3D Vision

20 December, 2018

Ajad Chhatkuli[1] and Bishesh Khanal[2]
1 ETH Zürich, Switzerland; NAAMII, Nepal
2 King's/Imperial College London, UK; NAAMII, Nepal

Computer Vision – 3D Vision

- What can you tell about the 3D object from their images?

  - Pattern Recognition – image pixel colors

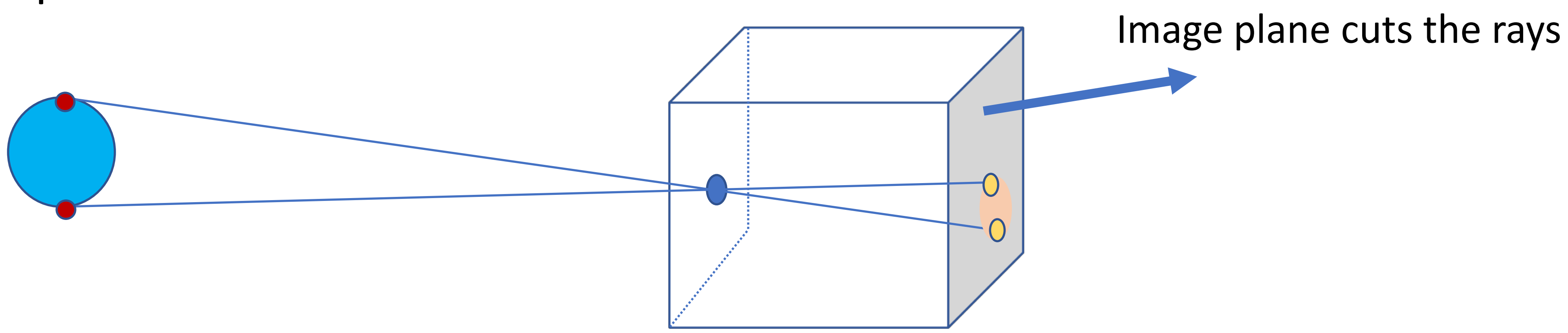
  - 3D vision – Pixel locations from projection

Main reference: Multiple View Geometry, [Hartley and Zisserman, 2004]

Pinhole Camera model

❑ Camera Projection

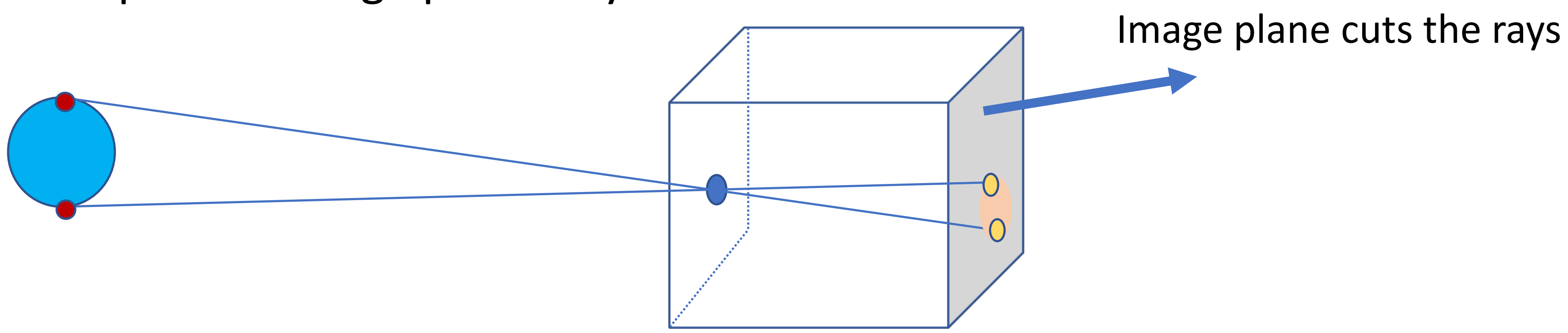
- The pin-hole camera model



The Image plane

❑ Camera Projection

- You can put the image plane anywhere

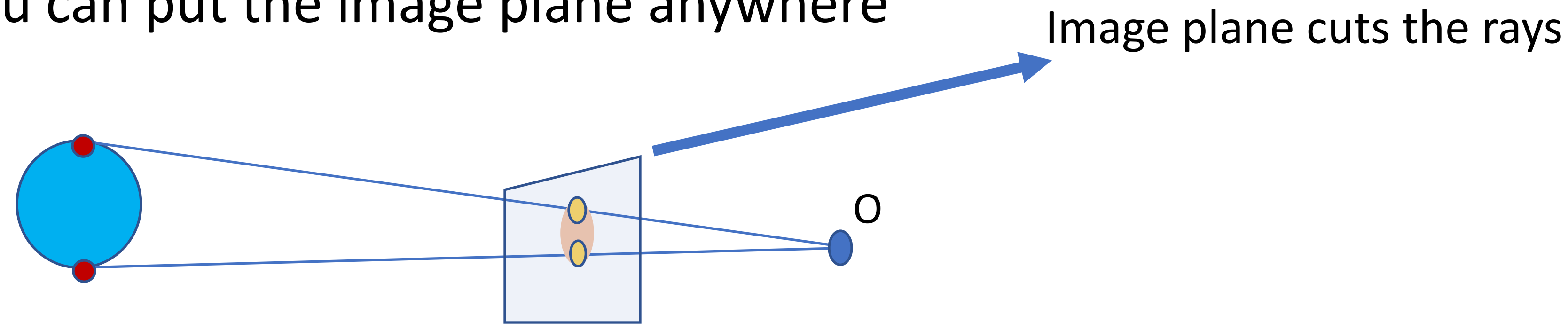


- Simpler to model if it assumed on the front at distance 1.

The Image plane

❑ Camera Projection

- You can put the image plane anywhere



- Simpler to model if it is assumed on the front at distance 1.

Any difference?

