CS4277 / CS5477
3D Computer Vision

Lecture 7:
Generalized Cameras

Assoc. Prof. Lee Gim Hee

AY 2022/23

Semester 2

Course Schedule

Week	Date	Topic	Assignments
1	11 Jan	2D and 1D projective geometry	**Assignment 0:** Getting started with Python (Ungraded)
2	18 Jan	3D projective geometry, Circular points and Absolute conic	
3	25 Jan	Rigid body motion and Robust homography estimation	
4	01 Feb	Camera models and calibration	**Assignment 1:** Metric rectification and robust homography (10%) **Due:** 2359hrs, 07 Feb
5	08 Feb	Single view metrology	**Assignment 2**: Affine 3D measurement from vanishing line and point (10%) **Due:** 2359hrs, 14 Feb
6	15 Feb	The Fundamental and Essential matrices	
-	22 Feb	Semester Break	No lecture
7	**01 Mar**	**Mid-term Quiz (20%)** Lecture: Generalized cameras	**In-person Quiz (LT 15, 1900hrs – 2000hrs)** Lecture: 2000hrs – 2130hrs
8	08 Mar	Absolute pose estimation from points or lines	
9	15 Mar	Three-view geometry from points and/or lines	
10	22 Mar	Structure-from-Motion (SfM) and bundle adjustment	**Assignment 3:** SfM and Bundle adjustment (10%) **Due:** 2359hrs, 28 Mar
11	29 Mar	Two-view and multi-view stereo	**Assignment 4:** Dense 3D model from multi-view stereo (10%) **Due:** 2359hrs, 04 Apr
12	05 Apr	3D Point Cloud Processing	
13	12 Apr	Neural Field Representations	

Final Exam: 03 MAY 2023

Learning Outcomes

- Students should be able to:

1. Use the Plücker line representation to derive the generalized epipolar geometry (GEC).

2. Apply the linear 17-point algorithm to obtain the relative pose of two-view generalized camera.

3. Explain the degenerate cases of the GEC, i.e. locally central projection, axial camera and locally-central-and-axial camera.

4. Compute the absolute pose of a generalized camera using 2D-3D point or line correspondences.

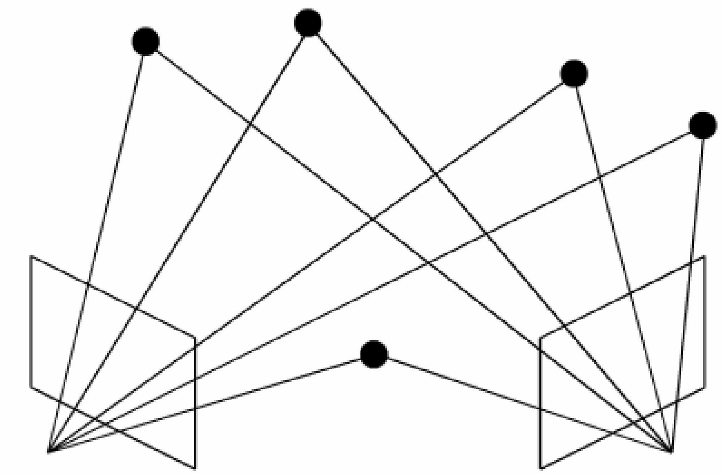
Acknowledgements

- A lot of slides and content of this lecture are adopted from:

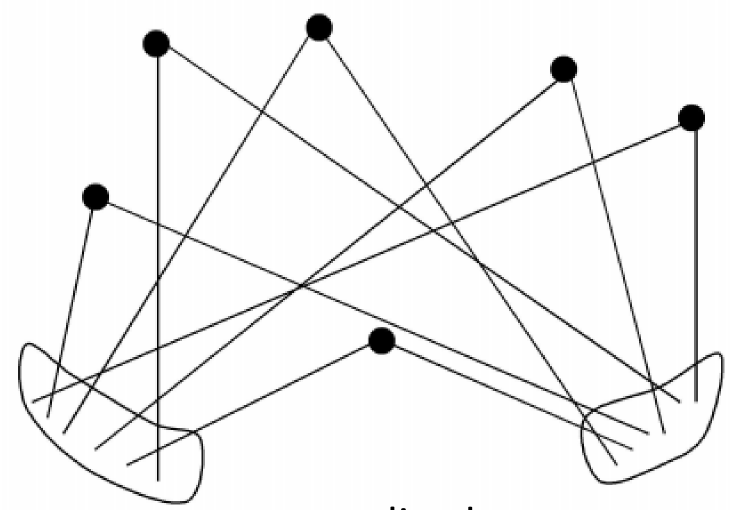
1. Robert Pless, Using Many Cameras as One, CVPR 2003.
2. Hongdong Li et al, A linear approach to motion estimation using generalized camera models, CVPR 2008.
3. Gim Hee Lee et al, Minimal Solutions for the Multi-Camera Pose Estimation Problem, IJRR 2015.
4. Gim Hee Lee, A Minimal Solution for Non-Perspective Pose Estimation from Line Correspondences, ECCV 2016.

What is a Generalized Camera?

- All light rays of the pinhole camera model converge at a single point, i.e. the camera center.
- In contrast, light rays do not meet at a single point in a generalized camera (a.k.a. non-central camera).



pinhole camera

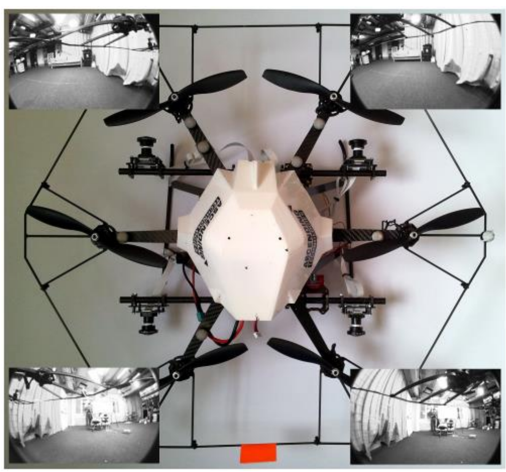


generalized camera

Image source: Henrik Stewenius et al, “Solutions to Minimal Generalized Relative Pose Problems”, ICCV workshops 2005

Example of Generalized Cameras

- Multi-camera systems with minimal or without overlapping field-of-view.



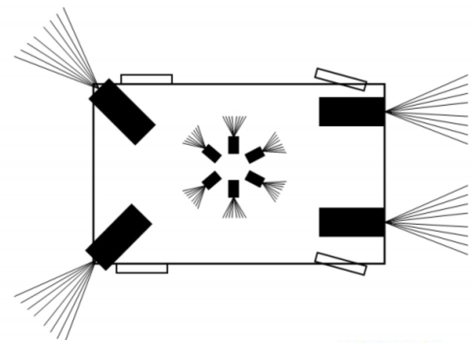


Image source: L. Heng et al, Self-Calibration and Visual SLAM with a Multi-Camera System on a Micro-Aerial Vehicle, RSS 2014
P. Furgale et al, Toward Automated Driving in Cities using Close-to-Market Sensors, an Overview of the V-Charge Project, IV 2013.
Robert Pless, Using Many Cameras as One, CVPR 2003.

Why Multi-Camera System?

Advantages:

- Cameras are low-cost and easy to maintain
- Configuration can be chosen to maximize field-of-view
- Absolute scale directly from the epipolar geometry

Challenges:

- No or minimal overlapping FOV means stereo cannot used.
- Processing each camera independently is inefficient.
- **Solution:** use the generalized camera formulation.

