No.3
ステレオ・エピポーラ幾何
Stereo and Epipolar Geometry

担当教員：向川康博・田中賢一郎

Slide credits

- Special Thanks: Some slides are adopted from other instructors' slides.
  - Tomokazu Sato, former NAIST CV1 class
  - James Tompkin, Brown CSCI 1430 Fall 2017
  - Ioannis Gkioulekas, CMU 16-385 Spring 2018

  - We also thank many other instructors for sharing their slides.

Today's topics

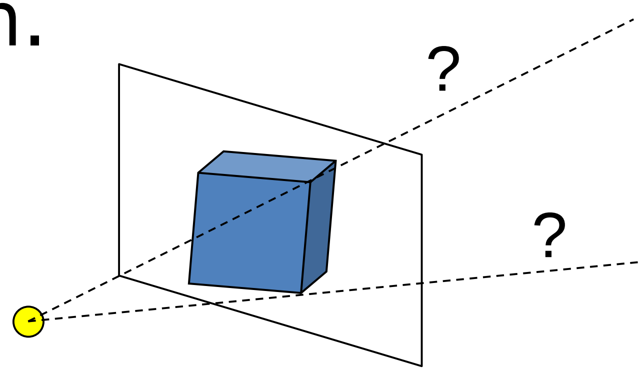
- 3D point from matched point pair
- Epipolar geometry
  - E and F matrices
  - 8-point algorithm
- Other techniques
  - Multi-view (Stereo, EPI, Space curving)
  - Active illumination (SL,ToF)

3D scene recovery

- Can you compute 3D scene from an image?

Geometry of stereo views

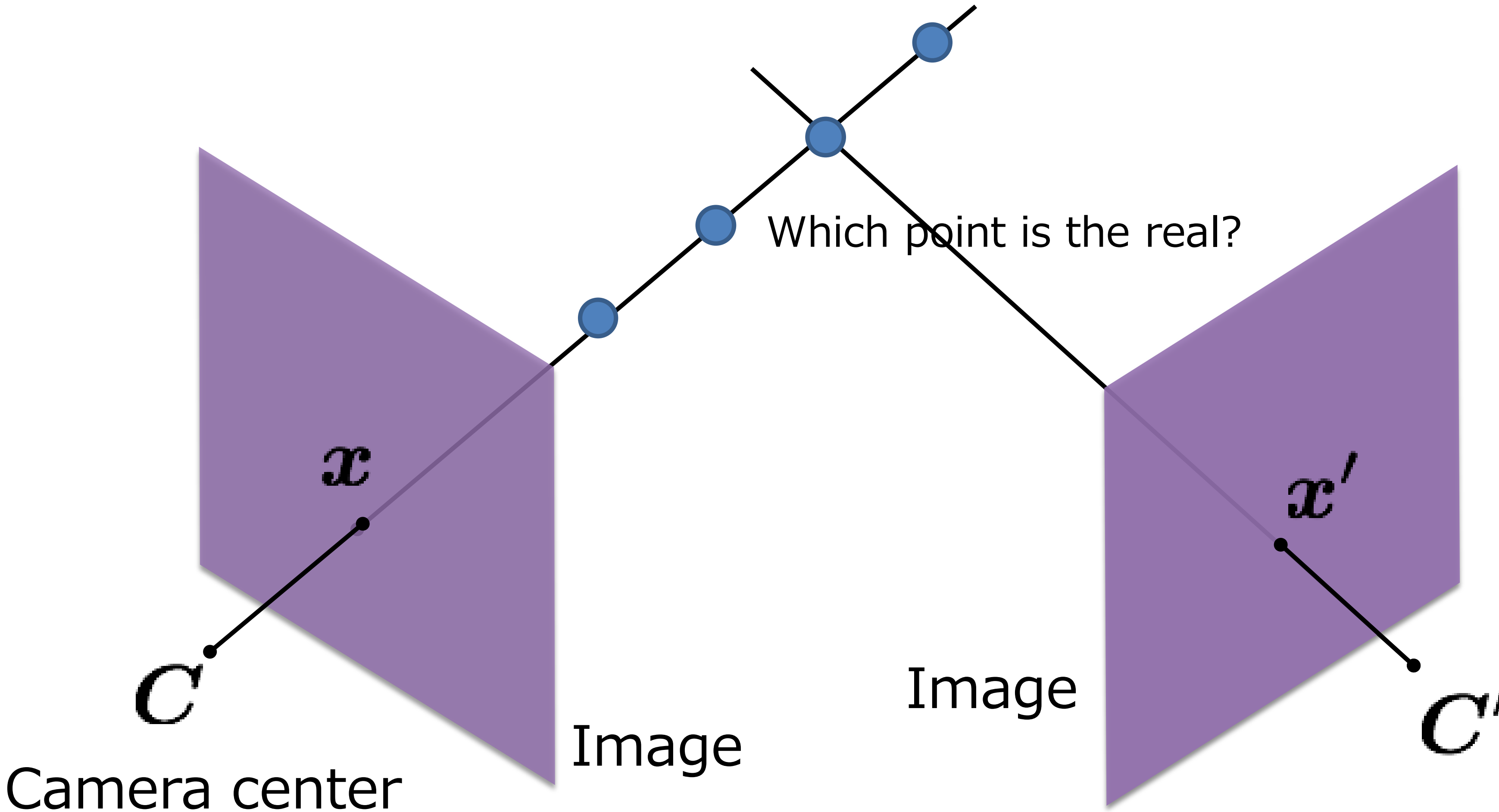
- A single image does not have direct depth information.



- With at least two images, depth can be measured through triangulation.
  - The most animals have at least two eyes.
  - A stereo vision system is equipped on an autonomous robot as eyes.

2-view geometry

- Image pixel corresponds to a ray in 3D space.
- Additional view gives the depth information.



Triangulation（三角測量）

	Structure (scene geometry)	Motion (camera geometry)	Measurements
Pose Estimation	known	estimate	3D to 2D correspondences
Triangulation	estimate	known	2D to 2D correspondences
Reconstruction	estimate	estimate	2D to 2D correspondences

Two views

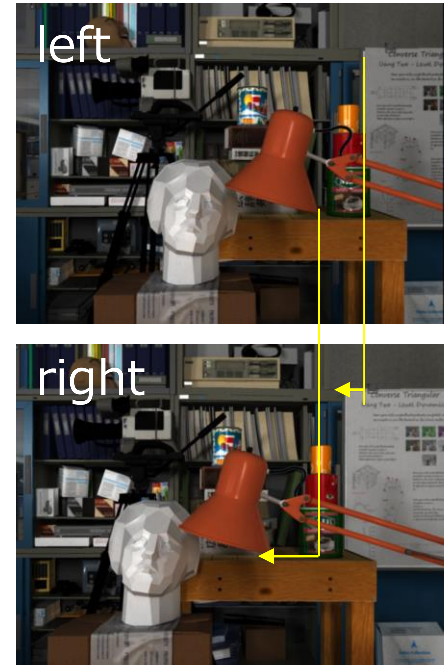
Tsukuba Stereo

Left

Right

mini-report1: What is the difference?

