Deep Learning for Computer Vision

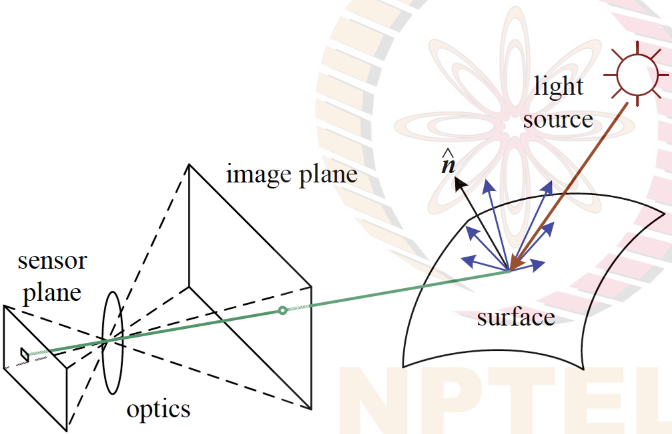
Image Formation

Vineeth N Balasubramanian

Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Image Formation: An Overview[1]

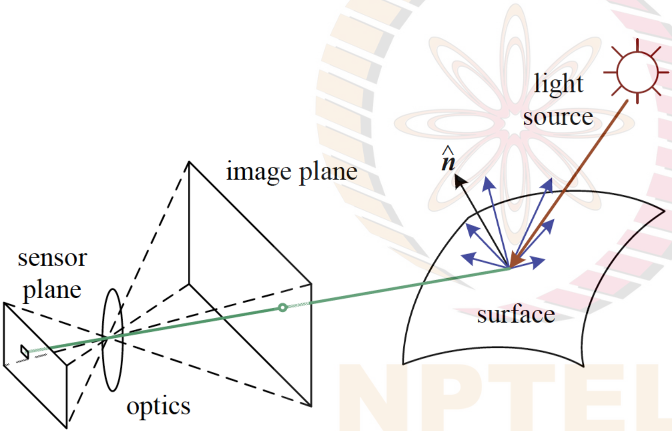


Factors

- Light source strength and direction
- Surface geometry, material and nearby surfaces
- Sensor capture properties
- Image representation and colour

[1]Credit: Szeliski, *Computer Vision: Algorithms and Applications*, 2010

Image Formation: An Overview[1]



Factors

- Light source strength and direction
- Surface geometry, material and nearby surfaces
- Sensor capture properties
- Image representation and colour

[1] Credit: Szeliski, *Computer Vision: Algorithms and Applications*, 2010

Related Topics

- **Geometry**: 2D Transformations, 3D Transformations, Camera Calibration, Distortion
- **Photometry**: Lighting, Reflectance and Shading, Optics
- **Colour**: Physics of Colour, Human Colour, Colour Representation
- **Sensor**: Human Perception, Camera Design, Sampling and Aliasing, Compression

Related Topics

Geometry	2D Transformations, 3D Transformations, Camera Calibration, Distortion
Photometry	Lighting, Reflectance and Shading, Optics
Colour	Physics of Colour, Human Colour, Colour Representation
Sensor	Human Perception, Camera Design, Sampling and Aliasing, Compression

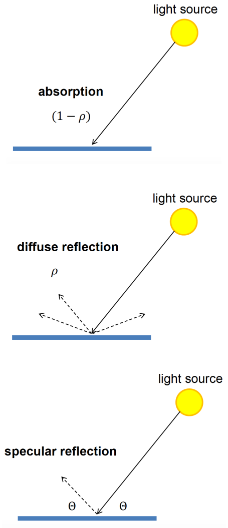
We will cover a few relevant topics from these

For a detailed understanding, read Chapters 1-5 of the book, *Computer Vision: A Modern Approach* by Forsyth and Ponce

Models of Reflection

When light hits a surface:

- Some light is absorbed ($1 - \rho$, $\rho =$ albedo);
  - More absorbed for low albedos
- Some light is reflected diffusively, independent of viewing direction
  - E.g.: Brick, cloth, rough wood
  - **Lambert's cosine law:** Amount of reflected light proportional to $\cos(\theta)$
- Some light is reflected specularly, depends on viewing direction
  - E.g.: Mirror

