55:148 Digital Image Processing

Chapter 11

3D Vision, Geometry

Topics:

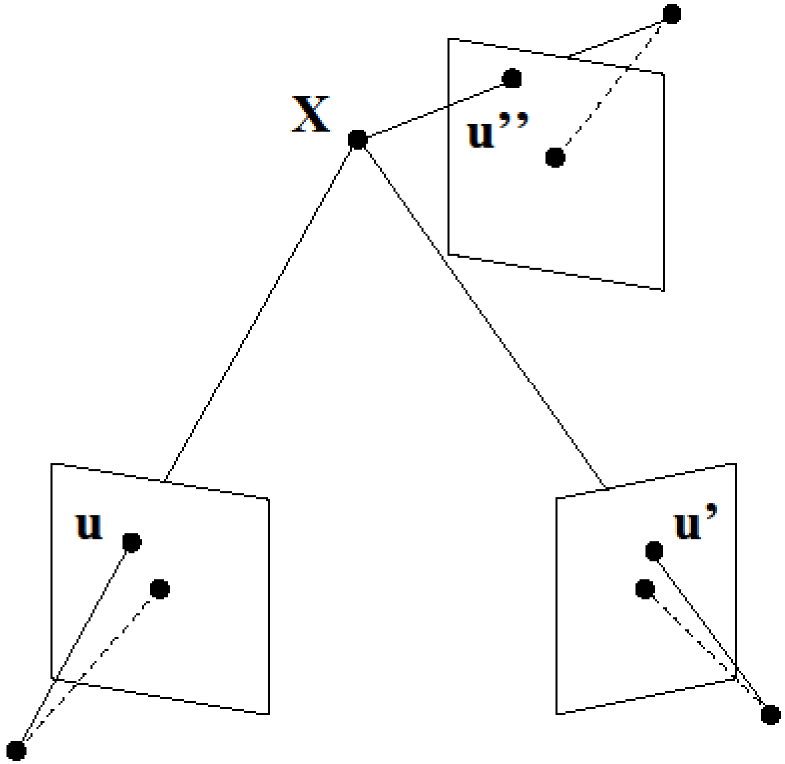
- Basics of projective geometry
  - Points and hyperplanes in projective space
  - Homography
  - Estimating homography from point correspondence
- The single perspective camera
  - An overview of single camera calibration
  - Calibration of one camera from the known scene
- Scene reconstruction from multiple views
  - Triangulation
  - Projective reconstruction
  - Matching constraints
  - Bundle adjustment
- Two cameras, stereopsis
  - The geometry of two cameras. The fundamental matrix
  - Relative motion of the camera; the essential matrix
  - Estimation of a fundamental matrix from image point correspondences
  - Applications of the epipolar geometry in vision
- Three and more cameras
  - Stereo correspondence algorithms

Scene reconstruction from multiple views

Task: Given matching points in n images. Determine the 3D scene point.

Basic Principle: Back-trace the ray in 3D scene for each image. The scene point is the common intersection of all rays.

Information needed: To back-trace a ray in the scene space, we need to know the corresponding camera matrix M_j.



Challenges:

- In an ideal condition, m back-traced rays intersect at a common point in the scene space.
- However, in real applications, due to noise and other source of errors, single-point intersection may not happen

How to proceed?

GO by Maximum likelihood estimation!

Triangulation

We want to locate the 3D scene point from its projections in several cameras.

The task is simple, if we know camera projection matrices $M_j \,|\, j = 1, \dots, n$

Problem formulation: Given image points $\mathbf{u}_j$ and camera projection matrices $M_j \,|\, j = 1, \dots, n$, solve the linear homogeneous system

$$\alpha_j \mathbf{u}_j = M_j \mathbf{X} \,|\, j = 1, \dots . n$$

Output: the 3D scene point $\mathbf{X}$

Formulate the problem into an ML optimization task (here, $[\hat{\mathrm{u}}_\mathrm{j}, \hat{\mathrm{v}}_\mathrm{j}]^\mathrm{T}$ are measures image points)

$$\min_{\mathbf{X}} \sum_{j=1}^{m} \left[\left(\frac{\mathbf{m}_{j,1}\mathbf{X}}{\mathbf{m}_{j,3}\mathbf{X}} - \hat{u}_j \right)^2 + \left(\frac{\mathbf{m}_{j,2}\mathbf{X}}{\mathbf{m}_{j,3}\mathbf{X}} - \hat{v}_j \right)^2 \right]$$

Q: Why the error factors in measured points $[\hat{\mathrm{u}}_\mathrm{j}, \hat{\mathrm{v}}_\mathrm{j}]^\mathrm{T}$ are not used here in the formulation ML optimization function?

