Introduction

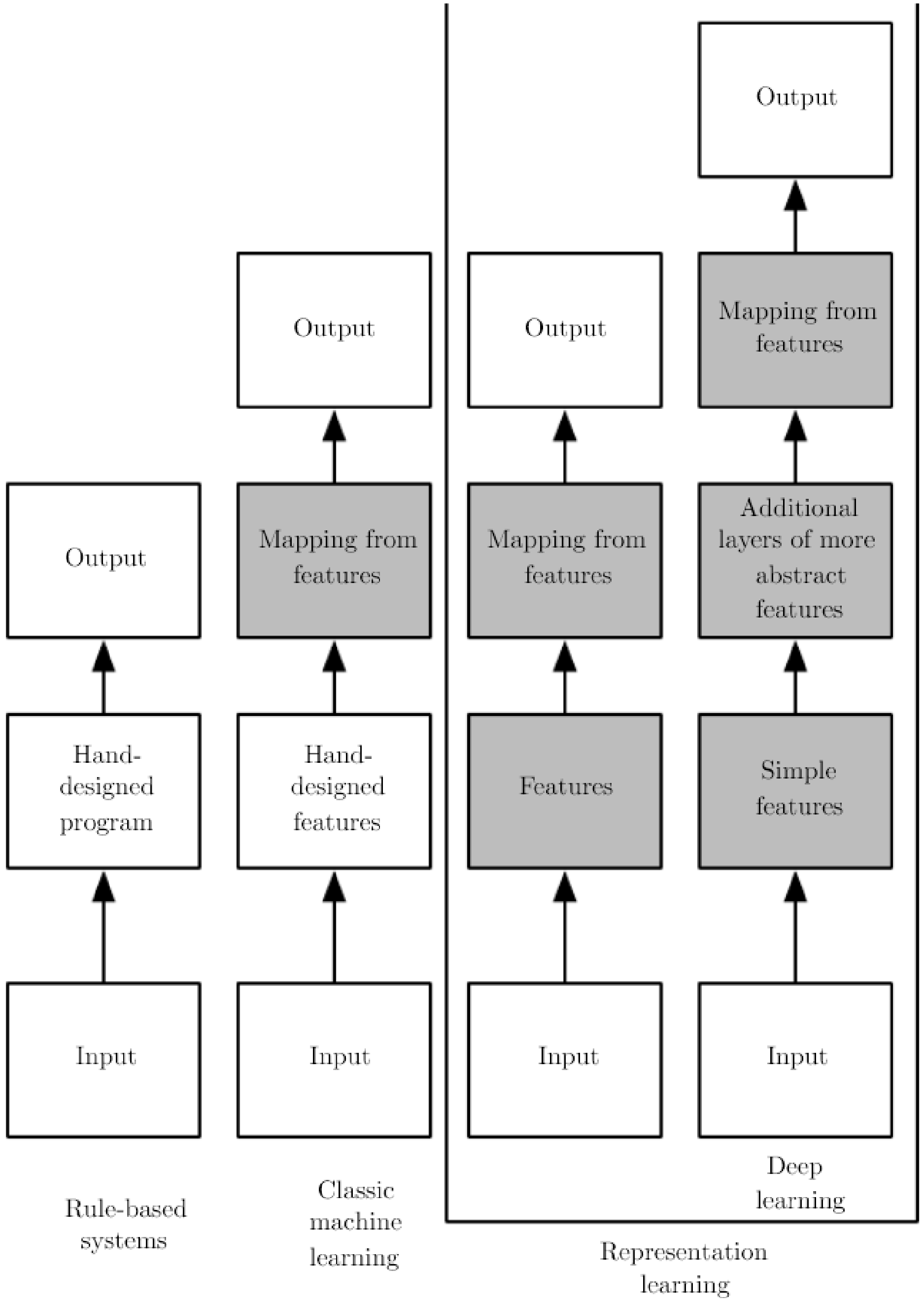
Computer Vision Apps in Deep Learning

ENEE 4584/5584

Slide Credits: Gonzalez & Woods

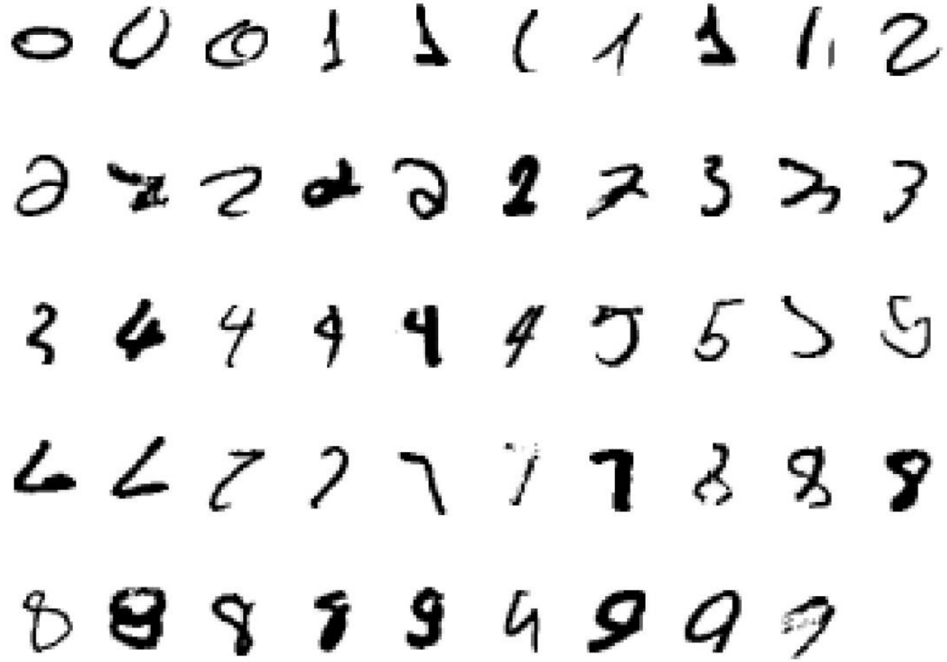
Why NN?

