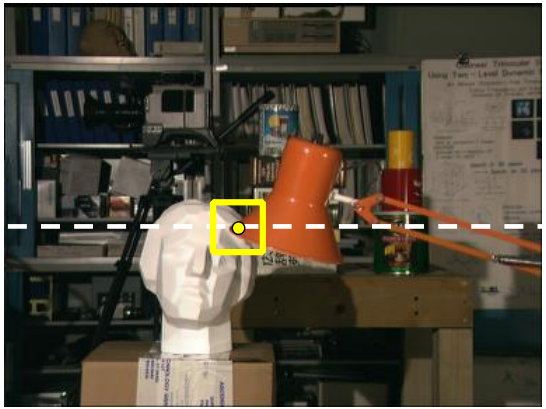
Special Case: Rectified Stereo Pair



I_1



Depth Image (D_2)



I_2

Brightness constancy assumption to find corresponding points:

$$I_2(x,y) = I_1\left(x + \underbrace{d_2(x,y)}_{\text{Disparity}}, y\right)$$

[Newcombe & Lovegrove, Geometric Reconstruction Lecture]

Stereo Correspondence as Energy Minimization

$I_1(x, y)$

$I_2(x, y)$

Pixel Error:

$e(x, y, d) = I_1(x + d, y) - I_2(x,y)$

Cost (with quadratic penalty):

$C(x, y, d) = (I_1(x + d, y) - I_2(x,y))^2$

y = 141

d

x

$C(x, y, d)$; the *disparity space image* (DSI)

[Newcombe & Lovegrove, Geometric Reconstruction Lecture]

Stereo Correspondence as Energy Minimization

Simple pixel / window matching: choose the minimum of each column in the DSI independently:

$$d(x, y) = \arg\min_{d'} C(x, y, d')$$

Find shortest path through cost volume via dynamic programming!

[Newcombe & Lovegrove, Geometric Reconstruction Lecture]

Stereo as energy minimization

- Better objective function

$$E(d) = \underbrace{E_d(d)}_{\text{match cost}} + \lambda \underbrace{E_s(d)}_{\text{smoothness cost}}$$

Want each pixel to find a good match in the other image

Adjacent pixels should (usually) move about the same amount

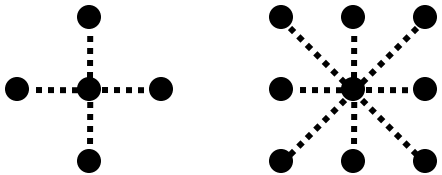
Stereo as energy minimization

$$E(d) = E_d(d) + \lambda E_s(d)$$

match cost: $$E_d(d) = \sum_{(x,y)\in I} C(x, y, d(x, y))$$

smoothness cost: $$E_s(d) = \sum_{(p,q)\in\mathcal{E}} V(d_p, d_q)$$

$\mathcal{E}$: set of neighboring pixels



4-connected neighborhood

8-connected neighborhood

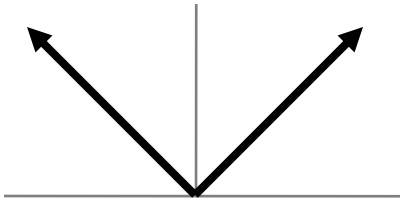
Smoothness cost

$$E_s(d) = \sum_{(p,q)\in\mathcal{E}} V(d_p, d_q)$$

How do we choose *V*?

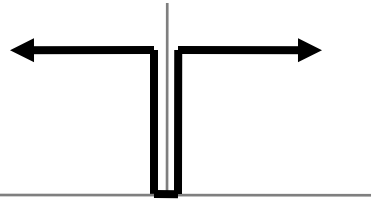
$$V(d_p, d_q) = |d_p - d_q|$$

L_1 distance



$$V(d_p, d_q) = \begin{cases} 0 & \text{if } d_p = d_q \\ 1 & \text{if } d_p \neq d_q \end{cases}$$

"Potts model"



Smoothness cost

$$E(d) = E_d(d) + \lambda E_s(d)$$

- If λ = infinity, then we only consider smoothness
- Optimal solution is a surface of constant depth/disparity
  - *Fronto-parallel* surface

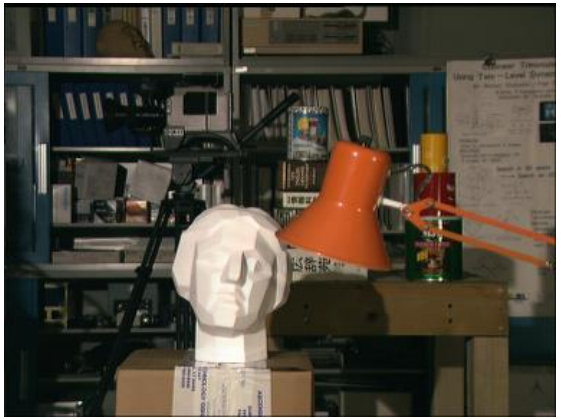
- In practice, want to balance data term with smoothness term

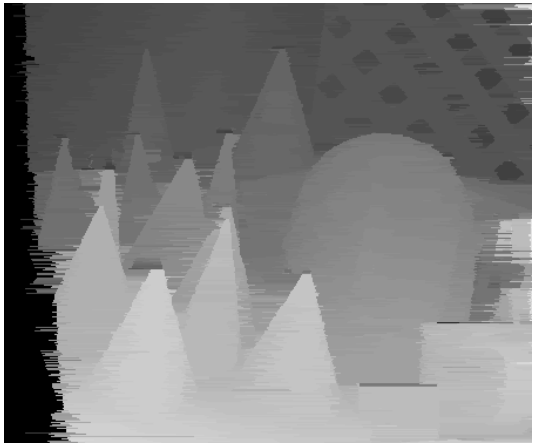
Dynamic programming

$$E(d) = E_d(d) + \lambda E_s(d)$$

- Can minimize this independently per scanline using dynamic programming (DP)

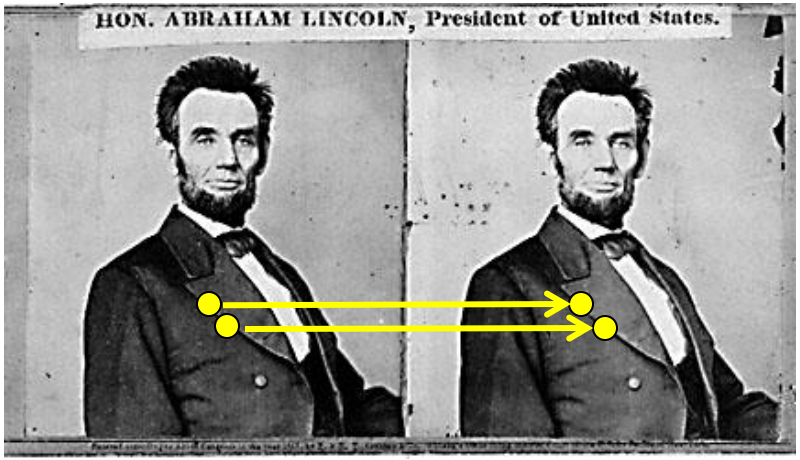
Stereo via Dynamic Programming





[Slide credit: Noah Snavely]

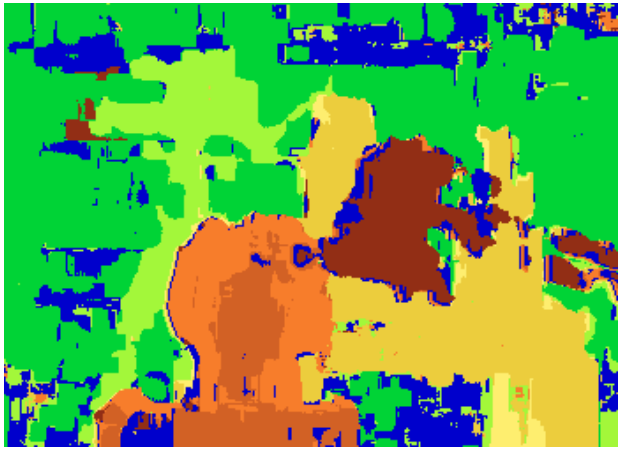
Stereo as energy minimization



- What defines a good stereo correspondence?
  1. Match quality
     - Want each pixel to find a good match in the other image
  2. Smoothness
     - If two pixels are adjacent, they should (usually) move about the same amount

[Slide credit: Noah Snavely]

Results with window search



Window-based matching
(best window size)

Graph cuts-based method

[Boykov et al., Fast Approximate Energy Minimization via Graph Cuts, ICCV 1999]

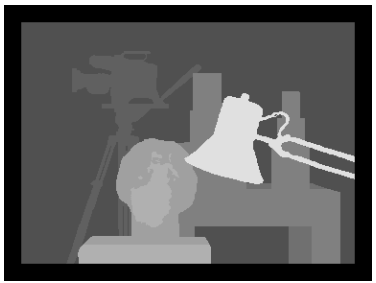
Ground truth

[Slide credit: Noah Snavely]

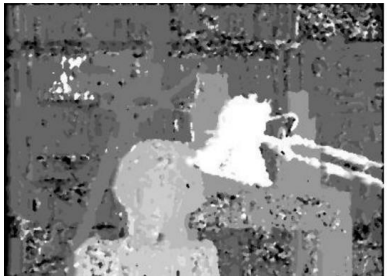
Stereo Correspondence as Energy Minimization

Effect of window size (W) for aggregating the photometric cost:

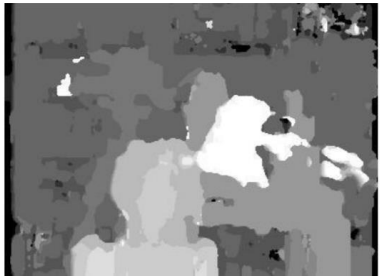
$$\sum_{(i,j)\in W} |I_1(i,j) - I_2(x+i, y+j)|$$



Ground truth



SAD W=3



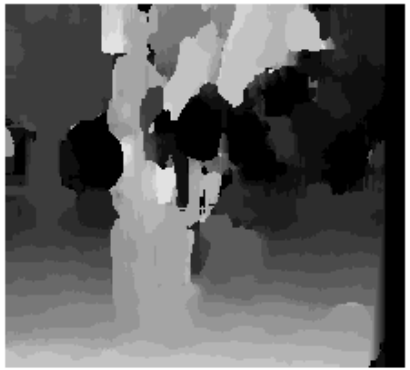
SAD W=11

SAD W=25

[Newcombe & Lovegrove, Geometric Reconstruction Lecture]

Window size

W = 3



W = 20

Effect of window size

- Smaller window
  - \+ more detail
  - \- more noise
- Larger window
  - \+ less noise
  - \- less detail

Better results with *adaptive window*

- T. Kanade and M. Okutomi, *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment*, ICRA 1991.
- D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. IJCV, July 1998

[Slide credit: Noah Snavely]

Stereo Correspondence as Energy Minimization

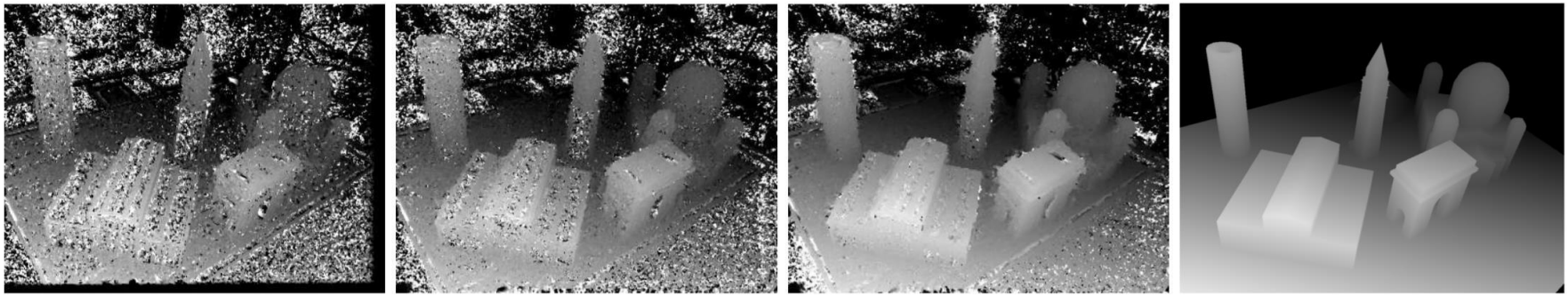
The design of the cost function, including *window size* for aggregation, *image error* function and *penalty* can improve quality of correspondence:

Similarity Measure	Formula	
Sum of Absolute Differences (SAD)	$\sum_{(i,j)\in W} \lvert I_1(i,j) - I_2(x+i,y+j)\rvert$	Outlier robustness
Sum of Squared Differences (SSD)	$\sum_{(i,j)\in W} \left(I_1(i,j) - I_2(x+i,y+j)\right)^2$	Outlier sensitive
Zero-mean SAD	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \bar{I}_1(i,j) - I_2(x+i,y+j) + \bar{I}_2(x+i,y+j)\rvert$	Invariant to additive brightness changes.
Locally scaled SAD	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \frac{\bar{I}_1(i,j)}{\bar{I}_2(x+i,y+j)} I_2(x+i,y+j)\rvert$	Somewhat invariant to brightness changes.
Normalized Cross Correlation (NCC)	$\frac{\sum_{(i,j)\in W} I_1(i,j).I_2(x+i,y+j)}{\sqrt[2]{\sum_{(i,j)\in W} I_1^2(i,j).\sum_{(i,j)\in W} I_2^2(x+i,y+j)}}$	Invariant to additive and multiplicative brightness changes.

