Lecture 3: CS677

Aug 29, 2017

Review

- Previous class
  - Some problems of vision
  - State-of-art examples
  - Evolution of eyes
  - Pin-hole camera model
- Today’s objective
  - Derivation of projection equations
  - Homogeneous coordinates
  - Coordinate transformations

Cloud Computing

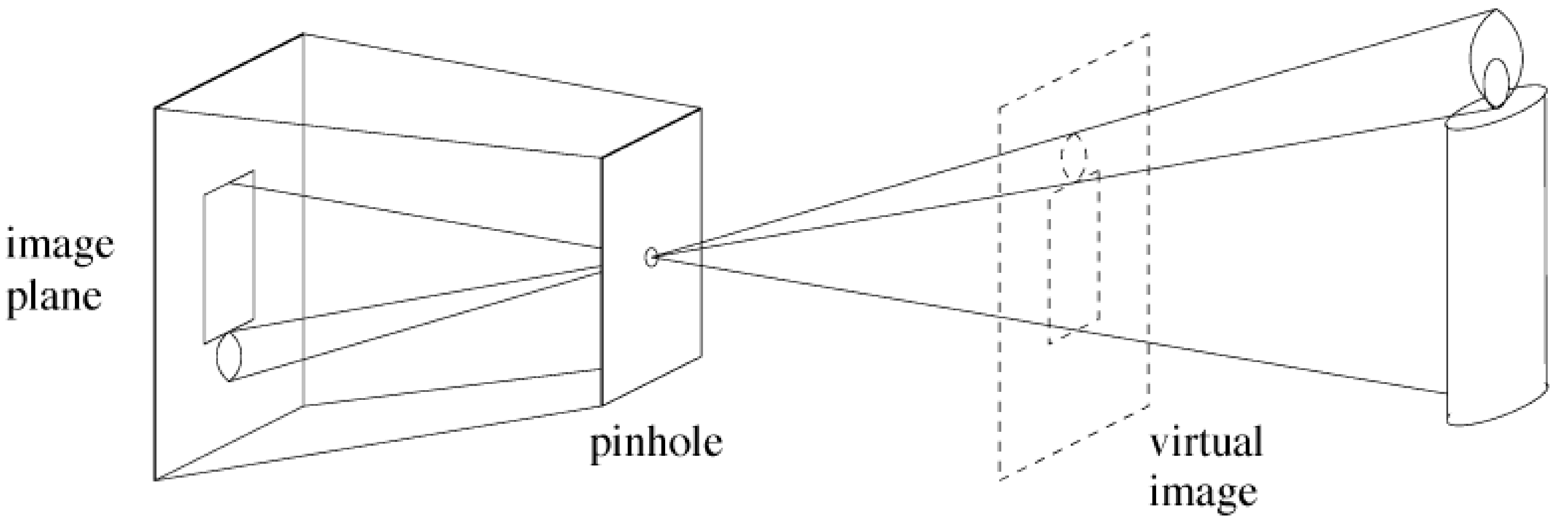
- Google has provided limited cloud computing account for class (worth $50). Students can get a better persona account at: https://console.cloud.google.com/freetrial?_ga=2.228461851.-722665125.1503520492&page=1
- Will ask for credit card but see note below.
- **Access to all Cloud Platform Products**
- Get everything you need to build and run your apps, websites, and services, including Firebase and the Google Maps API.
- **$300 credit for free**
- Sign up and get $300 to spend on Google Cloud Platform over the next 12 months.
- **No autocharge after free trial ends**
- We ask you for your credit card to make sure you are not a robot. You won't be charged unless you manually upgrade to a paid account.

Image Formation

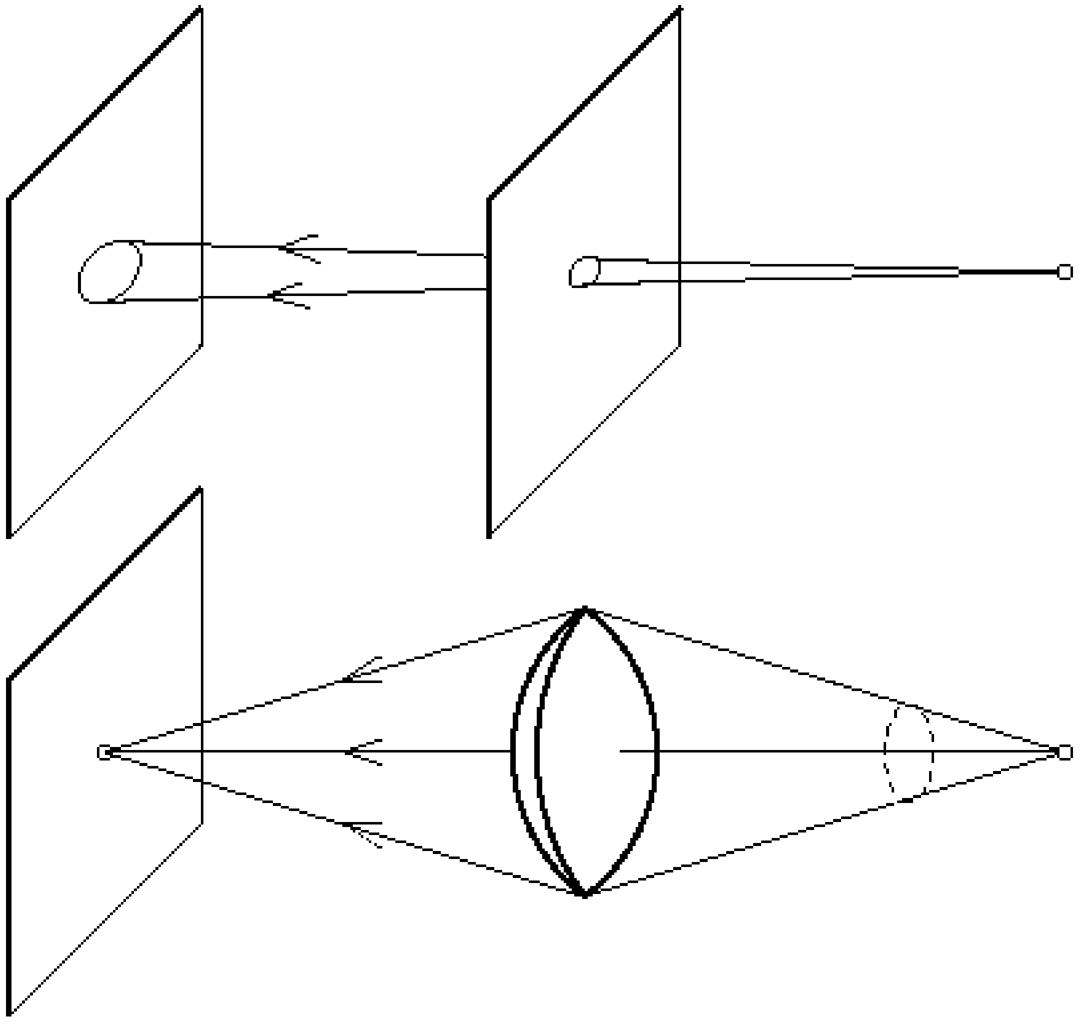
- Geometry
  - Where is the image of a point formed?
- Photometry/Colorimetry
  - How bright is the point?
  - What is its *color*?
- Ideal camera models
- Real lenses