Plücker Vectors

- In order to describe the line in space that each pixel samples in this more general camera setting, we need a mechanism to describe arbitrary lines in space.
- The Plücker vectors give a convenient mechanism for the types of transformations required.

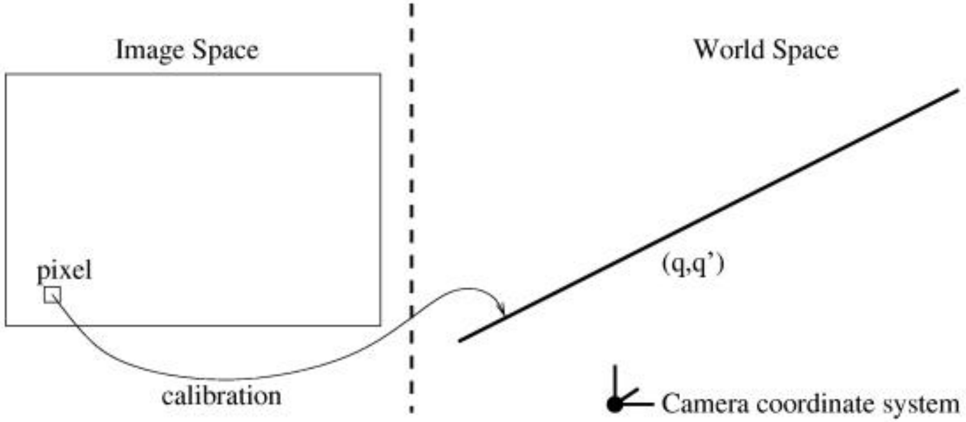


Image source: Robert Pless, Using Many Cameras as One, CVPR 2003.

Plücker Vectors

- The Plücker vectors of a line are a pair of 3-vectors: the (normalized) direction vector $\mathbf{q}$, and the moment vector $\mathbf{q}'$.

- $\mathbf{q}$ is a vector of any length in the direction of the line, and $\mathbf{q}' = \mathbf{P} \times \mathbf{q}$, for any point $\mathbf{P}$ on the line.

- Two constraints that this pair of vectors must satisfy:

1. $\mathbf{q}^\top \mathbf{q}' = 0$, and
2. the remaining five parameters are homogeneous, their overall scale does not affect which line they describe.

Plücker Vectors

- It is often convenient to force the direction vector to be a unit vector, which defines a scale for the homogeneous parameters.
- The set of all points that lie on a line with these Plücker vectors is given by:

$$(q \times q') + \alpha q, \forall \alpha \in R.$$

- If $\mathbf{q}$ is a unit vector, the point $(\mathbf{q} \times \mathbf{q}')$ is the point on the line closest to the origin and α is the (signed) distance from that point.

Plücker Vectors of a Multi-Camera System

- A pinhole camera $\mathbf{C}_i$ whose nodal point is at the origin samples a pencil of rays incident on the origin.

- A pixel (x, y) samples along a ray with Plücker vector $\left[\left(\mathrm{K}_{\mathbf{C}_i}^{-1}[x, y, 1]^\top\right)^\top, \mathbf{0}^\top\right]^\top$, where $\mathrm{K}_{\mathbf{C}_i}$ is the camera calibration matrix.

- The moment vector of the Plücker ray is zero because the point $[0, 0, 0]^\top$ is on the ray.

Plücker Vectors of a Multi-Camera System

- A camera $\mathbf{C}_i$ not at the origin has:

1. an internal calibration matrix $\mathrm{K}_{\mathbf{C}_i}$, and
2. a rotation $\mathrm{R}_{\mathbf{C}_i}$, and a translation $\mathbf{t}_{\mathbf{C}_i}$ which transform points from the camera coordinate system to the reference coordinate system.

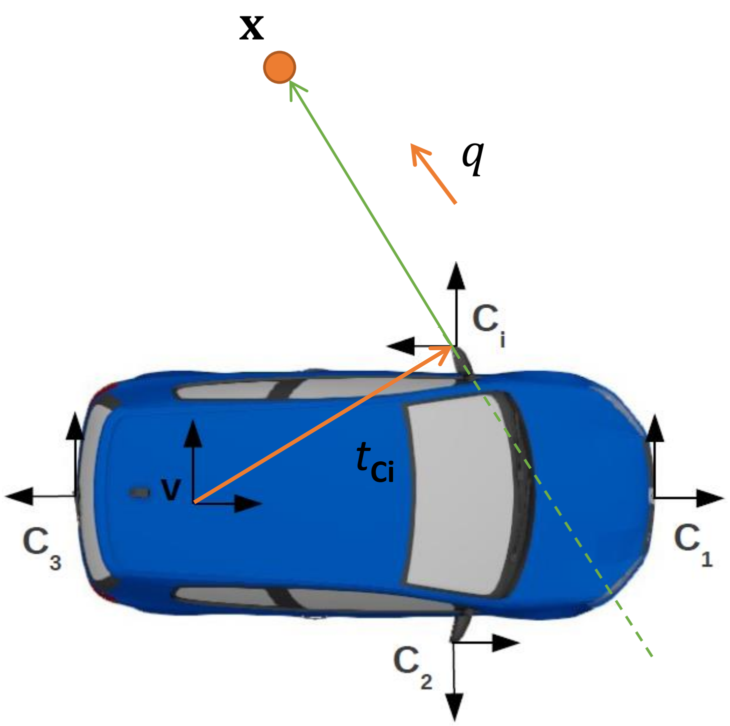
- In this case, the ray sampled by a particular pixel on the camera $\mathbf{C}_i$ a direction vector $\mathbf{q} = \mathrm{R}_{\mathbf{C}_i}\mathrm{K}_{\mathbf{C}_i}^{-1}[x, y, 1]^\top$, and a moment vector $\mathbf{t}_{\mathbf{C}_i} \times \mathbf{q}$.

Plücker Vectors of a Multi-Camera System

6-vector Plücker line $[\mathbf{q}^\top, \mathbf{q}'^\top]^\top$ to represent the light rays, where

$$\mathbf{q} = \mathrm{R}_{\mathbf{C}_i} \mathrm{K}_{\mathbf{C}_i}^{-1} [x, y, 1]^\top,$$

$$\mathbf{q}' = \mathbf{t}_{\mathbf{C}_i} \times \mathbf{q}\,.$$



Two-view Geometry

- Suppose, in two generalized images, we have a correspondence between pixel (x_1, y_1) in the first image and pixel (x_2, y_2) in a second image.

- This correspondence implies that the rays sampled by these pixels $[\mathbf{q}_1^\top, \mathbf{q}_1'^\top]^\top$, and $[\mathbf{q}_2^\top, \mathbf{q}_2'^\top]^\top$ must intersect in space.

- There is a rotation R and a translation $\mathbf{t}$ which takes points in the first coordinate system and transforms them into the new coordinate system.