- NN is a presentation learning

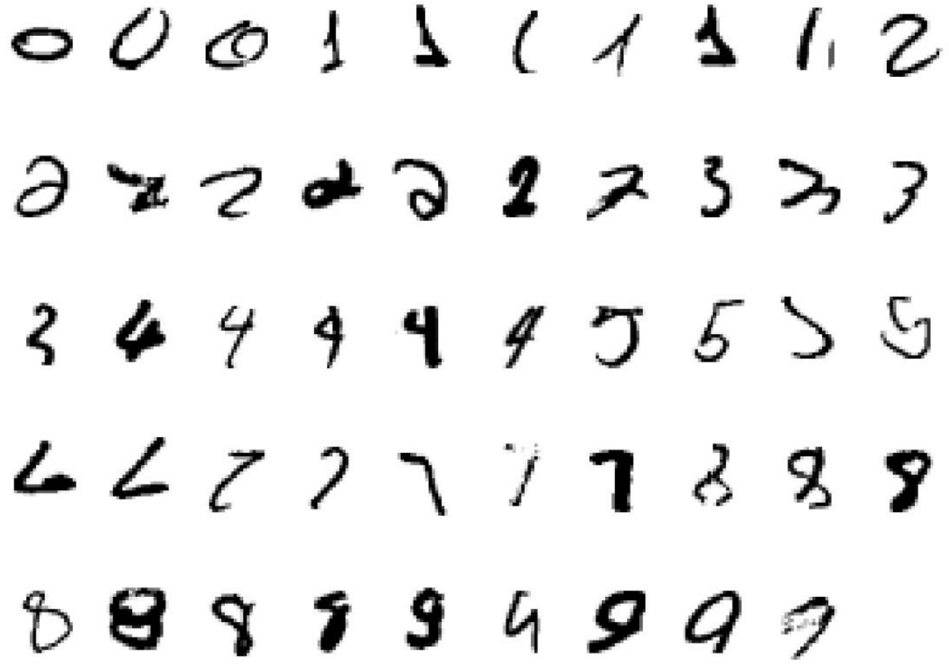


Example: Handwritten Digits

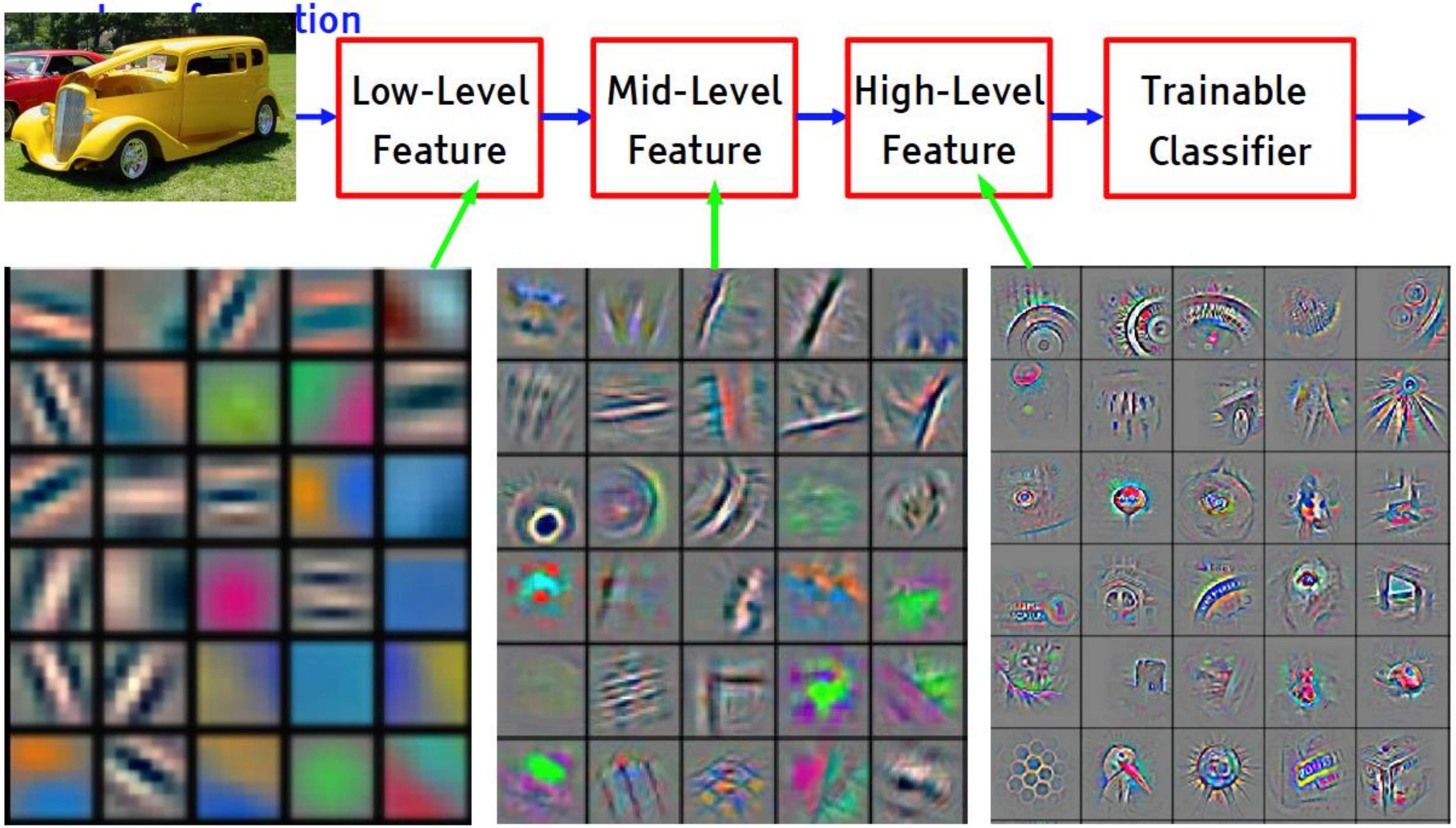
- Pixel images
- 2D Challenges:
  - Photometric: Noise, occlusions
  - Geometric: Character size, orientation (rotation)
  - Thickness, styles



Example: Handwritten Digits



- Classical ML:
  - Develop a set of invariant features to detect.
    - Vertical lines, horizontal lines
    - Closed loops, curves
    - Intersection, end points
  - Map features to class
- NN:
  - Feed raw images (presentation) to network
  - Control how the network updates weights and biases
  - Network learns features
    - Deep networks learns features within features
  - Network learns mapping



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

What’s New?

- Classical deep NN are difficult to train
  - More data is available
  - Better design software
  - Better generalization algorithms
  - Newer architectures

Computer Vision

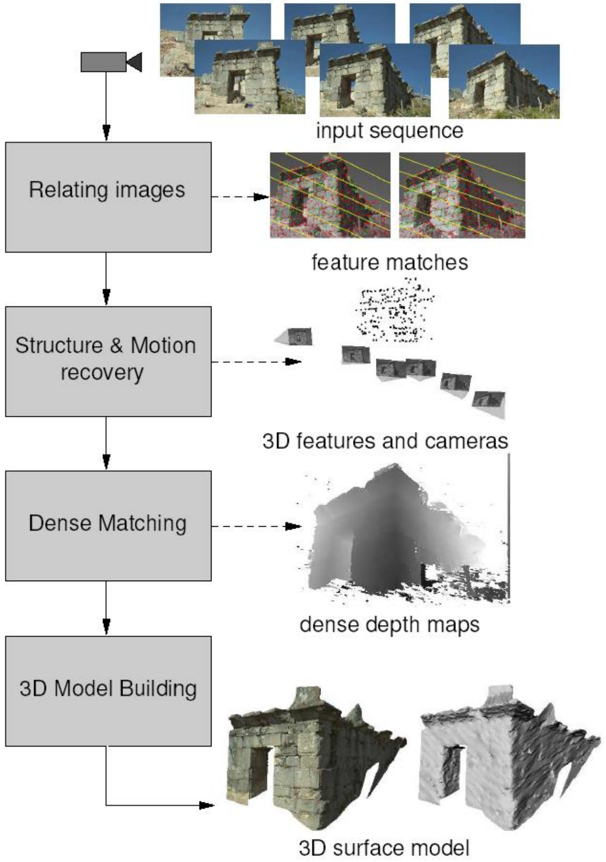
- Automatic
- Understanding
  - *Measurement*
  - *Perception and interpretation*.
- of Images:
  - Data set of x,y; Distorted z.
- and Video:
  - Sequence of images
  - Effect of time.

Measurement

Real-time stereo

Pollefeys et al.