Projection reconstruction and ambiguity

Suppose there are m scene points $\mathbf{X}_i \mid i = 1, \dots, m$ and n cameras $M_j \mid j = 1, \dots, n$

Given image points $\mathbf{u}_{i,j}$ and camera projection matrices $M_j \mid j = 1, \dots, n$, solve the linear homogeneous system

$$\alpha_{i,j}\mathbf{u}_{i,j} = M_j\mathbf{X}_i \,| i = 1, \dots, m, j = 1, \dots . n$$

Consider the task when both scene points $\mathbf{X}_i$ and camera matrices M_j are both unknown

The R.H.S. contains nonlinear terms of unknowns. Thus, its no more a linear system problem.

Projective ambiguity

Here, we identify the natural **ambiguity** in the system

Let M_j and $\mathbf{X}_i$ be a solution of the system and let T any non-singular 3×3 matrix. Then, assuming that $M_j' = M_j T^{-1}$ and $\mathbf{X}_i' = T\mathbf{X}_i$,

$$M_j'\mathbf{X}_i' = M_j T^{-1} T \mathbf{X}_i = M_j \mathbf{X}_i$$

i.e., M_j' and $\mathbf{X}_i'$ are also valid solutions to the same system.

So, there exists an ambiguity in the projective reconstruction.

More, specifically, the unknown true reconstruction $\{M_j, \mathbf{X}_i\}$ and the estimated reconstruction $\{M_j', \mathbf{X}_i'\}$ differ by a linear transformation

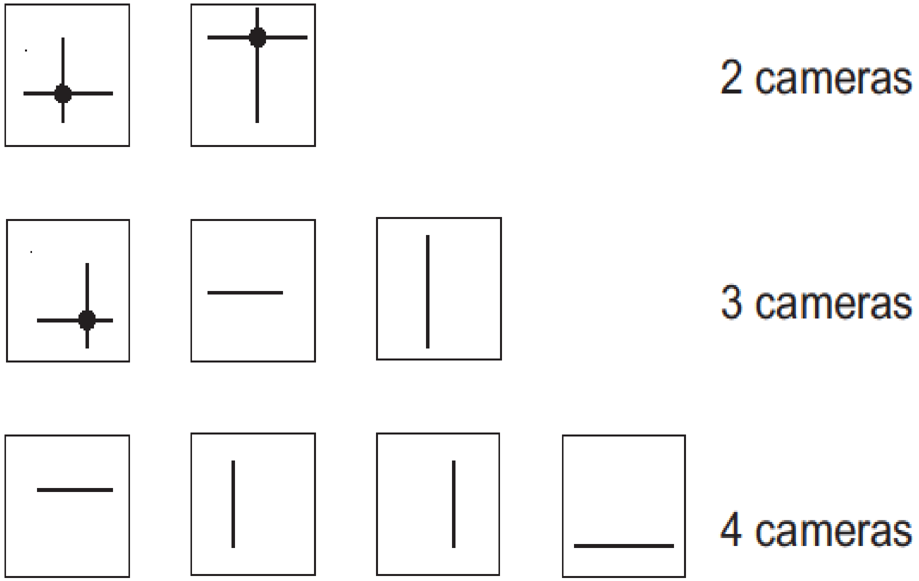
Matching constraints (Initial rough estimation)

- Relations satisfied by collections of corresponding image points in n views.
- It is used to solve initial and not very accurate estimates of camera matrices $M_j | j = 1, \dots, n$

Remember the equation

$$\left(\begin{bmatrix} S(\mathbf{u}_1)M_1 \\ \vdots \\ S(\mathbf{u}_n)M_n \end{bmatrix} = W\right)\mathbf{X} = W\mathbf{X} = \mathbf{0}$$

- To hold the equality, W must be a rank-deficient matrix
- Each row of $S(\mathbf{u}_j)$ is a line and each leads to a plane in the scene space with the transformation M_1 = a row of W
- Thus the determinant from any four rows of W is zero, i.e., the four planes have a common intersection



Bundle adjustment (optimum solution)

Nonlinear optimization function (here, $\left[\hat{\mathrm{u}}_{\mathrm{j}}, \hat{\mathrm{v}}_{\mathrm{j}}\right]^{\mathrm{T}}$ are measures image points)

$$\min_{\mathbf{X}} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\left(\frac{\mathbf{m}_{j,1}\mathbf{X}_i}{\mathbf{m}_{j,3}\mathbf{X}_i} - \hat{u}_{i,j} \right)^2 + \left(\frac{\mathbf{m}_{j,2}\mathbf{X}_i}{\mathbf{m}_{j,3}\mathbf{X}_i} - \hat{v}_{i,j} \right)^2 \right]$$

Two cameras, stereopsis

Objective: Create 3D machine vision using images from two cameras – similar to the principle of human vision

Major steps:

- Camera calibration
- Establishing point correspondence between two pairs of points from the left and the right images
- Reconstruction of 3D coordinates of the points in 3D scene space

