ME5411

ROBOT VISION AND AI

Dr. NG Hsiao Piau

Ng_h_p@nus.edu.sg

Lecture 2

Image Acquisition and Camera Calibration

Topics

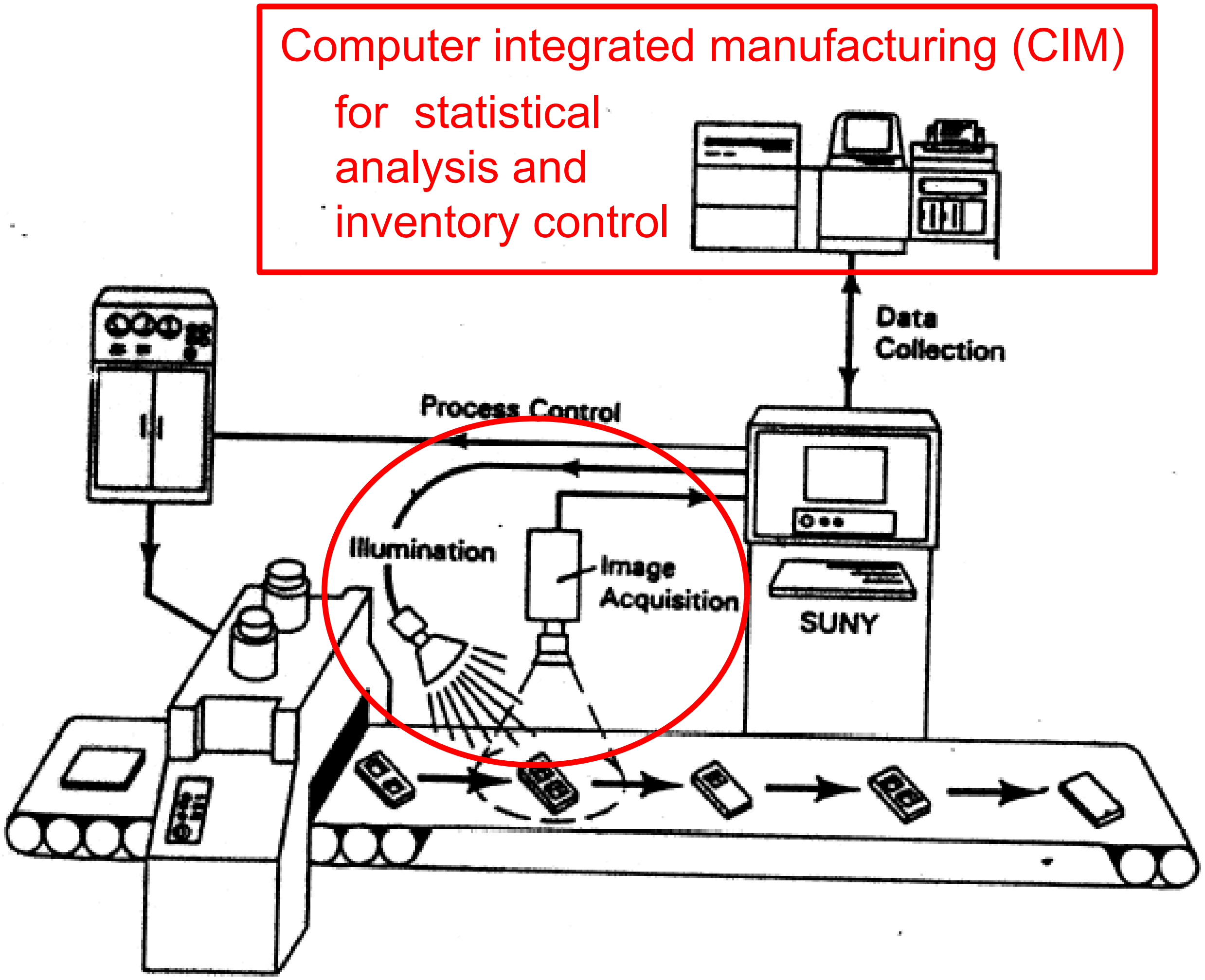
- Definitions
- Illumination
- Camera
- Camera Calibration

Key takeaways

- Image acquisition (illumination + camera + frame grabber)
- Lighting types: front, back, structured
- CCD/CMOS sensors use photovoltaic effect
- Optics: aperture, focal length, magnification
- Calibration: 3D-to-2D mapping via matrix transforms

1. Definitions

- Industrial manufacturing cell with vision cell
- Task: the vision system observes the object, determines if it is within specification, and generates command signals accordingly.
- <u>Image acquisition system: lights, camera and frame grabber.</u>
- Processing equipment
- Output equipment
- Control action



Computer integrated manufacturing (CIM)

for statistical analysis and inventory control

Image Acquisition

- The process to transform visual image of a physical object and its intrinsic characteristics into a set of digitized data which can be used by the processing unit of the system.
  - Lights (Illumination)
  - Camera
  - Frame Grabber

Four phases:

1. Illumination
2. Image formation or focusing
3. Image detection or sensing
4. Formatting camera output signal

Frame Grabber

Formatting camera output signal

- An electronic device that captures individual, digital still frames from an analog video signal or a digital video stream.
- In a typical machine/computer vision system, the frame grabber also displays, stores or transmits the captured video frame in raw or compressed digital form.

2. Illumination

- Refers to the science of the application of lighting.
  - sources of lighting
  - design of lighting systems

- Aims to produce an effective environment for camera to see in the context of machine vision.

Illumination

- Key parameter, often limiting factor
- May require 30% of application effort
- Customized to each application effort
- Can produce effects beyond human vision (infrared, ultraviolet, X-rays, 3D information in a 2D image)
- Performance of fluorescent lamp varies
  - Regular monitoring

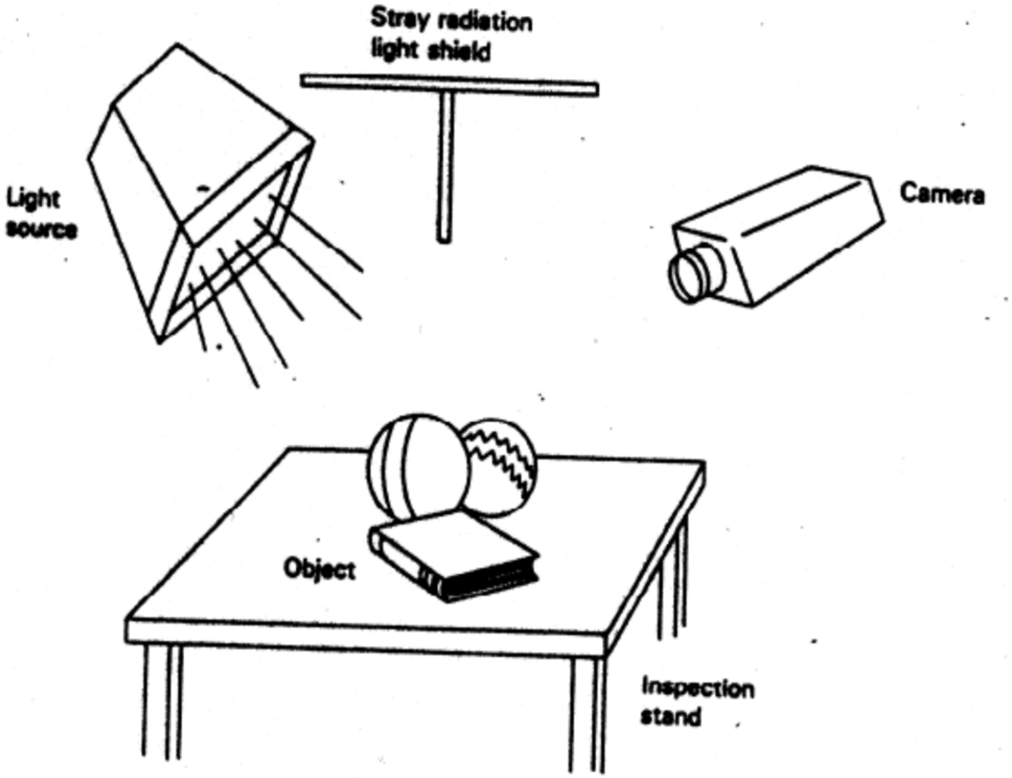
The fluorescent lamp output decreases as much as 15% during the first 100 hours and then continues to decrease everyday at a slower rate. The fluorescent lamps are brightest at 40^{o}C and are sensitive to the applied voltage.

Principal kinds of lighting

- Front lighting: light source and camera on the same side relative to object
  - Useful to obtain surface texture, features and for dimensioning
- Back lighting: object between light source and camera
- Structured light: illuminating the object with a grid or regular stripes

Front Lighting

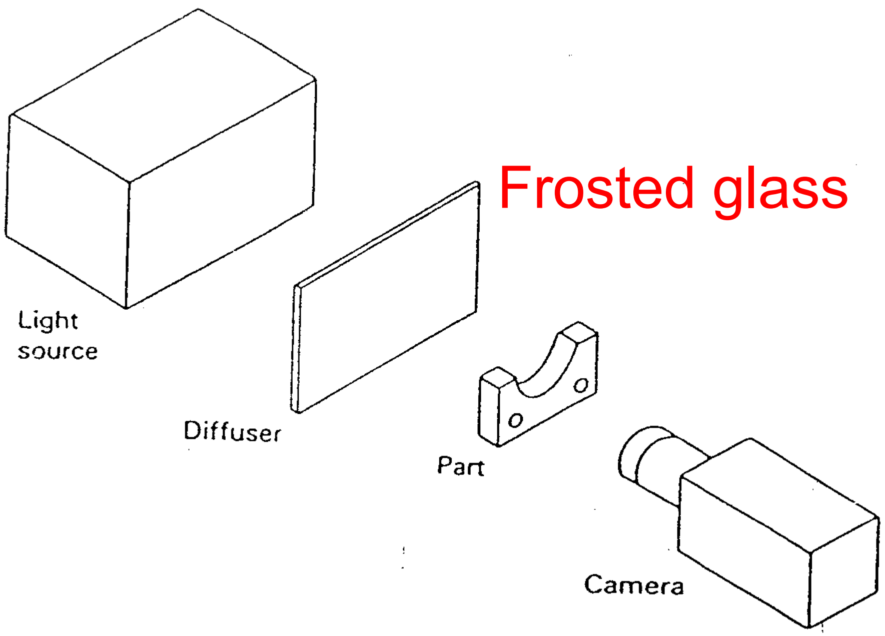
Front lighting employs light reflected from the object. The illumination source and the camera are both on the same side of the object. It is useful to obtain surface texture of features as well as dimensioning.



Back Lighting

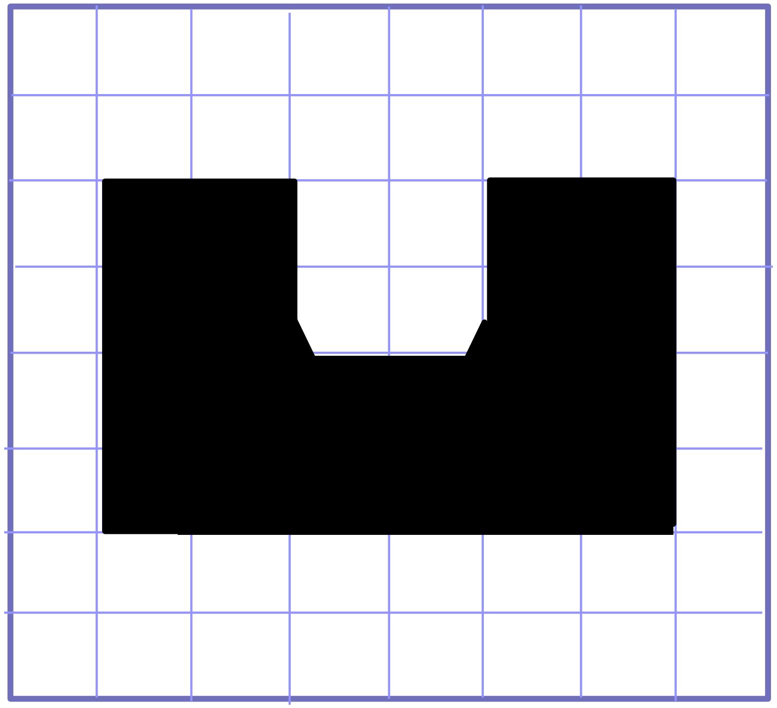
- Create a silhouette of the object
- Often used to locate parts moving on a conveyer belt
- Ideal to detect foreign material and fracture in transparent objects

Back lighting is when an object is located between the light source and the camera.



Frosted glass

Back Lighting



Back-lighted object

15	15	15	15	15	15	15	15
15	15	15	15	15	15	15	15
15	1	0	15	15	0	1	15
15	1	0	10	10	0	1	15
15	1	0	0	0	0	1	15
15	1	0	0	0	0	1	15
15	1	1	1	1	1	1	15
15	15	15	15	15	15	15	15

Image data

Produces high contrast images, minimizes processing tasks and reduces the sensitivity of the system to illumination source variation.

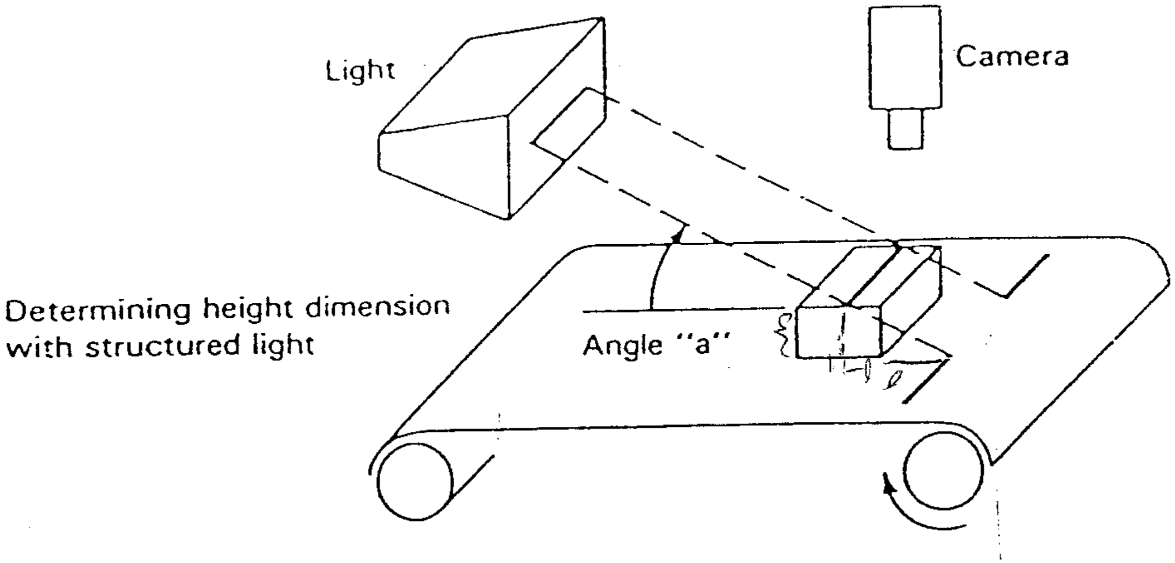
Cannot be employed to obtain information on surface characteristics, or features not visible in silhouette like the presence of bolts in blind hole, and objects located on top of each other.