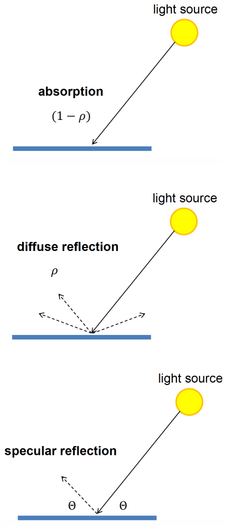


Credit: Derek Hoiem, UIUC

Models of Reflection

When light hits a surface:

- Some light is absorbed ($1 - \rho$, $\rho =$ albedo);
  - More absorbed for low albedos
- Some light is reflected diffusively, independent of viewing direction
  - E.g.: Brick, cloth, rough wood
  - **Lambert's cosine law:** Amount of reflected light proportional to $\cos(\theta)$
- Some light is reflected specularly, depends on viewing direction
  - E.g.: Mirror

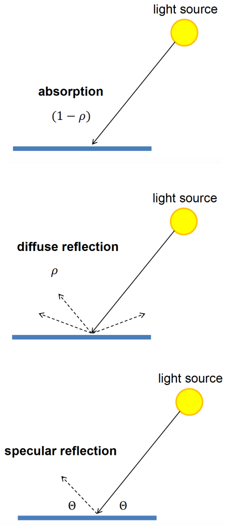


Credit: Derek Hoiem, UIUC

Models of Reflection

When light hits a surface:

- Some light is absorbed ($1 - \rho$, $\rho =$ albedo);
  - More absorbed for low albedos
- Some light is reflected diffusively, independent of viewing direction
  - E.g.: Brick, cloth, rough wood
  - **Lambert's cosine law:** Amount of reflected light proportional to $\cos(\theta)$
- Some light is reflected specularly, depends on viewing direction
  - E.g.: Mirror



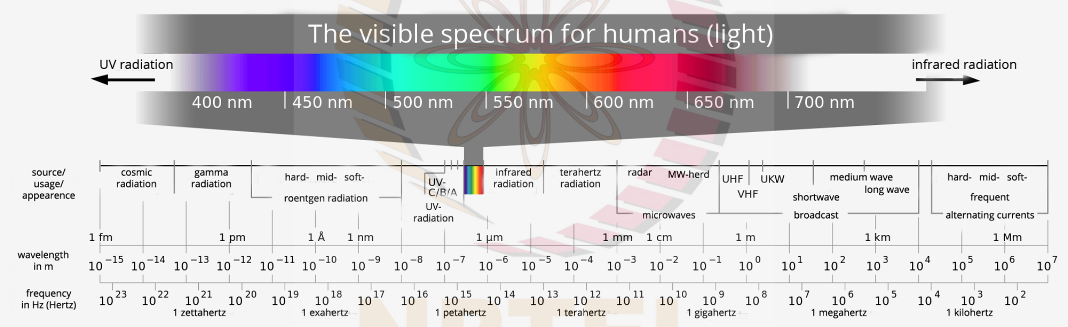
Credit: Derek Hoiem, UIUC

Models of Reflection

- Most surfaces have both specular and diffuse components
- Intensity depends on illumination angle because less light comes in at oblique angles
- Other possible effects:
  - Transparency
  - Refraction
  - Subsurface scattering
  - Fluorescence, phosphorescence
- BRDF - Bidirectional Reflectance Distribution Function: Model of local reflection that tells how bright a surface appears when viewed from one direction when light falls on it from another

Credit: Derek Hoiem, UIUC

Light is composed of a spectrum of wavelengths

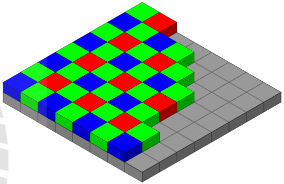


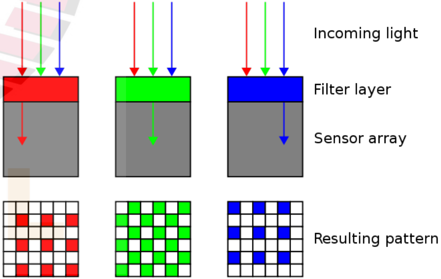
Coloured light arriving at sensor involves: (i) Colour of light source; (ii) Colour of surface

Credit: Electromagnetic spectrum by Horst Frank, Jailbird and Phrood. Under CC 3.0 License

Bayer Grid/Filter

- Bayer arrangement of color filters on a camera sensor
- Filter pattern is 50% green, 25% red and 25% blue
- To obtain full-colour image, **demosaicing** algorithms used - surrounding pixels used to estimate values for a particular pixel.





