Contents

TEST

Adam Hawley

April 9, 2019

0.1 Admin

0.1.1 Exam Format

One 2 hour exam at the beginning of Spring Term.

- One question worth 20% (basic knowledge, shorter questions)
- Two questions worth 40% (more advanced, longer questions)

Note: Programming will NOT be examined

0.1.2 Reading

- Nielsen, “Visual Computing: Geometry, Graphics and Vision”, Charles River Media, 2005
- Prince, “Computer Vision”, Cambridge University Press

1 Lecture 1: Introduction

Contents

1.1 Conventions

- Matrices written in uppercase bold eg: $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
- Matrices are also indexed (row, column) from top left starting at 1.
- Individual elements are written as lowercase:
$$a_{11} = 1, a_{12} = 2, a_{21} = 3, a_{22} = 4$$
- Vectors are simply matrices with a single row or column represented by a lower-case bold symbol:
$$\mathbf{a} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} OR \ \ \mathbf{a} = \begin{bmatrix} x & y & z \end{bmatrix}$$

1.2 Transformations

1.2.1 Translation

Translation involves ***component-wise addition***.

For example:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d_x \\ d_y \end{bmatrix}$$

1.2.2 Scaling

Scaling involves ***multiplication by a scalar***.

For example:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s \begin{bmatrix} x \\ y \end{bmatrix}$$

Or to have a separate dimensional scalars:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}$$

1.2.3 Rotation

Rotation involves ***matrix multiplication***. Using the formula below to give a 2D rotation counter clockwise by angle θ:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Rotation preserves distances:

$$||\mathbf{x}|| = ||\mathbf{Rx}||$$
$$||\mathbf{x}||^2 = x_1^2 + x_2^2 + ... + x_n^2 = \mathbf{x}^T\mathbf{x}$$
$$\mathbf{x}^T\mathbf{x} = (\mathbf{Rx})^T(\mathbf{Rx})$$
$$\mathbf{x}^T\mathbf{Ix} = (\mathbf{x})^T(\mathbf{R}^T)(\mathbf{Rx})$$
$$= \mathbf{x}^T(\mathbf{R}^T\mathbf{R})\mathbf{x}$$

And hence:

$$\mathbf{R}^T\mathbf{R} = \mathbf{I}$$
$$\mathbf{R}^T\mathbf{R}\mathbf{R}^{-1} = \mathbf{I}\mathbf{R}^{-1}$$
$$\mathbf{R}^T = \mathbf{R}^{-1}$$

So to get the inverse of a rotation you only need to transpose it.

1.2.4 Affine Transformations

All of the previously mentioned transformations have been **affine** transformations. These are transformations which can be expressed as matric multiplication and addition:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

For 2D, **A** is a 2x2 matrix and **b** a 2x1 vector.

For 3D, **A** is a 3x3 matrix and **b** a 3x1 vector.

1.3 Homogeneous Coordinates

1.3.1 Defining Homogenous Coordinates

Representing the affine transformations can be done by moving from Cartesian to homogeneous coordinates.

All 2D affine transformations can be then represented by a 3x3 matrix (or 4x4 for 3D)

$$\begin{bmatrix} x \\ y \end{bmatrix} \textit{becomes} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

1.3.2 Affine Transformations as Homogeneous Coordinates

The extra dimension allows uniform treatment of transformations:

Translation: $\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & d_x \\ 0 & 1 & d_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$

Scaling: $\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$

Rotation: $\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$

In general, any affine transformation:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

Can be expressed as:

$$\begin{bmatrix} y \\ 1 \end{bmatrix} = \left[\begin{array}{ccc|c} & \mathbf{A} & & \mathbf{b} \\ 0 & \dots & 0 & 1 \end{array}\right] \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}$$

Note: Two homogeneous points are equal if they differ only by a scale factor.