[Newcombe & Lovegrove, Geometric Reconstruction Lecture]

Multi View Stereo

- How to Integrate more information from Multiple Views?



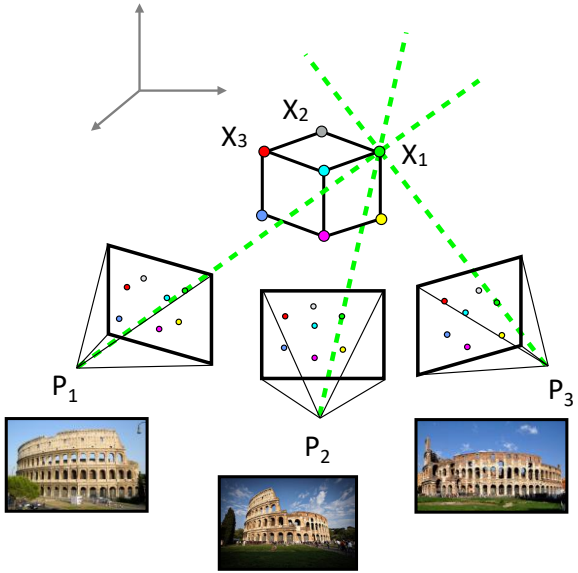
(a) 2 views (b) 5 views (c) 20 views (d) Ground Truth

[Newcombe & Lovegrove, Geometric Reconstruction Lecture]

Today's Agenda

- Shape from X
- Camera Calibration
- 2-view Case:
  - Epipolar Geometry
  - Stereo Vision
  - Image Rectification
  - Stereo Reconstruction
- Multi-view Case: Structure from Motion

Multi-View Geometry



Bundle Adjustment

Bundle Adjustment

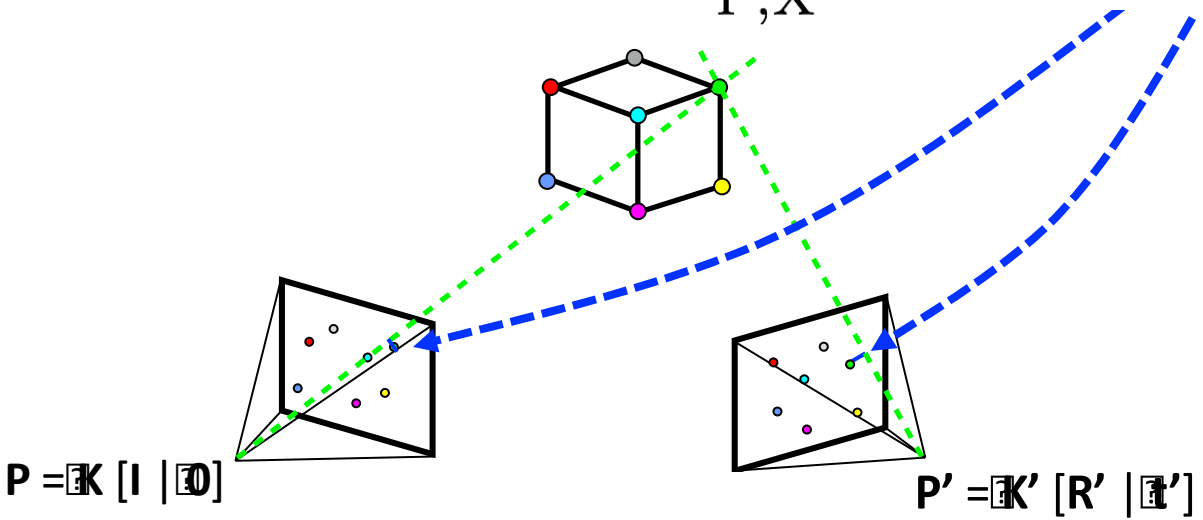
Goal: Adjust all bundles of viewing rays to minimize the reprojection error of 3D points in all 2D input images.

- Non-linear refinement of structure and motion 3x3 matrix

π	Projection mapping
P	Projection matrices
X	3D Points
x	2D Points

- Minimize reprojection error:

$$\min_{\mathrm{P},\mathrm{X}} \|x - \pi(\mathrm{P}, \mathrm{X})\|$$



Challenges of Bundle Adjustment

Initialization:

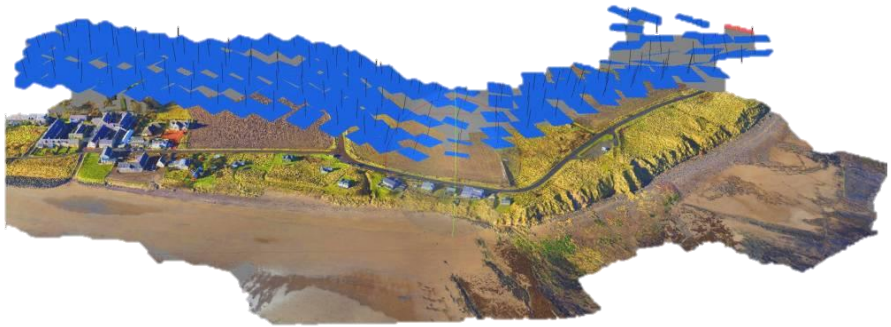
- Highly **non-convex** energy
- Good initialization is often crucial, e.g. factorization techniques or incremental strategy
- Incremental bundle adjustment initializes with carefully selected two-view reconstruction and iteratively adds new images

Optimization:

- For large number of cameras bundle adjustment is computationally demanding (cubic in #unknowns = 3D points, camera parameters)
- Leverage sparsity: The optimization problem is typically **sparse** since 3D points are only visible in a small amount of views. (→ Ceres Solver)

[Slide Credit: CV Lecture, Andreas Geiger] [Ceres-Solver, http://ceres-solver.org] [Triggs et al., "Bundle Adjustment – A Modern Synthesis"]

Structure from Motion



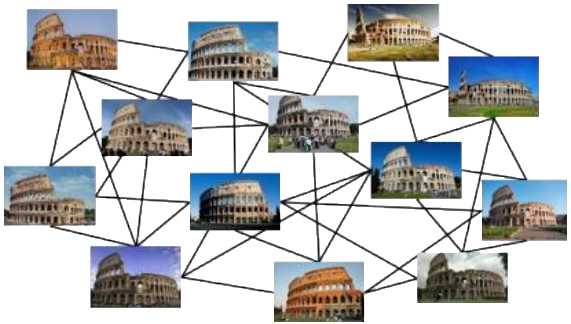
Google Earth / Maps

[https://www.flyingeye.fr/wp-content/uploads/2014/06/photogrammetrie.jpg]

Batch-based 3D Reconstruction Pipeline

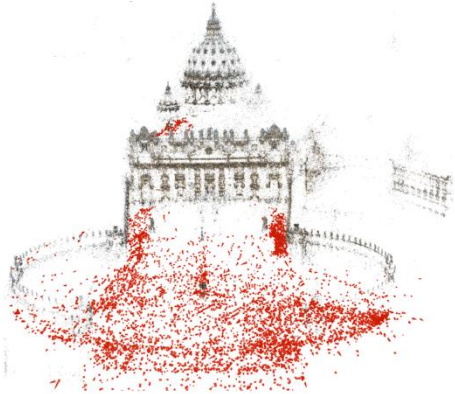
Image Set

Image Association



Scene Graph

SfM



Sparse Model

MVS

(Semi-) Dense Model

[Large-scale 3D Modeling Tutorial, CVPR 2017]

Structure-from-Motion (SfM)

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

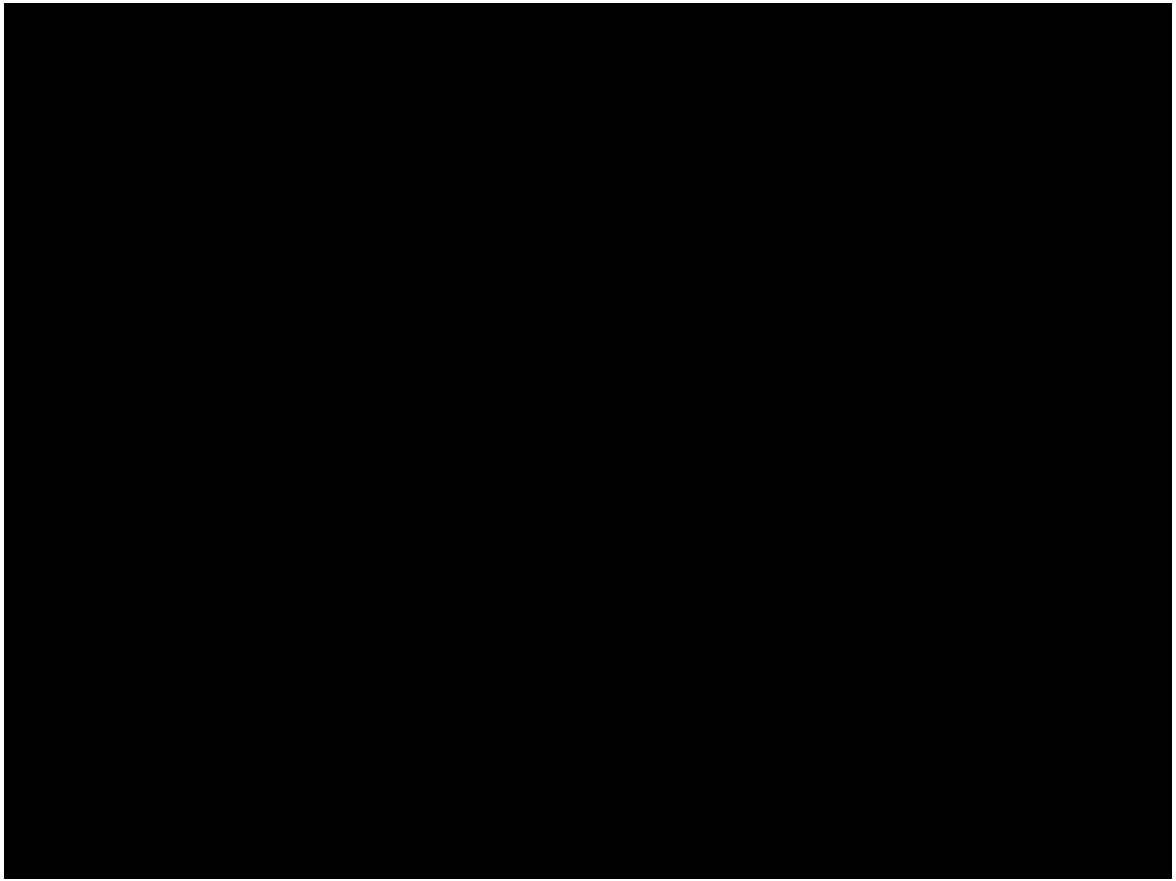
- Mutli view geometry
- Structure from motion

Unknow

Unknow

Unknow

Structure-from-Motion (SfM) - Pipeline



1. Capture many images of a scene
2. Find and describe feature points
3. Search for feature point correspondences across images
4. Compute camera positions
5. Triangulate feature points

Structure-from-Motion (SfM)

Rome dataset

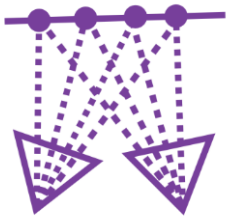
74,394 images

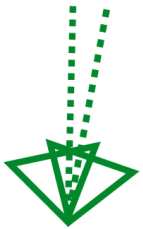
[Johannes L. Schönberger, Jan-Michael Frahm. **Structure-from-Motion Revisited**. CVPR, 2016]

Two-view Geometry

Model Selection

General	Planar	Panoramic
• Fundamental matrix F (*uncalibrated*) • Essential matrix E (*calibrated*)	• Homography H	• Homography H
• 7 correspondences • 5 correspondences	• 4 correspondences	• 4 correspondences





[Large-scale 3D Modeling Tutorial, CVPR 2017]