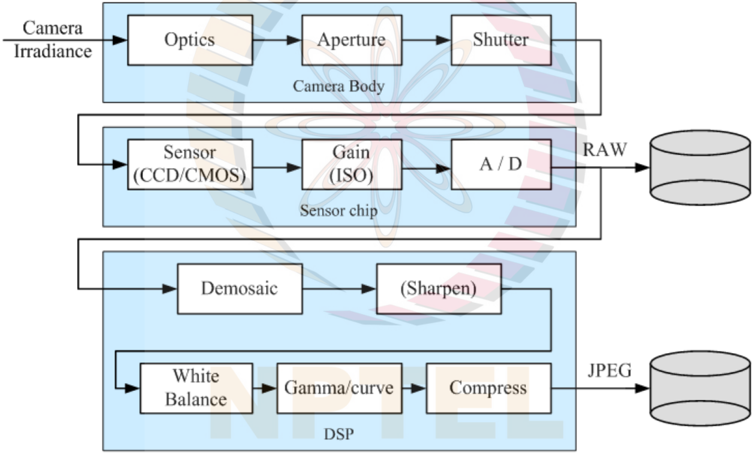
Credit: https://en.wikipedia.org/wiki/Bayer_filter

Question

On Colour

If visible light spectrum is VIBGYOR, why RGB colour representation?

Image Sensing Pipeline[2]



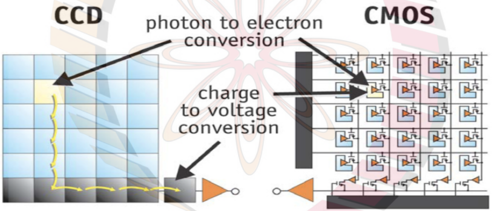
[2]Credit: Szeliski, *Computer Vision: Algorithms and Applications*, 2010

Digital Image Sensing

CCD vs CMOS: What's the difference?

Digital Image Sensing

CCD vs CMOS: What's the difference?



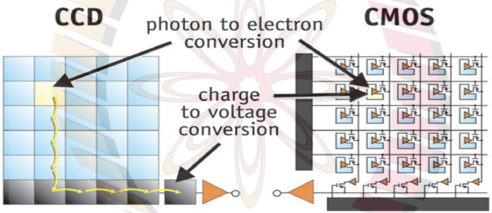
CCD

- Move photogenerated charge from pixel to pixel, and convert it to voltage at output node
- An analog-to-digital converter (ADC) then turns each pixel's value into a digital value

Photo Credit: Litwiller, *CMOS vs. CCD: Maturing Technologies, Maturing Markets*, 2005

Digital Image Sensing

CCD vs CMOS: What's the difference?



CMOS

- CMOS convert charge to voltage inside each element
- Uses several transistors at each pixel to amplify and move the charge using more traditional wires
- CMOS signal is digital, so it needs no ADC

Digital Image Sensor Properties

Shutter speed:	Controls the amount of light reaching the sensor (also called *exposure time*)
Sampling pitch:	Physical spacing between adjacent sensor cells on the imaging chip
Fill factor:	Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches)
Chip size:	Size/area of the chip
Analog gain:	Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras)
Sensor noise:	Noise from various sources in the sensing process
Resolution:	How many bits for each pixel, decided by analog-to-digital conversion module
Post-processing:	Digital image enhancement methods often used before compression and storage of captured image

Digital Image Sensor Properties

Shutter speed:	Controls the amount of light reaching the sensor (also called *exposure time*)
Sampling pitch:	Physical spacing between adjacent sensor cells on the imaging chip
Fill factor:	Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches)
Chip size:	Size/area of the chip
Analog gain:	Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras)
Sensor noise:	Noise from various sources in the sensing process
Resolution:	How many bits for each pixel, decided by analog-to-digital conversion module
Post-processing:	Digital image enhancement methods often used before compression and storage of captured image

Digital Image Sensor Properties

Shutter speed:	Controls the amount of light reaching the sensor (also called *exposure time*)
Sampling pitch:	Physical spacing between adjacent sensor cells on the imaging chip
Fill factor:	Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches)
Chip size:	Size/area of the chip
Analog gain:	Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras)
Sensor noise:	Noise from various sources in the sensing process
Resolution:	How many bits for each pixel, decided by analog-to-digital conversion module
Post-processing:	Digital image enhancement methods often used before compression and storage of captured image