2 Lecture 2: Transformations

2.1 Types of Sensor

There are many different types of sensors used in computer vision:

- Optical
  - CCD, Photodiodes, Photomultipliers
- Infra-red (thermal imaging cameras)
  - CCD (Cooled), Photodiodes
- Synthetic Aperture Radar (SAR)
  - Radar, Antenna
- Range Sensors
  - Laser & Photodiode
- MRI
  - Magnetic field gradients applied causing production of rotating magnetic field which can be measured.
- PET/CAT
  - Simulated radiation emission via magnetic field or radio isotope.

2.2 From Light to Images

2.2.1 Definitions

- ***Irradiance***: power incident on a surface (power per unit area).
- ***Radiance***: power travelling from a source (power per unit solid angle per unit projected source area).

2.2.2 Charge-Coupled Devices

The most common device for digitising image information is a charge-coupled device (CCD). They are made up of a square array of solid-state capacitors:

From the photo-electric effect photons of light knock out electrons from the upper plate and hence charge accumulates in the electron capture "*wells*". Using a shift register, capacitors transfer their contents to the appropriate neighbour. The final capacitor dumps the charge for each capacitor into the analogue-digital converter (ADC). The dump happens once for each capacitor until the contents of each one has been read.

The ideal reading from a CCD would be as simple as "$output = gain \cdot input$". See figure 2. However, a typical reading from a CCD also includes bias and ccd.png

Figure 1: The flow of electrons inside a CCD

ccdResponse1.png

Figure 2: Ideal response of a CCD

`ccdResponse2.png`

Figure 3: Typical response of a CCD

noise (as can be seen in figure 3). (“$output = gain \cdot input + bias + noise$”) The noise mostly comes from the following sources:

- Dark Current - Thermal noise.
- Photon Noise - Quantum noise.
- Quantisation of pixels.
- Amplifier

2.3 Saturation

The electron capture wells inside the CCD have a finite capacity $\approx 700\epsilon^{-}/\mu m^2$. That means that there is roughly a maximum of 100,000 e^{-} per well in a typical CCD. When a well is full, the pixel is said to be saturated. CCDs have a small illumination range (this is what *HDR imaging* attempts to overcome). In an 8-bit image, a pixel brightness of 255 indicates saturation.

2.4 Noise

2.4.1 Signal to Noise Ratio

The level of noise relative to the signal strength is measured by the *signal to noise ratio*. This is often expressed in Decibels (10 times the logarithm of the ratio):

$$SNR(dB) = 10 \log_{10} \frac{signalpower}{noisepower}$$

10dB is a tenfold increase in power, 3dB is a doubling of signal power.

2.4.2 Photon Noise (Shot noise)

- Light sources emit light in photons.
  - β photons/second on average
  - However the photons are emitted randomly.
- They follow a Poisson distribution of time t.
  - Therefore there is a mean intensity of $\mu = \beta t$
  - And a standard deviation of $\sigma = \sqrt{\beta t}$

Note: At mid-brightness levels the *Signal to Noise Ratio*(SNR) for photon noise is around 23dB.

2.4.3 Dark Current

Dark Current is a product of thermal energy from within the CCD. Electrons are displaced over time that are independent of light actually falling on the sensor (therefore this effect builds up with exposure). The effect also obviously increases at higher environmental temperatures. The manufacturing process means certain electron wells may have higher than average dark current (*hot pixels*). *Dark frame subtraction* can be used to remove an estimate of the mean pattern of dark current. However, dark current is also random and hence why it still affects CCD readings. The randomness of the dark current follows the Poisson distribution just as the photon noise did.

Dark current can be easily shown by taking a picture in complete darkness and inspecting the images to see that it is unlikely that all the pixels will be of absolute darkness.

The SNR for dark current in mid-light levels is approximately 38dB.

2.4.4 Quantisation noise

Pixels in photos are digitised to 2^B levels. There is a uniformly distributed error on values of $\pm\frac{1}{2}$. This distribution leads to the following properties:

- Mean: $\mu = 0$
- Standard Deviation: $\sigma = \frac{1}{\sqrt{12}}$
- $SNR = 10log_{10}(2^{B-1} \cdot \sqrt{12})$

$SNR = 26dB$ for 8-bit digitisation, which is better than the photon noise ratio. Photon noise is the limiting factor of CCDs.