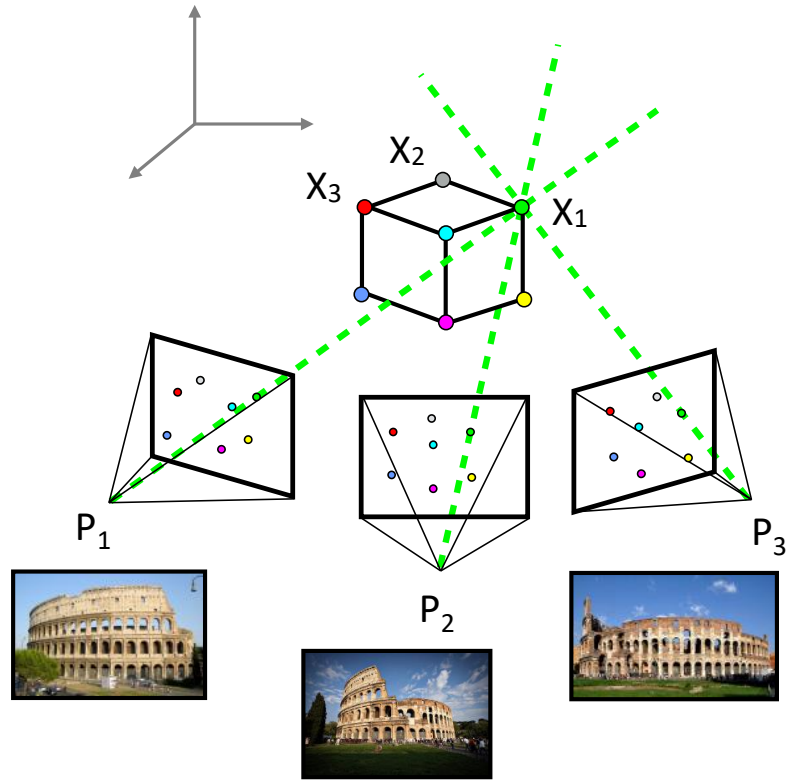
Structure-from-Motion (SfM)

Joint estimation of

- (Scene) structure X_i
- Cameras P_j

… from (camera) motion

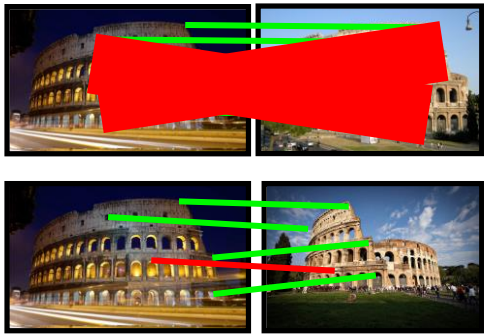
- images (of the same scene) from different viewpoints



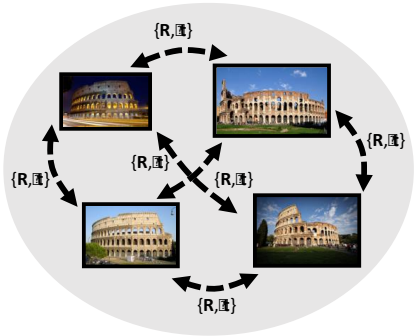
[Large-scale 3D Modeling Tutorial, CVPR 2017]

Structure-from-Motion (SfM)

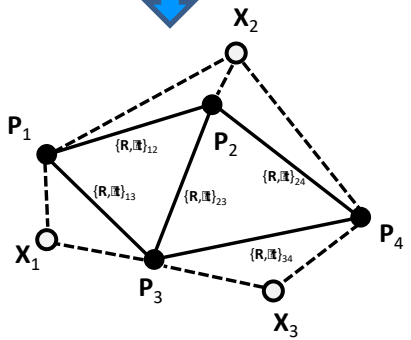
Image Set



Inlier/Outlier
Corresondences

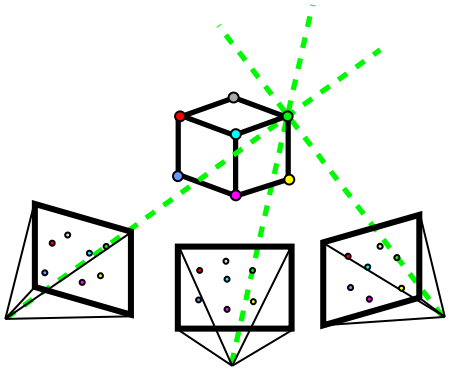


Pairwise (Relative) Relations



Scene Graph

- Scale of t unknown
- Contains outlier correspondences and image pairs



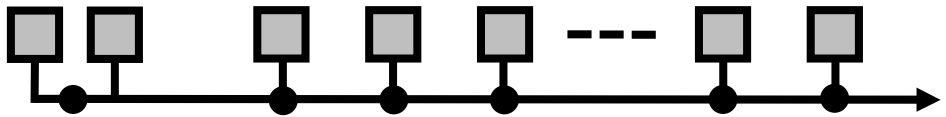
Absolute Camera Locations
& Scene Structure

[Large-scale 3D Modeling Tutorial, CVPR 2017]

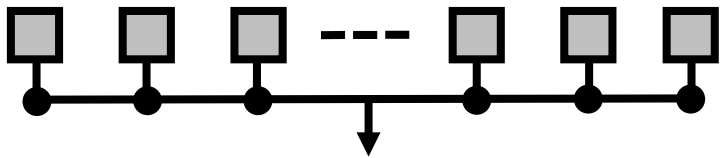
Structure-from-Motion (SfM)

3 paradigms

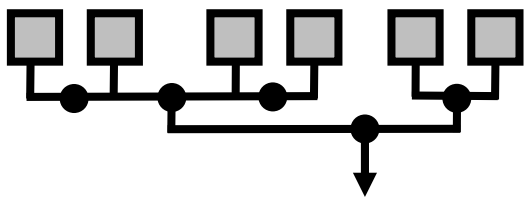
- Incremental



- Global



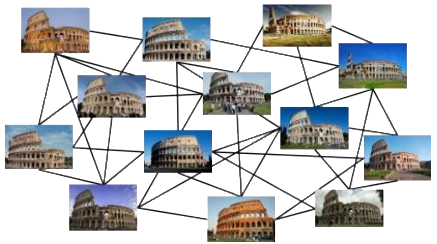
- Hierarchical

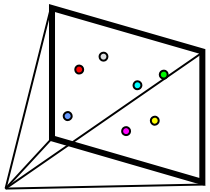


[Large-scale 3D Modeling Tutorial, CVPR 2017]

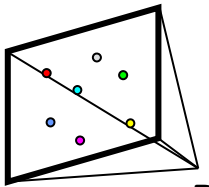
Incremental SfM

- Initialization
  1. Choose two non-panoramic views $(\|t\| \neq 0)$





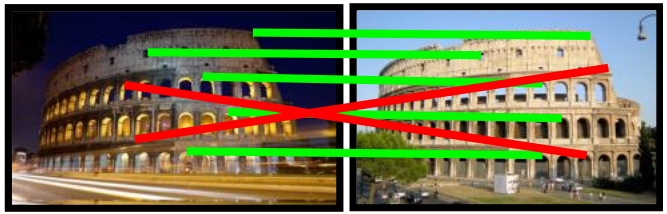
P = K [I | 0]

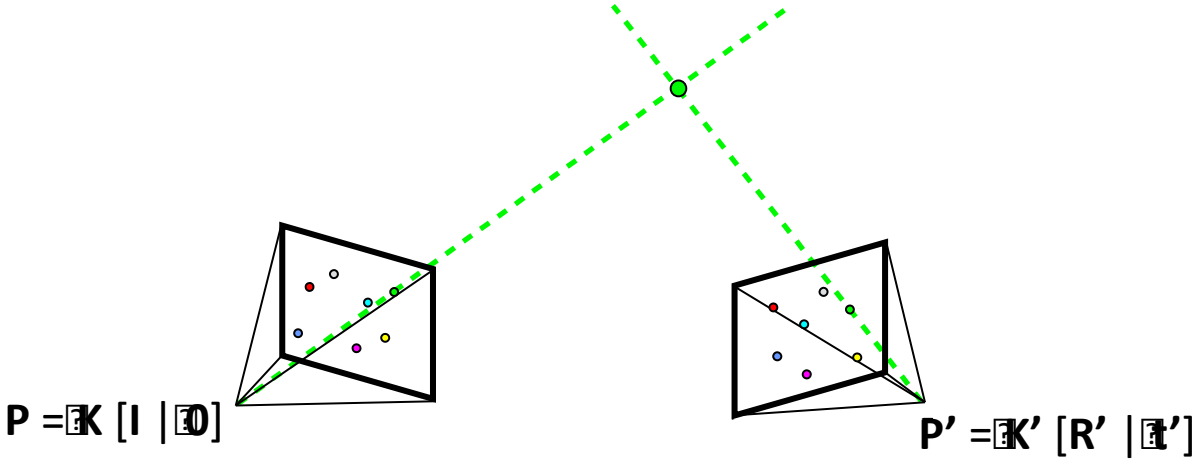


P' = K' [R' | t']

Incremental SfM

- Initialization
  1. Choose two non-panoramic views $(\|t\| \neq 0)$
  2. Triangulate inlier correspondences

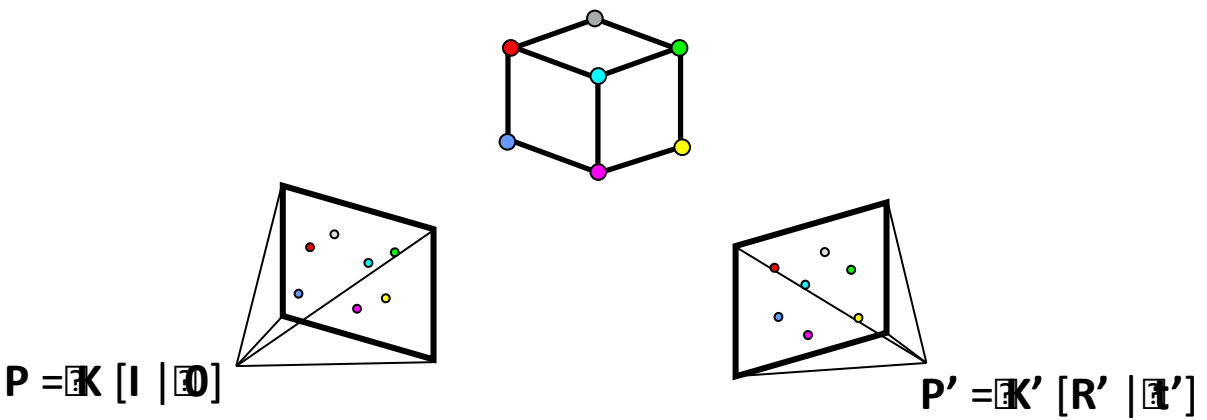




[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

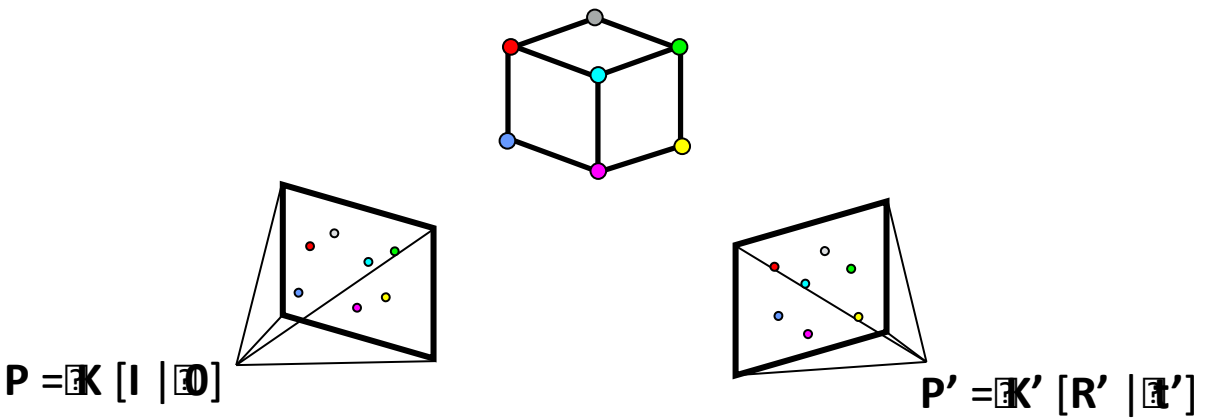
- Initialization
  1. Choose two non-panoramic views $(\|t\| \neq 0)$
  2. Triangulate inlier correspondences



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

- Initialization
  1. Choose two non-panoramic views $(\|t\| \neq 0)$
  2. Triangulate inlier correspondences
  3. Bundle adjustment

