Introduction to Computer Vision

CS 4243 Computer Vision & Pattern Recognition

Angela Yao

Today's Lecture

1. What is Computer Vision?
2. Why Is Computer Vision Difficult?
3. Digital Image Capture & Representations

What is Computer Vision?

Every Image Tells A Story

Goal of computer vision: perceive the "story" behind the picture.

Compute properties of the world

- 3D shape
- Names of people or objects
- What happened?
- When, how, why?

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

Slide Credit: Noah Snavely

Can Computers Match Human Perception?

Yes and no (mainly no)

- computers can be better at "easy" things
- humans are much better at "hard" things

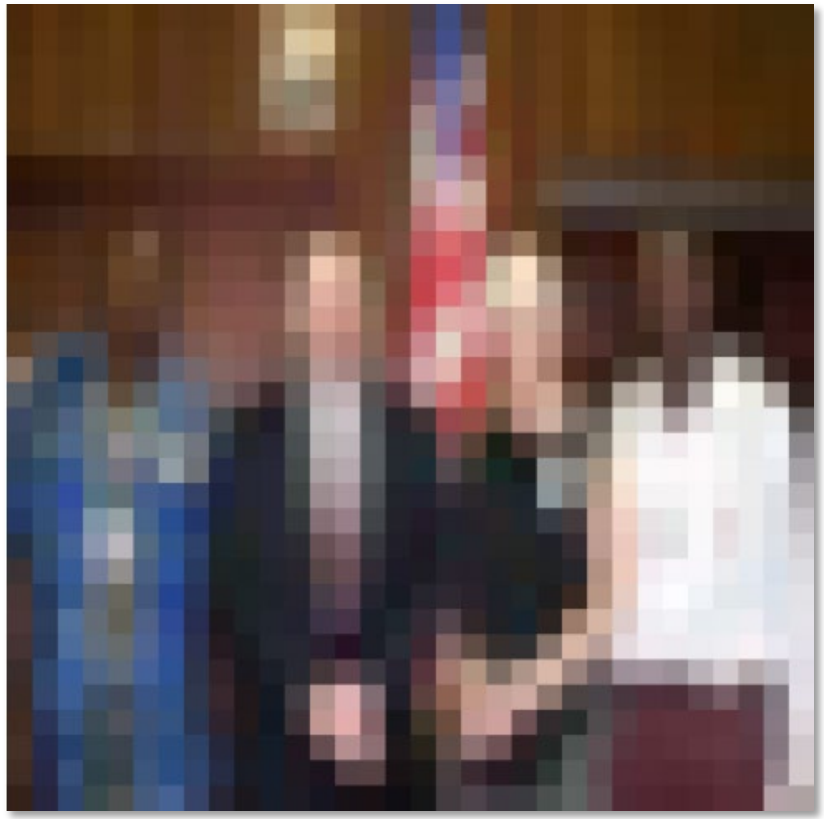
But huge progress has been made

- Accelerating in past 5 years due to deep learning
- What is considered "hard" keeps changing

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

Slide Credit: Noah Snavely

Humans can tell a lot from very little …



Source: "80 million tiny images" by Torralba, et al.

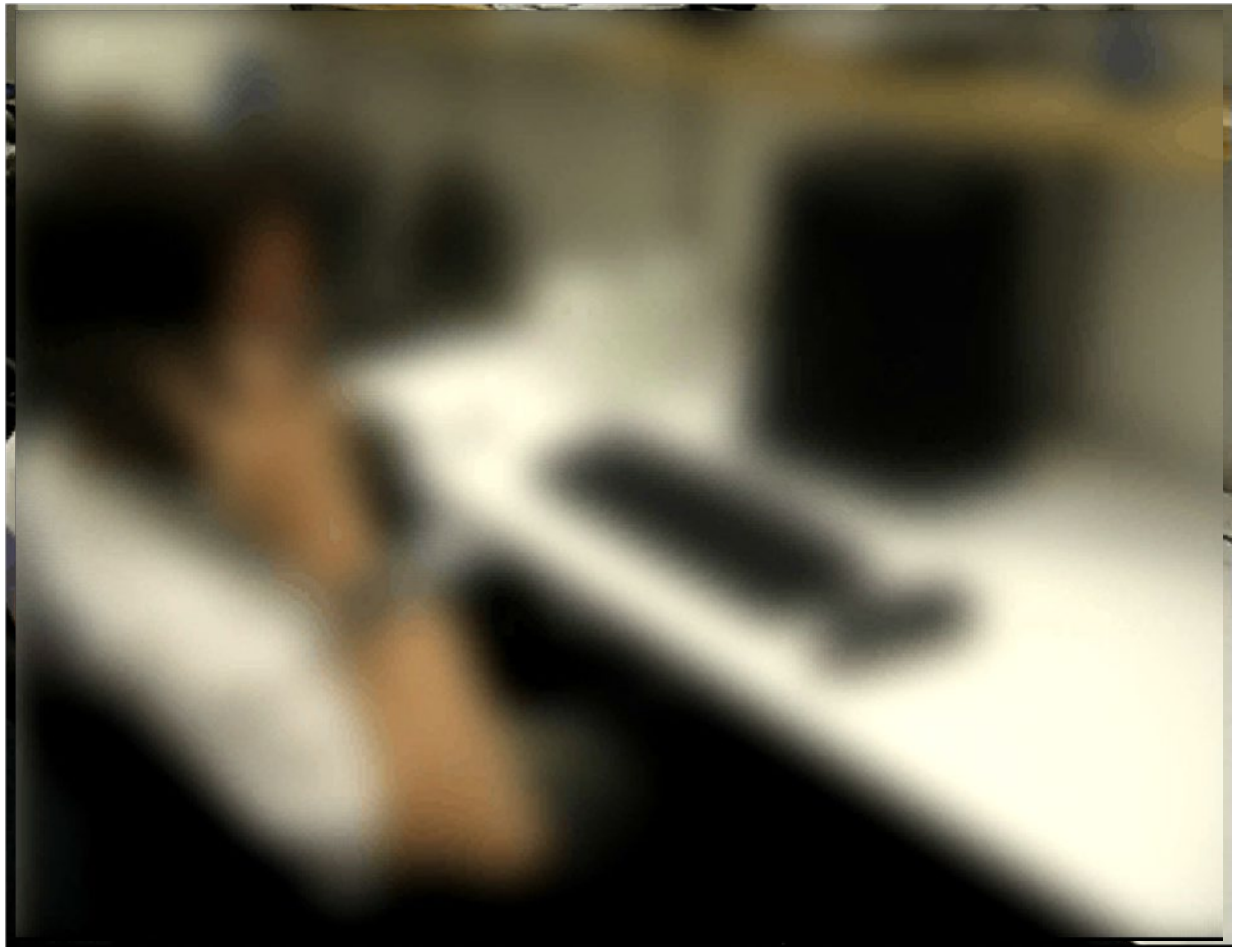
What do you see in this image?

~900 pixels (32x32) is all we need to see 4 people standing in front of some cabinets and a flag(?).

Two are white and two are black.
Two are men and two are women?

Adapted from Noah Snavely

But we rely heavily on prior knowledge …



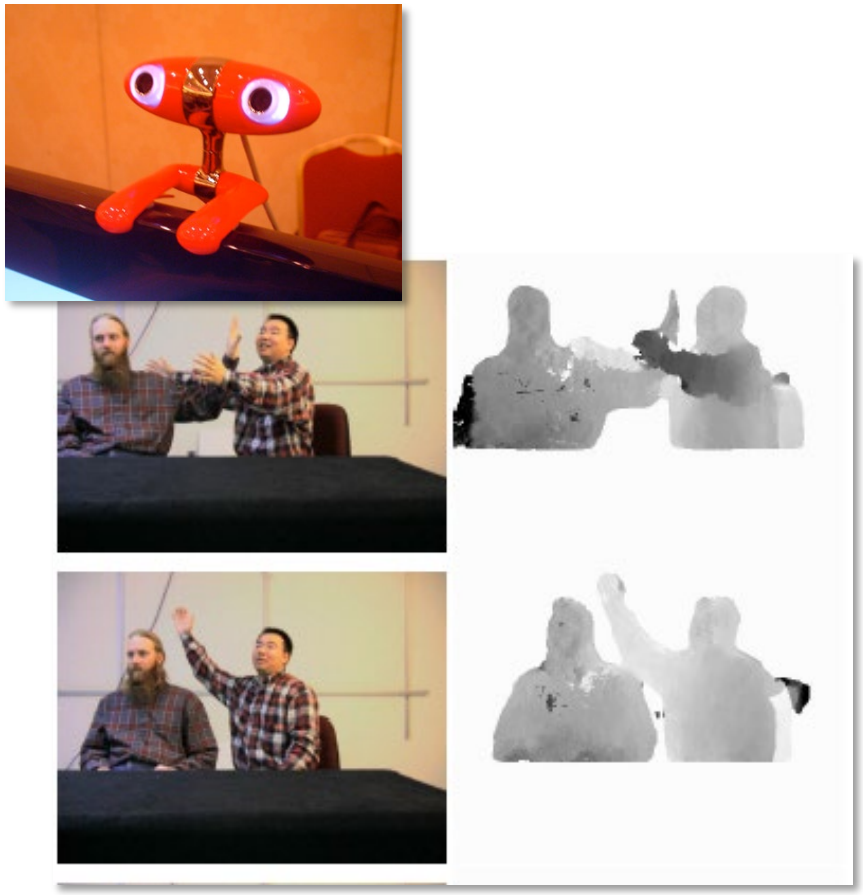
Slide Credit: Noah Snavely

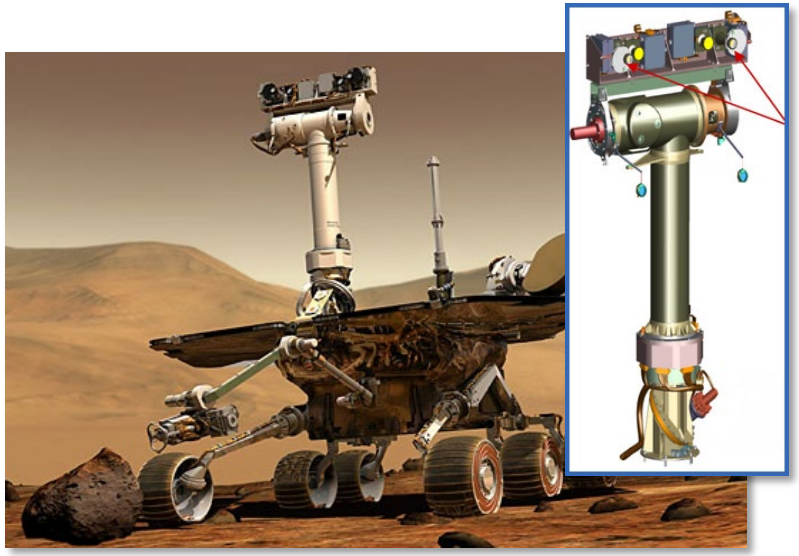
What is the goal of computer vision?