2.5 CCDs and Colour

CCDs record photons of all wavelengths, hence they are all monochrome. A colour image requires red, green and blue values.

To get colour when using a CCD, there are three options:

1. Take 3 separate images with a different filter places over the lens. This technique can use filters to allow other frequencies to pass and is hence a multispectral camera. However this method can have issues with motion between frames.
2. Use a beamsplitter and 3 different CCDs, each with their own colour filter. Using the beamsplitter means that one can have instantaneous colour image capture but careful calibration is required to get exact correspondence between pixels of 3 cameras and it is 3 times more expensive.
3. Place filters over individual pixels in a mosaic pattern. This is the most common approach. There are more green than red/blue filters since human vision has higher resolution in green than in red/blue. The colours must be interpolated (the process of *demosaicing*). This method also reduces the effective resolution of the CCD.

2.6 Conclusions

- Measuring a colour image digitally is difficult
- CCDs are generally the best choice
- There are many sources of noise, namely:
  - Photon noise

  - Dark current
  - Quantisation noise
- No easy way to capture colour with a CCD
- Most common option is a Bayer mosaic
  - But this definitely introduces artefacts
  - (in fact — these can be used to detect forgeries)

Hawley, Adam Lecture 3: Photometric Image Formation
Part 2

3 Lecture 3: Photometric Image Formation Part 1

3.1 Brightness

Light reaches the camera from a scene. The amount of light arriving from the scene depends on what is in the scene (strength of light sources, reflectance properties of objects in the scene etc). The pixel brightness reported in an image for a given scene radiance quantity depends on the *exposure* of the image. A camera can control *exposure* in a number of ways.

3.1.1 Formulae

Before discussing the several parameters which can be adjusted to increase total exposure, there are a couple of formulae to be explained and understood first.

- The amount of light captured by a lens is proportional to the area of the aperture:
  $$Area = \pi(\tfrac{D}{2})^2$$
  D is the diameter of the entrance pupil.
- The **f-number**(N) is the ratio of the focal length (f) to the diameter of the entrance pupil:
  $$N = \tfrac{f}{D}$$
  The higher the f-number, the less light that reaches the CCD (smaller diameter)

Substituting D gets:

$$Area = \pi(\tfrac{f}{2N})^2$$

- Area is proportional to the reciprocal square of the f-number
- Photographers have found it convenient to define a discrete set of f-numbers as "stops".
- Each stop represents a halving of area of the aperture relative to the previous stop.

$$A_1 = 2 \cdot A_2$$
$$\frac{\pi(d_1)^2}{4} = 2 \cdot \frac{\pi(d_2)^2}{4}$$
$$(d_1)^2 = 2 \cdot (d_2)^2$$
$$d_2 = \sqrt{(d1)^2 \cdot \frac{1}{2}} = (d_1) \cdot \frac{\sqrt{1}}{\sqrt{2}} = \frac{d_1}{\sqrt{2}}$$

Therefore to halve the area of the aperture the f-number must be multiplied by $\sqrt{2} \approx 1.4$.

Hence, the set of stops is:

$$\tfrac{f}{1}, \tfrac{f}{1.4}, \tfrac{f}{2}, \tfrac{f}{2.8} \cdots$$

Each stop represents a halving of the light intensity from the previous stop. Hence aperture gives us one way to control how much light reaches the camera. However, as one changes the aperture, other complex effects are introduced.

As the f-number reduces (aperture gets larger, more light enters lens), "depth of field" gets smaller and sharp focus is only possible for a limited range of distances.

3.1.2 Aperture

Aperture is the first parameter which can change the overall light exposure onto the CCD. The *aperture* of a camera is the hole through which light travels. Aperture size determines the cone angle of a bundle of rays that come to focus in the image plane.