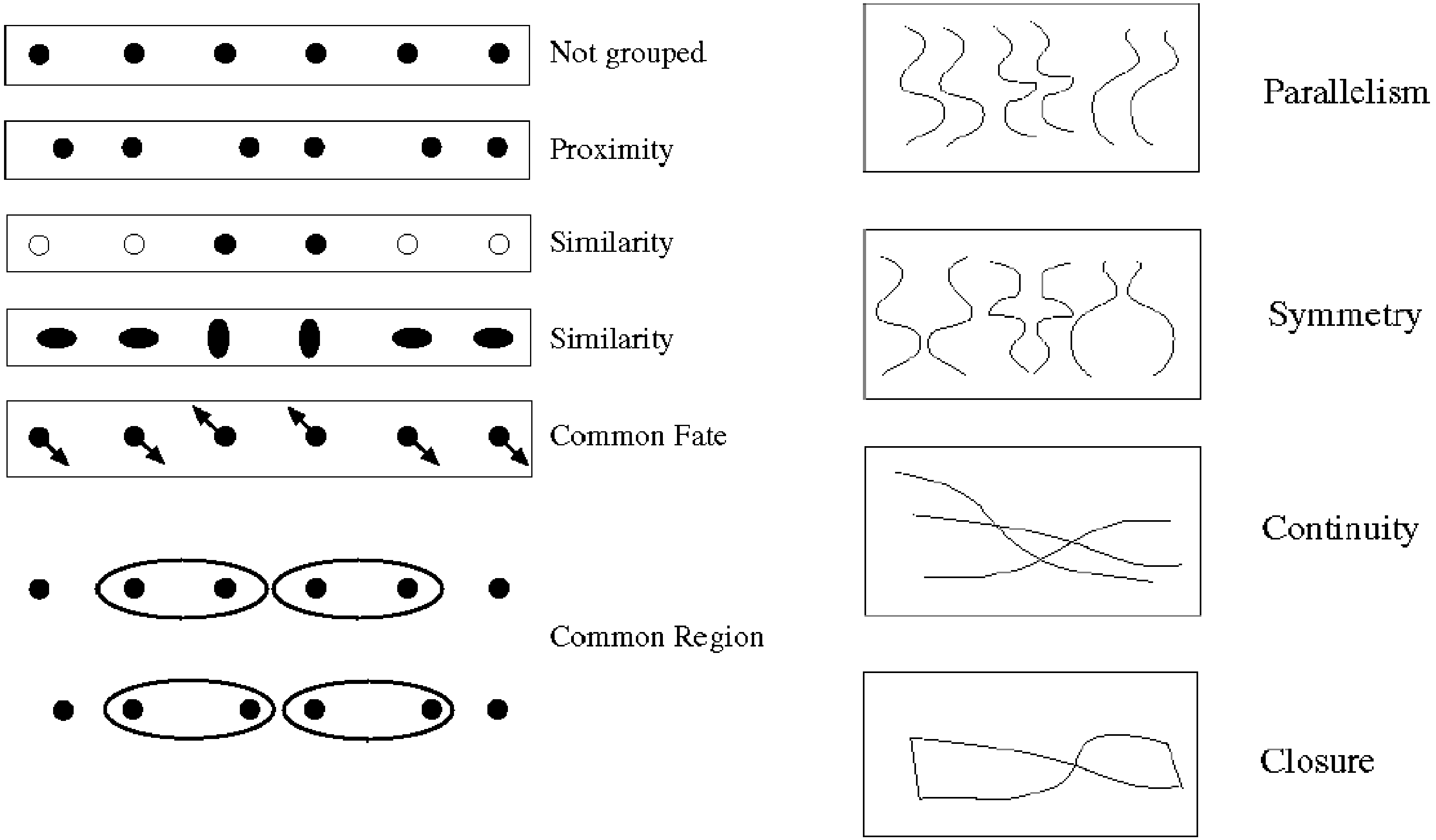
Structure from motion



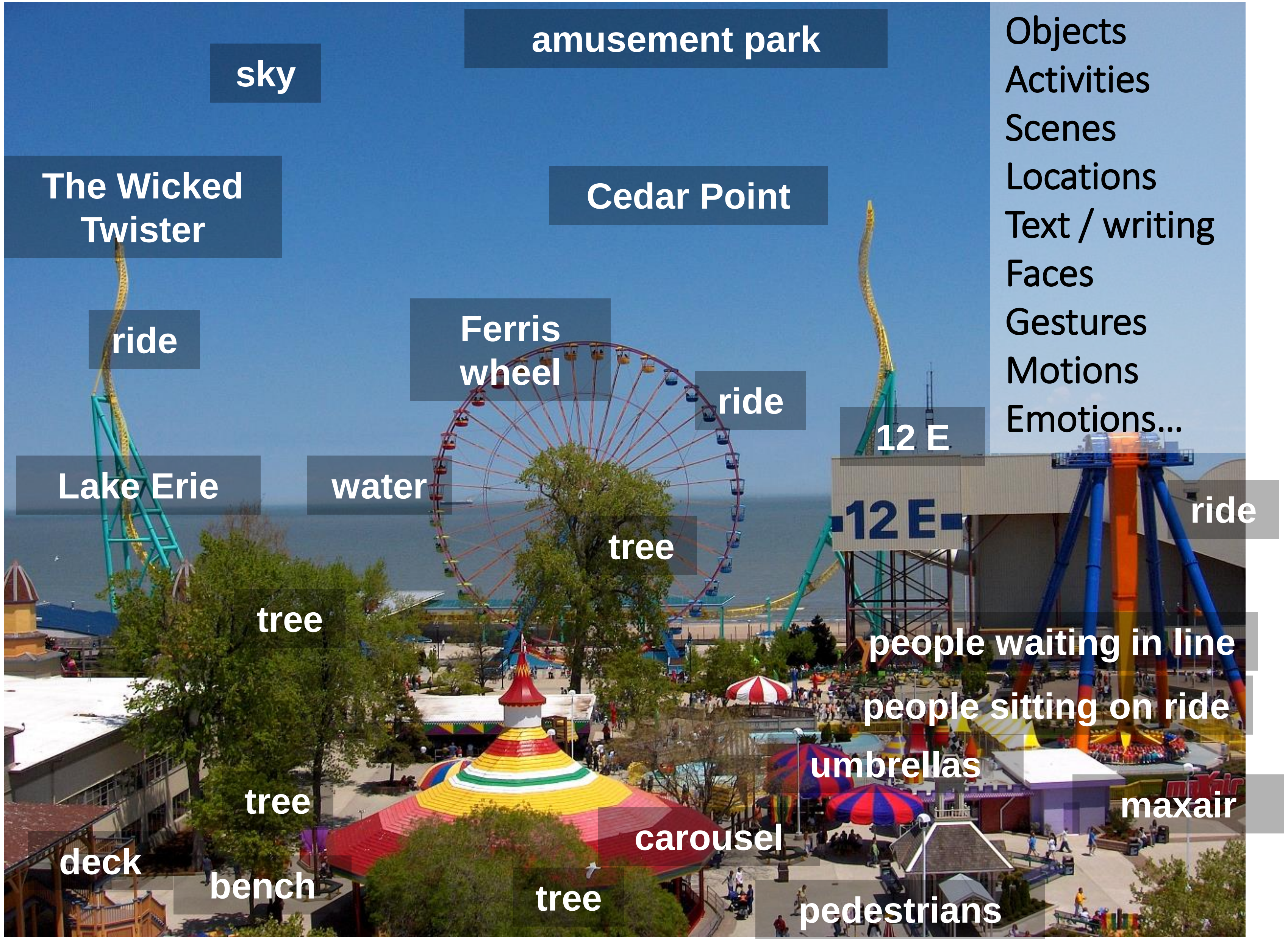
Multi-view stereo for community photo collections

Goesele et al.

