THE UNIVERSITY OF MELBOURNE

Imaging Basics

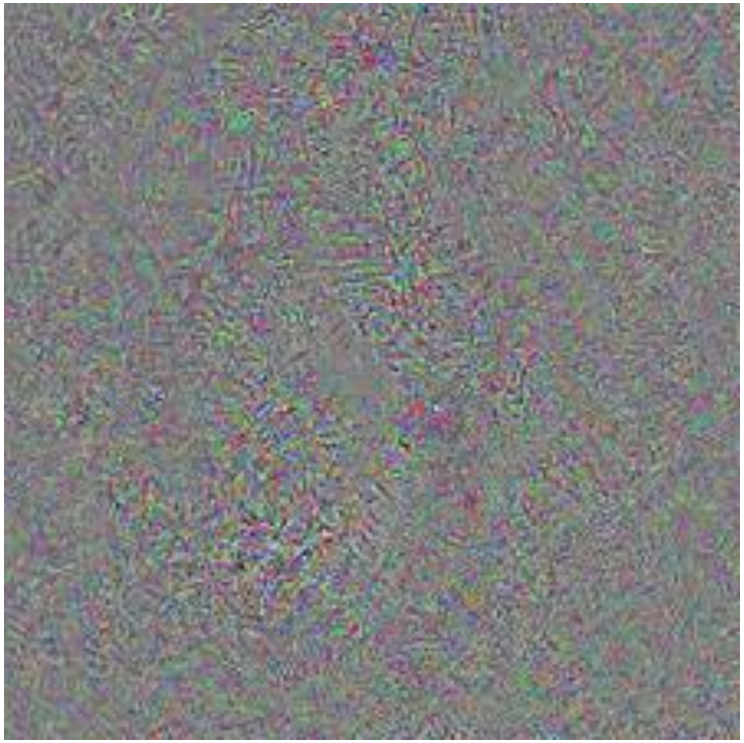
Semester 2, 2021

Kris Ehinger

Demo

- https://kennysong.github.io/adversarial.js/

Dog to hot dog



Network: MobileNet V2
Prediction: English springer (90.08%)

Model: MobileNet V2
Prediction: hot dog (68.88%)

Images: ImageNet, https://kennysong.github.io/adversarial.js/

Outline

- Basics of image formation
- How images are represented digitally
- Image manipulation: resampling

Learning outcomes

- By the end of the lecture, you should be able to:
  - Use the pinhole camera projection model to map between world and image points
  - Explain how camera parameters affect the appearance of an image
  - Explain how images are represented in a computer as 3D tensors
  - Explain why resampling is necessary for image manipulations, and the trade-offs of different methods

Image formation

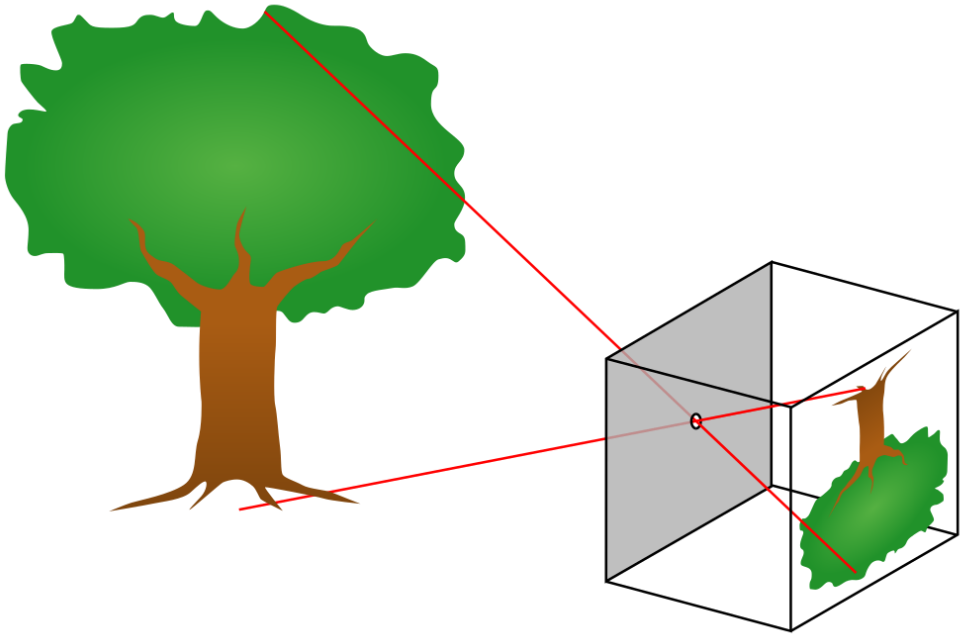
What is an image?