Camera axis

❑ Assume orientations

- Z – axis passes perpendicular to the image plane

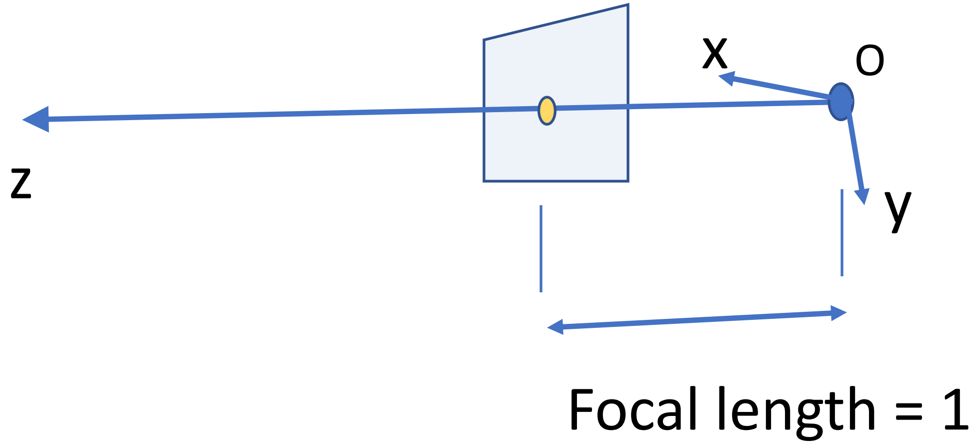
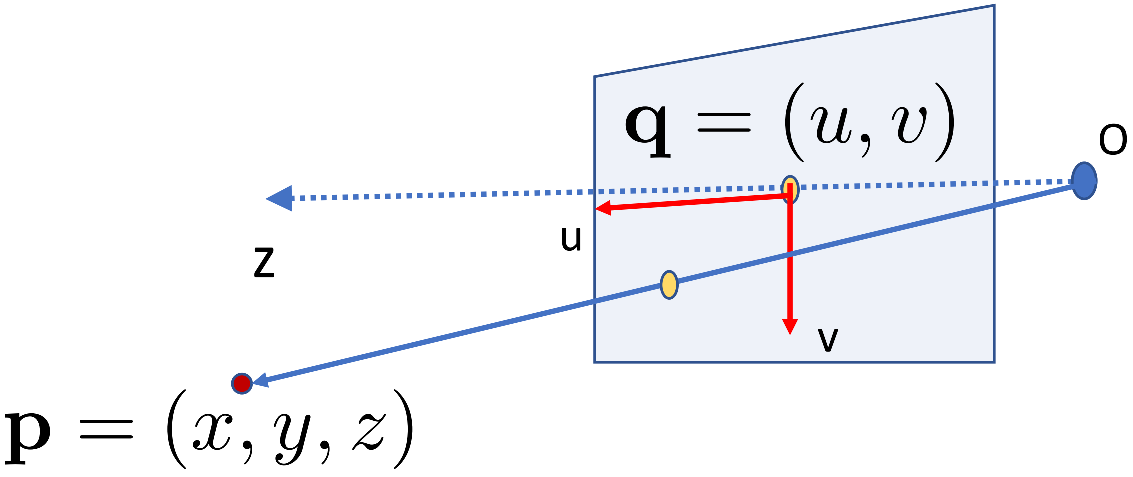


Image coordinates

❑ Projection equations

- Image plane coordinate system



$$u = \frac{x}{z}$$

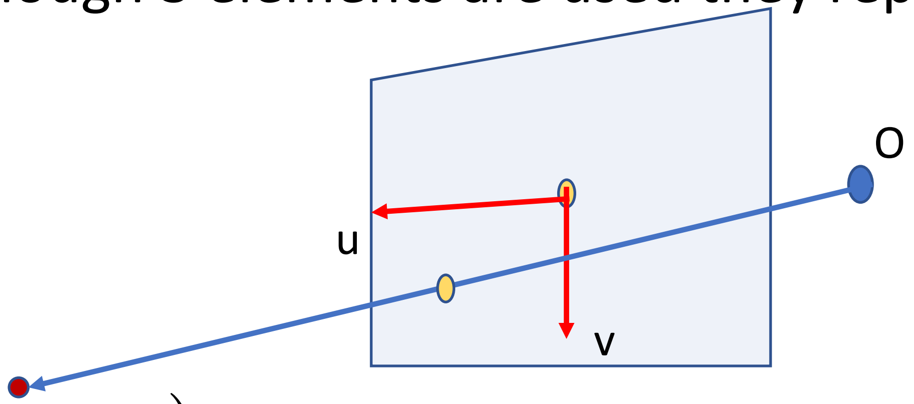
$$v = \frac{y}{z}$$

Can you prove them? – Exercise!

What kind of equations are they?

Image coordinates

- Linear form
  - Homogeneous coordinates
  - Although 3 elements are used they represent projected points in 2D!



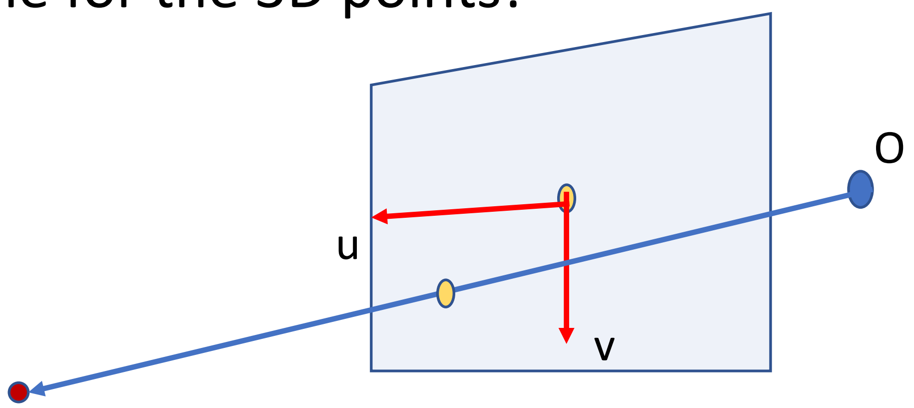
$$\mathbf{p} = (x, y, z)$$

$$\mathbf{q} = (u, v, 1) \in \mathbb{P}^2$$

Homogeneous coordinates!

Image coordinates

- Linear form
  - Homogeneous coordinates
  - Same for the 3D points!



$$\mathbf{p} = (x, y, z, 1)$$

$$\mathbf{q} = (u, v, 1) \in \mathbb{P}^2$$

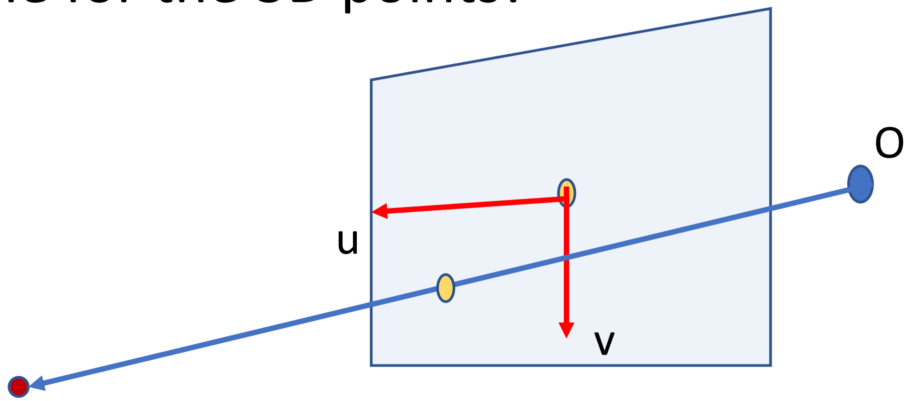
Homogeneous coordinates!

$$(u, v, 1) = (wu, wv, w)$$

Projection matrix

- Linear form
  - Homogeneous coordinates
  - Same for the 3D points!