Toy example

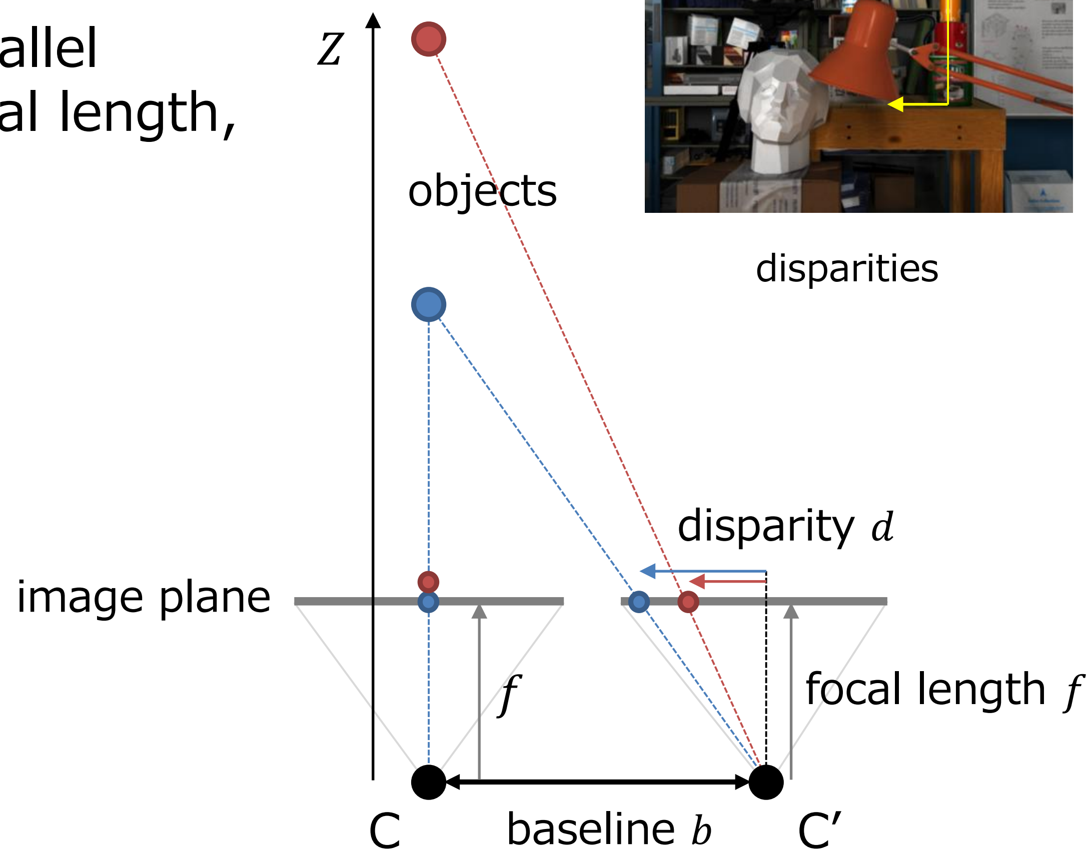


disparities

Disparity (視差):
near object moves larger

If two cameras are parallel
and have the same focal length,

$$Z = \frac{bf}{d}$$



(General) Triangulation Problem

Given a matched points

$$x, x'$$

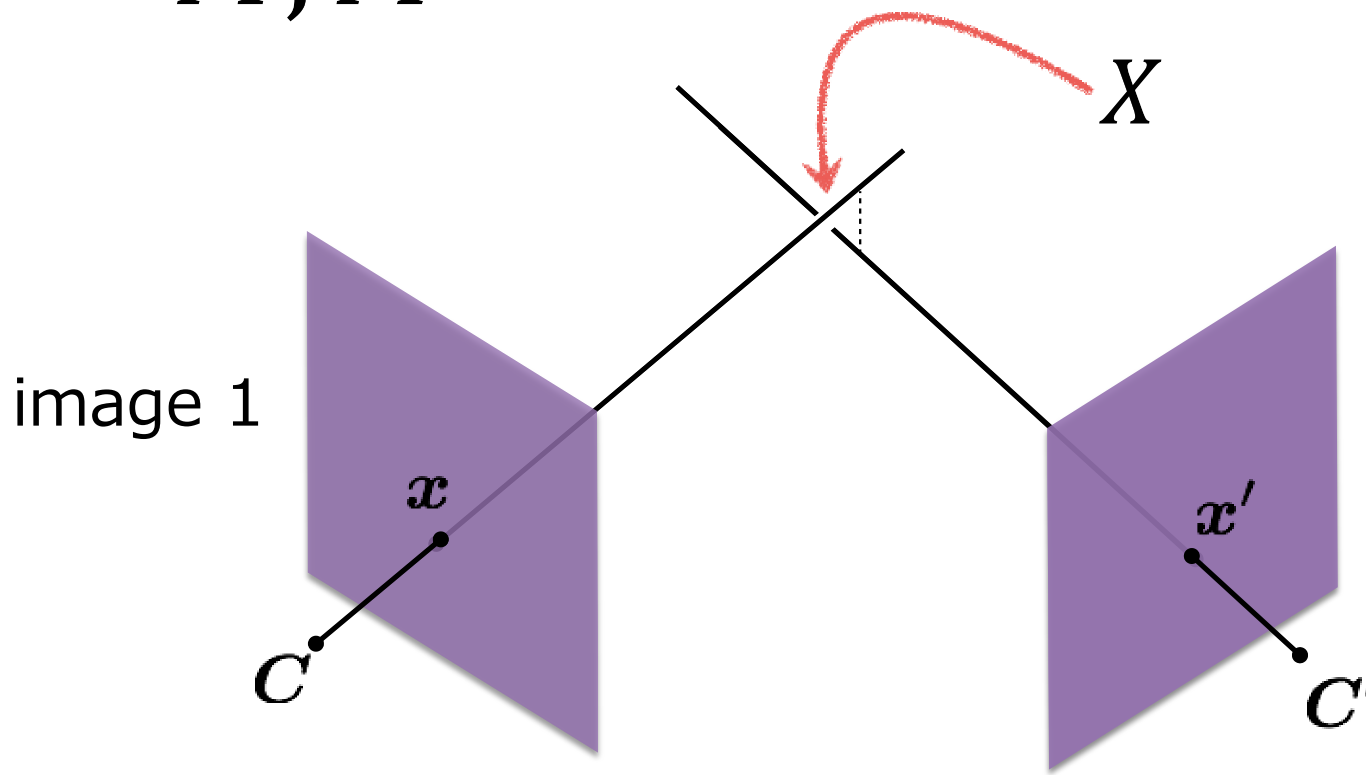
And camera projection matrices

$$M, M'$$

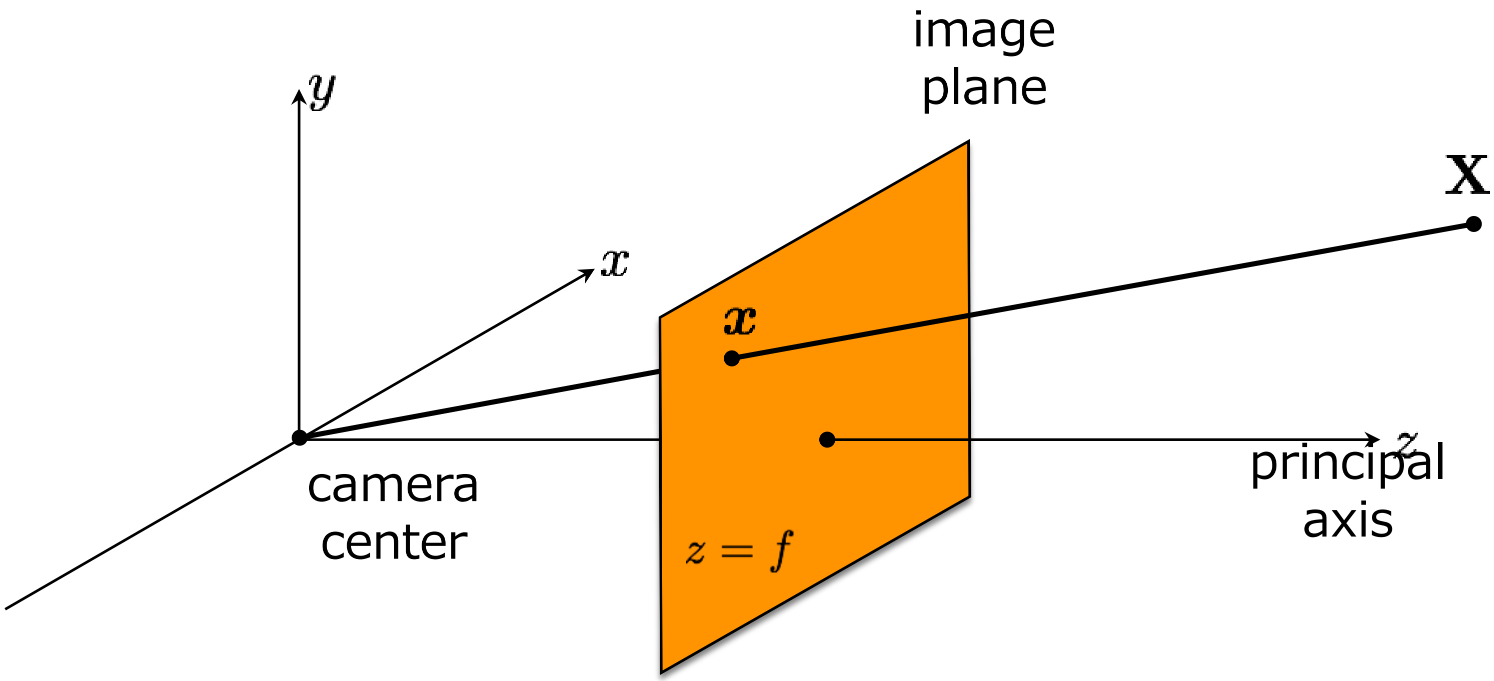
Estimate →

Best fit 3D point

$$X$$



Recall: camera projection



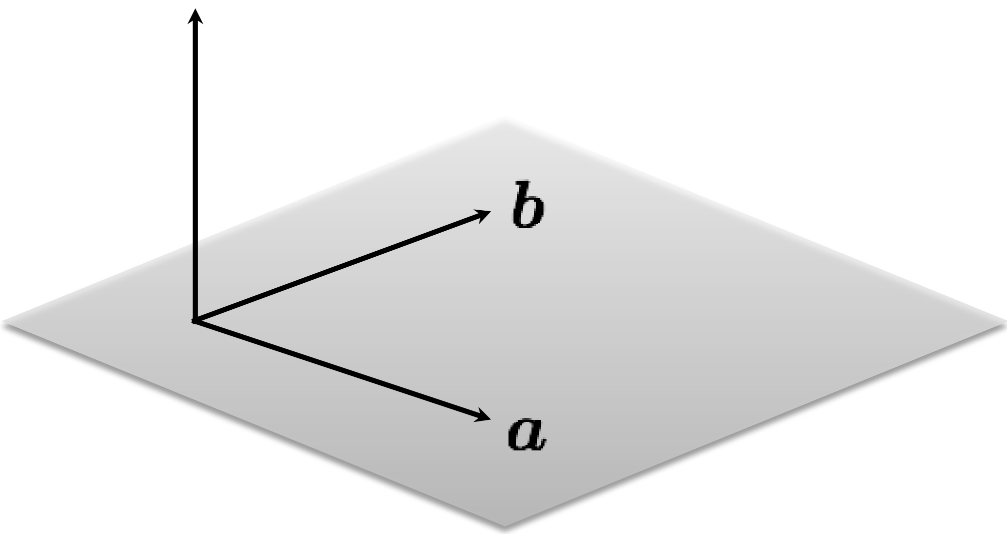
$$\boldsymbol{x} = \boldsymbol{M}\boldsymbol{X}$$

3D points to 2D image points

Recall: Cross Product

Vector (cross) product
takes two vectors and returns a vector perpendicular to both

$$\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}$$



$$\boldsymbol{c} \cdot \boldsymbol{a} = 0 \qquad \boldsymbol{c} \cdot \boldsymbol{b} = 0$$

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}$$

cross product of two vectors
in the same direction is zero

$$\boldsymbol{a} \times \boldsymbol{a} = 0$$

remember this!!!

Solve for **X**

- Start from camera projection

$$\underset{\text{known}}{\boldsymbol{x}} = \underset{\text{known}}{\boldsymbol{M}}\underset{\text{unknown}}{\boldsymbol{X}}$$

- Expand the right hand side

$$\boldsymbol{M}\boldsymbol{X} = \begin{bmatrix} - & \boldsymbol{m}_1^\top & - \\ - & \boldsymbol{m}_2^\top & - \\ - & \boldsymbol{m}_3^\top & - \end{bmatrix} \boldsymbol{X} = \begin{bmatrix} \boldsymbol{m}_1^\top \boldsymbol{X} \\ \boldsymbol{m}_2^\top \boldsymbol{X} \\ \boldsymbol{m}_3^\top \boldsymbol{X} \end{bmatrix}$$

- Same direction? → cross product is zero

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \times \begin{bmatrix} \boldsymbol{m}_1^\top \boldsymbol{X} \\ \boldsymbol{m}_2^\top \boldsymbol{X} \\ \boldsymbol{m}_3^\top \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} y\boldsymbol{m}_3^\top \boldsymbol{X} - \boldsymbol{m}_2^\top \boldsymbol{X} \\ \boldsymbol{m}_1^\top \boldsymbol{X} - x\boldsymbol{m}_3^\top \boldsymbol{X} \\ x\boldsymbol{m}_2^\top \boldsymbol{X} - y\boldsymbol{m}_1^\top \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

redundant

Number of unknown is 3 and equation is 2.

(Linear combination)

Solve for **X**

- Same for the other camera
  And superpose

came from
1st camera

came from
2nd camera

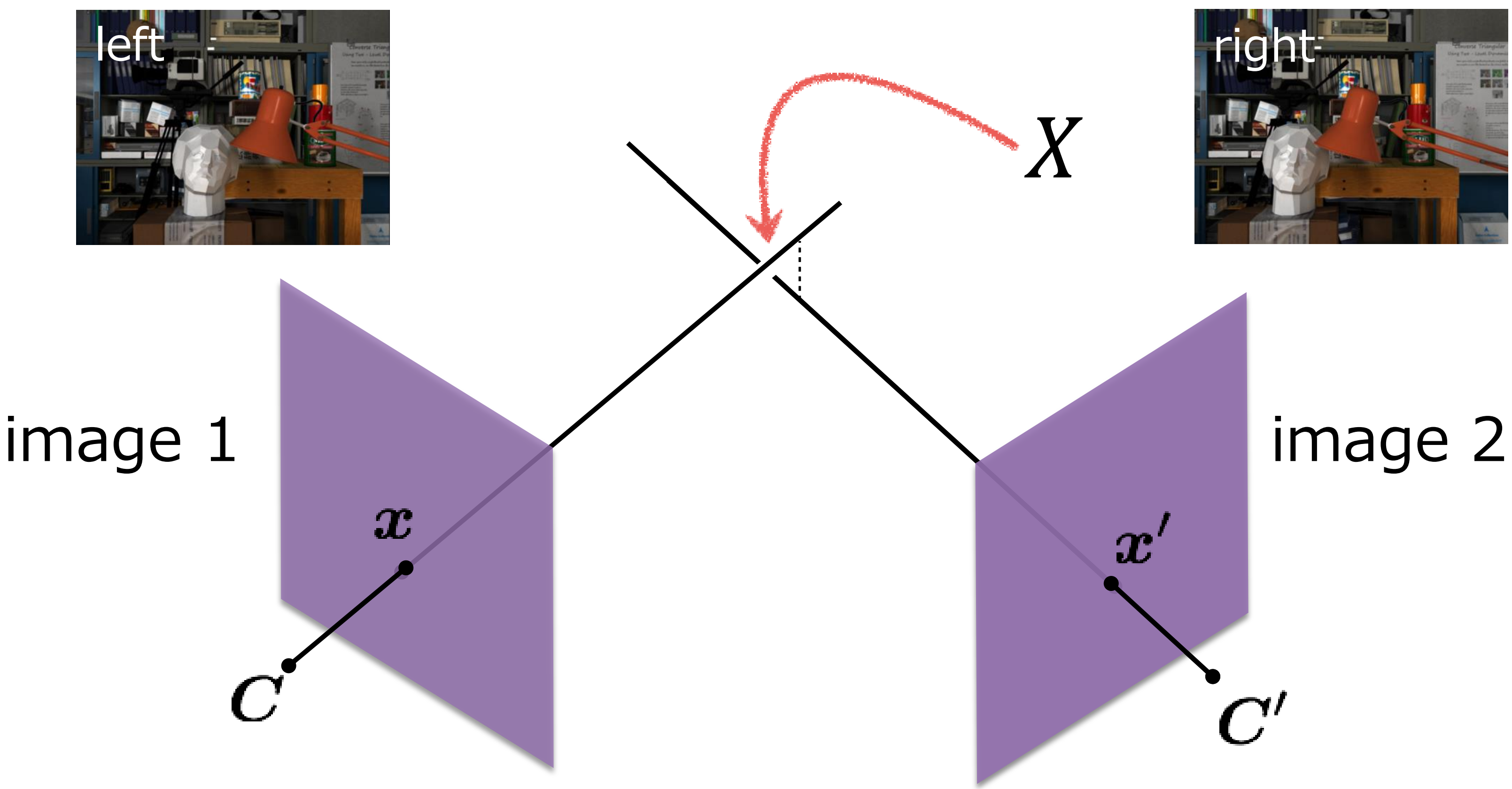
$$\begin{bmatrix} y\boldsymbol{m}_3^\top \boldsymbol{X} - \boldsymbol{m}_2^\top \boldsymbol{X} \\ \boldsymbol{m}_1^\top \boldsymbol{X} - x\boldsymbol{m}_3^\top \boldsymbol{X} \\ y'\boldsymbol{m'}_3^\top \boldsymbol{X} - \boldsymbol{m'}_2^\top \boldsymbol{X} \\ \boldsymbol{m'}_1^\top \boldsymbol{X} - x'\boldsymbol{m'}_3^\top \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Number of unknown is 3 and equation is 4.