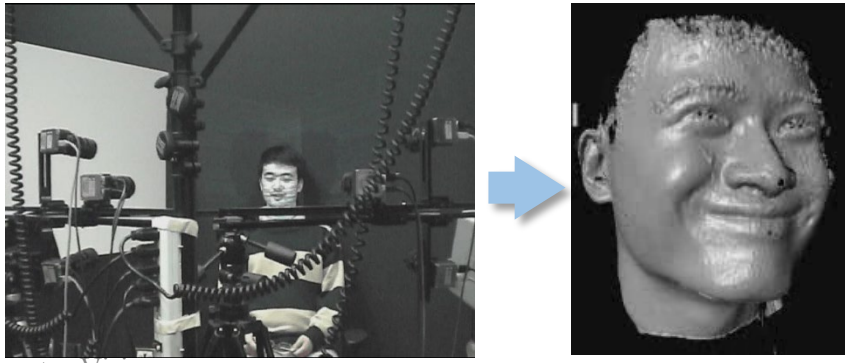
https://www.hudsandguis.com/home/2011/01/02/terminator-the-evolution-of-machine-vision

Slide Credit: Noah Snavely

Compute the 3D shape of the world



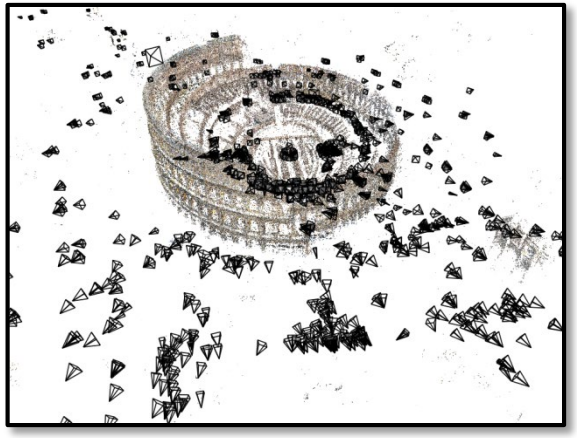




Slide Credit: Noah Snavely

Compute the 3D shape of the world

Internet Photos
("Colosseum")



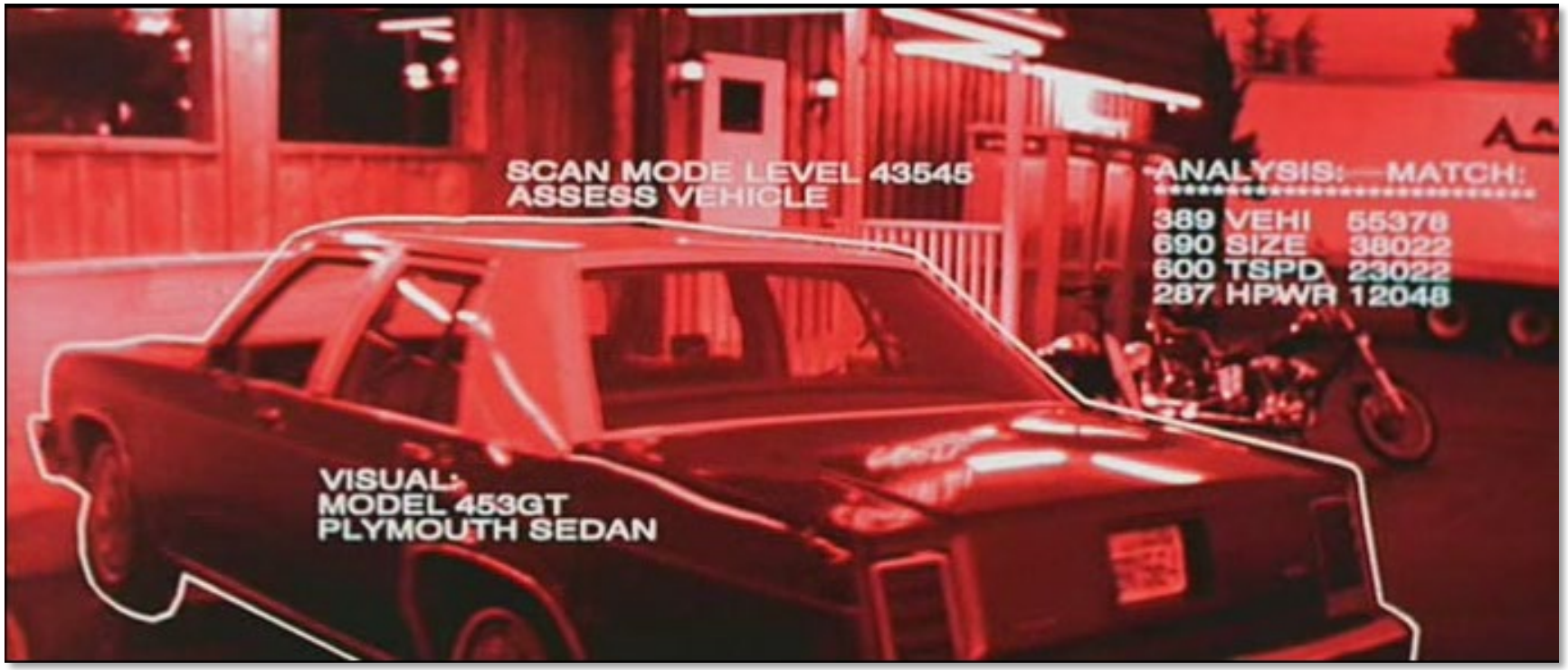
Reconstructed 3D
cameras and points

Dense 3D
model

https://demuc.de/colmap/

https://3dscanexpert.com/review-foldio360-3d-scanning-photogrammetry/

Recognize Objects & People



Terminator 2 (1991)

Slide Credit: Noah Snavely

Slide Credit: Kristin Grauman
















Check out the online web demo!

https://cs.stanford.edu/people/karpathy/deepimagesent/generationdemo/

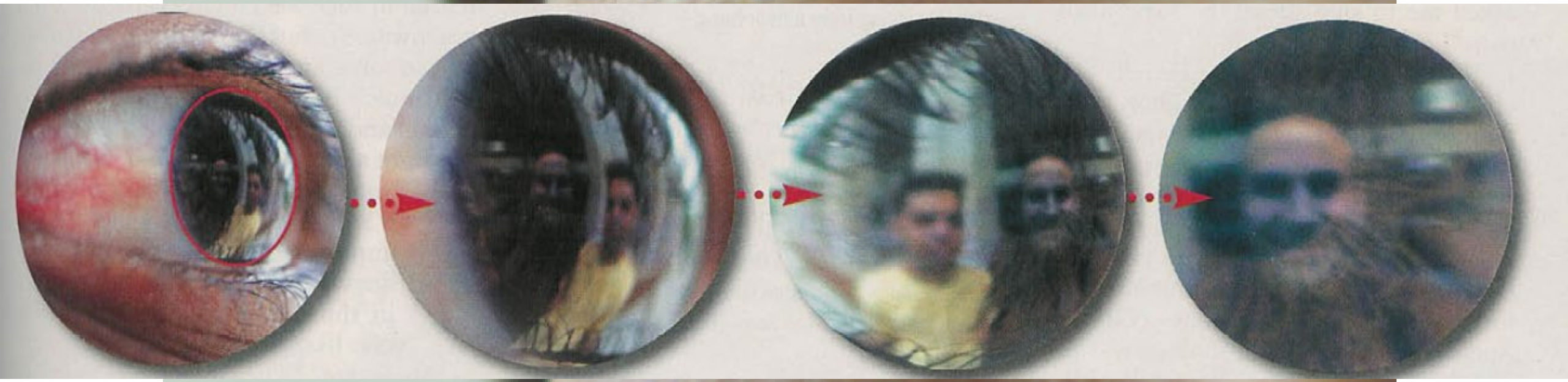
https://cs.stanford.edu/people/karpathy/deepimagesent/

“Enhance” Images

I cannot get the video clip to export correctly for the lecture video. Also the quality is awful... Just watch it on YouTube [here].

Slide Credit: Noah Snavely

“Enhance” Images



Source: Nayar and Nishino, “Eyes for Relighting” SIGGRAPH 2004. http://www.cs.columbia.edu/CAVE/projects/eyes_relight/

Researchers warn peace sign photos could expose fingerprints

But the likelihood of anyone actually using images to recreate prints is pretty slim.

Jamie Rigg, @jmerigg
01.13.17 in Security

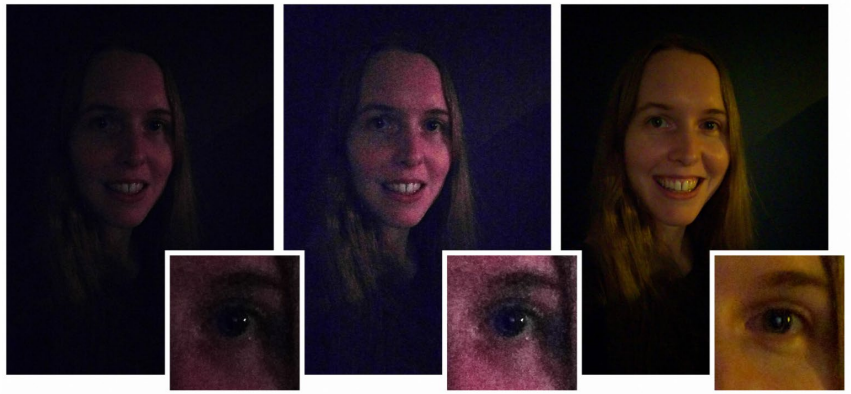
Comments

1721
Shares

Slide Credit: Noah Snavely

Improve photos ("Computational Photography")

Super-resolution (source: 2d3)



Low-light photography
(credit: Hasinoff et al., SIGGRAPH ASIA 2016)

Depth of field on cell phone camera
(source: Google Research Blog)

Inpainting / image completion
(image credit: Hays and Efros)

Slide Credit: Noah Snavely

-3 ev -2 ev -1 ev 0 ev +1 ev +2 ev

Source images straight out of the camera (Camera: Nikon D7000; Lens: Nikon 10-24mm; f/11; 1/400s - 1/15s)

Tone-mapped HDR

http://farbspiel-photo.com/wp-content/gallery/bna-images/hdr-before-and-after-izmir-harbor.jpg

