Image Formation

CSE P576

Vitaly Ablavsky
Image Formation

- Light, Optics, Sensing
- The Digital Camera

[ Szeliski Chapter 2 ]
2.2 Photometric image formation

Image formation is the process by which light emitted by one or more light sources is reflected from an object's surface and directed towards the camera. A portion of this light is captured by the sensor. This simplified model ignores multiple reflections, which often occur in real-world scenes.

The intensity of a light source falls off with the square of the distance between the source and the object being lit, because the same light is being spread over a larger (spherical) area. A light source may also have a directional falloff (dependence), but we ignore this in our simplified model.

Area light sources are more complicated. A simple area light source such as a fluorescent ceiling light fixture with a diffuser can be modeled as a finite rectangular area emitting light equally in all directions (Cohen and Wallace 1993; Sillion and Puech 1994; Glassner 1995). When the distribution is strongly directional, a four-dimensional lightfield can be used instead (Ashdown 1993).

A more complex light distribution that approximates, say, the incident illumination on an object sitting in an outdoor courtyard, can often be represented using an environment map (Greene 1986) (originally called a reflection map (Blinn and Newell 1976)). This representation maps incident light directions \( \hat{v} \) to color values (or wavelengths, \( L(\hat{v}); \)), (2.80)

and is equivalent to assuming that all light sources are at infinity. Environment maps can be represented as a collection of cubical faces (Greene 1986), as a single longitude–latitude map (Blinn and Newell 1976), or as the image of a reflecting sphere (Watt 1995). A convenient way to get a rough model of a real-world environment map is to take an image of a reflective mirrored sphere and to unwrap this image onto the desired environment map (Debevec 1998). Watt (1995) gives a nice discussion of environment mapping, including the formulas needed to map directions to pixels for the three most commonly used representations.
Light and Colour

• Light is electromagnetic radiation in the 400-700nm band
• This is the peak in the spectrum of sunlight passing through the atmosphere
• Newton’s Prism experiment showed that white light is composed of all frequencies
• Black is the absence of light!
Spectral Power Distribution

• The spectral distribution of energy in a light ray determines its colour

• Surfaces reflect light energy according to a spectral distribution as well

• The combination of incident spectra and reflectance spectra determine the light colour
Spectral Reflectance Example

\[ E(\lambda) \]

\[ E(\lambda)S(\lambda) \]

\[ S(\lambda) \]
Surface Reflectance

- Reflected intensity also depends on geometry: surface orientation, viewer position, shadows, etc.

It also depends on surface properties, e.g., diffuse or specular.
Diffuse and Specular

- A pure **mirror** reflects light along a line symmetrical about the surface normal.
- A pure **diffuse** surface scatters light equally in all directions.

**Pure Mirror Reflection**
\[ \theta_i = \theta_r \]

**Lambertian Reflection** (Diffuse)

**Specular** surfaces directly reflect over a small angle.
Contribution of each of the lighting/reflectance components

Figure credit: https://en.wikipedia.org/wiki/Phong_shading
Diffuse Reflection

- Light is reflected equally in all directions (Lambertian surface)
- But the amount of light reaching unit surface area depends on the angle between the light and the surface...
Radiance & Irradiance

Radiance: \( \text{Watts} \frac{\text{m}^2 \cdot \text{sr}}{\text{sr}} \) with projected area

Irradiance: \( \text{Watts} \frac{\text{m}^2}{\text{steradian}} \)

I: illuminance, \( i: \) light intensity

\( K_a, K_d, K_s \): surface reflection coefficients

\( \frac{1}{\cos \theta} \): unit length

10.1
Specular Reflection

- Light reflected strongly around the mirror reflection direction
- Intensity depends on viewer position
Model of Specular Reflection

\[ I_s = k_s \cdot i_s \cdot \cos \phi \]

reflection \hspace{1cm} intensity
Phong Illumination Model

- Includes ambient, diffuse and specular reflection

\[ I = k_a i_a + k_d i_d \cos \theta + k_s i_s \cos^\alpha \phi \]
Reflectance in Vision
Reflectance in Vision

• More complex models than Phong are possible with reflected intensity at a given ray an arbitrary function of the surface geometry and lights, see Szeliski 2.2.2 (BRDFs)
• For Computer Vision, understanding reflection can help us to infer shape, e.g., **shape from shading** and **photometric stereo**, we will revisit this later in the course
Optics

Camera Obscura = “dark room”
Clifton Observatory

A working camera obscura open to the public
Pinhole Camera

- All rays pass through a single point (the pinhole)

