Image Processing and Imaging
Overview and Image Acquisition

Dominik Söllinger
Fachbereich Computerwissenschaften
Universität Salzburg

Winter term 2022/23

PARIS LODRON UNIVERSITY SALZBURG

Topics discussed

What I'm gonna learn in this course?

In this course you will learn the basic techniques in image processing.

However, note that this is not a "Photoshop course"! More or less you will learn how the methods applied in Photoshop work.

Short overview about the topics

- Human Visual System
- Image Acquisition and formation
- Image Enhancement techniques (histogram modification, image sharpening and denoising, etc.)
- Edge detection techniques
- Morphological operations
- Image segmentation

Course material and contact details

Course material

Information and content for both course (VO+PS) will be published at
https://github.com/dsoellinger/teaching

How to get in touch?

In case of questions/problems you can drop me an e-mail (dsoellinger@cs.sbg.ac.at) or come to my office (Room 1.13).

Important note

If you expect a fast reply, do **NOT** use the PLUS e-mail address!

Overview and Related Terms

Vision

Allows humans to perceive and understand the world around them

Image

2D representation of a 3D scene as captured by an image sensor

Computer Vision

Aims to duplicate the effect of human vision by electronically perceiving and understanding an image. However, this not an easy task!
We live in a three dimensional world, but available sensors capture 2D images. Projection to a lower number of dimensions results in an enormous loss of information

Outline

Levels of Image Processing

- In general, we distinguish between two different levels of image processing:

Low Level Image Processing

Classical image processing
Methods use very little knowledge about (semantic) content of images
Main topic of this course

High Level Image Processing (Understanding)

Based on knowledge, goals and plans to achieve this goals; artificial intelligence (AI) methods often used; can be identified with computer vision and tries to imitate human cognition and the ability to make decisions according to information contained in an image

Low Level Digital Image Processing Tasks

- The following sequence of processing steps is common in digital image processing:

Image Acquisition

Image is captured by a sensor (e.g. digital camera) and digitised.

Pre-Processing

Suppress noise, enhance some object features relevant to understanding the image; edge extraction is carried out at this stage

Image Segmentation

Separate objects from the image background (focus on relevant information)

Object Description and Classification

Extract features from the relevant objects (in a totally segmented image)
May be considered as low level image processing; often seen as part of high level img. proc.

Outline

We see light – but what is that ? These three are the same:

- Light: pure energy
- Electromagnetic waves: energy-carrying waves emitted by vibrating electrons
- Photons: particles of light

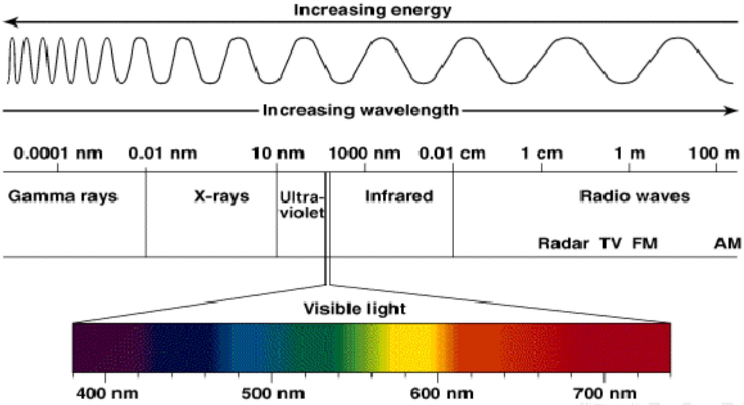


Figure: Electromagnetic spectrum

Electromagnetic spectrum

The electromagnetic spectrum is classically partitioned into the following seven bands (for each waveform, the speed is identical, i.e. the speed of light (299,792,458 m/s):

- Radio waves: communication
- Microwaves: used to cook
- Infrared: "Heat Waves"
- Visible light: what humans see
- Ultraviolet: Causes sunburn
- X-Rays: Penetrate tissue
- Gamma rays: Most energetic

Outline

The human eye acts like a camera (or is a camera ?):

- Eye has an iris like a camera (used to control the aperture by radial muscles)
- Focusing is done by changing the shape of the lens
- Photoreceptor cells (rods – "Stäbchen", and cones – "Zäpfchen") in the retina act as the image sensor

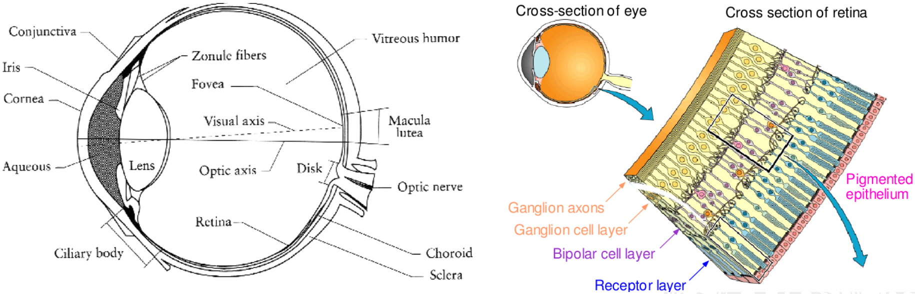


Figure: Human eye and retina.

- Fovea is a small region of high resolution containing mostly cones
- Optic nerve consists of 1 million flexible fibers used to transfer the acquired data for further processing
- Region where the optic nerve leaves the retina is called the "disc',
    - In this area there are no photoreceptors; thus, we are blind in this area.

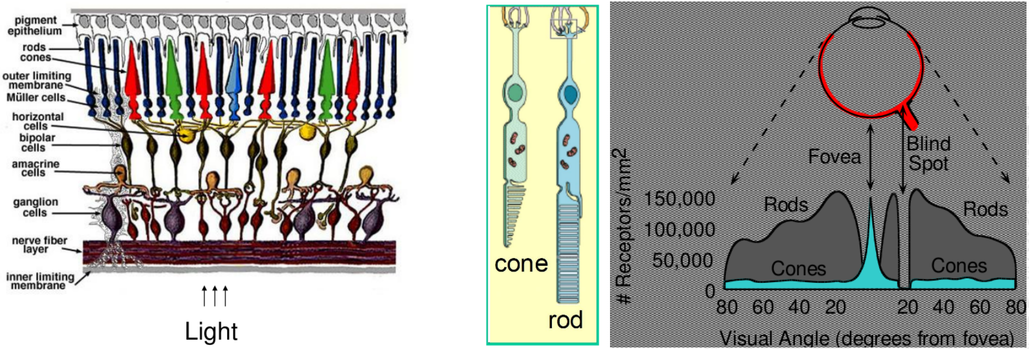


Figure: Colour und luminance perception: Rods and cones.