Digital Image Sensor Properties

Shutter speed:	Controls the amount of light reaching the sensor (also called *exposure time*)
Sampling pitch:	Physical spacing between adjacent sensor cells on the imaging chip
Fill factor:	Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches)
Chip size:	Size/area of the chip
Analog gain:	Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras)
Sensor noise:	Noise from various sources in the sensing process
Resolution:	How many bits for each pixel, decided by analog-to-digital conversion module
Post-processing:	Digital image enhancement methods often used before compression and storage of captured image

Digital Image Sensor Properties

Shutter speed:	Controls the amount of light reaching the sensor (also called *exposure time*)
Sampling pitch:	Physical spacing between adjacent sensor cells on the imaging chip
Fill factor:	Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches)
Chip size:	Size/area of the chip
Analog gain:	Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras)
Sensor noise:	Noise from various sources in the sensing process
Resolution:	How many bits for each pixel, decided by analog-to-digital conversion module
Post-processing:	Digital image enhancement methods often used before compression and storage of captured image

Digital Image Sensor Properties

Shutter speed:	Controls the amount of light reaching the sensor (also called *exposure time*)
Sampling pitch:	Physical spacing between adjacent sensor cells on the imaging chip
Fill factor:	Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches)
Chip size:	Size/area of the chip
Analog gain:	Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras)
Sensor noise:	Noise from various sources in the sensing process
Resolution:	How many bits for each pixel, decided by analog-to-digital conversion module
Post-processing:	Digital image enhancement methods often used before compression and storage of captured image

Digital Image Sensor Properties

Shutter speed:	Controls the amount of light reaching the sensor (also called *exposure time*)
Sampling pitch:	Physical spacing between adjacent sensor cells on the imaging chip
Fill factor:	Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches)
Chip size:	Size/area of the chip
Analog gain:	Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras)
Sensor noise:	Noise from various sources in the sensing process
Resolution:	How many bits for each pixel, decided by analog-to-digital conversion module
Post-processing:	Digital image enhancement methods often used before compression and storage of captured image

Digital Image Sensor Properties

Shutter speed:	Controls the amount of light reaching the sensor (also called *exposure time*)
Sampling pitch:	Physical spacing between adjacent sensor cells on the imaging chip
Fill factor:	Active sensing area size as a fraction of the theoretically available sensing area (product of horizontal and vertical sampling pitches)
Chip size:	Size/area of the chip
Analog gain:	Amplification of the sensed signal using automatic gain control (AGC) logic (controlled using ISO setting on cameras)
Sensor noise:	Noise from various sources in the sensing process
Resolution:	How many bits for each pixel, decided by analog-to-digital conversion module
Post-processing:	Digital image enhancement methods often used before compression and storage of captured image

Popular Question: *"With smartphones, do you need DSLR cameras?"*[3]

[3]Source: VSBytes.com

Popular Question: *"With smartphones, do you need DSLR cameras?"*[3]

- **DSLR - Digital Single Lens Reflex camera:** Uses a mirror mechanism to reflect light from lens to a viewfinder, or let light fully pass onto image sensor by moving the mirror out of the way
- Essentially a comparison between mirror and mirrorless cameras
- *Pros of mirrorless cameras:* Accessibility, portability, Low cost
- *Pros of DSLRs:* Picture quality, Versatility and functionality, Physical shutter, Variable focal length/aperture

Credit: http://www.pixelrajeev.com/

[3]Source: VSBytes.com

Sampling and Aliasing

Credit: Wikimedia Commons

- **Shannon's Sampling Theorem:** $f_s \geq 2f_{\text{max}}$, where f_s is sampling rate, and f_{max} is maximum frequency in signal, also called **Nyquist frequency**
- Frequencies above Nyquist frequency or when Shannon's sampling rate is not met $\implies$ *aliasing* happens

Sampling and Aliasing

Credit: Wikimedia Commons

- **Shannon's Sampling Theorem:** $f_s \geq 2f_{max}$, where f_s is sampling rate, and f_{max} is maximum frequency in signal, also called **Nyquist frequency**
- Frequencies above Nyquist frequency or when Shannon's sampling rate is not met $\implies$ *aliasing* happens
- Why is aliasing bad?