Final image after post-processing






Organization & Search

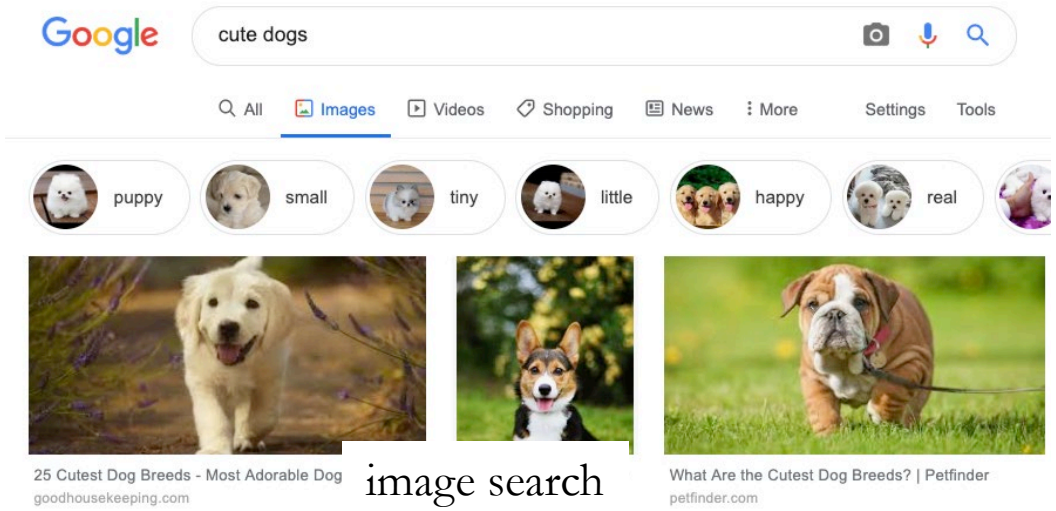


image search

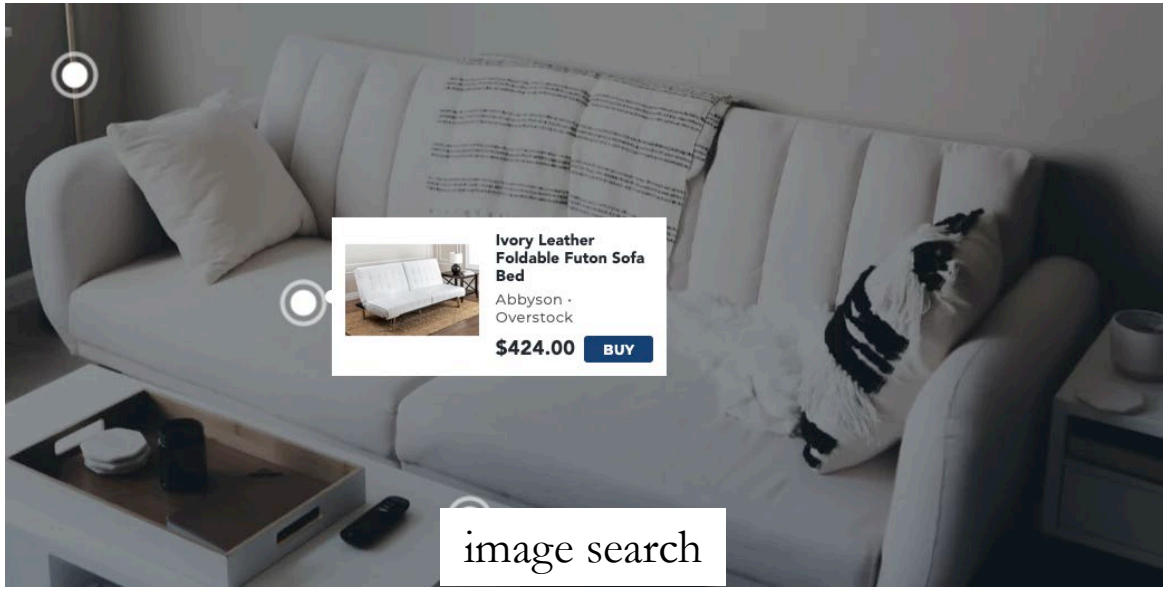
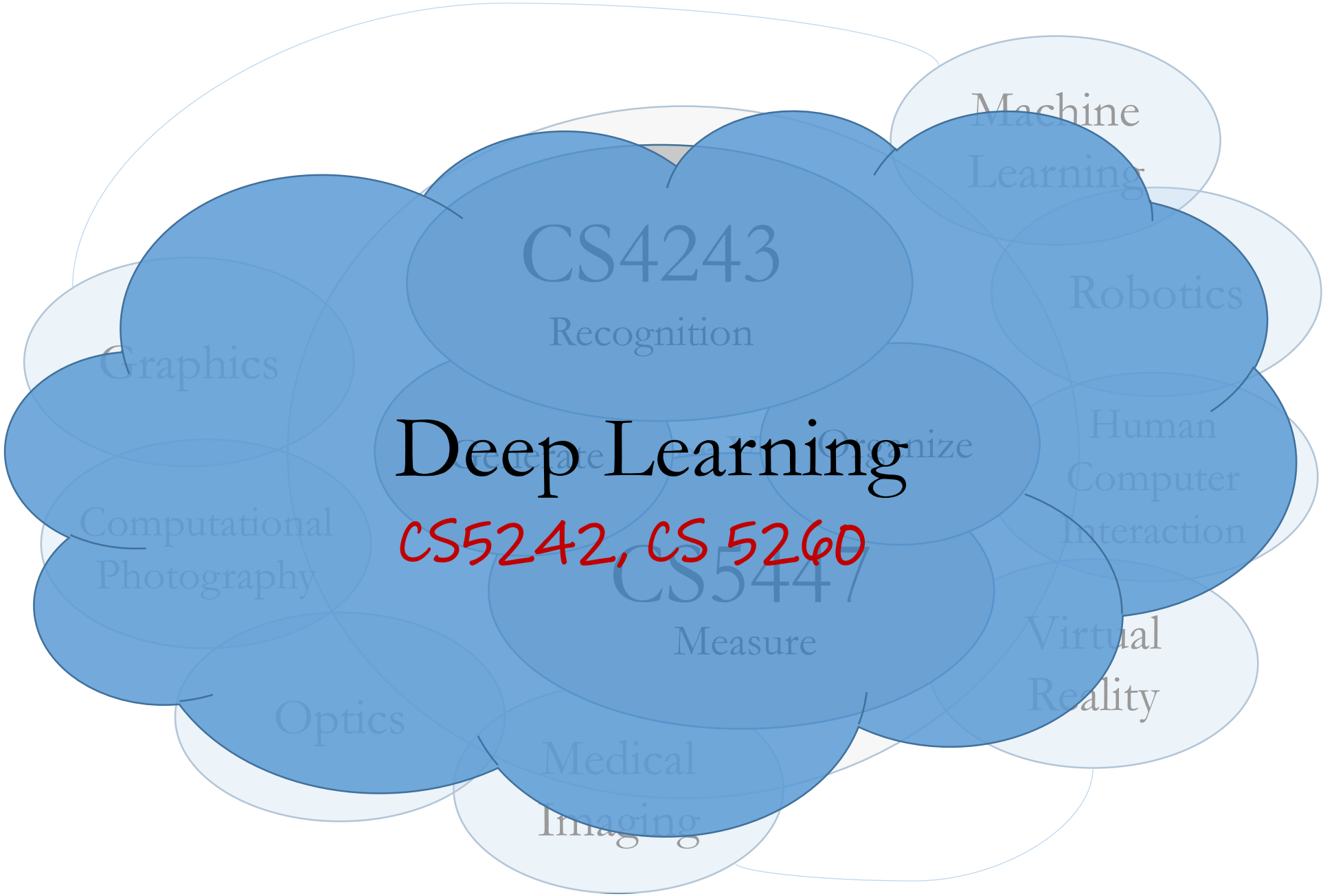


image search

Goals of Computer Vision

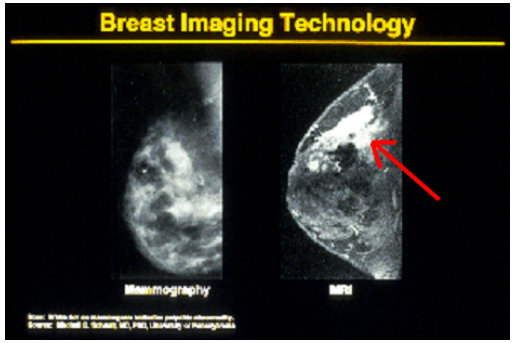
1. **Measure:** properties of the 3D world from visual data
2. **Recognize:** algorithms and representations to allow a machine to recognize objects, people, scenes, activities.
3. **Generate**: enhance, complete and manipulate images and videos
4. **Organize:** algorithms to mine, search, and interact with visual data

Adapted from K. Grauman


Machine Learning
CS4243
Recognition
Graphics
Robotics
Deep Learning
CS5242, CS 5260
Human Computer Interaction
Computational Photography
CS5447
Measure
Virtual Reality
Optics
Medical Imaging

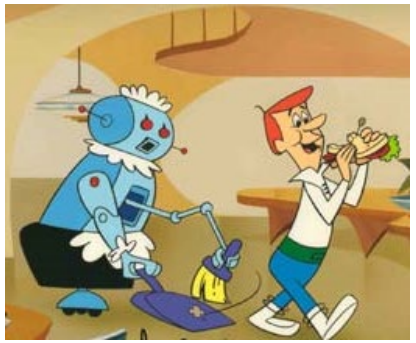
Why Does Computer Vision Matter?

Safety



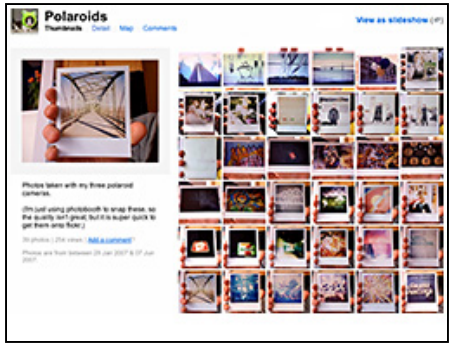
Health

Security



Comfort

Entertainment

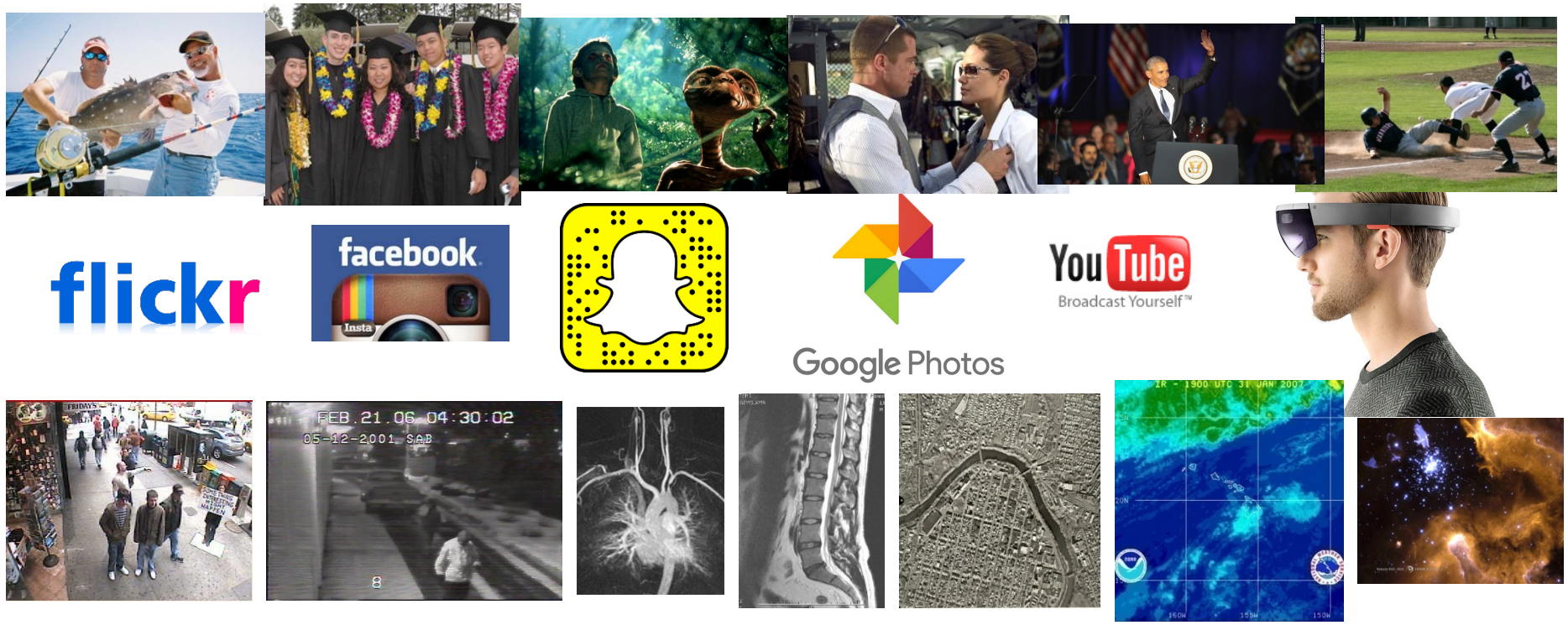


Access

Slide Credit: James Hays

Why Study Computer Vision?

Billions of images/videos captured per day



Huge number of useful applications…

Why Is Computer Vision Hard?

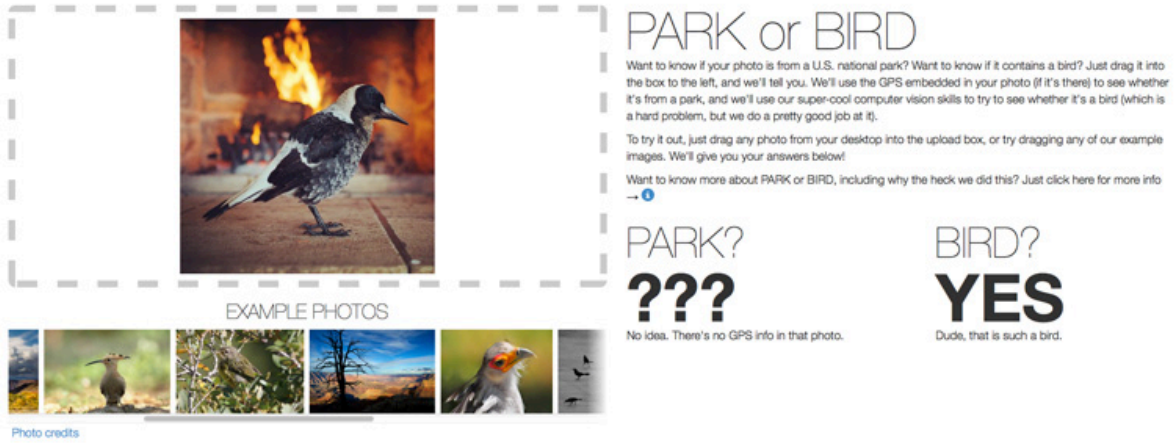
(After all, it comes so naturally to humans.)

Computer Vision is Easy?

Park or Bird? A National Park and Bird Identifying App Inspired by an xkcd Comic

by E.D.W. Lynch at 5:21 PM on October 21, 2014

Facebook Twitter Pinterest Tumblr Flipboard More



screenshot via Flickr, photo via penguinthemagpie

Park or Bird? is a web app by the Flickr Vision and Search team that can determine whether a photo depicts a bird or a National Park. The app stems from an xkcd comic about the often confusing difference between a simple computer task (determining whether a photo was shot in a National Park) and an extremely challenging task (determining whether a photo contains a bird).

Why is computer vision difficult?

Viewpoint variation

Illumination

Scale

Slide Credit: Noah Snavely

Why is computer vision difficult?