Two-view Geometry

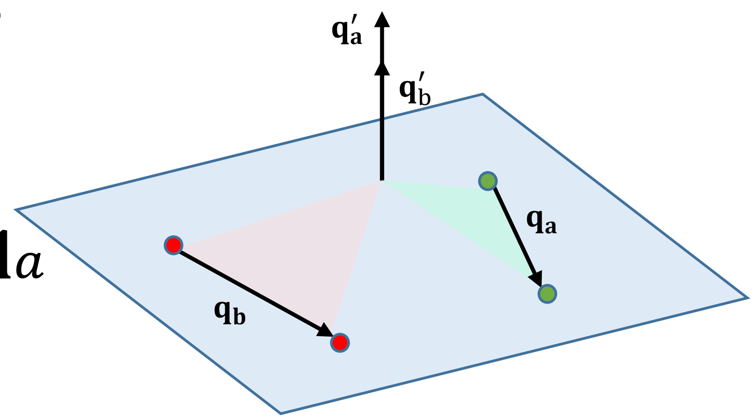
- After this rigid transformation, the Plücker vectors of the first line in the second coordinate system become:

$$\begin{bmatrix} \mathrm{R} & 0 \\ [\mathbf{t}]_\times \mathrm{R} & \mathrm{R} \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}'_1 \end{bmatrix} = \begin{bmatrix} \mathrm{R}\mathbf{q}_1 \\ [\mathbf{t}]_\times \mathrm{R}\mathbf{q}_1 + \mathrm{R}\mathbf{q}'_1 \end{bmatrix}.$$

- A pair of lines with Plücker vectors $[\mathbf{q}_a^\top, \mathbf{q}_a'^\top]^\top$, and $[\mathbf{q}_b^\top, \mathbf{q}_b'^\top]^\top$ intersect if and only if:

$$\begin{bmatrix} \mathbf{q}_b \\ \mathbf{q}'_b \end{bmatrix}^\top \begin{bmatrix} 0 & \mathrm{I} \\ \mathrm{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{q}_a \\ \mathbf{q}'_a \end{bmatrix} = \mathbf{q}_b^\top \mathbf{q}'_a + \mathbf{q}_b'^\top \mathbf{q}_a$$

$$= 0.$$



Two-view Geometry

- Combining the two equations, we get:

$$\mathbf{q}_2^\top [\mathbf{t}]_\times \mathrm{R}\mathbf{q}_1 + \mathbf{q}_2^\top \mathrm{R}\mathbf{q}_1' + \mathbf{q}_2'^\top \mathrm{R}\mathbf{q}_1 = 0.$$

- Or in matrix form:

$$\begin{bmatrix} \mathbf{q}_2 \\ \mathbf{q}_2' \end{bmatrix}^\top \begin{bmatrix} \mathrm{E} & \mathrm{R} \\ \mathrm{R} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_1' \end{bmatrix} = 0.$$

- where $\mathrm{E} = [\mathbf{t}]_\times \mathrm{R}$ is the essential matrix between the two frames.

Generalized Epipolar Geometry

- We call the equation generalized epipolar geometry:

$$\begin{bmatrix} \mathbf{q}_2 \\ \mathbf{q}_2' \end{bmatrix}^\top \begin{bmatrix} \mathrm{E} & \mathrm{R} \\ \mathrm{R} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_1' \end{bmatrix} = 0.$$

- The matrix $\begin{bmatrix} \mathrm{E} & \mathrm{R} \\ \mathrm{R} & 0 \end{bmatrix}$ is the 6×6 generalized essential matrix.

- And $[\mathbf{q}_1^\top, \mathbf{q}_1'^\top]^\top$ and $[\mathbf{q}_2^\top, \mathbf{q}_2'^\top]^\top$ are a point correspondence is the first and second view represented as Plücker coordinates.

Generalized Epipolar Geometry

- There are a total of 18 unique entries from E and R in the generalized essential matrix.

- The 18 unique entries are linear to the Plücker line correspondences, i.e. we can write the Generalized epipolar geometry into:

$$\mathbf{a}^\top \mathbf{g} = 0,$$

- $\mathbf{a}$ is a 18-vector made up of the known elements in $[\mathbf{q}_1^\top, \mathbf{q}_1'^\top]^\top$ and $[\mathbf{q}_2^\top, \mathbf{q}_2'^\top]^\top$, and $\mathbf{g}$ is an 18-vector made up of the unknown elements in E and R.

Generalized Epipolar Geometry

- We need a minimal of 17-point correspondences to solve for the unknown $\mathbf{g}$, i.e. we stack the equations to get:

$$\mathrm{A}\mathbf{g} = 0,$$

- A is an $n \times 18$ matrix made up of the known point correspondences with $n \geq 17$.

- We can then solve for the unknown $\mathbf{g}$ using the SVD method.

Generalized Epipolar Geometry

- The SVD method gives a one parameter λ family of solutions, i.e. $\mathbf{g} = \lambda\mathbf{v}$, where $\mathbf{v}$ is the right singular vector of A with the least singular value.
- Let $\mathbf{g} = [e_1, \dots, e_9, r_1, \dots, r_9]$, where e_i and r_i are the entries in E and R.
- We can solve for λ by enforcing the constraint that the $\det(\mathrm{R}) = 1$.
- The relative translation $\mathbf{t}$ can then be solved from $\mathrm{E} = [\mathbf{t}]_\times \mathrm{R}$.

Generalized Epipolar Geometry

- Note that in general, there is $n \geq 17$ point correspondences from different cameras with known extrinsics, i.e. $\{\mathrm{R}_{\mathbf{C}_1}, \mathbf{t}_{\mathbf{C}_1}, \ldots, \mathrm{R}_{\mathbf{C}_i}, \mathbf{t}_{\mathbf{C}_i}\}$.

- This means that there is no scale ambiguity in the relative translation $\mathbf{t}$ solved from the generalized epipolar geometry.

Generalized Point Reconstruction

- Given the camera transformation R, $\mathbf{t}$ and corresponding points, it is possible to determine the 3D coordinates of the world point in view.

- Using the points on the Plücker line equation defined earlier, we get:

$$\mathrm{R}(\mathbf{q}_1 \times \mathbf{q}_1') + \alpha_1 \mathrm{R}\mathbf{q}_1 + \mathbf{t} = (\mathbf{q}_2 \times \mathbf{q}_2') + \alpha_2 \mathbf{q}_2$$

- Which can be written into the over-determinate linear equations $\mathrm{A}[\alpha_1, \alpha_2]^\top = \mathbf{b}$, where the unknowns α_1, α_2 can be solved.

Generalized Point Reconstruction

- Finally, the 3D point can be reconstructed as:

$$\mathbf{P} = (\mathbf{q}_1 \times \mathbf{q}_1') + \alpha_1 \mathbf{q}_1.$$

- where the coordinate frame of the first view is used as the reference frame.

Analysis of Degeneracies

- We will now look at several configurations of the multi-camera setup that will lead to the degeneracy of the generalized epipolar geometry.

- Specifically, we will consider three configurations:

1. Locally central projection
2. Axial cameras
3. Locally-central-and-axial cameras

Analysis of Degeneracies

- An image ray passing through a point $\mathbf{v}$ (e.g., camera center) with unit direction $\mathbf{x}$ can be represented by a Plücker 6-vector $\mathbf{L} = (\mathbf{x}^\top, (\mathbf{v} \times \mathbf{x})^\top)^\top$.

- Using this representation, we then re-state the generalized epipolar geometry (GEC) as follows:

$$\mathbf{x}_i{}^\top \mathrm{E}\mathbf{x}'_i + \mathbf{x}_i{}^\top \mathrm{R}(\mathbf{v}'_i \times \mathbf{x}'_i) + (\mathbf{v}_i \times \mathbf{x}_i)^\top \mathrm{R}\,\mathbf{x}'_i = 0\ .$$

- Our initial goal is to examine the linear structure of the solution set to these equations under various camera geometries.

Analysis of Degeneracies

- We identify several degeneracies for the set of equations arising from the GEC, which cause the set of equations to have smaller than the expected rank.
- Assume that there are at least r equations arising from point correspondences via the GEC.
- If the linear family of solutions has rank 18−r, then the equation system must have a rank no greater than r.

The Most General Case

- In the most general case, the camera is simply a set of unconstrained image rays in general position.
- The rank is 17 and an unique solution can be obtained from the SVD method.