Pinhole cameras

- Abstract camera model - box with a small hole in it
- Note inverted image
- Pinhole cameras work in practice, ignoring diffraction

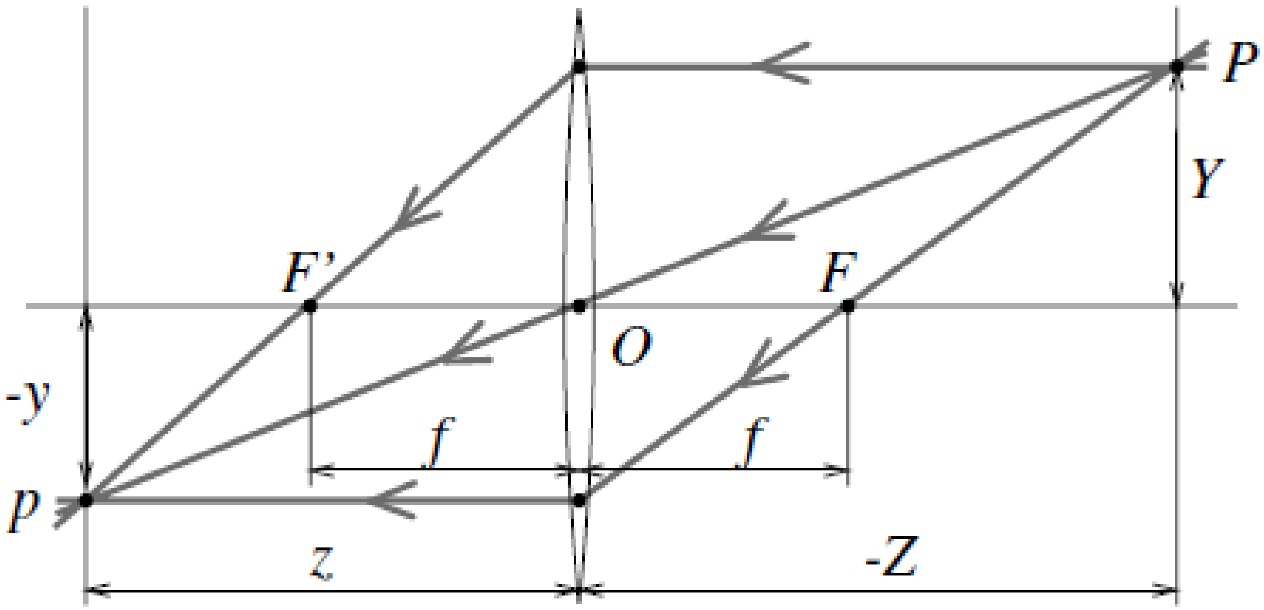


The reason for lenses



Sharper focus,
more light

The thin lens



$$\frac{1}{z} - \frac{1}{Z} = \frac{1}{f}$$

Thin Lens Properties

- Points at different depth focus at different positions of the image plane
  - With a fixed image plane, not all points will be in focus
  - "Depth of field", *i.e.* distance over which focus is acceptable depends on the *aperture* size
  - Larger aperture captures more light but has lower DOF
  - Defocus property can be used to infer depth
    - Limited accuracy
- Field of view: depends on imaging surface size, not lens aperture size

Field of View (FoV)