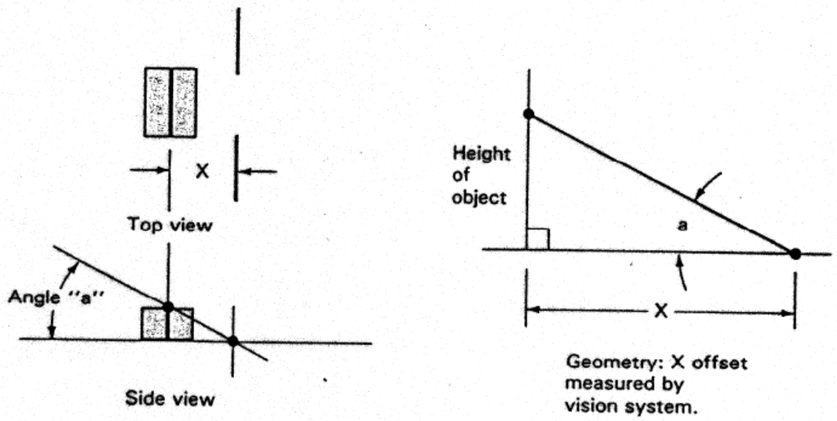
Structured Light

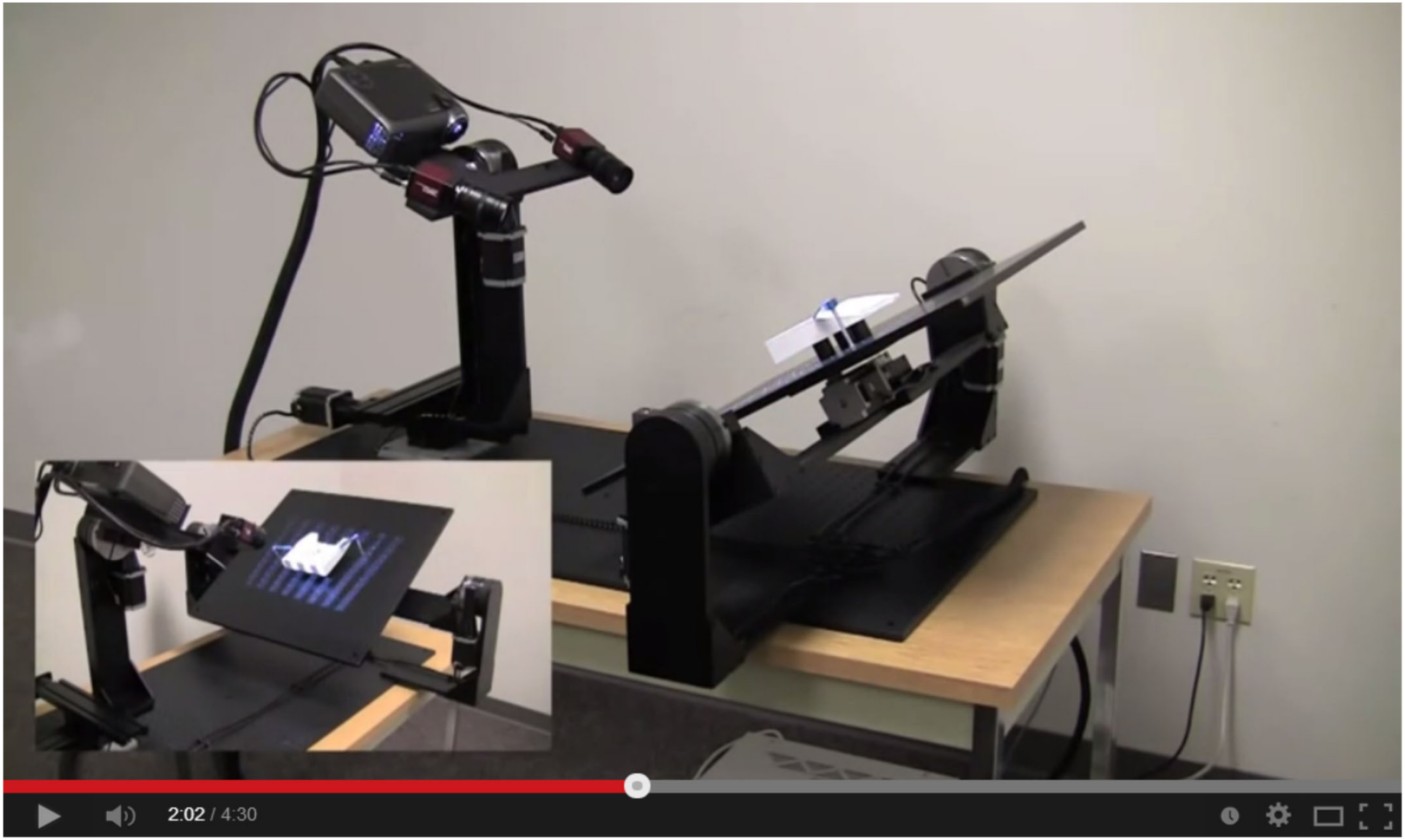
- 3D information in a 2D image
- Lighting angle and light structure (regular stripes, grids) depend on the application

Structured light is the use of the illumination of the object with a special pattern or grid. The intersection of the object and the projected illumination results in a unique pattern depending on the shape and dimensions of the object

A 3-D feature is converted to a 2-D image. Vertical and horizontal distances, as well as the shape of the surface features, can be measured.







Automated Inspection using Structured Light Scanning:
http://www.youtube.com/watch?v=IpxQBTrPBEg

3. Camera

- A camera is a device used to capture still images (photographs) or as sequences of moving images (movies or videos).
- Generally consists of
  - Lens system including an opening (aperture) at one end for light to enter.
  - Image capture for capturing the light at the other end.

Image Formation and Focusing

The image of the object is focused on the sensing element with a lens. The sensor converts the visual image to an electrical signal.

Image Detection

Image detection is done with by the image sensor of the camera. The basic concept of image sensors is that a separate electrical signal is produced for each pixel or area in the sensor depending on the amount of light energy falling on the device in each pixel area.

Camera

- Most cameras use either CCD or CMOS photosensitive elements to capture image, both using photovoltaic principles. They capture brightness of a monochromatic image.

- A charge-coupled device (CCD) is a sensor for recording images, consisting of an integrated circuit containing an array of linked, or coupled, capacitors.

- Complementary metal-oxide semiconductor (CMOS) is a major class of integrated circuits.

Camera

- Photovoltaic principles: The energy of a photon from a light source causes an electron to leave its valence band and changes to a conduction band. The quantity of incoming photons affects macroscopic conductivity. The excited electron is a source of electric voltage which becomes electric current. The current is directly proportional to the amount of incoming energy (photons).
- Cameras are equipped with necessary electronics to provide digitized images.
- Color cameras are similar to monochromatic ones and contain color filters.

CCD Camera

- CCD Camera is based on Charge Couple Device (CCD) technology.

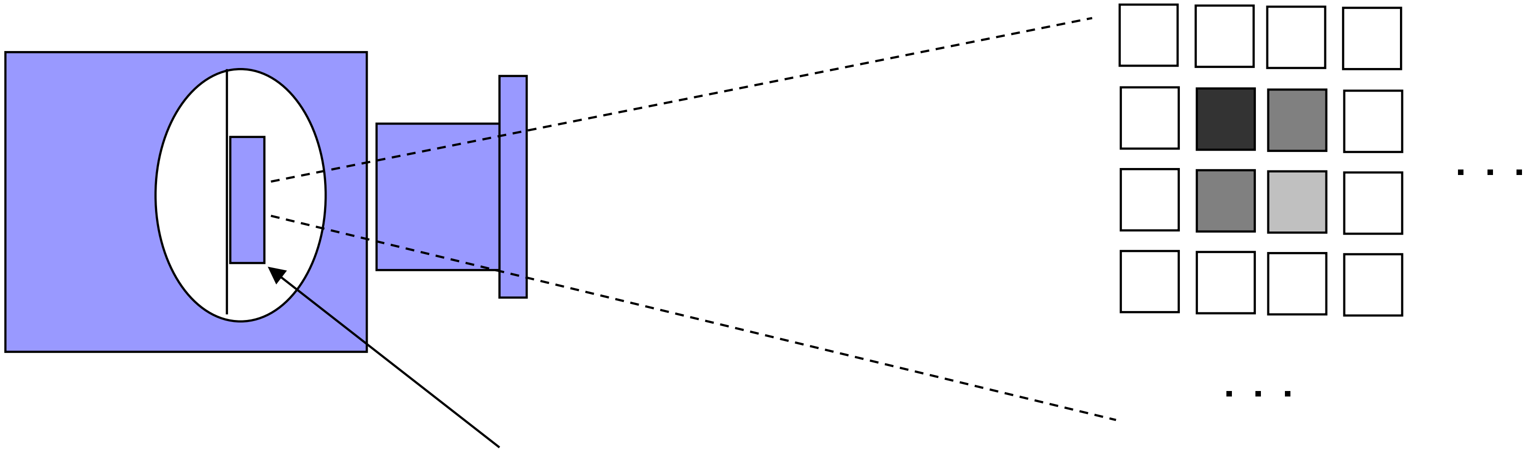
- A CCD is an analog shift register consisting of a series of closely spaced capacitors.
  - It enables analog signals (electric charges) to be transported through successive stages (capacitors) controlled by a clock signal.
  - It is used to serialize parallel analog signals in arrays of photoelectric light sensors.
  - For image capturing, there is a photoactive region made of silicon and a transmission region which is the CCD.

CCD

- Basics of image capturing operation
  - Image is projected onto the capacitor array via the photoactive region. Each capacitor will accumulate an electric charge proportional to the light intensity at the location.
  - When the array of capacitor is completely exposed to the image (exposure time), a control circuit causes each capacitor to transfer its contents to its neighbor.
  - The last capacitor in the array moves its charge to a charge amplifier which converts the charge into a voltage.
  - By repeating this process, the control circuit converts the entire semiconductor contents of the array to a sequence of voltages that are sampled, digitized and stored in some form of memory.

CCD

- Typical sensor size: 8.8 mm x 6.6 mm, almost 200,000 pixels



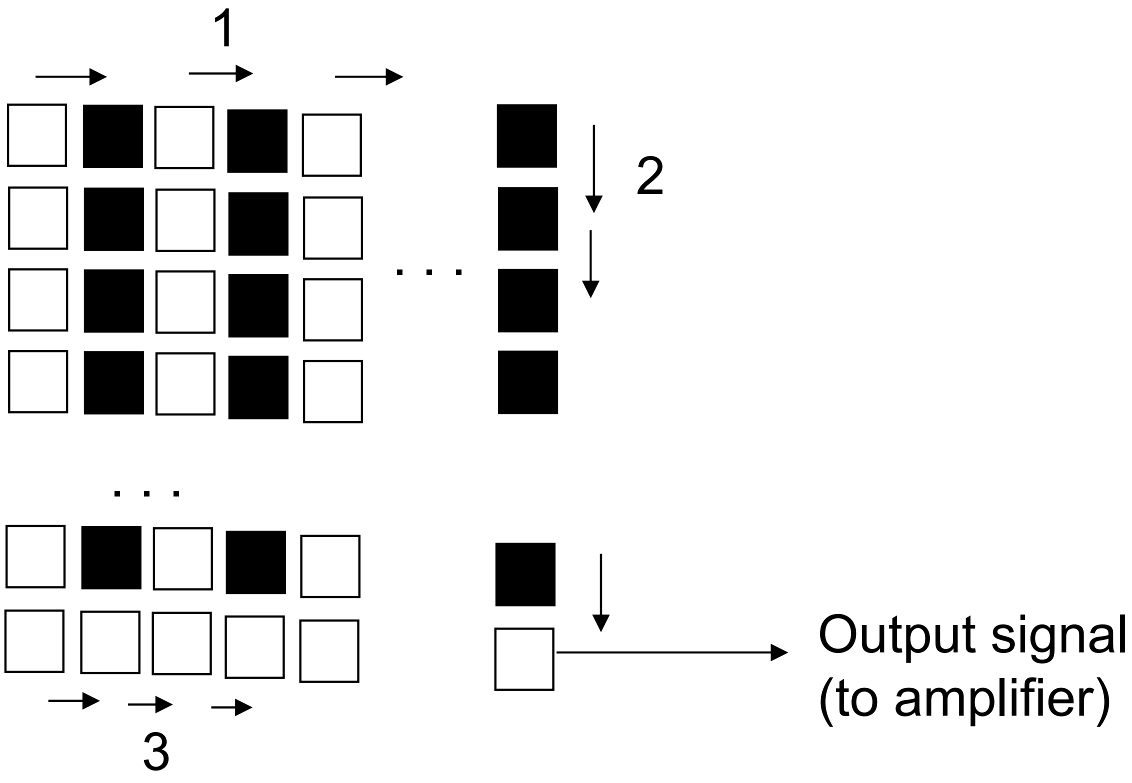
CCD sensor inside the camera

CCD

- Two methods to read out the charges that are accumulated by the capacitors of the sensor
  - Interline transfer
  - frame transfer

Interline transfer:

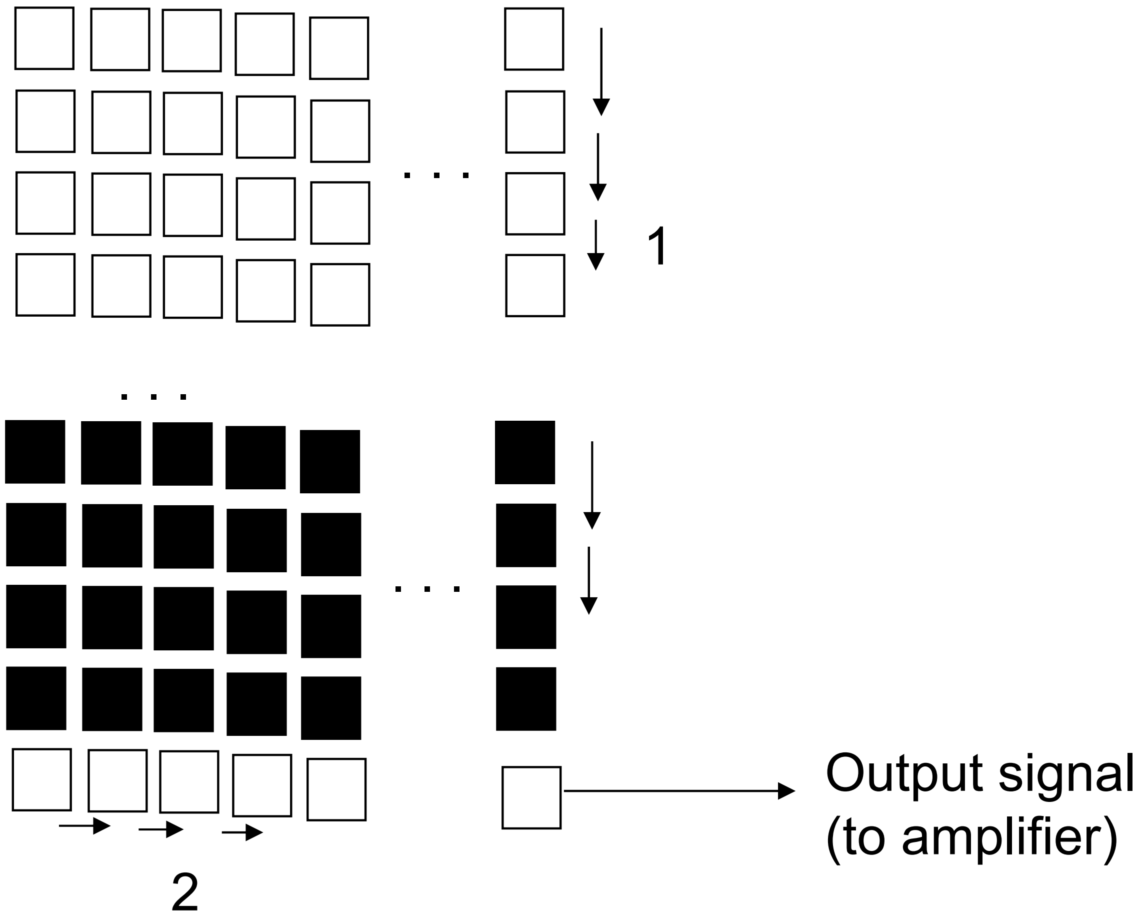
1. Charges are shifted to the shielded area.
2. The charges in the column are shifted down one cell.
3. A row of charge is then shifted out.



CCD

Frame transfer:

1. All charges are shifted down to the shielded area.
2. Each row of charge is then shifted out.



CCD

- Other common CCD sensor: 512x256 array of photoreceptors
  - A device that detects light by capturing photons.
  - Photoreceptor = photodetector = photosensor
- Customized arrangement of the receptors possible, e.g. similar to human retina
- Low geometric distortion, price depends on the quality of the photoreceptors
- Pixel transfer rate up to 20 mega pixel per second
- Voltage signal digitalized (processing) or converted to signal (direct to monitor)

Formatting camera output signal

- NTSC color television standard and PAL (Phase Alternating Line) are most widely used color encoding system.

Monochromatic Camera

Charge amplifier – converts the charge into a voltage

Automatic gain control – changes the gain of the camera according to the amount of light in the scene. g correction performs non linear transformation of the intensity scale.

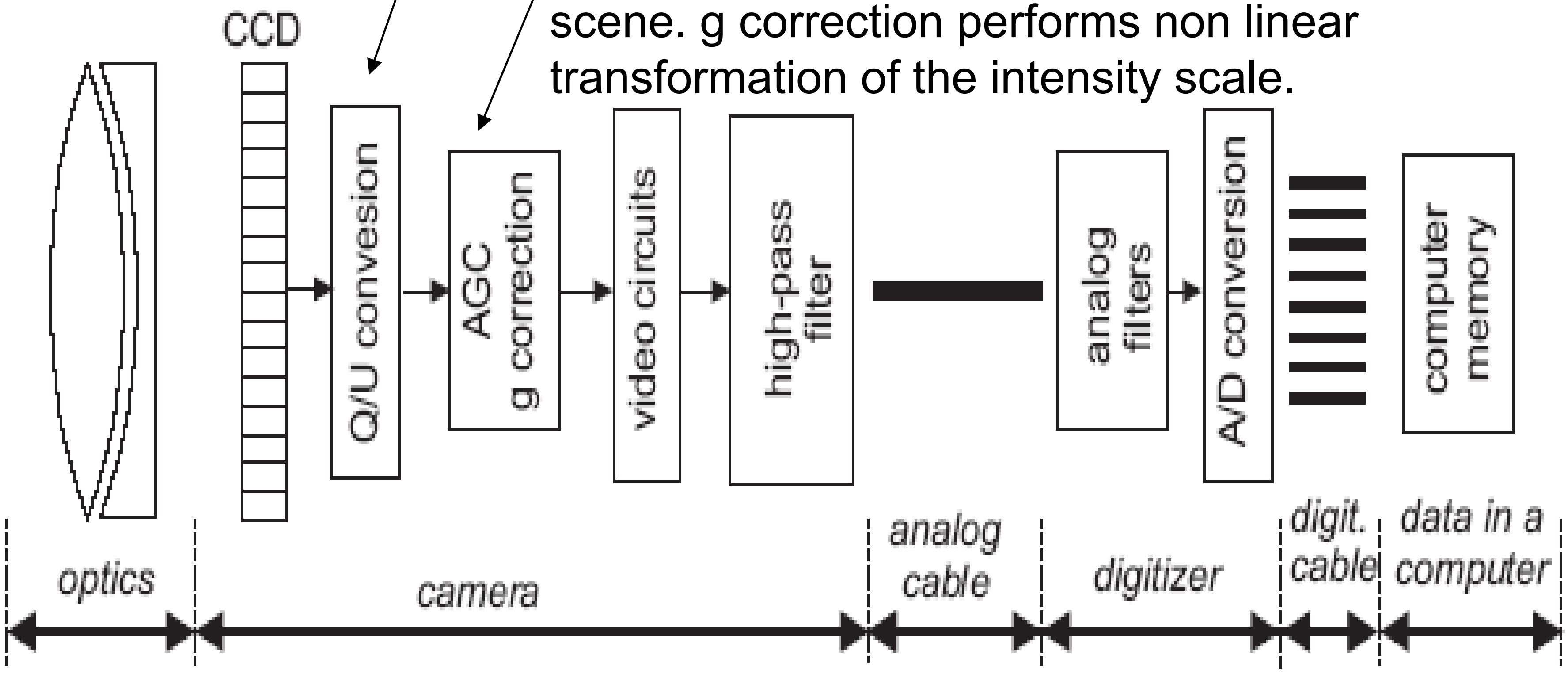


Figure 2.35: Analog CCD camera.

Monochromatic Camera

Gamma correction or encoding of images is to compensate for properties of human vision – to ensure not too many bits are allocated for highlights that human cannot see and too few bits for shadow values that human are sensitive to.

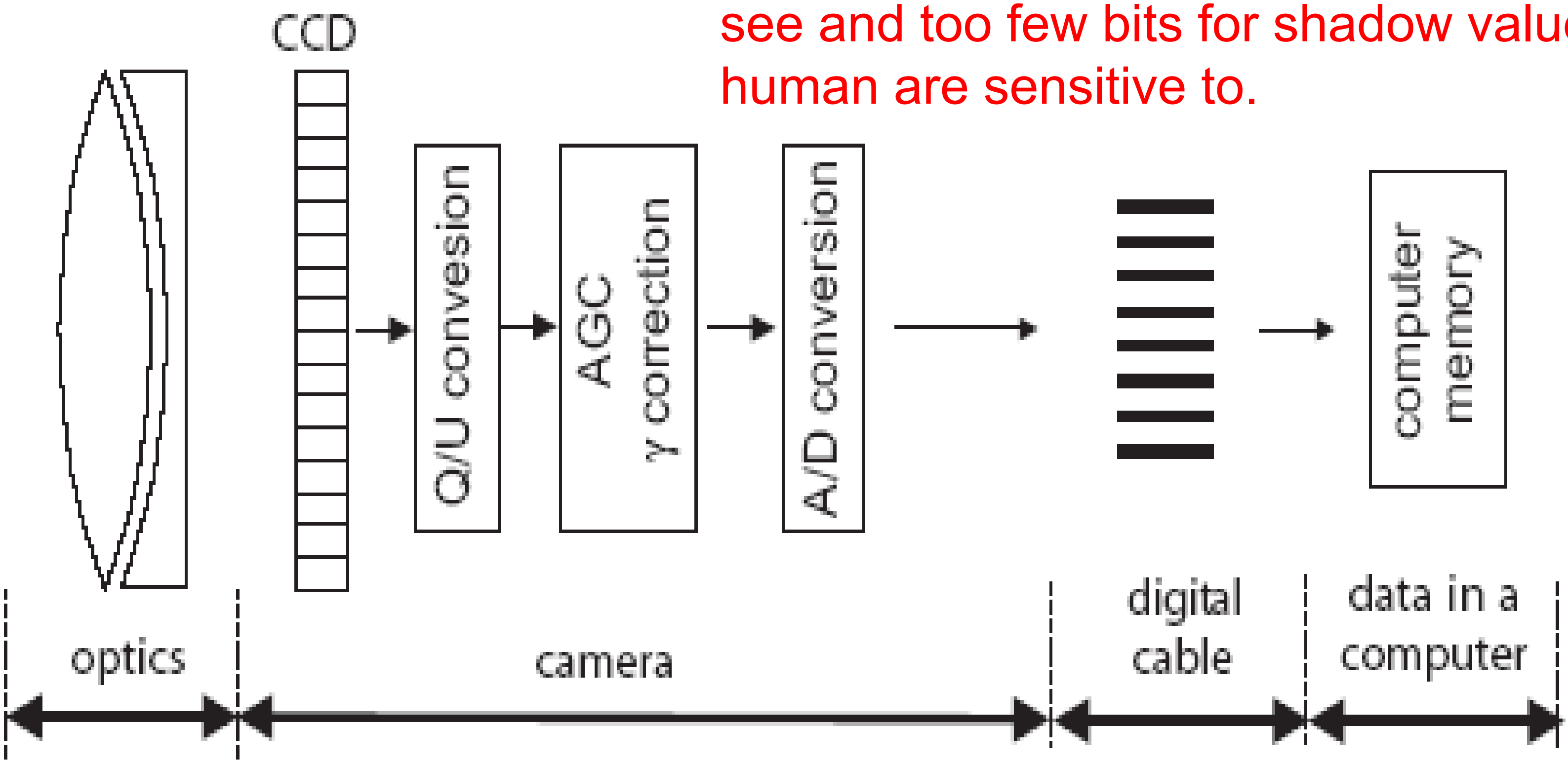
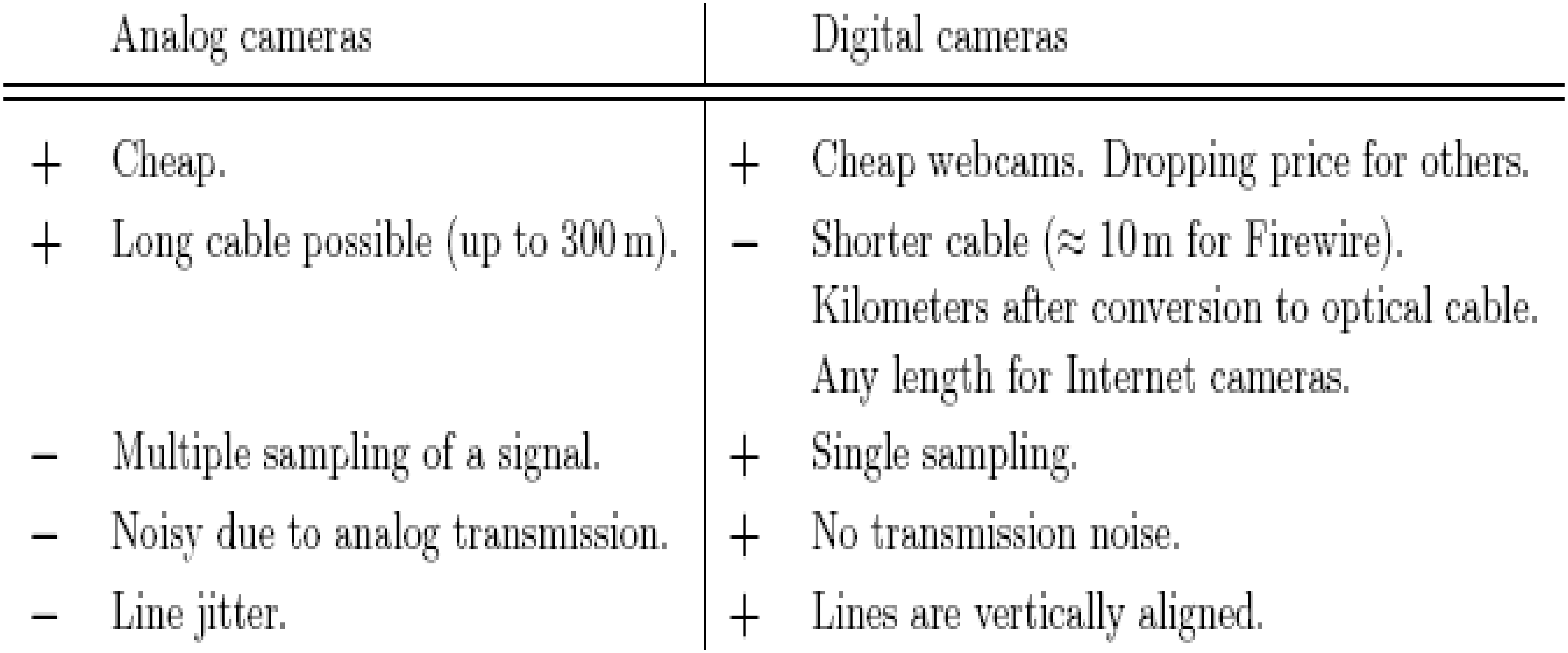


Figure 2.36: Digital CCD camera.