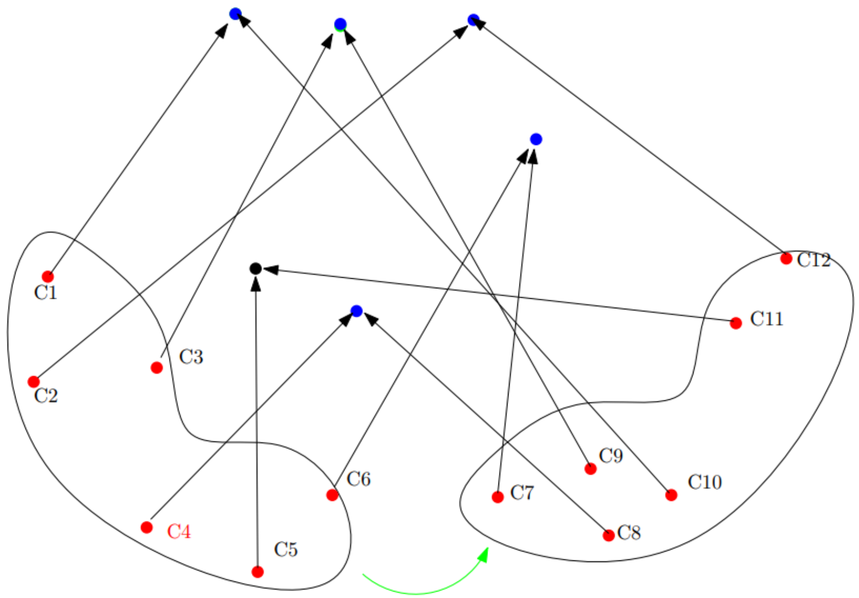


Image source: H. Li, A linear approach to motion estimation using generalized camera models, CVPR 2008.

Locally Central Projection

- This is the multi-camera setting where each component camera is a central projection camera, so that all rays go through the camera center.

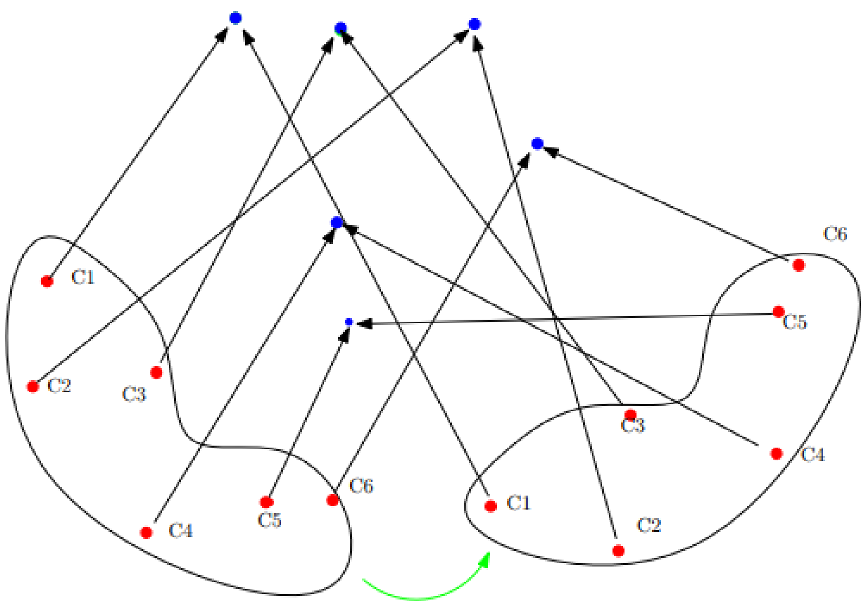


Image source: H. Li, A linear approach to motion estimation using generalized camera models, CVPR 2008.

Locally Central Projection

- Since rays are represented in a coordinate system attached to the camera rig.
- The correspondence is between points $(\mathbf{x}_i, \mathbf{v}_i) \leftrightarrow (\mathbf{x}'_i, \mathbf{v}'_i)$, where $\mathbf{v}_i$ is the camera center, and note that $\mathbf{v}'_i = \mathbf{v}_i$.
- The GEC now becomes:

$$\mathbf{x}_i{}^\top \mathrm{E}\mathbf{x}'_i + \mathbf{x}_i{}^\top \mathrm{R}(\mathbf{v}_i \times \mathbf{x}'_i) + (\mathbf{v}_i \times \mathbf{x}_i)^\top \mathrm{R}\,\mathbf{x}'_i = 0\,.$$

Locally Central Projection

- Now let (E, R) be a one solution to this set of equations, with $\mathrm{E} \neq 0$.
- It is easily seen that $(0, \mathrm{I})$ is also a (degenerate) solution, with $\mathrm{E} = 0$.

Proof:

Substituting $(0, \mathrm{I})$ in the GEC results in

$$\left(\mathbf{x}_i{}^\top (\mathbf{v}_i \times \mathbf{x}'_i) + (\mathbf{v}_i \times \mathbf{x}_i)^\top \mathbf{x}'_i\right),$$

which is zero because of the anti-symmetry of the triple-product. $\square$

Locally Central Projection

- Generically, the rank is not less than 16, so a complete solution to this set of equations is therefore of the form $(\lambda\mathrm{E}, \lambda\mathrm{R} + \mu\mathrm{I})$, a two-dimensional linear family.

Proof:

Substituting $(\lambda\mathrm{E}, \lambda\mathrm{R} + \mu\mathrm{I})$ into the GEC, we get

$$\mathbf{x}_i^\top \lambda\mathrm{E}\mathbf{x}'_i + \mathbf{x}_i^\top(\lambda\mathrm{R} + \mu\mathrm{I})(\mathbf{v}_i \times \mathbf{x}'_i) + \left(\mathbf{v}_i \times \mathbf{x}_i^\top\right)(\lambda\mathrm{R} + \mu\mathrm{I})\mathbf{x}'_i = 0,$$

$$\Rightarrow \lambda\underbrace{\left(\mathbf{x}_i^\top\mathrm{E}\mathbf{x}'_i + \mathbf{x}_i^\top\mathrm{R}(\mathbf{v}_i \times \mathbf{x}'_i) + \left(\mathbf{v}_i \times \mathbf{x}_i^\top\right)\mathrm{R}\mathbf{x}'_i\right)}_{\text{=0 for all } \lambda \text{(original GEC)}} +$$

$$\mu\underbrace{\left(\mathbf{x}_i^\top(\mathbf{v}_i \times \mathbf{x}'_i) + \left(\mathbf{v}_i \times \mathbf{x}_i^\top\right)\mathbf{x}'_i\right)}_{\text{=0 for all } \mu \text{ (triple product)}} = 0$$

$\square$

Locally Central Projection

- An interesting property of the set of solutions $(\lambda \mathrm{E}, \lambda \mathrm{R} + \mu \mathrm{I})$ to the GEC under locally central projection is found:

  The ambiguity is contained entirely in the estimation of R, while the essential matrix E is still able to be determined uniquely up to scale.

- **Remark:** Under pure translation, the GEC degenerates to the single camera epipolar geometry, i.e. $\mathbf{x}_i^\top \lambda \mathrm{E} \mathbf{x}'_i = 0$.

Axial Cameras

- This is defined as a generalized camera in which all the rays intersect in a single line, called the axis.