Intra-class variation

Motion (Source: S. Lazebnik)

Background clutter

Occlusion

Slide Credit: Noah Snavely

But there are lots of cues we can exploit…

Source: S. Lazebnik

Bottom line

- Perception is an inherently ambiguous problem
  - Many different 3D scenes could have given rise to a particular 2D picture

  - We often need to use prior knowledge about the structure of the world

Slide Credit: S. Lazebnik

Some historical Notes

1966: Papert assigns computer vision as an undergrad summer project

1960's: interpretation of synthetic worlds

1970's: some progress on interpreting selected images

1980's: ANNs come and go; shift toward geometry and increased mathematical rigor

1990's: face recognition; statistical analysis in vogue

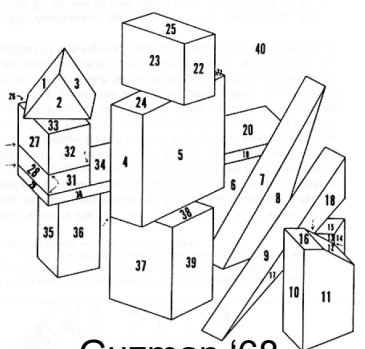
2000's: broader recognition; large annotated datasets available; video processing starts

2010's: Deep learning with CNNs

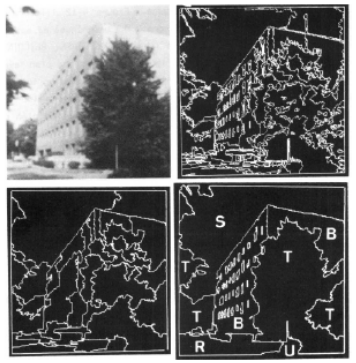
2020's: Widespread autonomous vehicles?

2030's: robot uprising?

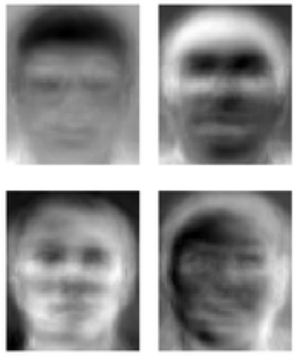
Are we here now?



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

Slide Credit: James Hays

The state of Computer Vision and AI: we are really, really far.

Andrej Karpathy blog

Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

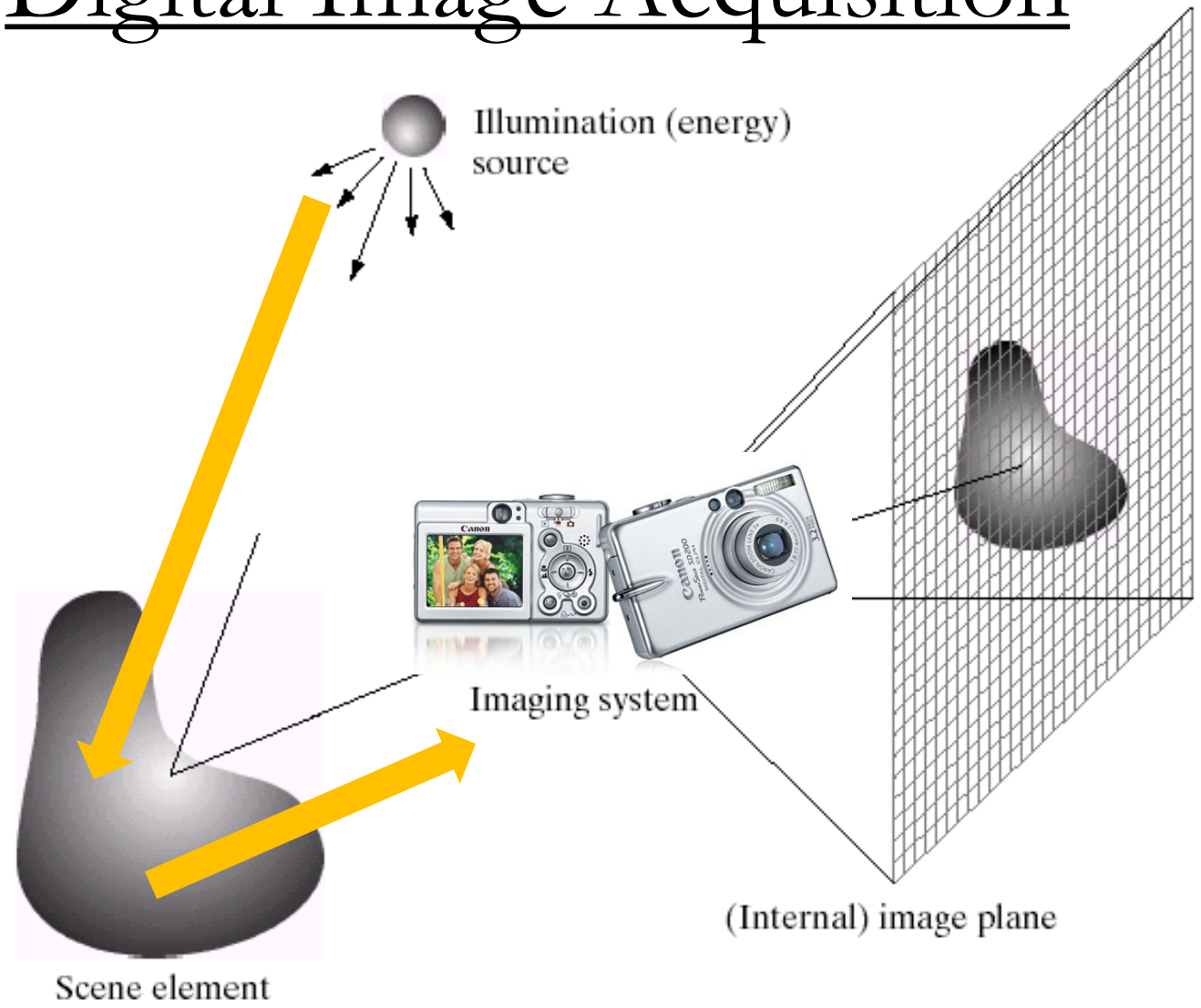
- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

http://karpathy.github.io/2012/10/22/state-of-computer-vision/

Digital Images

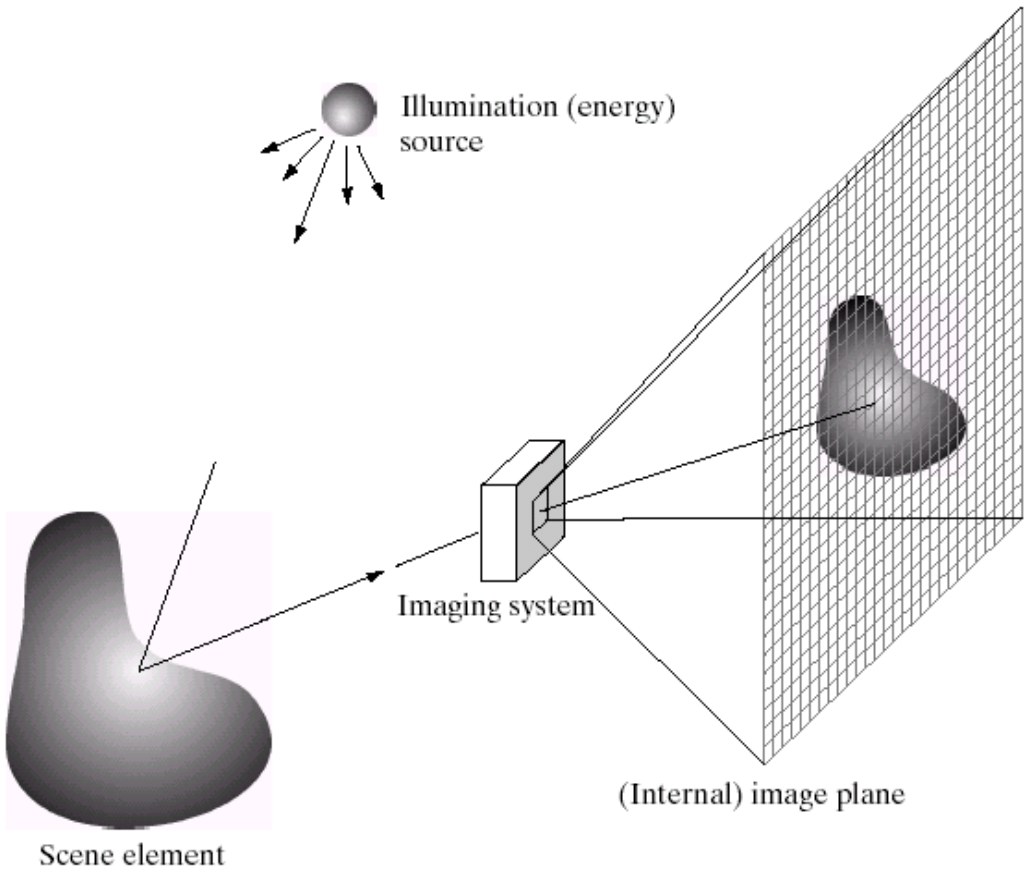
Imaging sensors, exposure time, spatial and intensity resolution

Digital Image Acquisition



1. Energy transfer from source (sun, light bulb) to scene to imaging system
2. Optics in imaging system focuses energy onto sensor (e.g. CCD).
3. Digital sensor measures amount of energy

Image Formation Model

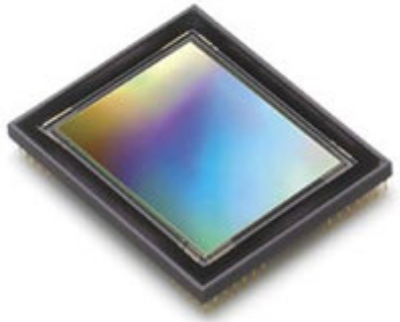


- $f(x, y)$ - 2D image function, where x, y are coordinates on the internal image plane
- Illumination component $i(x, y)$: amount of source illumination incident on the scene
- Reflectance component $r(x, y)$: amount of illumination reflected by objects in the scene

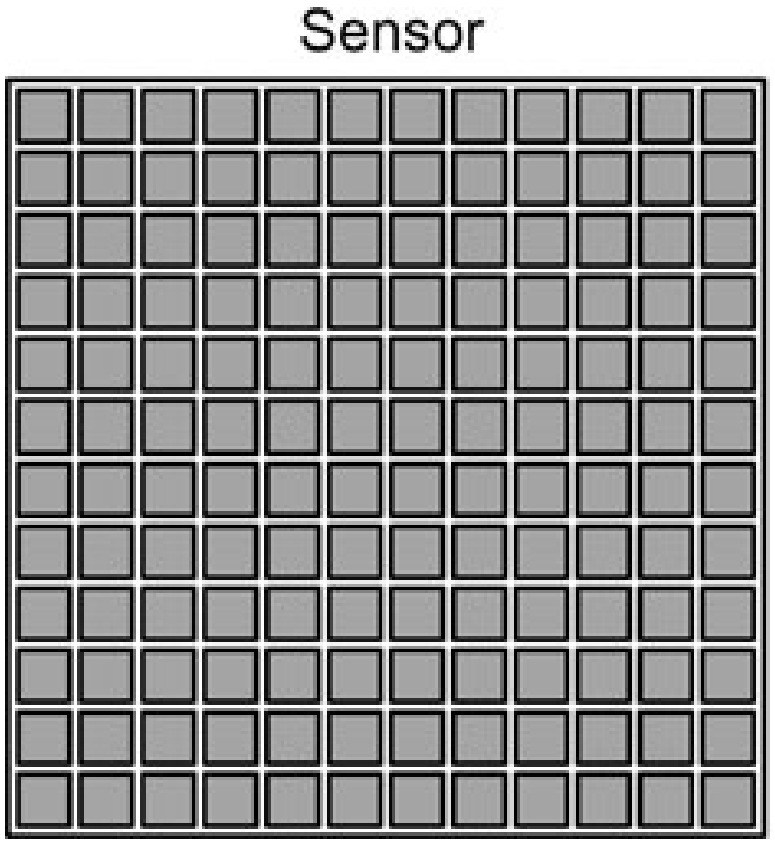
$$f(x, y) = i(x, y)r(x, y)$$

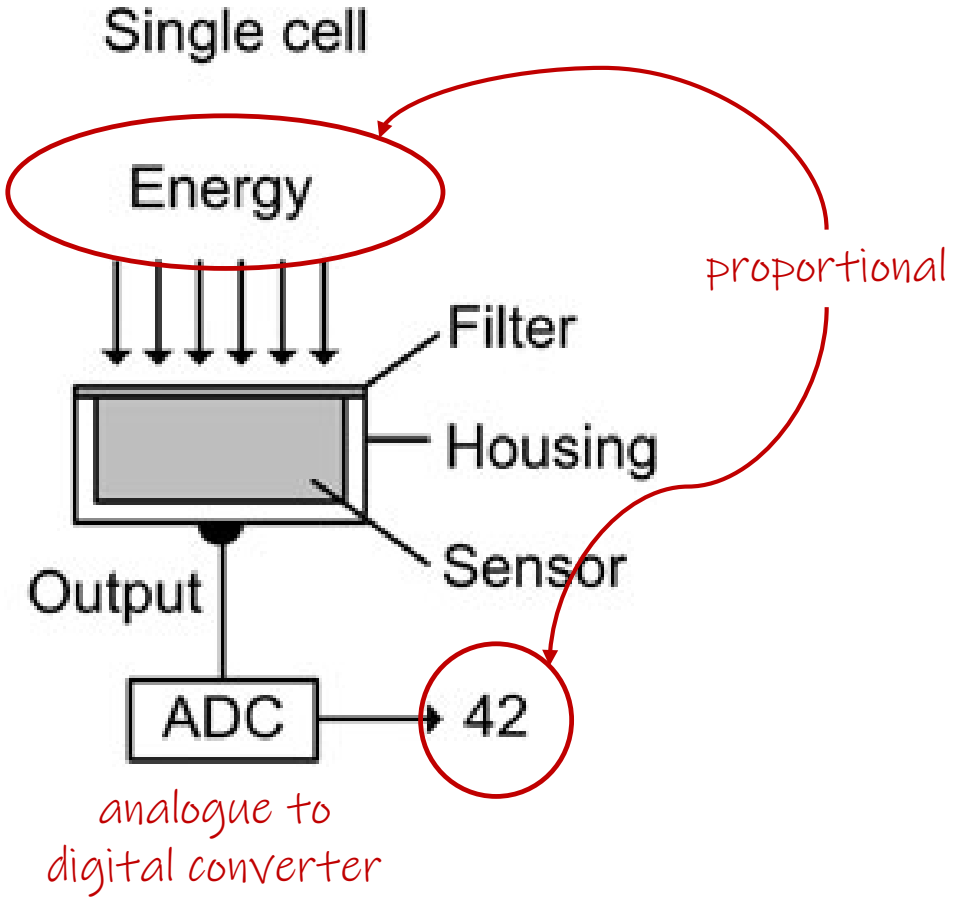
- $0 \leq i(x, y) < \infty$ - no upper bound
- $0 \leq r(x, y) \leq 1$
  0 for total absorption, 1 for total reflectance