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$



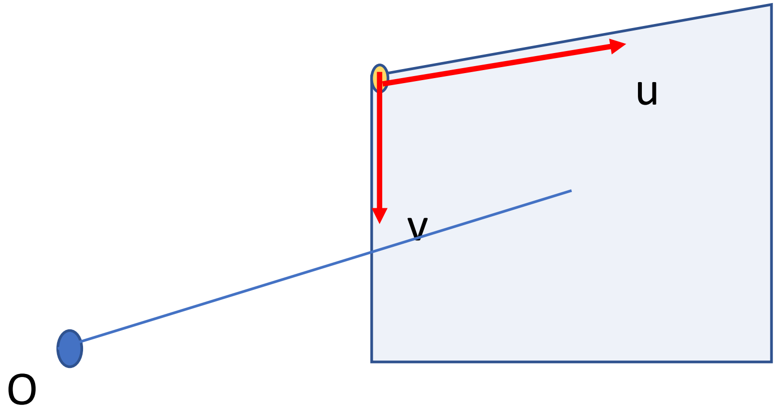
$$\mathbf{p} = (x, y, z, 1) \qquad \mathbf{q} = (u, v, 1) \in \mathbb{P}^2$$

Verify:

$$M\mathbf{p} = \mathbf{q}$$

Your cameras are different!

- Image origin is at the top left
  - This changes the projection equation!
  - Introduce the camera calibration matrix

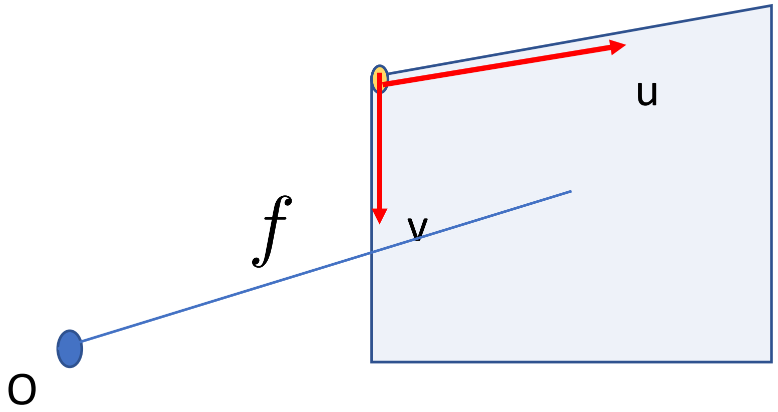


$$KM\mathbf{p} = \mathbf{q}'$$

$$K = \begin{bmatrix} 1 & 0 & c_1 \\ 0 & 1 & c_2 \\ 0 & 0 & 1 \end{bmatrix}$$

Your cameras are different!

- The focal length is much greater than 1 pixel.
  - Scale the image coordinates.



$$KM\mathbf{p} = \mathbf{q}'$$

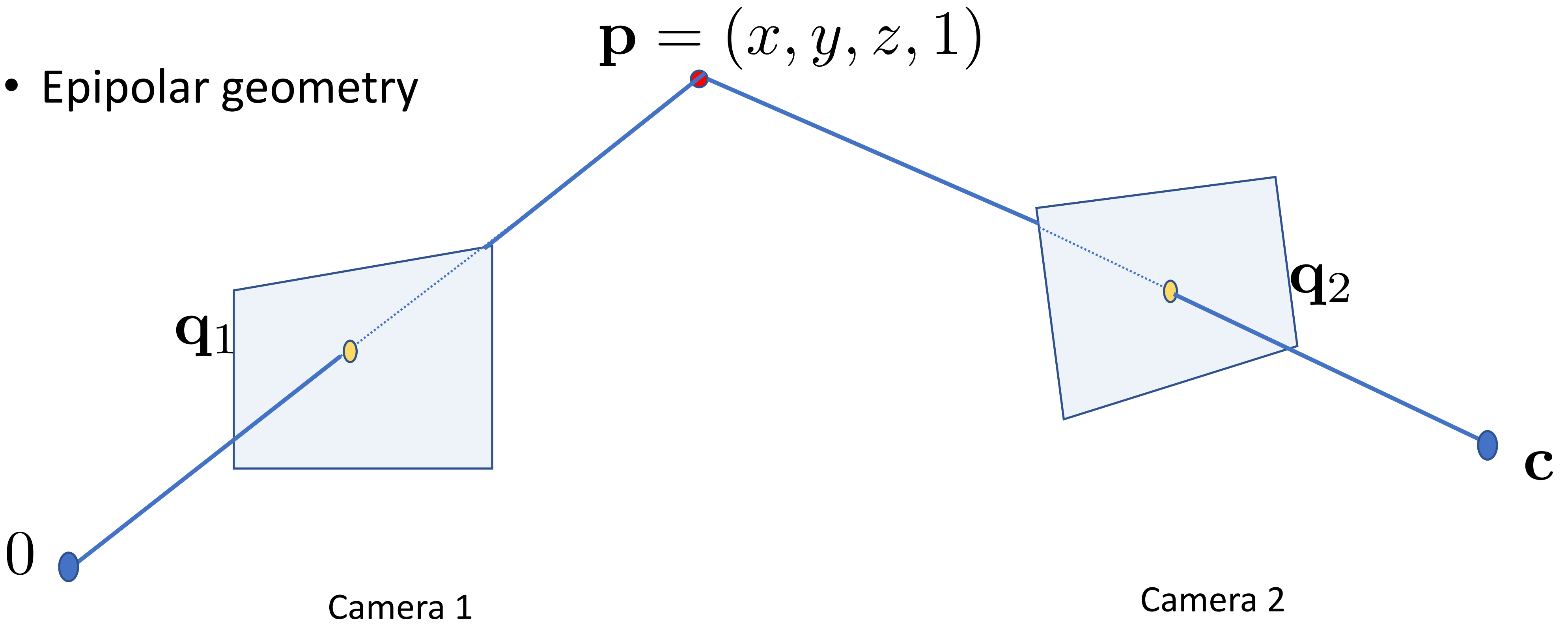
$$K = \begin{bmatrix} f & 0 & c_1 \\ 0 & f & c_2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{q} = K^{-1}\mathbf{q}'$$