We will start with understanding **Epipolar geometry** and **Fundamental matrix**

MATH: Points and lines in $\mathcal{P}^2$

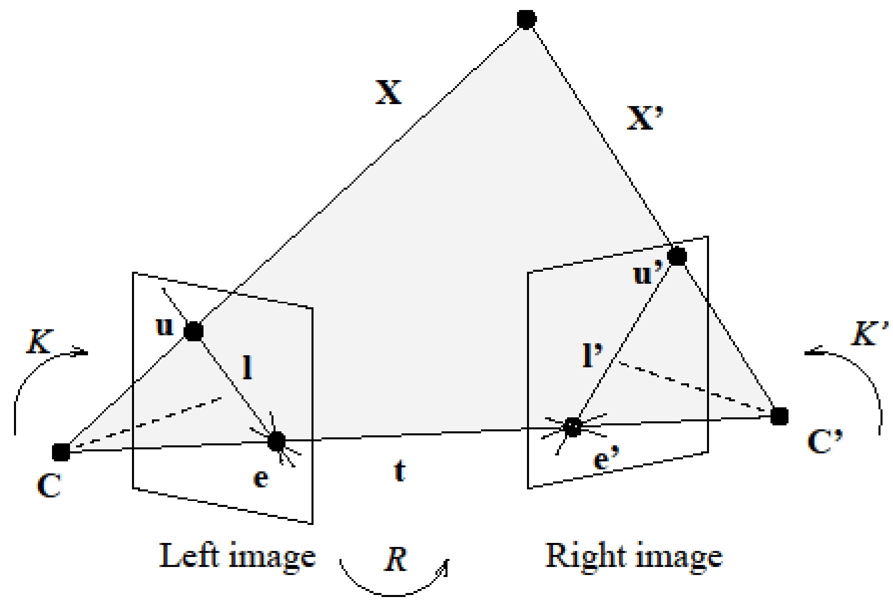
Let $\mathbf{u}$ and $\mathbf{v}$ be two points on a projection plane $\mathcal{P}^2$; a line $\mathbf{l}$ passing through the two points are expressed as $\mathbf{l} = \mathbf{u} \times \mathbf{v}$. Also, it may be shown that

$$\mathbf{l} = S(\mathbf{u})\mathbf{v}$$

Any point $\mathbf{w}$ lying on the line satisfies $\mathbf{l}^{\mathrm{T}}\mathbf{w} = 0$

Epipolar geometry and Fundamental matrix

- **Optical centers**
- **Baseline**
- **Epipoles**
- **Epipolar plane**
- **Epipolar line**



Epipolar constraints:

$$\mathbf{l}'^{\mathrm{T}}\mathbf{u}' = 0$$
$$\mathbf{l}^{\mathrm{T}}\mathbf{u} = 0$$

Fundamental matrix (F): The transformation matrix relating matching points in two images.

Find the relation between fundamental matrix and camera geometry

$$\mathbf{l}' = \mathbf{e}' \times \mathbf{u}' = \mathbf{e}' \times M'\mathbf{X} = \mathbf{e}' \times M'M^{+}\mathbf{u}$$

$$\mathbf{l}' = S(\mathbf{e}')M'M^{+}\mathbf{u} = F\mathbf{u}, \qquad \text{where, } F = S(\mathbf{e}')M'M^{+}$$

Using the epipolar constraint, $\mathbf{l}'^{\mathrm{T}}\mathbf{u}' = 0 \Rightarrow \mathbf{u}'^{\mathrm{T}}\mathbf{l}' = 0 \Rightarrow \mathbf{u}'^{\mathrm{T}}F\mathbf{u} = 0$

Also, $\mathbf{u}^{\mathrm{T}}F^{\mathrm{T}}\mathbf{u}' = 0$

A closer look at the Fundamental matrix

Consider **Case I**

$$M = [I|\mathbf{0}]$$

Following the projective ambiguity, we can always find a T s.t. the first camera matrix satisfies the above form.