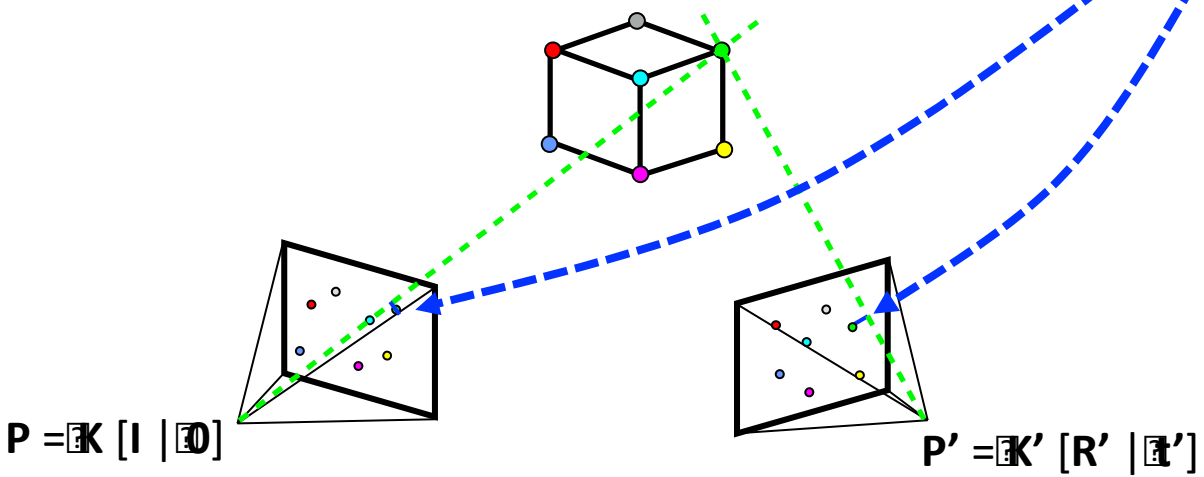


[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

- Bundle adjustment
  1. Non-linear refinement of structure and motion
  2. Minimize reprojection error:

$$\min_{\mathrm{P,X}} \|x - \pi(\mathrm{P}, \mathrm{X})\|$$



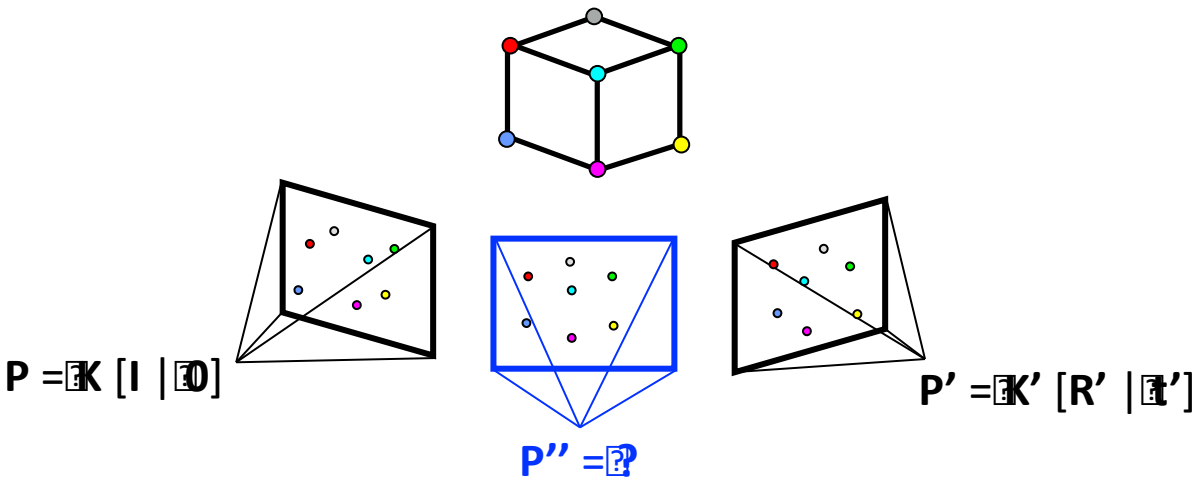
[Ceres-Solver, http://ceres-solver.org/]

[Triggs et al., "Bundle Adjustment – A Modern Synthesis"]

[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

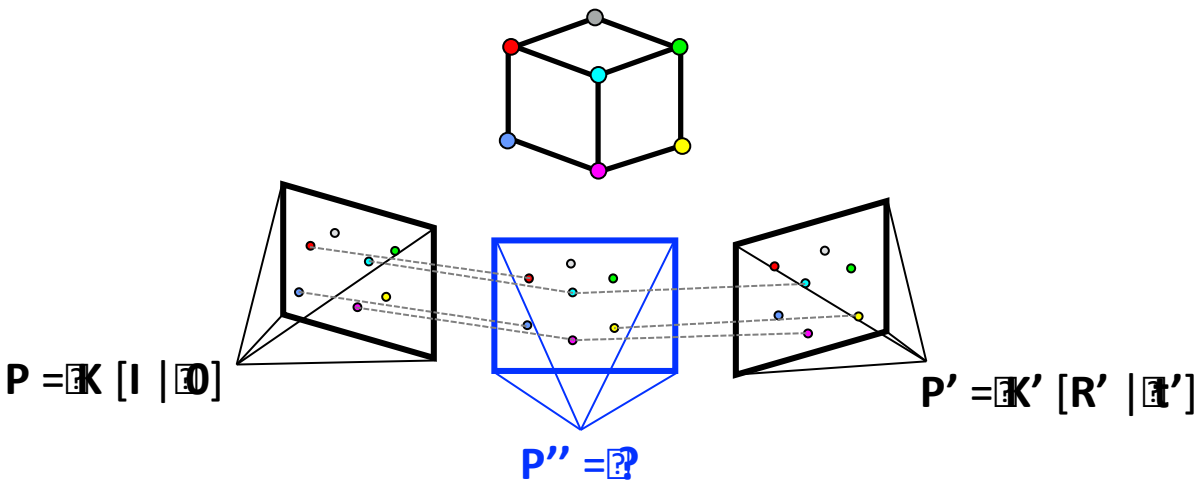
- Absolute camera registration



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

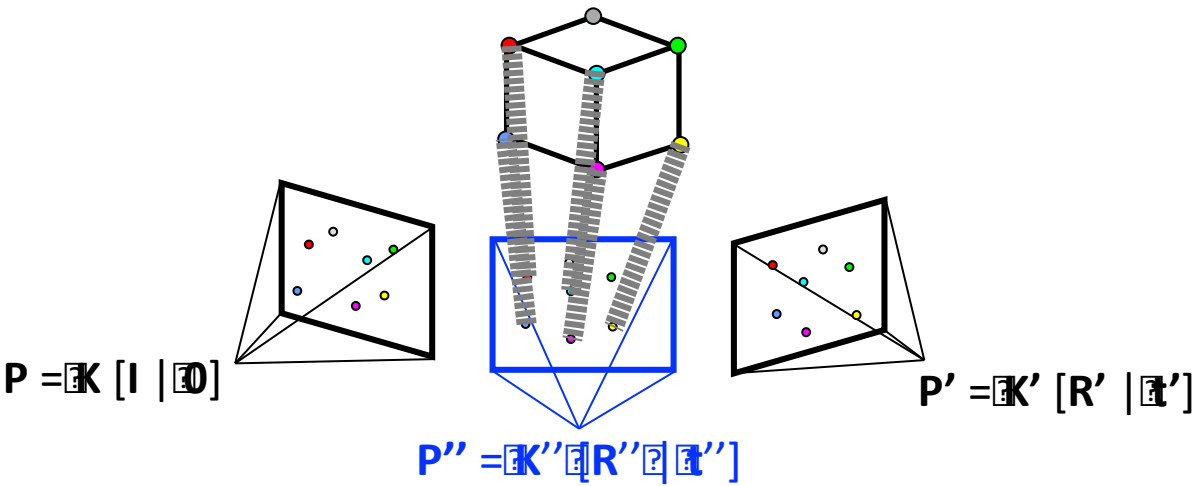
- Absolute camera registration
  1. Find 2D-3D correspondences



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

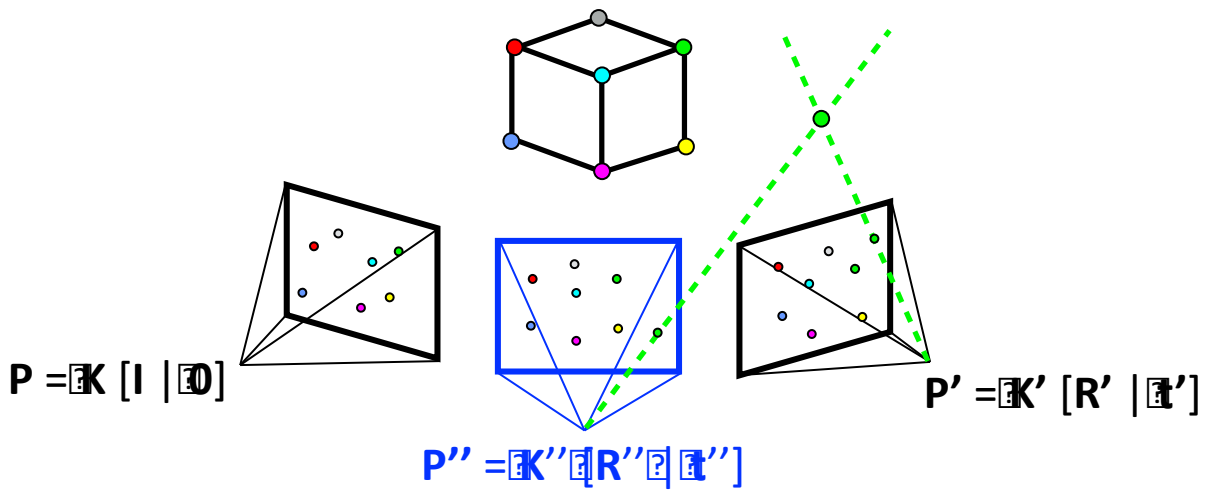
- Absolute camera registration
  1. Find 2D-3D correspondences
  2. Solve Perspective-n-Point problem



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

- Absolute camera registration
  1. Find 2D-3D correspondences
  2. Solve Perspective-n-Point problem
  3. Triangulate new points

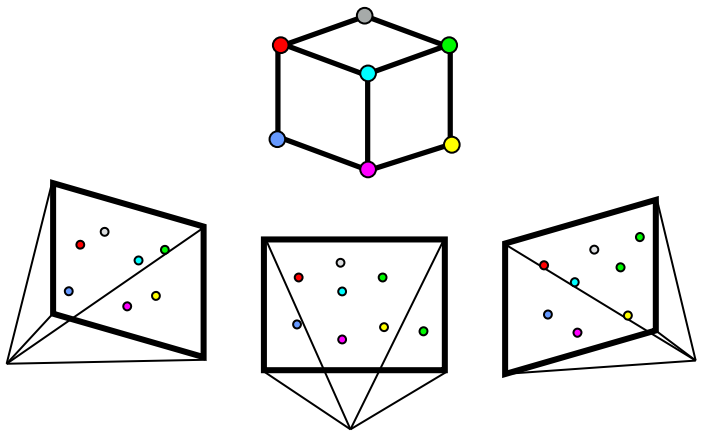


[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

- Bundle Adjustment

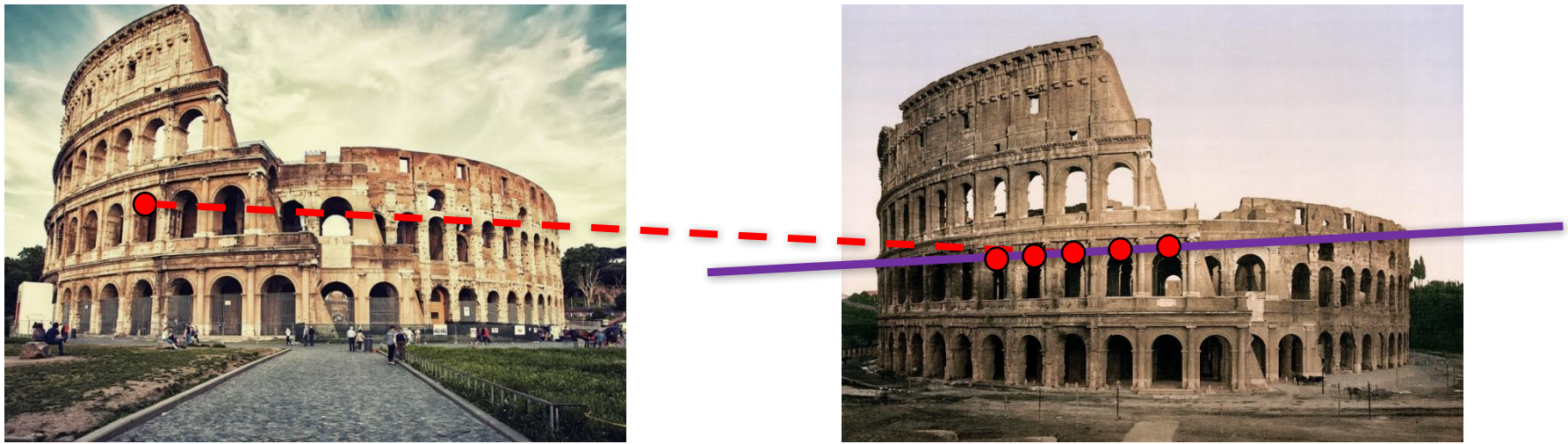
$$\min_{\mathrm{P,X}} \|x - \pi(\mathrm{P}, \mathrm{X})\|$$



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

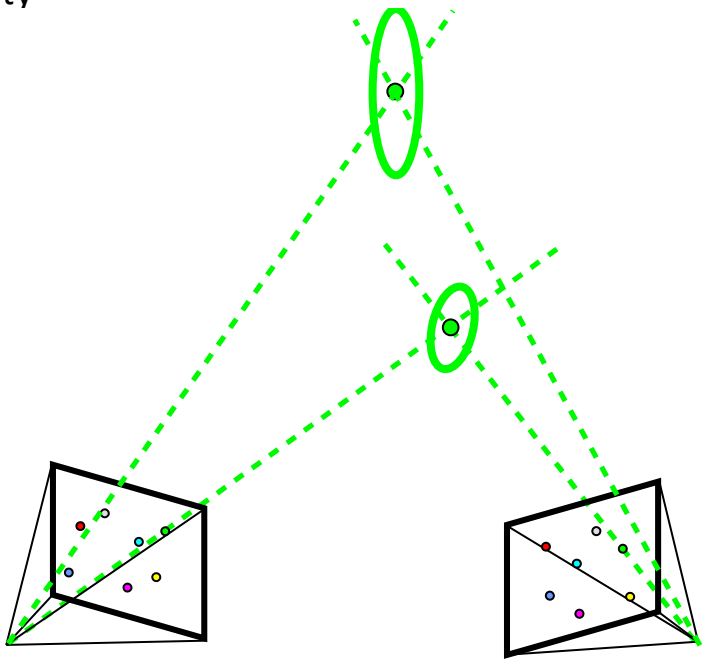
- Outlier filtering
  - Remove points with large reprojection error