Human Eye - Rods and Cones (2)

Cones

Less sensitive, operate in high light, and are responsible for colour vision
Three types of receptors, which perceive different portions of the visible light spectrum: Red (L — long wavelengths), green (M — middle wavelengths), and blue (S — short wavelengths)

Rods

Highly sensitive, operate at night, and provide gray-scale vision only
Note that these support peripheral vision only (see the distribution) – so why are there more stars off-center ?

Human Eye - Light intensities

- Human visual system can perceive approximately 10^{10} different light intensity levels
- At any one time we can only discriminate between a much smaller number – *brightness adaptation*
- Perceived intensity of a region is related to the light intensities of the regions surrounding it (so called "Mach Bands")

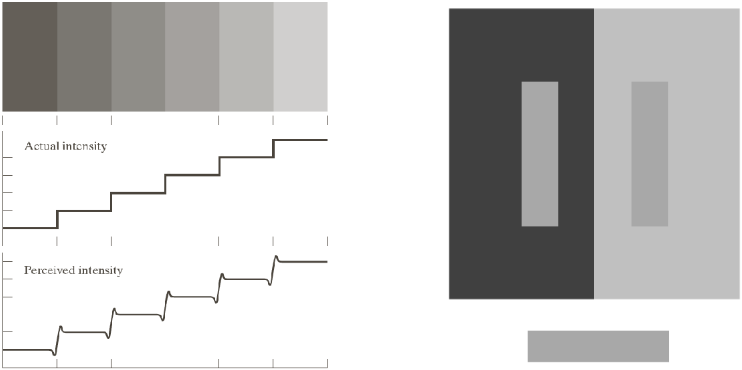


Figure: Perceived intensity depends on surroundings.

Muscles within the eye are used to change the shape of the lens

- Allowing us focus on objects that are near or far away
- An image is focused onto the retina
- Causing rods and cones to become excited which ultimately send signals to the brain
- Focus function of the lens has to be taken into account

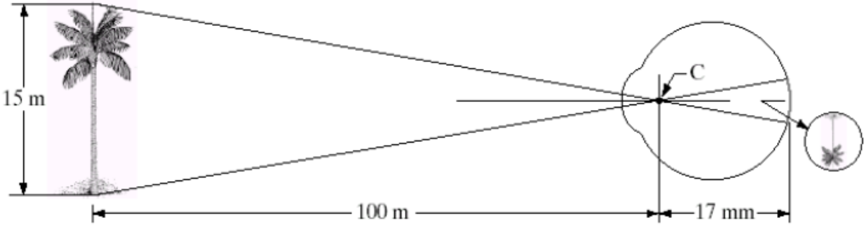
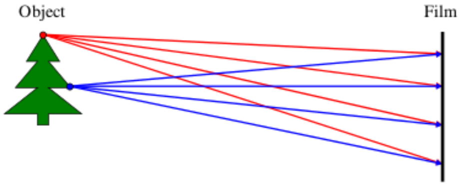


Figure: Image formation in the eye

Outline

- Some medium to emulate / replace the retina is needed
- In the last millennium, film was used; able to record light due to a chemical reaction
- Nowadays, digital sensor took over

The first idea to build a camera is to put a sensor / film in front of an object:



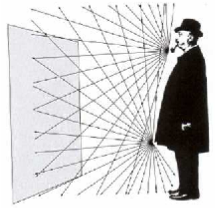
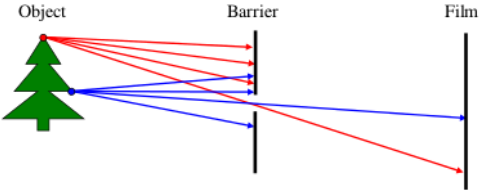


Figure: Using a plain sensor as a camera does not work.

Problem: no reasonable image as the information on the sensor is extremely blurred, basically each single object point is projected to each sensor element

Motivates the use of a barrier with a small aperture only:

- A small amount of rays is able to pass
- This concept is called "pinhole camera"
- Significantly reduces the blurring effect
- Each point in the scene projects to a single point (or very small area) point the image



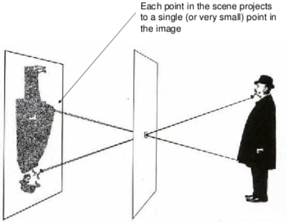


Figure: Using a barrier with hole improves the situation.

- The focal length f is the distance between the pinhole and the sensor.
- If doubled f, the size of the projected object is doubled as well

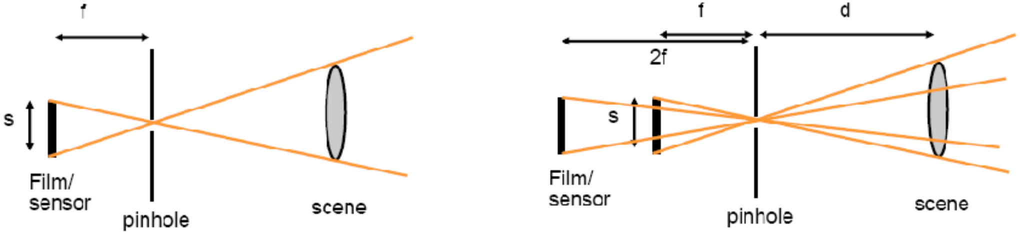
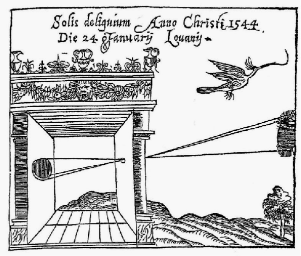


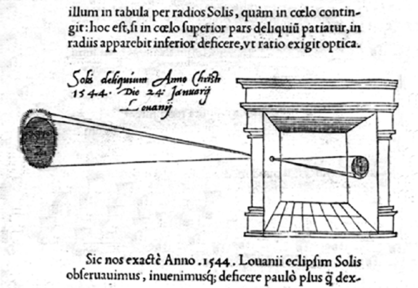
Figure: Influence of focal length on object size.

Pinhole cameras (despite already being known long ago) have severe limitations:

- Aperture (the "hole") must be very small to obtain a clear image
- As the pinhole size is reduced, less light is received by the image plane (i.e. sensor)
- If the pinhole size is comparable to wavelength of incoming light, "diffraction" effects blur the image
- Strategy to obtain differently sized objects on the sensor is not very convenient:
  - Varying the barrier – sensor distance significantly



Camera Obscura, Gemma Frisius, 1558


illum in tabula per radios Solis, quàm in cœlo contingit: hoc est, si in cœlo superior pars deliquiũ patiatur, in radiis apparebit inferior deficere, vt ratio exigit optica.

Solis deliquium Anno Christi 1544. Die 24 Januarij Louanij

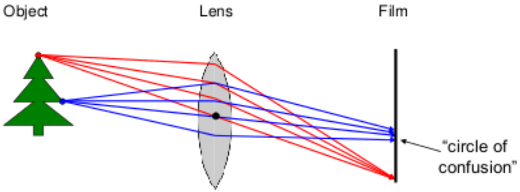
Sic nos exactè Anno .1544. Louanii eclipsim Solis obseruauimus, inuenimusq; deficere paulò plus q̃ dex-

Figure: Historic pinhole cameras and tradeoff size of pinhole / sharpness.