Perception

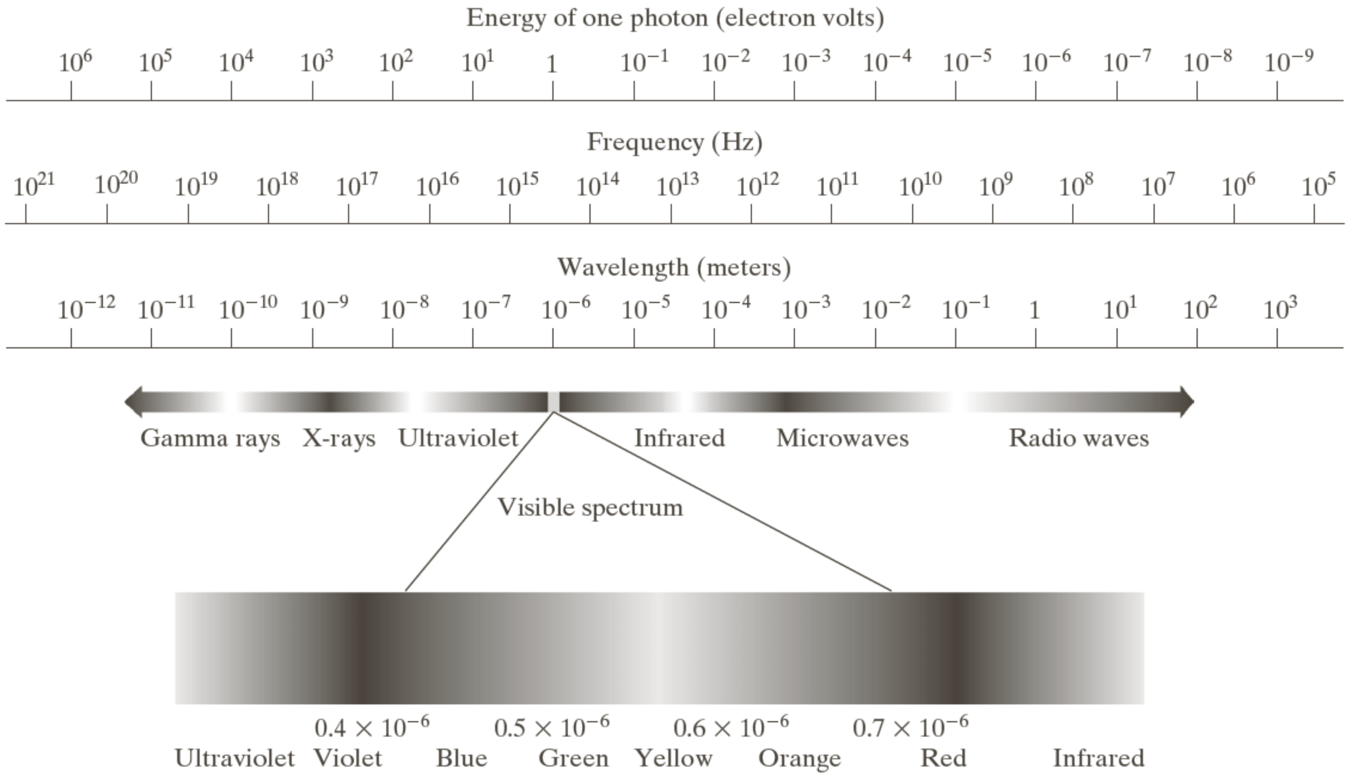

Not grouped
Proximity
Similarity
Similarity
Common Fate
Common Region
Parallelism
Symmetry
Continuity
Closure

Perception, Interpretation



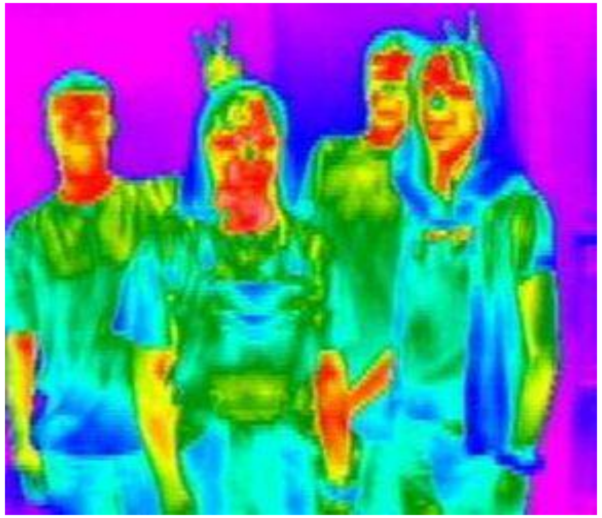
Image

- Image
  - 2D spatial signal: f(x,y)
  - Contains a lot of information: “worth a thousand words”
  - Can be captured or generated
  - Can be captured from a infinitely continuous 3D dynamic source
    - Projection of 3D world to 2D images => loss of information
      - S(x,y,z) => f(x,y)
    - dynamic scenes => occlusion, transformation, motion, bright sources, noise
      - S(x,y,z,t) => f(x,y)
  - Presented in the visible light range
  - Can be captured from a non-visible range

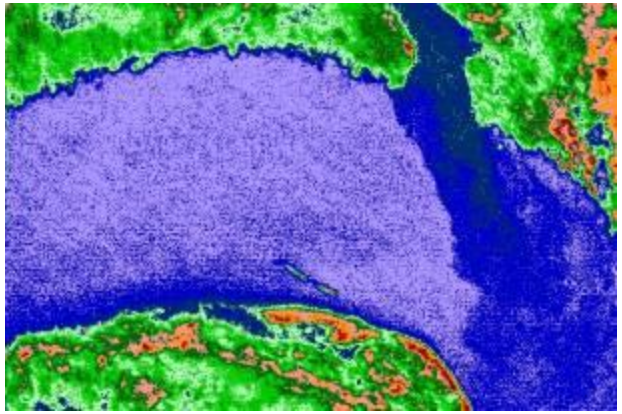


Band No.	Name	Wavelength (μm)	Characteristics and Uses
1	Visible blue	0.45–0.52	Maximum water penetration
2	Visible green	0.52–0.60	Good for measuring plant vigor
3	Visible red	0.63–0.69	Vegetation discrimination
4	Near infrared	0.76–0.90	Biomass and shoreline mapping
5	Middle infrared	1.55–1.75	Moisture content of soil and vegetation
6	Thermal infrared	10.4–12.5	Soil moisture; thermal mapping
7	Middle infrared	2.08–2.35	Mineral mapping