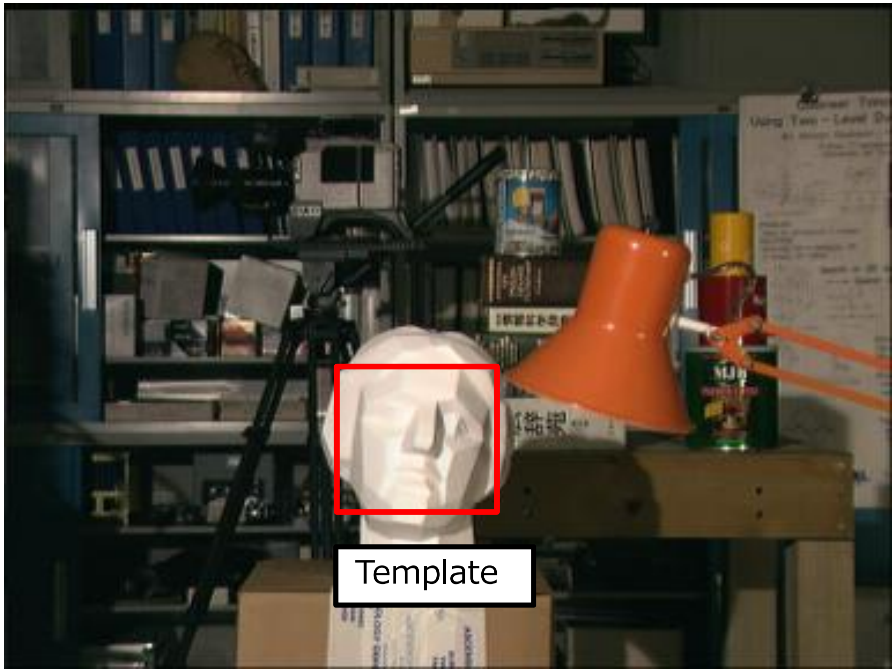
$$\boldsymbol{AX} = \boldsymbol{0}$$

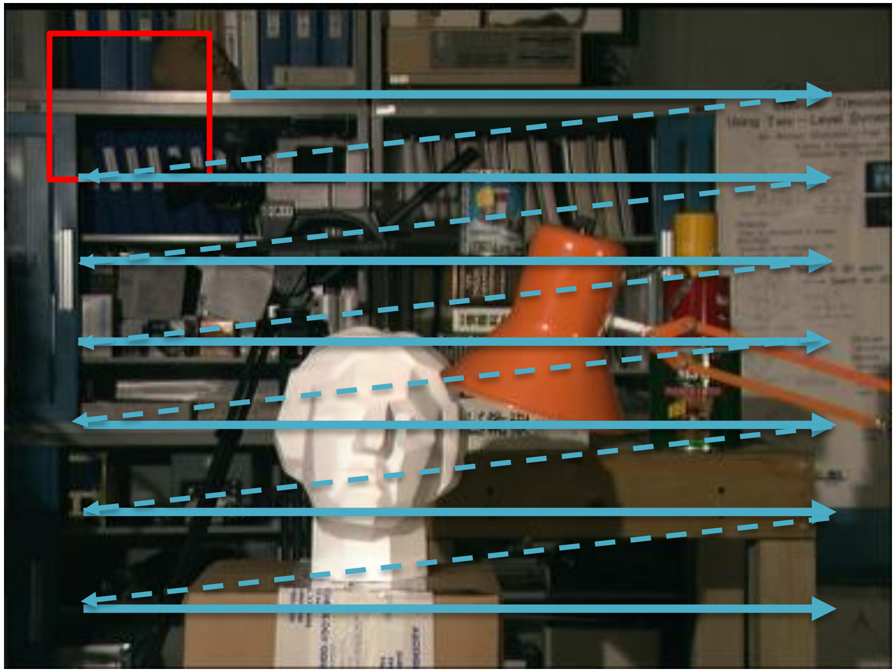
- How to solve? → SVD!

How to obtain the matched pair?



Naïve method: Template matching





For each point of the left image, find the best similar point from the right hand image.

How many evaluation is required?

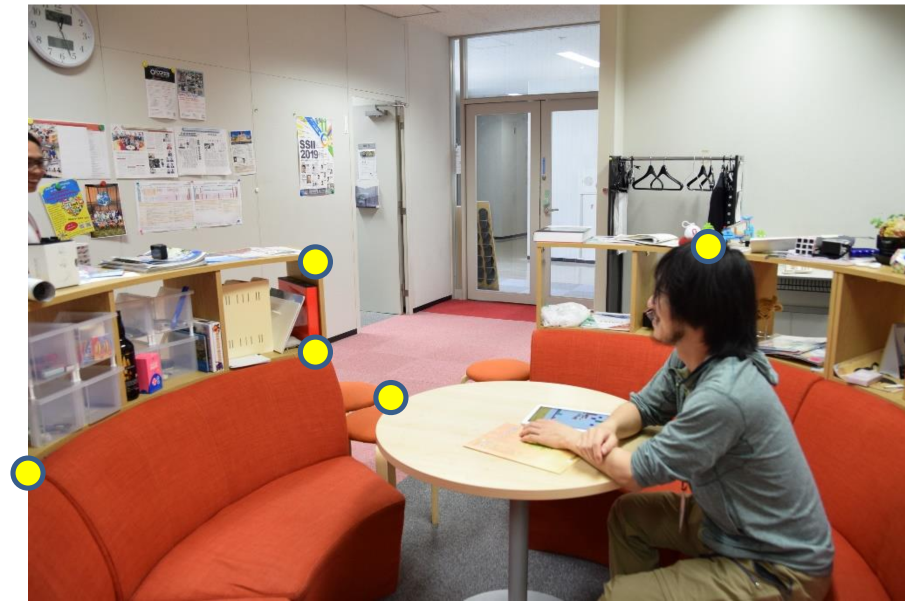
Width * height * width * height * [cost of template comparison]
Example) 1000 * 1000 * 1000 * 1000 = 10^{12} times template comparison.

Today's topics

- 3D point from matched point pair
- **Epipolar geometry**
  - **E and F matrices**
  - **8-point algorithm**
- Other techniques
  - Multi-view (Stereo, EPI, Space curving)
  - Active illumination (SL,ToF)

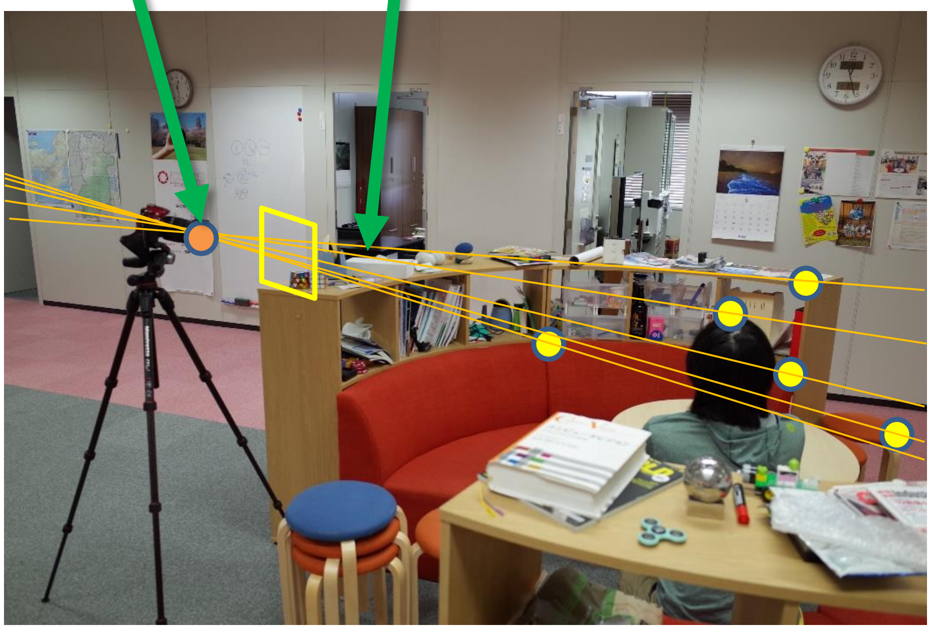
Another camera

- One camera looks at this scene.
- How does another camera look?