Outline

Pinhole concept is replaced by introducing a lens:

- Now we finally arrive very close to the human model, which focuses light onto the sensor
- At a specific distance, objects are "in focus", other points project to a "circle of confusion" in the image → results in blur
- Changing the shape of the lens (as is done in the human eye) changes the focal distance.
- Parallel rays are focused onto a single point, the focal point.
- With a lens, f denotes the distance between the plane of the lens and the focal point
- An aperture of certain size restricts the range of rays passing through the lens and influences required exposure time (compare: pinhole size)



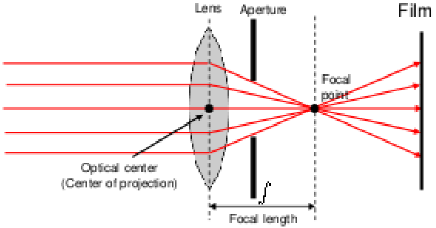


Figure: Introduction of the lens.

Changing the aperture also affects the depth of field:

- A smaller aperture increases the range in which the object is approximately in focus

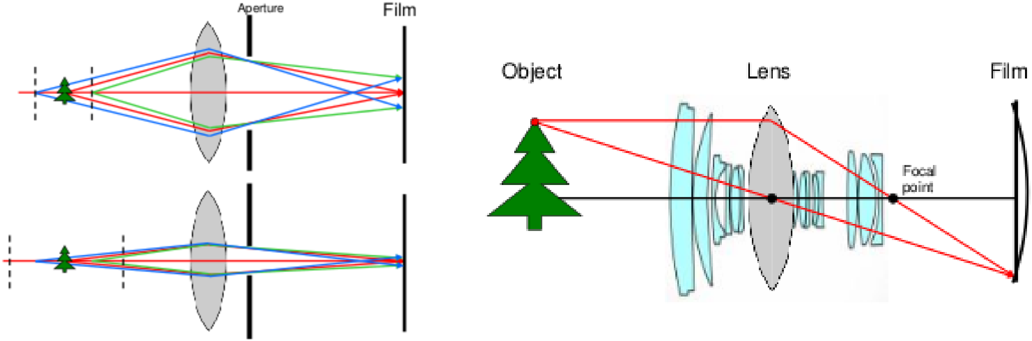


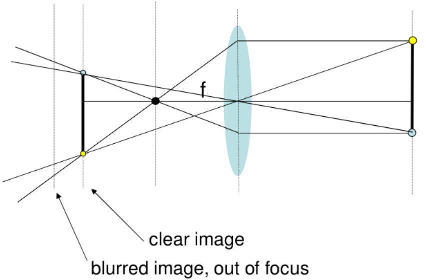
Figure: Effects of changing aperture and thick lenses.

Usually, we consider lenses to be “thin” lenses, i.e. that the thickness of the lens is negligible compared to the focal length → in modern lenses, this is hardly the case

Thin Lens Scenario

Prrinciple of generating clear and blurred images depending on the position of the sensor plane relative to the focal point and the lens plane:

- y is the object size
- y' is the size of the object in the image
- d is the distance between object and lens plane
- d' is the distance between sensor and lens plane
- f is the focal length



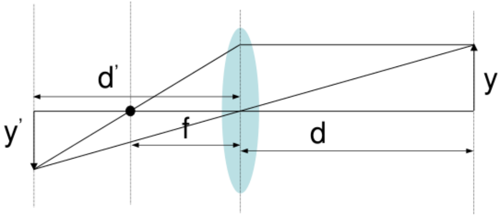


Figure: Thin lens scenario.

Deriving the "thin lens formula" using simple geometric principles (i.e. triangle similarity properties):
Exploiting the yellow triangles' similarity, we obtain

$$\frac{y'}{d'} = \frac{y}{d} \Longrightarrow \frac{y'}{y} = \frac{d'}{d} \ .$$

Similarly, exploiting the green triangles' similarity, we obtain

$$\frac{y'}{d'-f} = \frac{y}{f} \Longrightarrow \frac{y'}{y} = \frac{d'-f}{f} \ .$$

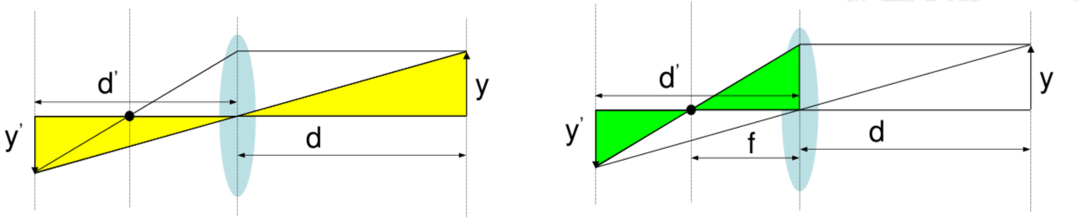


Figure: Deriving the thin lens formula.

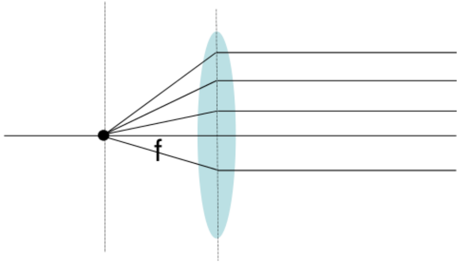
Equating the right side of both equations we result in

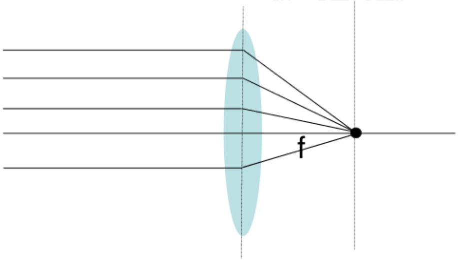
$$\frac{y'}{y} = \frac{d'}{d} = \frac{d' - f}{f} \Longrightarrow \frac{d'}{d} = \frac{d'}{f} - 1 \Longrightarrow \frac{1}{d} = \frac{1}{f} - \frac{1}{d'} \,.$$

Under thin lens assumptions objects are in focus for which $\frac{1}{d'} + \frac{1}{d} = \frac{1}{f}$ holds.