Monochromatic Camera

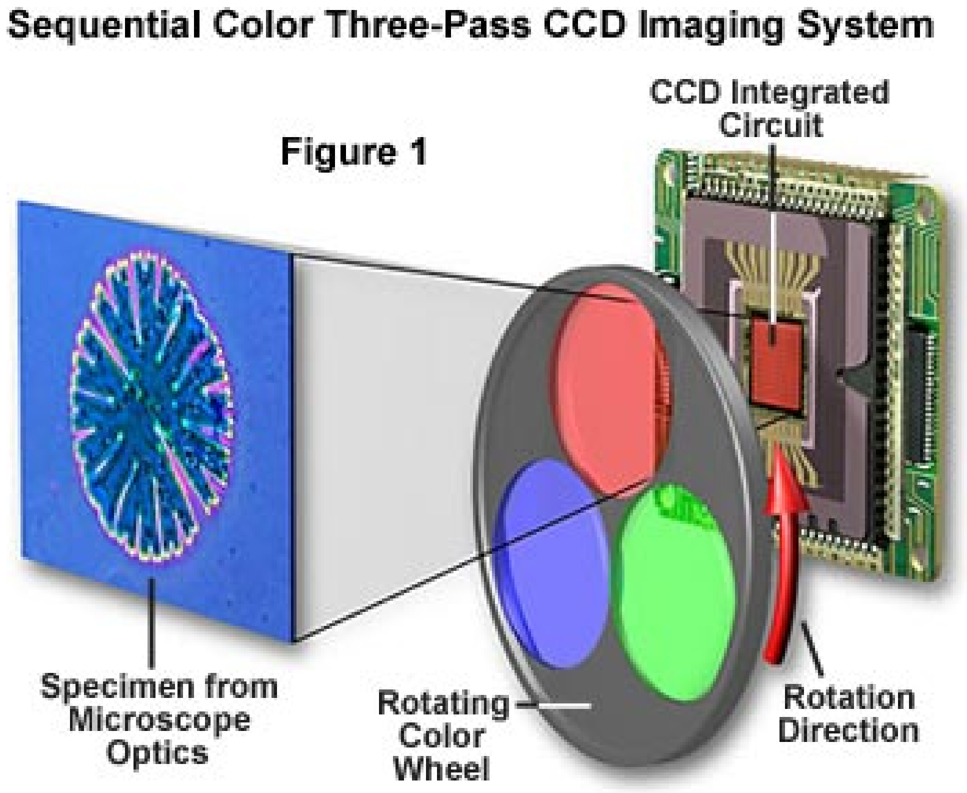


	Analog cameras		Digital cameras
+	Cheap.	+	Cheap webcams. Dropping price for others.
+	Long cable possible (up to 300 m).	–	Shorter cable (≈ 10 m for Firewire). Kilometers after conversion to optical cable. Any length for Internet cameras.
–	Multiple sampling of a signal.	+	Single sampling.
–	Noisy due to analog transmission.	+	No transmission noise.
–	Line jitter.	+	Lines are vertically aligned.

Color Camera

- Three strategies to capture color images:
  - Record three different images in succession by employing color filters in front of monochromatic camera.
    - Color sequential capture – switching colors with a color filter wheel.
  - Using a color filter array on a single sensor.
    - Integral color filter arrays (CFA) – filters of the appropriate characteristics of R,G and B are placed on the chip.
  - The incoming light is split into several color channels using a prism-like device.
    - Three-chips color – use optics to split the scene into three separate image planes and onto three CCD chips

Color sequential capture
– switching colors with a color filter wheel


Sequential Color Three-Pass CCD Imaging System
Figure 1
CCD Integrated Circuit
Specimen from Microscope Optics
Rotating Color Wheel
Rotation Direction

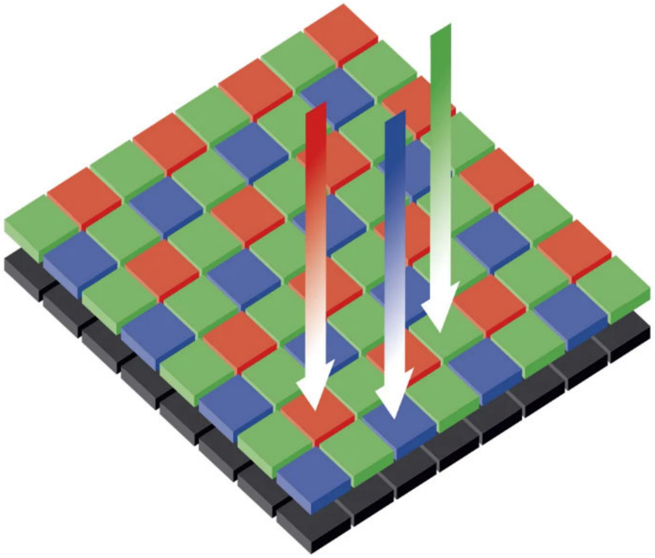
Olympus Life Science
Digital Imaging in Optical Microscopy - Concepts in Digital Imaging - Sequential Color Imaging Systems | Olympus LS

https://www.olympus-lifescience.com/es/microscope-resource/primer/digitalimaging/concepts/threepass/

Disadvantage: Mechanical complexity of the system

Integral color filter arrays (CFA)

- filters of the appropriate characteristics of R,G and B are placed on the chip

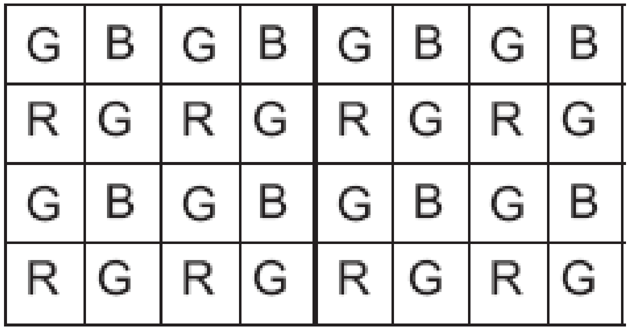


Integral color filter arrays

Used in almost all color digital cameras

Bayer filter: What is it and how does it work?
https://www.whatdigitalcamera.com/technology_guides/bayer-filter-work-60461

The property that human eye is most sensitive to green, less to red, and least to blue is used by the most common color filter for single chip camera.

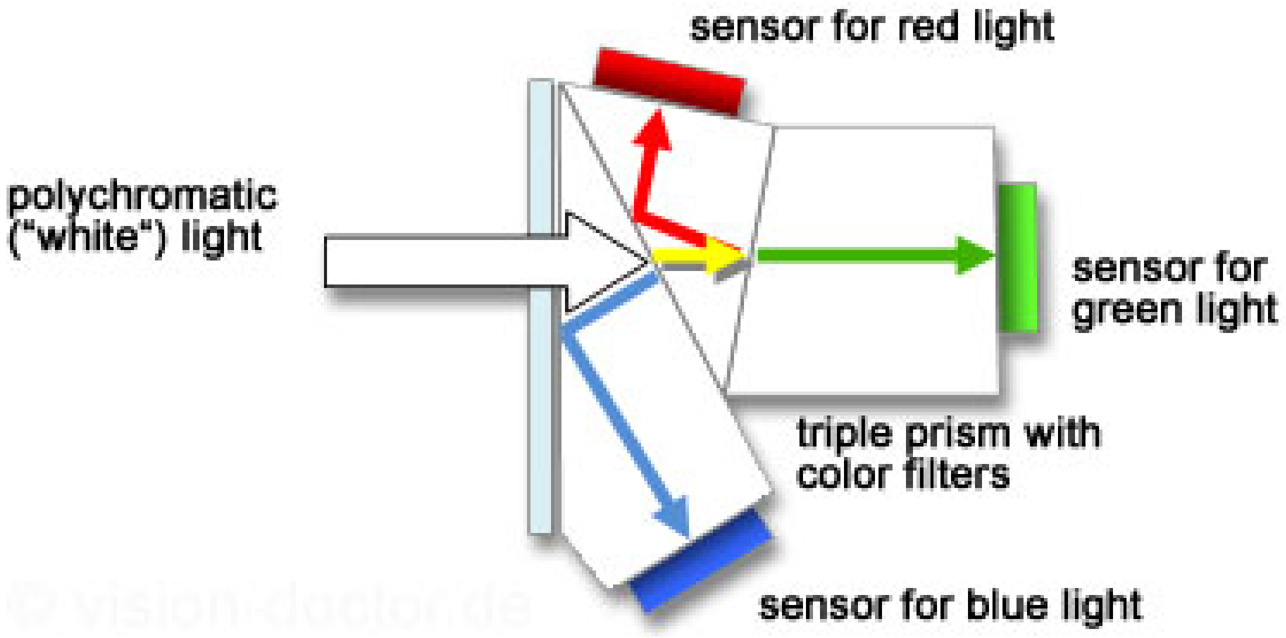


G	B	G	B	G	B	G	B
R	G	R	G	R	G	R	G
G	B	G	B	G	B	G	B
R	G	R	G	R	G	R	G

Figure 2.37: Bayer filter mosaic for single chip color cameras.

Three-chips color

– use optics to split the scene into three separate image planes and onto three CCD chips



https://www.vision-doctor.com/en/area-scan-cameras/three-chip-colour-cameras.html

Disadvantage: using smaller sensors and hence lower image resolution.

CMOS Camera

- CMOS camera system uses CMOS image sensors or CMOS active pixel sensor (APS). CMOS (Complementary metal-oxide-semiconductor) is a class of integrated circuit.

- An APS is an image sensor consisting of an integrated circuit containing an array of pixel sensors, each pixel containing a photodetector and an active amplifer.

- Most commonly used in mobile phone cameras and web cameras.

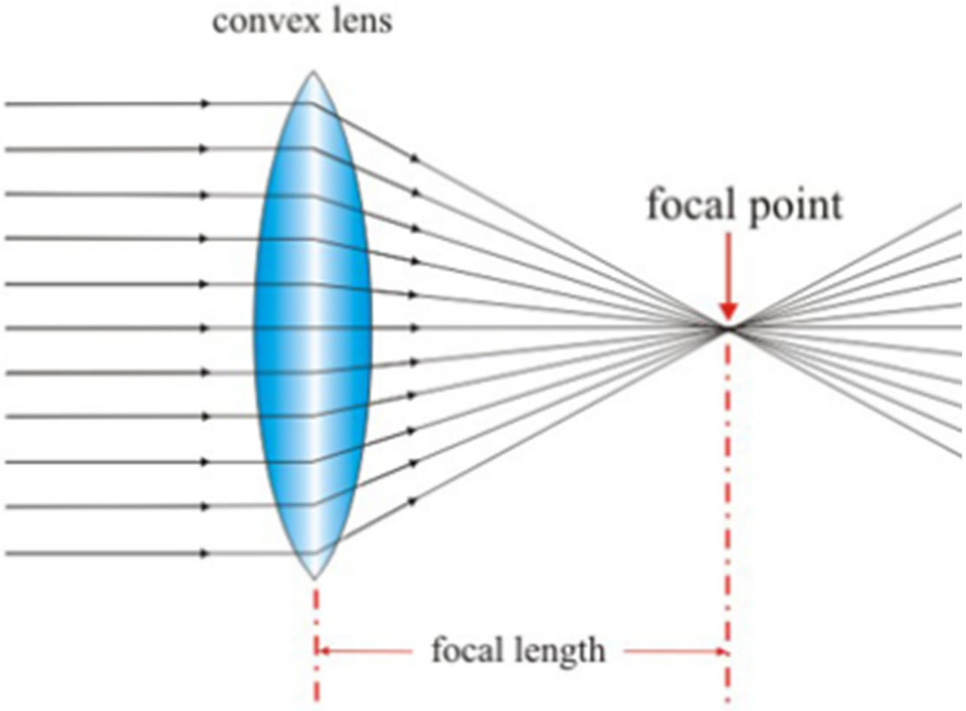
Smart Camera

- A smart camera is an integrated machine vision system which, in addition to image capture circuitry, includes a processor, which can extract information from images without need for an external processing unit, and interface devices used to make results available to other devices.

- It is an embedded system usually with CMOS image sensor.

Aperture and Focal Point

- Aperture is an opening through which light is admitted. It controls the amount of light that passes through the camera lens.
- Focal point is a point onto which collimated light parallel to the axis is focused.
  - Collimated light is light whose rays are nearly parallel.



Focal length is the distance between the optical center of a lens and the focal point at which parallel rays of light converge or appear to converge after passing through the lens.

Focal Length

- *f* is a measure of how strongly an optical system converges or diverges light.
- A system with a shorter focal length has greater optical power than one with a long focal length.
- From the focal length *f*, the distance D_o at which the object is sharp can be computed.
- Example: *m*=0.05, *f*=35.7mm, D_o=750 mm
  - For a lens system with *m*=0.05, distance between object and lens (camera) is 750 mm, we need a focus length *f* = 35.7 .

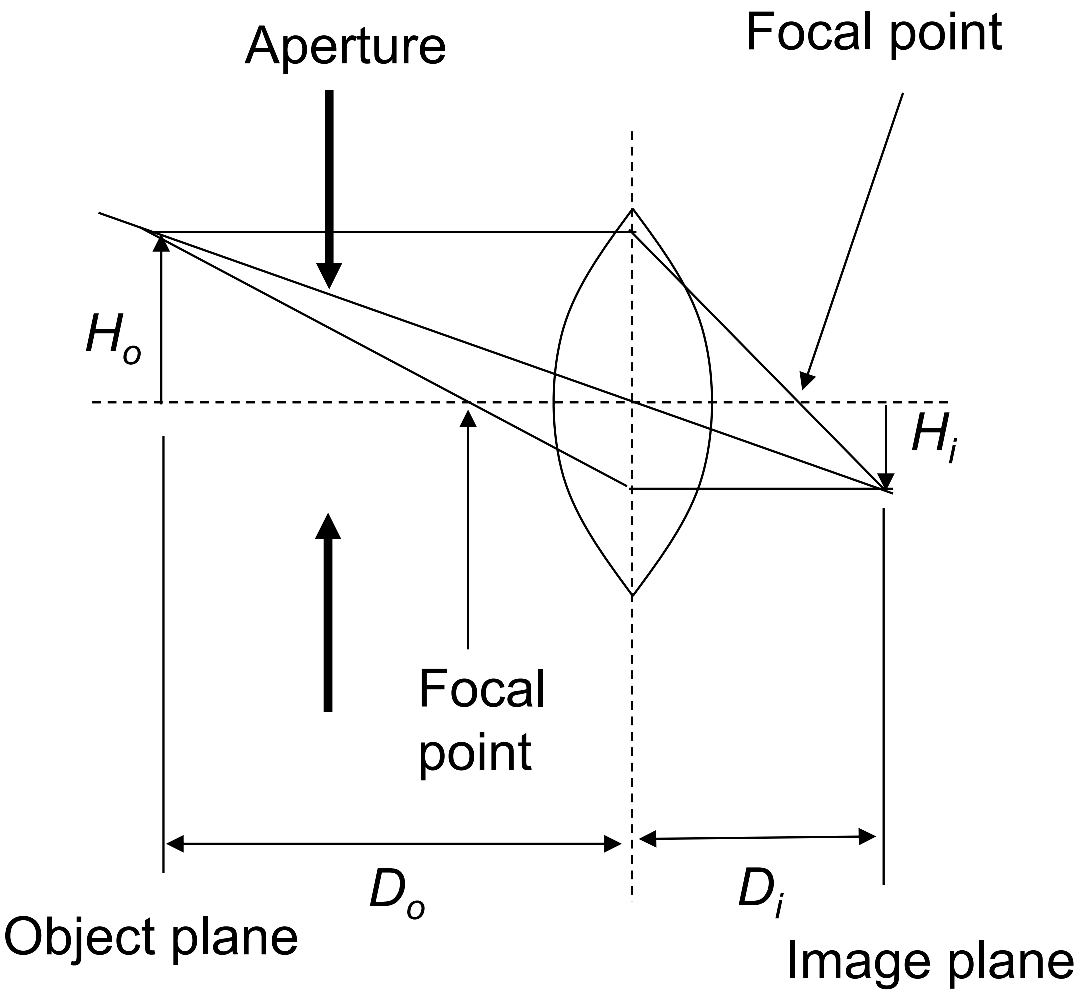
p = optical power (or refractive power) = 1/*f*

$$\boxed{D_o = f(1+\frac{1}{m})}$$

Magnification

- m = magnification
  - H_i*: image size*
  - H_o*: object size*
- m << 1: macroworld
  - typical industrial automation systems
- m >> 1: microscope

$$\boxed{m = \frac{H_i}{H_o} = \frac{D_i}{D_o}}$$



Given a 100 x 100 array of photoreceptors to acquire an image of size 4 x 4 mm of an object of size 80 x 80 mm. What are the magnification, pixel size and pixel size on the object?