[Large-scale 3D Modeling Tutorial, CVPR 2017]

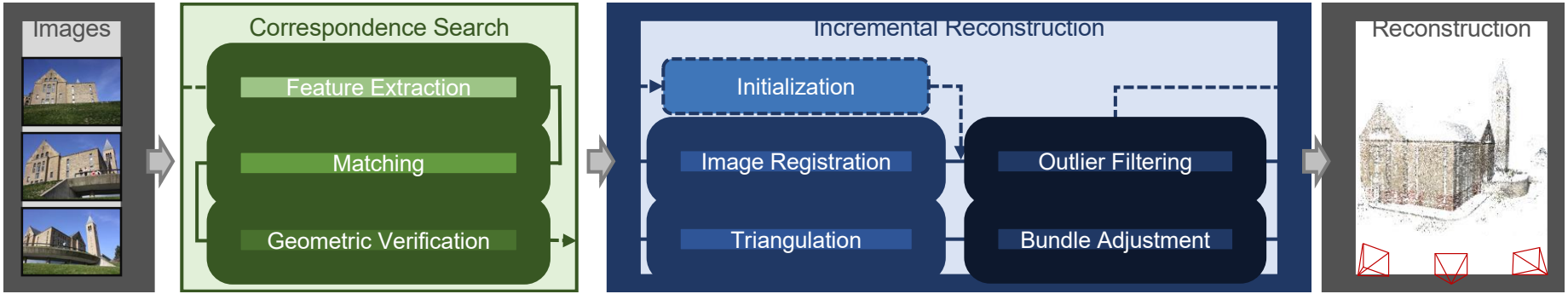
Incremental SfM

- Outlier filtering
  - Remove points with large reprojection error
  - Remove points at “infinity”



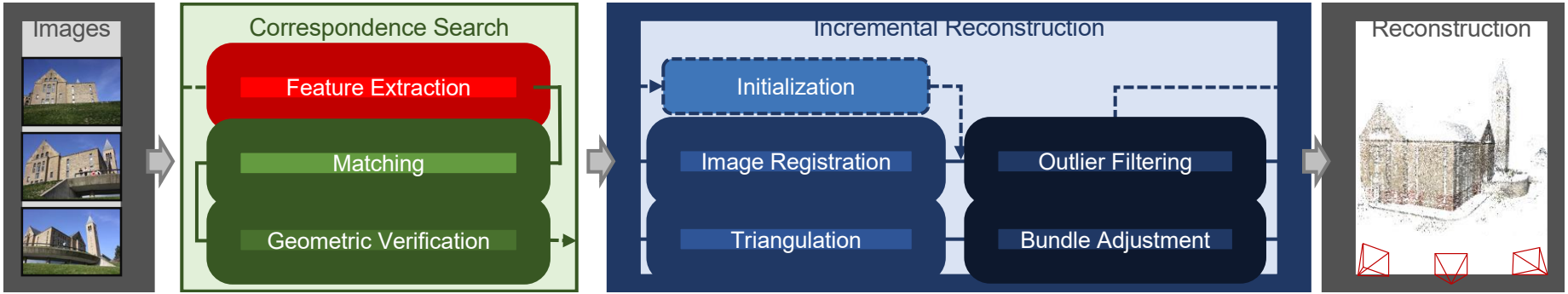
[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM



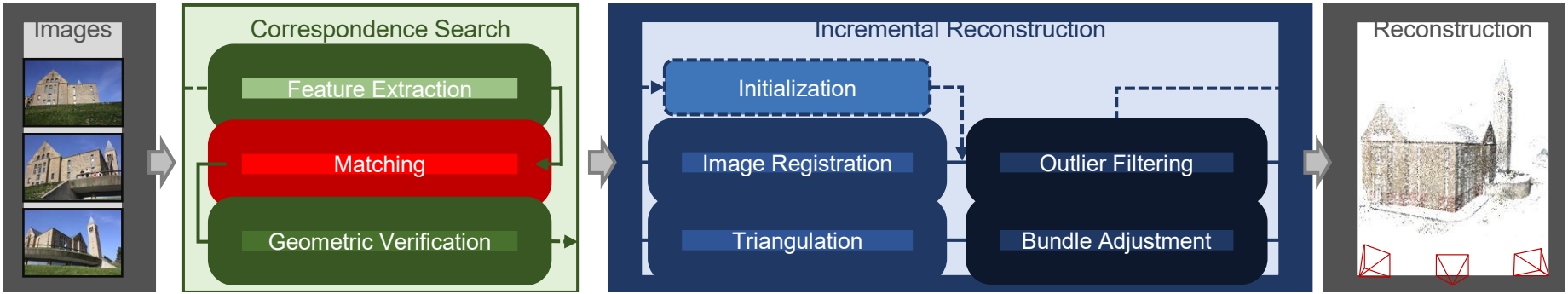
[Large-scale 3D Modeling Tutorial, CVPR 2017]

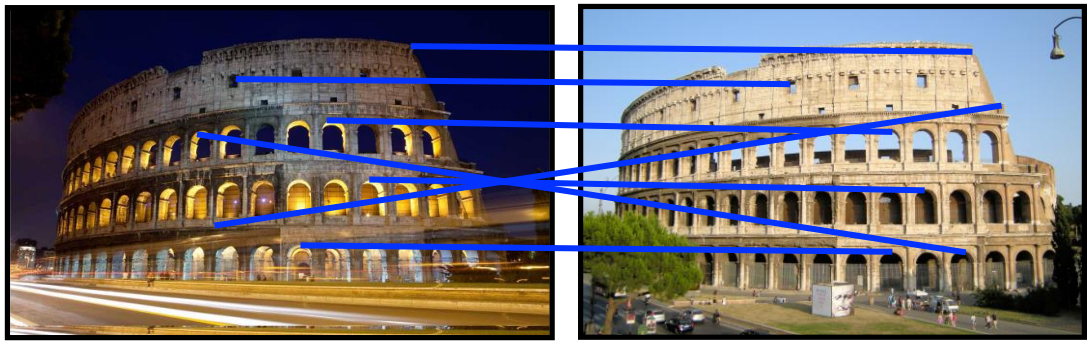
Incremental SfM



[Large-scale 3D Modeling Tutorial, CVPR 2017]

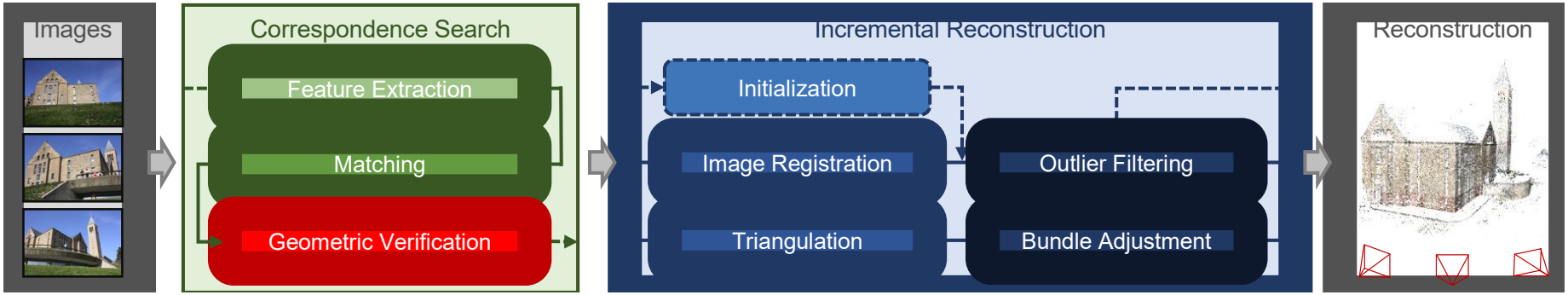
Incremental SfM



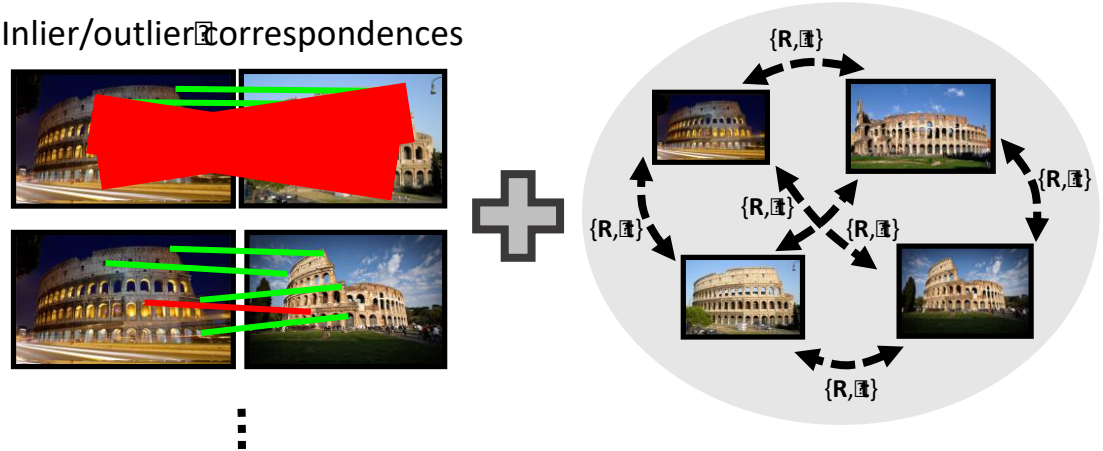


[Large-scale 3D Modeling Tutorial, CVPR 2017]

Incremental SfM

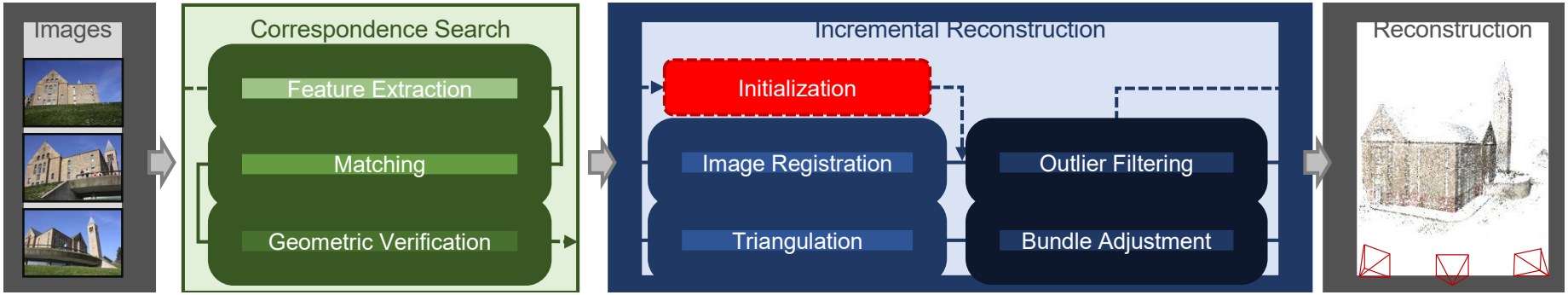


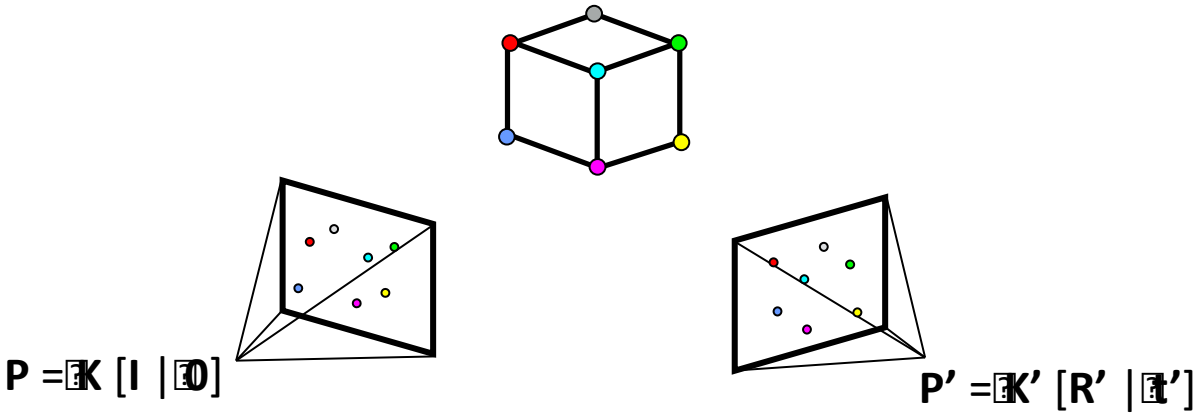
Inlier/outlier correspondences



[Large-scale 3D Modeling Tutorial, CVPR 2017]