Two-view geometry

❑ Two pinhole cameras

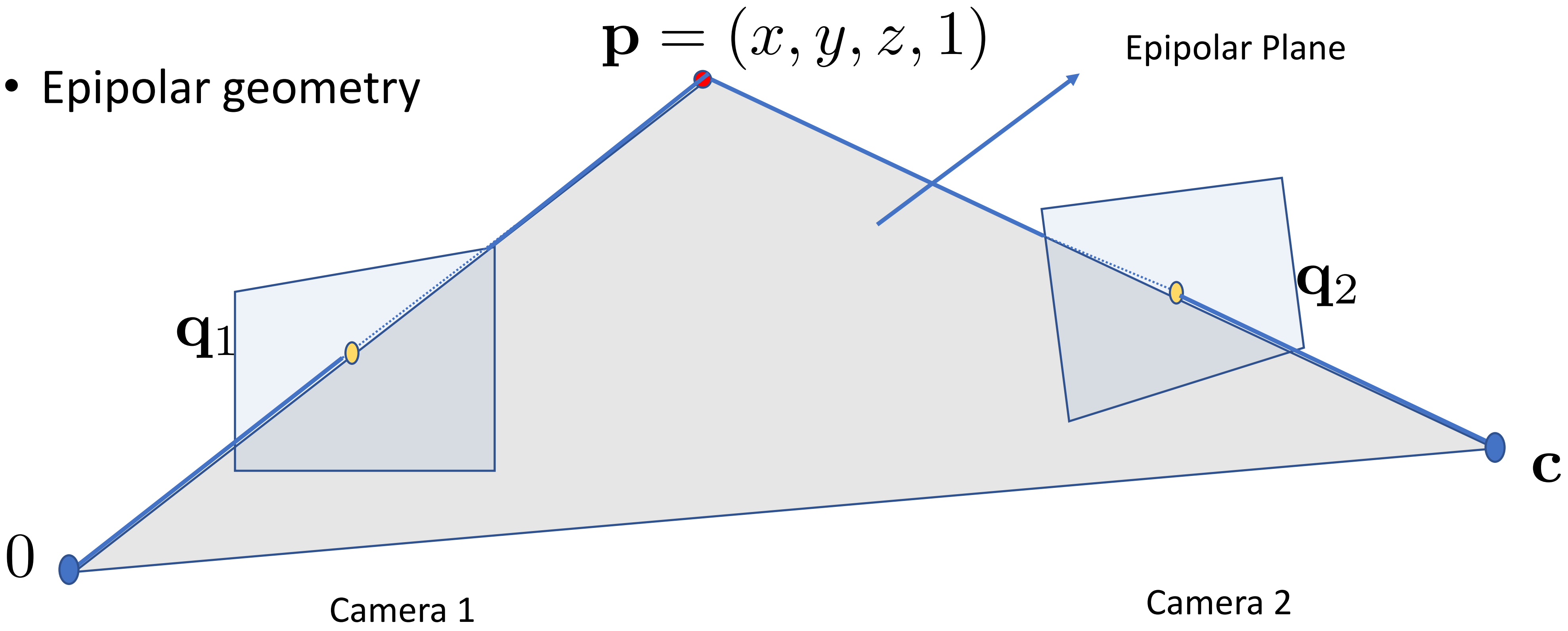
• Epipolar geometry



Two-view geometry

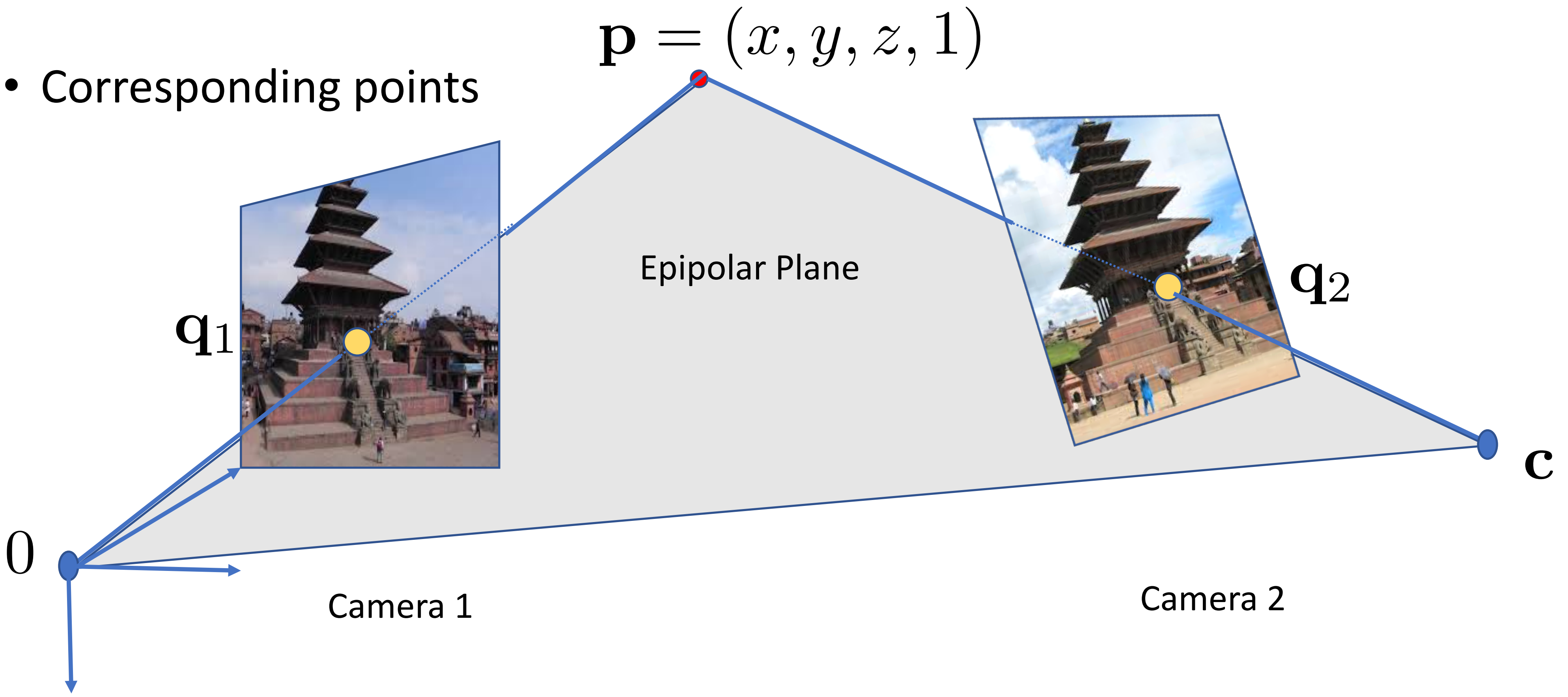
❑ Two pinhole cameras

- Epipolar geometry