Consequences of this result for extreme situations:

- Objects at infinity focus at f: If $d \longrightarrow \infty$ then $d' \longrightarrow f$
- By analogy, when the object gets closer, the focal plane moves away from f
- At the limit: If $d \longrightarrow f$ then $d' \longrightarrow \infty$, i.e. an object at distance f requires the focal plane to be at infinity





Effects of Focal Length on the Image Size

Varying the focal length f results in a different image size as a consequence:

- Corresponds to applying a zoom or changing the lens
- Set M as the relation between y and y', i.e. $M = \frac{y'}{y} = \frac{d'}{d}$
- As a consequence of the thin lens equation ($\frac{1}{d'} + \frac{1}{d} = \frac{1}{f}$):
- $M = \frac{f}{d-f}$ for $d > f$, i.e. M gets larger for increasing f

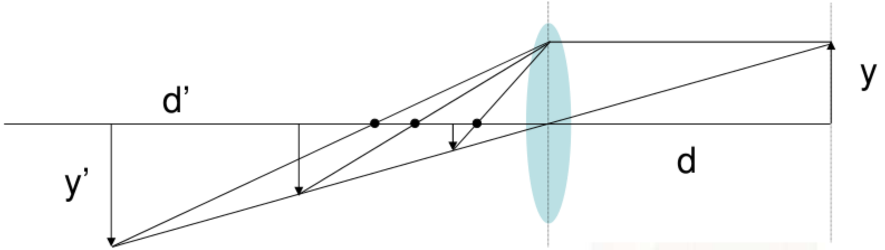


Figure: Effect of focal length on image size.

The sensor has a fixed size:

- As f gets larger, smaller parts of the world project onto the sensor, the image becomes more *telescopic*
- As f gets smaller, more world points project onto the sensor, the images become more wide angle in nature

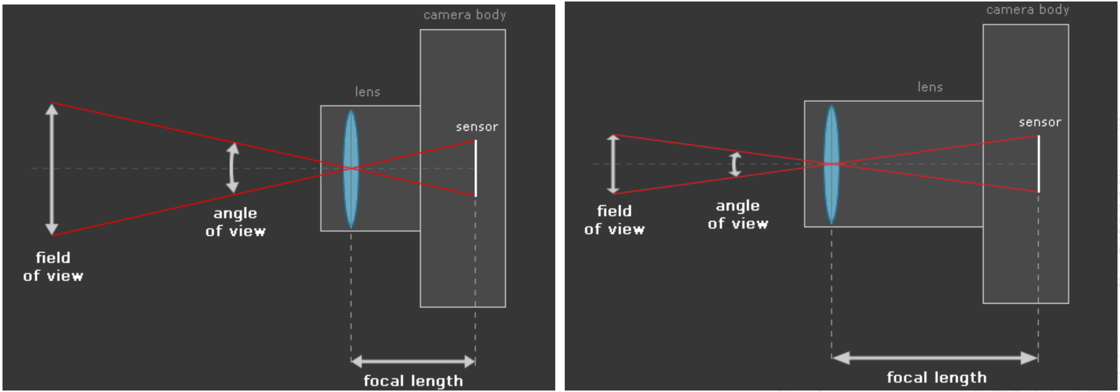


Figure: Effect of focal length on image content.

Depth of Field ("Tiefenschärfe")

Is the range of object distances d over which the object in the image are sufficiently well focused.

Connection between the circle of confusion (the area of the sensor in the focal plane, onto which out of focus objects are projected) and the depth of field:

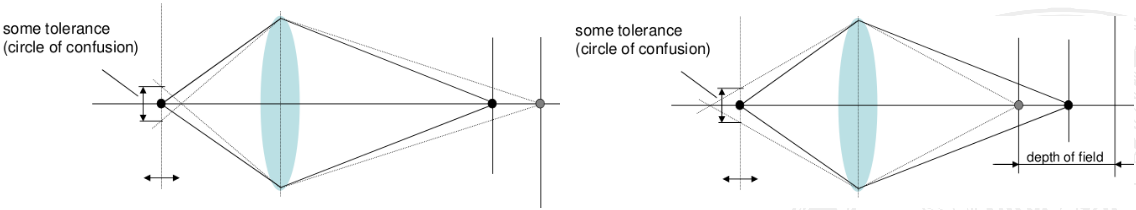


Figure: Depth of field and circle of confusion.

Outline

- When using an aperture, less rays arrive at the focal plane
- Not much difference except that image is darker (requiring a longer exposure time).

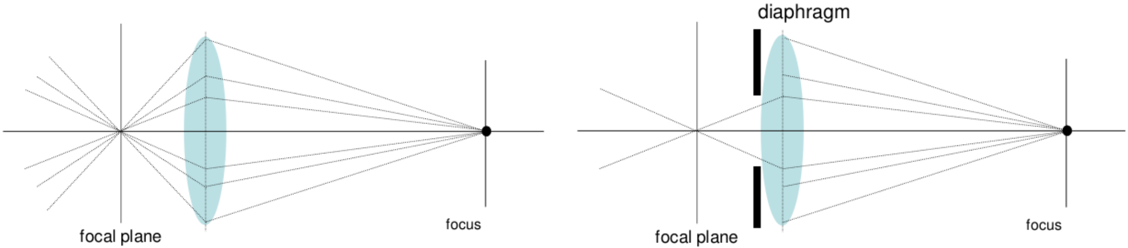


Figure: Effect of using an aperture (object in focus).

Considering the situation of out-of-focus points:

- The usage of the aperture has an additional effect:
- The circle of confusion is smaller
- Leading to a sharper (less blurry) image

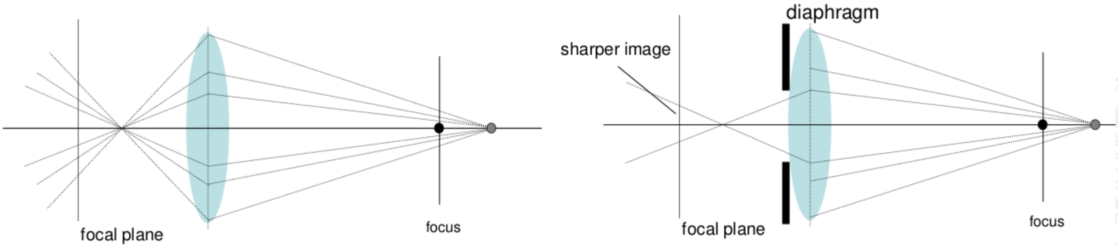


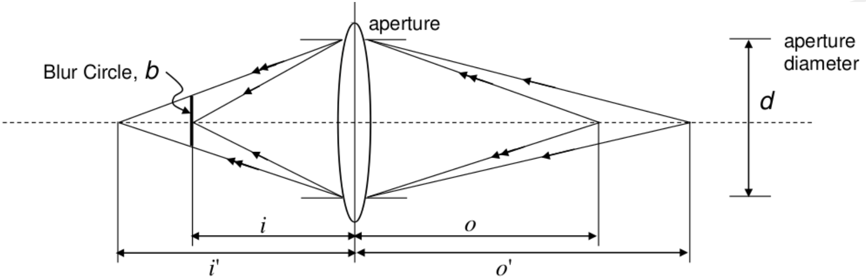
Figure: Effect of using an aperture (object non fucused).