Operate in infrared frequency

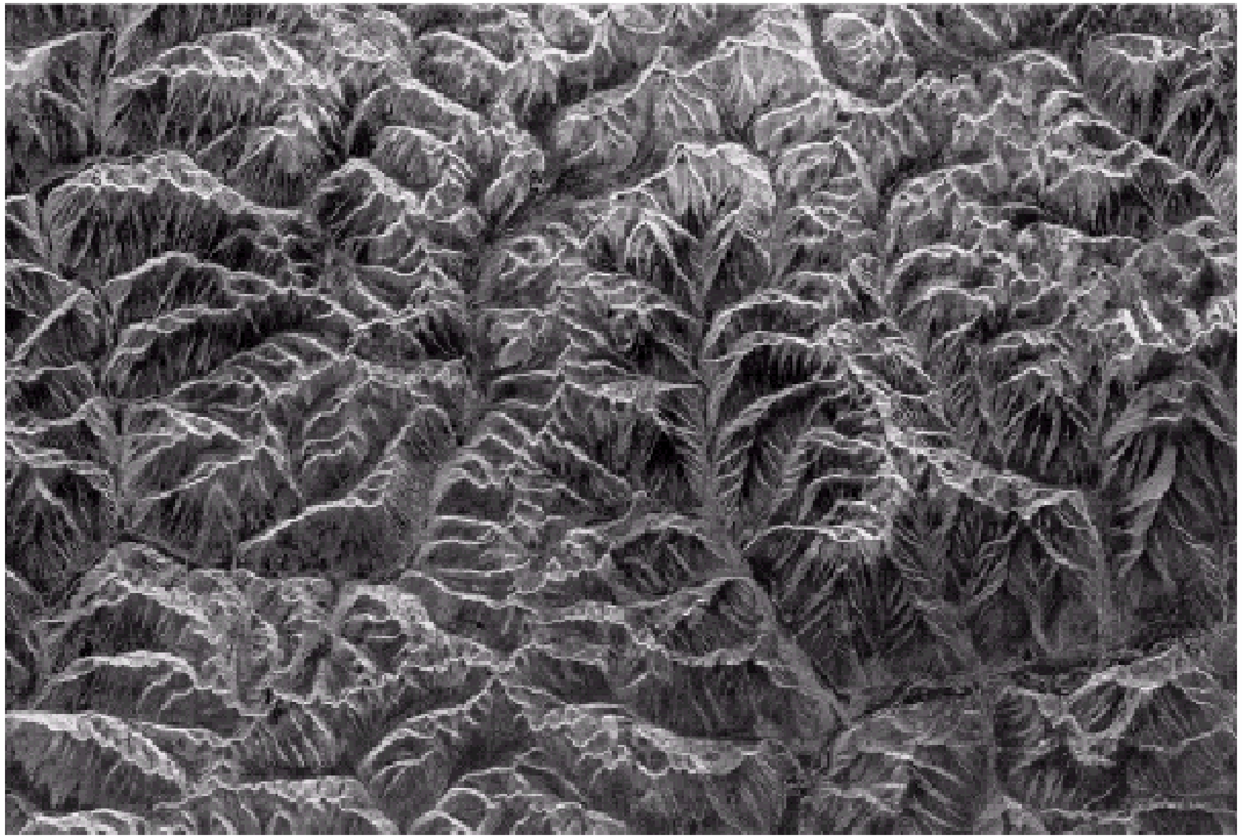


Human body disperses heat (red pixels)



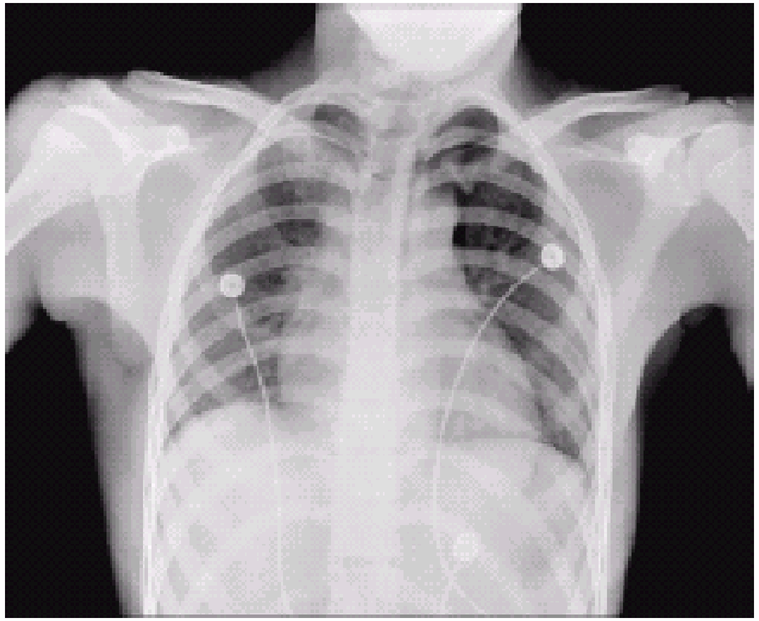
Different colors indicate varying temperatures

Operate in microwave frequency

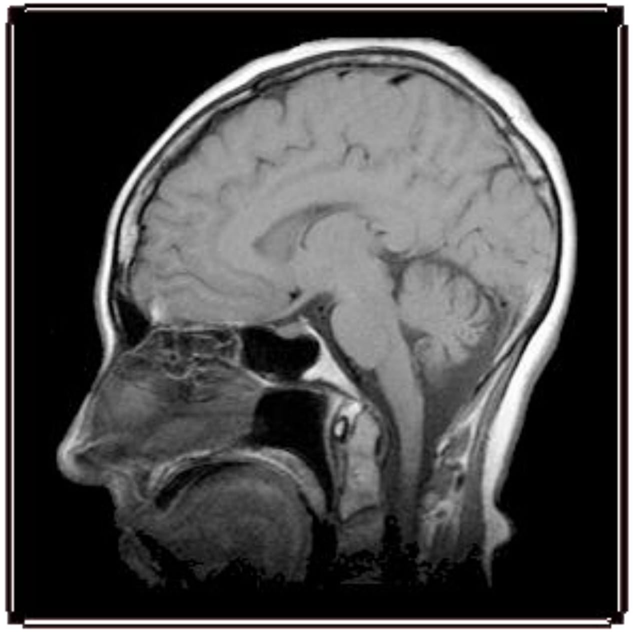


Moutains in Southeast Tibet

Operate in X-ray frequency



Operate in radio frequency

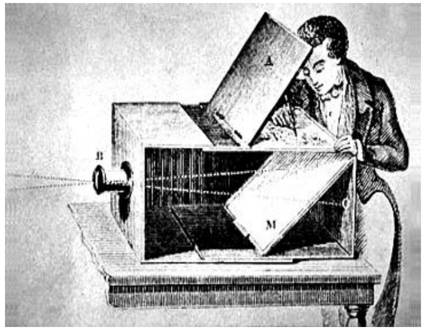


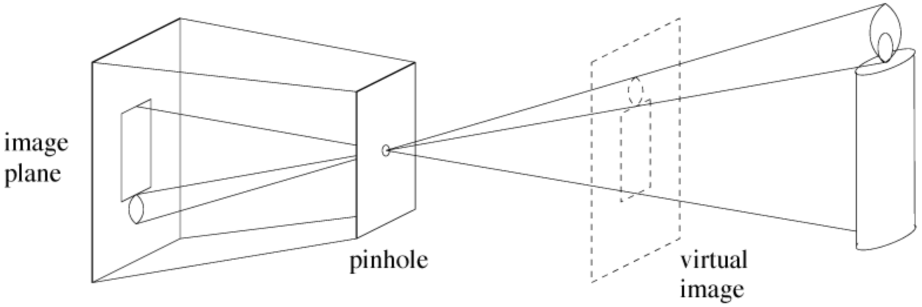
Camera History

- Pinhole model: Mozi (470-390 BCE), Aristotle (384-322 BCE)
- Principles of optics (including lenses): Ibn Al-Haytham (965-1039 CE)
- Camera obscura: Leonardo da Vinci (1452-1519), Johann Zahn (1631-1707)
- First photo: Joseph Nicephore Niepce (1822)
- Daguerréotypes (1839)
- Photographic film (Eastman, 1889)
- Cinema (Lumière Brothers, 1895)
- Color Photography (Lumière Brothers, 1908)
- First consumer camera with CCD: Sony Mavica (1981)
- First fully digital camera: Kodak DCS100 (1990)

Digital Image Acquisition

- First camera was invented in 16th century.
- Camera Obscura
- It used a *pinhole* to focus light rays onto a wall or translucent plate.