Imaging System Sensors



CCD sensor





Introduction to Video & Image Processing, Thomas B. Moeslund, Springer 2012

Exposure Time

Exposure is the amount of time that incident light can reach the imaging sensor.

HWQ: Why does motion blur occur and what adjustments should we make to compensate?

Correctly exposed

Over exposed

exposure time too long
too much light reaches sensor

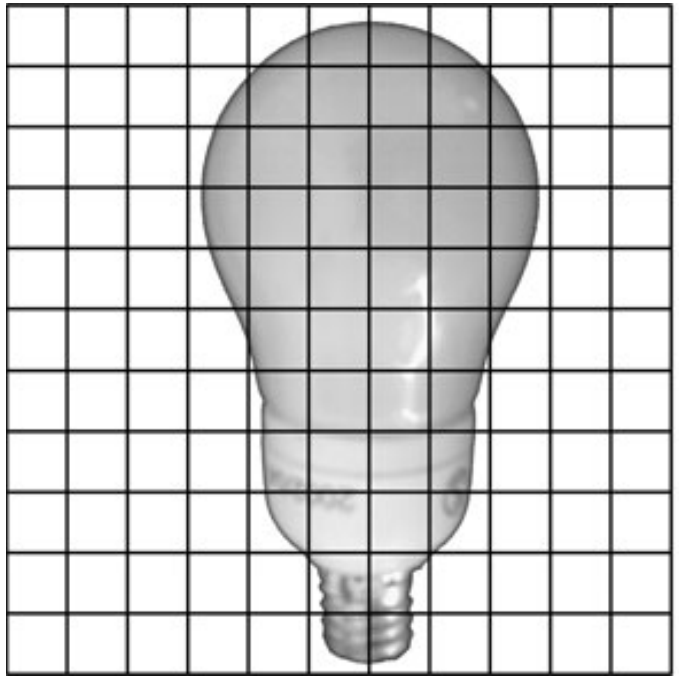
Under exposed

exposure time too short
not enough light reaches sensor

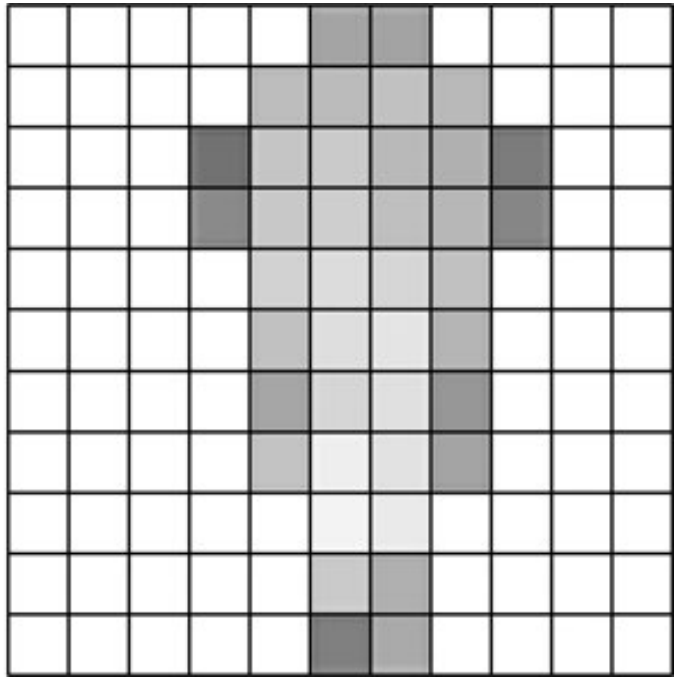
Motion blur

Introduction to Video & Image Processing, Thomas B. Moeslund, Springer 2012

Image Formation



continuous projection
onto imaging sensor



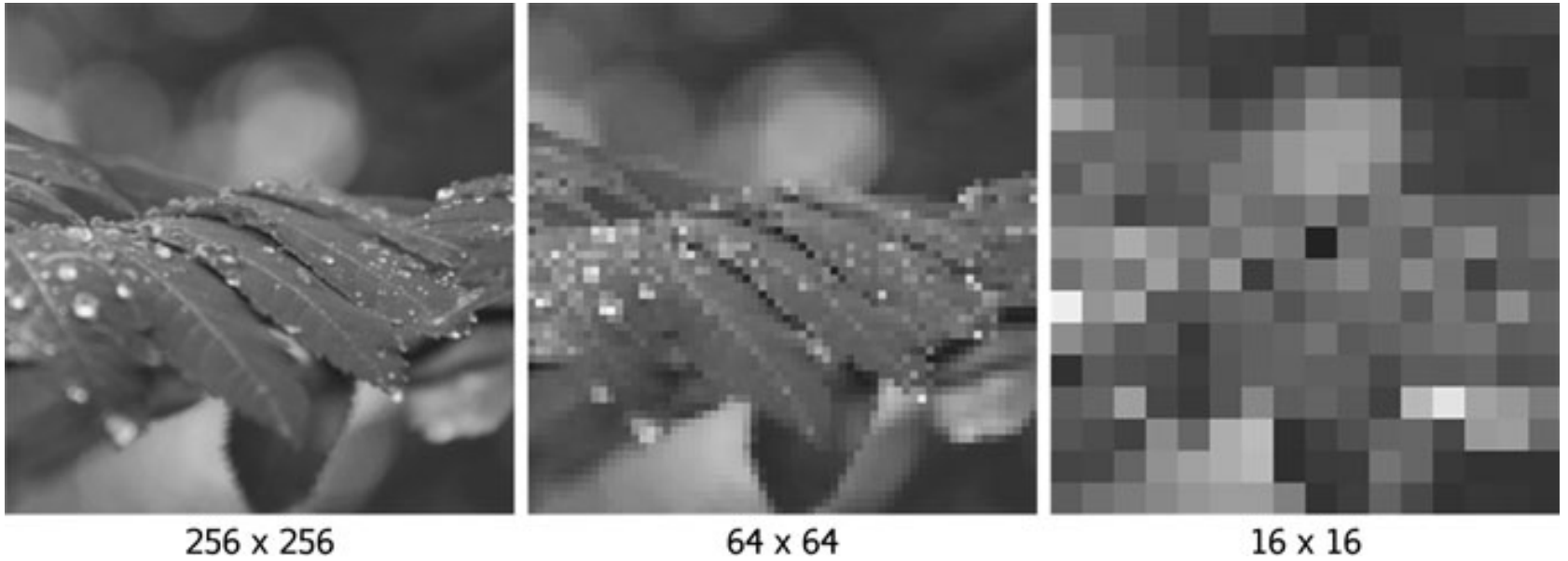
digital sampling

spatial resolution
intensity resolution

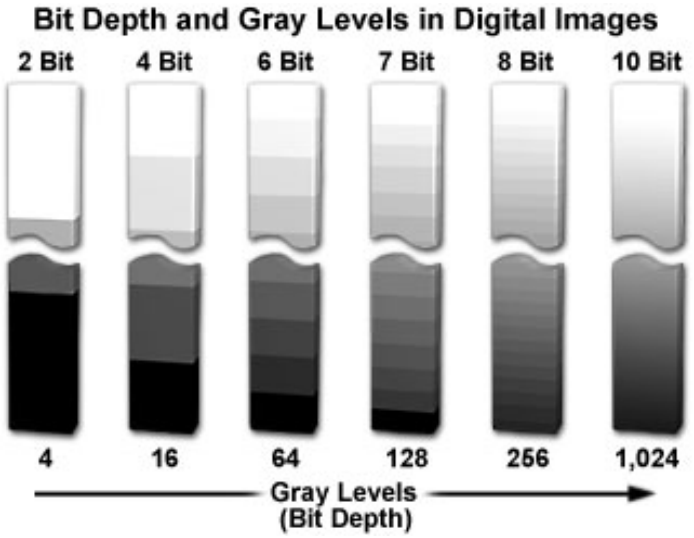
Introduction to Video & Image Processing, Thomas B. Moeslund, Springer 2012