- Pattern formed by light falling on a photosensitive surface
  - Examples of photosensitive surfaces = camera sensor, retina
- Light is reflected off of objects in the world
- 2D projection of a 3D scene

A short account of the eye and nature of vision, James Ayscough (1755)

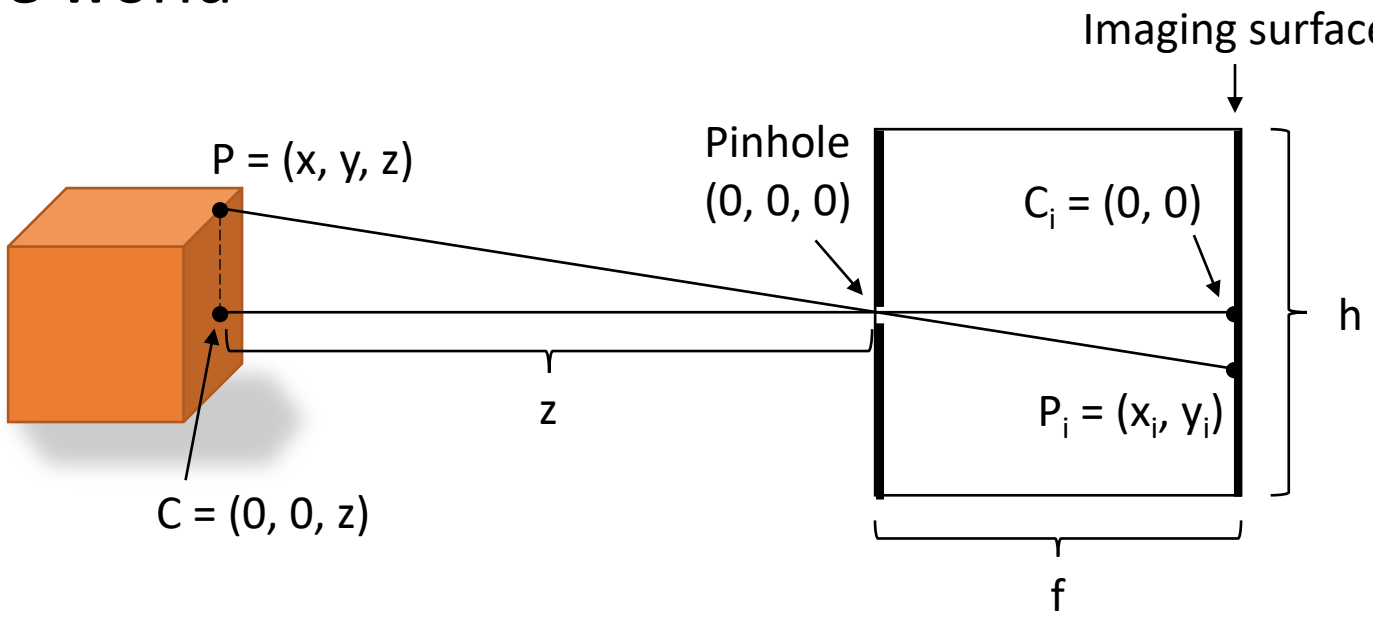
Creating an image

- Simple imaging system: pinhole camera or “camera obscura”



Creating an image

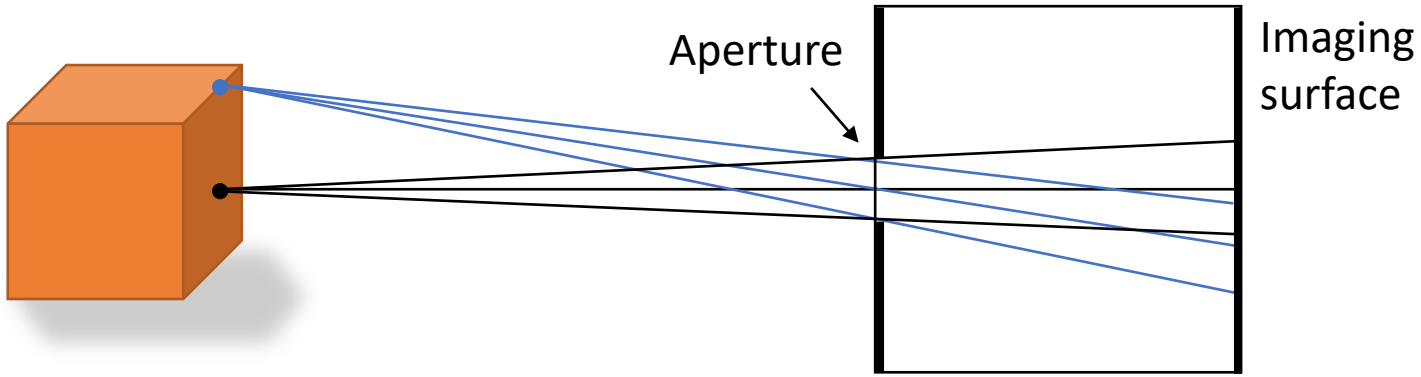
- Every point in the image corresponds to a point in the world



$$x_i = f\frac{x}{z} \qquad y_i = f\frac{y}{z}$$

Creating an image

- Pinhole camera: simple design, not common in practice
  - Why not?
- Instead of a pinhole, most cameras use lenses
  - Why?