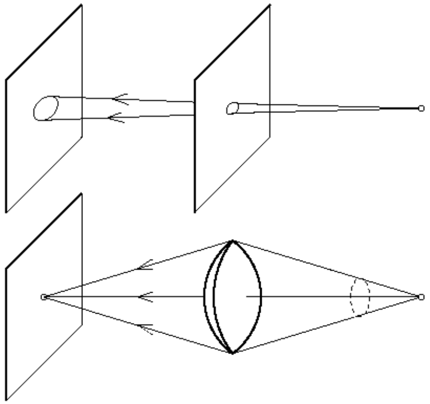




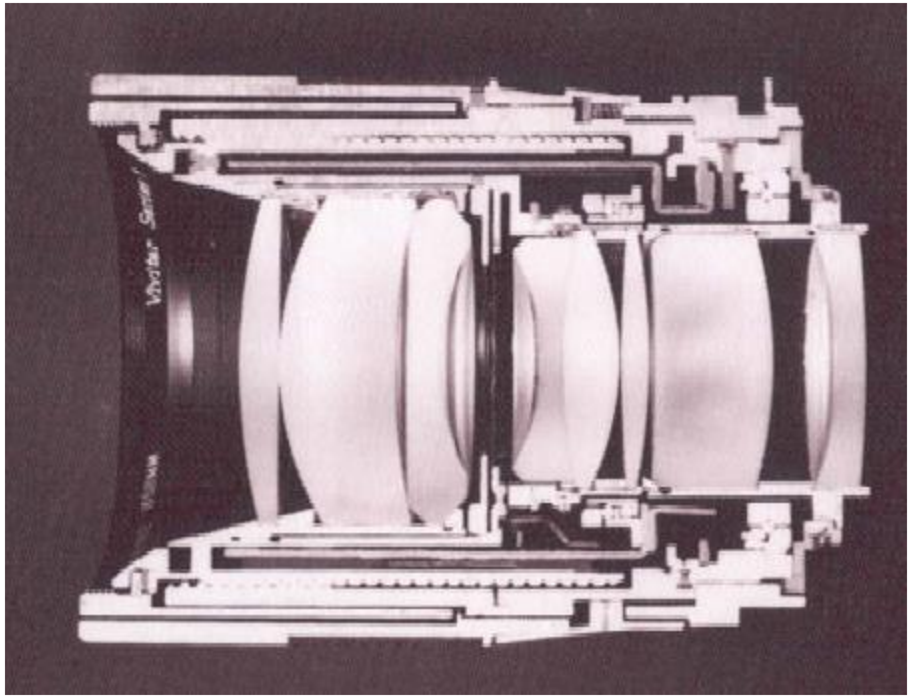
- Take a box, prick a small hole in one of its sides with a pin, and then replace the opposite side with a translucent plate.
- Place a candle on the pinhole side, you will see an inverted image of the candle on the translucent plate.

Cameras With Lenses

- Most cameras are equipped with lenses.
- There are two main reasons for this:
  - To gather light. For an ideal pinhole, a single light ray would reach each point the image plane. Real pinholes have a finite size, so each point in the image plane is illuminated by a cone of light rays. The larger the hole, the wider the cone and the brighter the image => blurry pictures. Shrinking the pinhole produces sharper images, but reduces the amount of light and may introduce diffraction effects.
  - To keep the picture in sharp focus while gathering light from a large area.

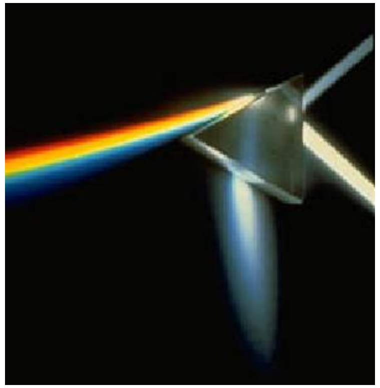


Compound Lens Systems



Real Lenses

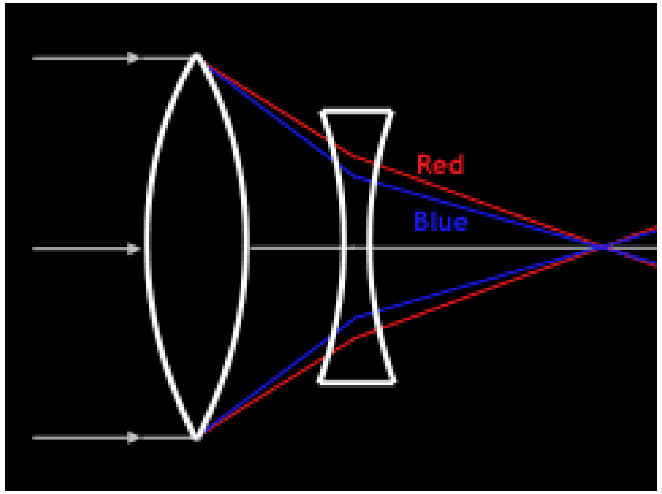
- The index of refraction is a function of wavelength.
- Light at different wavelengths follow different paths.



Chromatic aberration

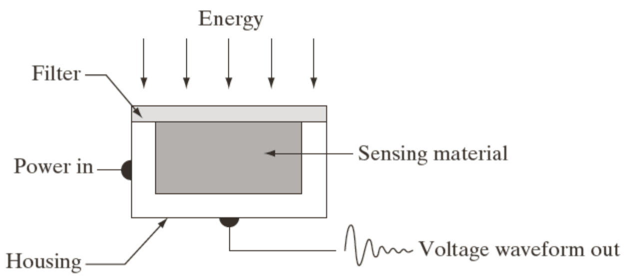
Real Lenses

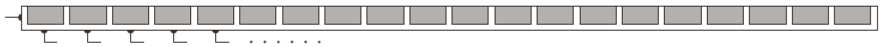
- Special lens systems using two or more pieces of glass with different refractive indices can reduce or eliminate this problem. However, not even these lens systems are completely perfect and still can lead to visible chromatic aberrations.