Relation between the Circle of Confusion and the Aperture (1)

We can derive an explicit relation between the diameter of the circle of confusion b ("blur circle") and the diameter of the aperture d:

- Similar triangles, $b = \frac{d}{i'}(i' - i)$ for two object distances o, o' and their respective images distances i', i
- Distance $(i' - i)$ cannot be determined reasonably $\rightarrow$ we express b with f and o, o' instead
- For both pairs o, i' and o', i the thin lens equation holds:

$$\frac{1}{i'} + \frac{1}{o} = \frac{1}{f} \text{ and } \frac{1}{i} + \frac{1}{o'} = \frac{1}{f} \; .$$



(note that object and image distance to the lens are inconsistently denoted as o and i')

Relation between the Circle of Confusion and the Aperture (2)

From these formulas we can isolate i and i' and may derive

$$(i' - i) = \frac{f}{(o' - f)} \frac{f}{(o - f)} (o - o') \ .$$

- Now we are able to compute the diameter of the circle of confusion based on objects' distance and the focal length f
- This is important to determine the exact depth of field:
  - Range for which the diameter of the circle of confusion is smaller than the resolution of the sensor
  - As long as all information within the circle of confusion is mapped to a single pixel, there is no blur

Problems Arising with (Thick) Lenses

Several problems arise with the usage of lenses:

- Some of which can be corrected
- Others are hard to correct
- Some of which may be used to identify certain specific lenses as done in image forensics

Spherical Aberration

Spherical lenses are the only easy shape to manufacture, but are not correct for perfect focus, different refractive indices leading to different focal points for different parts of the lens.

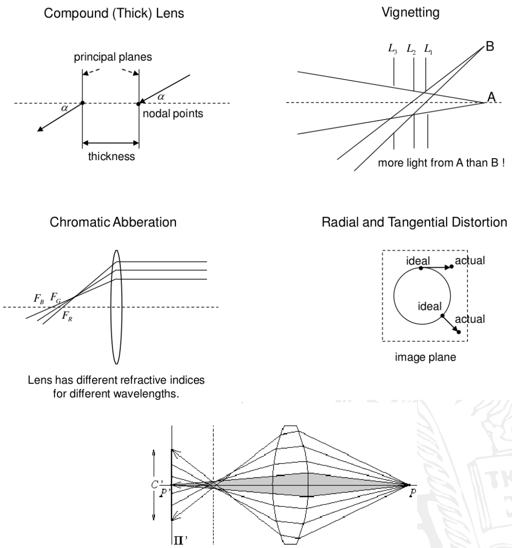


Figure: Some selected problems with lenses.

Problems Arising with (Thick) Lenses - Examples

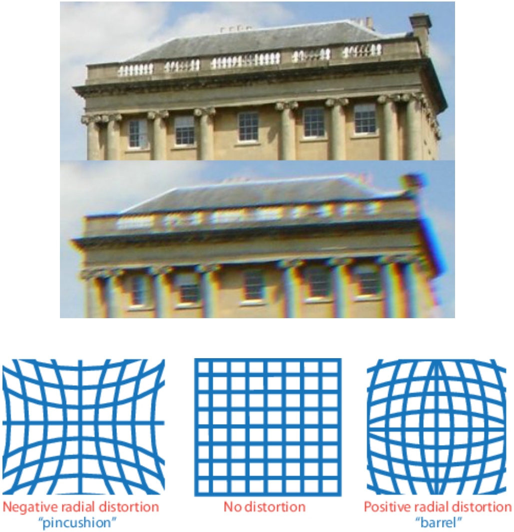


Figure: Chromatic aberration and radial distortion

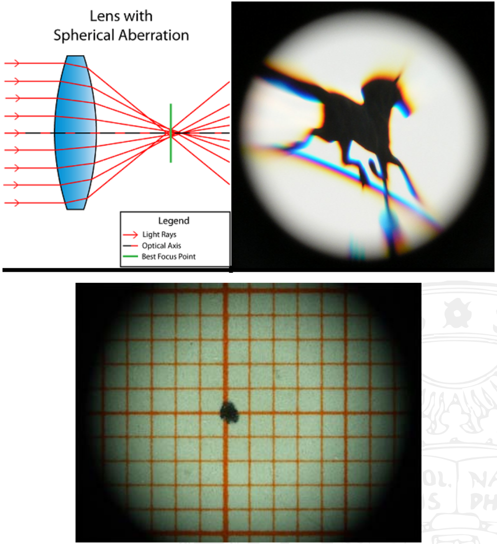


Figure: Spherical aberration and vignetting

Outline

- Classical stationary array sensors as found in digital cameras
- Sensing devices with moving sensor elements
- Moving sensor elements are found in scanners
- Toroidal sensor areas

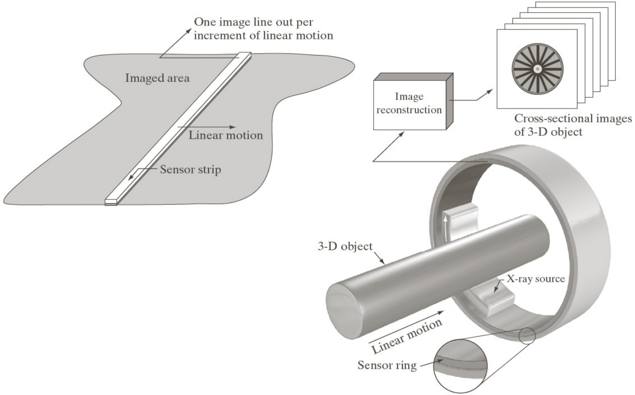


Figure: 1D and 3D sensing devices.

- A source of radiation is moved to generate different perspectives of the data
- Often used in medical imaging
- In the shown tomography example to produce a set of cross section images
- Can be applied to any organic material, e.g. also for generating cross section images of wood-logs

Figure: Sensors in Medical Imaging, left: CT scanner, right: MRI scanner

Outline

- Effect of radiation on the semiconductor is to kick electrons from the valence to the conducting band (still inside the material)
- This effect can be read out and digitised.
- These electrons are collected and accumulated in the photosensitive cells
- Once a control switch is activated, they are transferred out of these cells to be converted into measurable electrical quantities, i.e. charge/voltage.
- Linearity between incoming light and resulting charge/voltage is required and usually given, if not, a linearisation is being applied

- Sensors with sensitivity against different bands of the electromagnetic spectrum can be constructed by using different types of semiconductor material
- Silicium is well suited for the visible range and near-infrared (NIR filters needed in cameras)
- Material like e.g. Germanium and others are suited for the far-infrared range

