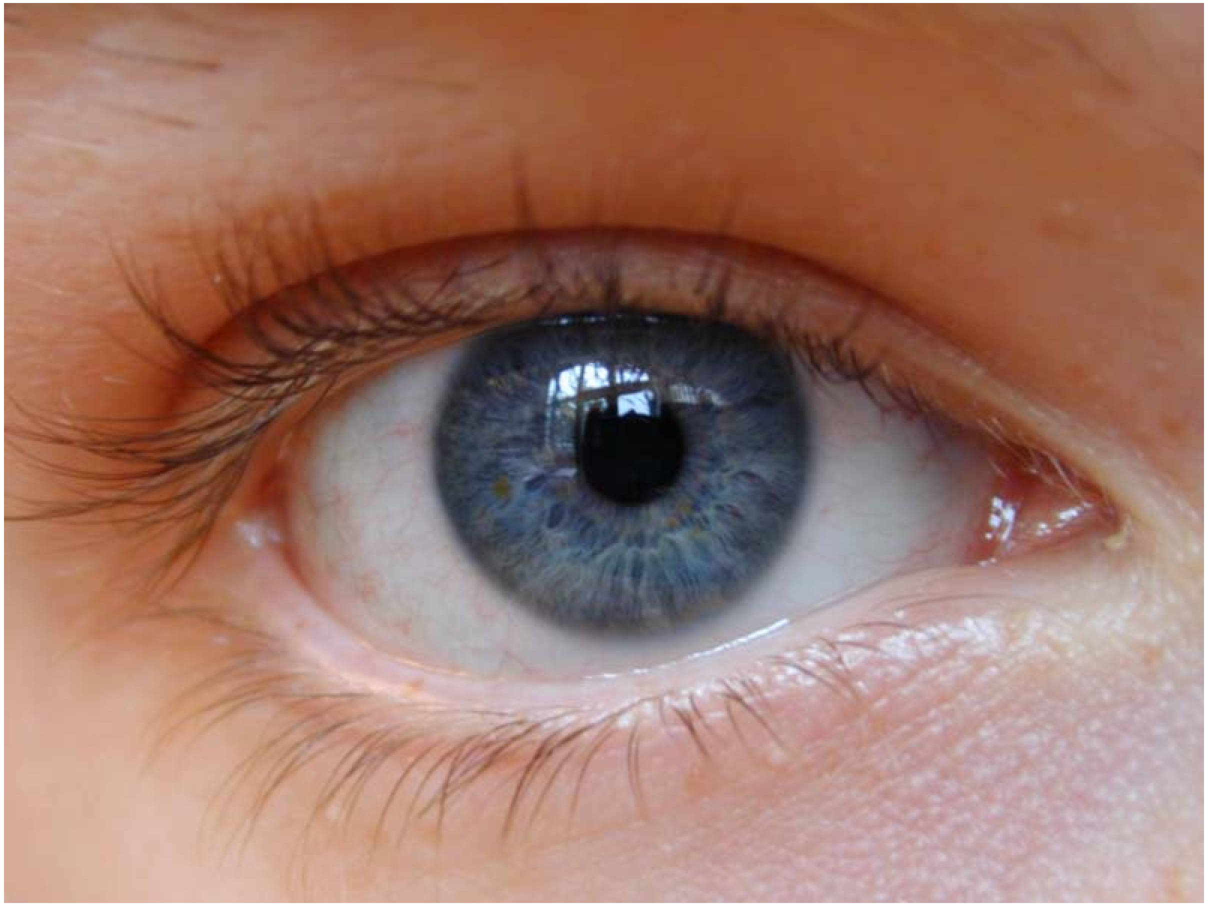
Leow Wee Kheng

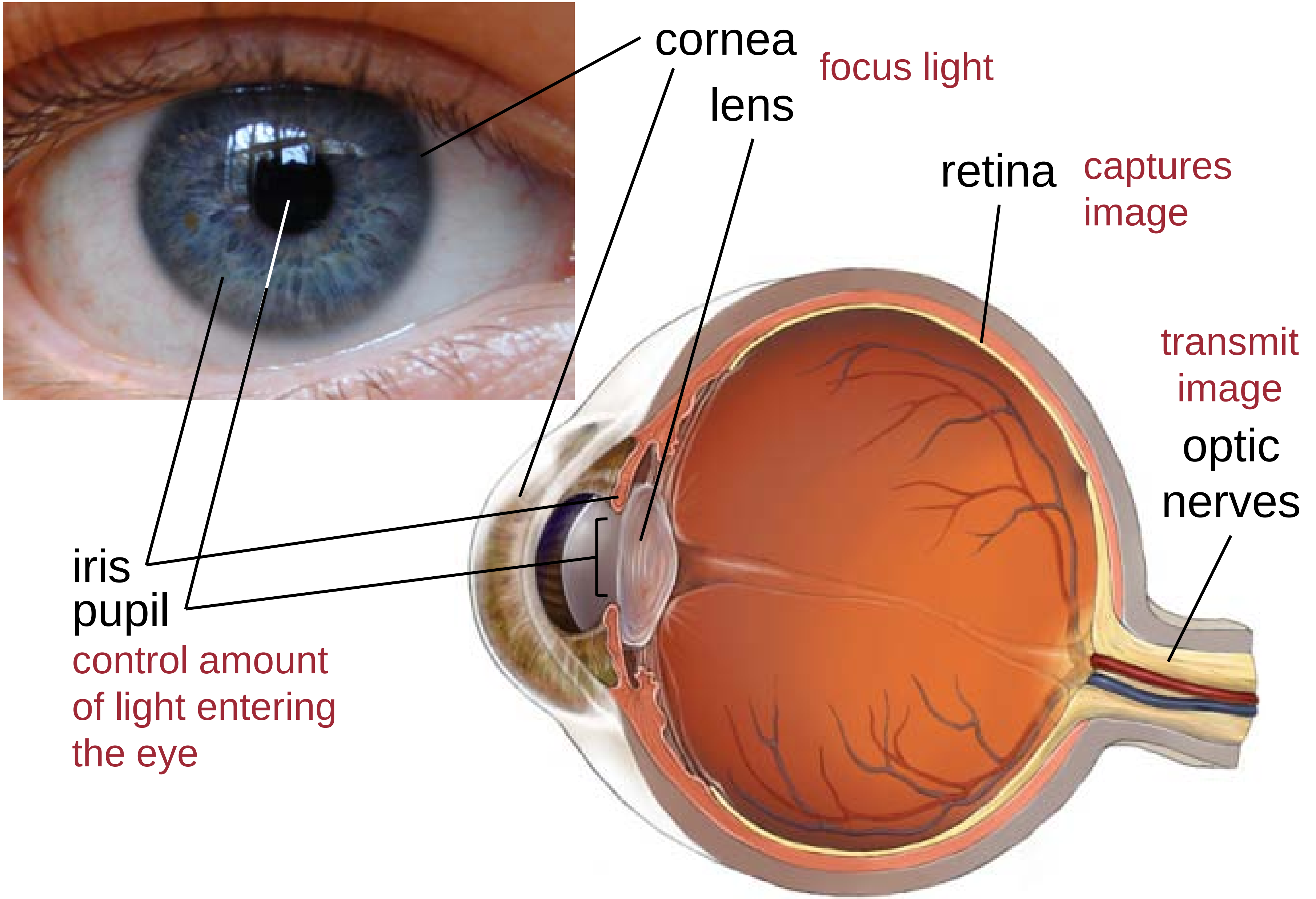
CS4243 Computer Vision and Pattern Recognition

Camera Models and Imaging

Through our eyes…

- Through our eyes we see the world.