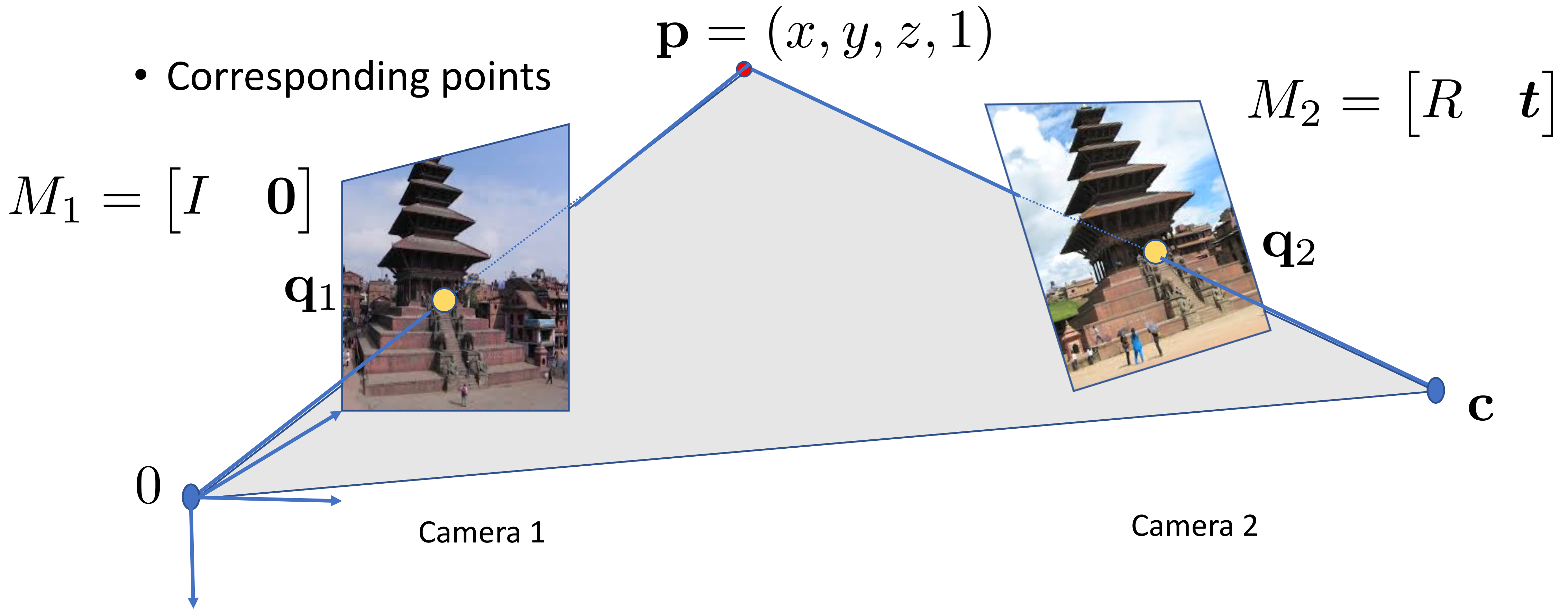
Two-view geometry

- ❑ Two pinhole cameras
  - Corresponding points



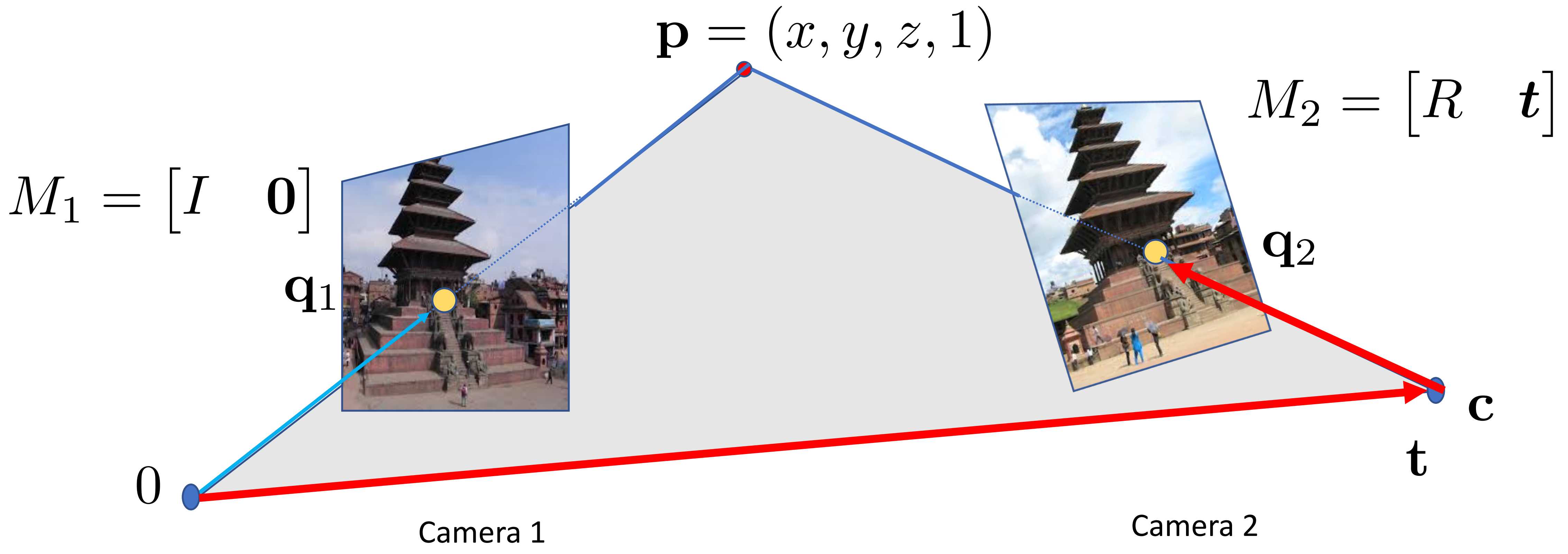
Two-view geometry

- ❑ Camera 2 has been rotated and translated.