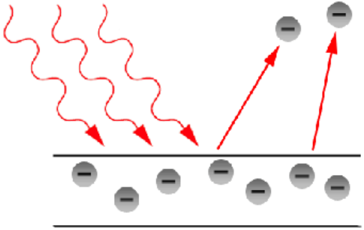


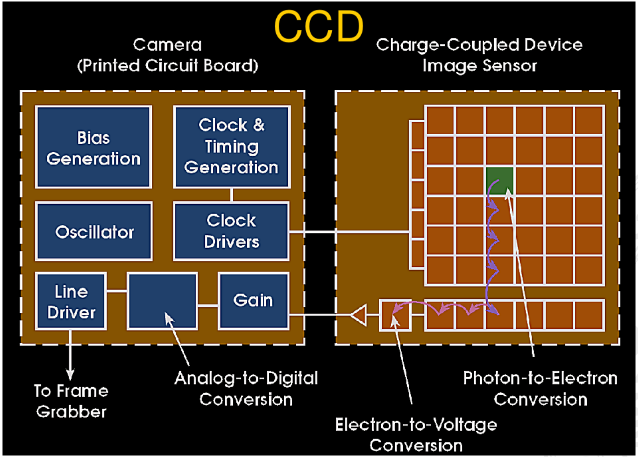
Figure: Photoelectric effect

Outline

CCD Sensor

The two major types of sensors are CCD and CMOS; CCD (charge coupled device) sensor:

- Photosensitive cells able to store charge produced by the light-to-electron conversion
- In addition, the charge can be transferred to an interconnected, adjacent cell
- In this case charges are shifted out of the sensor
- Bigger sensors, better quality, but additional circuitry necessary



- Complimentary metal oxide semiconductor
- Consist of transistors within the photosensitive cell
- Perform charge-to-voltage conversion and allow the pixels to be read individually
- Higher integration, less power consumption, but less sensitive, higher noise

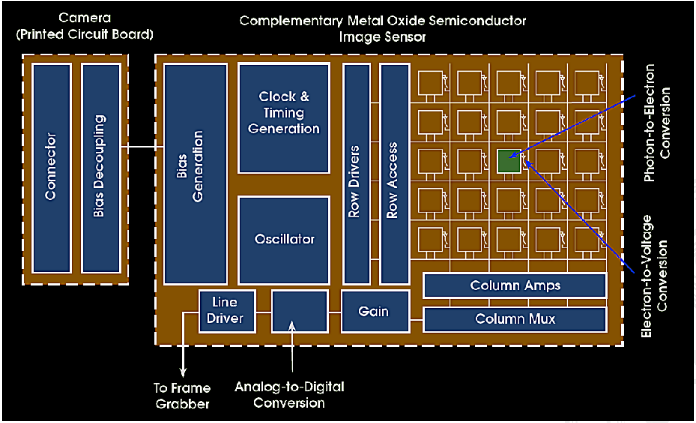


Figure: CMOS sensor

Schematic comparison of the two technologies:

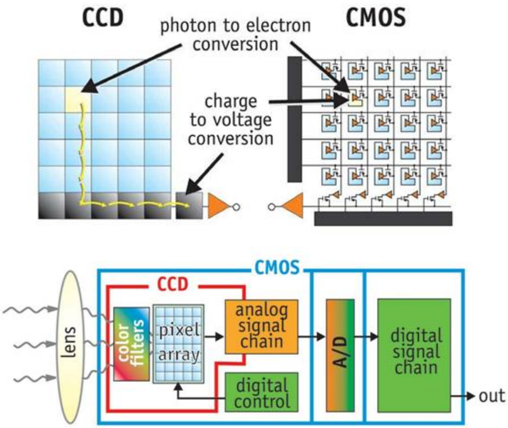


Figure: Schematic comparison of CCD and CMOS sensors.

FEATURE	CCD	CMOS
Pixel Signal	Charge	Voltage
Chip Signal	Voltage	Bits
Fill Factor	High	Moderate
System Complexity	High	Low
Sensor Complexity	Low	High
Relative R&D Cost	Lower	Higher
Power consumption	Higher	Lower
Dynamic Range	Very High	High
Quality	Very High	High*

Figure: Comparing CCD and CMOS sensors: Respective properties

CCD Configurations / Techniques

- Different types of CCD configurations / techniques which have some distinct properties
- Useful for different types of cameras / applications:
  - Full frame (FF)
  - Frame transfer (FT)
  - Interline transfer (IT)
  - Frame interline transfer (FIT).

- Above ones are front side illuminated
- Light enters on the same surface of the device that holds the circuitry
- Almost all single-tap, having only one output point

- Note: in all of the figures, the light gray areas are sensitive to light and the dark gray areas are covered with aluminum
- The arrows show the direction of charge motion when the electrodes are operating

FF - Full Frame CCD Image Sensor (1)

- Simplest version: Photosensitive registers arranged next to each other in columns
- A transport register at the bottom configured so that each cell can receive charge from a different column
- Typical operating cycle: light is prevented from reaching the sensor by a shutter, then the charge is cleared from the photosensitive registers while the image sensor is in the dark
- Shutter is opened/closed for the desired exposure interval
- In the dark, the charge in all of the columns is shifted down by one row so that the last row moves into the horizontal transport register
- A faster clock then moves the charge packets in the horizontal transport register to the sampling node to generate the output voltage
- When the row is complete, the next row is moved down for readout
- This process is repeated until all rows have been read. The sensor is then ready for another exposure.

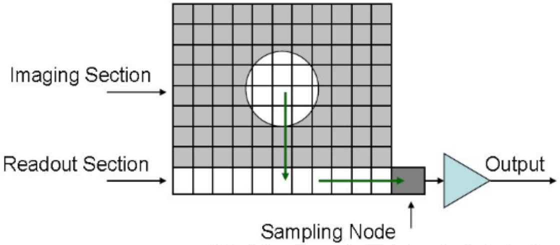


Figure: FF CCD configuration

Full Frame CCD Image Sensor (2)

- Without shutter, a disadvantage is “charge smearing”:
  - Caused by light falling on the sensor whilst the accumulated charge signal is being transferred to the readout register.
- Mechanical shutters have lifetime issues and are relatively slow
- FF are typically the most sensitive CCD’s available but charge readout is rather slow
- There are two strategies to cope with the disadvantages of the FF CCD: FT and IT

Figure: Smearing effect in CCDs configurations.