Lenses

- Lenses focus light rays onto a single point (F) at a distance (*f*) beyond the lens
- Aperture diameter (D) restricts the range of rays

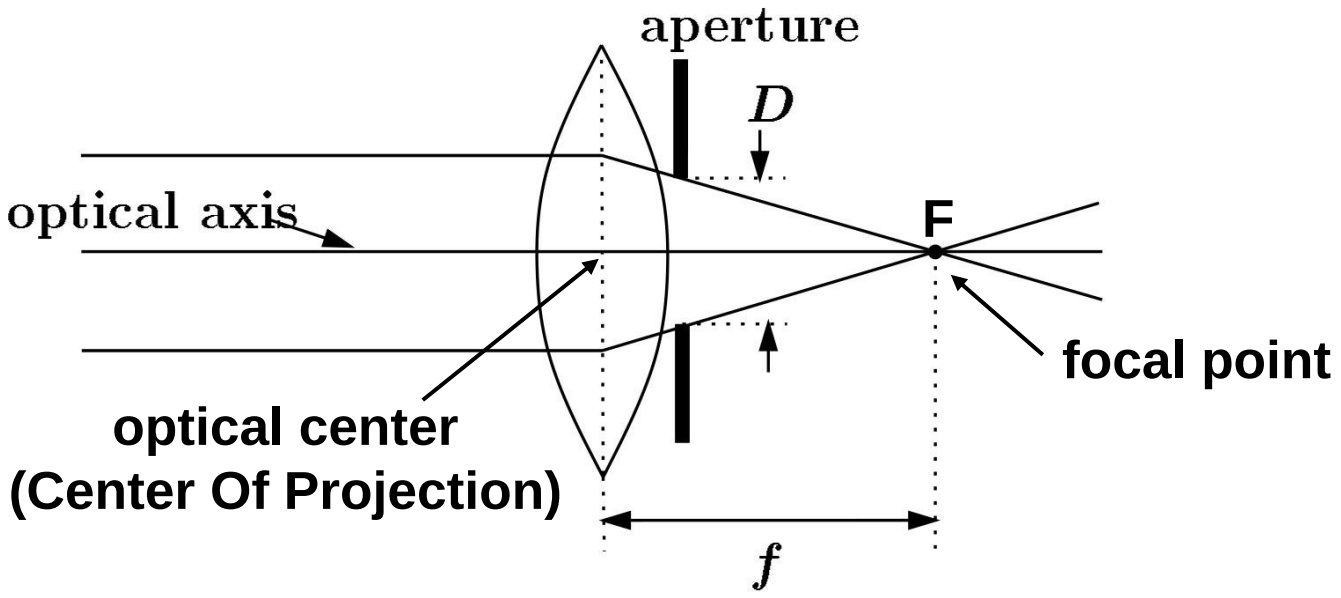


Image: Steve Seitz

Focal length / angle of view

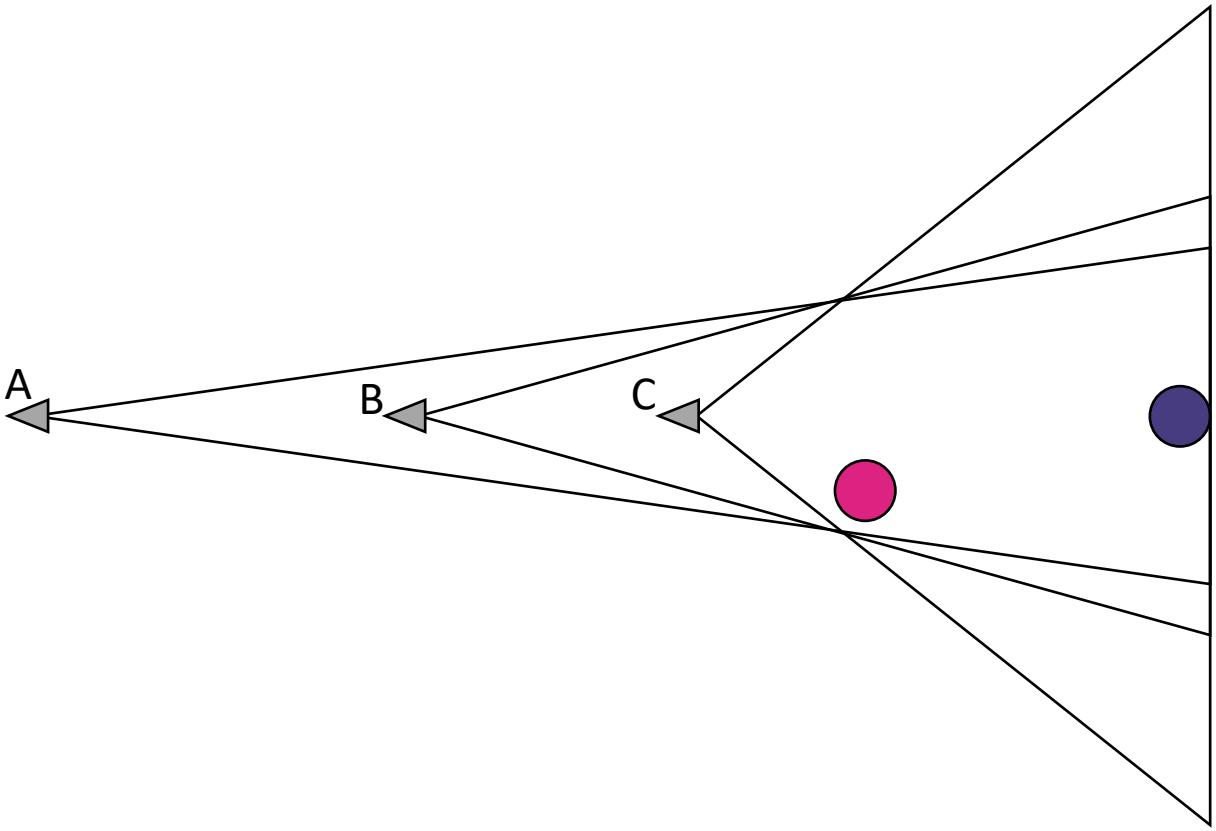


Image: Jcbrooks

Focal length / angle of view