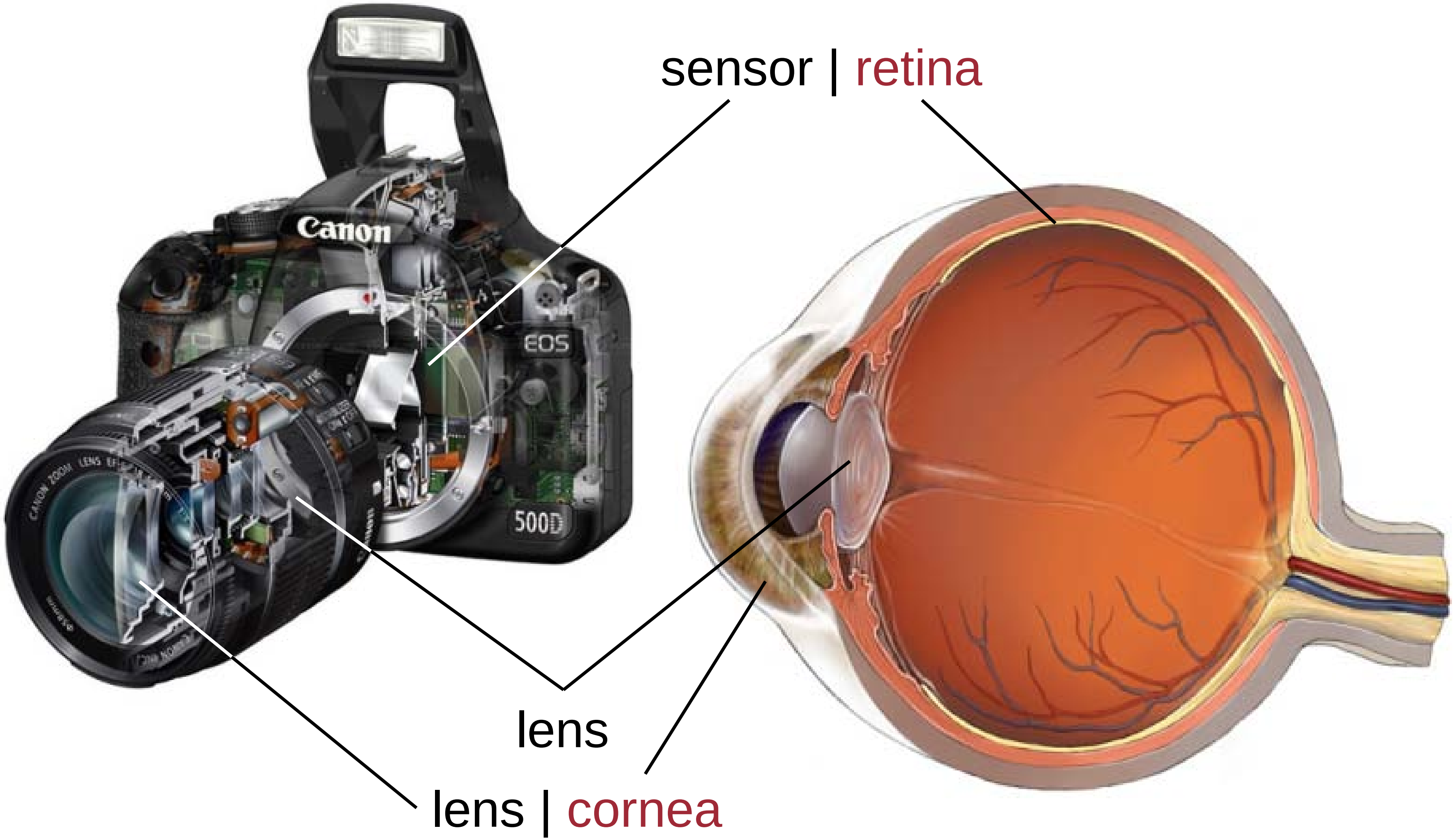


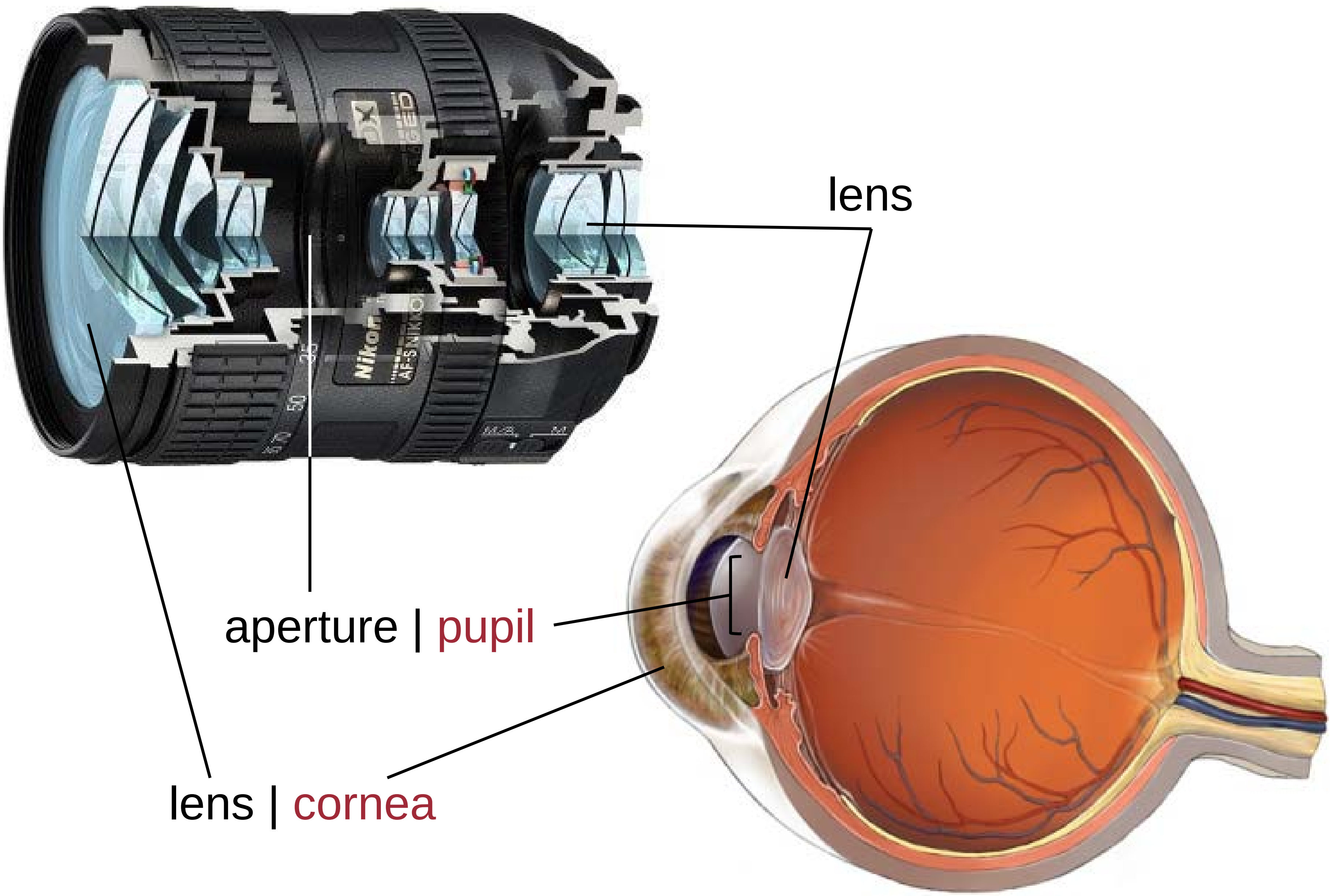

cornea
focus light
lens
retina
captures image
transmit image
optic nerves
iris
pupil
control amount of light entering the eye

Through their eyes…

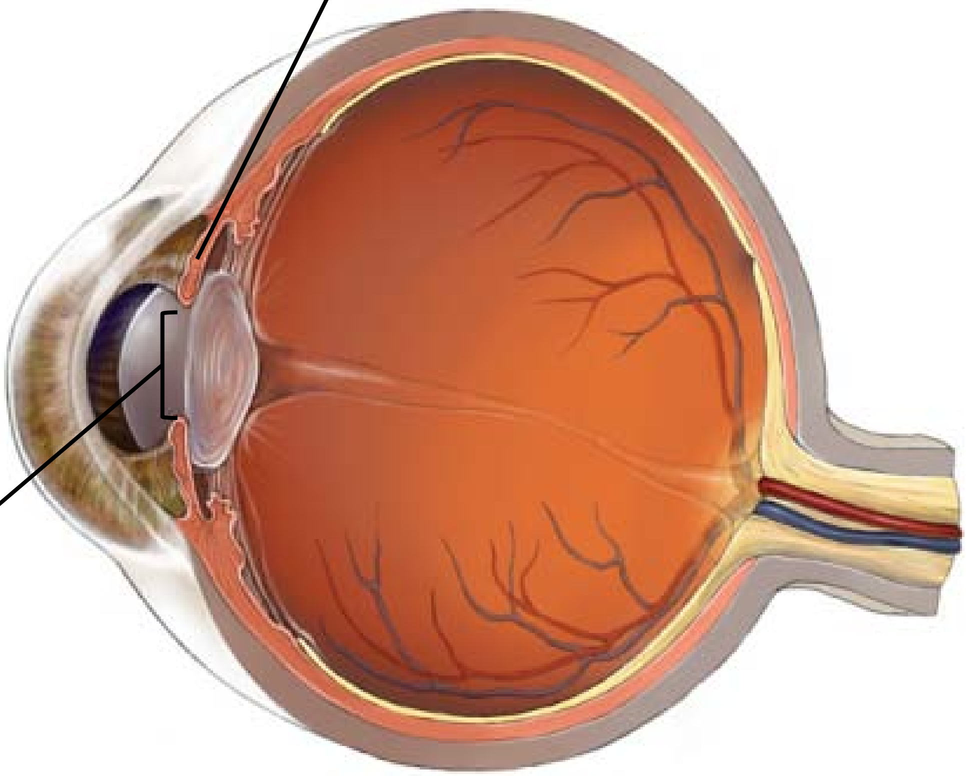
- Camera is computer's eye.

- Camera is structurally the same the eye.




lens
aperture | pupil