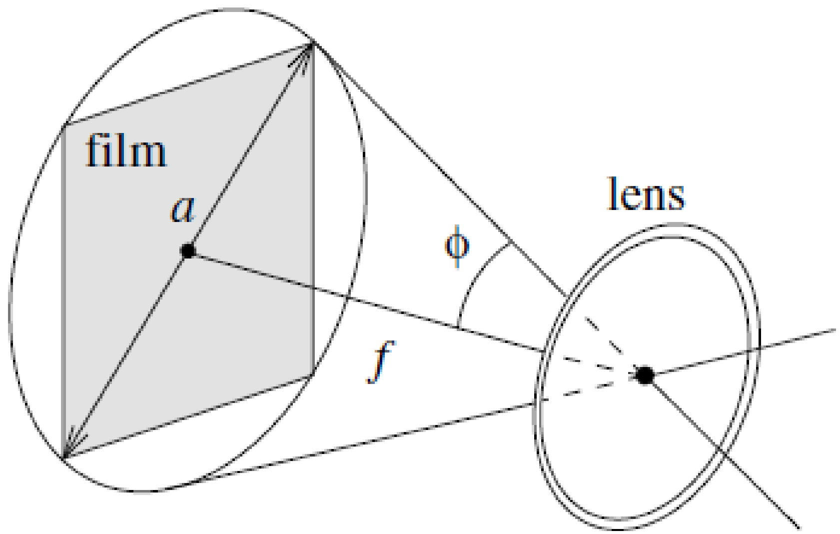
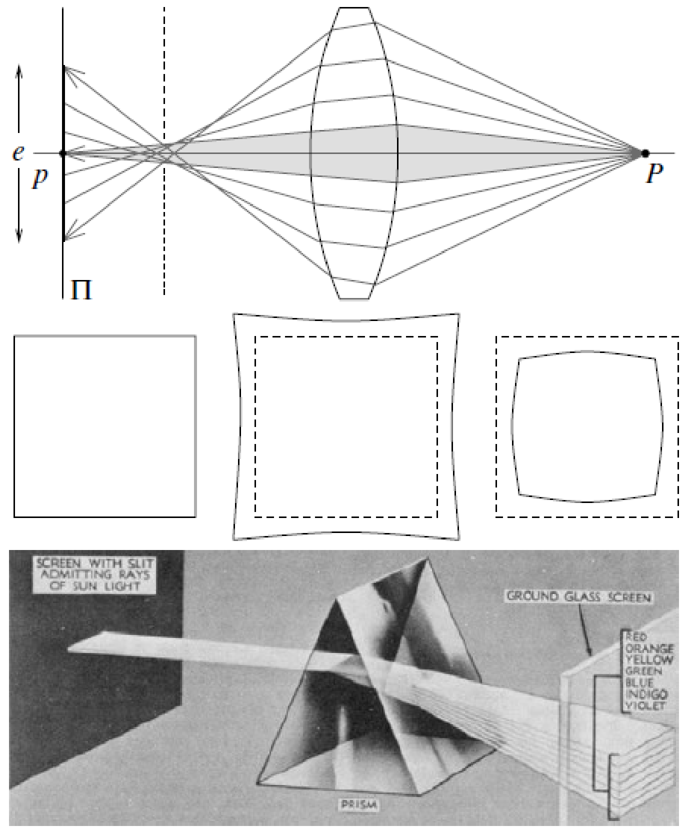


FIGURE 1.9: The field of view of a camera. It can be defined as 2ϕ, where $\phi \stackrel{\text{def}}{=} \arctan \frac{a}{2f}$, a is the diameter of the sensor (film, CCD, or CMOS chip), and f is the focal length of the camera.

Lens Distortions

- Real lenses suffer from various errors/distortions
- Chromatic aberration (not all wavelengths focus at the same point)
- Geometric distortions: complex lens systems used to reduce distortion
- Usually we will assume that complex lenses behave as ideal pinhole models but without the negative effects
  - No diffraction effects, sufficient light collection, all points in focus

Distortion Illustrations



(a)

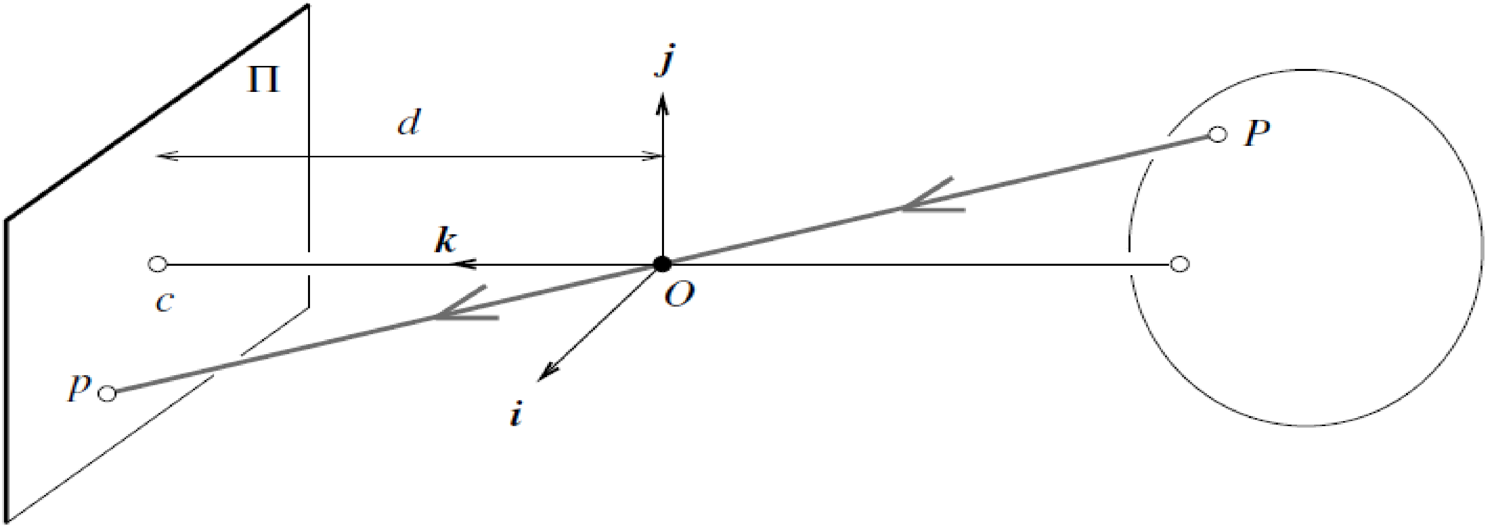
(b)

(c)

FIGURE 1.11: Aberrations. (a) Spherical aberration: The gray region is the paraxial zone where the rays issued from P intersect at its paraxial image p. If an image plane π were erected in p, the image of p in that plane would form a circle of confusion of diameter e. The focus plane yielding the circle of least confusion is indicated by a dashed line. (b) Distortion: From left to right, the nominal image of a fronto-parallel square, pincushion distortion, and barrel distortion. (c) Chromatic aberration: The index of refraction of a transparent medium depends on the wavelength (or color) of the incident light rays. Here, a prism decomposes white light into a palette of colors. *Figure from US NAVY MANUAL OF BASIC OPTICS AND OPTICAL INSTRUMENTS, prepared by the Bureau of Naval Personnel, reprinted by Dover Publications, Inc. (1969).*

The equation of projection

- Note: *k*-axis *towards* the camera (right handed coordinate system $\boldsymbol{k} = \boldsymbol{i} \times j$).