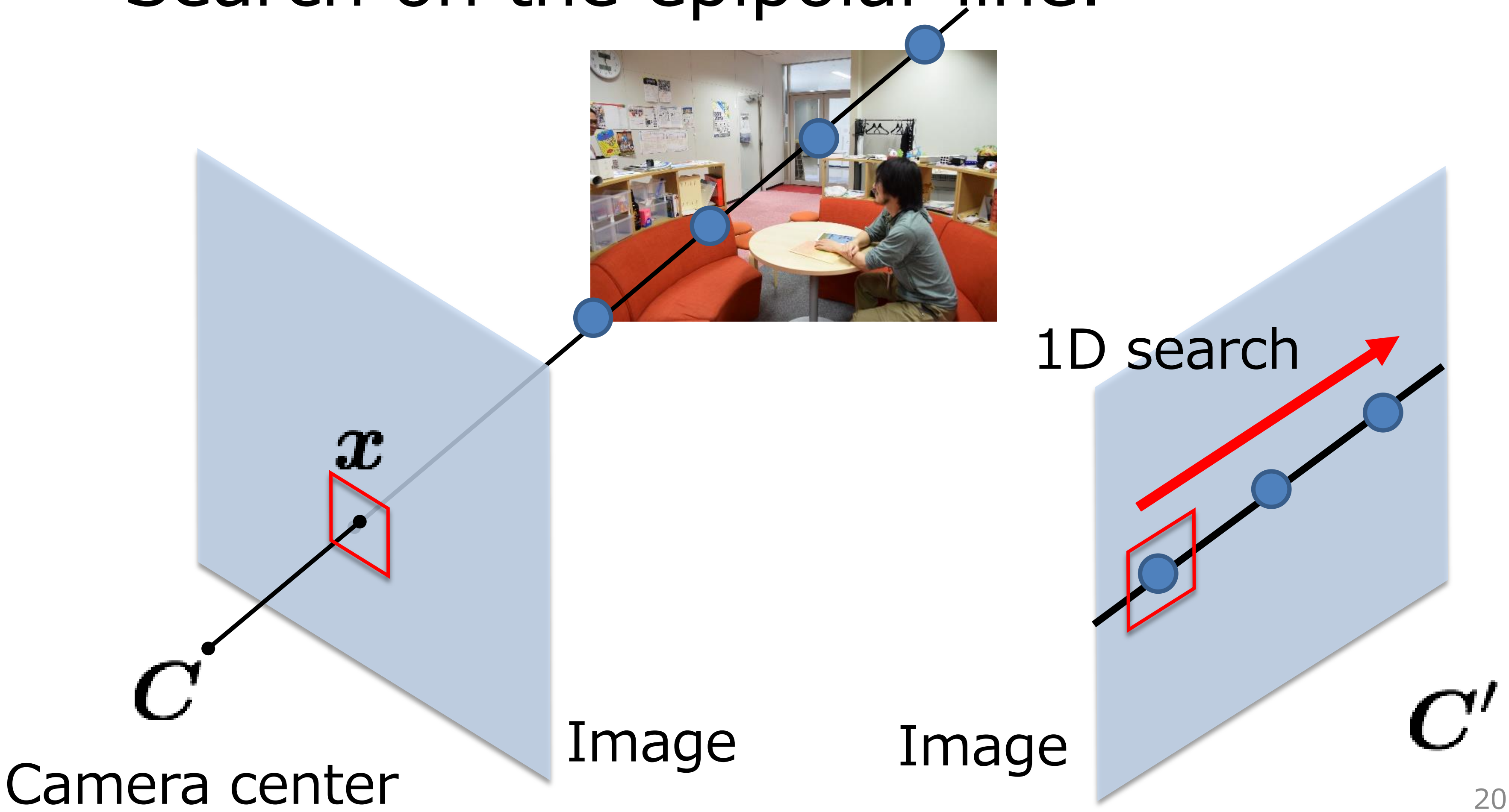
Another camera
Epipole

Line corresponds to point
Epipolar line

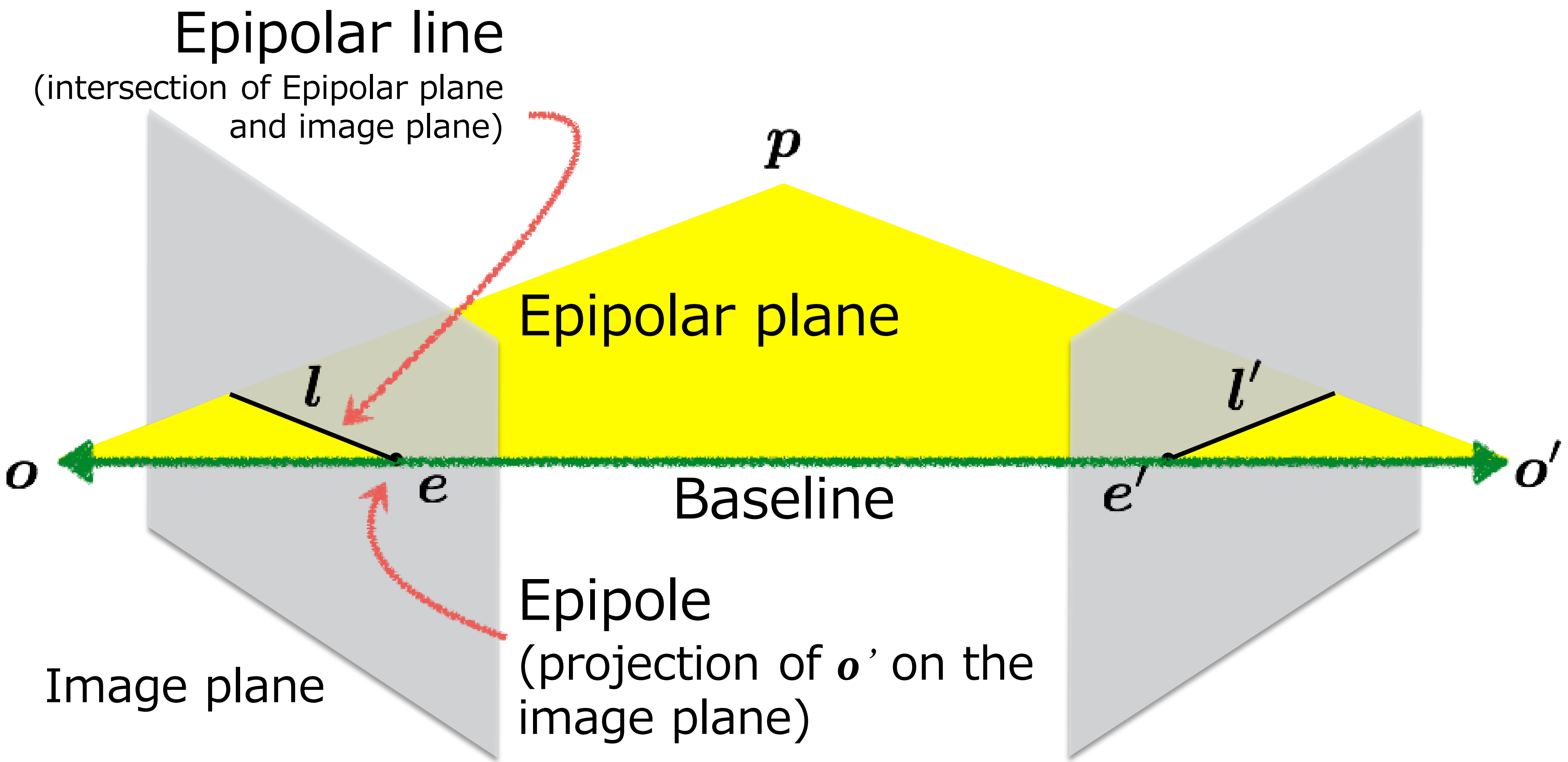


Epipolar geometry

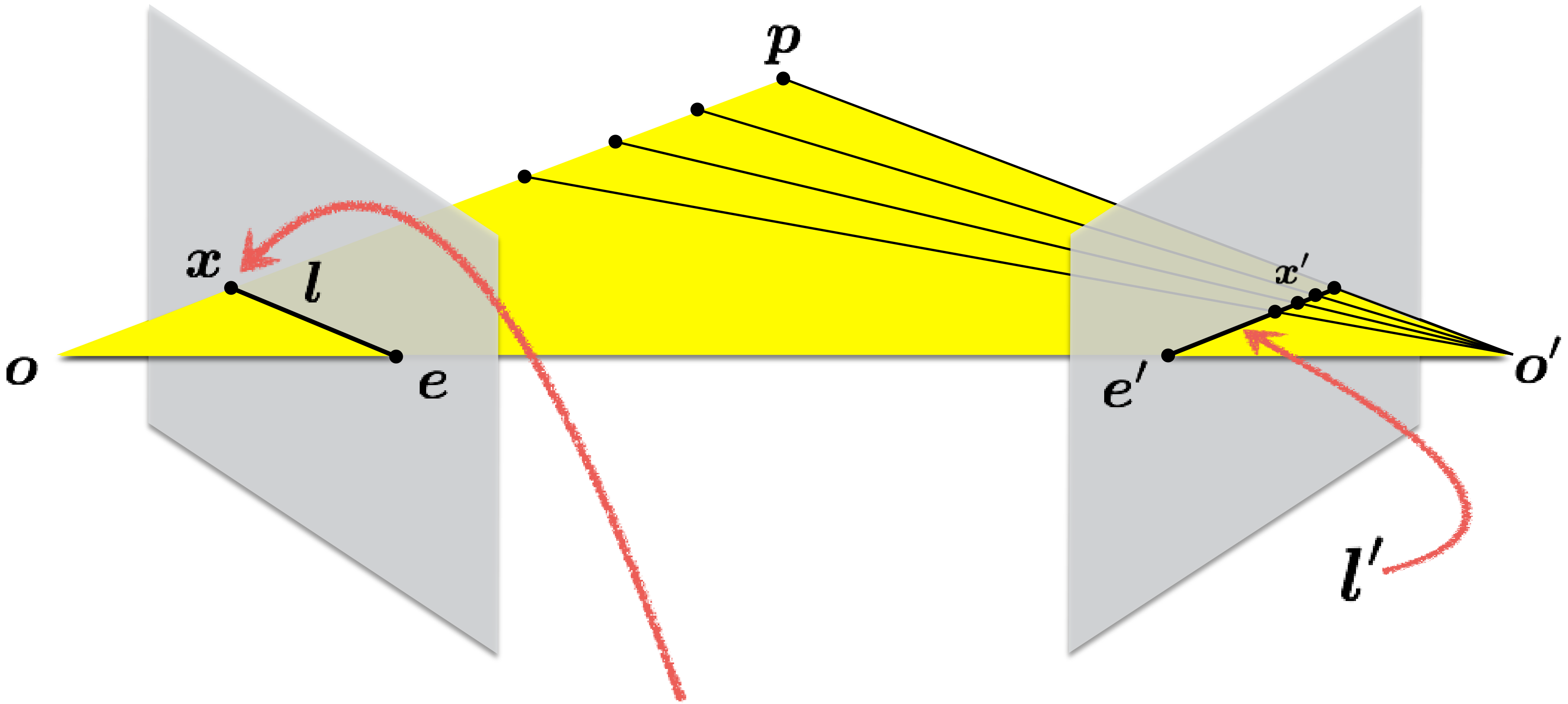
- Where is the corresponding point?
- Search on the epipolar line.