lens | cornea

aperture plates | iris



aperture | pupil

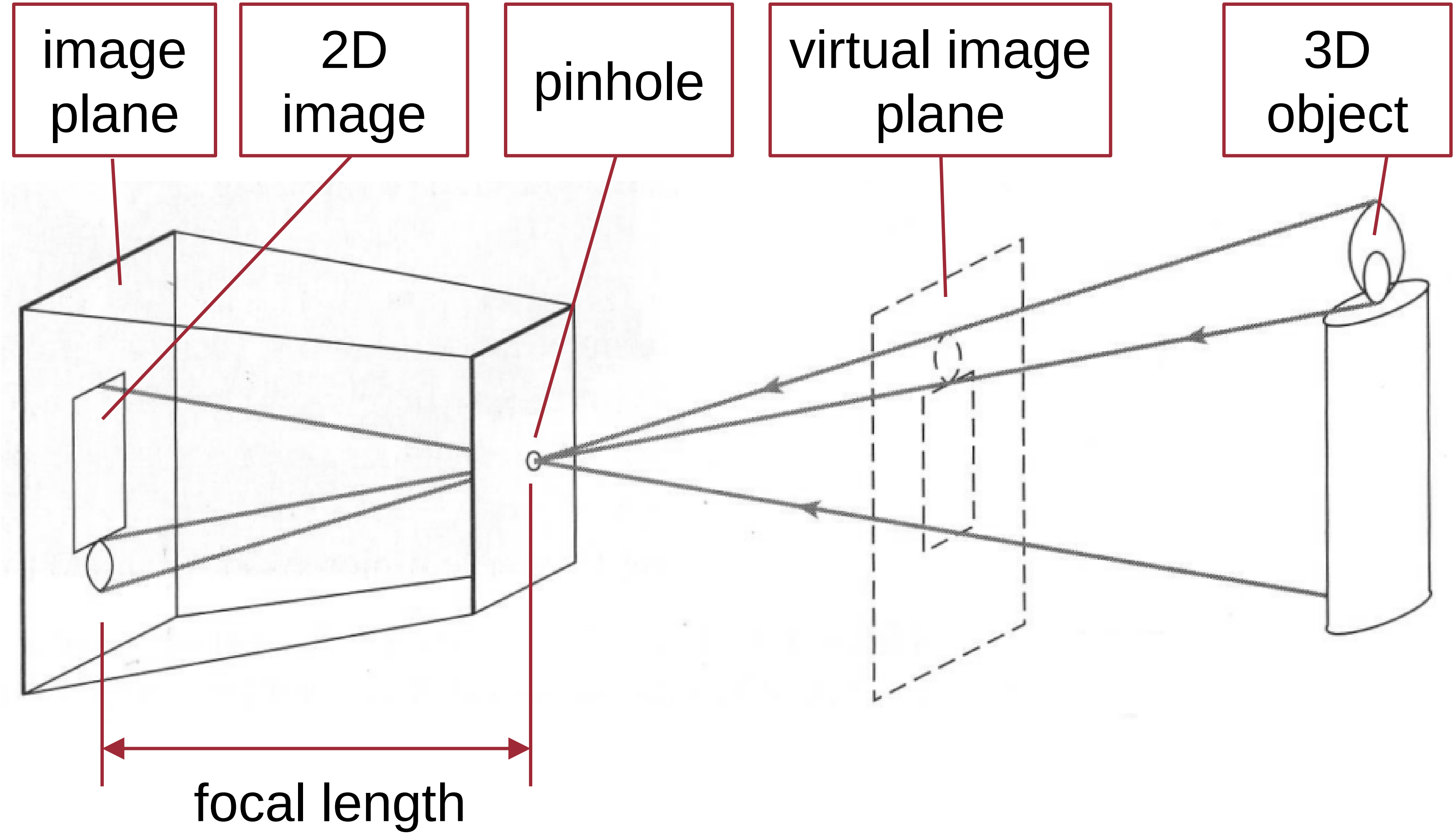
Imaging

- Images are 2D projections of real world scene.
- Images capture two kinds of information:
  - Geometric: positions, points, lines, curves, etc.
  - Photometric: intensity, colour.
- Complex 3D-2D relationships.
- Camera models approximate relationships.