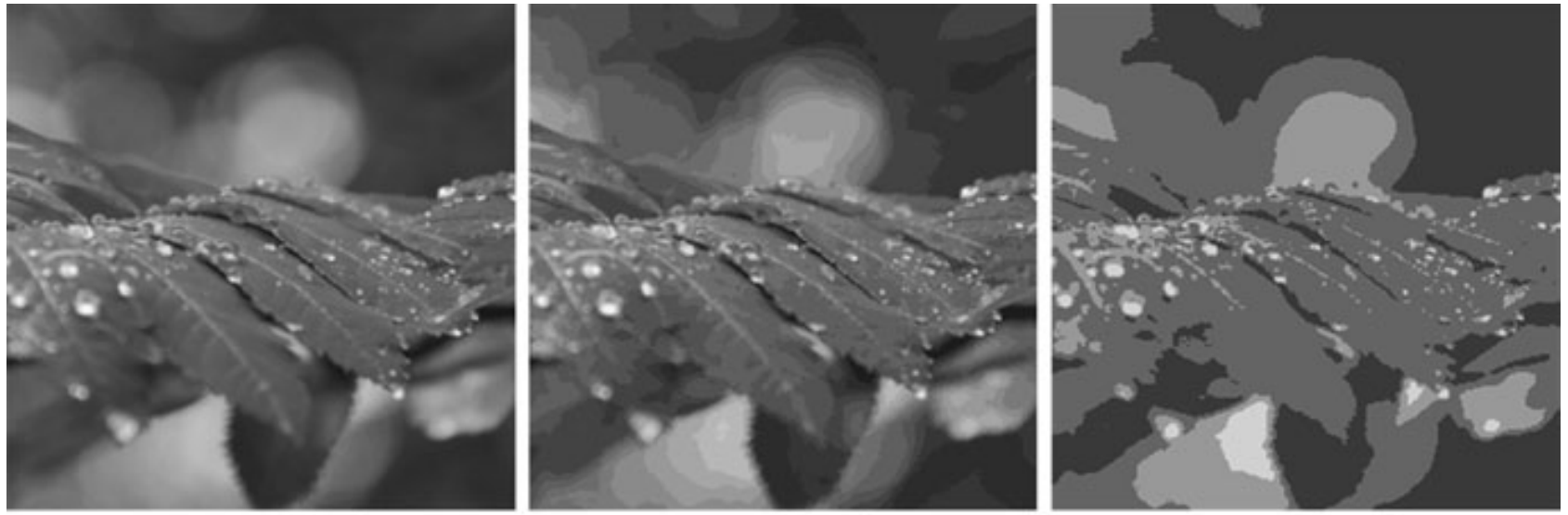
Resolution

Spatial Resolution



Intensity Resolution





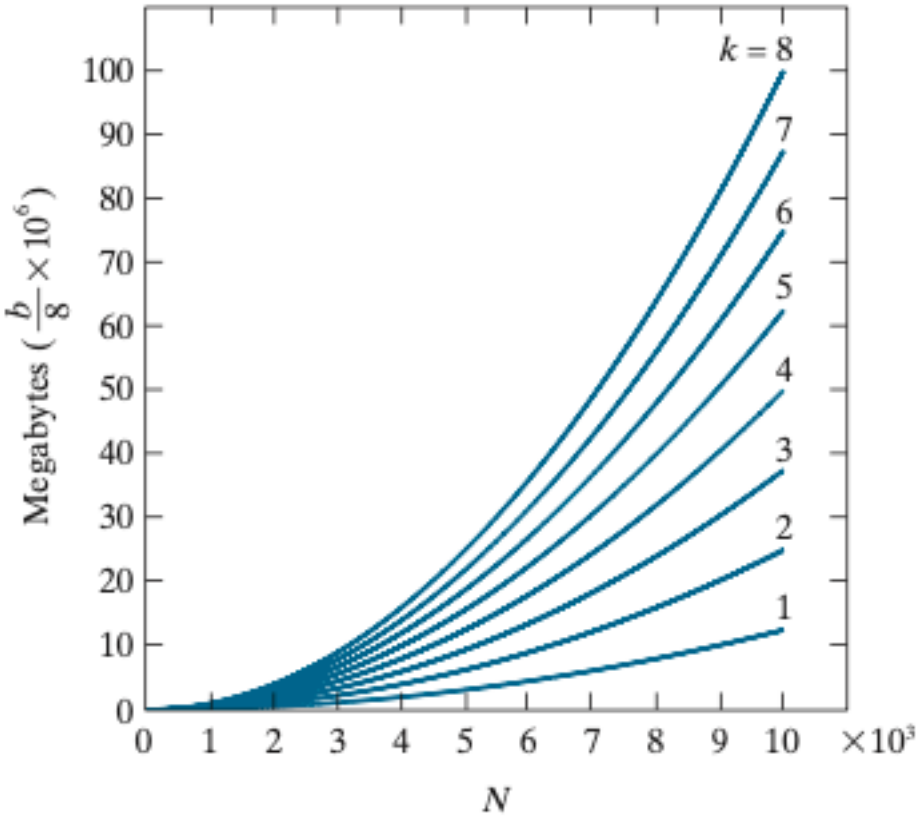
we almost always work with 8 bits, i.e. from 0-255

some visible false contouring

Introduction to Video & Image Processing, Thomas B. Moeslund, Springer 2012

Resolution Determines File Size

Number of megabytes required to store images for various values of N and k.



- Spatial and intensity resolution determines the file size and also transmission time if we were to send images over the internet
- An image of size $M \cdot N$ with a bit-depth of k requires b bits to store, where

$$b = M \cdot N \cdot k$$

- b is the storage requirement for a raw bitmap (.bmp). Compressed formats (.png, .jpg) can reduce file sizes

Colour Representations

RGB images, Bayer patterns, colour to greyscale,
HSV colour space, colour-based applications

Colour Images

[R, G, B] :[120, 30, 24]

Compare

- Sky vs. building in the center
- Top of the fountain vs. clouds
- [0,0,0] is black, [255,255,255] is white

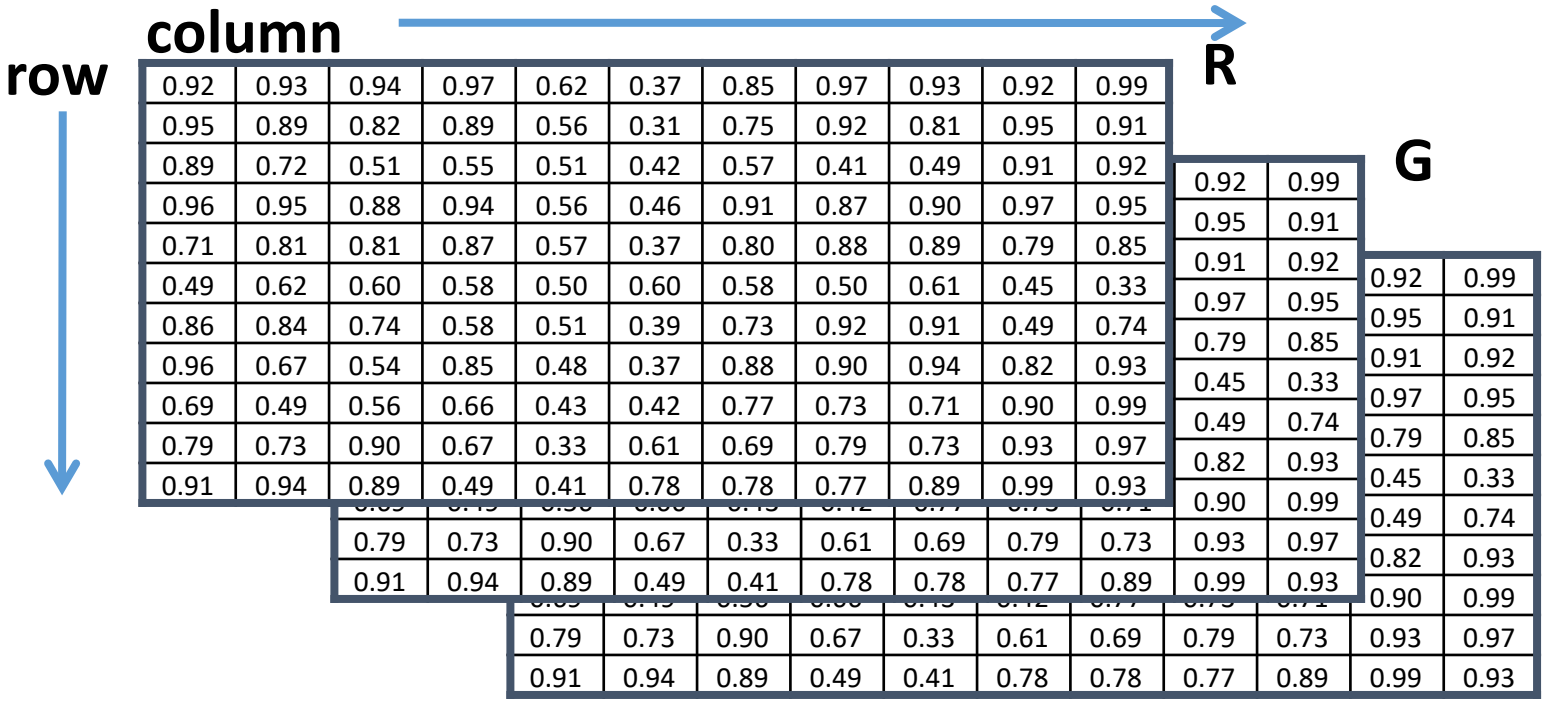
red channel

green channel

blue channel

Image Source: R. Klette

Images in Python



[0,255] → [0,1] for arithmetic convenience

represented as a matrix

N x M RGB image "im"

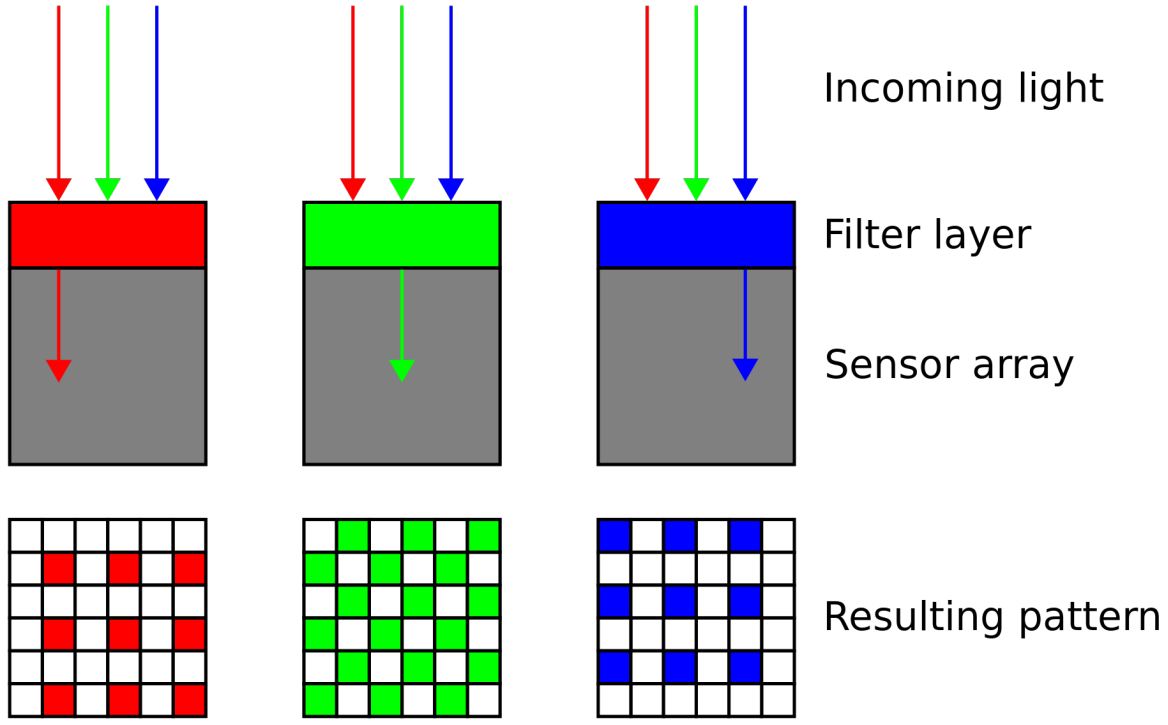
im(0,0,0) = top-left pixel value in R-channel

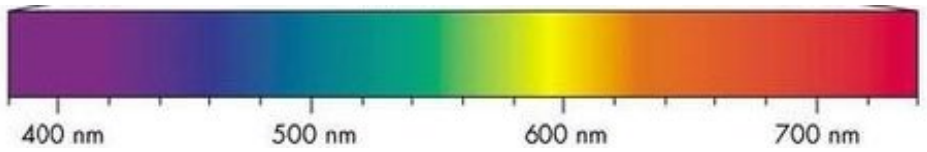
im(y, x, b) = y pixels down, x pixels to right in the bth channel

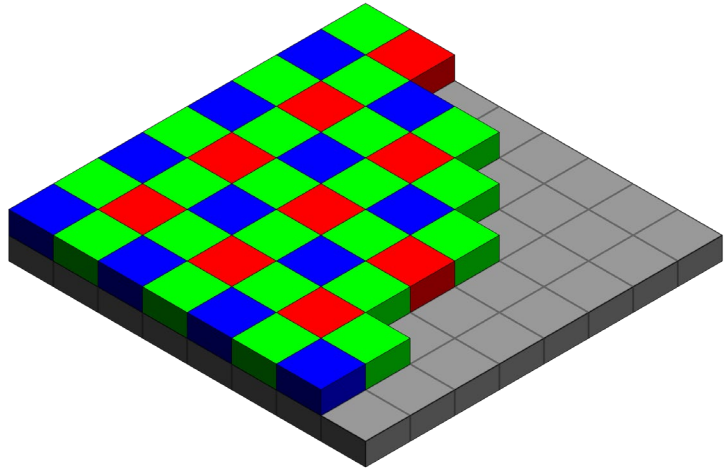
im(N-1, M-1, 2) = bottom-right pixel in B-channel