28 mm lens, 65.5° × 46.4°

50 mm lens, 39.6° × 27.0°

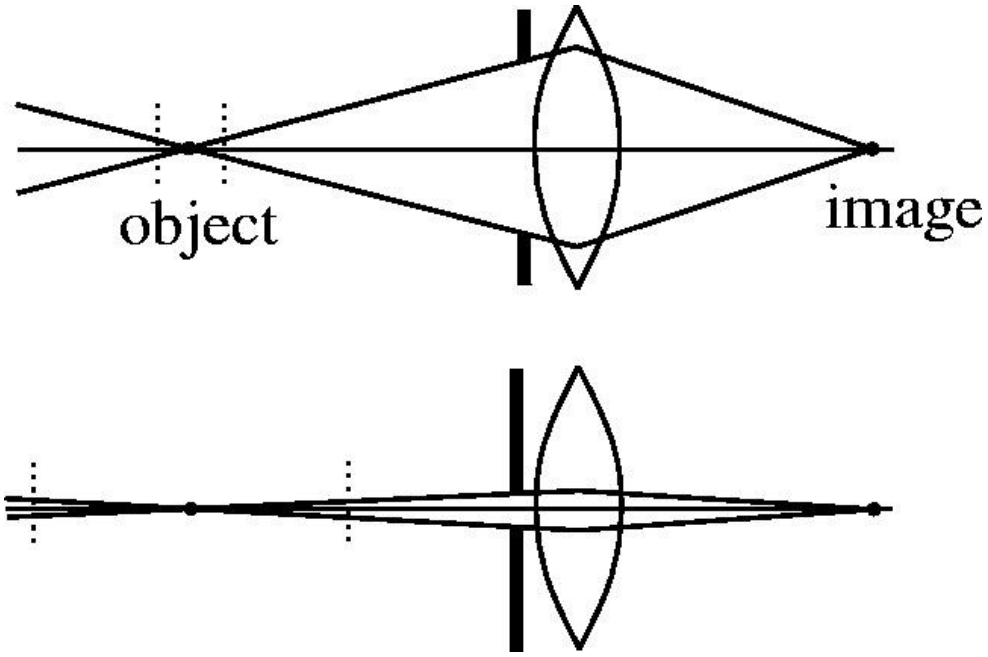
70 mm lens, 28.9° × 19.5°

210 mm lens, 9.8° × 6.5°

https://en.wikipedia.org/wiki/Angle_of_view

Depth of field

- In cameras, aperture size controls depth of field (smaller aperture = greater range of depth in focus)