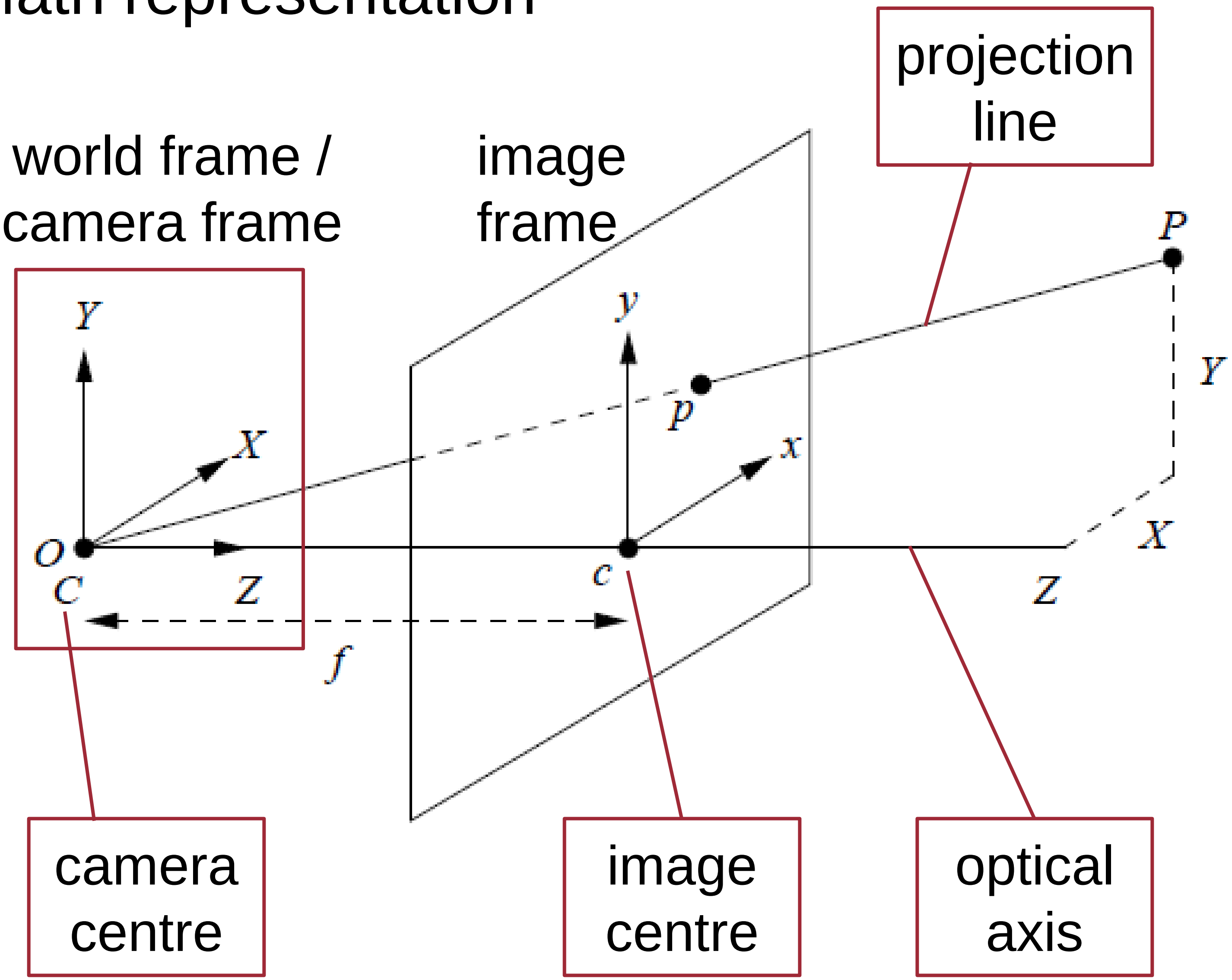
Camera Models

- Pinhole camera model
- Orthographic projection
- Scaled orthographic projection
- Paraperspective projection
- Perspective projection

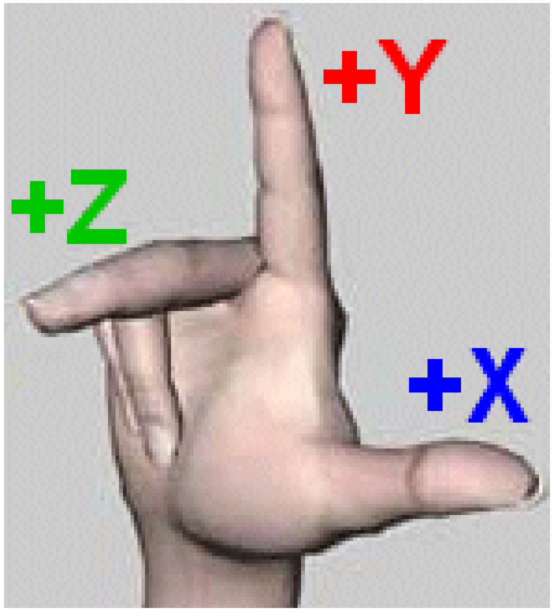
Pinhole Camera Model

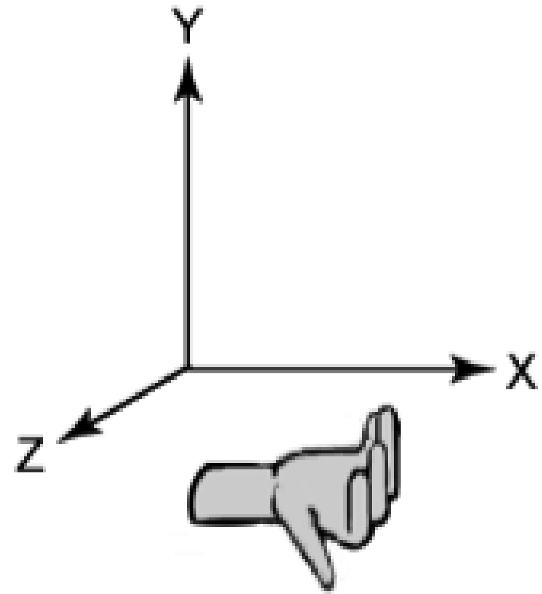


- Math representation



- By default, use right-handed coordinate system.