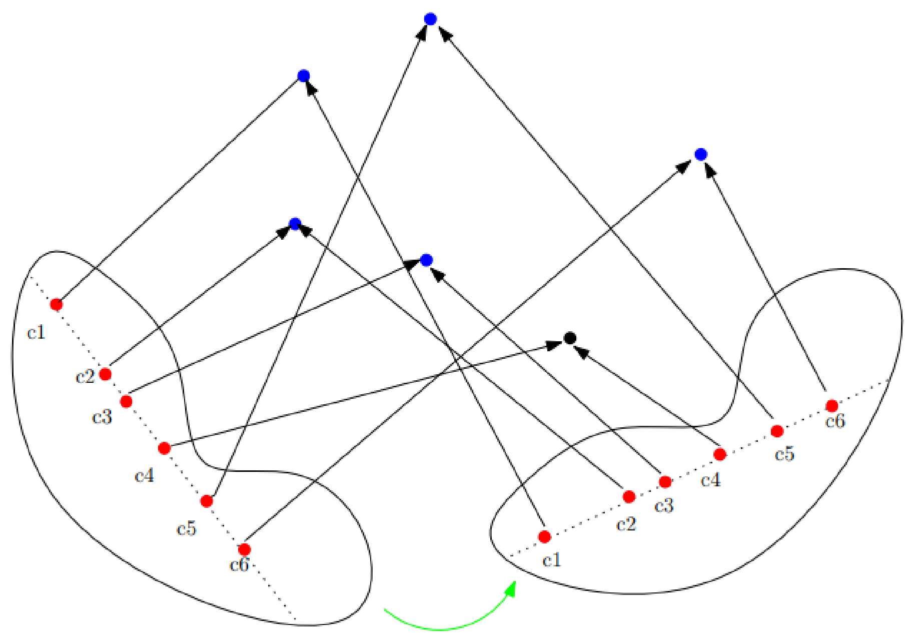


Image source: H. Li, A linear approach to motion estimation using generalized camera models, CVPR 2008.

Axial Cameras

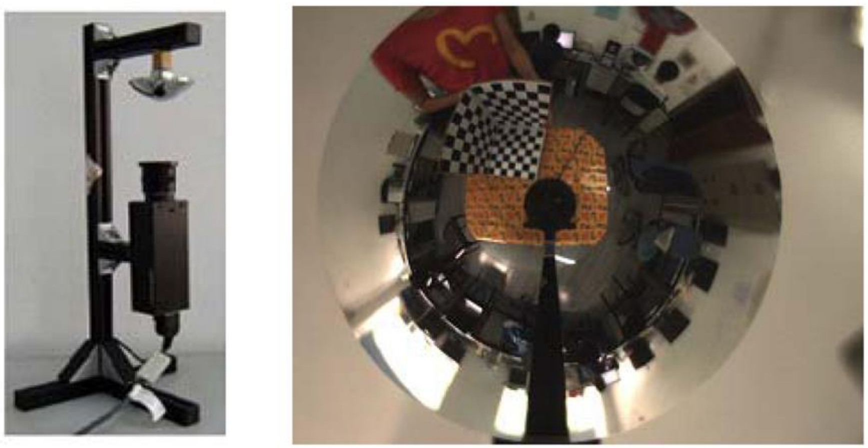
- There are several examples of this which may be of practical interest:

1. A pair of rigidly mounted central projection cameras (for instance, ordinary perspective cameras).
2. A set of central projection cameras with collinear centers. We call this a linear camera array.
3. A set of non-central catadioptric or fisheye cameras mounted with collinear axes.

- The first two cases are also locally central projections, provided that points are not tracked between cameras.

Catadioptric and Fisheye Cameras

Catadioptric Camera



(a) (b)

Fisheye Cameras

Image source: https://www.mdpi.com/1424-8220/12/6/7299/htm

https://www.teknistore.com/en/camera-lenses-and-accessories/12829-universal-235-detachable-clip-fisheye-lens-camera-for-iphone-6-6-plus-all-phones.html

Axial Cameras

- Let us assume that the origin of the world coordinate system lies on the axis.
- We may write $\mathbf{v}_i = \alpha_i \mathbf{w}$ and $\mathbf{v}'_i = \alpha'_i \mathbf{w}$, where $\mathbf{w}$ is the direction vector of the axis, the GEC then takes the form:

$$\mathbf{x}_i{}^\top \mathrm{E} \mathbf{x}'_i + \alpha_i (\mathbf{w} \times \mathbf{x}_i)^\top \mathrm{R}\, \mathbf{x}'_i + \alpha'_i \mathbf{x}_i{}^\top \mathrm{R}(\mathbf{w} \times \mathbf{x}'_i) = 0\,.$$

- Generically, the equation system has rank 16, so the general solution for an axial camera is $(\lambda \mathrm{E}, \lambda \mathrm{R} + \mu \mathbf{w}\mathbf{w}^\top)$.

Axial Cameras

- Note the most important fact that the E part of the solution is constant, and the ambiguity only involves the R part of the solution.

- Thus, we may retrieve the matrix E without ambiguity from the degenerate system of equations.

- It is important to note that this fact depends on the choice of coordinate system such that the origin lies on the axis.

- Without this condition, there is still a two-dimensional family of solutions, but the solution for the matrix E is not invariant.

Locally-Central-and-Axial Cameras

- We have seen already that for locally central projections, (0, R) is also a solution.
- However, in the case of a locally central and axial camera array, a further degeneracy occurs.
- The condition of local centrality means that $\alpha_i = \alpha_i'$ in

$$\mathbf{x}_i{}^\top \mathbf{E}\mathbf{x}_i' + \alpha_i(\mathbf{w} \times \mathbf{x}_i)^\top \mathrm{R}\,\mathbf{x}_i' + \alpha_i' \mathbf{x}_i{}^\top \mathrm{R}(\mathbf{w} \times \mathbf{x}_i') = 0\,.$$

Locally-Central-and-Axial Cameras

- We may now identify a further solution $(0, [\mathbf{w}]_\times)$, since

$$
\begin{aligned}
& (\mathbf{w} \times \mathbf{x}_i)^\top [\mathbf{w}]_\times \mathbf{x}'_i + {\mathbf{x}_i}^\top [\mathbf{w}]_\times (\mathbf{w} \times \mathbf{x}'_i) \\
= \quad & (\mathbf{w} \times \mathbf{x}_i)^\top (\mathbf{w} \times \mathbf{x}'_i) + (\mathbf{x}_i \times \mathbf{w})^\top (\mathbf{w} \times \mathbf{x}'_i) = 0 \, .
\end{aligned}
$$

- In summary, in the case of a locally central axial camera the complete solution set is of the form

$$
(\alpha \mathrm{E}, \; \alpha \mathrm{R} + \beta \mathrm{I} + \gamma [\mathbf{w}]_\times + \delta \mathbf{w}\mathbf{w}^\top) \, .
$$

Locally-Central-and-Axial Cameras

- The complete solution set is under the assumption that the coordinate origin lies on the camera axis.

- Once more, the E part of the solution is determined uniquely up to scale, even though there is a 4-dimensional family of solutions.

Linear Algorithm Under Degeneracy

- As we have seen earlier, each point correspondence gives one equation:

$$\mathbf{x}_i{}^\top \mathrm{E}\mathbf{x}'_i + \mathbf{x}_i{}^\top \mathrm{R}(\mathbf{v}'_i \times \mathbf{x}'_i) + (\mathbf{v}_i \times \mathbf{x}_i)^\top \mathrm{R}\,\mathbf{x}'_i = 0\ .$$

- Given sufficiently many point correspondences, we may solve for the entries of matrices E and R linearly from the set of equations:

$$\mathrm{A}(\mathrm{vec}(\mathrm{E})^\top, \mathrm{vec}(\mathrm{R})^\top) = 0.$$

Linear Algorithm Under Degeneracy

- However, we have seen that the standard SVD solution to this set of equations gives a whole family of solutions.
- If one ignores the rank deficiency of the equations, totally spurious solutions may be found.
- **Observation:** The E part of the solution is unchanged by the ambiguity.