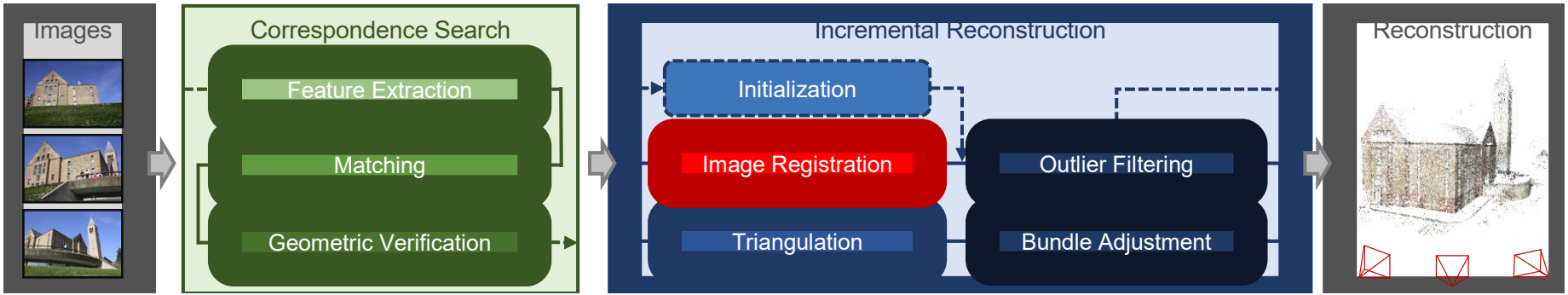
Incremental SfM

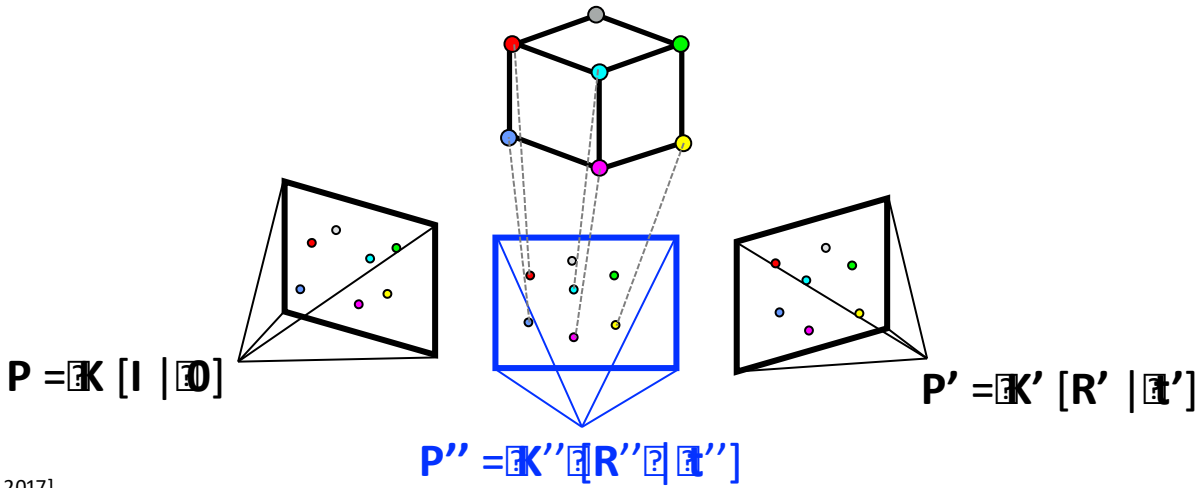




[Large-scale 3D Modeling Tutorial, CVPR 2017]

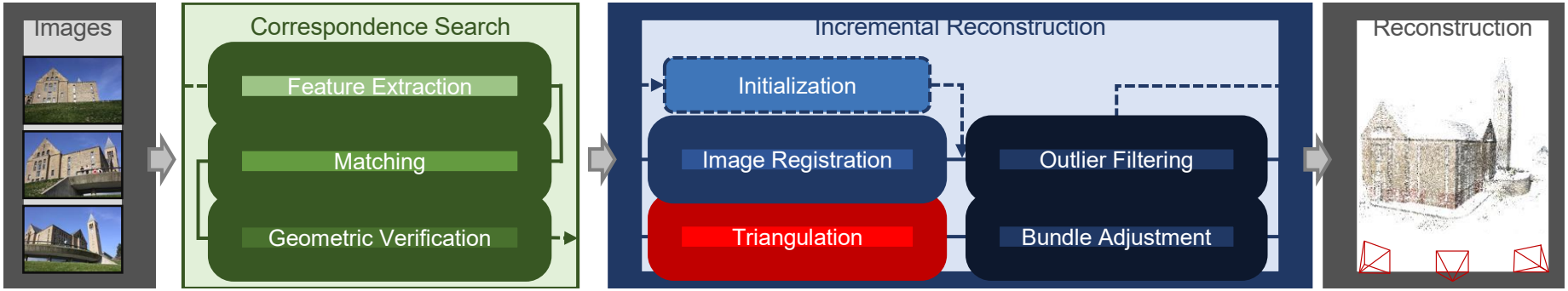
Incremental SfM

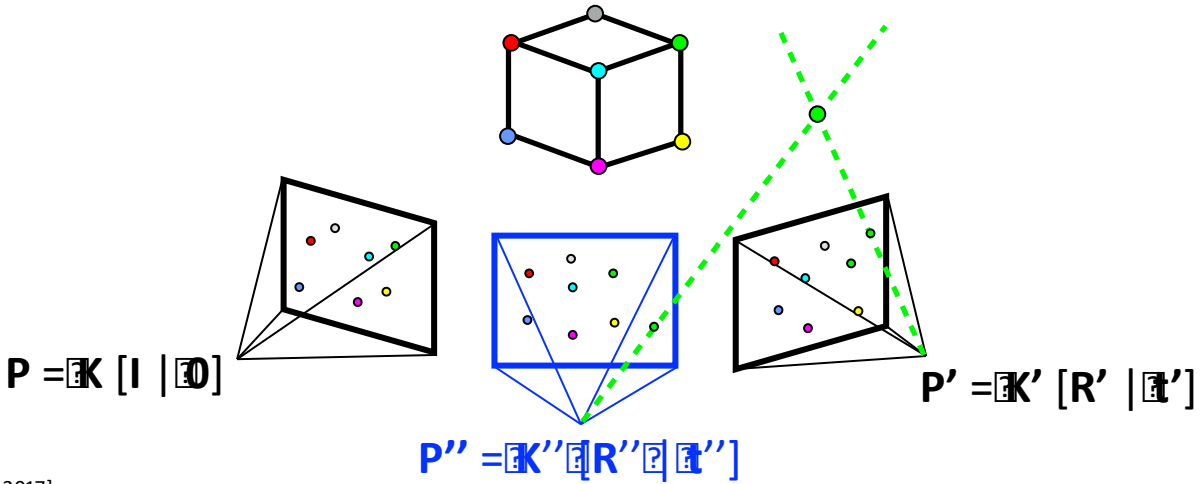




[Large-scale 3D Modeling Tutorial, CVPR 2017]

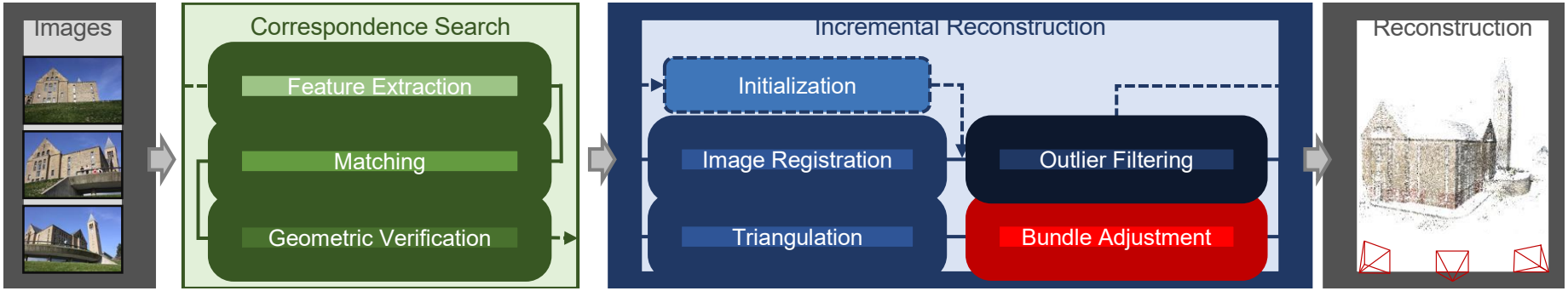
Incremental SfM





[Large-scale 3D Modeling Tutorial, CVPR 2017]

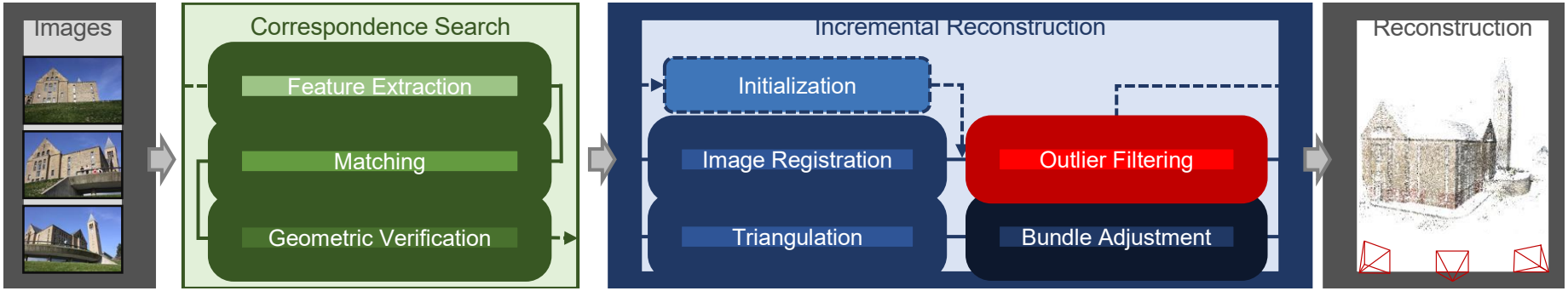
Incremental SfM

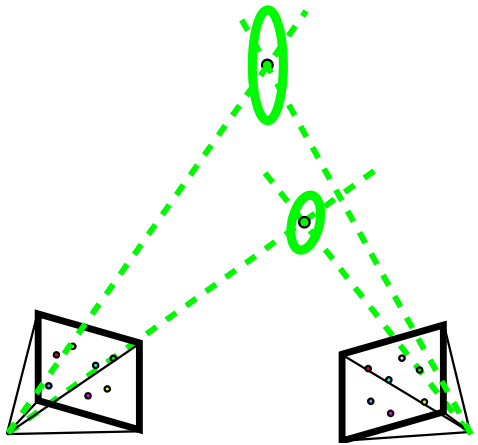


$$\min_{\mathrm{P,X}} \|x - \pi(\mathrm{P}, \mathrm{X})\|$$

[Large-scale 3D Modeling Tutorial, CVPR 2017]

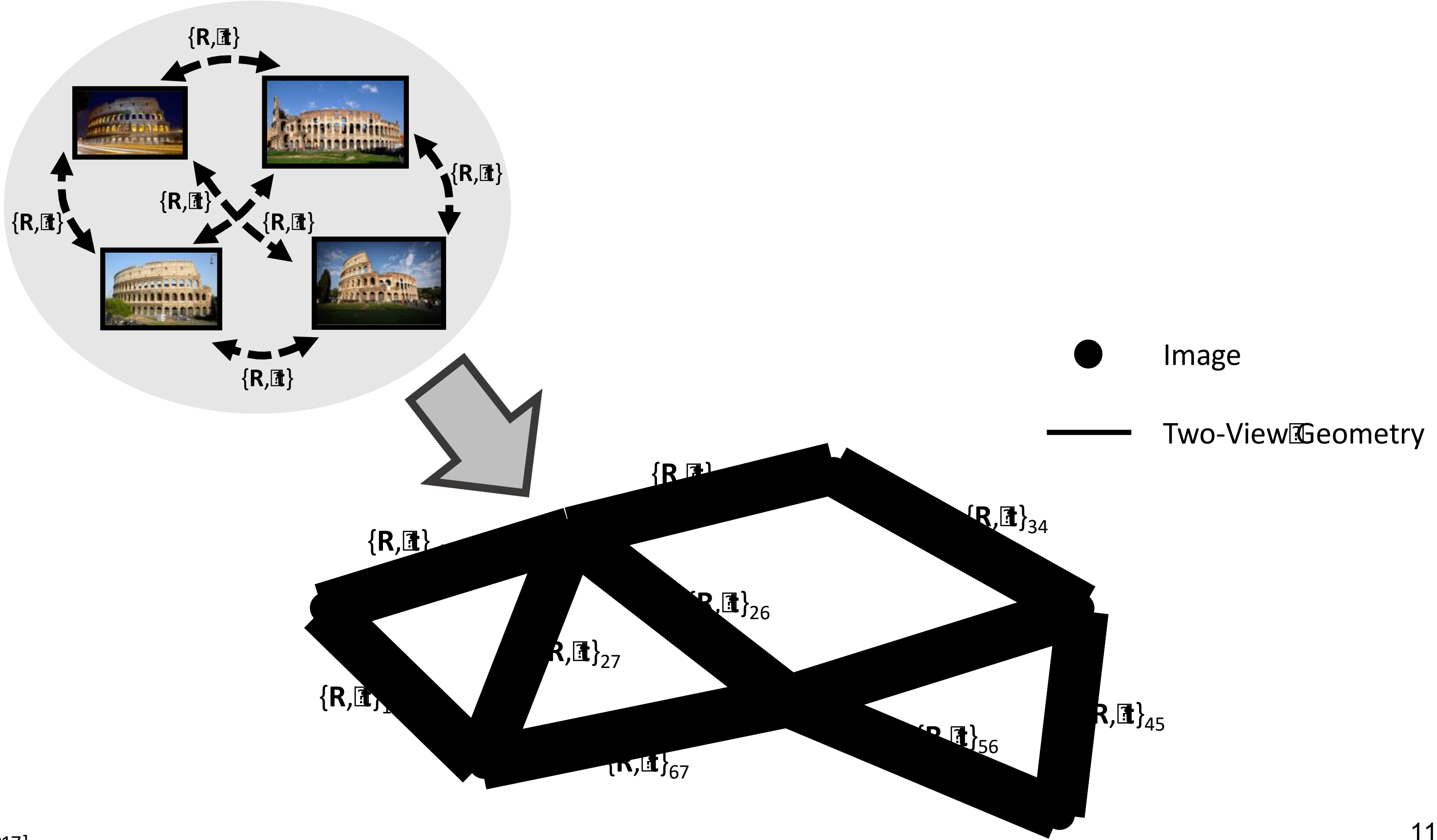
Incremental SfM





[Large-scale 3D Modeling Tutorial, CVPR 2017]