- 3D point $\mathbf{P} = (X, Y, Z)^\mathsf{T}$ projects to 2D image point $\mathbf{p} = (x, y)^\mathsf{T}$.
- By symmetry,

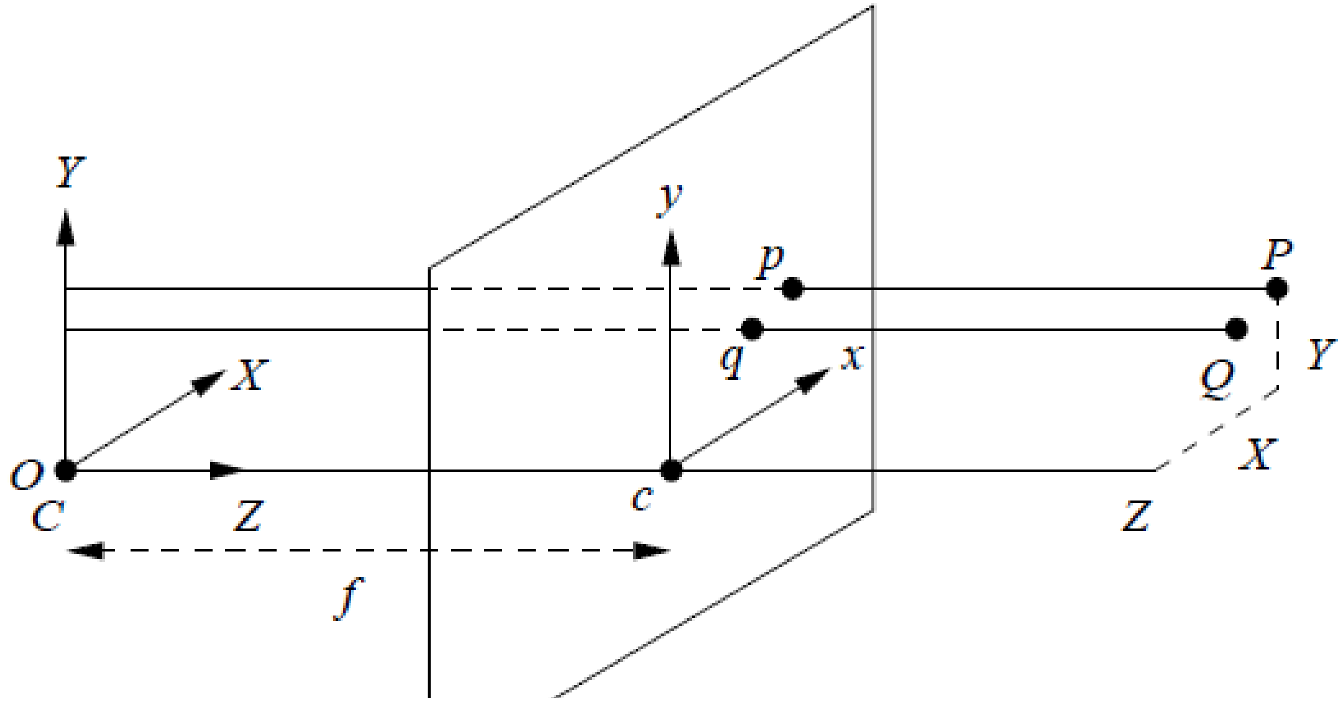
$$\frac{X}{Z} = \frac{x}{f}, \quad \frac{Y}{Z} = \frac{y}{f}$$

  i.e.,

$$x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z}$$

- Simplest form of perspective projection.

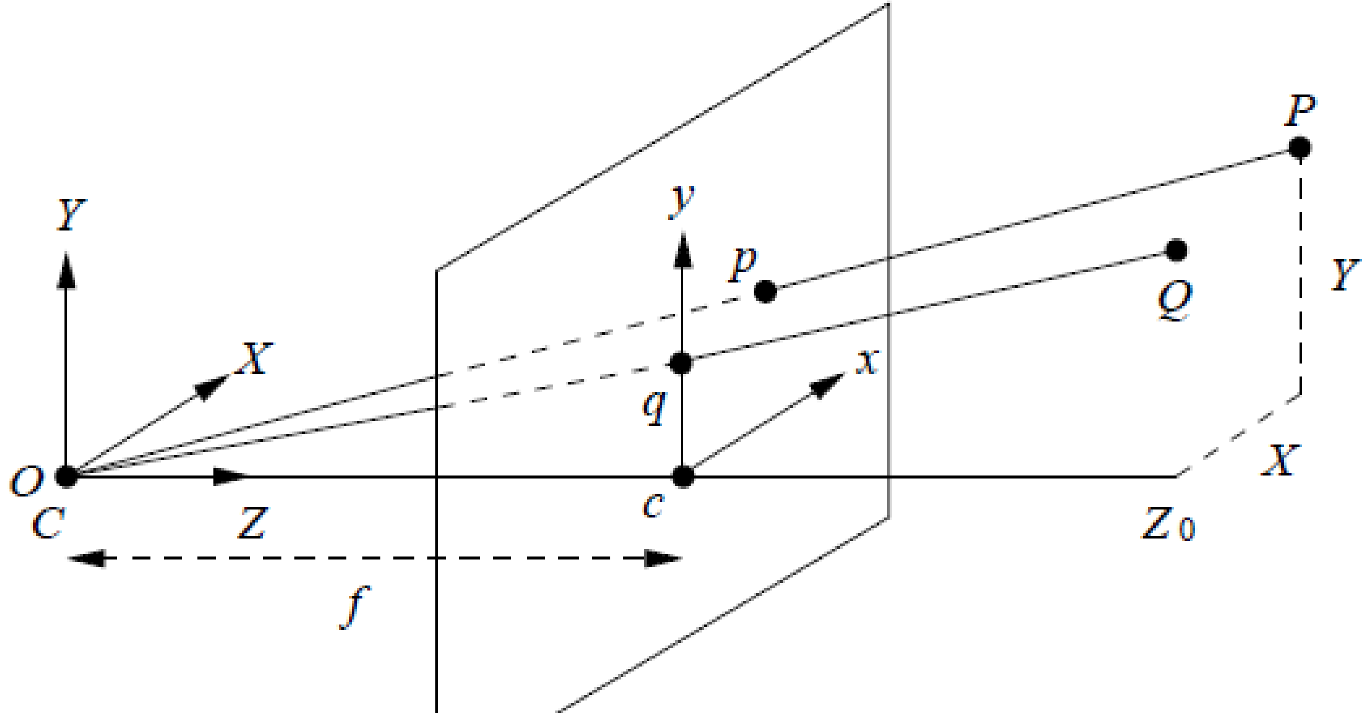
Orthographic Projection



- 3D scene is at infinite distance from camera.
- All projection lines are parallel to optical axis.
- So,

$$x = X, \quad y = Y$$

Scaled Orthographic Projection



- Scene depth << distance to camera.
- Z is the same for all scene points, say Z_0