Sampling and Aliasing

Credit: Wikimedia Commons

- **Shannon's Sampling Theorem:** $f_s \geq 2f_{\max}$, where f_s is sampling rate, and $f_{\max}$ is maximum frequency in signal, also called **Nyquist frequency**
- Frequencies above Nyquist frequency or when Shannon's sampling rate is not met $\implies$ *aliasing* happens
- Why is aliasing bad? Creates issues while downsampling and upsampling an image
- More coming in later lectures

Colour Space Representations

- Popular colour spaces: RGB, CMYK
- Additive colours: R, G, B
- Subtractive colours: C, M, Y
- Other colour spaces: XYZ, YUV, Lab, YCbCr, HSV
- Standards established by Commission Internationale d'Eclairage (CIE)
- Understanding of colour spaces important in printing industry


Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

For more information:

- https://www.tutorialspoint.com/dip/introduction_to_color_spaces.htm
- https://ciechanow.ski/color-spaces/

Image Compression

- Last stage in a camera's processing pipeline
- Convert signal into YCbCr (or variant), compress luminance with higher fidelity than chrominance
- Most common compression technique: Discrete Cosine Transform (DCT), used in MPEG and JPEG
- DCT, a variant of Discrete Fourier Transform - a reasonable approximation of eigendecomposition of image patches
- Videos also use block-level motion compensation
- Compression quality measured using Peak Signal-to-Noise Ratio (PSNR):

$$PSNR = 10 \log_{10} \frac{I_{max}^2}{MSE}$$

$$\text{where } MSE = \frac{1}{n} \sum_x \left[I(x) - \hat{I}(x) \right]^2$$

where $\hat{I}$ is compressed version of I

Image Compression

- Last stage in a camera's processing pipeline
- Convert signal into YCbCr (or variant), compress luminance with higher fidelity than chrominance
- Most common compression technique: **Discrete Cosine Transform (DCT)**, used in MPEG and JPEG
- DCT, a variant of Discrete Fourier Transform - a reasonable approximation of eigendecomposition of image patches

- Videos also use block-level motion compensation
- Compression quality measured using **Peak Signal-to-Noise Ratio (PSNR)**:

$$PSNR = 10 \log_{10} \frac{I_{max}^2}{MSE}$$

where $MSE = \frac{1}{n} \sum_x \left[I(x) - \hat{I}(x) \right]^2$

where $\hat{I}$ is compressed version of I

Image Compression

- Last stage in a camera's processing pipeline
- Convert signal into YCbCr (or variant), compress luminance with higher fidelity than chrominance
- Most common compression technique: **Discrete Cosine Transform (DCT)**, used in MPEG and JPEG
- DCT, a variant of Discrete Fourier Transform - a reasonable approximation of eigendecomposition of image patches
- Videos also use block-level motion compensation
- Compression quality measured using **Peak Signal-to-Noise Ratio (PSNR)**:

$$PSNR = 10 \log_{10} \frac{I_{max}^2}{MSE}$$

$$\text{where } MSE = \frac{1}{n} \sum_x \left[I(x) - \hat{I}(x) \right]^2$$

where $\hat{I}$ is compressed version of I

Image Compression

- Last stage in a camera's processing pipeline
- Convert signal into YCbCr (or variant), compress luminance with higher fidelity than chrominance
- Most common compression technique: **Discrete Cosine Transform (DCT)**, used in MPEG and JPEG
- DCT, a variant of Discrete Fourier Transform - a reasonable approximation of eigendecomposition of image patches
- Videos also use block-level motion compensation
- Compression quality measured using **Peak Signal-to-Noise Ratio (PSNR)**:

$$PSNR = 10 \log_{10} \frac{I_{max}^2}{MSE}$$

$$\text{where } MSE = \frac{1}{n} \sum_x \left[I(x) - \hat{I}(x) \right]^2$$

where $\hat{I}$ is compressed version of I

Homework

Readings

- Chapter 2, Szeliski, *Computer Vision: Algorithms and Applications*
- Other links provided on respective slides
- (Optional, if you want to know about geometric formation of images) Chapters 1-5, Forsyth and Ponce, *Computer Vision: A Modern Approach*

References

- Dave Litwiller. "CMOS vs. CCD: Maturing Technologies, Maturing Markets-The factors determining which type of imager delivers better cost performance are becoming more refined.". In: *Photonics Spectra* 39.8 (2005), pp. 54–61.
- Richard Szeliski. *Computer Vision: Algorithms and Applications*. Texts in Computer Science. London: Springer-Verlag, 2011.
- David Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. 2 edition. Boston: Pearson Education India, 2015.
- VSBytes Team. *DSLR Cameras vs Smartphone - Which of the two cameras is better?* May 2019. URL: https://vsbytes.com/dslr-vs-smartphone-camera/ (visited on 04/14/2020).