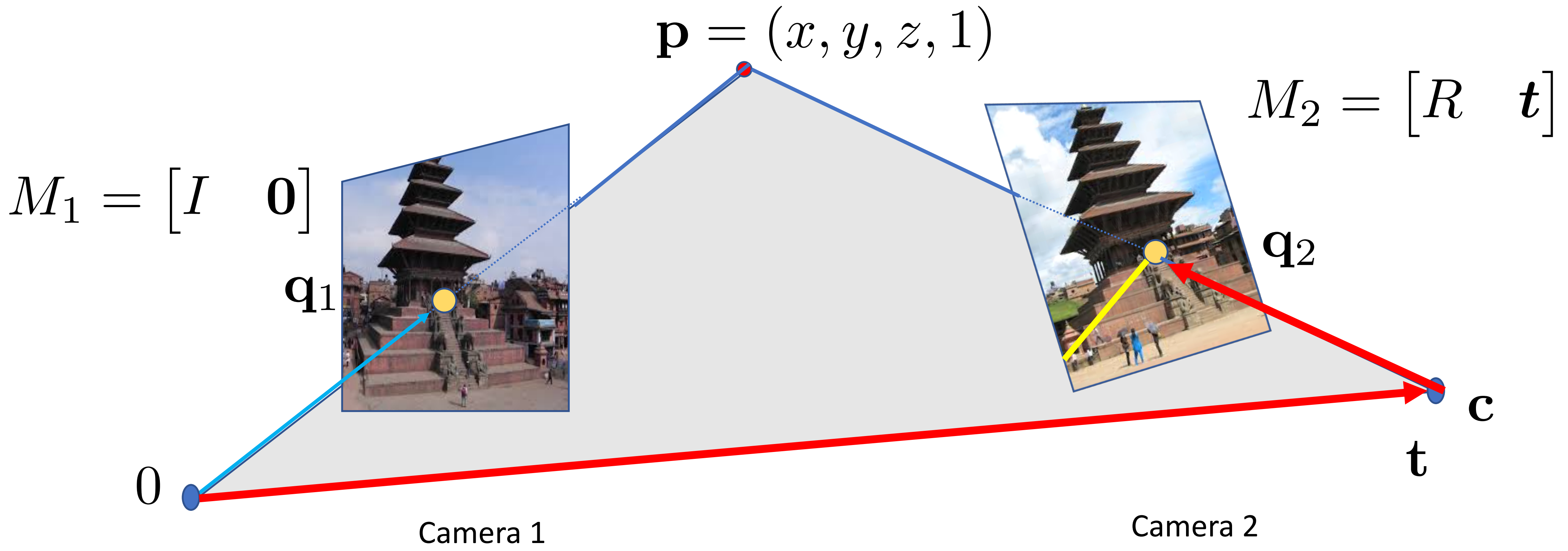
Two-view geometry

$$\mathbf{q}_2 \cdot (\mathbf{t} \times \mathbf{q}_2) = 0$$



Two-view geometry

$$\mathbf{q}_2 \cdot (\mathbf{t} \times R\mathbf{q}_1) = 0 \quad \Longrightarrow \quad \mathbf{q}_2^\top [\mathbf{t}]_\times R\mathbf{q}_1 = 0$$



The Essential Matrix

$$\mathbf{q}_2 \cdot (\mathbf{t} \times R\mathbf{q}_1) = 0 \quad \Longrightarrow \quad \mathbf{q}_2^\top [\mathbf{t}]_\times R\mathbf{q}_1 = 0$$

$$E = [t]_\times R$$

❑ Essential matrix is:

- Rank 2 matrix
- 2 equal non-zero singular values
- What about the last singular value?

The Essential Matrix

- True for all corresponding points:

$$\mathbf{q}_2^\top E \mathbf{q}_1 = 0$$

- Mathematically,

$$\mathrm{rank}(E) = 2, \quad \det(E) = 0$$

$$2EE^\top E - \mathrm{trace}(EE^T)E = 0.$$

The Fundamental matrix

For cameras with different focal length and principal point

$$\mathbf{q} = K^{-1}\mathbf{q}'$$

$$(K^{-1}\mathbf{q}_2')^\top [\mathbf{t}]_\times R(K^{-1}\mathbf{q}_1') = 0$$

$$\mathbf{q}'^\top_2 (K^{-\top}[\mathbf{t}]_\times RK^{-1})\mathbf{q}'_1 = 0$$

$$\mathbf{q}'^\top_2 F\mathbf{q}'_1 = 0$$

Too many equations!!

The Fundamental matrix

- For all corresponding points:

$$\mathbf{q'}_2^\top F \mathbf{q'}_1 = 0$$

- Properties:

$$\text{rank}(F) = 2, \quad \det(F) = 0$$

- Lets discuss!! What does it mean by the rank being 2?

The 8-point Method

- Given many correspondences, compute F or E

$$A \qquad\qquad\qquad \mathbf{f} = 0$$

$$\begin{bmatrix} u_2u_1 & u_2 & v_1 & u_2 & v_2u_1 & v_2v_1 & v_2 & u_1v_1 & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0 \qquad \|\mathbf{f}\| = 1$$

Requires at least 8 points to solve!

The 8-point Method

- Given many correspondences, compute F or E

$$\begin{bmatrix} u_2 & v_2 & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0$$

$$u_2 u_1 f_{11} + u_2 v_1 f_{12} + u_2 f_{13} + v_2 u_1 f_{21} + v_2 v_1 f_{22} + v_2 f_{23} + u_1 f_{31} + v_1 f_{32} + f_{33} = 0$$

The 8-point Method

- Given many correspondences, compute F or E

$$u_2u_1f_{11} + u_2v_1f_{12} + u_2f_{13} + v_2u_1f_{21} + v_2v_1f_{22} + v_2f_{23} + u_1f_{31} + v_1f_{32} + f_{33} = 0$$

$$\begin{bmatrix} u_2u_1 & u_2v_1 & u_2 & v_2u_1 & v_2v_1 & v_2 & u_1 & v_1 & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

The 5-point Method

- Can you do better for the Essential matrix?