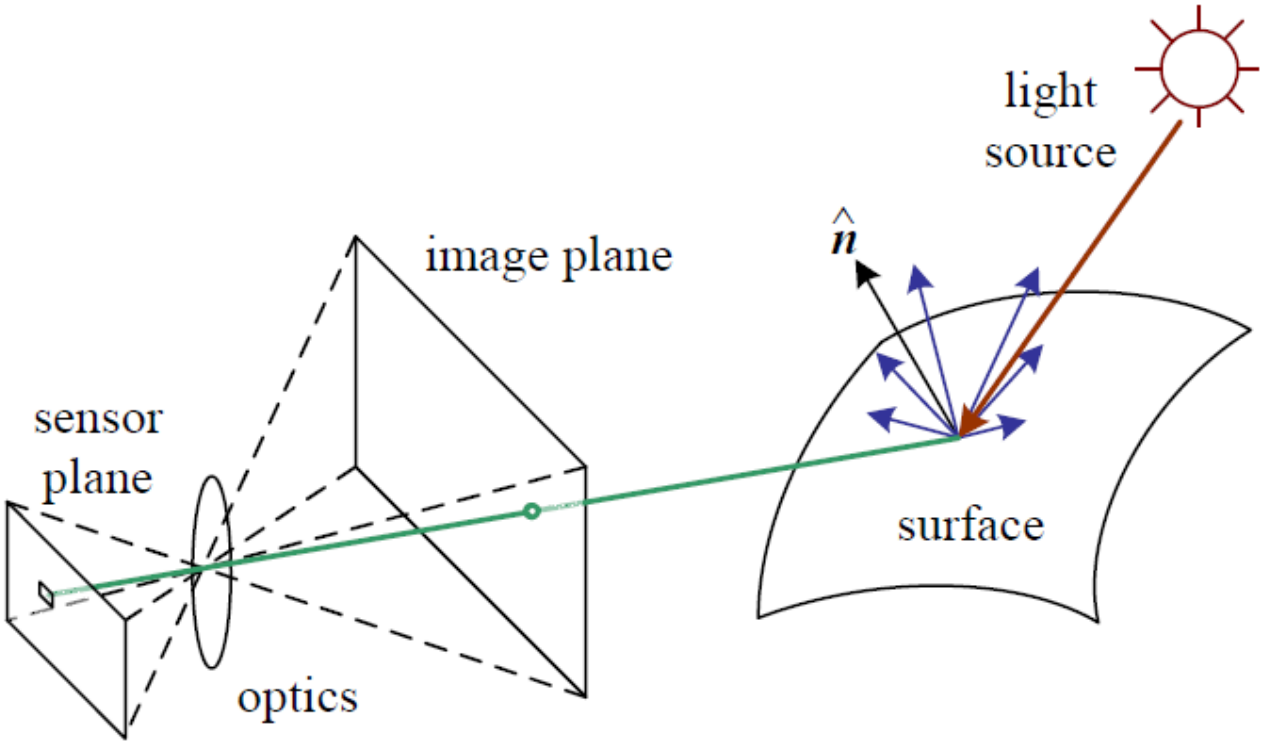
f / 5.6

f / 32

Images: Steve Seitz, Fir0002/Flagstaffotos

Distortion

Image formation model



Computer Vision, Richard Szelisky (2021)

Image formation

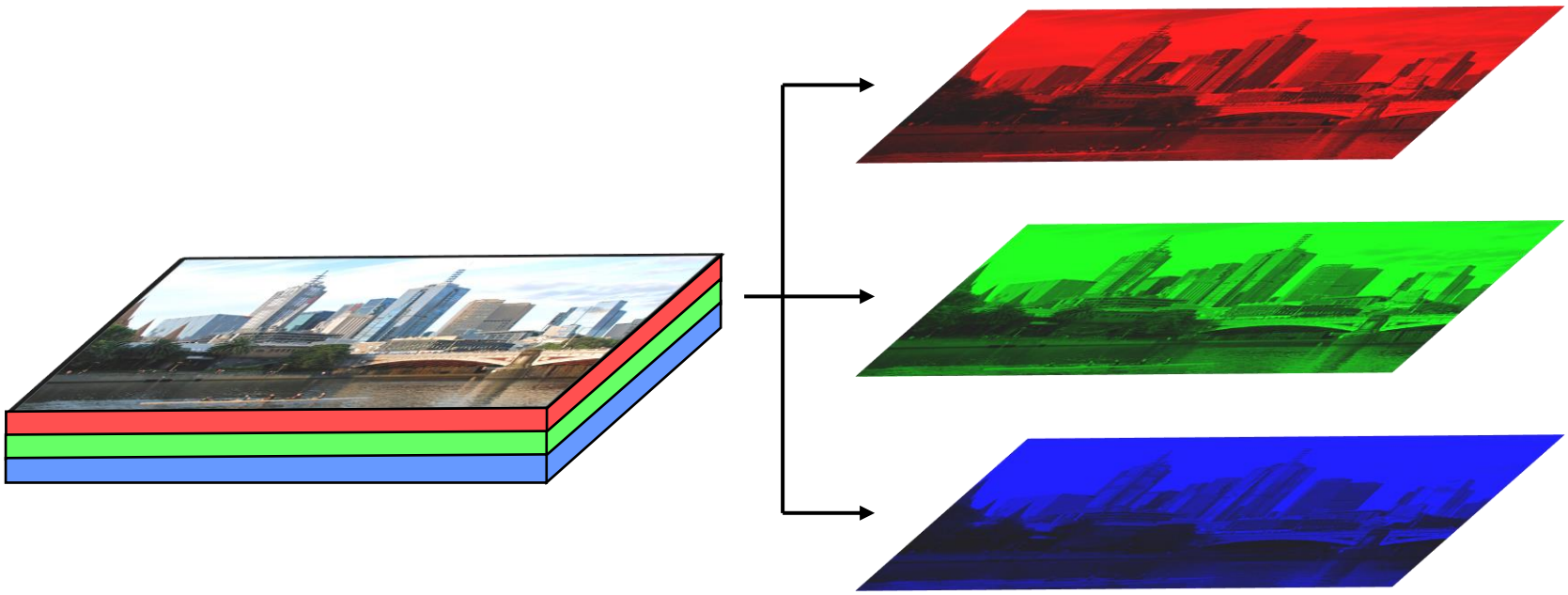
- World parameters
  - Light source
  - Surface properties
- Camera parameters
  - Focal length / angle of view
  - Aperture size / depth of field
  - Lens distortion