$$x = sX, \quad y = sY, \quad s = \frac{f}{Z_0} \text{ for all scene points.}$$

Perspective Projection

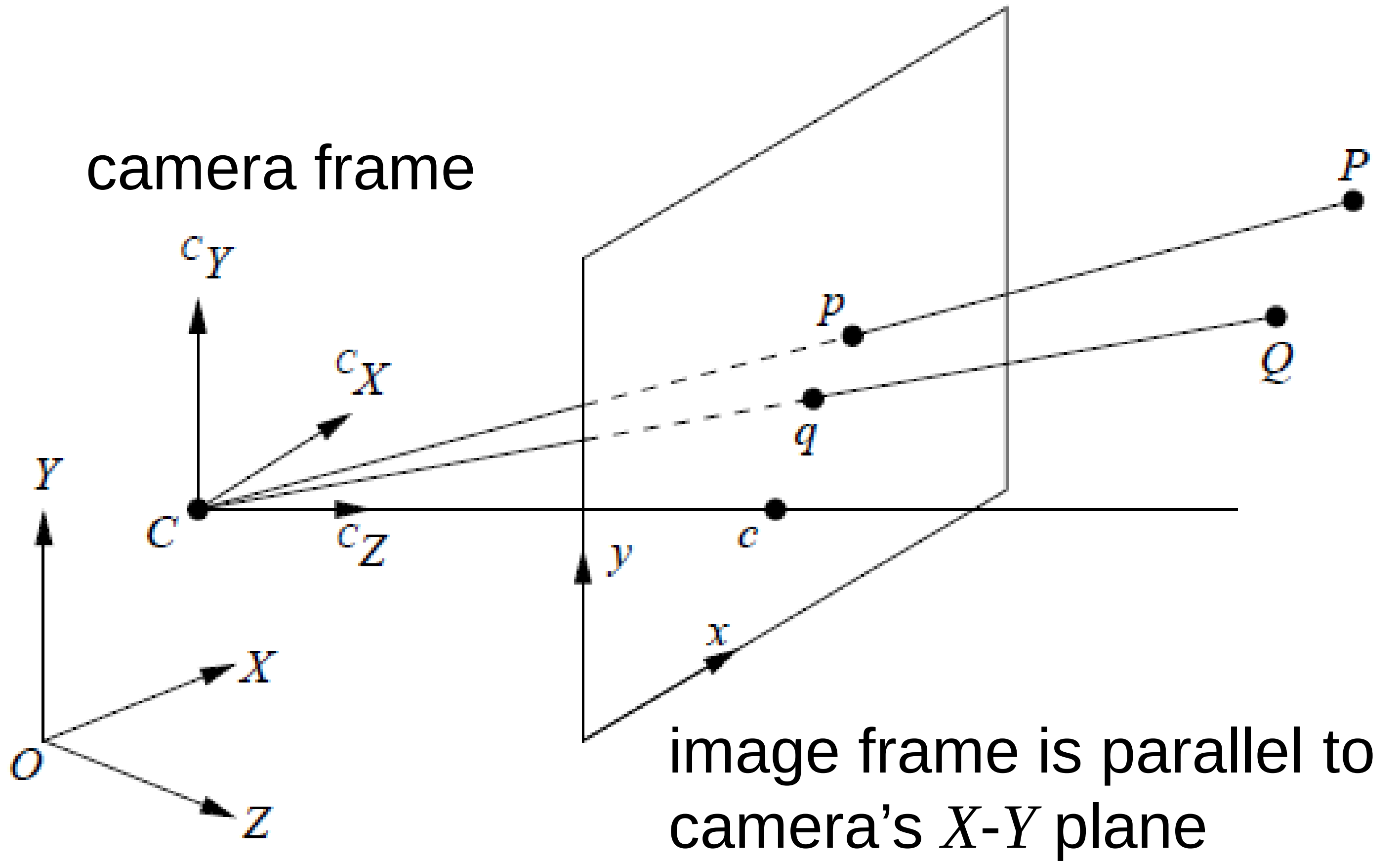


image frame is parallel to camera's *X*-*Y* plane

world frame

Intrinsic Parameters

- Pinhole camera model

$$x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z}$$

- Camera sensor's pixels not exactly square

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

  - x, y : coordinates (pixels)
  - k, l : scale parameters (pixels/m)
  - f : focal length (m or mm)

- f, k, l are not independent.
- Can rewrite as follows (in pixels):

$$f_x = kf \qquad f_y = lf$$

- So,

$$x = f_x \frac{X}{Z}, \quad y = f_y \frac{Y}{Z}$$

- Image centre or principal point c may not be at origin.
- Denote location of c in image plane as c_x, c_y.
- Then,

$$x = f_x \frac{X}{Z} + c_x, \quad y = f_y \frac{Y}{Z} + c_y$$

- Along optical axis, $X = Y = 0$, and $x = c_x$, $y = c_y$.

- Image frame may not be exactly rectangular.
- Let θ denote skew angle between x- and y-axis.
- Then,

$$x = f_x \frac{X}{Z} - f_x \cot\theta \frac{Y}{Z} + c_x, \quad y = \frac{f_y}{\sin\theta}\frac{Y}{Z} + c_y$$

- Combine all parameters yield

$$\tilde{\mathbf{x}} = \frac{1}{Z}\mathbf{K}\mathbf{X}, \quad \mathbf{K} = \begin{bmatrix} f_x & -f_x \cot\theta & c_x \\ 0 & \dfrac{f_y}{\sin\theta} & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Intrinsic parameter matrix

$$\tilde{\mathbf{x}} = [x, y, 1]^\top$$

homogeneous coordinates

- Denoting ρ = Z, we get

$$\rho\,\tilde{\mathbf{x}} = \mathbf{K}\mathbf{X}$$

- Some books and papers absorb ρ into $\tilde{\mathbf{x}}$:

$$\tilde{\mathbf{x}} = [\rho\,x, \rho\,y, \rho]^\top$$

  giving

$$\tilde{\mathbf{x}} = \mathbf{K}\mathbf{X}$$

  Be careful.
- A simpler form of $\mathbf{K}$ uses skew parameter s:

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Extrinsic Parameters

- Camera frame is not aligned with world frame.
- Rigid transformation between them:

frame

$$ {}^{C}\mathbf{X} = {}^{C}_{W}\mathbf{R}\,{}^{W}\mathbf{X} + {}^{C}_{W}\mathbf{T} $$

object

  - Coordinates of 3D scene point in camera frame.
  - Coordinates of 3D scene point in world frame.
  - Rotation matrix of world frame in camera frame.
  - Position of world frame's origin in camera frame.