Pixel size on array is the distance between sensor elements, and is given by dividing the dimension of array by the number of elements in the same direction.

$$\text{magnification} = \frac{\text{image}}{\text{object}} = \frac{4}{80} = \frac{1}{20}$$

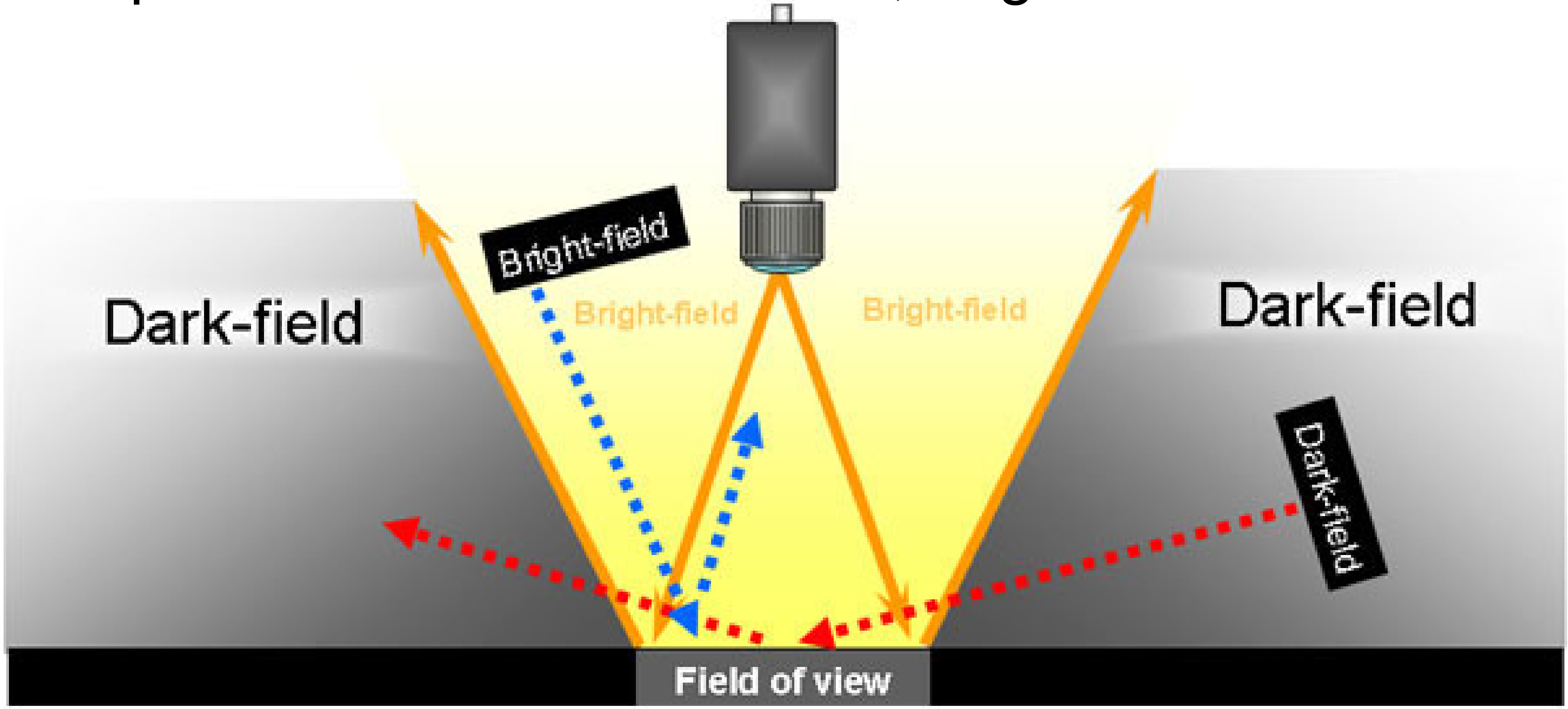
$$\text{pixel size} = \frac{4}{100} = 0.04\text{mm}$$

$$\text{pixel size on object} = \text{pixel size} \times \frac{\text{object}}{\text{image}} = 0.04 \times \frac{20}{1} = 0.8\text{mm}$$

Field of View

The field of view or field of vision (FOV) of a sensor or camera is the size of the scene that the sensor can sense or the camera can observe.

Examples: FOV = 10 cm x 10 cm; Angular FOV = 50^o x 40^o



Source: www.microscan.com

Depth of Field

- A camera focuses its lens at a single point but there is a zone that stretches in front of and behind this focus point that still appears sharp. This zone is known as the depth of field.

Depth of field (DoF) is the range of distances in a photograph where objects appear sharp and in focus. It is influenced by the size of the aperture.

Depth of Field

Depth of field is the portion of a scene that appears sharp in image.

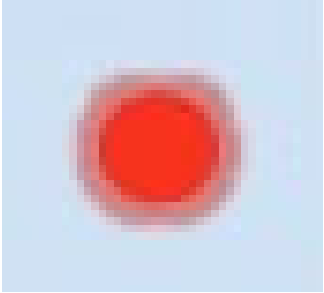
Circle of confusion is an optical spot caused by a cone of light rays from a lens not coming to a perfect focus when imaging a point source

$$\text{depth of field} = \frac{2\,\alpha\ f/\ (m+1)}{m^2}$$

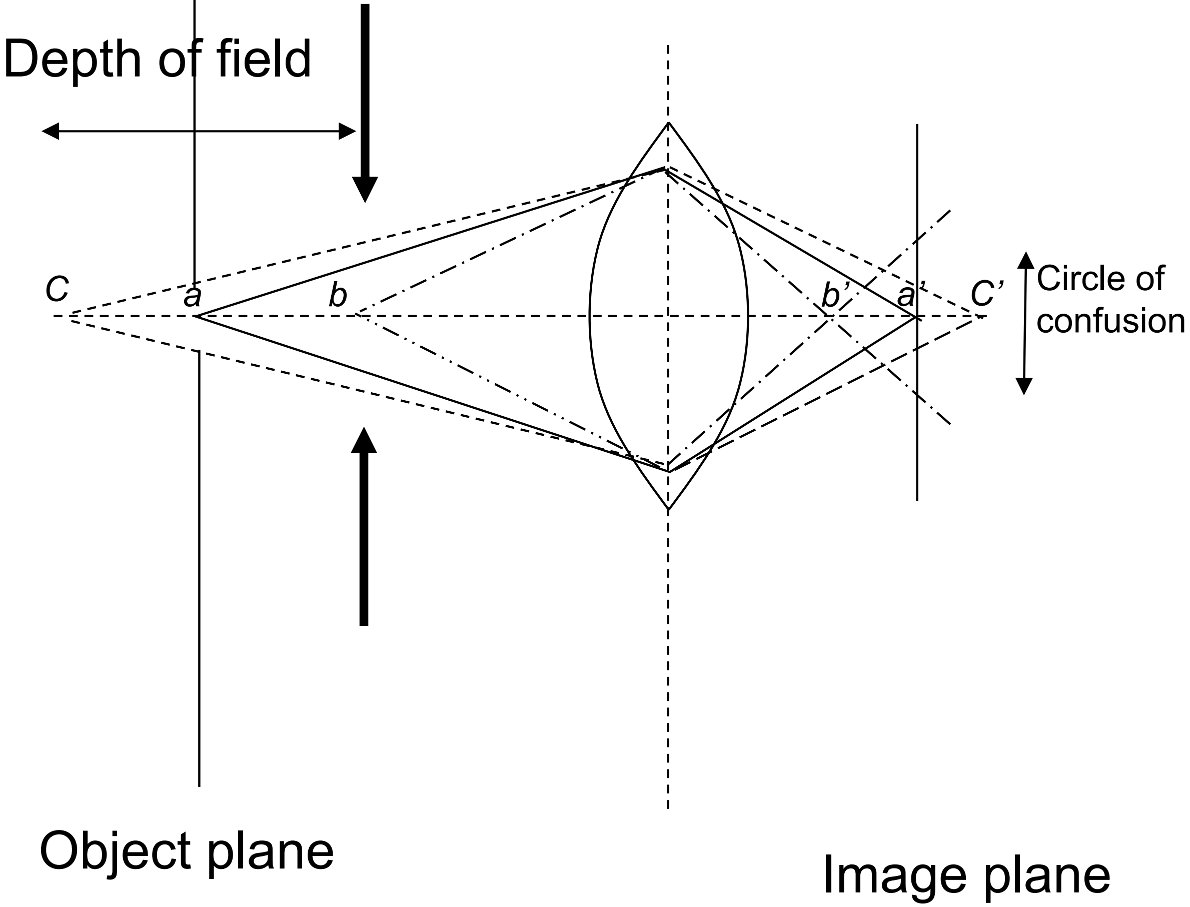
α : pixel size

$f/$: aperture size(f – stop)

m : magnification



Aperture (f stop)



Depth of Field

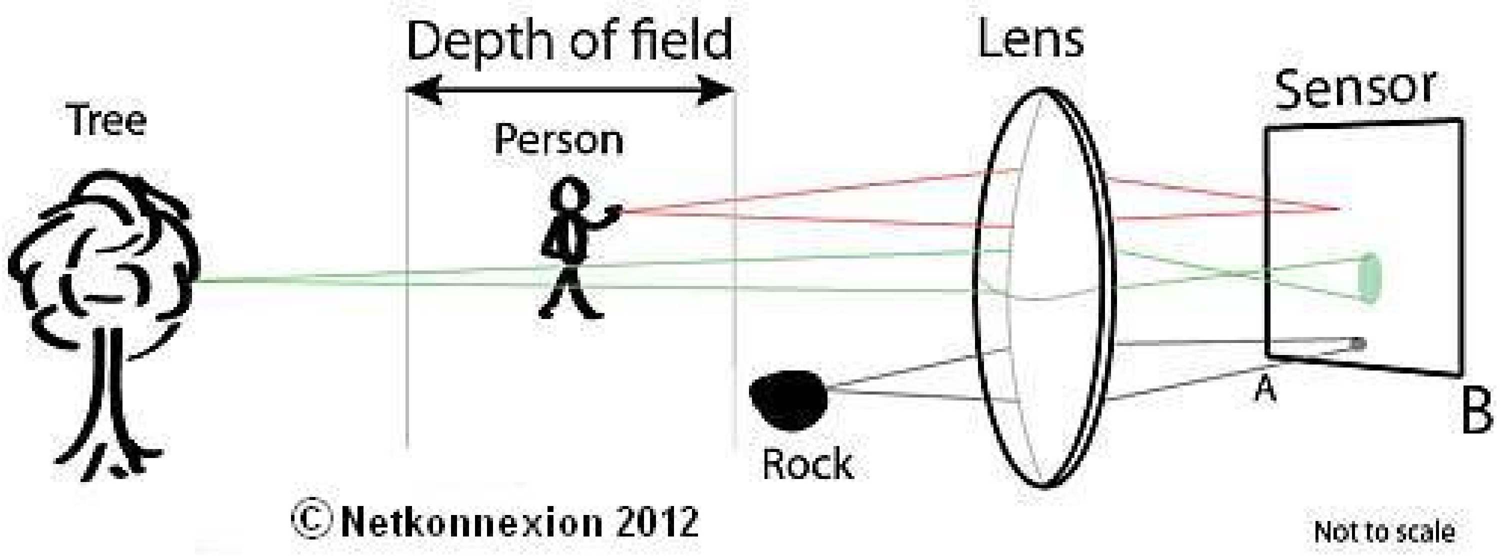


Diagram showing various sizes of Circles of Confusion (CoC). They are sized according to the focus (not to scale). Only those CoCs projected from inside the Depth of Field are sharp. Our eyes cannot perceive them well as they form sharp points. The ones we do see are projected from outside the depth of field. When we are able to see the CoC it is said to be unsharp.

Source: http://www.photokonnexion.com/?page_id=4373

Depth of Field

- f-stop is the discrete step the f-number is adjusted in a camera.
- f-number is the focal length divided by the effective aperture diameter.
- Standard f-stop scale: f/1, f/1.4, f/2, f/2.8, f/4, f/5.6, f/8, f/11, f/16, f/22, f/32, …

f/4: f-number = 4

200 mm f/4 lens: aperture diameter = 200/4 = 50 mm and 200 mm/50 mm = 4.

Depth of Field

Given f=40 mm and f-stop = 8, what are the diameter of aperture opening and f-number?