Let P = (X,Y,Z), p = (x, y, z)

- We know that z= d, find values of x and y
- Op = λ.OP for some λ, λ = d/Z

hence:

$$\begin{cases} x = d\dfrac{X}{Z}, \\ y = d\dfrac{Y}{Z}. \end{cases}$$

Comments on projection equation

- Note: if X is a positive number , *x* will be negative since Z is negative
- If image plane is in front (virtual plane), image is not inverted; change signs of *x* and *y* .
- Some authors (*e.g.* RS book) assume that the *z*-axis points towards the object; change signs to accommodate
- How to compute image of a curve?
  - Project points along the curve
    - How many points to sample?
  - Analytical equations may be possible in some cases if the original curve has an analytical equation
- How to project a surface?
  - All points on the surface? All points may not be visible.

Projections of Certain Shapes

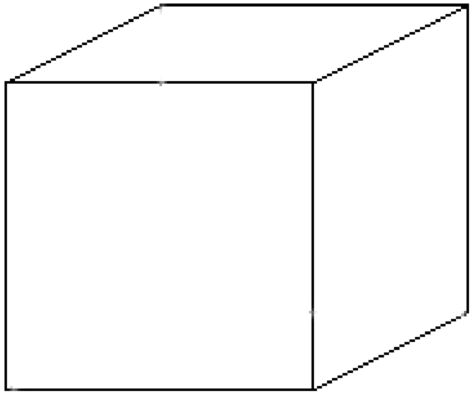
- Projection of a straight line
  - Straight line
  - How to show/prove? Geometrically? Algebraically?
- Projection of a circle?
  - A conic section
  - How to show prove? Geometrically? Algebraically?
- Image of a sphere
  - A conic?
- Images of a set of parallel lines?
  - Do images remain parallel?

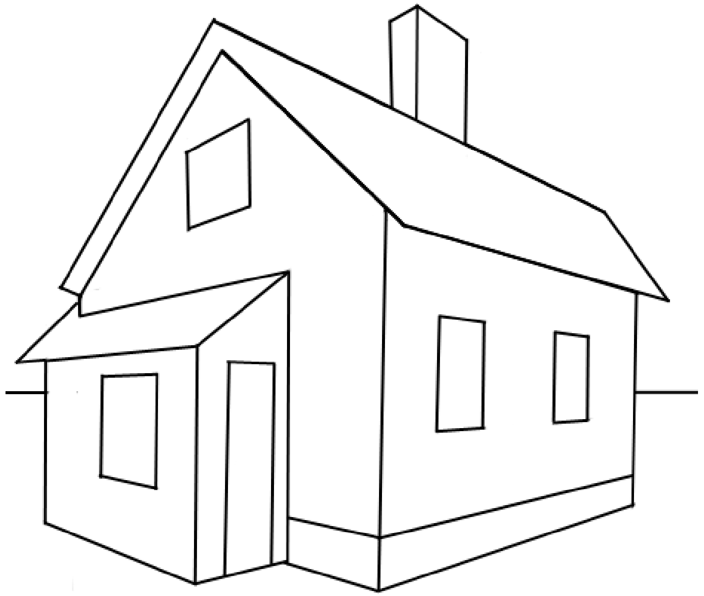
Converging Lines

Back Projection

- Given an image of an object, what can we infer about the 3-D object casting the image?
- Given a single 2-D image point?
  - A line (orientation) along which the 3-D point must lie, but we can not fix a unique distance
- Given a straight line in the image?
  - Must the object also be a straight line?
    - Not necessarily, but likely (except for accidental viewpoints)
  - Constraints on the object line?
    - Must line in a specific plane (given by pinhole or lens center and the image line)
- Back projection of an ellipse
  - Another ellipse; if we assume it is projection of a circle, we can estimate orientation of the plane
- Is back projection of more complex shapes more constrained?

How do we see Depth in Simple Drawings?





From:
http://www.drawinghowtodraw.com/stepbystepdrawinglessons/2014/01/how-to-draw-a-house-with-easy-2-point-perspective-techniques/

- What assumptions do we make?
- 2-D properties are not accidental: parallel lines in image also parallel in 3-D; intersections are real; symmetry/simplicity of objects…
- Significant theories developed but apply only to very clean drawings as shown here; not topic of serious study at this time.
- Will color, intensity help? We will address this a bit later.

Multiple Cameras

- Each camera specifies a line on which the 3-D point must lie
- Point must be at intersection of these rays
- Issues: How to find the corresponding points? What if camera relative positions are not known?