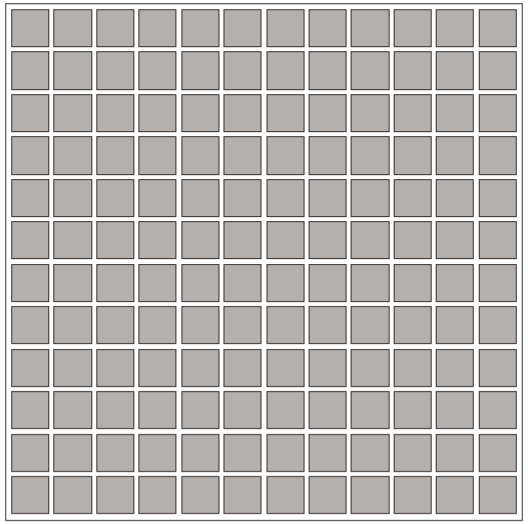


Digital Detectors

- Photon detectors
  - When photons strike, electron-hole pairs are generated on sensor sites.
  - Electrons generated are collected over a certain period of time.
  - The number of electrons are converted to pixel values.







Digital Detectors

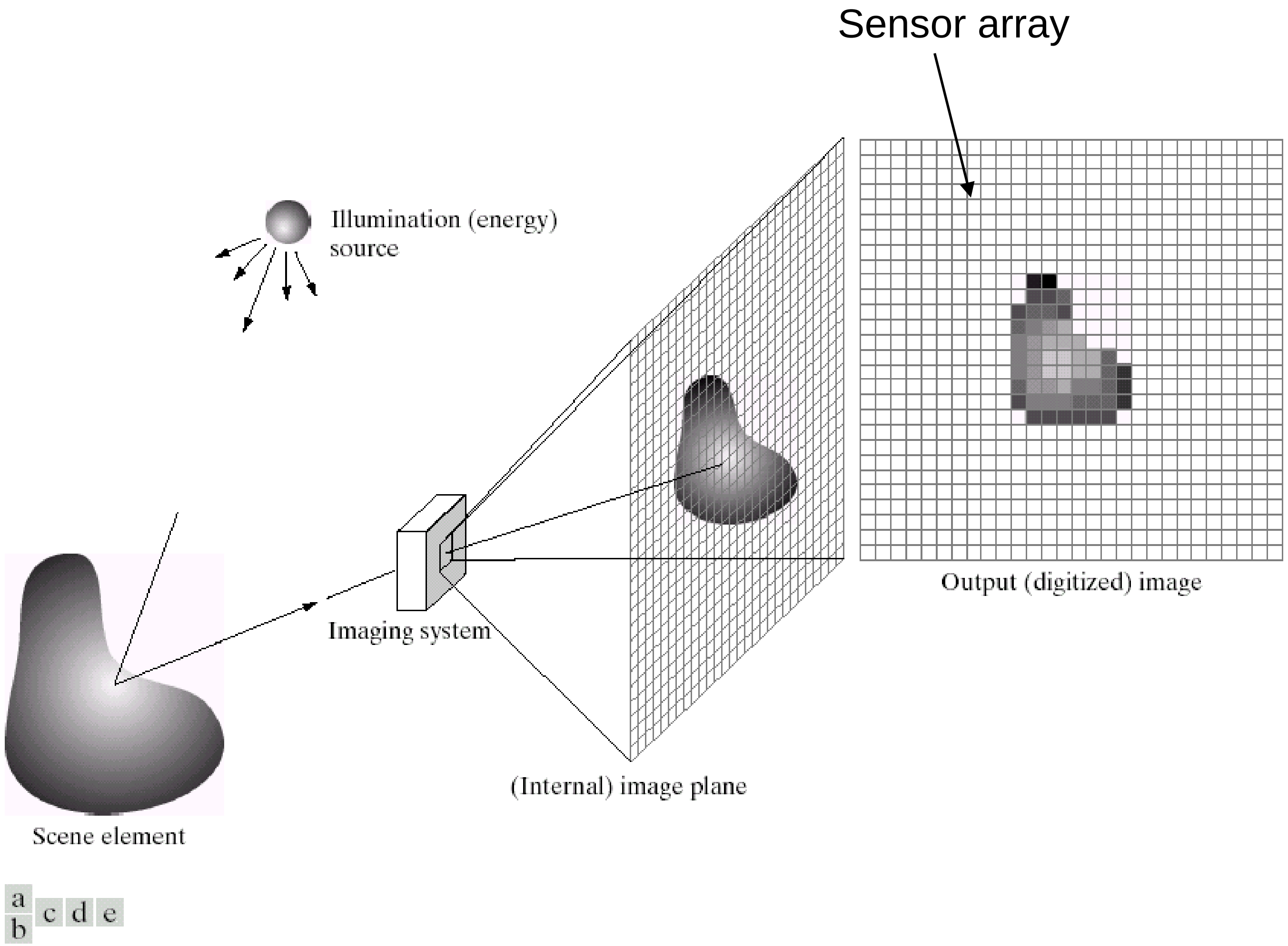
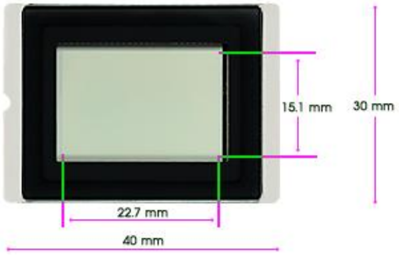
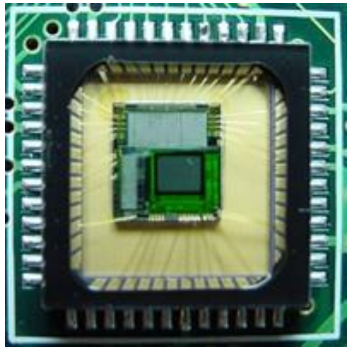


FIGURE 2.15 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.





Resolution

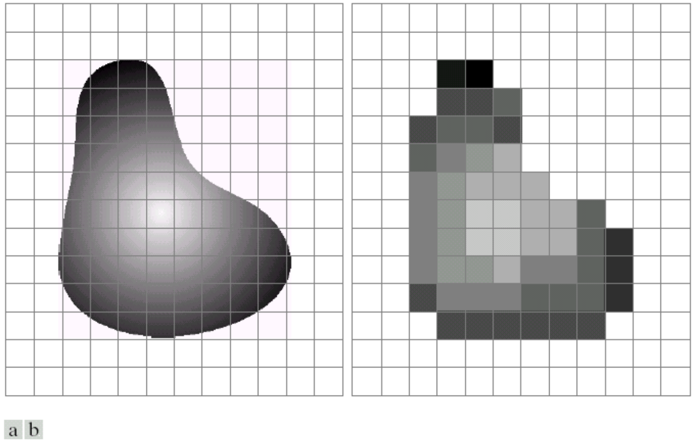
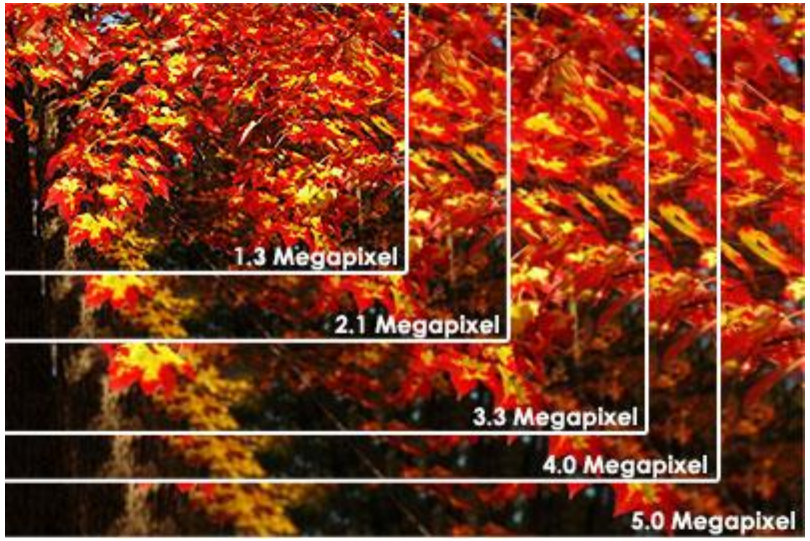


FIGURE 2.17 (a) Continuos image projected onto a sensor array. (b) Result of image sampling and quantization.