Digital images

What is a (digital) image?

- A tensor (3D dimensional array of values)
  - Width x height x channel
  - 3 channels = RGB colour image (red, green, blue)
  - 1 channel = grayscale image

What is a (digital) image?



Colour channels

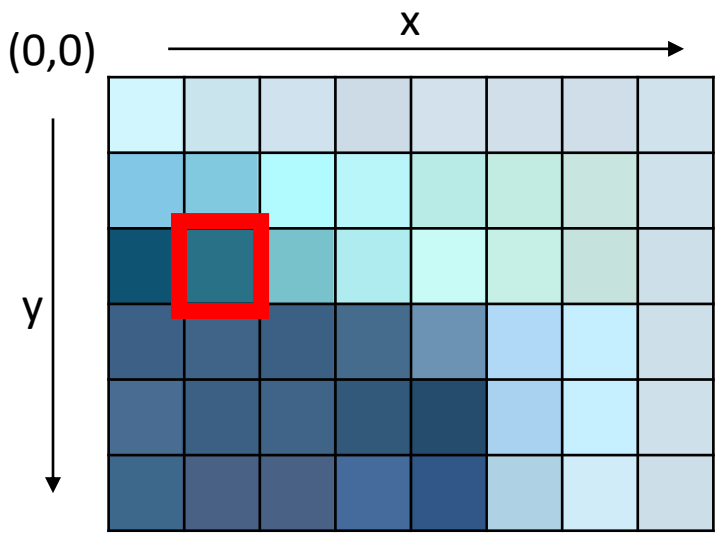
What is a (digital) image?

- A tensor (3D dimensional array of values)
  - Width x height x channel
  - 3 channels = RGB colour image (red, green, blue)
  - 1 channel = grayscale image

- Note: the exact format can vary across libraries / languages!
  - E.g., “channel-first” = channel x height x width
  - Height x width x channel
  - BRG = blue, red, green

What is a (digital) image?

- Pixel = smallest unit of an image
  - Grayscale image: pixel is a grayscale value
  - Colour image: pixel is a 1x3 vector



Pixel location:

x =

y =

Values = [41, 113, 135]

What is a (digital) image?

- Most common data type is uint8 (unsigned 8-bit integers)
  - Range = 0 – 255
  - “24-bit colour” = 3 uint8 channels
- But you may encounter other data types:
  - double (range 0.0-1.0)
  - uint16, uint32: medical images
  - float32: high dynamic range (HDR) images

Common file formats

- Lossy compression:
  - JPEG (.jpg, .jpeg)
- Lossless compression:
  - PNG (.png), BMP (.bmp), GIF (.gif), TIF (.tif, .tiff)


Original: 760 KB


Compressed: 144 KB

JPEG compression

Digital images

- Stored as a tensor (3D array) of values
- Colour is represented through multiple colour channels (typically red, green, blue = RGB)
- Values are typically uint8 (0-255)
- Some image formats discard information to save space (lossy compression)

Image manipulation

Image scaling

Original: 768 x 512 pixels

Crop to 512 x 512

Resize to 512 x 512 (does not preserve aspect ratio)

Resize to 512 x 341 (preserves aspect ratio)