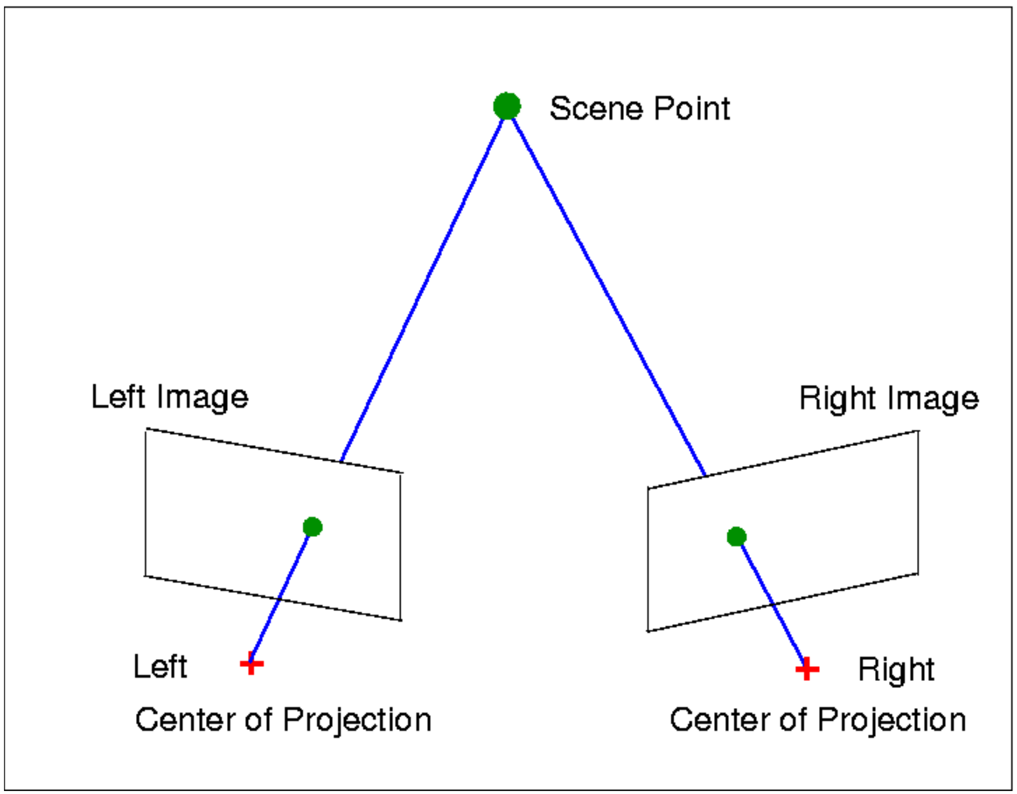


Figure from:
http://www.eng.tau.ac.il/~nk/computer-vision/stereo/index.html

Homogeneous Coordinates

- Add an extra coordinate
  - $(X,Y,Z) \Rightarrow (X_h, Y_h, Z_h, w) = (wX, wY, wZ, w)$, w is any constant (in the FP book, w is usually set to 1)
- Advantage: allows perspective transformation to be *linearized*, *i.e.* expressed as a matrix equation

$$\begin{bmatrix} x_h \\ y_h \\ w_h \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/d & 0 \end{bmatrix} \begin{bmatrix} X_h \\ Y_h \\ Z_h \\ w \end{bmatrix}$$

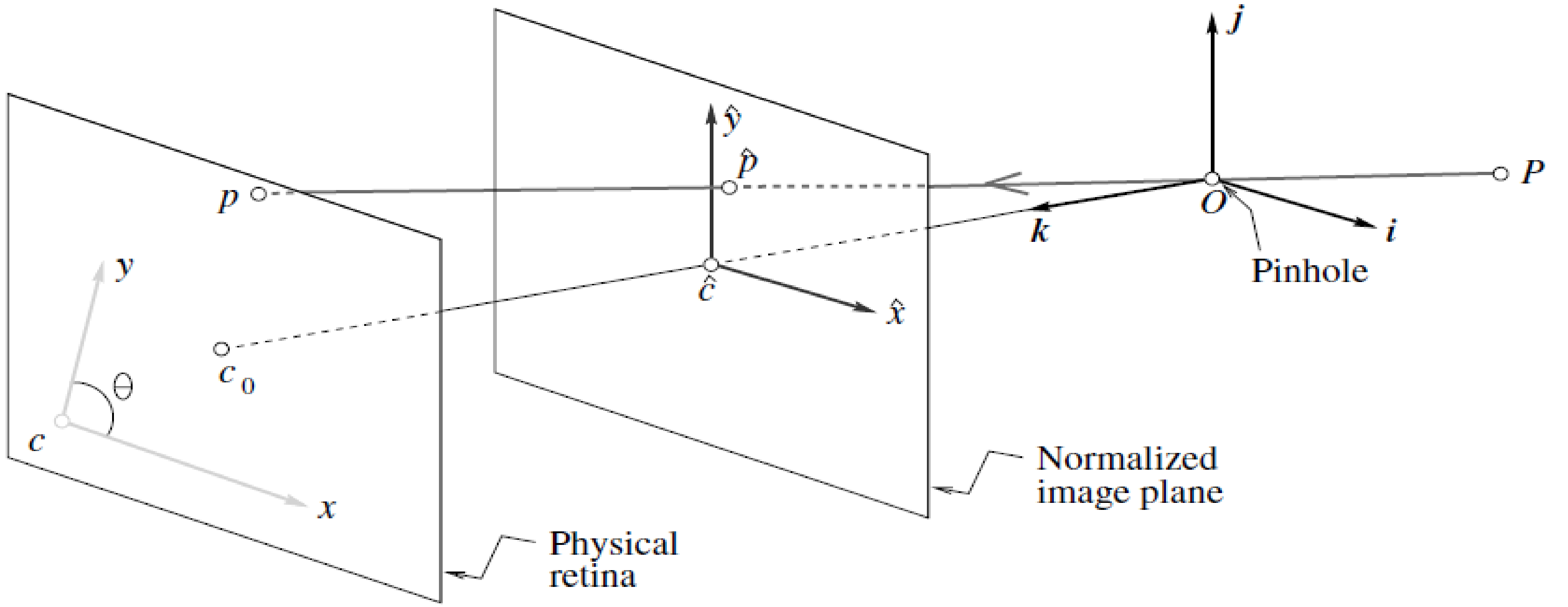
$x_h = X_h$, $y_h = Y_h$, $w_h = 1/d * Z_h$

$x = x_h / w_h = d*X_h / Z_h = d * X/Z$, $y = d * Y/Z$

Also common to represent focal length by variable f; also to write matrix as $\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