Global SfM

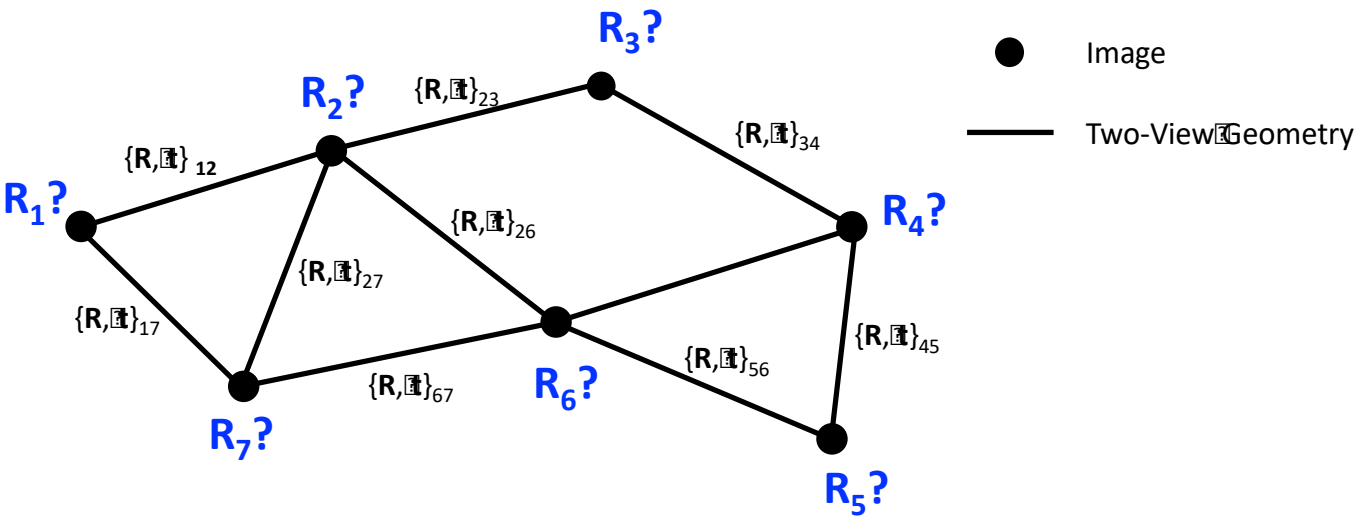


[Large-scale 3D Modeling Tutorial, CVPR 2017]

Global SfM

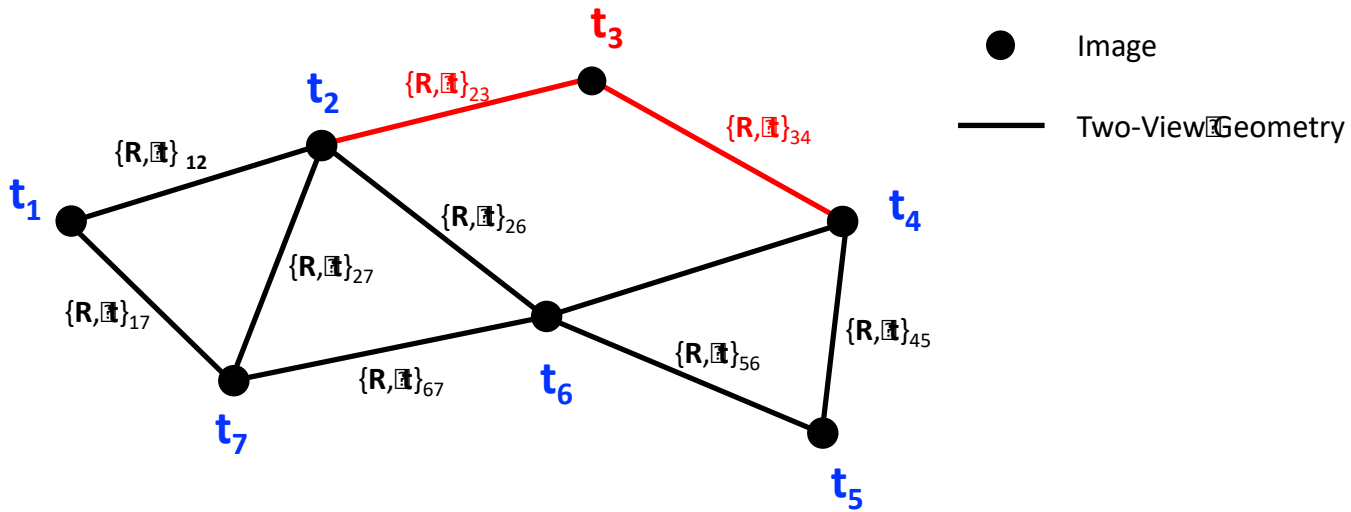
1. Estimate and filter global rotations: $\min_{\theta} \|R_{ij} - R_j R_i^T\|$

[Chatterje and Govindu 2013, "Efficient and Robust Large-Scale Rotation Averaging"]



Global SfM

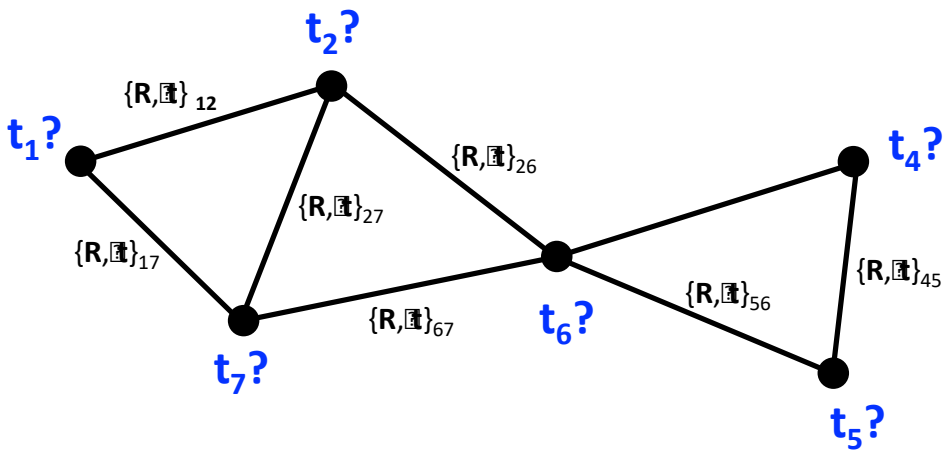
1. Estimate and filter global rotations: $\min_{R} \sum_{ij} \|R_{ij} - R_j R_i^T\| \quad \forall \|R_{ij} - R_j R_i^T\| > d$



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Global SfM

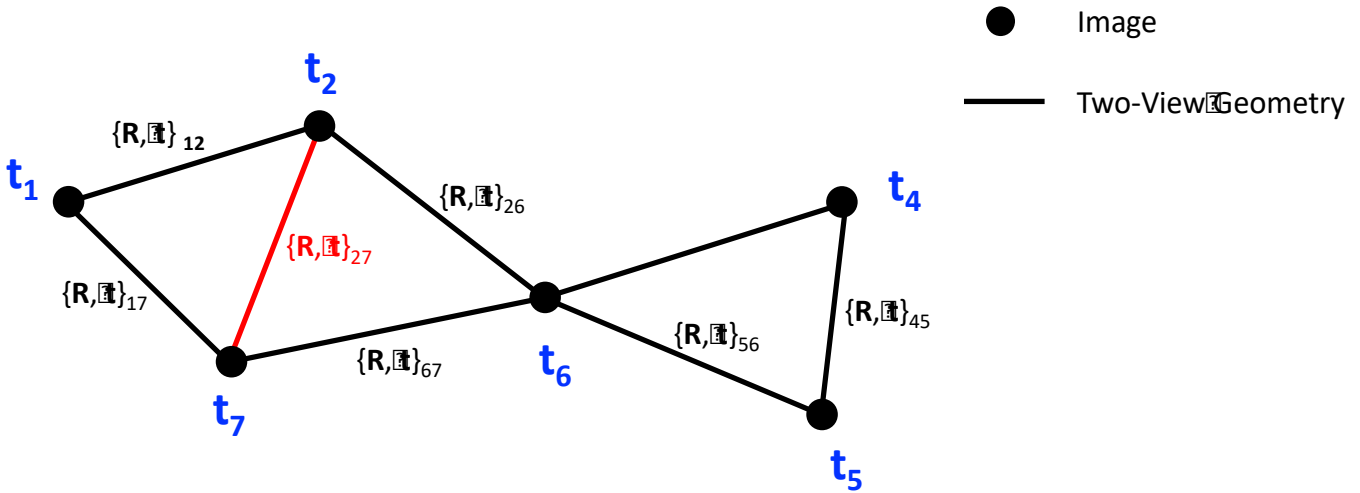
1. Estimate and filter global rotations: $\min_{R} \|R_{ij} - R_j R_i^T\| \quad \|R_{ij} - R_j R_i^T\| > t$
2. Estimate and filter global translations: $\min_{t} \left\| t_{ij} - \frac{t_j - t_i}{\|t_j - t_i\|} \right\|$



[Large-scale 3D Modeling Tutorial, CVPR 2017]

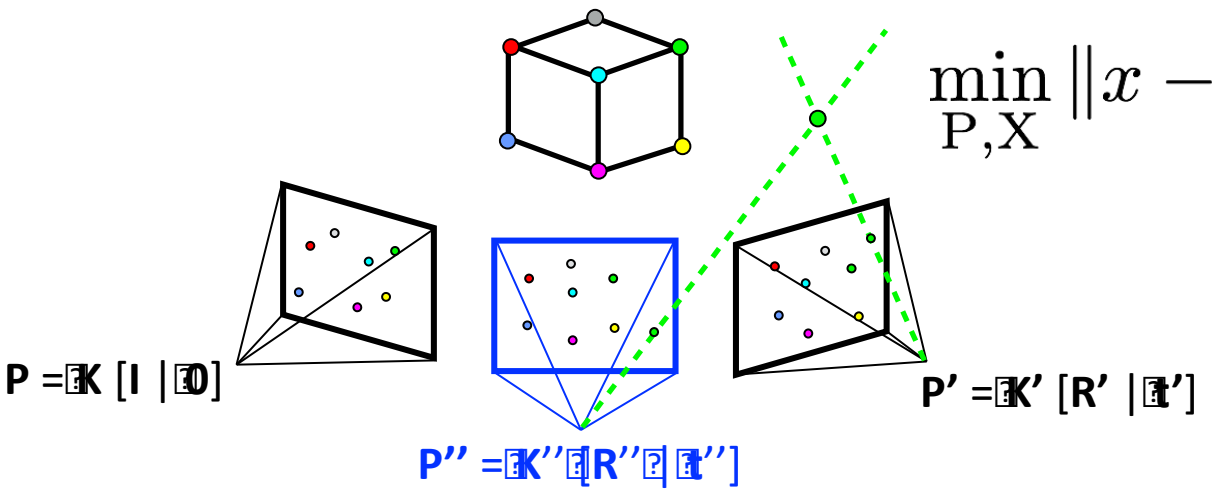
Global SfM

1. Estimate and filter global rotations: $\min_{R} \|R_{ij} - R_j R_i^T\| \quad \|R_{ij} - R_j R_i^T\| > \epsilon$
2. Estimate and filter global translations: $\min_{t} \left\| t_{ij} - \frac{t_i - t_j}{\|t_i - t_j\|} \right\| \quad \left\| t_{ij} - \frac{t_i - t_j}{\|t_i - t_j\|} \right\| > \epsilon$