$$A\mathbf{f} = 0, \quad \|\mathbf{f}\| = 1$$

- Matrix constraint

$$2EE^\top E - \mathrm{trace}(EE^T)E = 0.$$

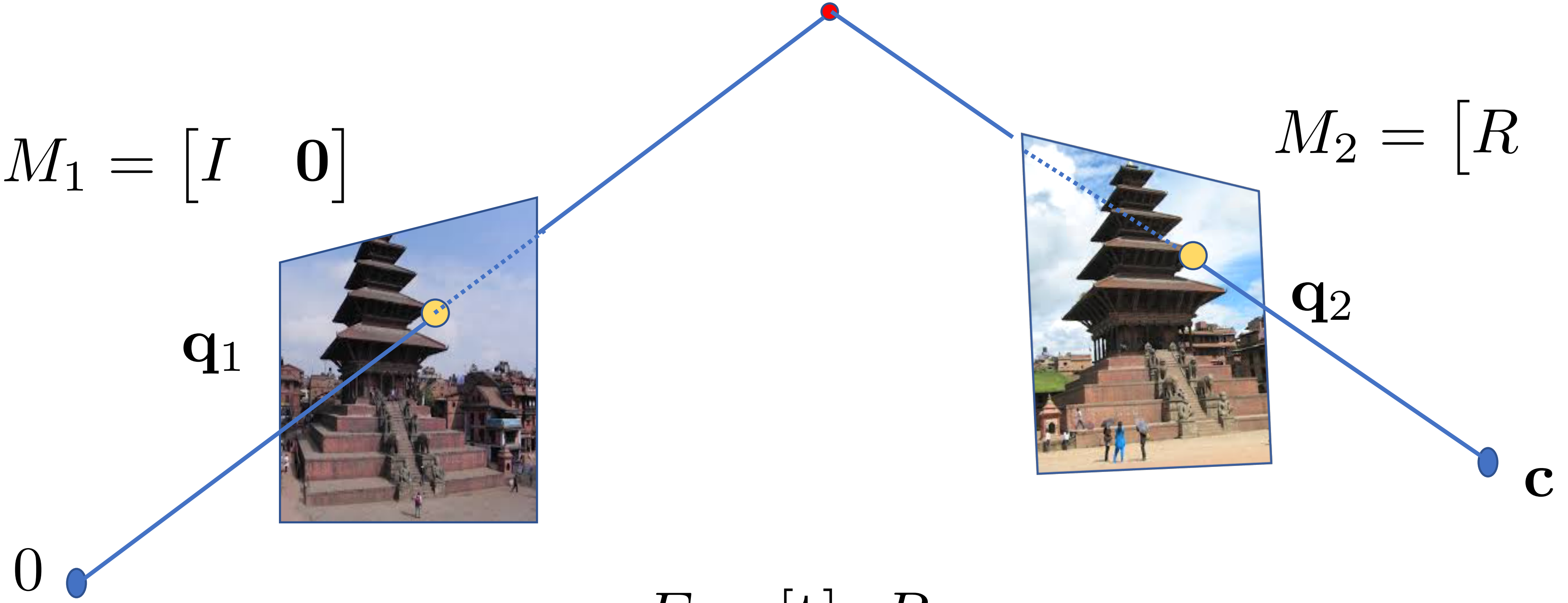
Triangulation

- Obtain 3D from 2D images

$$\mathbf{p} = (x, y, z, 1)$$

$$M_1 = [I \quad \mathbf{0}]$$

$$M_2 = [R \quad \boldsymbol{t}]$$



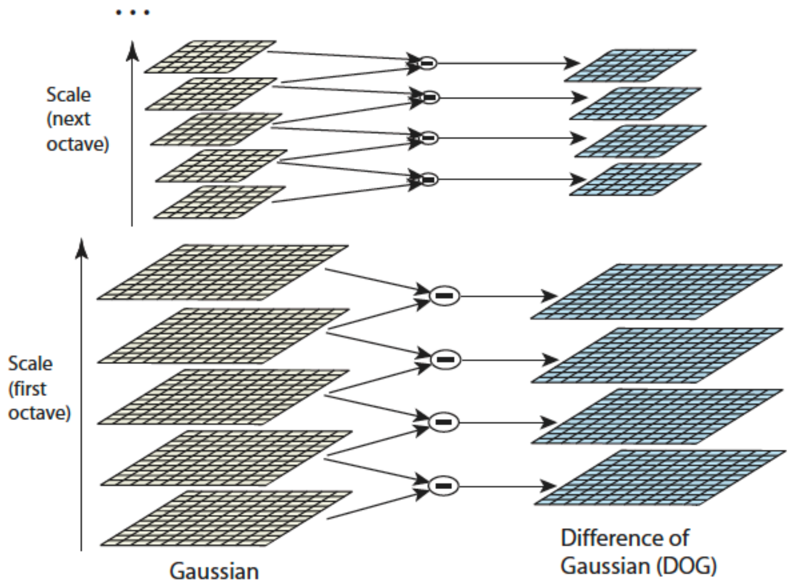
$$E = [\mathbf{t}]_\times R$$

Corresponding points

- Find “keypoints” in images
  - Scale Invariant Feature Transform (SIFT) – [Lowe, 2004]

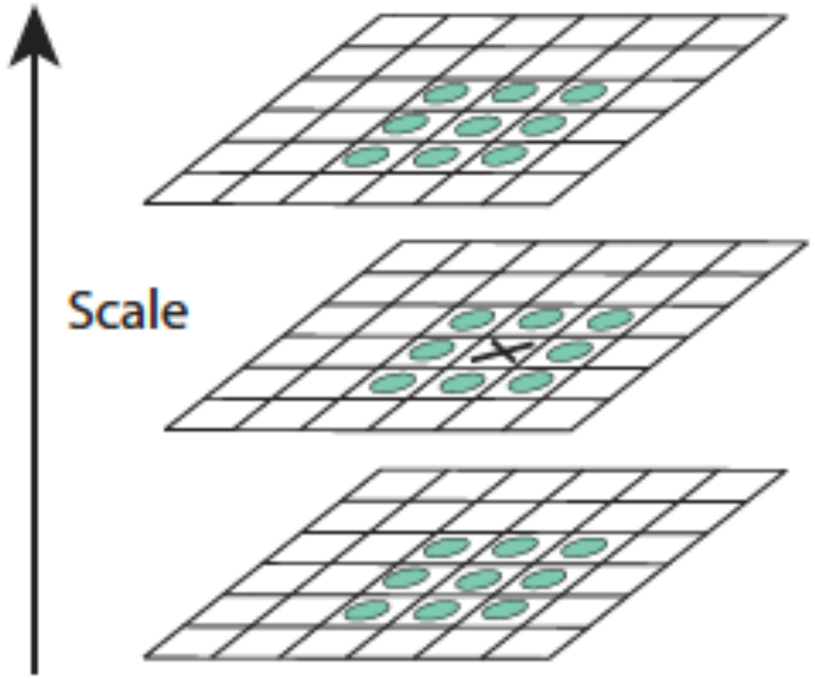
SIFT

- ❑ Scale Invariant Feature Transform [Lowe 2004]