Epipolar geometry



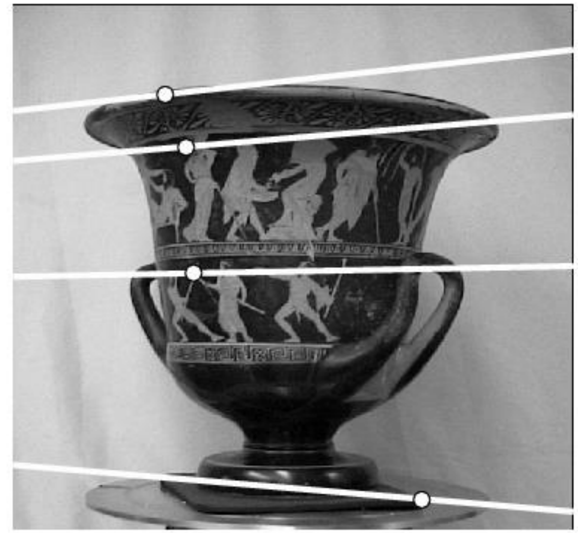
Epipolar constraint

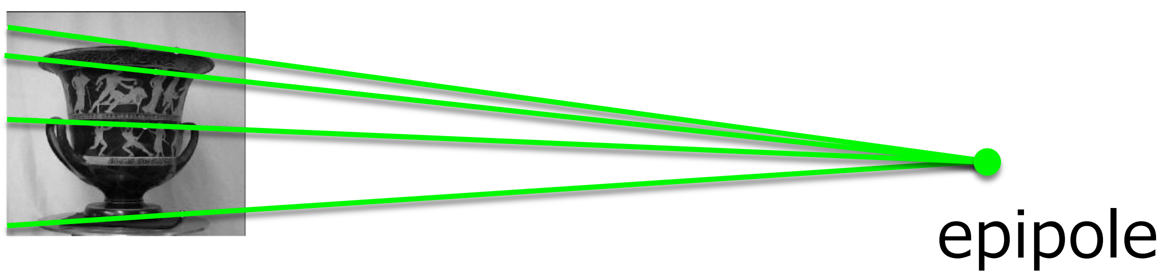


Potential matches for x lie on the epipolar line

Examples

Converging (向かい合った) camera

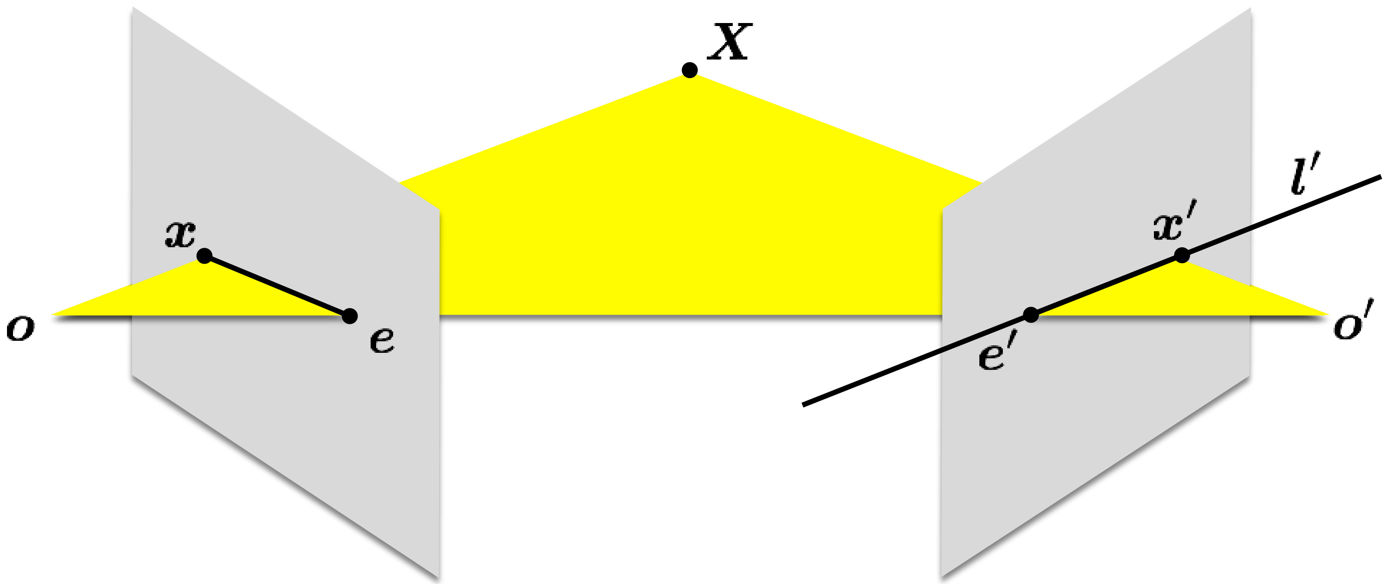




Calibrated case

- The intrinsic parameters of each camera are known.
- **Essential matrix** (基本行列): $\mathbf{E}$ (3x3 matrix)
  - Epipolar line $\mathbf{E}\boldsymbol{x} = \boldsymbol{l}'$
  - Relationship between two corresponding points

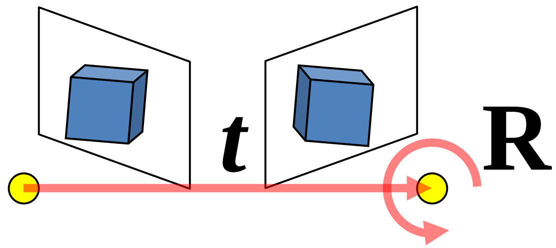
$$\boldsymbol{x}'^{\top}\mathbf{E}\boldsymbol{x} = 0$$



Parameters in essential matrix $\mathbf{E}$

- Parameterized by
  - the three DOF of the rotation matrix
  - the two DOF of the direction of the translation vector (defined up to scale)
- Include only **five** independent parameters.
- The $\mathbf{E}$ can be decomposed into $\mathbf{R}$ and $\boldsymbol{t}$.

$$\mathbf{E} = \mathbf{R}[\boldsymbol{t}_\times]$$



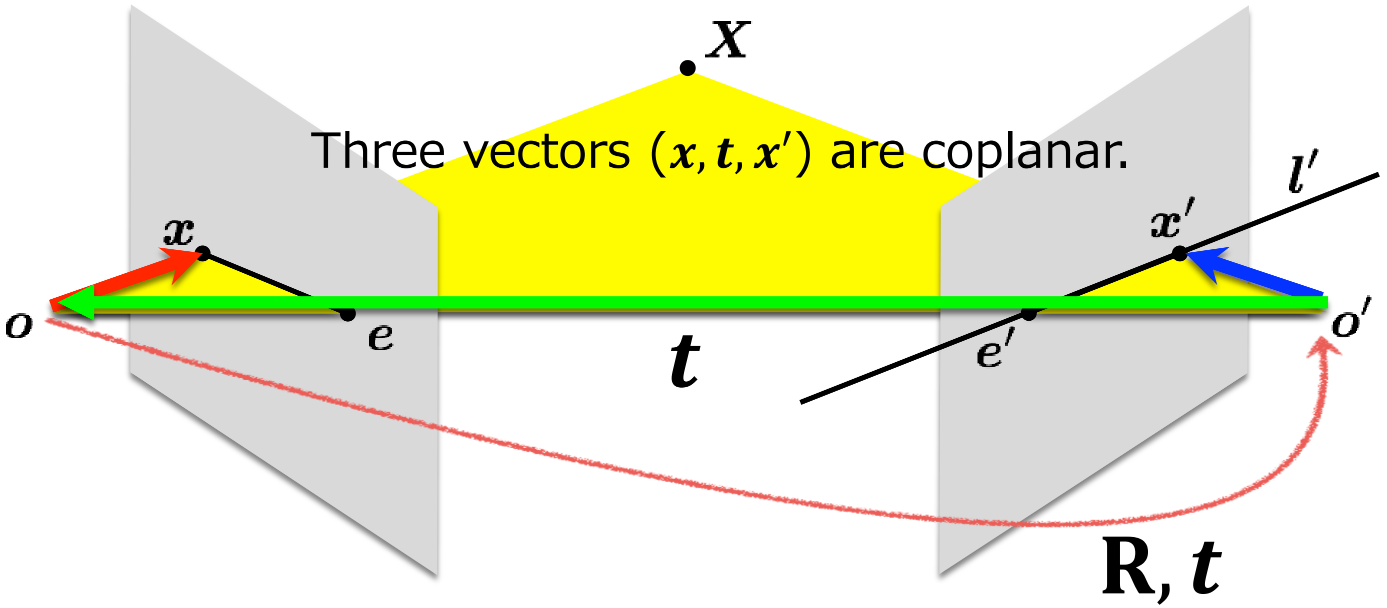
Memo

Cross product can be expressed by inner product using skew-symmetric matrix (歪対称行列) such that

$$\boldsymbol{x} \times \boldsymbol{y} = [\boldsymbol{x}_\times]\boldsymbol{y}$$

Derivation of $\mathbf{E}$

- coplanarity $\quad (\boldsymbol{x} - \boldsymbol{t})^\top (\boldsymbol{t} \times \boldsymbol{x}) = 0$
- rigid motion $\quad \boldsymbol{x}' = \mathbf{R}(\boldsymbol{x} - \boldsymbol{t})$



$$\boldsymbol{x}' = \mathbf{R}(\boldsymbol{x} - \boldsymbol{t})$$

$$\boldsymbol{x}'^\top = (\boldsymbol{x} - \boldsymbol{t})^\top \mathbf{R}^\top$$

$$\boldsymbol{x}'^\top \mathbf{R} = (\boldsymbol{x} - \boldsymbol{t})^\top$$

$$\boldsymbol{x}'^\top \mathbf{R}(\boldsymbol{t} \times \boldsymbol{x}) = 0$$

$$\boldsymbol{x}'^\top \mathbf{R}([\boldsymbol{t}_\times]\boldsymbol{x}) = 0$$

$$\boldsymbol{x}'^\top (\mathbf{R}[\boldsymbol{t}_\times])\boldsymbol{x} = 0$$

$$\boldsymbol{x}'^\top \mathbf{E}\boldsymbol{x} = 0$$

properties of the $\mathbf{E}$ matrix

Longuet-Higgins equation	$\boldsymbol{x}'^{\top}\mathbf{E}\boldsymbol{x} = 0$	
Epipolar lines	$\boldsymbol{x}^{\top}\boldsymbol{l} = 0$ $\boldsymbol{l}' = \mathbf{E}\boldsymbol{x}$	$\boldsymbol{x}'^{\top}\boldsymbol{l}' = 0$ $\boldsymbol{l} = \mathbf{E}^{T}\boldsymbol{x}'$
Epipoles	$\boldsymbol{e}'^{\top}\mathbf{E} = \mathbf{0}$	$\mathbf{E}\boldsymbol{e} = \mathbf{0}$

(points in normalized camera coordinates)

Uncalibrated case

- The intrinsic parameters $\boldsymbol{K}$ of each camera are **unknown**.
- The physical coordinates $\boldsymbol{x}$ are unknown. Only the image coordinates $\boldsymbol{p}$ are known.

$$\boldsymbol{p} = \boldsymbol{K}\boldsymbol{x}$$

- **Fundamental matrix**(基礎行列): $\boldsymbol{F}$ (3 x 3 matrix)

$$\boldsymbol{p}'^{\top}\boldsymbol{F}\boldsymbol{p} = 0$$

The **Fundamental matrix** is a **generalization** of the **Essential matrix**, where the assumption of **calibrated cameras** is removed.

Fundamental Matrix

Recall: Intrinsic parameters $\mathbf{K}$

$$\boldsymbol{p} = \mathbf{K}\boldsymbol{x} \qquad \boldsymbol{x} = \mathbf{K}^{-1}\boldsymbol{p}$$

Recall: Essential matrix $\mathbf{E}$

$$\boldsymbol{x}'^{\top}\mathbf{E}\boldsymbol{x} = 0$$

Substitute

$$\boldsymbol{p}'^{\top}\underbrace{\mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}}_{=\mathbf{F}}\boldsymbol{p} = 0$$

Properties

- Potential matches for $\boldsymbol{p}$ lie on the epipolar line $\mathbf{F}\boldsymbol{p}$
- $\mathbf{F}$ is defined up to scale, and rank($\mathbf{F}$) = 2.