  This suggests using the set of equations to solve only for E, and forget about the constraint on R.

Linear Algorithm Under Degeneracy

- Thus, given a set of equations

$$\mathtt{A}(\mathrm{vec}(\mathtt{E})^\top, \mathrm{vec}(\mathtt{R})^\top)^\top = \mathbf{0}$$

- we find the solution that minimizes

$$\|\mathtt{A}\,(\mathrm{vec}(\mathtt{E})^\top, \mathrm{vec}(\mathtt{R})^\top)^\top\| \text{ subject to } \|\mathtt{E}\| = 1\ ,$$

- instead of

$$\|(\mathrm{vec}(\mathtt{E})^\top, \mathrm{vec}(\mathtt{R})^\top)\| \ = \ 1$$

as in the standard SVD algorithm.

Linear Algorithm Under Degeneracy

- Write the equations as

$$\mathrm{A_E}\mathrm{vec}(\mathrm{E})^\top + \mathrm{A_R}\mathrm{vec}(\mathrm{R})^\top = 0,$$

- where $\mathrm{A_E}$ and $\mathrm{A_R}$ are submatrices of A consisting of the first and last 9 columns.

- Finding the solution that satisfies $\|\mathrm{vec}(\mathrm{E})\| = 1$ is equivalent to solving

$$(\mathrm{A_R}\mathrm{A_R^+} - \mathrm{I})\mathrm{A_E}\mathrm{vec}(\mathrm{E}) = \mathbf{0}$$

where $\mathrm{A_R^+}$ is the pseudo-inverse of $\mathrm{A_R}$.

Linear Algorithm Under Degeneracy

- This equation is then solved using the standard SVD method, and it gives a unique solution for E.
- Decompose E to get the pair of rotation matrices R and R'.
- Once R is known, we can solve for the translation $\mathbf{t}$ linearly using the GEC:

$$\mathbf{x}_i{}^\top [\mathbf{t}]_\times (\mathrm{R}\,\mathbf{x}'_i) + \mathbf{x}_i{}^\top \mathrm{R}(\mathbf{v}'_i \times \mathbf{x}'_i) + (\mathbf{v}_i \times \mathbf{x}_i)^\top \mathrm{R}\,\mathbf{x}'_i = 0\,.$$

Linear Algorithm Under Degeneracy

- The correct pose is the one where the 3D points are in front of both cameras.

- **Remarks:** We do not take the translation from the decomposition of the essential matrix because the unknown scale.

  Using the GEC to solve for the translation gives us absolute scale.

Generalized Pose Estimation Problem

- Given a set of three:

1. 3D points $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ in F_W seen by arbitrary cameras on the multi-camera system, and
2. Their corresponding 2D image coordinates $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$.

- Find the rigid transformation R and $\mathbf{t}$ that brings the multi-camera frame F_G into the world frame F_W.

Generalized Pose Estimation Problem

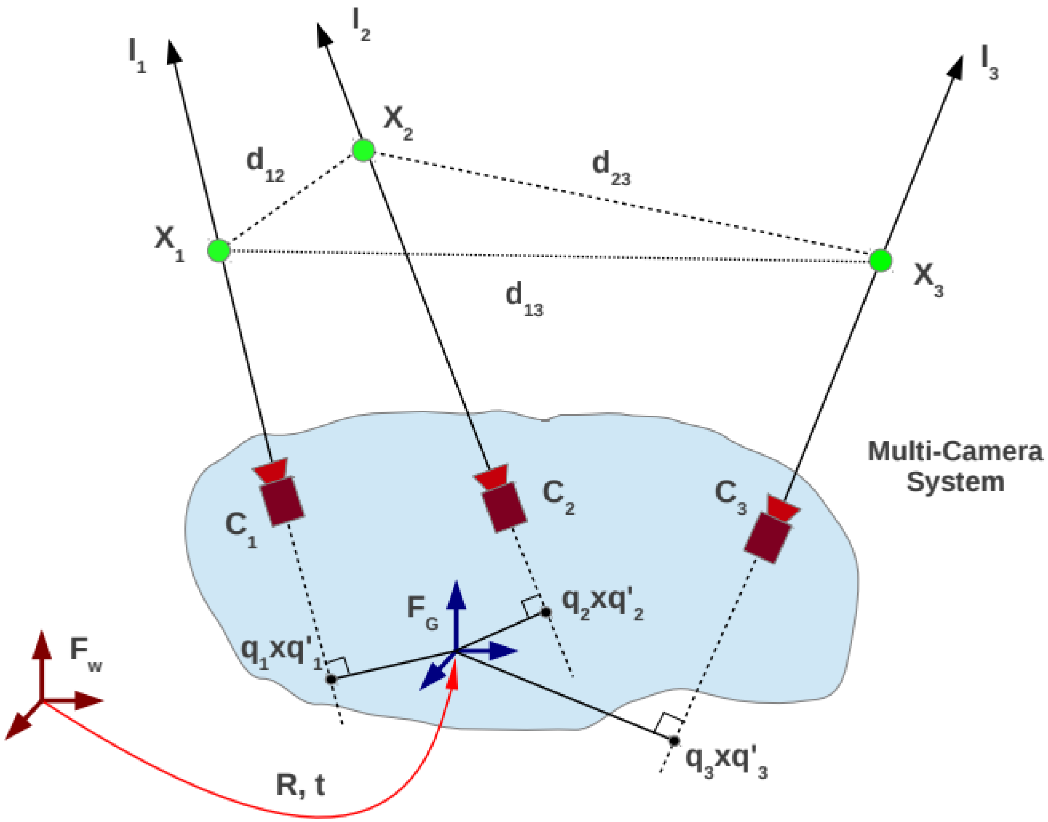


Image source: G. H. Lee et al, Minimal Solutions for the Multi-Camera Pose Estimation Problem, IJRR 2015.

Plücker Line Representation

- As seen earlier, any point $\mathbf{X}_i^G$ expressed in the multi-camera frame F_G is given by

$$\mathbf{X}_i^G = \mathbf{q}_i \times \mathbf{q}_i' + \lambda_i \mathbf{q}_i,$$

- where λ_i is the signed distance from $\mathbf{q}_i \times \mathbf{q}_i'$ to $\mathbf{X}_i^G$.

- Note that λ must be always positive for the 3D point to appear in front of the camera.

Distance Constraints

- The distances d_{ij}, $(i, j) \in \{(1,2), (1,3), (2,3)\}$ between the 3D points $\mathbf{X}_i$ in F_W are the same as the distances between the 3D points $\mathbf{X}_i^{\mathrm{G}}$ in F_G , i.e.

$$\left\|\mathbf{X}_i - \mathbf{X}_j\right\|^2 = \left\|\mathbf{X}_i^{\mathrm{G}} - \mathbf{X}_j^{\mathrm{G}}\right\|^2.$$

- Substituting the Plücker line equation into the distance constraint, we get:

$$\left\|\mathbf{X}_i - \mathbf{X}_j\right\|^2 = \left\|(\mathbf{q}_i \times \mathbf{q}_i' + \lambda_i \mathbf{q}_i) - (\mathbf{q}_j \times \mathbf{q}_j' + \lambda_j \mathbf{q}_j)\right\|^2,$$

- Where λ_1, λ_2 and λ_3 are the unknown signed distances.