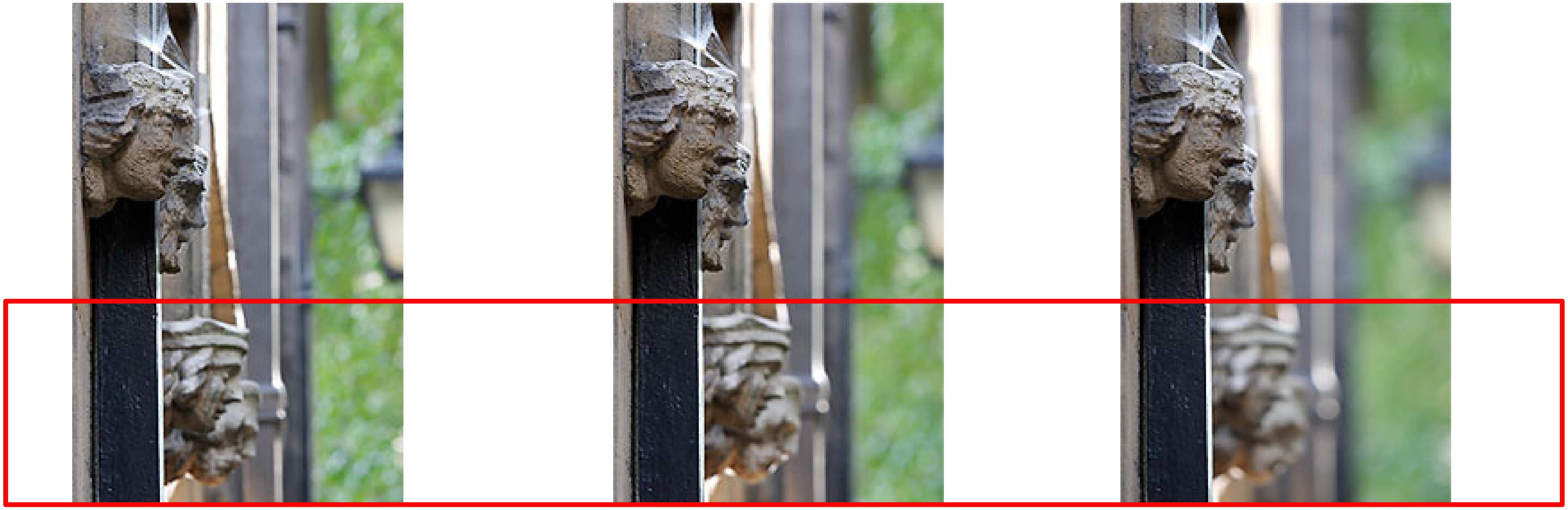
- Diameter of aperture opening (or effective aperture diameter) = f/f-stop = 40/8 = 5 mm
- f-number = f/(effective aperture diameter) = 40 mm/5 mm = 8

The aperture is the opening in the camera lens through which light enters. The size of the aperture can be adjusted to control the amount of light that reaches the camera's sensor. A larger aperture (smaller f-number, e.g., f/1.8) results in a shallow depth of field, where only a small portion of the scene is in focus, and the background appears blurred. A smaller aperture (larger f-number, e.g., f/16) creates a deeper depth of field, with more of the scene appearing sharp from foreground to background.

“Smaller F-stop number (Larger apertures) produce a shallower depth of field.”

Source: http://www.cambridgeincolour.com/tutorials/depth-of-field.htm

Image taken on a 200 mm lens



f/8.0

f/5.6

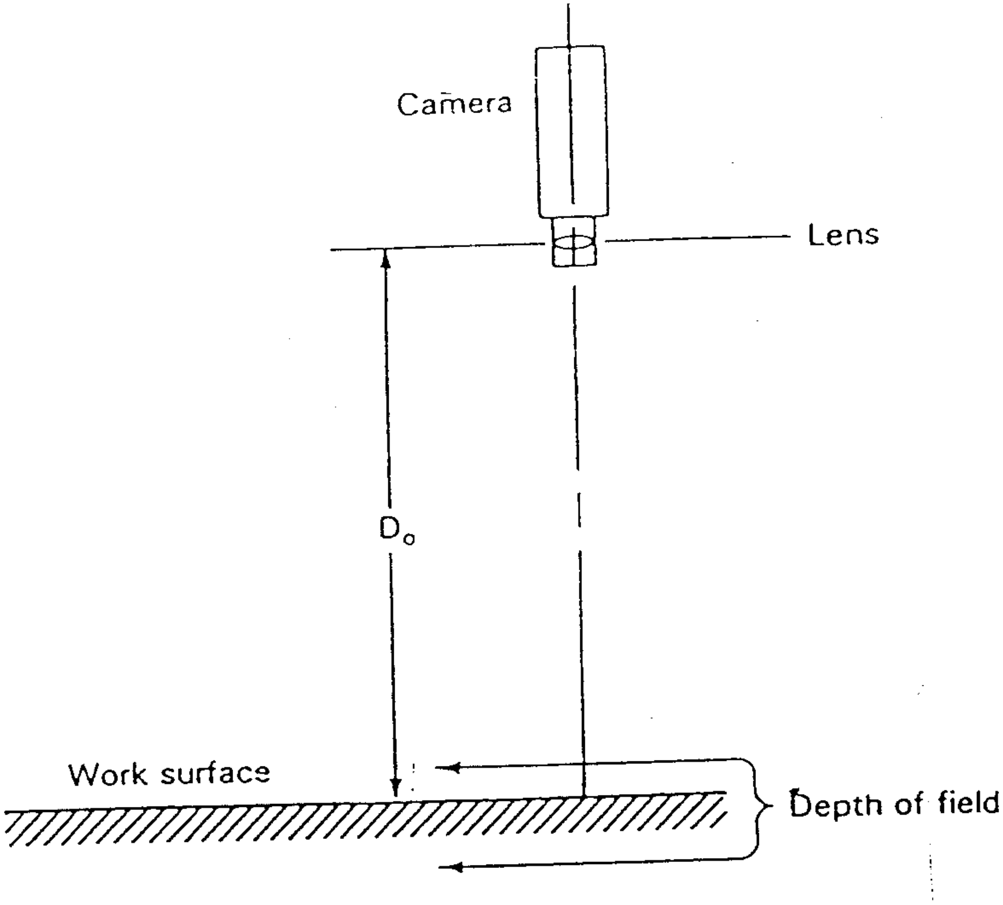
f/2.8

Depth of field is ‘deep’ – more of the picture appears sharp.

Depth of field is ‘shallow’ – a narrow zone appears sharp.

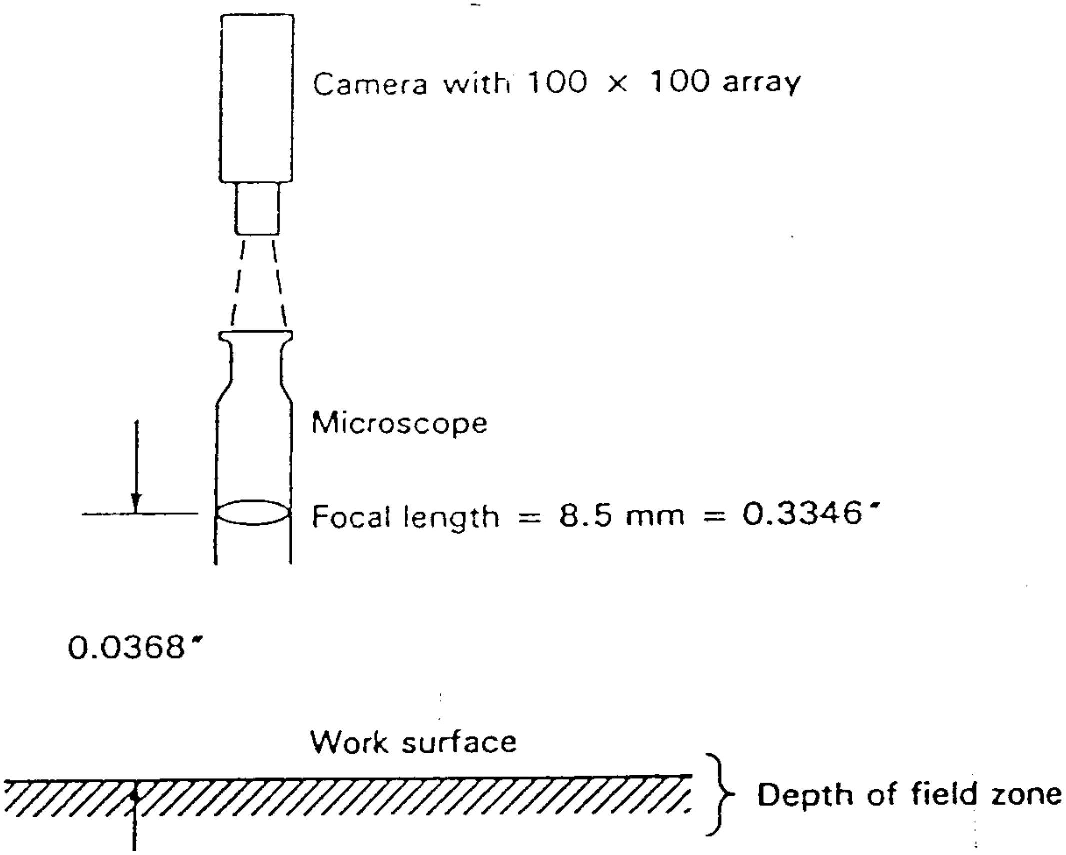
Depth of Field: Macroworld

- 7.6 x 7.6 mm size 200 x 200 array sensor
- f-stop = 16, magnification = 1/20
  - pixel size = 7.6/200 = 0.038 mm
  - depth of field = (2*0.038*16*(1+1/20))* 20^2 = 511 mm)
- In typical industrial applications (m << 1), the depth of field problem is not sufficiently important.

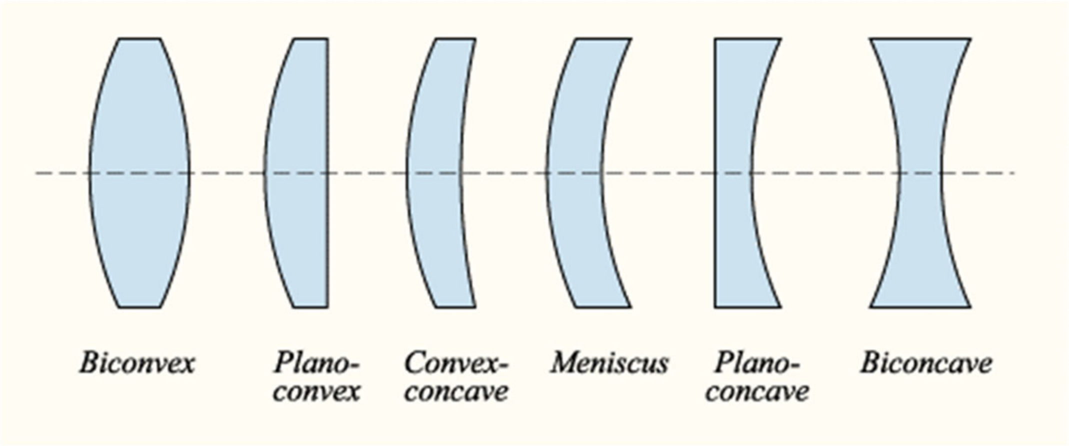


Depth of Field: Optical Microscope

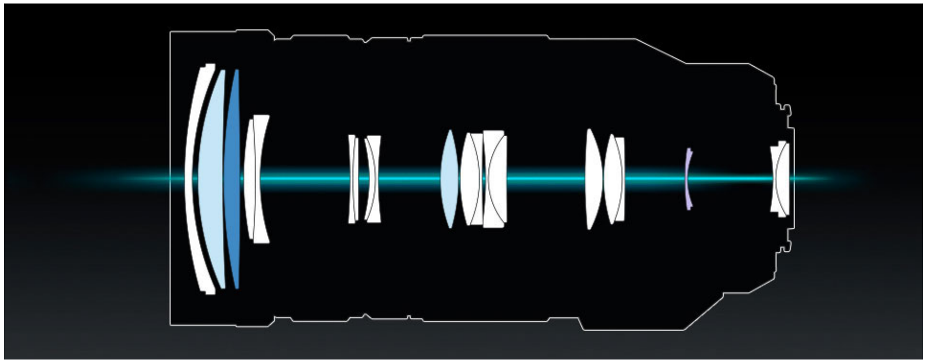
- 3.8 x 3.8 mm sensor array
  - Pixel size = 3.8/100 = 0.038 mm
  - If m = 100 and f-stop=16
    - Depth = (2*0.038*16*(1+100))/10000 = 12 micrometers
- In microscopy, the depth is limited, 3D vision is hardly possible.



There are many different types of lenses.



Panasonic super telephoto zoom – an assembly of lens



There is a need to have a relatively simple method to design complex lens system.
- Ray transfer matrix / ray matrix / optical system matrix according to Geometrical Optics.

Understanding Focal Length – Nikon USA

https://www.nikonusa.com/en/learn-and-explore/a/tips-and-techniques/understanding-focal-length.html#:~:text=The%20focal%20length%20of%20a,and%20the%20higher%20the%20magnification.

“Lens focal length tells us the **angle of view**—how much of the scene will be captured—and the **magnification**—how large individual elements will be. The longer the focal length, the narrower the angle of view and the higher the magnification. The shorter the focal length, the wider the angle of view and the lower the magnification.”

Zoom (variable focal length) vs. Prime Lens (fixed focal length)

FX format is full-frame sensor equivalent in size to a 35 mm frame; DX format is a smaller size sensor and hence the field of view is narrower.

Wide-angle lens: FX – 14-35 mm; DX – 10-24 mm
Standard lens: FX – 50-60 mm; DX – 35 mm
Telephoto lens: FX – 70-200 mm; DX – 55-200 mm
Super Telephoto lens: FX – 300-600 mm; DX – 200-600 mm
Macro lens: FX – 60, 105 and 200 mm; DX – 85 mm

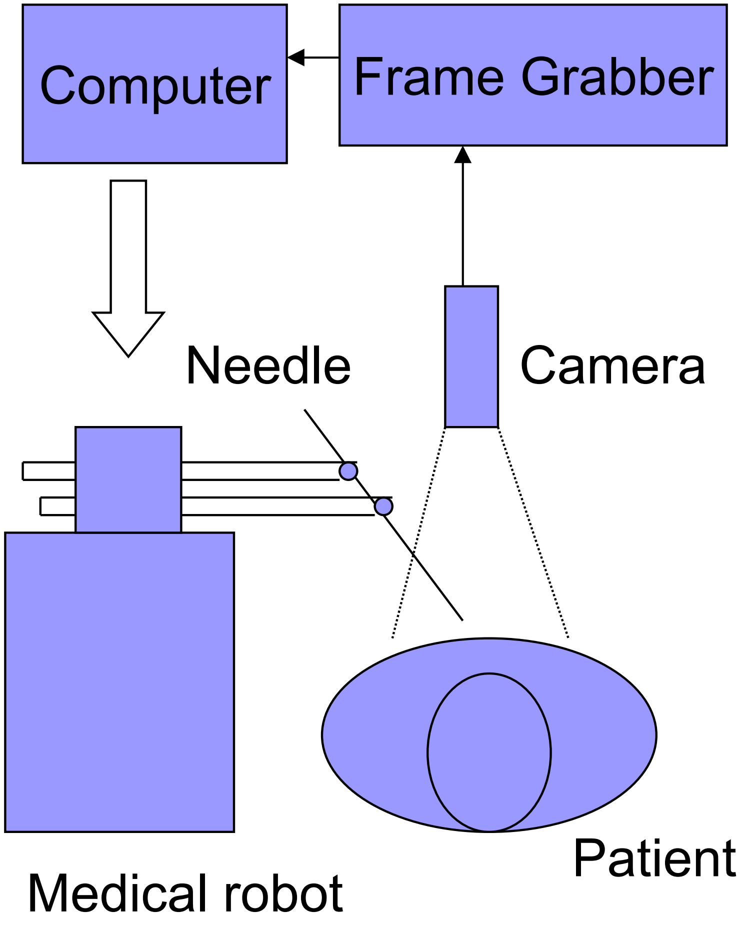
4. Camera Calibration

- Camera calibration is the process of estimating the intrinsic and extrinsic parameters of a camera.