Image scaling

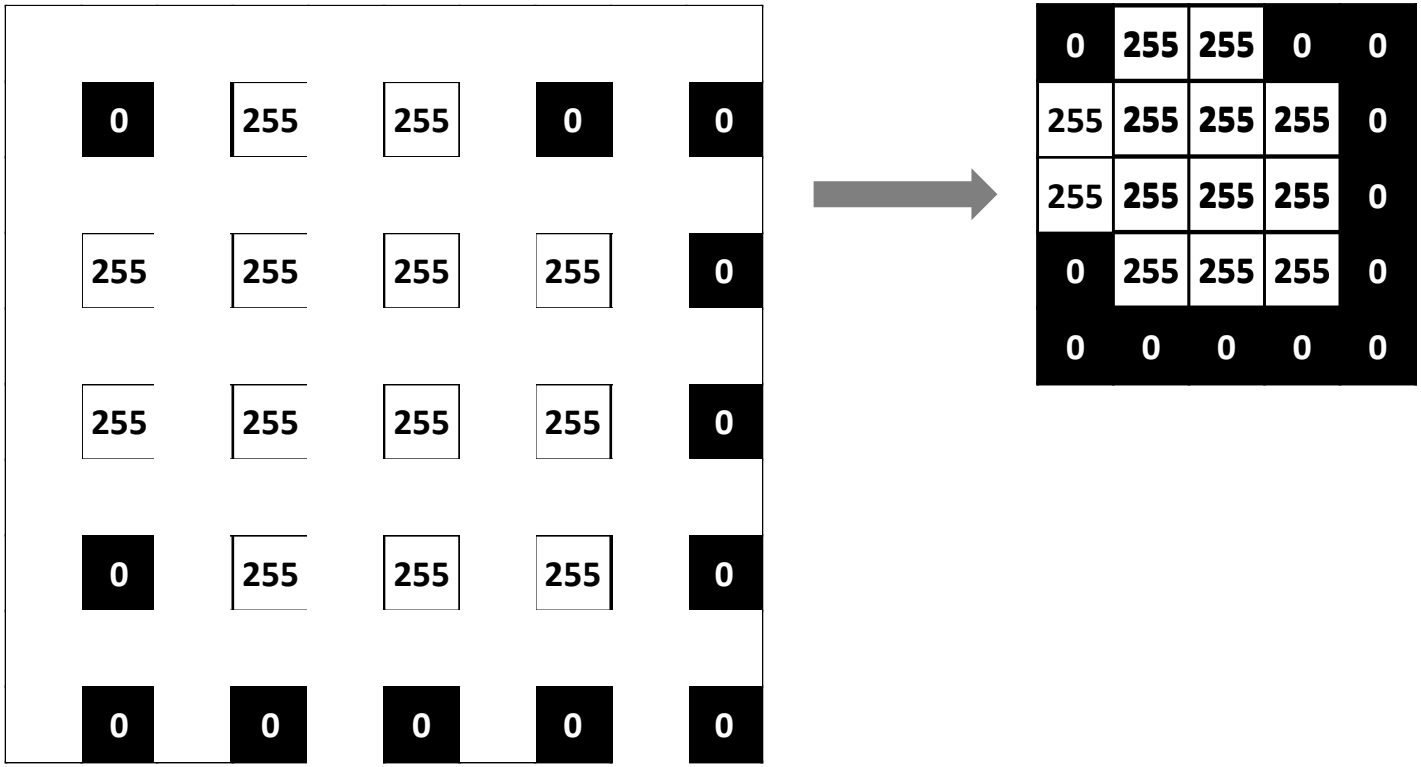
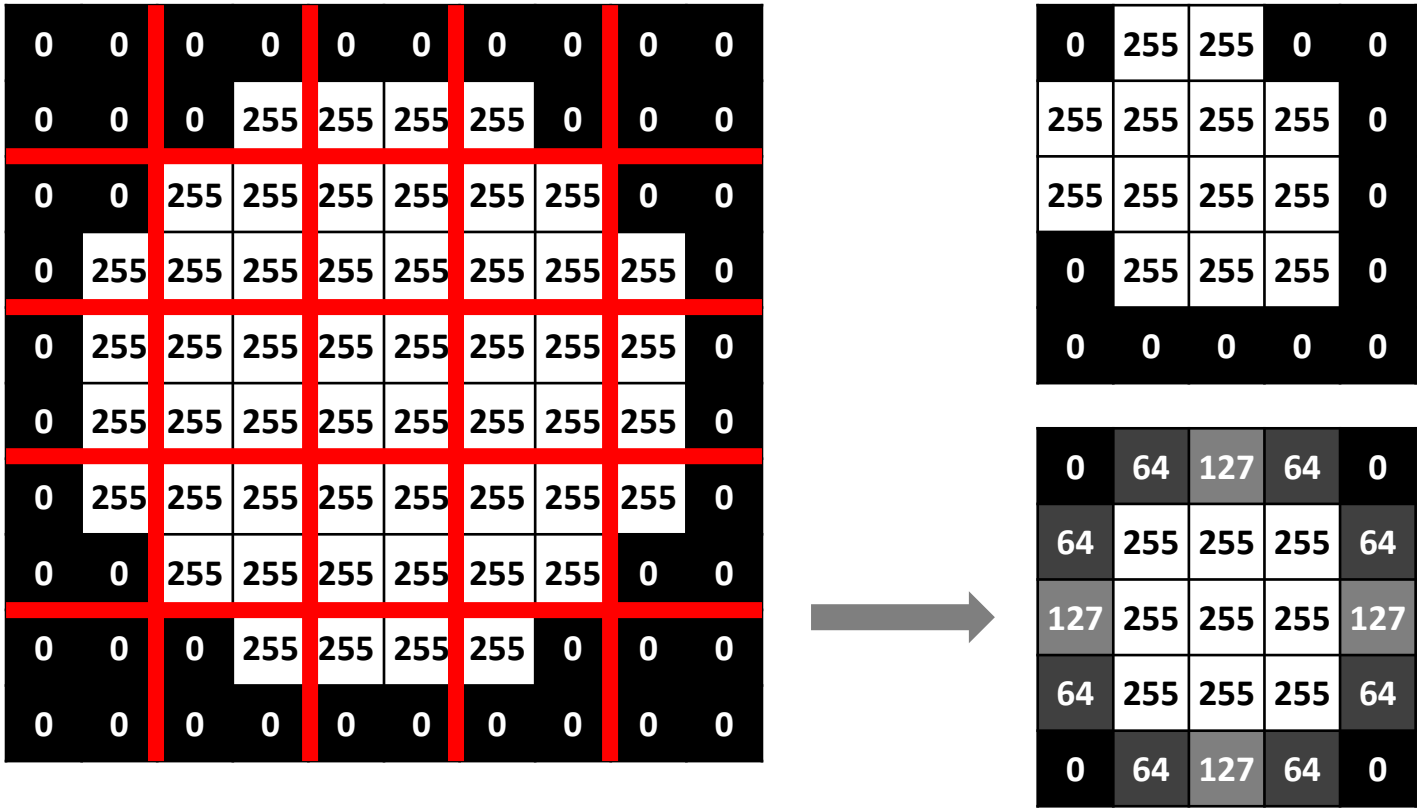