Slide Credit: James Hays

Colour Capture – Bayer Filter

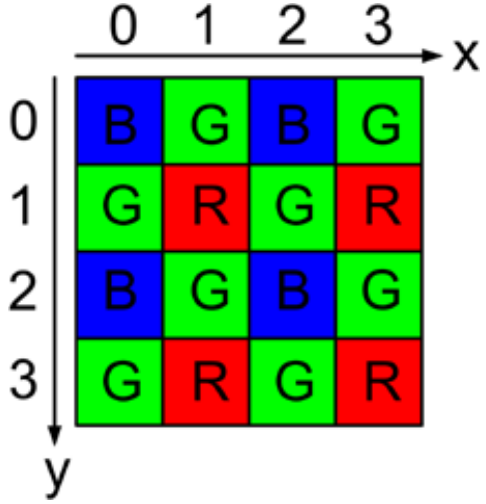


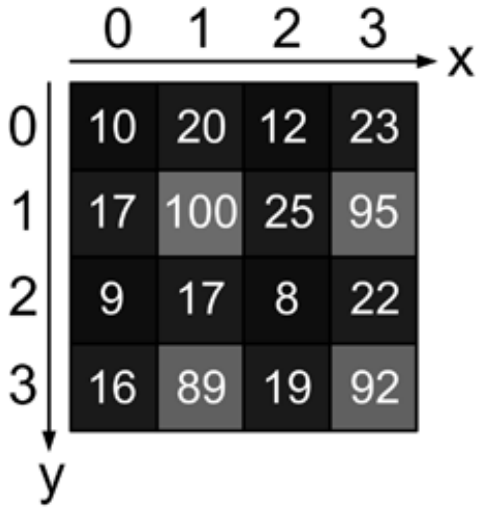


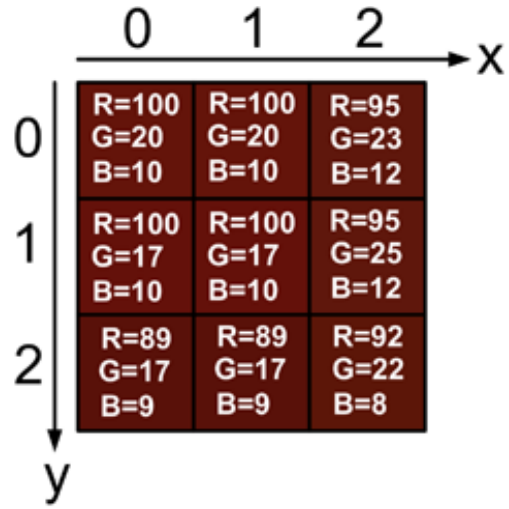


https://en.wikipedia.org/wiki/Bayer_filter

Demosaicing Bayer Patterns







$$g(x,y) \begin{cases} [R,G,B]_B = [f(x+1,y+1), f(x+1,y), f(x,y)] \\ [R,G,B]_{GB} = [f(x,y+1), f(x,y), f(x-1,y)] \\ [R,G,B]_{GR} = [f(x+1,y), f(x,y), f(x,y-1)] \\ [R,G,B]_R = [f(x,y), f(x-1,y), f(x-1,y-1)] \end{cases}$$

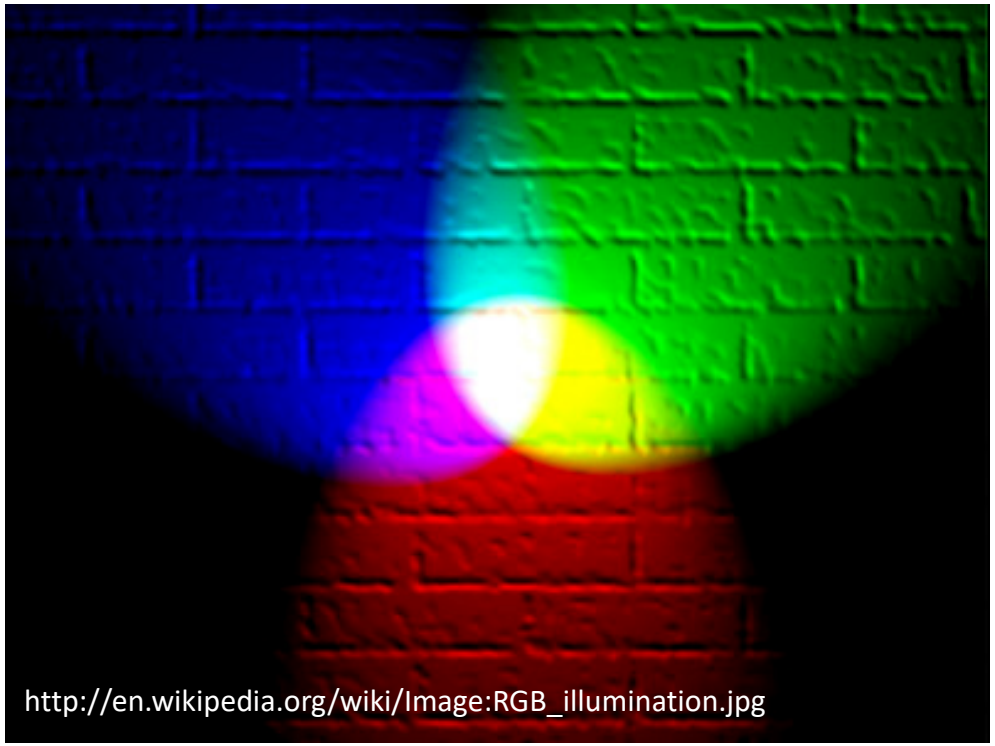
sensitive to blue

sensitive to green followed by blue / red in row

sensitive to red

Introduction to Video & Image Processing, Thomas B. Moeslund, Springer 2012

RGB Colour Model



- "default" colour space
- **additive** model used in electronic displays & photography
- based on **trichromatic** human perception of colours

Colour to Greyscale

greyscale intensity

weight factor

$$I = W_R \cdot \mathrm{R} + W_G \cdot \mathrm{G} + W_B \cdot \mathrm{B}$$

$$W_R + W_G + W_B = 1$$

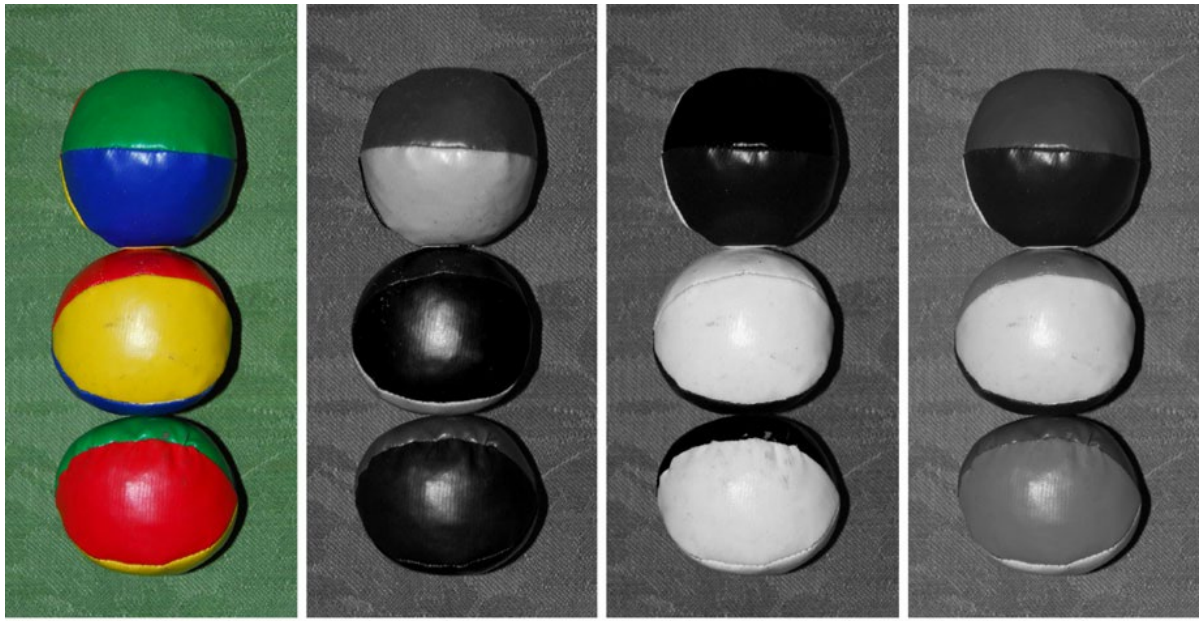
keeps values within [0, 255]

$$W_R = W_G = W_B = \frac{1}{3}$$

default, but can be adjusted depending on application

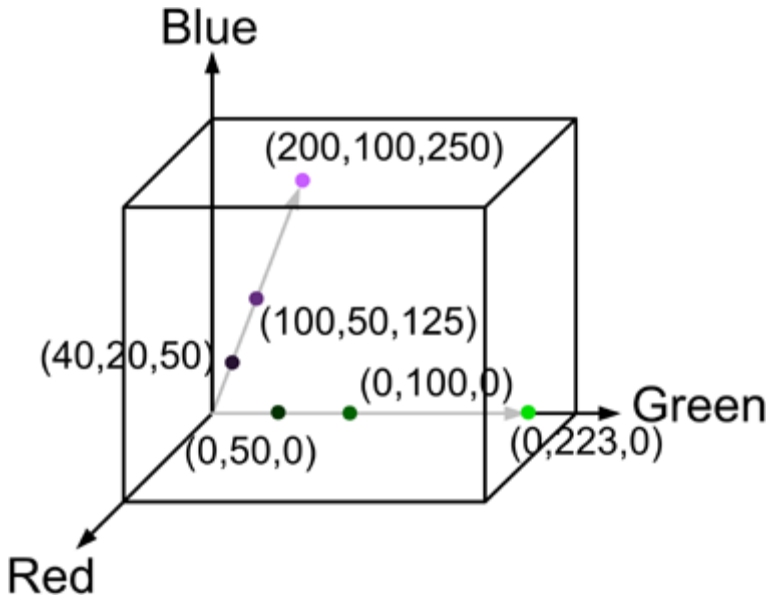
$$W_R = 0.299, \quad W_G = 0.587, \quad W_B = 0.114$$

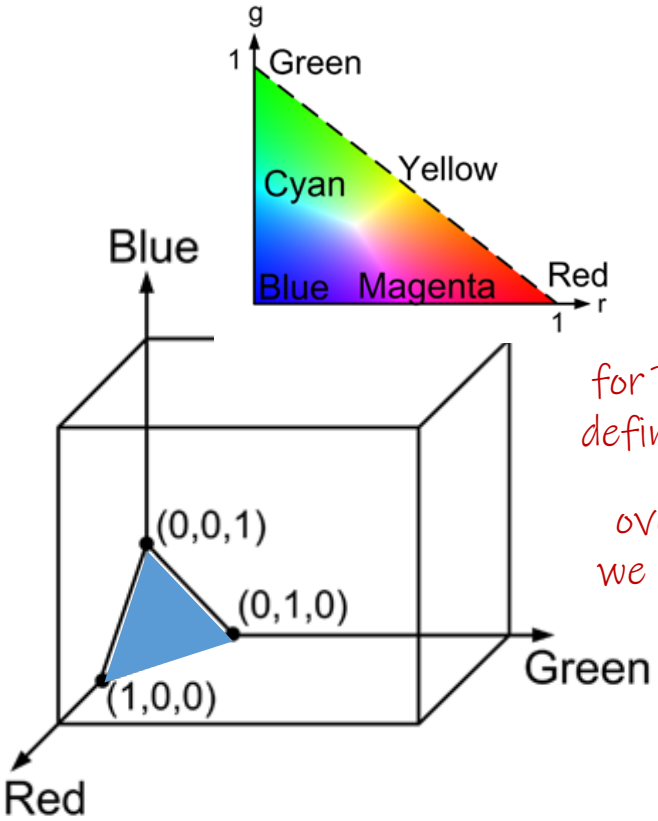
good compromise for visualization purposes



Introduction to Video & Image Processing, Thomas B. Moeslund, Springer 2012

Normalized RGB





for RGB =(0,0,0), we define (r,g,b) = (0,0,0)

over-specified since we know r + g + b = 1

(0, 50, 0), (0,100, 0), (0, 223, 0) all lie on the same vector spanned by (0,255,0) → 3 shades of **green**

same colour, different illumination

$$(r, g, b) = \left(\frac{R}{R+G+B}, \frac{G}{R+G+B}, \frac{B}{R+G+B} \right)$$

true "colour"

Intensity (amount of light)

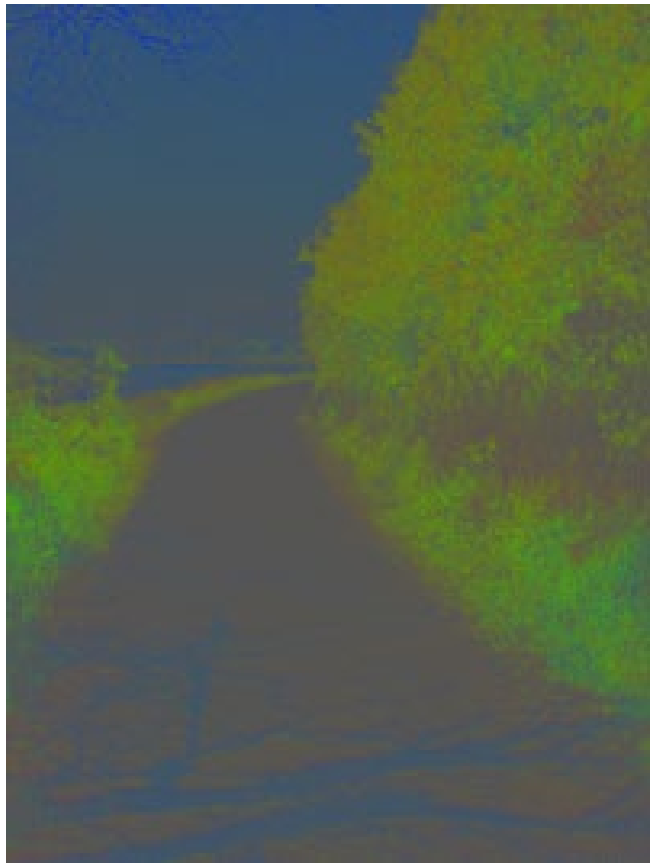
$$(R, G, B) \Leftrightarrow (r, g, I) \qquad I = \frac{R+G+B}{3}$$

Why separate colour and intensity?