- Translation matrix

$$\mathbf{T} = \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

- Rotation matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

- Rotation matrix is orthonormal:

$$\mathbf{R}^\top \mathbf{R} = \mathbf{I}$$

- So, $\mathbf{R}^{-1} = \mathbf{R}^\top$.
- In particular, let

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1^\top \\ \mathbf{R}_2^\top \\ \mathbf{R}_3^\top \end{bmatrix}$$

Then,

$$\mathbf{R}_i^\top \mathbf{R}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

- Combine intrinsic and extrinsic parameters:

$$\rho\,\tilde{\mathbf{x}} = \mathbf{K}\,{}^{C}\mathbf{X} = \mathbf{K}\,({}^{C}_{W}\mathbf{R}\,{}^{W}\mathbf{X} + {}^{C}_{W}\mathbf{T})$$

- Simpler notation:

$$\rho\,\tilde{\mathbf{x}} = \mathbf{K}(\mathbf{R}\mathbf{X} + \mathbf{T})$$

Lens Distortion

- Lens can distort images, especially at short focal length.

barrel

pin-cushion

fisheye

- Radial distortion is modelled as

$$
\begin{aligned}
x_d &= x\,(1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6) \\
y_d &= y\,(1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6)
\end{aligned}
$$

distorted coordinates | undistorted coordinates | distortion parameters

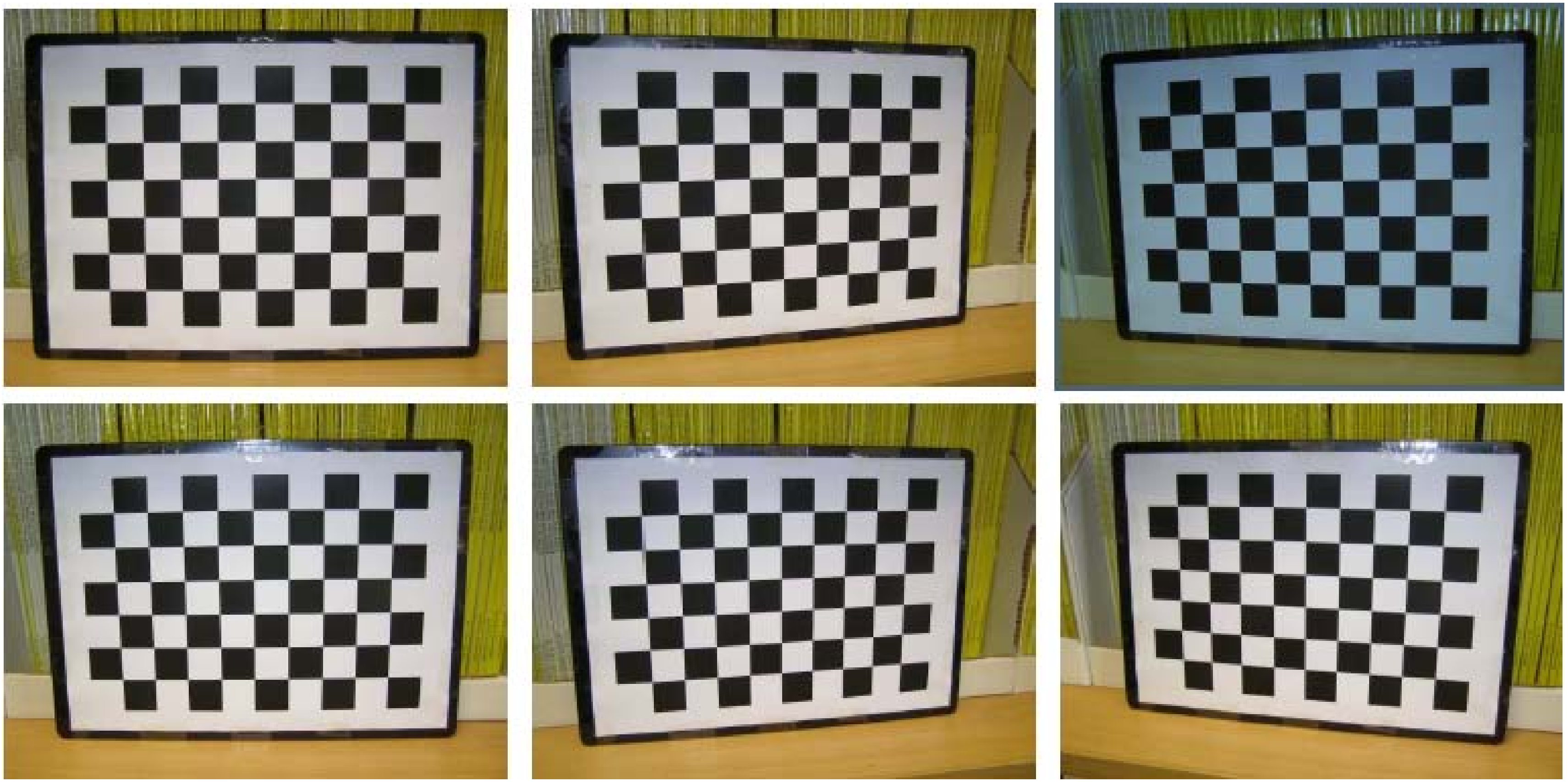
with $r^2 = x^2 + y^2$.

- Actual image coordinates

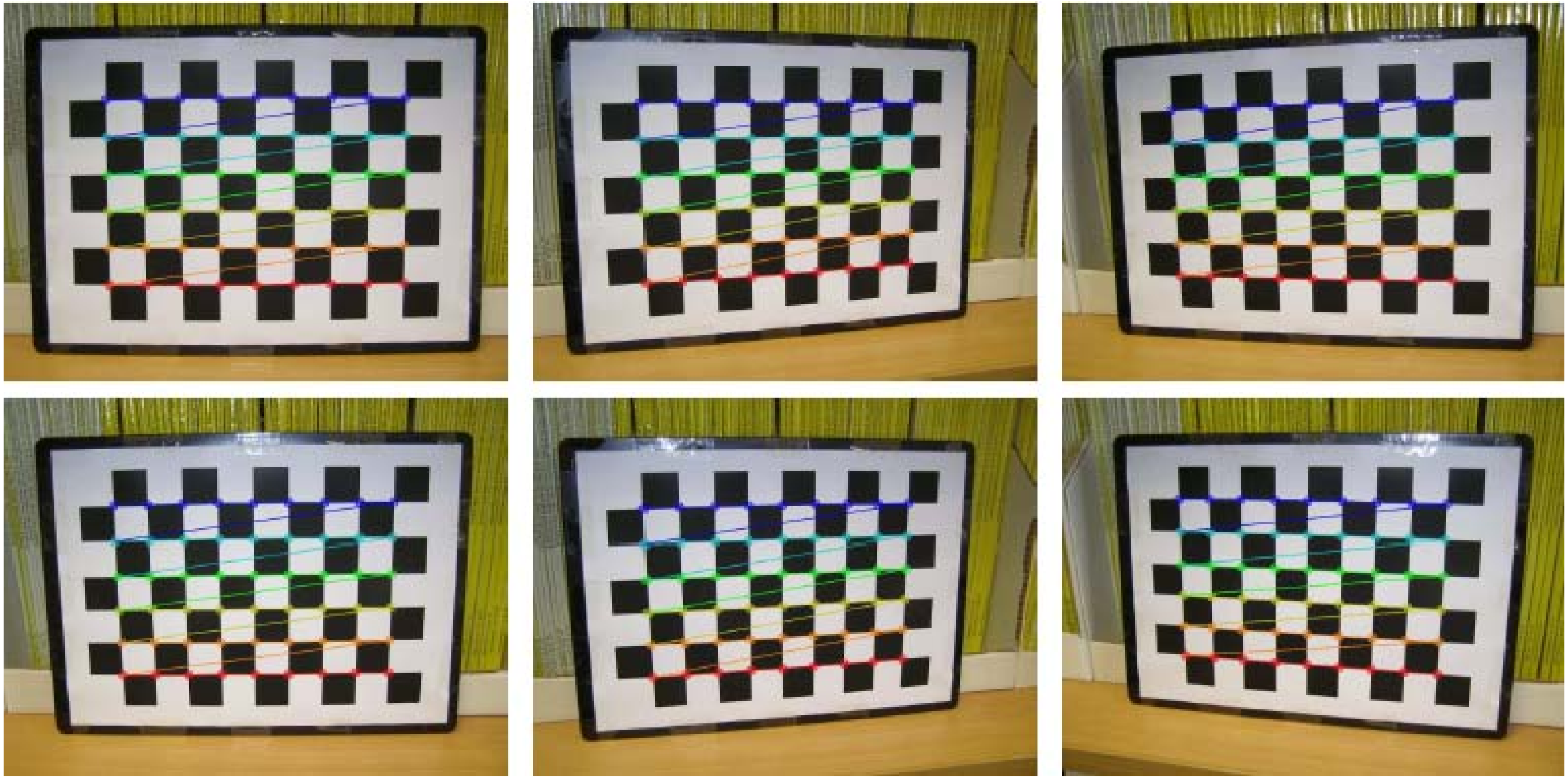
$$
\begin{aligned}
x_a &= f_x\, x_d + c_x \\
y_a &= f_y\, y_d + c_y
\end{aligned}
$$