SIFT

- Scale Invariant Feature Transform



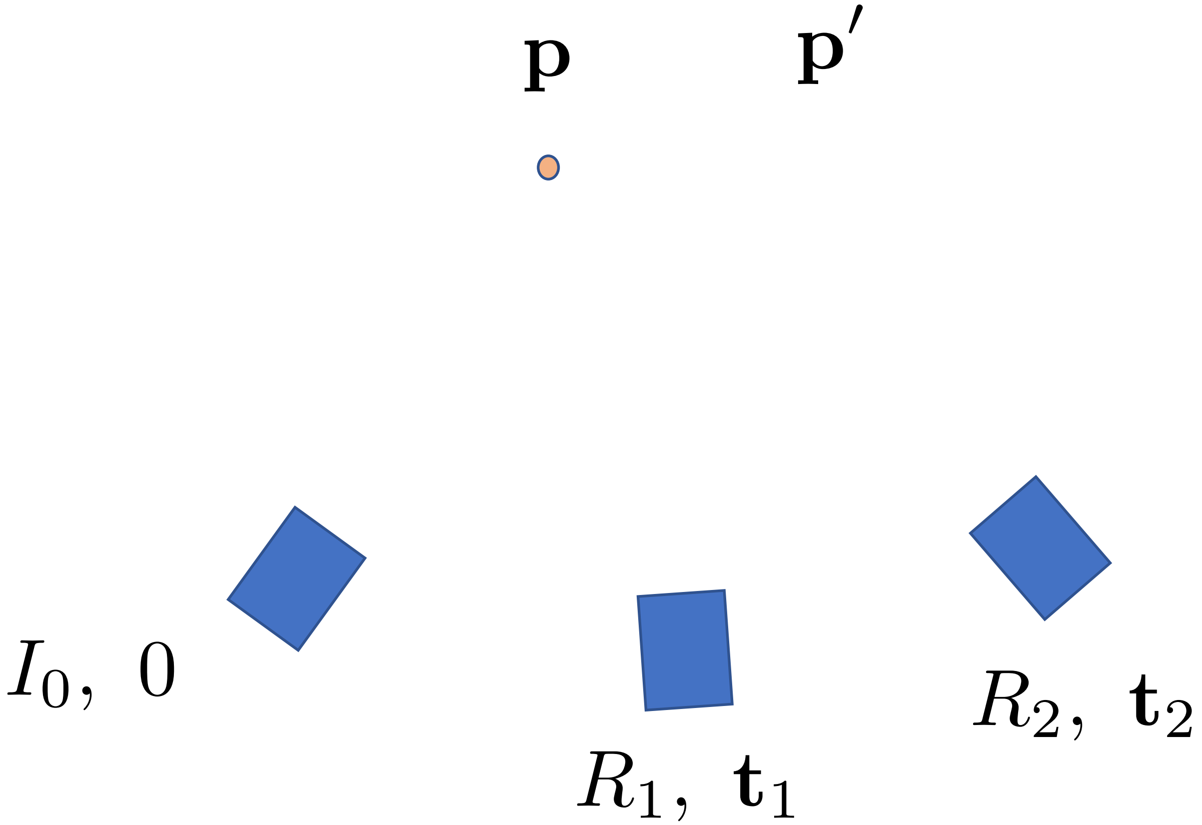
Structure-from-Motion

❑ Multi-view triangulation

- Triangulate with two views
- Transfer points to common reference
- Minimize triangulation error – Bundle adjustment
- Sparse 3D points
- Compute dense 3D points

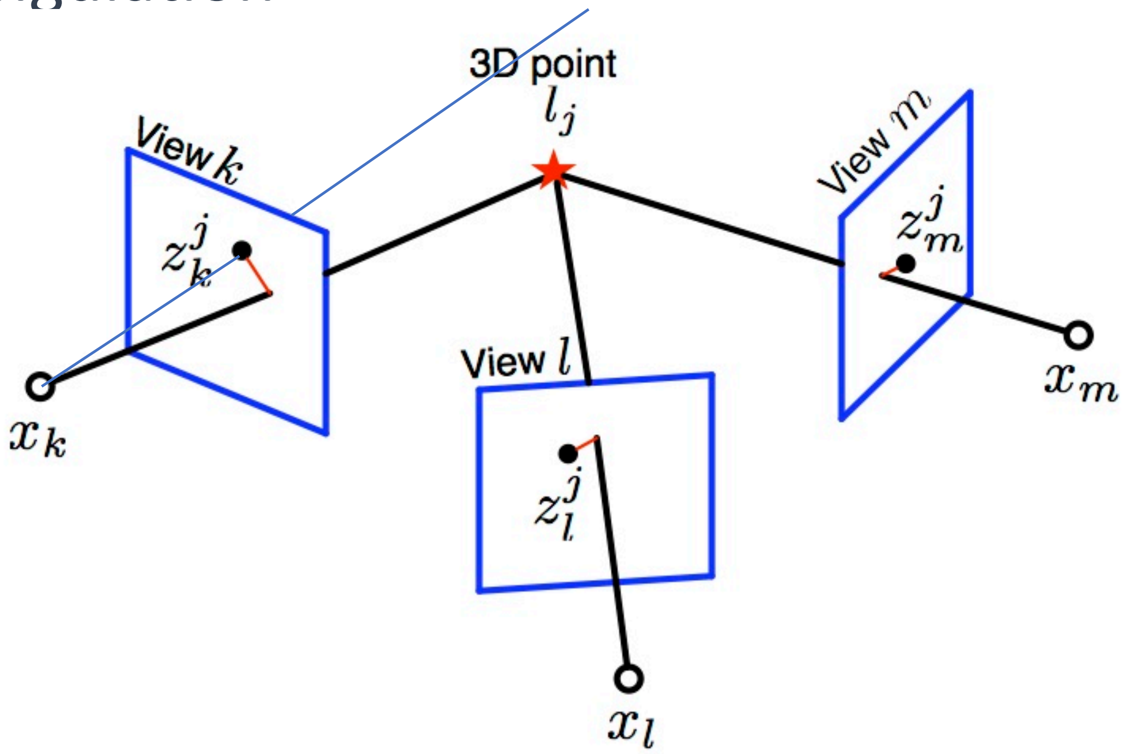
Structure-from-Motion

- Pose computation



Structure-from-Motion

- ❑ Multi-view triangulation



Source: Vadim Indelman, ANPL