Introduction to Video & Image Processing, Thomas B. Moeslund, Springer 2012

Intensity carries more "information"

Original RGB Image



Only color shown - Constant intensity

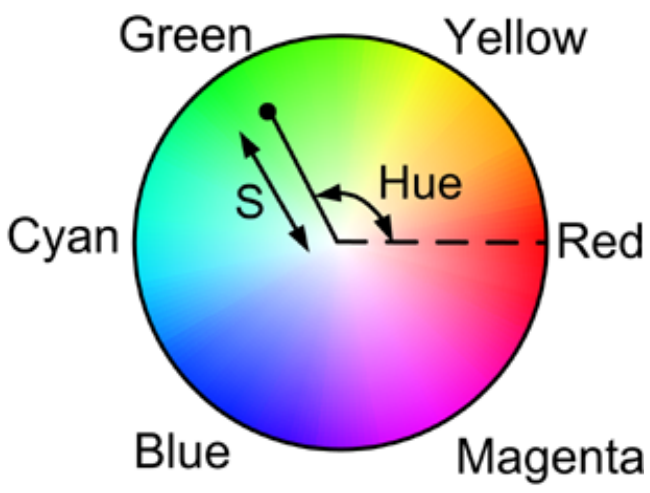
Only intensity shown – Constant color

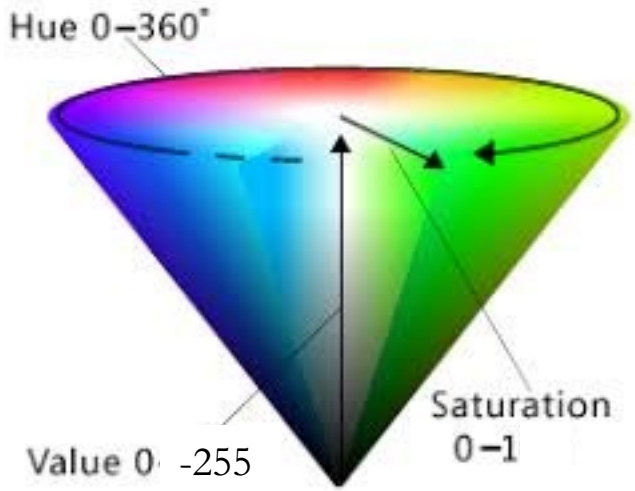
https://en.wikipedia.org/wiki/Rg_chromaticity

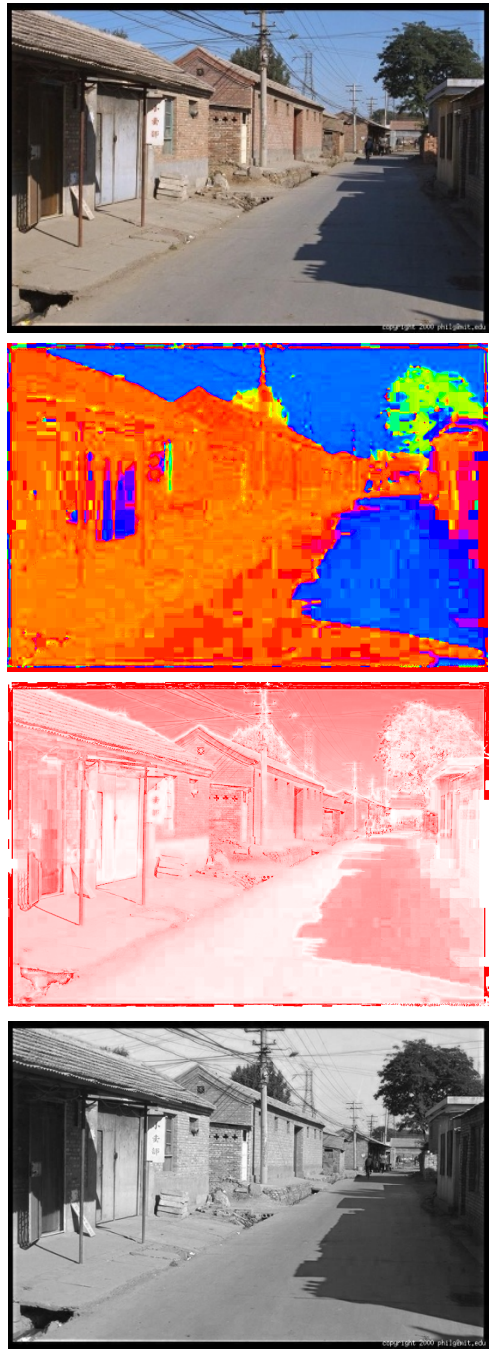
Adapted from James Hays

HSV Colour Space

Hue: "pure" colour (0 to 360)
Saturation: "purity", mixing pure colour (1) with white light (0)
Value: achromatic mixing from black (0) to white (255)







H
(S=1,V=1)

S
(H=1,V=1)

V
(H=1,S=0)

Adapted from James Hays

HSV Colour Space

$$H = \begin{cases} \frac{G-B}{V-\min\{R,G,B\}} \cdot 60°, & \text{if } V = R \text{ and } G \geq B; \\ \left(\frac{B-R}{V-\min\{R,G,B\}} + 2\right) \cdot 60°, & \text{if } G = V; \\ \left(\frac{R-G}{V-\min\{R,G,B\}} + 4\right) \cdot 60°, & \text{if } B = V; \\ \left(\frac{R-B}{V-\min\{R,G,B\}} + 5\right) \cdot 60°, & \text{if } V = R \text{ and } G < B \end{cases} \qquad H \in [0°, 360°[$$

$$S = \frac{V - \min\{R, G, B\}}{V} \quad S \in [0, 1]$$

$$V = \max\{R, G, B\} \quad V \in [0, 255]$$

Introduction to Video & Image Processing, Thomas B. Moeslund, Springer 2012

What is Colour Good for?

Indexing and Retrieval!

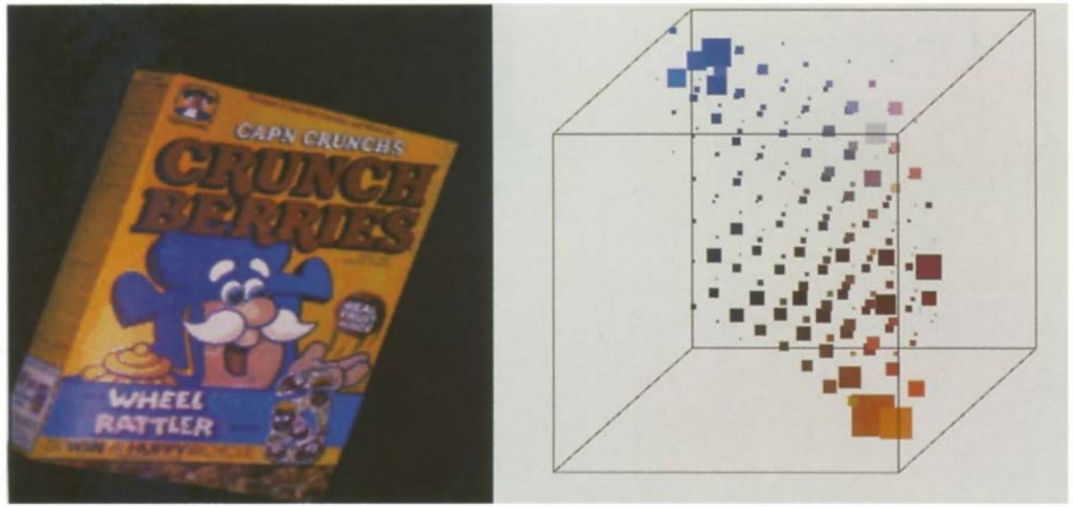


Fig. 1. Left: Image of a Crunchberries cereal box. Right: Three dimensional color histogram of the Crunchberries image with the black background substrated.

Fig. 2. Four views of Snoopy.

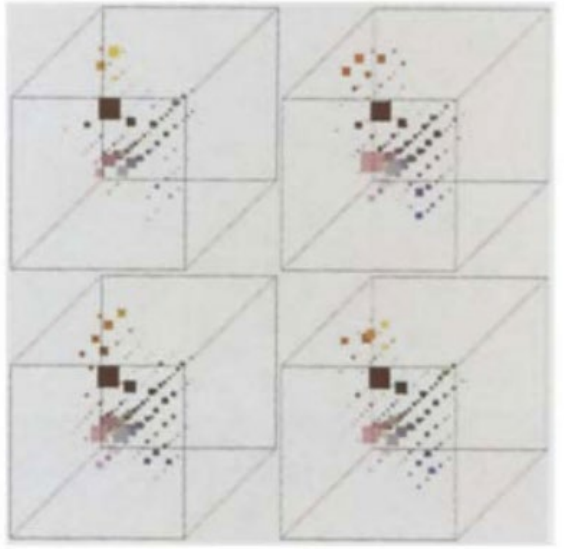


Fig. 3. Histograms of the four views of Snoopy.

This article demonstrates that color histograms of multicolored objects provide a robust, efficient cue for indexing into a large database of models. It shows that color histograms are stable object representations in the presence

http://www.inf.ed.ac.uk/teaching/courses/av/LECTURE_NOTES/swainballard91.pdf

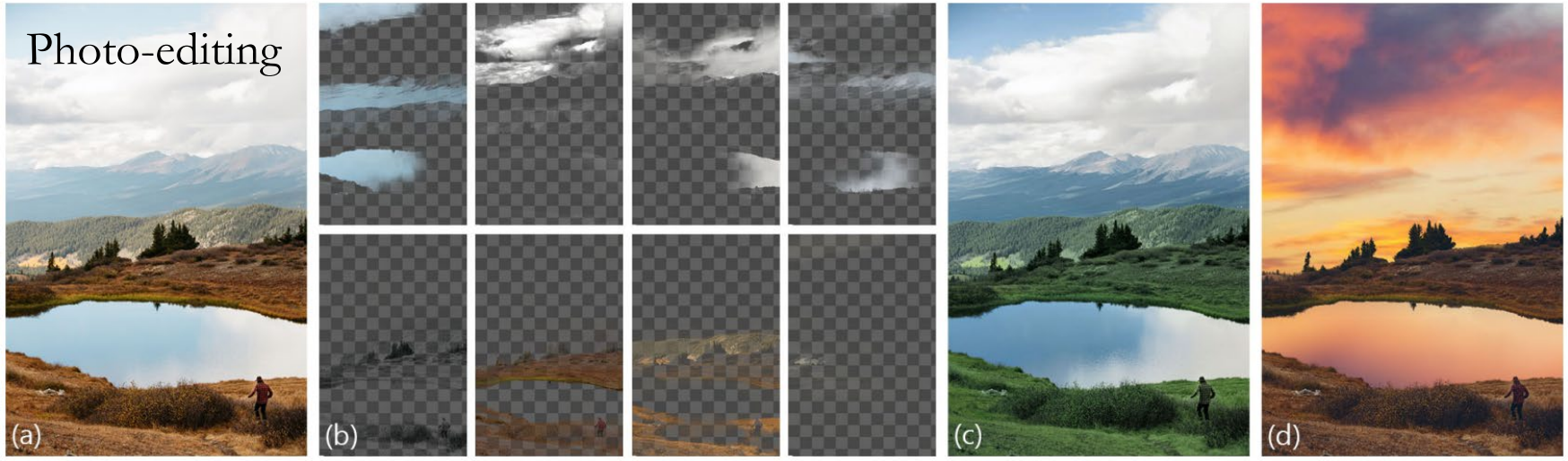
Colour-Based Segmentation

HWQ: Why is green preferred for special effects?

Green-screen for virtual backgrounds

Skin detection

Photo-editing



http://yaksoy.github.io/scs/

Summary

1. Computer vision aims to automatically interpret images (and video):
   - Real-world measurements, semantics & recognition
   - Manipulation & generation, organization & search
2. Visual perception is inherently ambiguous, algorithms have a hard time generalizing to all the challenges of the real visual world
3. Digital images
   - Exposure time & gain are two critical parameters of image capture
   - Quantization determines resolution in space and intensity
4. Colour Representations
   - Colour: Bayer mosaic, different colour spaces (RGB, HSV)
   - Intensity more informative than colour
   - Colour can be used for indexing & retrieval, segmentation (e.g. skin)