Camera Calibration

- Compute intrinsic / extrinsic parameters.
- Capture calibration pattern from various views.



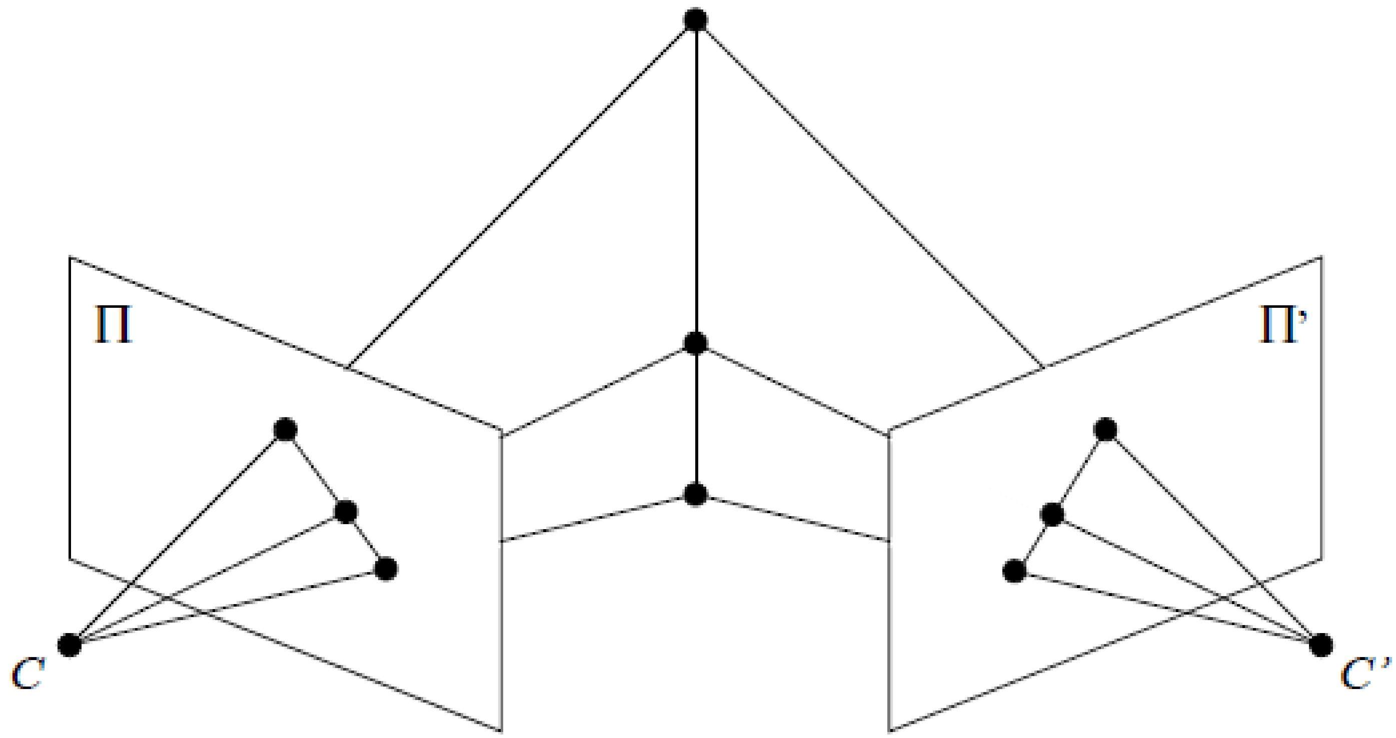
- Detect inner corners in images.



- Run calibration program (available in OpenCV).

Homography

- A transformation between projective planes.
- Maps straight lines to straight lines.



- Homography equation:

$$\begin{bmatrix} \rho_i' x_i' \\ \rho_i' y_i' \\ \rho_i' \end{bmatrix} = \tilde{\mathbf{x}}_i' = \mathbf{H}\tilde{\mathbf{x}}_i = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

image point — homography — image point

- Affine is a special case of homography:
  - $h_{31} = h_{32} = 0,\ h_{33} = 1$

- Scaling $\mathbf{H}$ by s does not change equation:

$$(s\mathbf{H})\tilde{\mathbf{x}}_i = s\tilde{\mathbf{x}}'_i = \tilde{\mathbf{x}}'_i$$

- Homography is defined up to unspecified scale.
- So, can set $h_{33} = 1$.
- Expanding homography equation gives

$$\begin{aligned} h_{11}x_i + h_{12}y_i + h_{13} - h_{31}x_i x'_i - h_{32}y_i x'_i &= x'_i \\ h_{21}x_i + h_{22}y_i + h_{23} - h_{31}x_i y'_i - h_{32}y_i y'_i &= y'_i \end{aligned}$$

- Assembling equations over all points yields

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 x'_1 & -y_1 x'_1 \\ & & & \vdots & & & & \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_n x'_n & -y_n x'_n \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 y'_1 & -y_1 y'_1 \\ & & & \vdots & & & & \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_n y'_n & -y_n y'_n \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \\ y'_1 \\ \vdots \\ y'_n \end{bmatrix}$$

- System of linear equations. Easy to solve.

Example

- Circles: input corresponding points.
- Black dots: computed points, match input points.

Before homographic transform

After homographic transform

Summary

- Simplest camera model: pinhole model.
- Most commonly used model: perspective model.
- Intrinsic parameters:
  - Focal length, principal point.
- Extrinsic parameters:
  - Camera rotation and translation.
- Lens distortion
- Homography: maps lines to lines.

Further Reading

- Camera models: [Sze10] Section 2.1.5, [FP03] Section 1.1, 1.2, 2.2, 2.3.
- Lens distortion: [Sze10] Section 2.1.6, [BK08] Chapter 11.
- Camera calibration: [BK08] Chapter 11, [Zha00].
- Image undistortion: [BK08] Chapter 11.

References

- Bradski and Kaehler. *Learning OpenCV*. O'Reilly, 2008.
- D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Pearson Education, 2003.
- R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2010.
- Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. PAMI, 22(11):1330–1334, 2000.*