Weak calibration

- The simple problem of estimating the epipolar geometry from point correspondences between two images with **unknown intrinsic parameters**.
- Corresponding pixel pairs

$$\{p'_m = ({x'}_m, y'_m),\, p_m = (x_m, y_m)\}$$

- Unknown 3x3 fundamental matrix

$$\boldsymbol{F} = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}$$

One equation from one correspondence

- One correspondence

$$\{p'_m = (x'_m, y'_m), p_m = (x_m, y_m)\}$$

- Epipolar constraint

$$p'^{\top}_m \boldsymbol{F} p_m = 0$$

$$\begin{bmatrix} x'_m & y'_m & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \\ 1 \end{bmatrix} = 0$$

- One equation

$$x_m x'_m f_1 + x_m y'_m f_2 + x_m f_3 +$$
$$y_m x'_m f_4 + y_m y'_m f_5 + y_m f_6 +$$
$$x'_m f_7 + y'_m f_8 + f_9 = 0$$

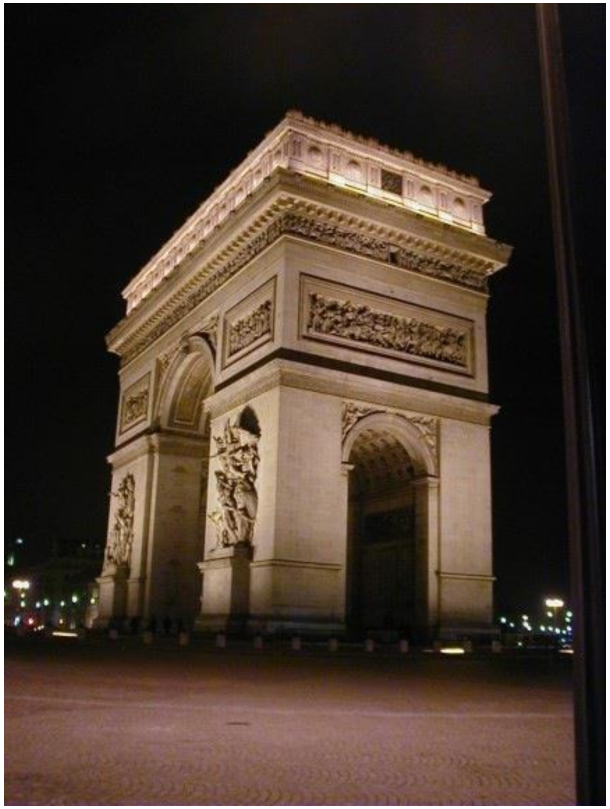
Eight-point algorithm

- $\boldsymbol{F}$ is 3x3 matrix (9 parameters)
- Since $\boldsymbol{F}$ is defined up to scale, we can set f_9=1.
- 8 unknowns
- We need at least **8 points**

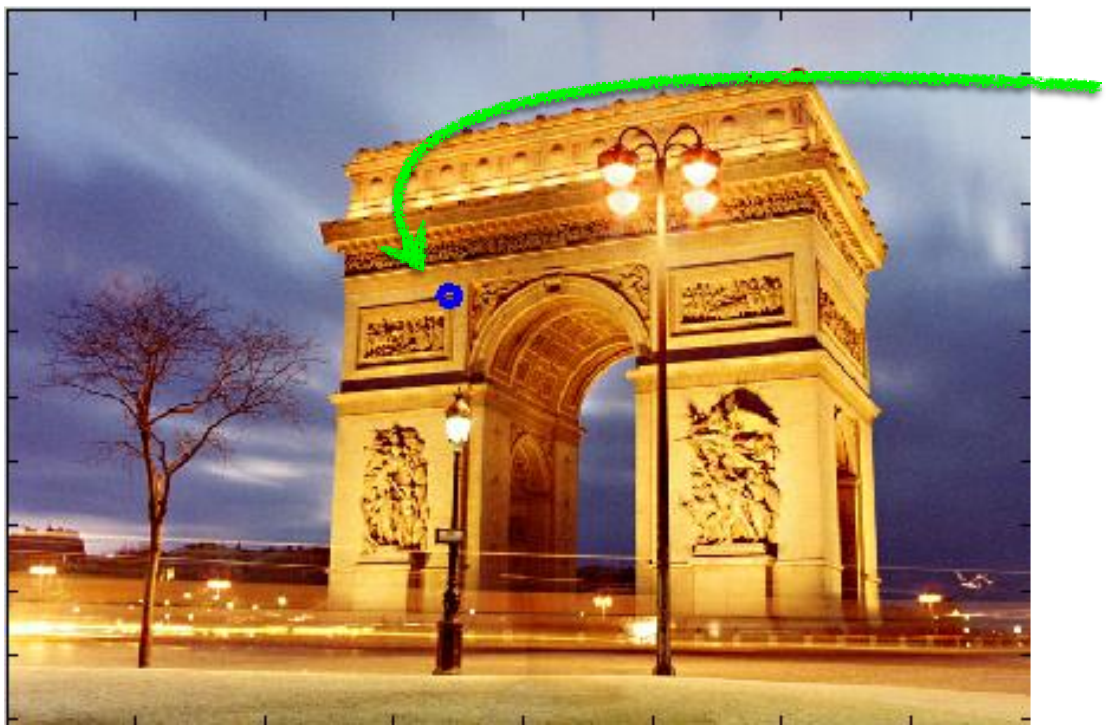
$$\begin{bmatrix} x_1 x'_1 & x_1 y'_1 & x_1 & y_1 x'_1 & y_1 y'_1 & y_1 & x'_1 & y'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_M x'_M & x_M y'_M & x_M & y_M x'_M & y_M y'_M & y_M & x'_M & y'_M & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \end{bmatrix} = \mathbf{0}$$

- How to solve $\boldsymbol{Ax} = 0$? → SVD!

Example



$$\mathbf{F} = \begin{bmatrix} -0.00310695 & -0.0025646 & 2.96584 \\ -0.028094 & -0.00771621 & 56.3813 \\ 13.1905 & -29.2007 & -9999.79 \end{bmatrix}$$



$$\boldsymbol{x} = \begin{bmatrix} 343.53 \\ 221.70 \\ 1.0 \end{bmatrix}$$

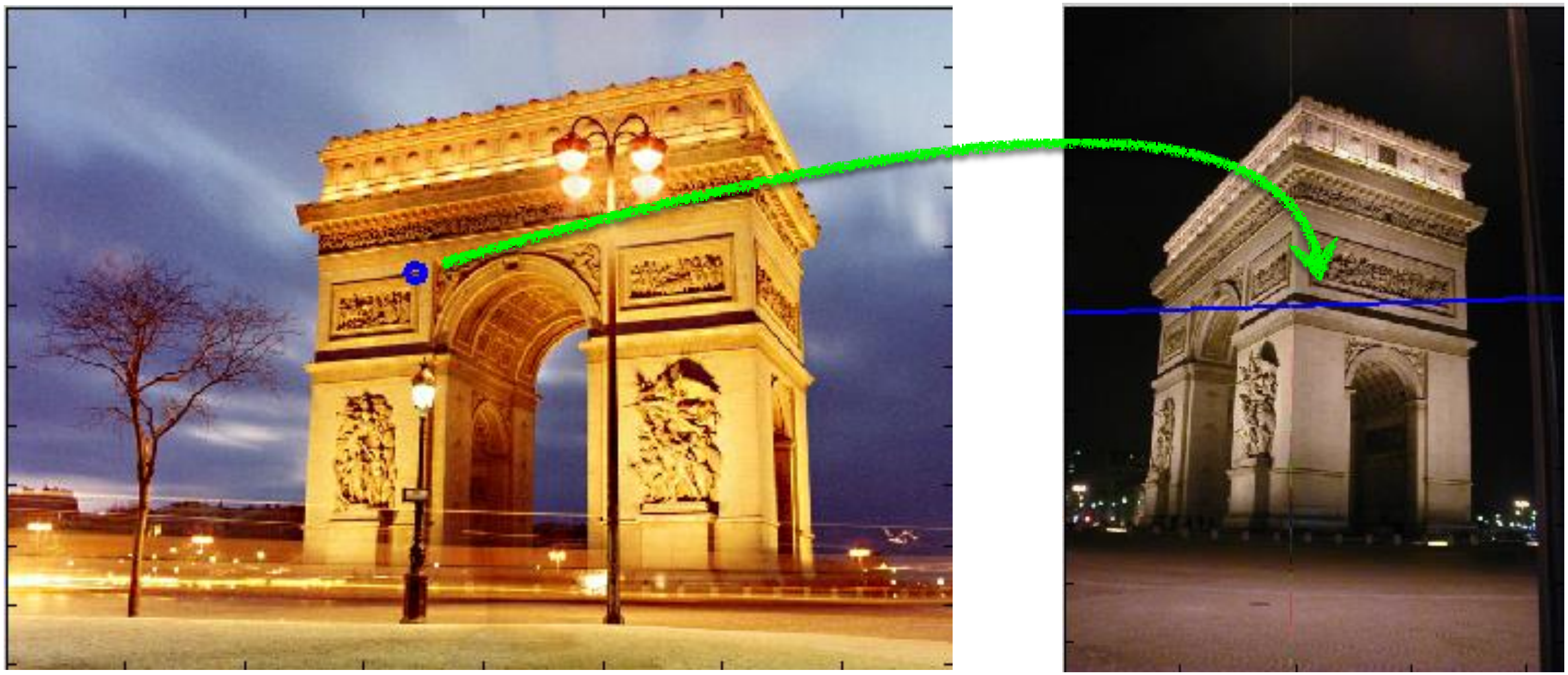
$$\boldsymbol{l}' = \mathbf{F}\boldsymbol{x}$$

$$= \begin{bmatrix} 0.0295 \\ 0.9996 \\ -265.1531 \end{bmatrix}$$

$$\boldsymbol{l}' = \mathbf{F}\boldsymbol{x}$$

$$= \begin{bmatrix} 0.0295 \\ 0.9996 \\ -265.1531 \end{bmatrix}$$

line: $0.0295x + 0.9996y - 265.1531 = 0$



Where is the epipole?

How would you compute it?

$$\mathbf{F}e = \mathbf{0}$$

The epipole is in the right null space of **F**

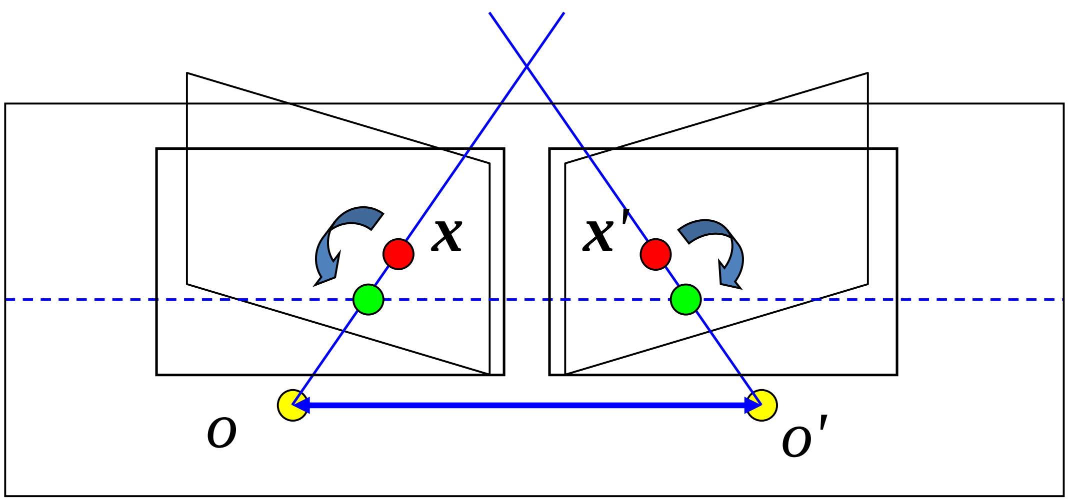
S V D !