Solving for the Unknown Signed Distances

- Expanding and rearranging the unknowns in the previous equation, we get:

$$k_{11}\lambda_1^2 + (k_{12}\lambda_2 + k_{13})\lambda_1 + (k_{14}\lambda_2^2 + k_{15}\lambda_2 + k_{16}) = 0 \qquad \text{(a)}$$

$$k_{21}\lambda_1^2 + (k_{22}\lambda_3 + k_{23})\lambda_1 + (k_{24}\lambda_3^2 + k_{25}\lambda_3 + k_{26}) = 0 \qquad \text{(b)}$$

$$k_{31}\lambda_2^2 + (k_{32}\lambda_3 + k_{33})\lambda_2 + (k_{34}\lambda_3^2 + k_{35}\lambda_3 + k_{36}) = 0 \qquad \text{(c)}$$

- where k are the coefficients made up of the known Plücker line coordinates $\mathbf{q}_i$ and $\mathbf{q}'_i$, and 3D world points $\mathbf{X}_i$.

Solving for the Unknown Signed Distances

- Eliminating λ_1 from Equation (a) and (b), we get a polynomial of $f(\lambda_2, \lambda_3) = 0$.

- We further eliminate λ_2 in $f(\lambda_2, \lambda_3) = 0$ and Equation (c) to get an eight-degree polynomial:

$$A\lambda_3^8 + B\lambda_3^7 + C\lambda_3^6 + D\lambda_3^5 + E\lambda_3^4 + F\lambda_3^3 + G\lambda_3^2 + H\lambda_3 + I = 0$$

- where A, B, C, D, E, F, G, H and I are coefficients made up of k from Equations (a), (b) and (c).

Solving for the Unknown Signed Distances

- Eight solutions for λ_3 can be obtained from the eigenvalues of the Companion matrix.

- λ_2 can be found by back-substituting λ_3 in Equation (c), i.e.

$$\lambda_2 = \frac{1}{2a}(-b \pm \sqrt{b^2 - 4ac})$$

- Where $a = k_{31}, \; b = k_{32}\lambda_3 + k_{33}, \; c = k_{34}\lambda_3^2 + k_{35}\lambda_3 + k_{36}.$

- Similarly, λ_1 can be found by back-substituting λ_2 into Equation (a) which takes a similar form.

Solving for the Unknown Signed Distances

- A total of up to 32 (i.e.8×2×2) solution triplets of λ_1, λ_2 and λ_3 can be obtained.

- A solution triplet is discarded if any one of the λs is an imaginary or negative value.

- For each of the solution triplet, we solve for the pose using absolute orientation.

- Finally, the solution that yields the highest inlier count is chosen.

Generalized Pose Estimation from Line Correspondences

- Given the 2D-3D line correspondences $L_j^W \leftrightarrow l_j^c$ defined in the world frame and image coordinate frame, respectively.

- Find the pose of the multi-camera system with respect to the fixed world frame, i.e. relative transformation

$$T_G^W = \begin{pmatrix} R_G^W & t_G^W \\ 0_{1\times 3} & 1 \end{pmatrix} .$$

Generalized Pose Estimation from Line Correspondences

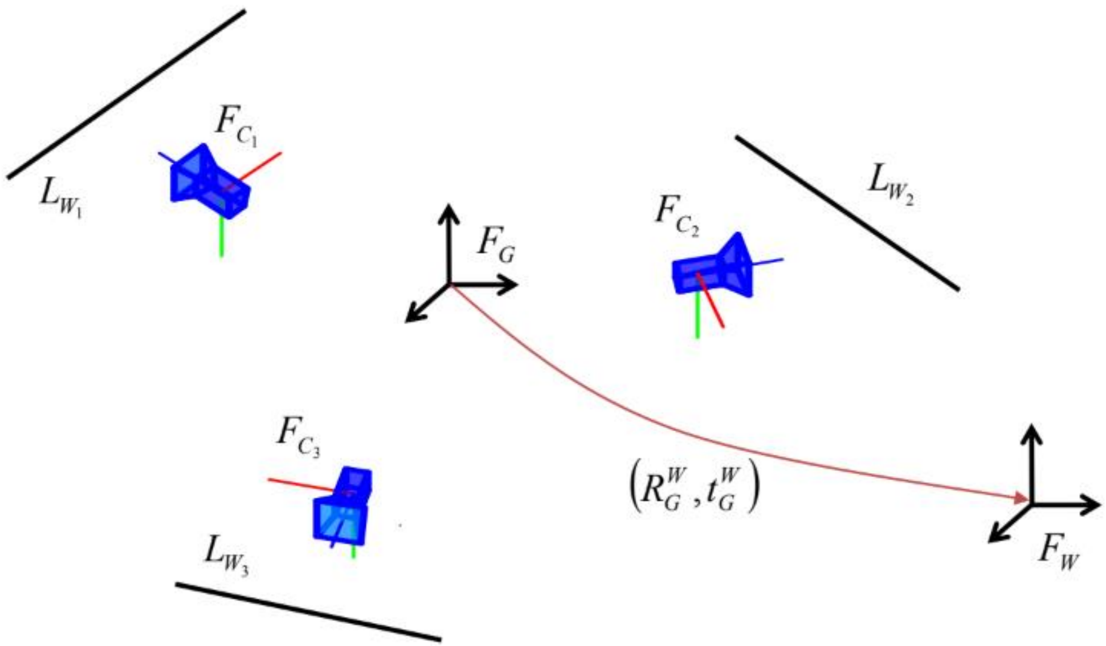


Image source: G. H. Lee, A Minimal Solution for Non-Perspective Pose Estimation from Line Correspondences, ECCV 2016.

Plücker Representation of 3D Lines

- Let $\mathrm{P}_a^W = [\mathrm{P}_{ax}^W \; \mathrm{P}_{ay}^W \; \mathrm{P}_{az}^W \; 1]^\top$ and $\mathrm{P}_b^W = [\mathrm{P}_{bx}^W \; \mathrm{P}_{by}^W \; \mathrm{P}_{bz}^W \; 1]^\top$ be the homogeneous coordinates in F_W that represent the two end points of the 3D line segment L_W.

- The 6-vector Plücker line of the 3D line segment is given by $L_W = [U_W^\top \; V_W^\top]^\top$, where

$$V_W = \frac{P_b^W - P_a^W}{\left\| P_b^W - P_a^W \right\|}, \quad U_W = P_a^W \times V_W.$$

- V_W is the unit direction and U_W is the moment of the first 3D line segment end point P_a^W and unit direction V_W.

Plücker Representation of 3D Lines

- $L_W = [U_W^\top \; V_W^\top]^\top$ is known and it is expressed in the coordinate frame of the fixed world frame F_W.

- L_W can be expressed in the camera reference frame F_C as follows:

$$L_C = \mathcal{T}_W^C L_W = \begin{pmatrix} R_W^C & \lfloor t_W^C \rfloor_\times R_W^C \\ 0_{3\times 3} & R_W^C \end{pmatrix} L_W ,$$

- $\mathcal{T}_W^C$ is the transformation matrix that brings a Plücker line defined in F_W to F_C.

Plücker Representation of 3D Lines

- Specifically,

$$T_W^C = \begin{pmatrix} R_W^C & t_W^C \\ 0_{1\times 3} & 1 \end{pmatrix} = \begin{pmatrix} R_G^C & t_G^C \\ 0_{1\times 3} & 1 \end{pmatrix} \begin{pmatrix} R_W^G & t_W^G \\ 0_{1\times 3} & 1 \end{pmatrix} = \begin{pmatrix} R_G^C R_W^G & R_G^C t_W^G + t_G^C \\ 0_{1\times 3} & 1 \end{pmatrix},$$

- where (R_G^C, t_G^C) is the known camera extrinsics, and (R_G^W, t_G^W) is the unknown pose of the multi-camera system.