Image scaling



Resampling methods

- Nearest-neighbour: closest value to sample point
  - Simple, preserves hard edges
  - Smooth curves may be blocky/distorted
- Bilinear: weighted average of 4 pixels around sample point
  - Smoother curves, but blurs hard edges
  - Slower to compute
- Other options: bicubic, Lanczos

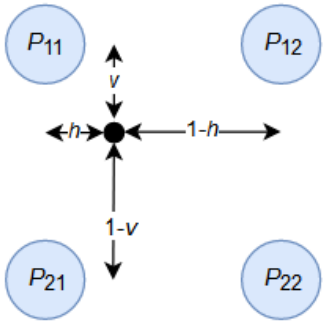
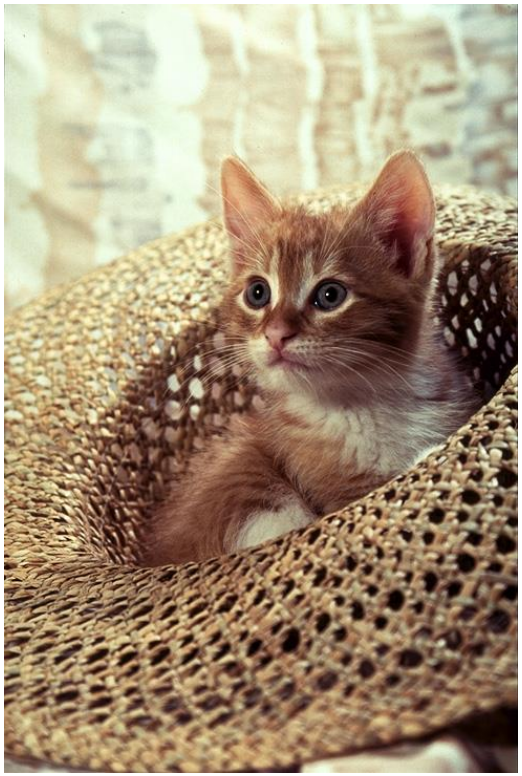


Image: Mathworks

Image resampling



Original

1/10 size
Nearest neighbour

1/10 size
Bilinear

Image resampling

Nearest neighbour

Bilinear

Bicubic

Image manipulations

- Crop = extract a subset of the image array (doesn't require resampling)
- Resize = change the dimensions of the image array (requires resampling)
- Different resampling methods give different results

Summary

- An image is a pattern of light from the world, projected onto a 2D surface
- A digital image is a sample of this pattern, represented as a tensor
- Images of the same scene can vary widely at the pixel level, due to:
  - Camera parameters (focus, field of view)
  - Digital processing steps (compression, resampling)