- Developed as the first CCD type suitable for continuous (video) imaging
- Exposure and readout functions of the sensor are physically separated

- Exposure section built essentially identical to an FF sensor and a storage section almost a copy of the exposure section but covered with an opaque layer (usually aluminum).
- Exposure and readout can occur almost simultaneously
- While the exposure section collects light for a new image, the previous image, held in the storage section, is shifted out

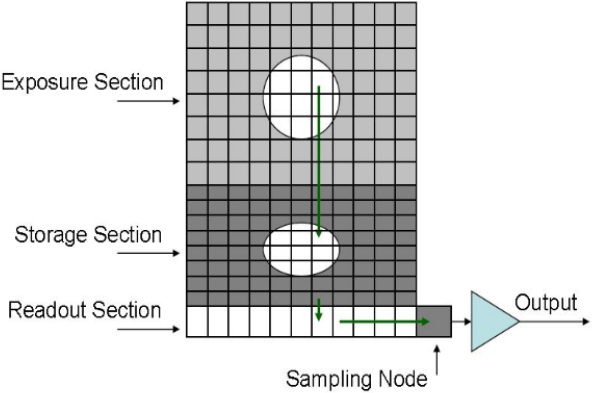


Figure: FT CCD configuration

FT - Frame Transfer CCD Image Sensor (2)

- FT sensor is slightly more complicated to operate that an FF sensor:
  - Exposure and storage section need separate shift drivers
- Typical cycle for a simple 30 frame-per-second progressive scan video camera:
  - Exposure section is collecting light
  - Charge is not moving while the previous information in the storage section is shifting down to the readout section line by line
  - After shifting is complete, the two sections are connected to the same clocks and the entire image is quickly moved down from the exposure section to the storage section
  - Line shift rate during the transfer is typically several hundred times faster than readout line shifting rate
  - Charge pattern moves down very rapidly → vertical charge smearing is reduced (but not eliminated) relative to the FF case
  - After the shift, the scanning is disconnected from the exposure section and the readout resumes at a slower shift rate
- Due to the high speed of the transfer, no mechanical shutter is required / used

- CCD cells in the storage section are smaller than the pixels in the exposure section:
  - No need to efficiently collect light
  - No need to provide a square matrix
  - No need to provide anti-blooming protection (protection against cross-photosensitive cell charge shift for high energetic light points)

Blooming

Occurs when the charge in a pixel exceeds the saturation level and the charge starts to fill adjacent pixels

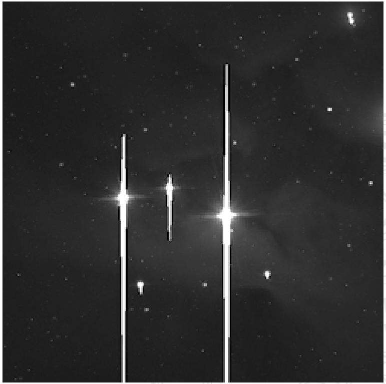


Figure: Blooming effect in CCDs configurations.

FT - Frame Transfer CCD Image Sensor (4)

- A few extra rows are usually added at the top of the storage section and covered with the opaque layer. These rows act as a buffer between the edge of the light-collecting pixels and the storage cells
- Prevents stray light from getting into the first few rows of the storage area that contain image data
  - It could generate spurious charge that can produce white background patches in the image

- FT devices have typically faster frame rates than FF devices
- Advantage of a high duty cycle i.e. the sensor is always collecting light
- FTs have the sensitivity of the full frame device and do not need a mechanical shutter
- Typically more expensive due to the larger sensor size needed to accommodate the frame storage region

IT - Interline Transfer CCD Image Sensor (1)

- Exposure and storage sections are alternated column by column in same area of silicon
- Each column of photosensitive elements has next to it a vertical transfer register covered with aluminum
- Group of transfer registers makes up the storage area

- Light is collected in the photoelements
- Accumulated charge is moved from all photoelements simultaneously into the neighboring transfer registers
- After this move, the charge is shifted vertically, one line at a time, into the output register for readout

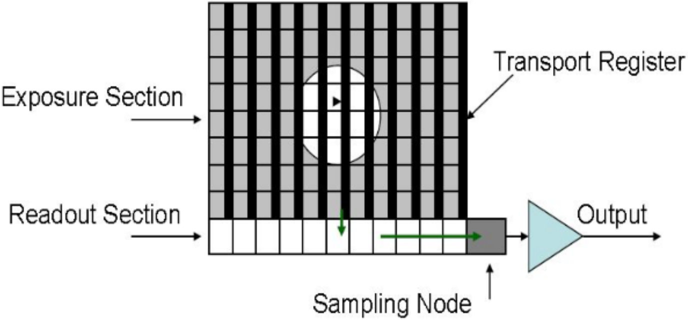


Figure: IT CCD configuration

IT - Interline Transfer CCD Image Sensor (2)

- Photoelements may appear to be in columns as in FT sensors
- BUT: photoelements in IT sensors are not connected together vertically
  - There is no need to shift vertically in the photosensitive areas

- Very rapid image acquisition virtually eliminates image smear, at least for long exposure times
- Photoelements are all isolated vertically from one another:
  - The possibility of blooming along the photoelement columns is essentially eliminated
- Antiblooming drains can be added between each column of photoelements and the transfer register for the column to the left as part of the isolation structure required between them
- If photoelements are isolated on all sides like this, blooming can be kept under very tight control

- Electronically shutter interline-transfer CCDs:
  - Altering the voltages at the photodiode
  - Generated charges are injected into the substrate, rather than shifted to the transfer channels

- Disadvantage: the interline mask effectively reduces the light sensitive area of the sensor
- This can be partially compensated by the use of microlens arrays to increase the photodiode fill factor
- Compensation usually works best for parallel light illumination
- For some applications which need wide angle illumination the sensitivity is significantly compromised

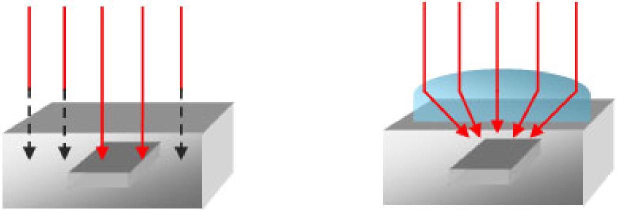
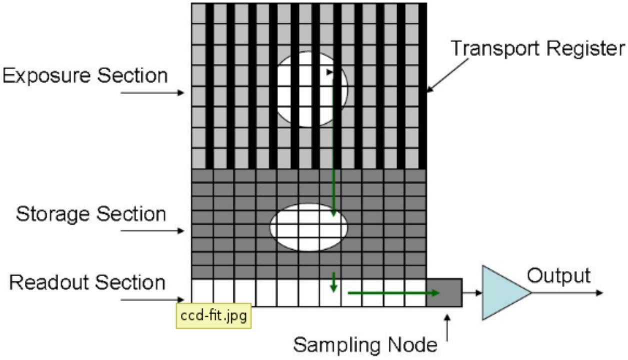


Figure: Microlens principle.