- This is essential for 3D computer vision tasks such as 3D reconstruction, augmented reality, and robot vision.

Transformation from Object Space to Image Space

- Object in world coordinate, image in image coordinate
- Example: image guided robotic needle insertion
    - Task: to perform multiple needle insertions into organ according to perspective projection image of the organ
    - The targets in world coordination are captured by camera and stored as images in image coordinate.
        - Assumption: all targets are on the same plane.
    - The robot has to insert the needle into the target precisely in world coordinate.



Calibration: Transformation Between 2D Object and 2D Image Coordinates

$$\text{image}: \begin{bmatrix} x \\ y \end{bmatrix}, \quad \text{object}: \begin{bmatrix} X \\ Y \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathrm{PRT} \begin{bmatrix} X \\ Y \end{bmatrix}$$

T : translation, R : rotation, P : perspective projection

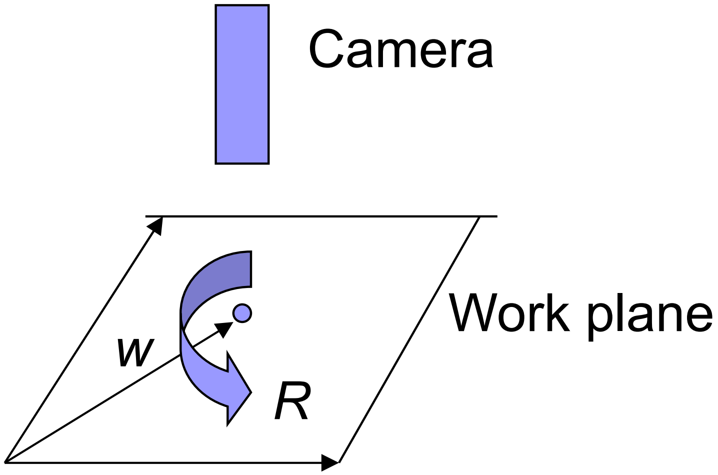
$$\mathrm{T}(-w): \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} -v_x \\ -v_y \end{bmatrix} = \begin{bmatrix} X' \\ Y' \end{bmatrix}$$

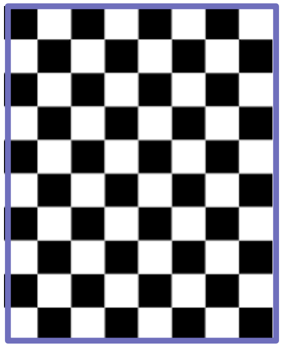
$$\mathrm{R}(-\alpha): \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} X' \\ Y' \end{bmatrix} = \begin{bmatrix} X'' \\ Y'' \end{bmatrix}$$

$$\mathrm{P}: -\beta \begin{bmatrix} X'' \\ Y'' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Rightarrow T: \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

a_{ij} : transformation parameters





Checkerboard for camera calibration

Calibration is the process of identifying the transformation parameters

Calibration

a

1. $$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} X & Y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X & Y & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix}$$

X (object)

2. Identify *N* image points that corresponds to *N* locations on the object space.

x (image)

3. $$\begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ \\ x_N \\ y_N \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_1 & Y_1 & 1 \\ X_2 & Y_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_2 & Y_2 & 1 \\ & & & & & \\ X_N & Y_N & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_N & Y_N & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix}$$

Least squares solution or least square approximation

4. $\mathbf{a} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{x}$ is the solution to the minimal error $\|\mathbf{x} - \mathbf{X}\mathbf{a}\|^2$

Least Squares Approximation

- Suppose a system *ax* = *b* has *N* equations, *N* > 1, in only one unknown. (Equation 1.1)
- One possibility of solving this inconsistent equation is to determine *x* from a part of the system, and ignore the rest.
- A more reasonable method is to determine *x* that minimizes the average error in the *N* equations.
- Sum of squares is a popular method to define average errors (Equation 1.2)
- If there is an exact solution, minimum *E*=0.
- If *b* is not proportional to *a*, the function *E*2 is a parabola with its minimum at the point defined by Equation 1.3.
- Solving for x, Equation 1.4 is the least squares solution of the system

$$\begin{cases} 2x = b_1 \\ 3x = b_2 \\ 4x = b_3 \end{cases} \Rightarrow \mathbf{a} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \qquad (1.1)$$

$$E^2 = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2 \qquad (1.2)$$

$$\frac{dE^2}{dx} = 2\left[(2x - b_1)2 + (3x - b_2)3 + (4x - b_3)4\right] = 0 \qquad (1.3)$$

$$x = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} \qquad (1.4)$$

Least Squares Approximation

- General formula: given any *a* not equal to 0 and any *b*, the error *E* is the length of the vector *ax-b* (Equation 1.5)

- The parabola (Equation 1.6) has a minimum at the point defined by Equation 1.7.

- The least squares solution to *ax=b* can be determined using Equation 1.8.

$$E = \|\mathbf{a}x - \mathbf{b}\| = [(a_1 x - b_1)^2 + \cdots + (a_N x - b_n)^2]^{\frac{1}{2}} \qquad (1.5)$$

$$E^2 = (\mathbf{a}x - \mathbf{b})^T(\mathbf{a}x - \mathbf{b}) = \mathbf{a}^T\mathbf{a}x^2 - 2\mathbf{a}^T\mathbf{b}x + \mathbf{b}^T\mathbf{b} \qquad (1.6)$$

$$\frac{dE^2}{dx} = 2\mathbf{a}^T\mathbf{a}x - 2\mathbf{a}^T\mathbf{b} = 0 \qquad (1.7)$$

$$\mathbf{a}^T\mathbf{a}x = \mathbf{a}^T\mathbf{b} \Rightarrow x = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}} \qquad (1.8)$$

Least Squares Approximation

- The least squares solution is given by Equation 1.9.
- If the matrix A^{T}A is invertible, Equation 1.10 is the unique least squares solution.

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{b} \qquad (1.9)$$

$$\mathbf{x} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} \qquad (1.10)$$

Least Squares Approximation

- Suppose we do a series of experiments, and expect the output *y* to be approximated by a linear function of *t*, *y*=*C*+*Dt*.

- Example, measure the strain *y* of a linear elastic material subjected to a load *t*. If there is no error, only two measurement of (*y*,*t*) are sufficient. But there will be experimental errors, a series of more than two sets of (*y*,*t*) measured from the experiments are unlikely to fall on a straight line. We need to “average” all the experiments to determine the optimal line.

Least Squares Approximation

From the experimental results,

$$\begin{aligned} y_1 &= C + Dt_1 \\ y_2 &= C + Dt_2 \\ &\vdots \\ y_N &= C + Dt_N. \end{aligned} \Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_N \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} \Rightarrow \mathbf{b} = \mathbf{Ax}$$

The best solution in the least squares sense is the one that minimizes

the sum of the squares of the errors

$\Rightarrow$ We want to choose C and D to minimize

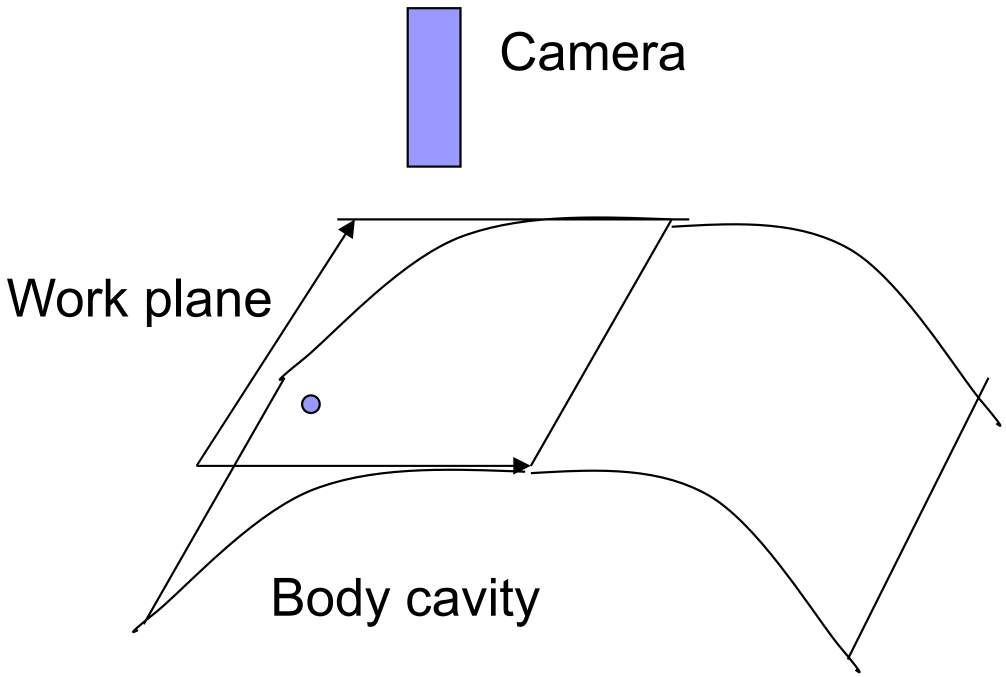
$$E^2 = \|b - Ax\|^2 = (y_1 - C - Dt)^2 + \cdots + (y_N - C - Dt_N)^2$$

$\Rightarrow$ We choose $\mathbf{x}$ so that $\mathbf{p} = \mathbf{Ax}$ is as close as possible to $\mathbf{b}(E^2 \to 0)$.

The straight line $y = \overline{C} + \overline{D}t$ which minimizes $E^2$is the least squares solution of $\mathbf{Ax} = \mathbf{b}$.

$$\mathbf{A}^{\mathrm{T}}\mathbf{Ax} = \mathbf{A}^{\mathrm{T}}\mathbf{b} \Rightarrow \mathbf{A}^{\mathrm{T}}\mathbf{A}\begin{bmatrix} \overline{C} \\ \overline{D} \end{bmatrix} = \mathbf{A}^{\mathrm{T}}\mathbf{b} \Rightarrow \begin{bmatrix} \overline{C} \\ \overline{D} \end{bmatrix} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} \Rightarrow \begin{bmatrix} \overline{C} \\ \overline{D} \end{bmatrix} = \begin{bmatrix} N & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i \\ \sum t_i y_i \end{bmatrix}$$

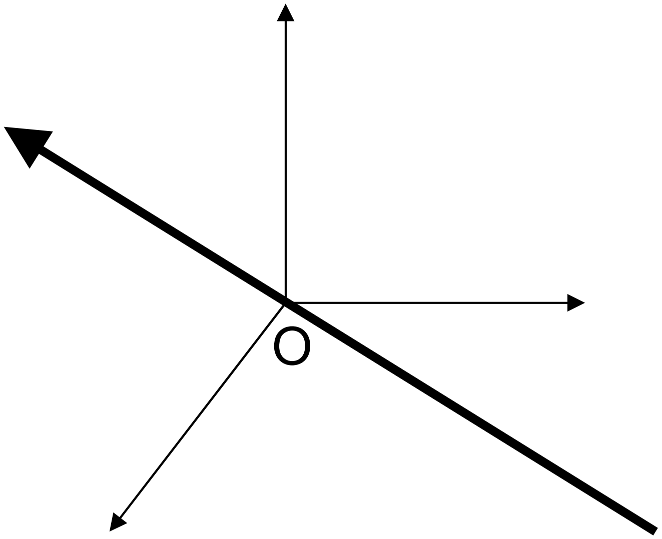
- In practice, the assumption in the example of image guided medical robot that all target points are on a single horizontal work plane may not hold.



A typical 2D image from a single camera does not have sufficient information to recover the 3D information of the object.

Homogeneous Coordinates

- A homogeneous vector corresponds to a straight line through the origin.



$$\begin{bmatrix}x\\y\\z\end{bmatrix} \leftrightarrow \begin{bmatrix}kx\\ky\\kz\\k\end{bmatrix}_h , \qquad k \neq 0$$

Cartesian coordinate

Homogeneous coordinate

Scale factor

Advantages of Homogeneous Coordinates

- Translation and perspective transformation can be expressed in matrix form:
  - **Unified matrix representation: v' = T v**
  - **v: column vector containing the coordinate**
  - **T: 4x4 Homogeneous Transformation Matrix (HTM)**
- Consecutive geometric transformation can be expressed as series of matrices with matrix multiplication

Image and Object Points in Homogeneous Coordinates

$$\text{image point}: \quad \mathbf{u}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \Rightarrow \hat{\mathbf{u}}_i = \begin{bmatrix} kx_i \\ ky_i \\ k \end{bmatrix}$$

$$\text{object point}: \quad \mathbf{X}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} \Rightarrow \hat{\mathbf{X}}_o = \begin{bmatrix} kx_o \\ ky_o \\ kz_o \\ k \end{bmatrix}$$

To convert from *n* x 1 vector in homogeneous coordinates to coordinate of dimension *n*-1, we have to divide by the *n*th element and then delete the *n*th component.

Translation in Homogeneous Coordinates

$$\begin{bmatrix}x\\y\\z\end{bmatrix}_{\text{new}} = \begin{bmatrix}x\\y\\z\end{bmatrix} + \begin{bmatrix}t_x\\t_y\\t_z\end{bmatrix}$$

In homogeneous coordinates:

Using unified matrix representation

$$\begin{bmatrix}x\\y\\z\\1\end{bmatrix}_{\text{new}} = \begin{bmatrix}1&0&0&t_x\\0&1&0&t_y\\0&0&1&t_z\\0&0&0&1\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix}$$

fake coordinate

Translation as invertible matrix:

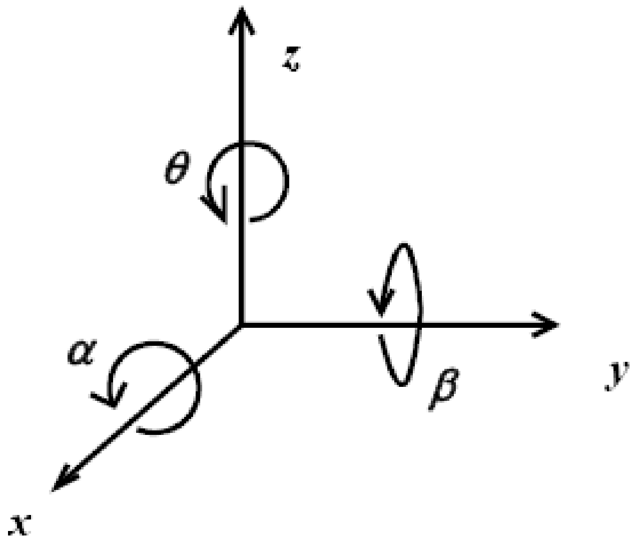
$$\mathbf{T}^{-1} = \begin{bmatrix}1&0&0&-t_x\\0&1&0&-t_y\\0&0&1&-t_z\\0&0&0&1\end{bmatrix}$$

Scaling in Homogeneous Coordinates

Scaling by factors S_x, S_y and S_z along the x, y, and z axes.

$$\mathbf{S} = \begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Rotation in Homogeneous Coordinates



Rotation of a point about each of the coordinate axes. Angles are measured clockwise.