Today's topics

- What is 3D points recovery problem?
- 3D point from matched point pair
- Epipolar geometry
  - E and F matrices
  - 8-point algorithm
- **Other techniques**
  - **Multi-view (Stereo, EPI, Space curving)**
  - **Active illumination (SL, ToF)**

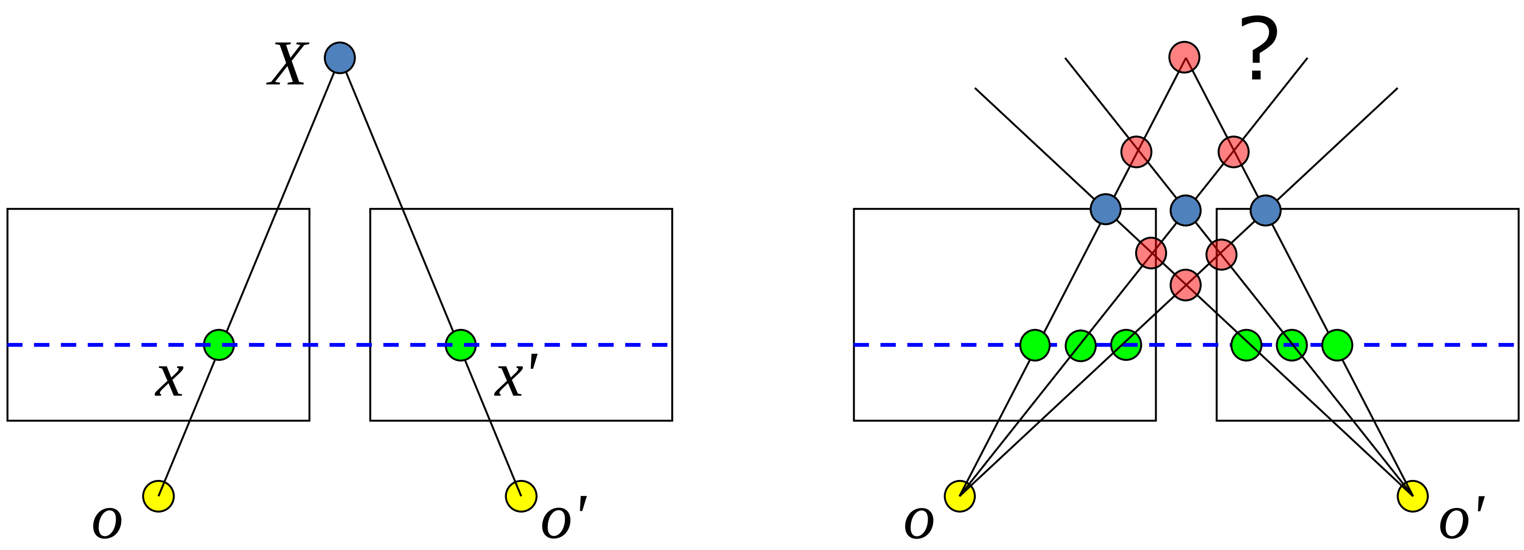
Image rectification

- The two image planes are reprojected onto a common plane parallel to the base line.
- The rectified epipolar lines are scanlines of the new images, and they are also parallel to the base line.



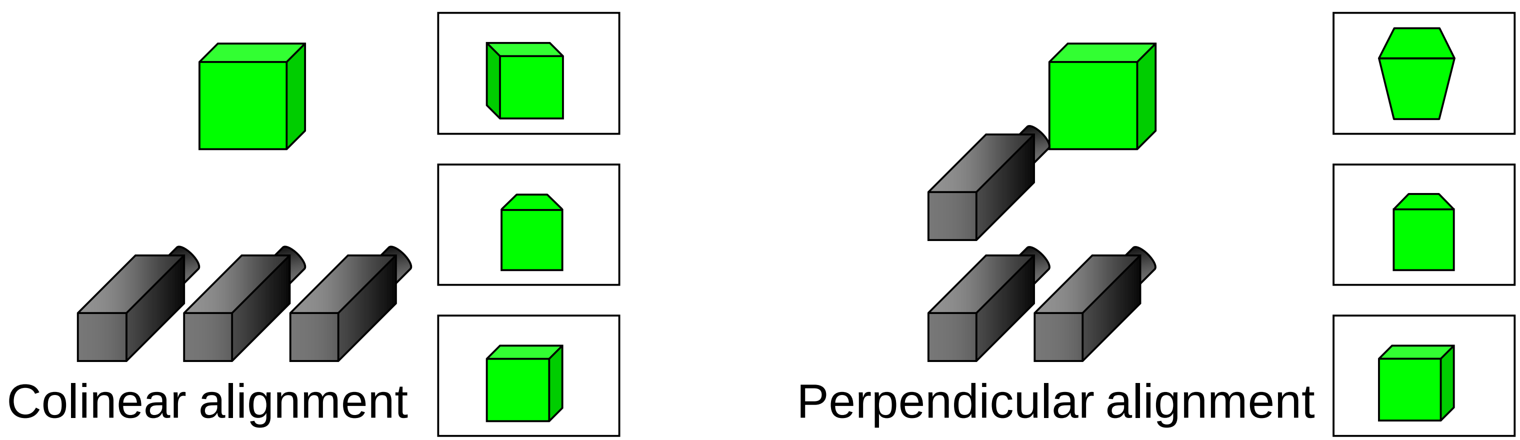
Ambiguity (曖昧性)

- When a single image feature is observed, there is no ambiguity.
- In the more usual case, wrong correspondences yield incorrect reconstructions.



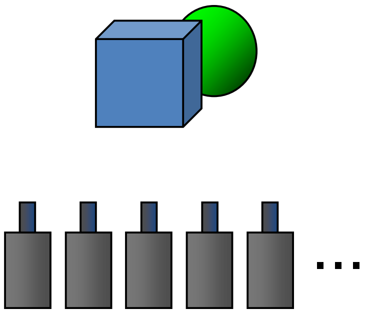
Three cameras (trinocular stereo)

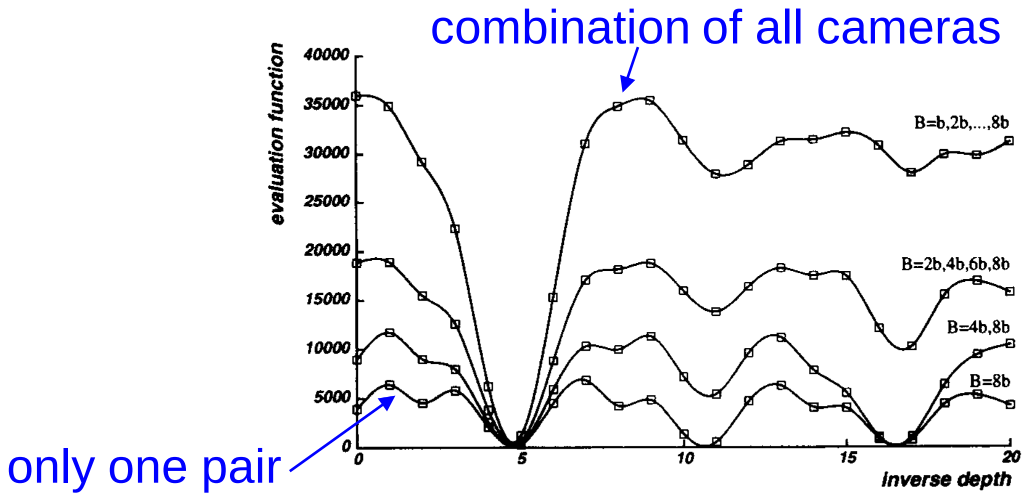
- Adding a third camera eliminates the ambiguity
- The third image can be used to check hypothetical matches between the first two pictures.
  - Colinear alignment: robust to occlusion
  - Perpendicular alignment: robust for all directional edges



Multiple cameras

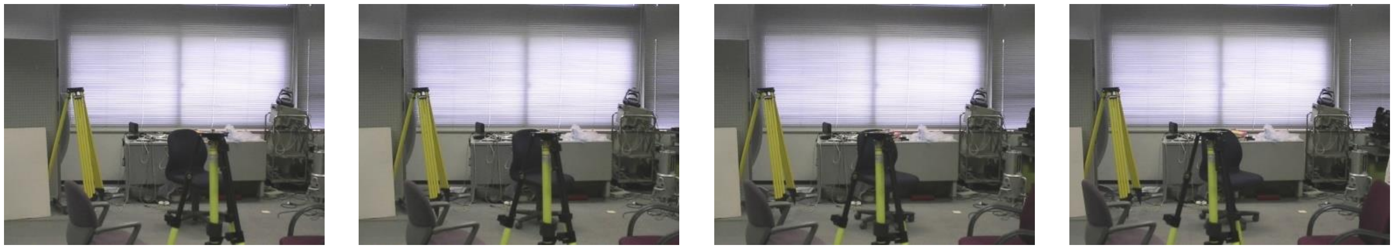
- Matches are found using all pictures
- Picking the first image as a reference, sums of squared differences associated with all other cameras are added into a global evaluation function.
- Robust to repetitive pattern



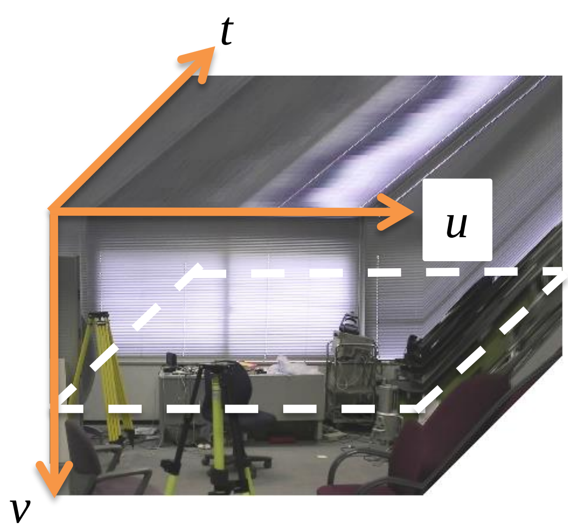


Combining multiple baseline stereo pairs

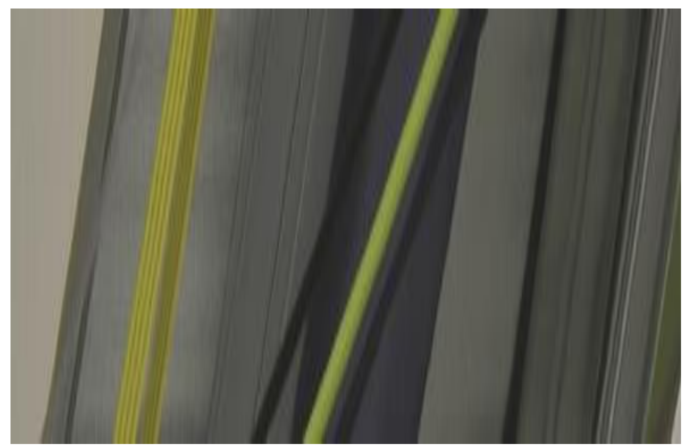
Epipolar plane image (EPI)



(a) Images taken under linear motion of camera with constant speed or by a light-field camera



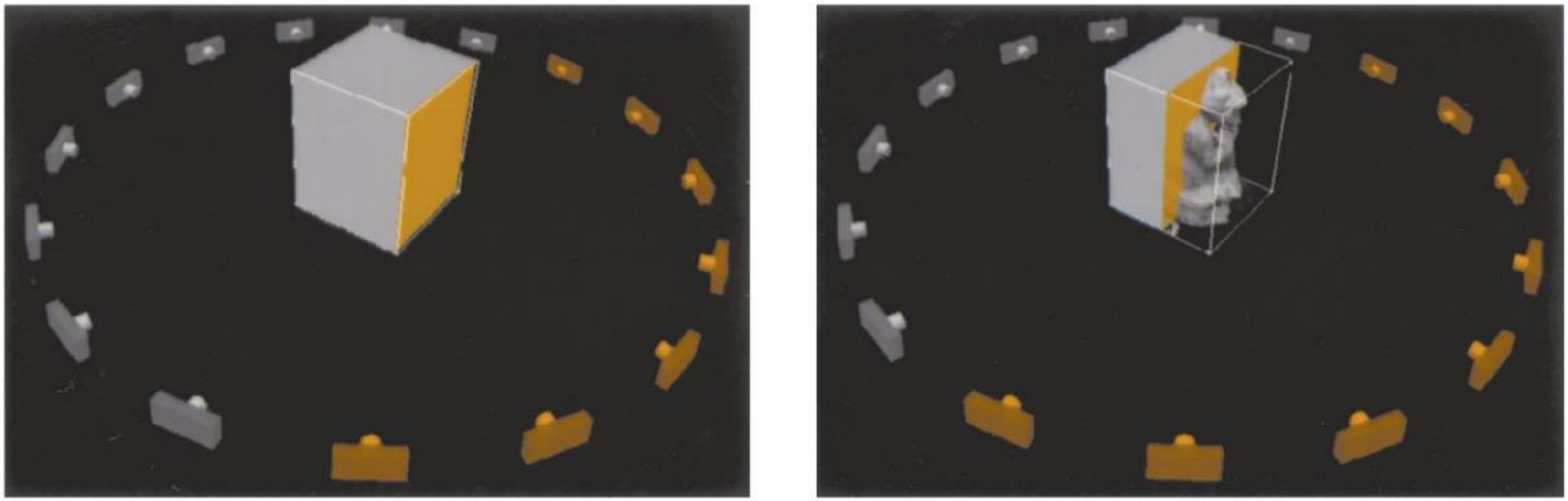
(b) Space time volume



(c) Epipolar plane image

Slope level of edge depends on the depth of the object.