Intrinsic Camera Parameters

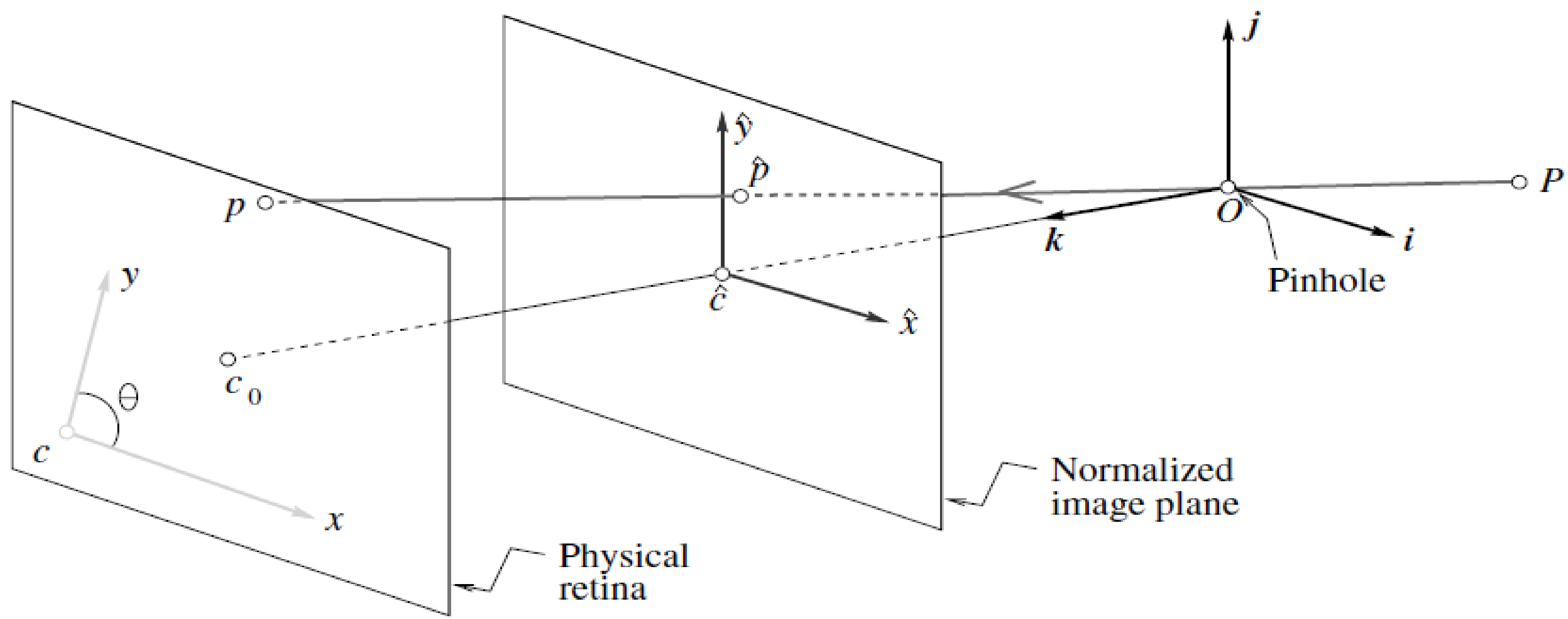
- Figure 1.14



- Measurement in image coordinate system may be in "pixel" units (*x*,*y*), pixels may not be rectangular, origin of image coordinate system may not be at the center of *image* (projection of lens center), axis may be *skewed* .
- *Normalized* image plane: parallel to physical retina but unit distance from lens center

Normalized Coordinates

- Figure 1.14



- *Normalized* image plane: parallel to physical retina but unit distance from lens center
- *Normalized coordinates*:
  - Origin at the intersection of normalized plane and the principal ray
  - Image plane axes parallel to the *i* and *j* axes

Projection in Normalized Coordinates

- In normalized coordinate system:

$$\begin{cases} \hat{x} = \dfrac{X}{Z} \\ \hat{y} = \dfrac{Y}{Z} \end{cases} \Longleftrightarrow \hat{p} = \frac{1}{Z}\begin{pmatrix} \mathrm{Id} & 0 \end{pmatrix} P$$

- Both $\hat{p}$ and P are expressed in homogeneous coordinates with the last term being set to "1"
- If we let the last term be "w", we would not need to carry 1/Z in our equations (it would come from the homogeneous representation) but we will follow book's notation.

Intrinsic Parameters

- We can go from normalized coordinates to actual camera coordinates by a series of transformations.
- Let *f* be focal length, *k* and *l* be scale parameters along *x* and *y* directions

$$x = kf\frac{X}{Z} = kf\hat{x},$$
$$y = lf\frac{Y}{Z} = lf\hat{y}.$$

- Image coordinates commonly expressed not in meters but in pixel units; *k* and *l* take care of this unit transformation. Let $\alpha = kf$, $\beta = lf$.
- Image center need not be at (0,0), let it be at c_0. Now,

$$x = \alpha\hat{x} + x_0,$$
$$y = \beta\hat{y} + y_0.$$

Intrinsic Parameters

- Let θ be the angle between axes in image plane, then

$$x = \alpha \hat{x} - \alpha \cot\theta \hat{y} + x_0,$$
$$y = \frac{\beta}{\sin\theta}\hat{y} + y_0.$$

- In matrix form:

$$\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}, \quad \text{where} \quad \boldsymbol{p} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad \text{and} \quad \mathcal{K} \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & -\alpha\cot\theta & x_0 \\ 0 & \dfrac{\beta}{\sin\theta} & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

- *K* is called the *internal calibration matrix*; ($\alpha,\beta,\theta,x_0,y_0$) are the *intrinsic parameters*.
- Including projection from *P* to *p*,

$$\boldsymbol{p} = \frac{1}{Z}\mathcal{K}\begin{pmatrix}\text{Id} & \mathbf{0}\end{pmatrix}\boldsymbol{P} = \frac{1}{Z}\mathcal{M}\boldsymbol{P}, \quad \text{where} \quad \mathcal{M} \stackrel{\text{def}}{=} \begin{pmatrix}\mathcal{K} & \mathbf{0}\end{pmatrix}$$

- Note: division by Z is an artifact of setting last term in p to be 1.

Object and World Coordinate Systems

- Previous transformation matrix requires object coordinates to be expressed in the *camera* coordinate system (with origin at lens center)
  - This, in general, is not very convenient
- *Object* coordinate system
  - Aligned with some components of the object, *e.g.* the three sides of a rectangular solid
- *World* coordinate system
  - Chosen for global convenience, *e.g.* lines forming corner of a room , or earth coordinates (latitude, longitude, height)
- Coordinate transformations define relations between different coordinate systems
- *Extrinsic* parameters relate world coordinate system to camera coordinates

Rigid Transformations

- Notation

$$^{F}P \qquad \text{Point } P \text{ in Frame } F$$

$$(A) = (O_A, \boldsymbol{i}_A, \boldsymbol{j}_A, \boldsymbol{k}_A)$$

$$(B) = (O_B, \boldsymbol{i}_B, \boldsymbol{j}_B, \boldsymbol{k}_B)$$

- In general, two coordinate systems can be aligned by
  - Translation of origin (3 parameters)
  - Rotation
    - 3 rotations about the 3 axes (*e.g.* rotate about z-axes, then about the new y-axis, then about the new x-axis); called Euler angles
    - One direction about which rotation occurs and one angle
      - Screw representation, quaternions