- Was developed to provide both very low vertical smearing and electronic exposure control:
  - Smearing in FT sensors (have no exposure control) can be reduced by moving the image rapidly from the imaging section to the storage section (limited by the charge transfer speed)
  - IT sensors (no exposure control) have more smearing as the exposure time is reduced (the relation between exposure time and transfer time gets problematic)
- The FIT sensor is a combination of IT and FT:
  - Exposure section of an FT sensor is replaced by an IT sensor array

- After the exposure, the charge can be shifted to the vertical transfer registers under the aluminum shields as in an IT device
- It is then immediately shifted at high speed into an FT storage array below



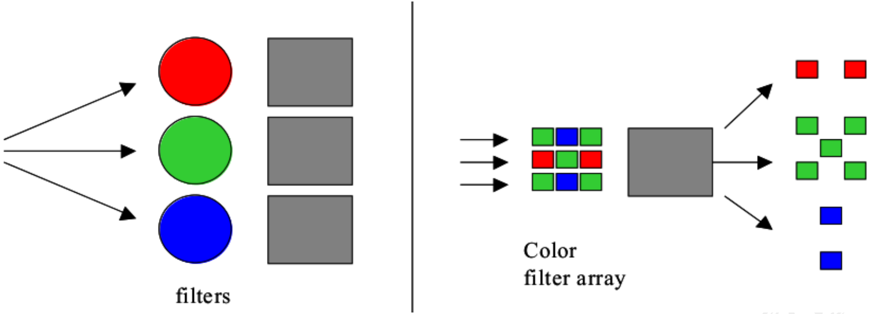
- Time available for accumulation of unwanted charge from light leaking under the shields is reduced to typically less than 1 millisecond with any exposure setting
- Smearing is reduced by both:
  - Effect of the shields
  - Rapid transfer out of the exposure section

- Disadvantage: high cost due to the large sensor area and the reduction of the sensor area by analogy to the IT case

Outline

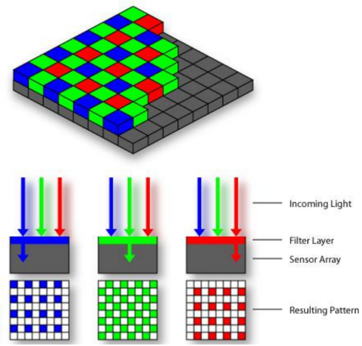
What about Colour?

- In both CMOS and CCD all photosensitive cell are sensitive to visible light, detect only brightness, not color. How to sense colour then?



- Colour Filter Array (CFA): Spatial multiplexing – sensors are made sensitive to red, green or blue using a filter coating that blocks the complementary light (note: similar to human cones)
- 3 detectors: Foveon system using three layer sensors.
- Take 3 shots: temporal multiplexing or three different sensors.

- Using colour filters, each single pixel is made sensitive to a specific colour (colour band) only
- Most prominent example: Bayer pattern



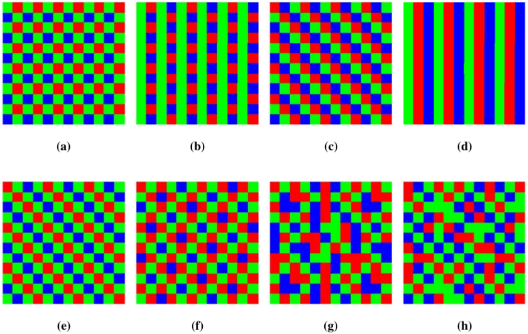


Figure: CFA examples.

- Using three different sensors (usually CCD) avoids spatial multiplexing
- But the camera needs to be bulky and therefore gets expensive
- In addition: the amount of light reaching the sensor is reduced

Nikon Dichrioc:

- Uses a microlens on top of a triplet of photoreceptors
- Using dichroic filters, wavelengths of light are separated to reach specific photoreceptors which record red, green, and blue wavelengths

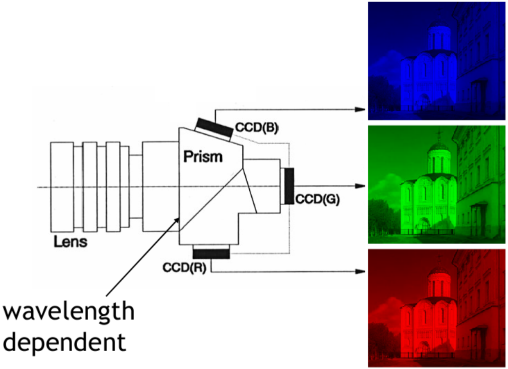
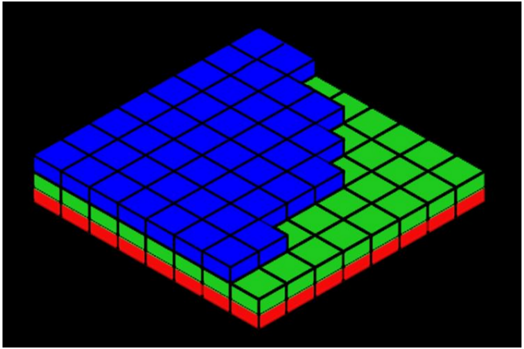


Figure: Non-spatial multiplexing colour sensing (multichip system).

- Foveon X3 is used in Sigma cameras and based on CMOS technology
- Three layers of photodiodes, silicon absorbs different "colors" at different depth
- Each layer captures a different color
- No spatial multiplexing required



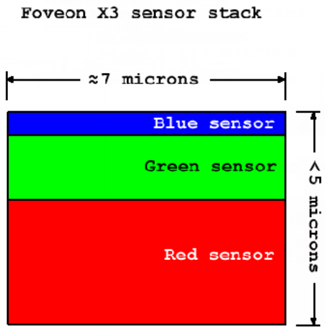


Figure: Non-spatial multiplexing colour sensing (Foveon system).

Outline

Analogies (to the Human Visual System)

- So far, camera development has really followed the human visual system quite closely
- Same strategy in the area of developing vehicles or means of transportation:
  - No cars, no bikes, no planes, no ships, no spacecrafts but bipedal robots

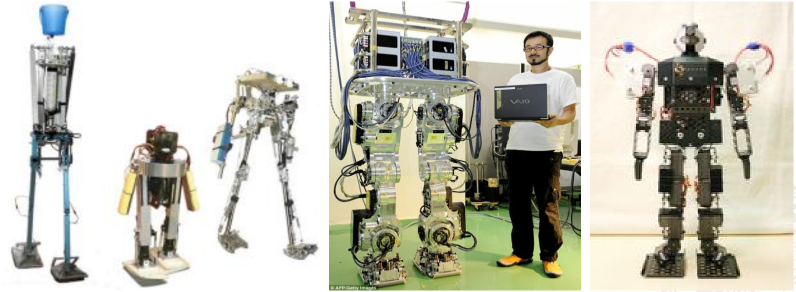


Figure: Means of transportation following the classical digital camera approach.

- Luckily imaging modalities have advanced beyond classical digital still image cameras:
  - Stereo vision, video, multi-view video, range scanners like LIDAR / time of flight, structured light cameras, etc.

Focusing done in cameras is fundamentally different as compared to the HVS:

- Lens does not change its shape (due to restrictive material properties)
  - Recently lenses that can change their shape based on an electric signal (voltage) have been developed.
- Cameras change the distance between lens and sensor surface to adjust for different object distances
- Cannot be accomplished with our physiology
- Focusing can be done:
  - Manually (eventually controlled by the HVS depending on the type of camera)
  - Automatically (i.e. auto-focus systems)