Plücker Representation of 3D Lines

- Since $L_C = [U_C^\top \ V_C^\top]^\top$,

$$U_C = \left(R_W^C \ \lfloor t_W^C \rfloor_\times R_W^C\right) \begin{pmatrix} U_W \\ V_W \end{pmatrix}$$

- is a vector in F_C perpendicular to the plane formed by the projection of the 3D line onto the camera image.

$$V_C = R_W^C V_W$$

- is the unit direction vector of the 3D line in the camera reference frame F_C.

Plücker Representation of 3D Lines

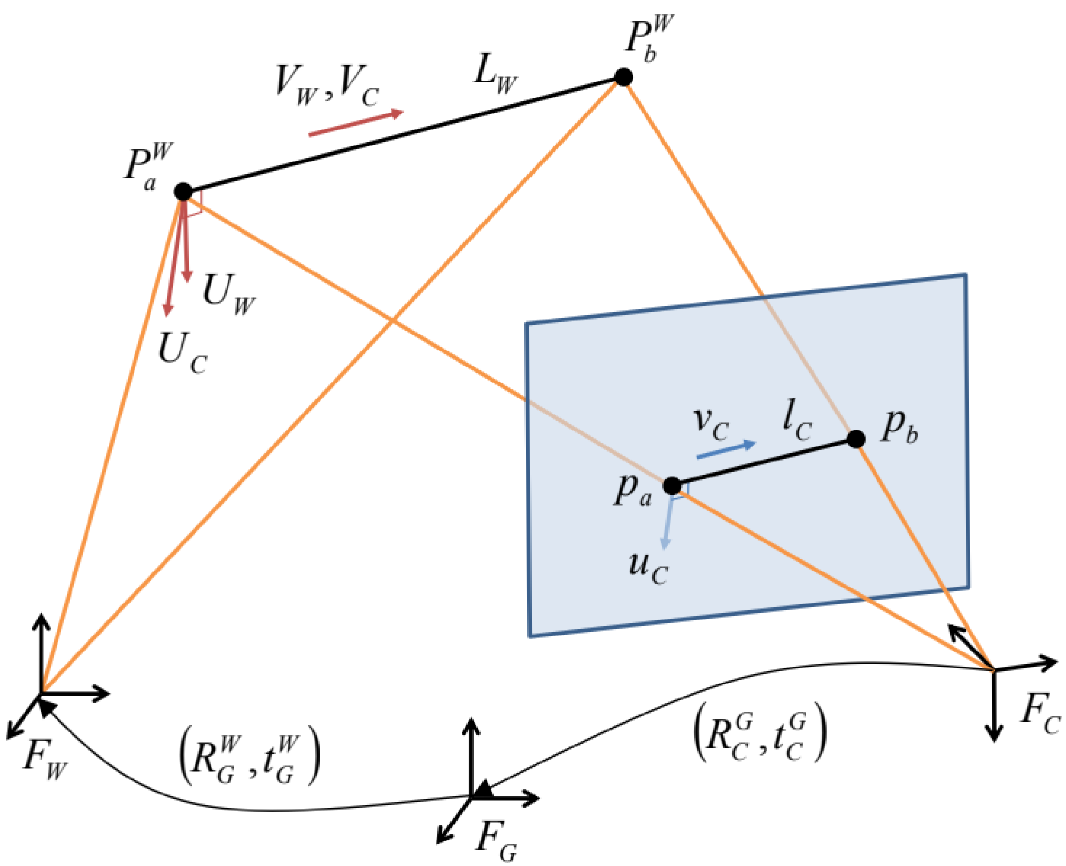


Image source: G. H. Lee, A Minimal Solution for Non-Perspective Pose Estimation from Line Correspondences, ECCV 2016.

Plücker Representation of 2D Lines

- We are also given the image coordinates of the end points $p_a = [p_{ax}\ p_{ay}\ 1]^\top$ and $p_b = [p_{bx}\ p_{by}\ 1]^\top$ from the 2D image line correspondence l_C of the 3D line.

- Similar to L_W, we can also represent l_C as a Plücker line $[u_C^\top\ v_C^\top]^\top$, where

$$v_C = \frac{\hat{p}_b - \hat{p}_a}{\|\hat{p}_b - \hat{p}_a\|}, \quad u_C = \hat{p}_a \times v_C.$$

- $\hat{p}_a = K^{-1} p_a$ and $\hat{p}_b = K^{-1} p_b$ are the camera matrix normalized image coordinates.

Solving for R_G^W

- The dot product of U and V of the Plücker line must be zero since they are perpendicular, hence we get:

$$U_C^\top R_W^C V_W = 0.$$

- Since we know that u_C is parallel to U_C , and $R_W^C = R_G^C R_W^G$, we can rewrite the dot product into:

$$u_C^\top R_G^C R_W^G V_W = 0,$$

- where the only unknown is R_W^G.

Solving for R_G^W

- The constraint can be rearranged into the form of a homogeneous linear equation $ar = 0$.
- a is 1×9 vector made of the known variables $u_C^\top$, R_G^C and V_W, and r is the 9×1 vectorized representation of the unknown R_G^W.
- To solve for r linearly, we need to have ≥ 8 2D-3D line correspondences to form

$$Ar = 0,$$

- which can be solved using the SVD method.

Solving for t_W^G

- Since u_C and U_C are parallel, we can write $\lambda u_C = U_C$ where λ is a is a scalar value.

- Substituting into $U_C = \left(R_W^C \ \lfloor t_W^C \rfloor_\times R_W^C\right) \begin{pmatrix} U_W \\ V_W \end{pmatrix}$, we get

$$\lambda u_C = \left(R_W^C \ \lfloor t_W^C \rfloor_\times R_W^C\right) \begin{pmatrix} U_W \\ V_W \end{pmatrix}.$$

- Taking the cross product of u_C on both sides to get rid of λ, we get:

$$\lfloor u_C \rfloor_\times \left(R_W^C \ \lfloor t_W^C \rfloor_\times R_W^C\right) \begin{pmatrix} U_W \\ V_W \end{pmatrix} = 0.$$

Solving for t_W^G

- Using the ≥ 8 2D-3D line correspondences we used to solve R_G^W, we get an overdetermined linear system:

$$Bt = 0$$

- Where B is made of the known variables R_G^C, t_G^C, R_W^G, u_C, U_W and V_W.

- $t = \begin{bmatrix} t_x \ t_y \ t_z \ 1 \end{bmatrix}^\top$ is a 4-vector made up of the three entries of the unknown translation vector t_G^W.

- We solve for t using the SVD method.

Special Cases

One Camera:

- The problem becomes the perspective pose estimation problem with line correspondences when all correspondences are seen by only one camera.
- Here the camera extrinsics (R_G^C, t_G^C) vanishes.
- And we directly solve for the camera orientation R_W^C without the need to decompose the orientation into $R_G^C R_W^G$.
- Similarly, we can solve for the camera translation t_W^C directly.

Special Cases

Parallel 3D Lines:

- Since the unit directions V_W are the same for parallel lines, the rank of matrix A drops below 8.
- Consequently, R_W^G cannot be solved.
- Fortunately, we can easily prevent this degenerate case by omitting parallel lines.

Summary

- We have looked at how to:

1. Use the Plücker line representation to derive the generalized epipolar geometry (GEC).

2. Apply the linear 17-point algorithm to obtain the relative pose of two-view generalized camera.

3. Explain the degenerate cases of the GEC, i.e. locally central projection, axial camera and locally-central-and-axial camera.

4. Compute the absolute pose of a generalized camera using 2D-3D point or line correspondences.