Now, the center $\mathbf{C}$ is projected at the origin, i.e.,

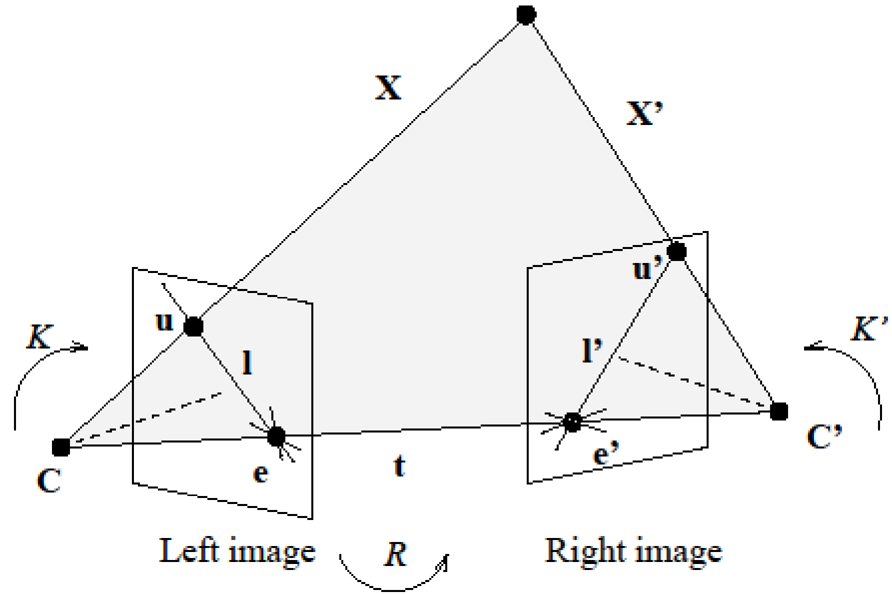
$$M\mathbf{C} = \mathbf{0} \Rightarrow \mathbf{C} = [0,0,0,1]^{\mathrm{T}}$$

Assume, $M' = \left[\widetilde{M}'|\mathbf{d}\right]$

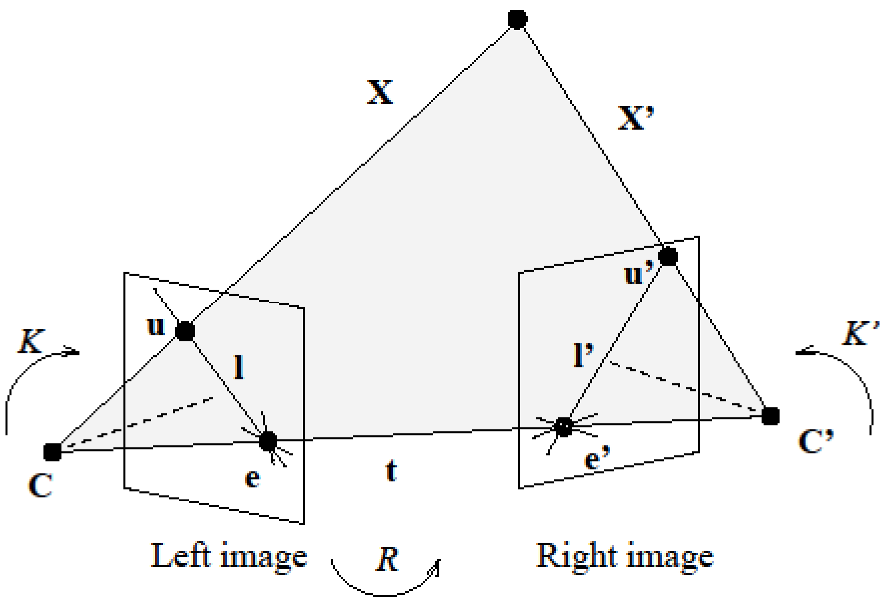
Then following, $M'\mathbf{C} = \mathbf{e}'$, $\mathbf{d}$ must be equal to $\mathbf{e}'$

Now, $M^{+}=M^{\mathrm{T}}\left(MM^{\mathrm{T}}\right)^{-1} = [I|\mathbf{0}]^{\mathrm{T}}$

Thus, $F = S(\mathbf{e}')M'M^{+} = S(\mathbf{e}')M'[I|\mathbf{0}] = S(\mathbf{e}')M' = S(M'\mathbf{C})\widetilde{M}'$



A closer look at the Fundamental matrix



Case II

Case I ignores the affine transform between image Euclidean space (ideal image space) and image affine space (acquired image space).

Case II solves the fundamental matrix under more realistic environment

$$M = K[I|\mathbf{0}] \;\; | \; K\text{: intrinsic callib. matrix}$$

The second camera matrix may be expressed in the form

$$M' = K'[R| - R\mathbf{t}]$$ Explain!

As in Case I,

$$M\mathbf{C} = \mathbf{0} \Rightarrow \mathbf{C} = [0,0,0,1]^{\mathrm{T}}$$

Now, $M^{+} = \begin{bmatrix} K^{-1} \\ \mathbf{0}^{\mathrm{T}} \end{bmatrix}$

Thus, $F = S(\mathbf{e}')M'M^{+} = S(M'\mathbf{C})K'[R| - R\mathbf{t}]\begin{bmatrix} K^{-1} \\ \mathbf{0}^{\mathrm{T}} \end{bmatrix} = S(-K'R\mathbf{t})K'RK^{-1}$

Using $S(H\mathbf{u}) = H^{-1\mathrm{T}}S(\mathbf{u})H^{-1}$, $F = K'^{-\mathrm{T}}RS(\mathbf{t})K^{-1}$