- Similar triangles

\[ f \quad Z \]

\[ u \]

\[ X \]
Pinhole Camera (Matrix Form)
Focal Length

- For a fixed sensor size, focal length determines the field of view (fov)

\[ \theta_1, \theta_2 \]

\[ f_1, f_2 \]

Q: What is the field of view of a full frame (35mm) camera with 50mm lens? 100mm lens?
Field-of-View Computations

\[ \tan \frac{\theta}{2} = \frac{y}{2f} \]

\[ \theta = 2 \cdot \arctan \frac{y}{2f} \quad \text{e.g.} \quad 35 \text{mm} \]

\[ \theta \big| y=35 \text{mm}, f=100 \text{mm} = 2 \cdot \arctan \frac{35}{200} = 19.8^\circ \]
Focal Length

28 mm

35 mm

50 mm

70 mm
Finite Aperture

- A real camera must have a finite aperture to get enough light, but this causes **blur** in the image.

Solution: use a **lens** to focus light onto the image plane.
Lens Basics

- A lens focuses rays from infinity at the focal length of the lens.
- Points passing through the centre of the lens are not bent.

To focus closer, we have to move the image plane back.

We can use these 2 properties to find the lens equation.
Thin Lens Model
Lens Basics

• Note that lenses focus all rays from a **plane** in the world.

• Objects off the plane are blurred depending on distance.
Effect of Aperture

Smaller aperture $\Rightarrow$ smaller blur, larger \textit{depth of field}
Depth of Field

- Photographers use large apertures to give small depth of field

Aperture size = \( f/N \), \( \Rightarrow \) large \( N \) = small aperture
Shutter Speed

- Shutter Speed: 1/1000
- Shutter Speed: 1/125
- Shutter Speed: 1/25
- Shutter Speed: 1/15
Real Lenses

- Multiple stages of positive and negative elements with differing refractive indices
- Deal with issues such as chromatic aberration (different colours bent by different amounts), vignetting (light fall off at image edge) and sharp imaging across the zoom range
Sensors

CMOS (or CCD)  Retina
Colour Perception

Cone responses

\[ E(\lambda) \]

\[ S(\lambda) \]

Cone excitation (multiply and add):

\[ \rho_{\text{red}} = \int R_{\text{red}}(\lambda)E(\lambda)S(\lambda) \, d\lambda \]
Digital Sensor

- Analogue image is sampled by a CMOS (or CCD) sensor
- RGB colour filters arranged in a “Bayer” pattern
- Counts from this sensor are camera RAW
- For viewing we need an RGB value per pixel
Demosaicing

- Each colour channel has different information:

How can we fill in the missing information?
Demosaicing

• Simple interpolation causes colour errors

Bilinear interpolation

Bennet et al 2006
(local 2 colour prior)

• Many techniques use edge information from the densely sampled green channel, and some form of image prior
• It can also been tackled via a data-driven approach, e.g., [Gharbi et al. 2016]
The Digital Image

- e.g., arranged in memory with RGB pixels stored in rows:
  
  \[
  \begin{bmatrix}
  \text{Pixel} & \text{Pixel} & \text{Pixel} & \ldots \\
  \text{Row} & \text{Row} & \text{Row} & \ldots \\
  \end{bmatrix}
  \]

- Many other possibilities, e.g., BGR, RGBA pixels, row/column major ordering, and rows or columns aligned to power of 2 boundaries.
Digital Camera Processing

- Main stages in a digital camera

![Image Sensing Pipeline Diagram](image_url)

1. Light falling on an imaging sensor is usually picked up by an active sensing area, integrated for the duration of the exposure (usually expressed as the shutter speed in a fraction of a second, e.g., \( \frac{1}{125}, \frac{1}{60}, \frac{1}{30} \)), and then passed to a set of sense amplifiers.

2. The two main kinds of sensor used in digital still and video cameras today are charge-coupled device (CCD) and complementary metal oxide on silicon (CMOS).

3. In a CCD, photons are accumulated in each active well during the exposure time. Then, in a transfer phase, the charges are transferred from well to well in a kind of “bucket brigade” until they are deposited at the sense amplifiers, which amplify the signal and pass it to an analog-to-digital converter (ADC). Older CCD sensors were prone to blooming, when charges from one over-exposed pixel spilled into adjacent ones, but most newer CCDs have anti-blooming technology (“troughs” into which the excess charge can spill).

4. In CMOS, the photons hitting the sensor directly affect the conductivity (or gain) of a photodetector, which can be selectively gated to control exposure duration, and locally amplified before being read out using a multiplexing scheme. Traditionally, CCD sensors outperformed CMOS in quality sensitive applications, such as digital SLRs, while CMOS was better for low-power applications, but today CMOS is used in most digital cameras.