Space carving (視体積交差法)



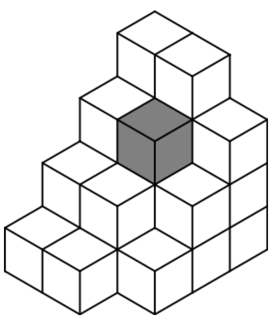


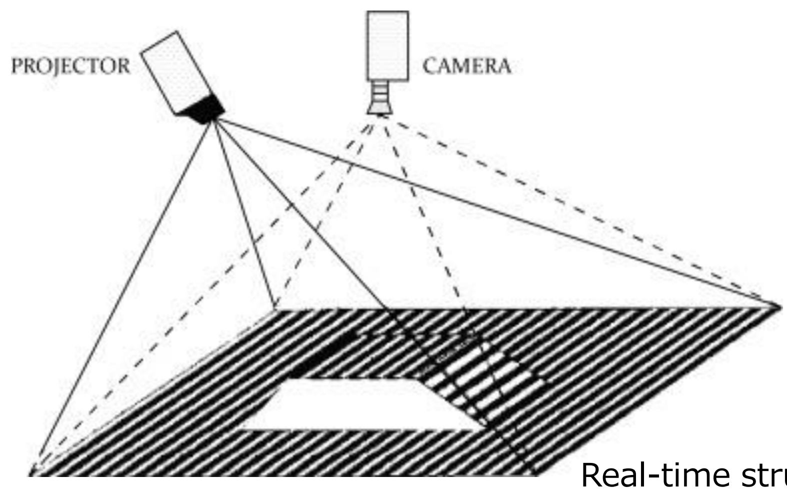
*image of voxel from wiki

1. Make voxel space and put initial voxels
2. Select one of surface voxel
3. Check the photo-consistency with considering occlusions
4. Remove voxel if photo-consistency is lower than threshol
5. Repeat 2 to 5 until any voxel is not removed

*Kiriakos N. Kutulakos , Steven M. Seitz, A Theory of Shape by Space Carving, ICCV99

Structured light

- Substitute a projector to a camera.
  – Easy to obtain the correspondence.

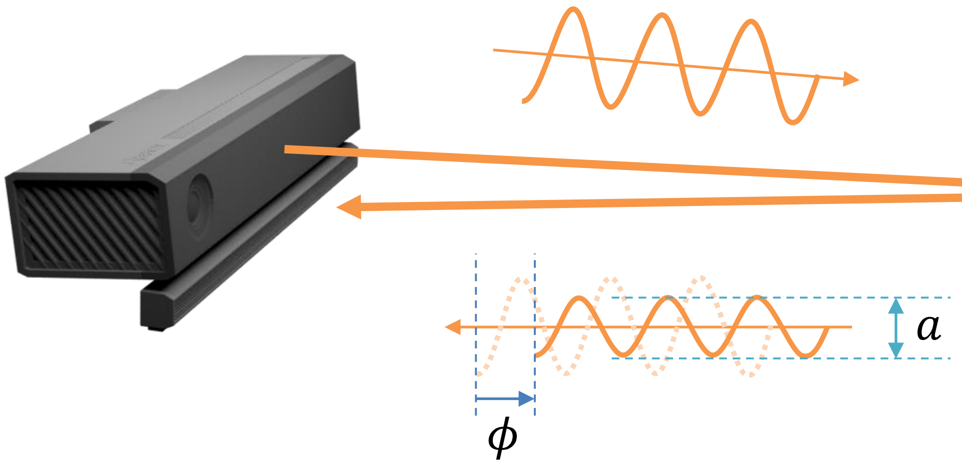


Real-time structured light profilometry: a review

- What pattern is used?
  – Gray code, XOR, phase shifting, micro PS, etc.

Time-of-Flight

- Measures the delay of returning light.



$$d = \frac{c\phi}{4\pi f}$$

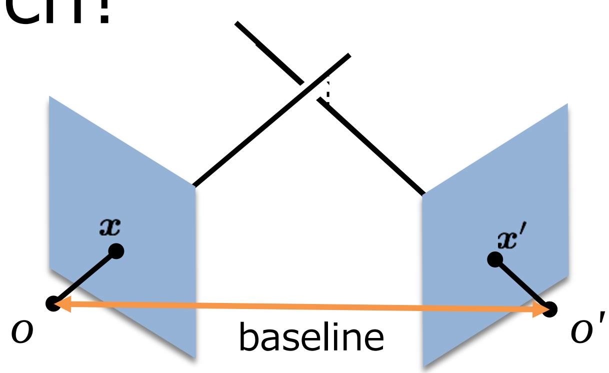
c: speed of light

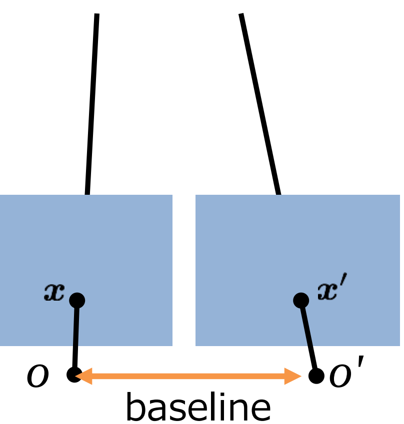
f: frequency of modulation

https://blogs.msdn.microsoft.com/uk_faculty_connection/2017/05/15/image-based-motion-analysis-with-kinect-v2-and-opencv/

Final mini report

- To obtain accurate 3d points by stereo, should the baseline be wide or narrow?
  - How fine depth can be described by 1 pixel disparity?
  - Are all pixels visible from both cameras? (occlusion problem: 隠れ問題)
  - How about the difficulty of correspondence search?

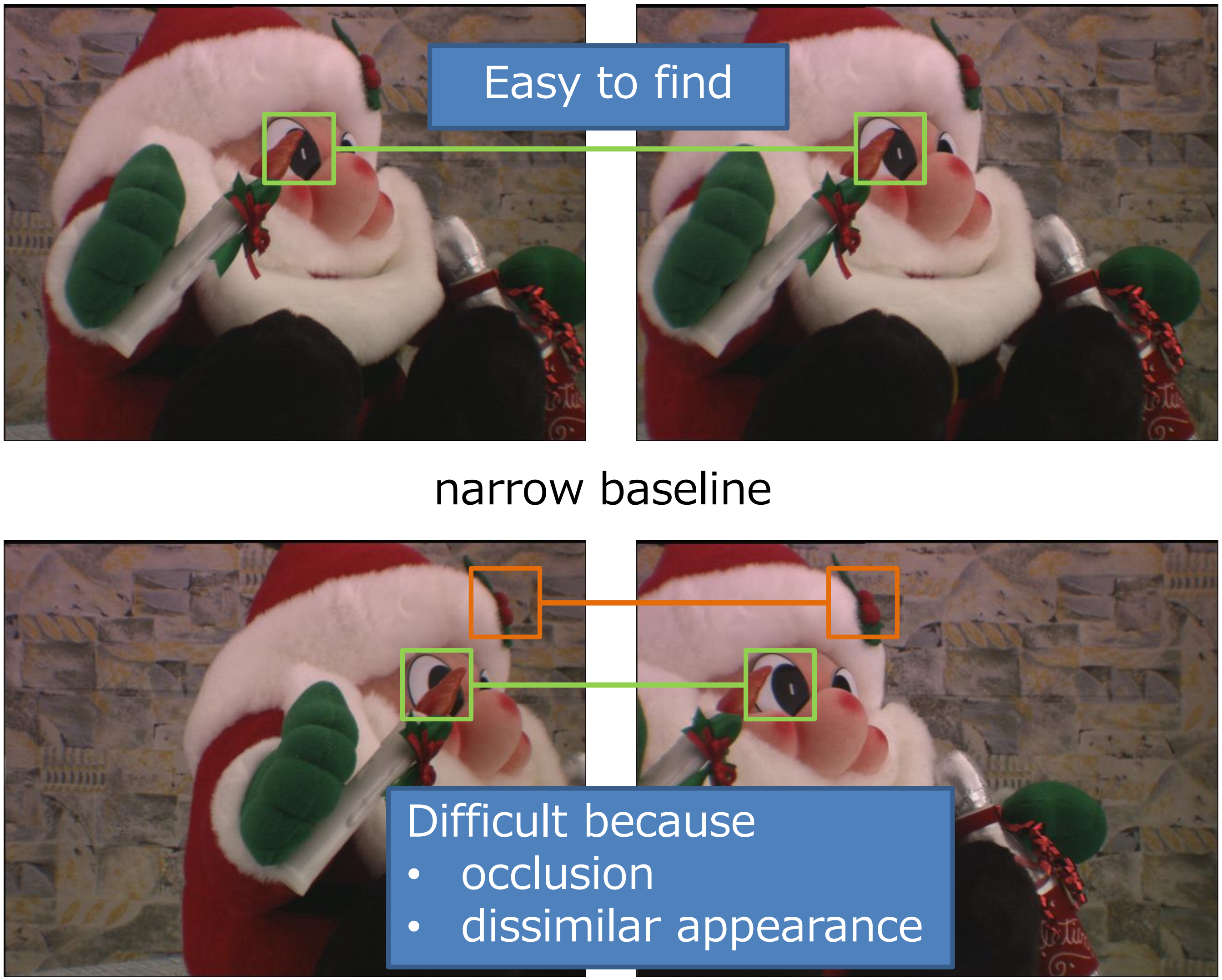




narrow baseline

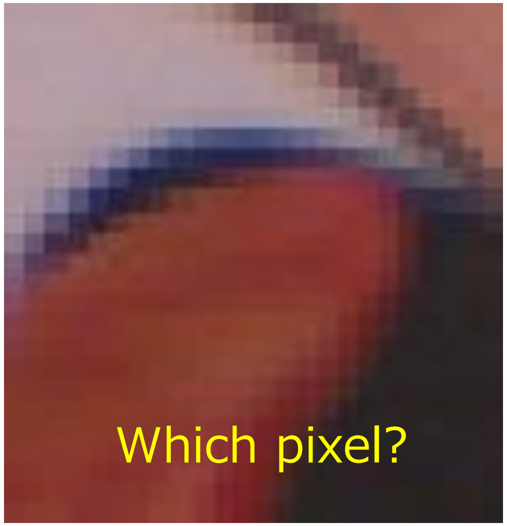
wide baseline

Corresponding pair



wide baseline

Accuracy



- baseline: b
- focal length: f
- disparity: d

$$Z = \frac{bf}{d}$$

wide baseline

narrow baseline

$$\left.\begin{array}{l} f = 10 \\ b = 10 \\ d = 10 \end{array}\right\} Z = \frac{10 \times 10}{10} = 10 \qquad \left.\begin{array}{l} f = 10 \\ b = 1 \\ d = 1 \end{array}\right\} Z = \frac{1 \times 10}{1} = 10$$

Error of 1 pixel

$$\left.\begin{array}{l} f = 10 \\ b = 10 \\ d = 11 \end{array}\right\} Z = \frac{10 \times 10}{11} = 9.1 \qquad \left.\begin{array}{l} f = 10 \\ b = 1 \\ d = 2 \end{array}\right\} Z = \frac{1 \times 10}{2} = 5$$