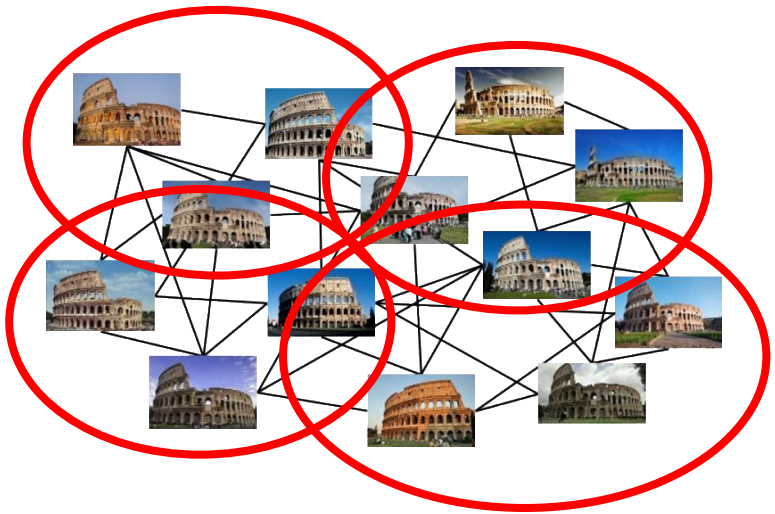
Global SfM

1. Estimate and filter global rotations: $\min_{R} \|R_{ij} - R_j R_i^T\|$ $\|R_{ij} - R_j R_i^T\| > \epsilon$
2. Estimate and filter global translations: $\min_{t} \left\| t_{ij} - \frac{t_i - t_j}{\|t_i - t_j\|} \right\|$ $\left\| t_{ij} - \frac{t_i - t_j}{\|t_i - t_j\|} \right\| > \epsilon$
3. Triangulate and refine with bundle adjustment:



Hierarchical SfM

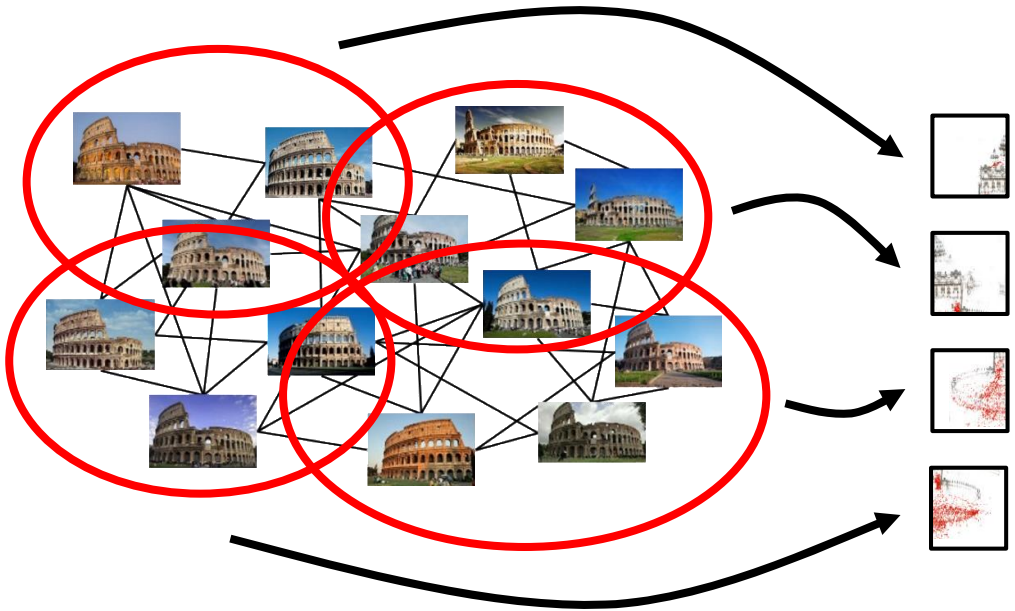
1. Hierachical clustering of scene graph



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Hierarchical SfM

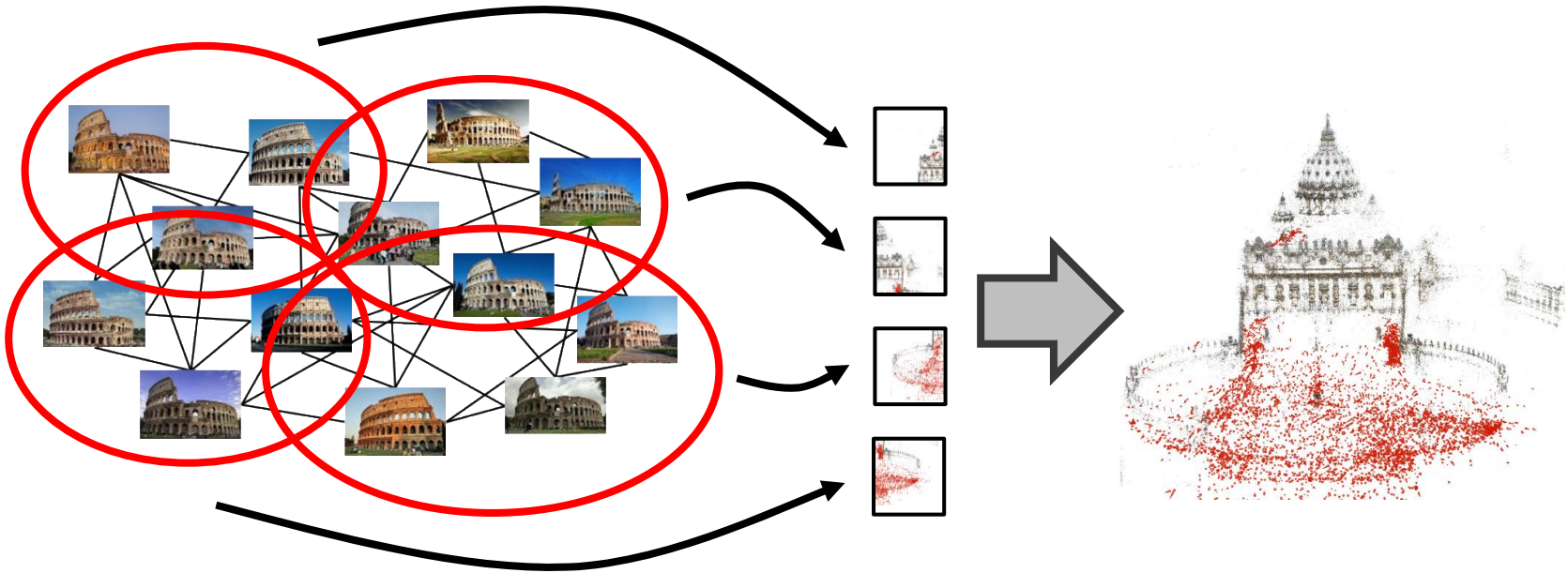
1. Hierachical clustering of scene graph
2. Reconstruct clusters independently



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Hierarchical SfM

1. Hierachical clustering of scene graph
2. Reconstruct clusters independently
3. Merge clusters using similarity transformations



[Large-scale 3D Modeling Tutorial, CVPR 2017]

Structure-from-Motion

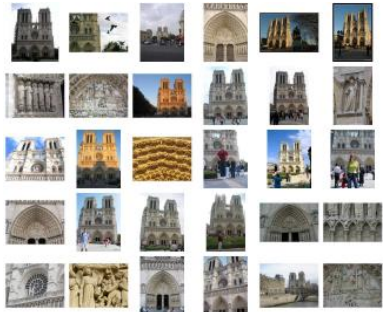
Method	Efficiency	Robustness	Accuracy
Incremental	-	++	+
Global	+	+	+
Hierarchical	++	-	-

[Large-scale 3D Modeling Tutorial, CVPR 2017]

SfM Applications

Photo Tourism

(SIGGRAPH 2006)

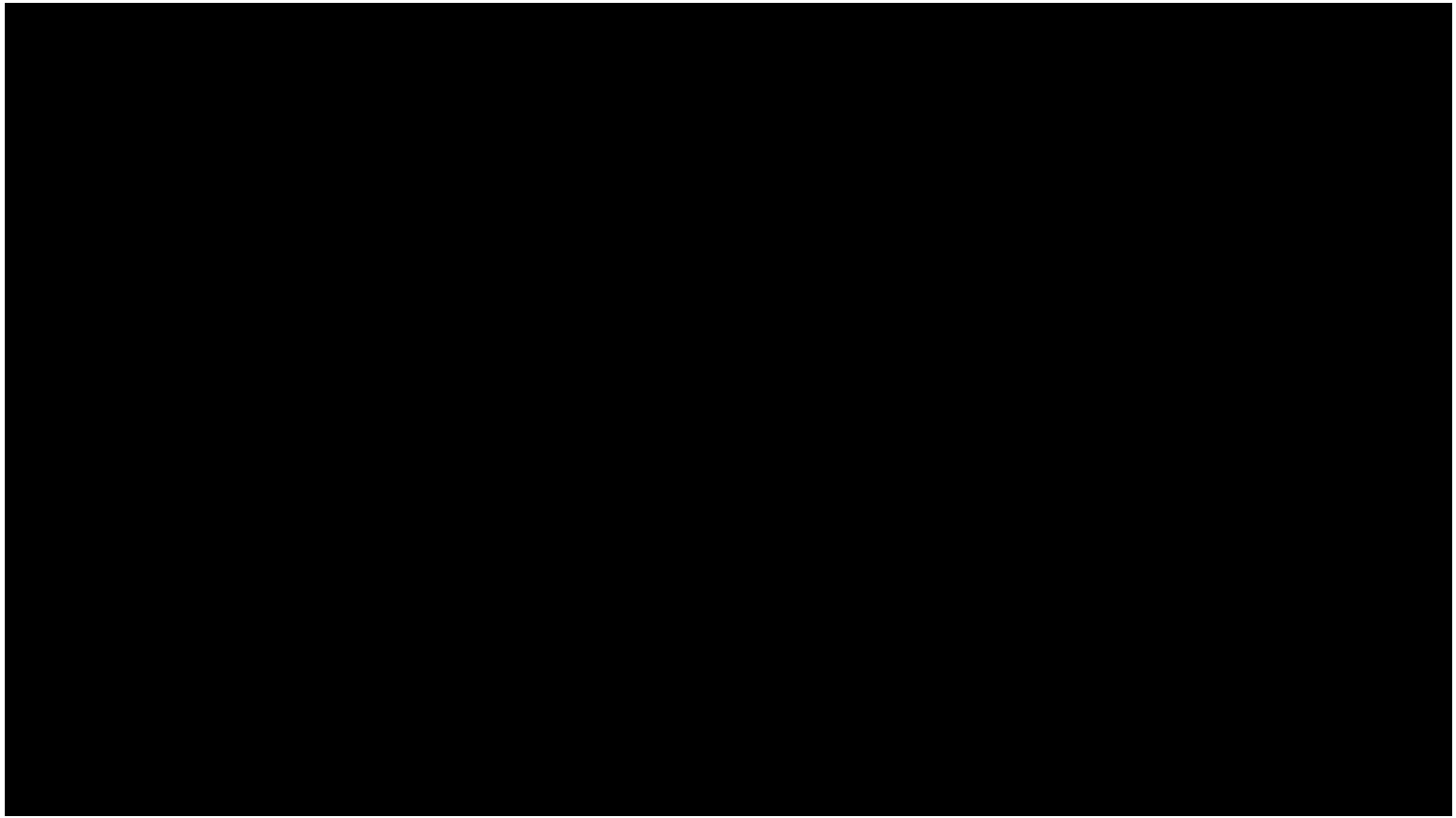


MS PhotoSynth

(2008-2017 discontinued product)

SfM Applications

Microsoft Flight Simulator / Google Earth



[https://www.youtube.com/watch?v=nd9l6EzuGF0]

Structure-from-Motion Software

Free/Opensource Software

- Bundler https://www.cs.cornell.edu/~snavely/bundler by Noah Snavely (Open Source, Freeware)
- VisualSFM http://ccwu.me/vsfm by Changchang Wu (Closed Source, Freeware)
- COLMAP https://colmap.github.io by Johannes Schönberger (Open Source, Freeware)
- AliceVision Meshroom https://alicevision.org/ (Open Source, Freeware)
- Theia http://www.theia-sfm.org by Chris Sweeney (Open Source, Freeware)
- OpenMVG https://github.com/openMVG/openMVG by Pierre Moulon (Open Source, Freeware)
- Regard3D http://www.regard3d.org (Open Source, Freeware)

Commercial Software

- Pix4Dmapper https://www.pix4d.com
- RealityCapture https://www.capturingreality.com
- Agisoft Metashape https://www.agisoft.com
- Photomodeler https://www.photomodeler.com
- Audodesk ReCap https://www.pix4d.com
- 3DF ZEPHYR https://www.3dflow.net
- Bentley ContextCapture
- Trimble Inpho

Disclaimer

Many of the slides used here are obtained from online resources (including many open lecture materials) without appropriate acknowledgement. They are used here for the sole purpose of classroom teaching. All the credit and all the copyrights belong to the original authors. You should not copy it, redistribute it, put it online, or use it for any other purposes than for this course.