Transformation Equations

- In non-homogeneous coordinates:

$$^{A}P = \mathcal{R}\,^{B}P + t$$

- Where *t* is translation vector (coordinates of origin of B in A); R is given by: $\mathcal{R} \stackrel{\text{def}}{=} \left({}^{A}\boldsymbol{i}_B, {}^{A}\boldsymbol{j}_B, {}^{A}\boldsymbol{k}_B\right)$

- Note that detailed matrix given in textbook, eq. 1.8 is wrong; correct answer is transpose of the given matrix

$$\mathcal{R} \stackrel{\text{def}}{=} \left({}^{A}\boldsymbol{i}_B, {}^{A}\boldsymbol{j}_B, {}^{A}\boldsymbol{k}_B\right) = \begin{pmatrix} \boldsymbol{i}_A \cdot \boldsymbol{i}_B & \boldsymbol{j}_A \cdot \boldsymbol{i}_B & \boldsymbol{k}_A \cdot \boldsymbol{i}_B \\ \boldsymbol{i}_A \cdot \boldsymbol{j}_B & \boldsymbol{j}_A \cdot \boldsymbol{j}_B & \boldsymbol{k}_A \cdot \boldsymbol{j}_B \\ \boldsymbol{i}_A \cdot \boldsymbol{k}_B & \boldsymbol{j}_A \cdot \boldsymbol{k}_B & \boldsymbol{k}_A \cdot \boldsymbol{k}_B \end{pmatrix}^{\mathrm{T}}$$

  - e.g. first column should be ($\boldsymbol{i}_A.\boldsymbol{i}_B$, $\boldsymbol{j}_A.\boldsymbol{i}_B$, $\boldsymbol{k}_A.\boldsymbol{i}_B$)

- In homogeneous coordinates:

$$^{A}P = \mathcal{T}\,^{B}P, \quad \text{where} \quad \mathcal{T} = \begin{pmatrix} \mathcal{R} & t \\ 0^T & 1 \end{pmatrix}$$

Combined Projection Equations

- Let (W) be a world coordinate frame, (C) a camera coordinate frame
- World to Camera coordinate transformation given by

$$^{C}P = \begin{pmatrix} \mathcal{R} & t \\ 0^T & 1 \end{pmatrix} {}^{W}P$$

- In camera coordinate frame

$$p = \frac{1}{Z}\mathcal{M}\,{}^{C}P,$$

- Combining the two, we get

$$p = \frac{1}{Z}\mathcal{M}P, \quad \text{where} \quad \mathcal{M} = \mathcal{K}(\mathcal{R} \quad t)$$

where P is in world coordinates, p in image coordinates.

Projection Equation

- Let m_1^T, m_2^T and m_3^T denote the 3 rows of M, then $Z = m_3 \cdot P$
- Alternate form:

$$x = \frac{m_1 \cdot P}{m_3 \cdot P},$$
$$y = \frac{m_2 \cdot P}{m_3 \cdot P}.$$

- Let r_1^T, r_2^T, and r_3^T denote the 3 rows of R, and t_1,t_2,t_3 denote the three components of t, then:

$$\mathcal{M} = \begin{pmatrix} \alpha r_1^T - \alpha \cot\theta r_2^T + x_0 r_3^T & \alpha t_1 - \alpha \cot\theta t_2 + x_0 t_3 \\ \dfrac{\beta}{\sin\theta} r_2^T + y_0 r_3^T & \dfrac{\beta}{\sin\theta} t_2 + y_0 t_3 \\ r_3^T & t_3 \end{pmatrix}$$

Properties of Matrix M

- Can any arbitrary 3 x 4 matrix be a perspective projection matrix (corresponding to some internal and external parameters)?

Theorem 1. Let $\mathcal{M} = (\mathcal{A} \quad \boldsymbol{b})$ be a 3×4 matrix, and let $\boldsymbol{a}_i^T$ $(i = 1, 2, 3)$ denote the rows of the matrix $\mathcal{A}$ formed by the three leftmost columns of $\mathcal{M}$.

- A necessary and sufficient condition for $\mathcal{M}$ to be a perspective projection matrix is that $\text{Det}(\mathcal{A}) \neq 0$.
- A necessary and sufficient condition for $\mathcal{M}$ to be a zero-skew perspective projection matrix is that $\text{Det}(\mathcal{A}) \neq 0$ and