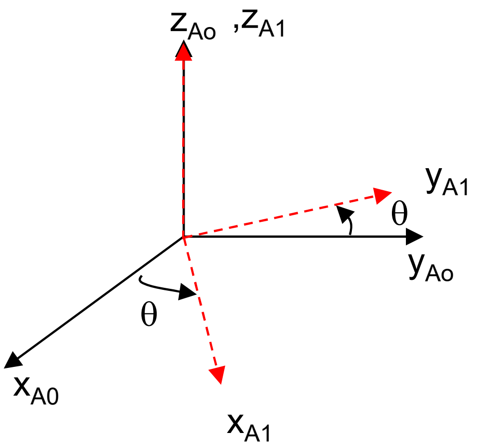
Non-standard orientation of coordinate system:
Common in 2D computer graphics with the origin at the top left corner and the *y*-axis down the screen.
Angles are clockwise positive.
Computer vision and computer graphics are closely related.

$$R_{z,\theta} = \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_{x,\alpha} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 \\ 0 & -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_{y,\beta} = \begin{bmatrix} \cos\beta & 0 & -\sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Standard right-hand Cartesian coordinate system in mathematics: *x*-axis to the right and *y*-axis up the screen.



$$R_Z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$R_X(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$$

$$R_Y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

Composite Transformation

- With homogeneous coordinates, a sequence of transformations: translation T, scaling S, followed by rotation about z axis can be represented in matrix equations.
- P’ is the transformed point corresponding to the point P.
- The order of application is important.
- Many transformations have inverse matrices that can be used to perform the opposite transformation.

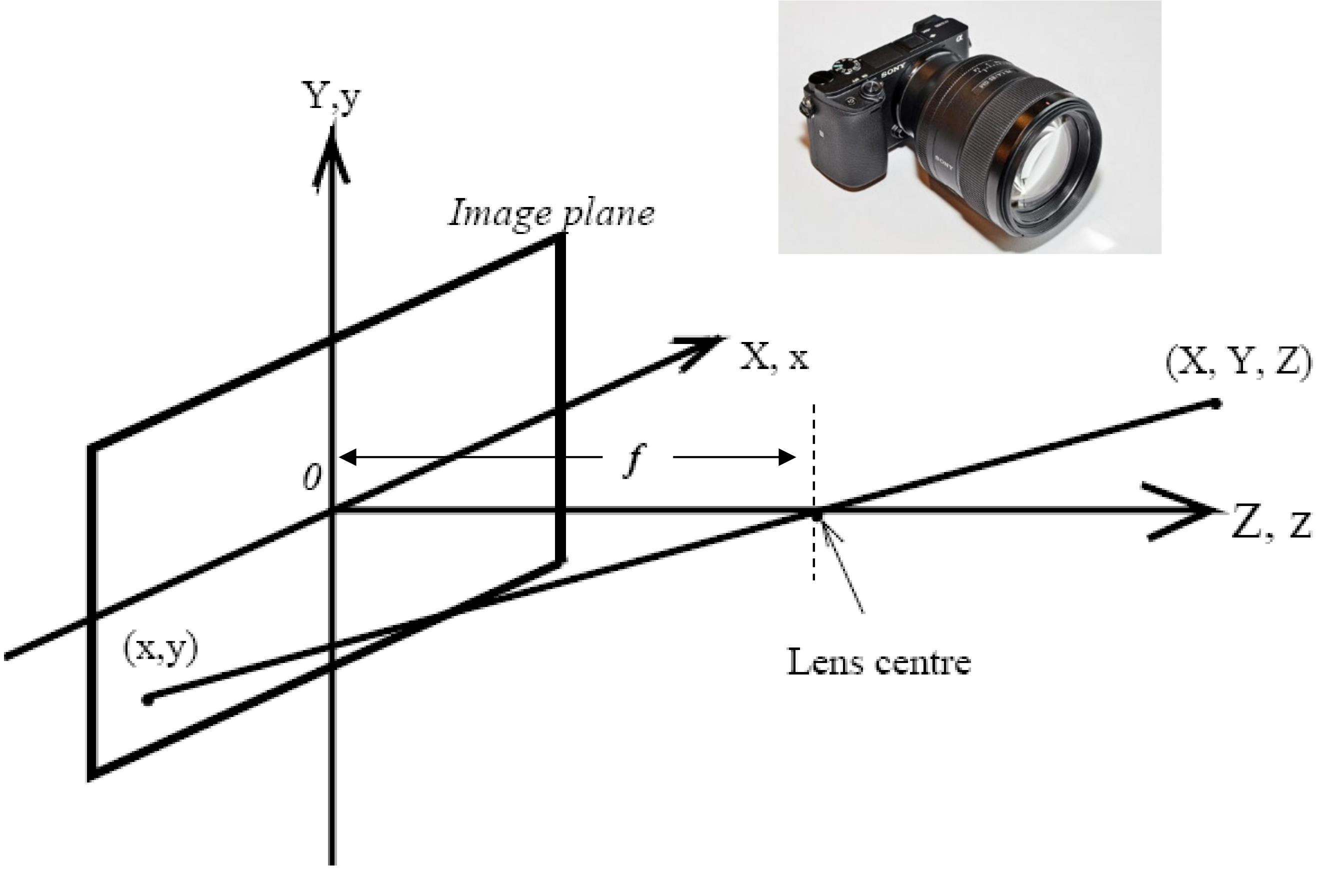
$$\boxed{\mathbf{A} = \mathbf{R}_{z,\theta}\mathbf{S}\mathbf{T}}$$

$$\boxed{\mathbf{P}' = \mathbf{A}\mathbf{P}}$$

Perspective Projection

Perspective transformation (image transformation) projects 3D points onto a plane.

- Projection from 3D points onto a plane



$Z >> f$

similar triangles:

$$\frac{x}{f} = -\frac{X}{Z-f} = \frac{X}{f-Z}$$

$$\frac{y}{f} = -\frac{Y}{Z-f} = \frac{Y}{f-Z}$$

The inverse perspective transformation from a point in plane to 3D space corresponds to a line, i.e. to an homogeneous vector.

Perspective Projection in Matrix Form

$\mathbf{P}$: perspective projection matrix

$\mathbf{c_h}, \mathbf{w_h}$: homogeneous coordinates for image and workspace respectively

$$\mathbf{c_h} = \mathbf{P}\mathbf{w_h} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix} \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} = \begin{bmatrix} kX \\ kY \\ kZ \\ -\frac{kZ}{f} + k \end{bmatrix}$$

Cartesian coordinates of the image point

$$\mathbf{c} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{fX}{f-Z} \\ \frac{fY}{f-Z} \\ \frac{fZ}{f-Z} \end{bmatrix}$$

$\mathbf{c}$ can be recovered from $\mathbf{c_h}$ by dividing the first three elements of $\mathbf{c_h}$ by the fourth element.

$$\mathbf{c} = \mathbf{c_h} \frac{1}{\mathbf{c_h}(4)}$$

Perspective Projection in Matrix Form

Inverse perspective transformation maps image point back into 3 - D

$$\mathbf{w_h} = \mathbf{P}^{-1}\mathbf{c_h}$$

$$\mathbf{P}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 1 \end{bmatrix}$$

Inverse perspective transformation problem:

Given the coordinates of a point in image plane, we cannot determine the corresponding coordinates in 3D scene.

Mapping a 3D scene into a 2D image plane is a many-to-one transformation.

Homogeneous Coordinates

- To represent rotation, translation, scaling and perspective transformation in a similar way, using matrices
- Composed operations can be realized using matrix multiplication.

3x3 rotation matrix

3x1 translation vector

$$\left[\begin{array}{ccc|c} R_{11} & R_{12} & R_{13} & t_1 \\ R_{21} & R_{22} & R_{23} & t_2 \\ R_{31} & R_{32} & R_{33} & t_3 \\ \hline 0 & 0 & 0 & 1 \end{array}\right]$$

1x3 perspective projection vector

Perspective Projection: -1/*f*

1x1 scaling (stretching) matrix

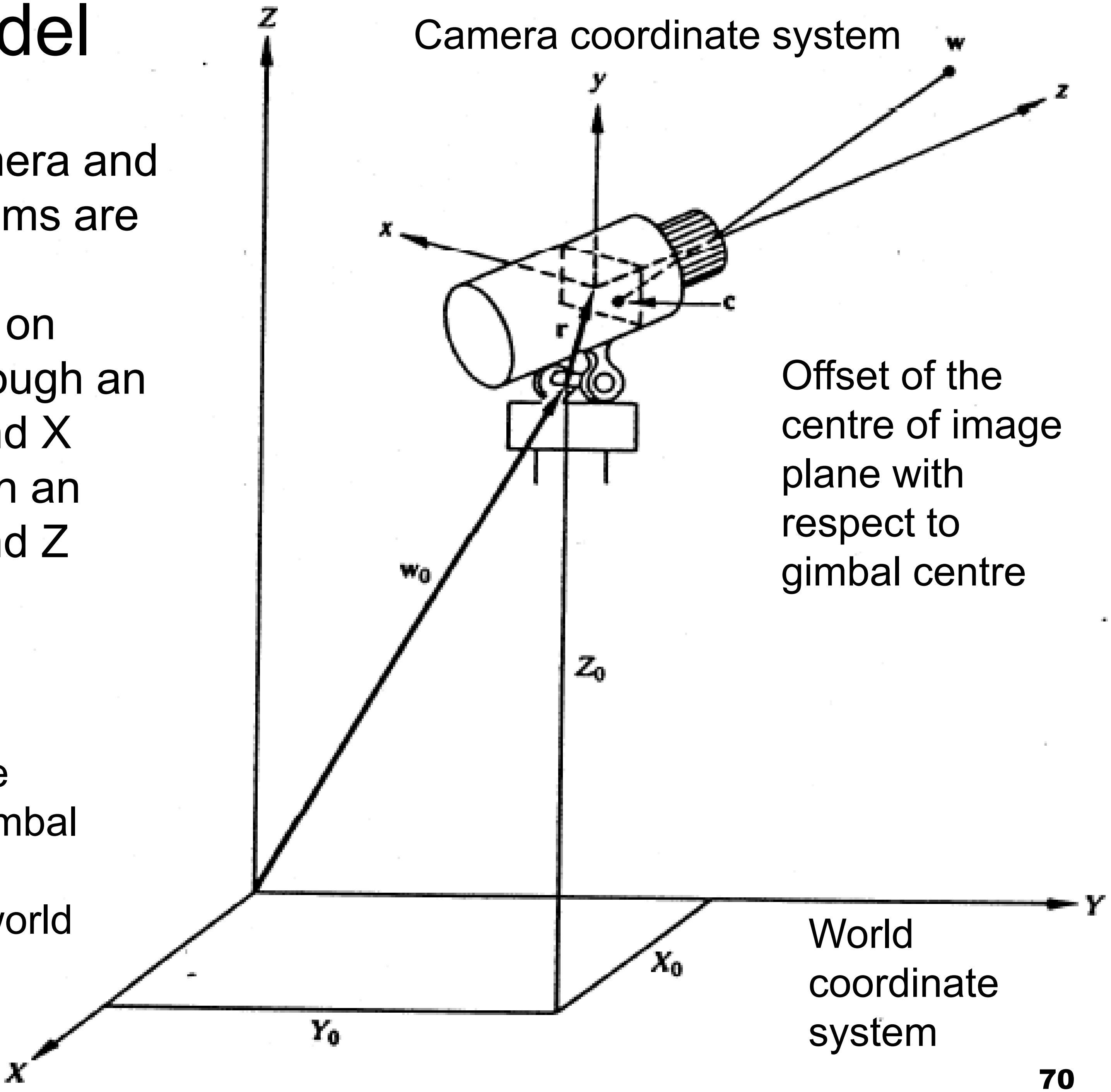
Camera Model

In practice, the camera and world camera systems are not coincident.

A camera mounted on gimbal can pan through an angle between x and X axes and tilt through an angle between z and Z axes.

w: a point in 3D

c: image point



Camera coordinate system

Offset of the centre of image plane with respect to gimbal centre

Offset of the centre of gimbal centre from original of world coordinate system

World coordinate system

To develop a camera model

The initial orientation of the camera has the x-y-z axes parallel to X-Y-Z axes (the camera is pointing upward initially).

1. A translation from the origin of the world coordinates to the gimbal centre
2. A rotation about z-axis by an angle θ
   - Pan angle is measured between the x and X axes.
3. A rotation about x-axis by an angle α
   - Tilt angle is measured between the z and Z axes.
4. A translation from the gimbal centre to the origin of the camera coordinate frame.
5. A perspective transform from the image point to the world point.

$$\boxed{\mathbf{c_h} = \mathbf{P}\mathbf{T}(r)\mathbf{R_x}(\alpha)\mathbf{R_z}(\theta)\mathbf{T}(w_0)\mathbf{w_h}}$$

Camera coordinates

Transformations

World coordinates

Camera Calibration of Linear Model

- In homogeneous coordinates, the projection of a scene point or object point in world coordinate X to an image point u is given by a simple linear mapping. (Equation 4.1). M is the camera projection matrix.

- If the scale factor = 1 in the homogeneous coordinates, we have Equation 4.2.

$$\mathbf{u} = \mathbf{MX} \Rightarrow \begin{bmatrix} kx \\ ky \\ kz \\ k \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad (4.1)$$

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad (4.2)$$

Camera Calibration of Linear Model

- Ignoring the term of z since z=0 on the image plane (Equation 4.3).

- There are 11 free parameters in M.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{41} & a_{42} & a_{43} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad (4.3)$$

Camera Calibration of Linear Model

1. Measure *N* points in workspace and record the corresponding points in the image
2. Identify the 11 parameters *α* using the least square solution method
3. Improve the calibration using a look up table or a neural network