5. The main factors affecting the performance of a digital image sensor are the shutter speed, sampling pitch, fill factor, chip size, analog gain, sensor noise, and the resolution (and quality).

6. In digital still cameras, a complete frame is captured and then read out sequentially at once. However, if video is being captured, a rolling shutter, which exposes and transfers each line separately, is often used. In older video cameras, the even fields (lines) were scanned first, followed by the odd fields, in a process that is called interlacing.

[ Szeliski 2.3 ]
White Balance

- Humans are good at adapting to global illumination conditions: you would still describe a white object as white whether under blue sky or candle light.
- However, when the picture is viewed later, the viewer is no longer correcting for the environment and the illuminant colour typically appears too strong.
- **White balancing** is the process of correcting for the illuminant

- A simple white balance algorithm is to assume the scene is grey on average “greyworld”, state of the art methods use learning, e.g., Barron ICCV 2015
Gamma Correction

• Equal steps in luminance ≠ equal in perceived brightness

linear luminance (raw) 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
equal brightness steps 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

• Equal steps in perceived brightness are achieved by increasingly large steps in luminance (sensor counts)
• Human brightness perception (V) follows a power law:

\[ \mathcal{L} = V^\gamma \]

• Using raw sensor counts wastes bits as we can’t differentiate the large values → use **gamma corrected encoding** that allocates more bits to smaller values
Contrast Sensitivity

• Human visual system is most sensitive to mid-frequencies
Discrete Cosine Transform

- Basis functions used in JPEG

\[ X(m, n) = \alpha_m \alpha_n \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} x(k, l) \cos \left( \frac{(2k + 1)m\pi}{2K} \right) \cos \left( \frac{(2l + 1)n\pi}{2L} \right) \]

- Energy is concentrated in the low frequency components

- Efficient algorithm to compute (similar to FFT)

8x8 basis functions
Coefficient Quantisation

- DCT coefficients $F(u, v)$ are quantised according to a quantisation table.
- High frequencies are less important (high factor).
- Quantisation table entries determine the “lossiness” of the compression.

$$F^Q[u, v] = \text{Round} \left( \frac{F[u, v]}{Q[u, v]} \right)$$

<table>
<thead>
<tr>
<th>$Q[u, v]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 11 10 16 24 40 51 61</td>
</tr>
<tr>
<td>12 12 14 19 26 58 60 55</td>
</tr>
<tr>
<td>14 13 16 24 40 57 69 56</td>
</tr>
<tr>
<td>14 17 22 29 51 87 80 62</td>
</tr>
<tr>
<td>18 22 37 56 68 109 103 77</td>
</tr>
<tr>
<td>24 35 55 64 81 104 113 92</td>
</tr>
<tr>
<td>49 64 78 87 103 121 120 101</td>
</tr>
<tr>
<td>72 92 95 98 112 100 103 99</td>
</tr>
</tbody>
</table>
The quantized DC coefficient is encoded as the difference from the DC term of the previous block to account for the strong correlation between adjacent CD coefficient.
YCbCr

- Separates luminance (Y) from chrominance (Cb, Cr) = colour

\[
Y' = 16 + 65.5R' + 128.6G'' + 25.0B''
\]
\[
Cb = 128 - 37.8R' - 74.2G'' + 112B''
\]
\[
Cr = 128 + 112.0R' - 93.8G'' - 18.2B''
\]

- Linear transform of RGB
- Primes = gamma correction

[ https://en.wikipedia.org/wiki/YCbCr ]
Blurring CbCr

sigma = 1.0
Blurring CbCr

sigma = 2.0
Blurring CbCr

\[ \text{sigma} = 4.0 \]
Blurring CbCr

\[ \text{sigma} = 8.0 \]
Blurring CbCr

\[ \text{sigma} = 16.0 \]
Blurring CbCr

sigma = 32.0
Blurring Y

\[ \text{sigma} = 1.0 \]
Blurring Y

\[ \text{sigma} = 2.0 \]
Blurring Y

\[ \text{sigma} = 4.0 \]
Blurring Y

\[ \text{sigma} = 8.0 \]
Blurring Y

\[ \text{sigma} = 16.0 \]
Subsampling CbCr vs Y

Original

Chrominance 1/8 scale

Luminance 1/8 scale
Compressibility of Chrominance

- Cb+Cr are transmitted at 1/2 size for JPEG

- Note that human vision uses a similar transform to this (opponent colours), also we have fewer cones than rods
Next Lecture

• Filtering and Pyramids
• Features and Matching