- Small aperture = less light and more highly collimated rays admitted, sharp focus at image plane.
- Large aperture = more light and uncollimated rays admitted, sharp focus only for rays with a certain focal length.

Aperture sizes are measured in units of *stops* with cameras offering a discrete set of aperture sizes.

3.1.3 Exposure

Exposure is the accumulated physical quality of light applied to the image-plane over a given time.

Product of the image irradiance (light arriving at the CCD) and time:

$$H = E \cdot t$$

Note: In practice, cameras apply a non-linear transformation to intensities so the discretised pixel values will be a function of H. However, this is ignored under the *"linear camera"* assumption.

Hence, another way to control how much light reaches the CCD is to change the exposure time.

However, increasing the exposure can cause motion blur and dark current noise (see Lecture 3 notes) will also increase with time.

3.1.4 Gain

The brightness of an image can also be increased by passing the analogue signal from the CCD through an amplifier. The signal gain of a camera is measured on the *"ISO"* scale. Increasing ISO from its default value of 100 corresponds to increasing signal gain. Modern CCDs can support incredibly high ISO values and hence operation in very low light. Remember, increasing the gain will not affect any saturated pixels, they will simply remain saturated.

3.1.5 Summary

To summarise, there are three main ways one can increase the brightness of an image:

- Increasing the aperture (smaller f-number). (Reduced depth of field)
- Longer exposure (motion blur is worsened).
- Increasing the gain through a large ISO (noise is also amplified)

3.2 High Dynamic Range Imaging

3.2.1 History

- **1850s**: Gustave Le Gray - Chemical process for combining two exposures.
- **1950s**: "Dodging and burning", during film development, make local adjustments to sensitivity.
- **1988-1993**: Technion, Israel — local tonemapping, first HDR video.
- **Modern HDR**: Global HDR **(1993)** — idea of combining images to compute a global luminance map then tonemap, Debevac et al. **(1997)** — the method explained below.

3.2.2 HDR — How is it done today?

High dynamic range imaging captures a sequence of images with different exposures. Combining the images captures a much wider dynamic range. The brightness of pixel i in image j is given by:

$$H_{ij} = E_i \cdot t_j$$
$$\log(H_{ij}) = \log(E_i) + \log(t_j)$$

- t is varied for each image
- H are the known pixel values
- E are the unknown pixel radiances

Each image gives us an estimate of the radiance at a pixel.

$$\log(E_i) = \log(H_{ij}) - \log(t_j)$$

The pixel brightness tells us something about how reliable the estimate is:

- ≈ 255 If a pixel is very bright (close to being saturated then the estimate is unreliable.
- ≈ 0 If a pixel is very dark it will be noisy and unreliable.

`hdrweights.png`

Figure 4: Formula for calculating weights based on pixel value

We give more weight to pixel values near the middle of the range: To estimate the true radiance at every pixel we then simply take a weighted average of the estimates from each image:

$$\log(E_i) = \frac{\Sigma_{j=1}^{K} \omega(H_{ij})(\log(H_{ij}) - \log(t_j))}{\Sigma_{j=1}^{K} \omega(H_{ij})}$$

Saturated pixels are completely excluded. Radiance estimates for a pixel come from images where the exposure is appropriate for the scene radiance. For colour images, separate channels can be processed independently.

3.3 Tonemapping

3.3.1 Gamma Compression

Gamma compression is an effective global (same function applied to all pixels) technique.

$$I_{tonemapped} = cE^{\gamma}$$

- c is a constant scaling where $c > 0$.
- γ is a constant that determines the contrast of the image where $0 < \gamma < 1$.

3.3.2 Gradient-based Techniques

The best techniques are local (output colour only depends on the local region around the pixel). Gradient-based techniques try to preserve local gradient with less concern about overall brightness. This means that they solve for a set of brightness values that minimises error in gradient while remaining between max and min allowable values.