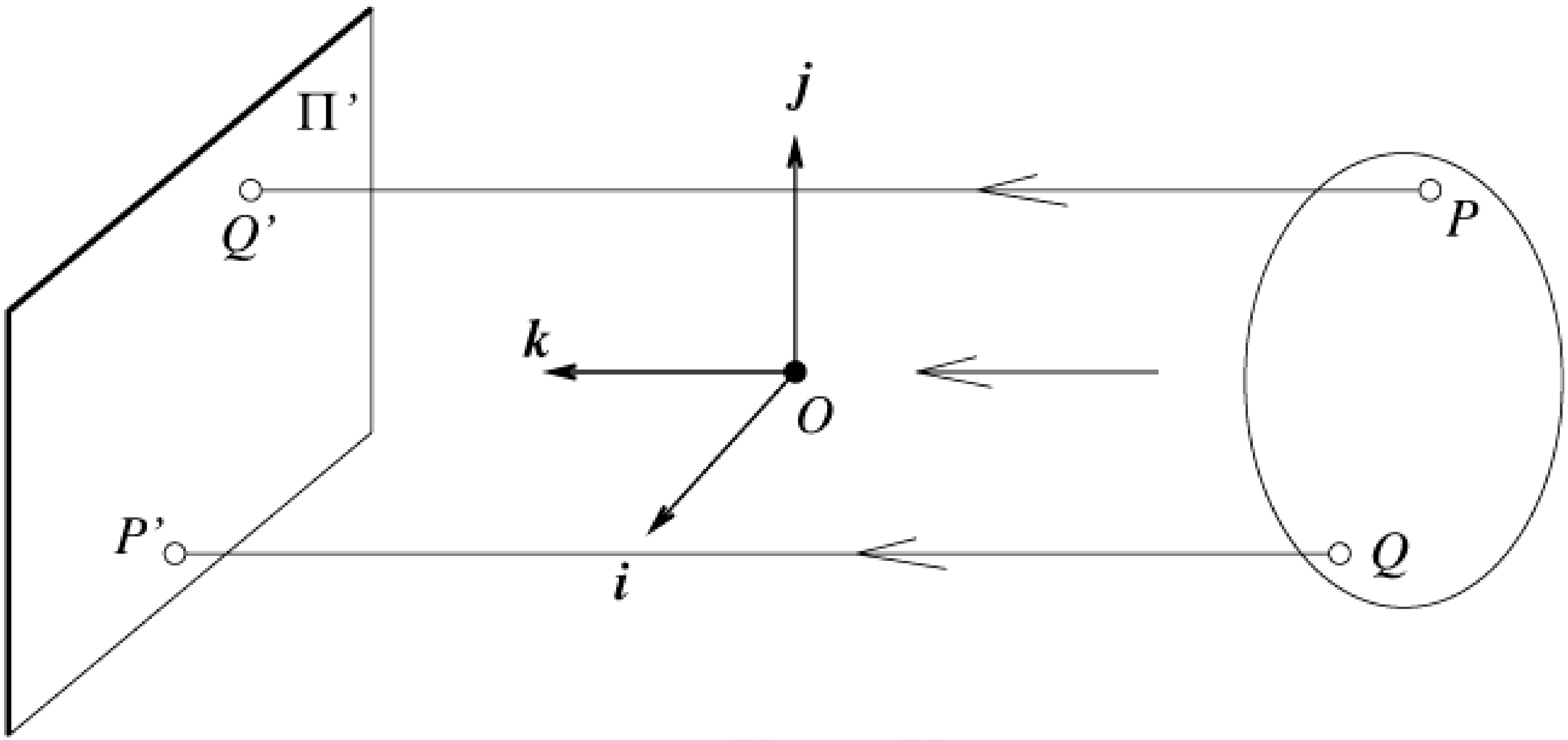
$$(\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3) = 0.$$

- A necessary and sufficient condition for $\mathcal{M}$ to be a perspective projection matrix with zero skew and unit aspect-ratio is that $\text{Det}(\mathcal{A}) \neq 0$ and

$$\begin{cases} (\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3) = 0, \\ (\boldsymbol{a}_1 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_1 \times \boldsymbol{a}_3) = (\boldsymbol{a}_2 \times \boldsymbol{a}_3) \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3). \end{cases}$$

Orthographic Projection

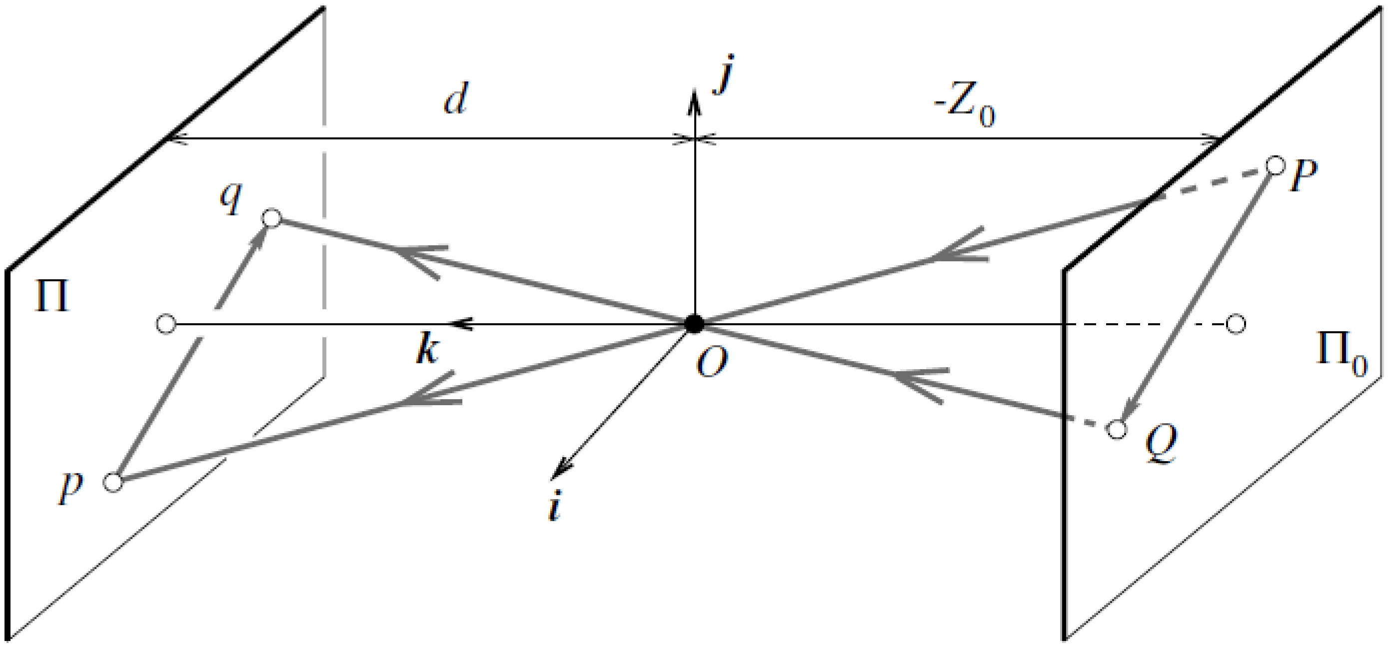
Assumes projection rays are parallel, and along the z-axis.



$$x = \mathrm{X}, \;\; y = \mathrm{Y}$$

Weak Perspective

Perspective projection but assume all points have the same z-value (object sizes small, compared to distance from camera)



$$x = -mX, \quad y = -mY, \quad \text{where} \quad m = -\frac{d}{Z_0}$$

Matrix form developed in next slide

Next Class

- FP: Sections 1.3, 2.1, 2.3.4, 2.4