- **256x256**
  - cheap
  - 65,000 total pixels.
- **640x480**
  - low end
  - ideal for e-mailing
- **1216x912**
  - "megapixel" = 1,109,000 pixels
  - good for printing pictures.
- **1600x1200**
  - ~2 M pixels,
  - "high resolution
  - 4x5 inch print same quality from a photo lab.
- **2240x1680**
  - 4 megapixel cameras
  - the current standard
  - good quality for prints up to 16x20 inches.
- **4064x2704**
  - A top-of-the-line
  - 11.1 megapixels
  - 13.5x9 inch prints with no loss of picture quality.

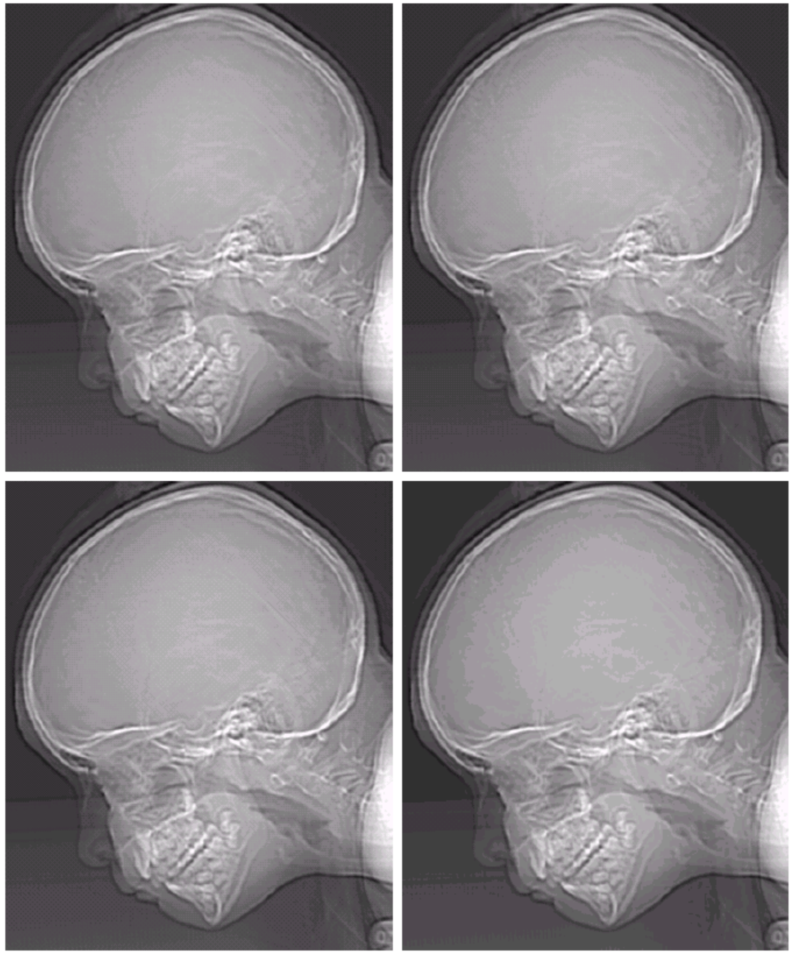
Digital Images

- Discrete and Quantized
  - Matrix of elements
  - Each element of an image is a pixel (Picture element).
- Sampled => discrete => resolution (tricky)
  - Pixel resolution => nxk pixels
  - Image resolution is not necessarily pixel resolution
  - Spatial resolution = DPI
- Quantized levels => intensity resolution
  - Half tone = black white
  - Gray scale => shades of black
- Stored
  - Size determined by number of pixels, quantization levels, colors, compression

- Newspapers 75dpi
- Magazines 133dpi
- Glossy brochures 175dpi

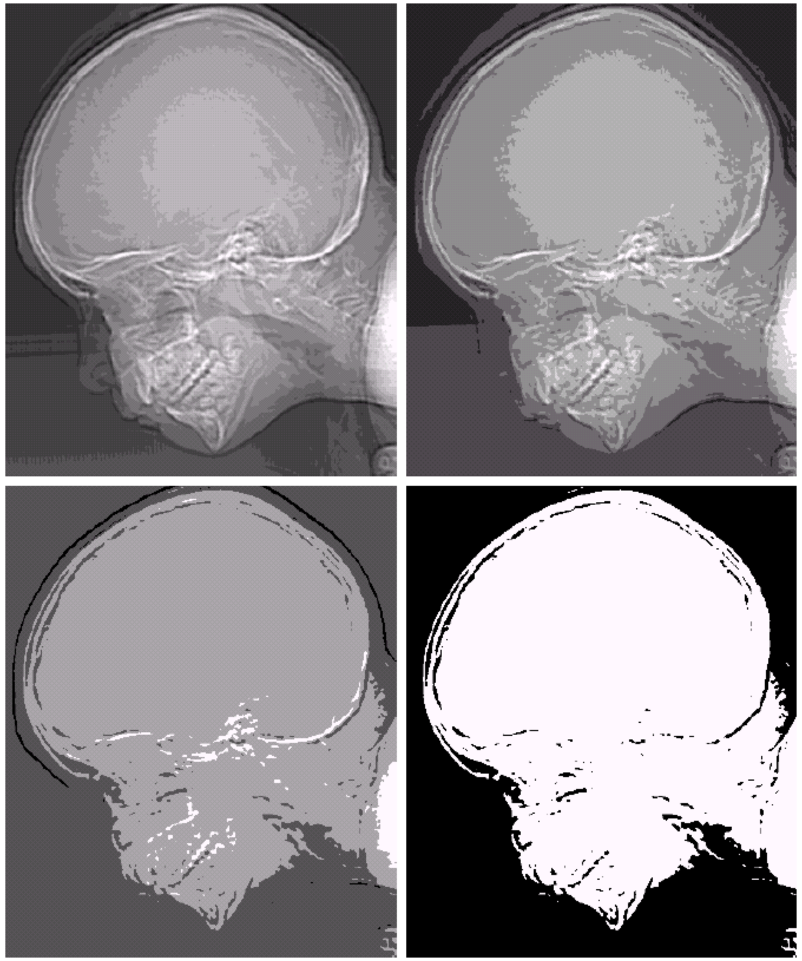
a b
c d

FIGURE 2.20 Typical effects of reducing spatial resolution. Images shown at: (a) 1250 dpi, (b) 300 dpi, (c) 150 dpi, and (d) 72 dpi. The thin black borders were added for clarity. They are not part of the data.



a b
c d

FIGURE 2.21
(a) 452 × 374, 256-level image. (b)–(d) Image displayed in 128, 64, and 32 gray levels, while keeping the spatial resolution constant.



e f
g h

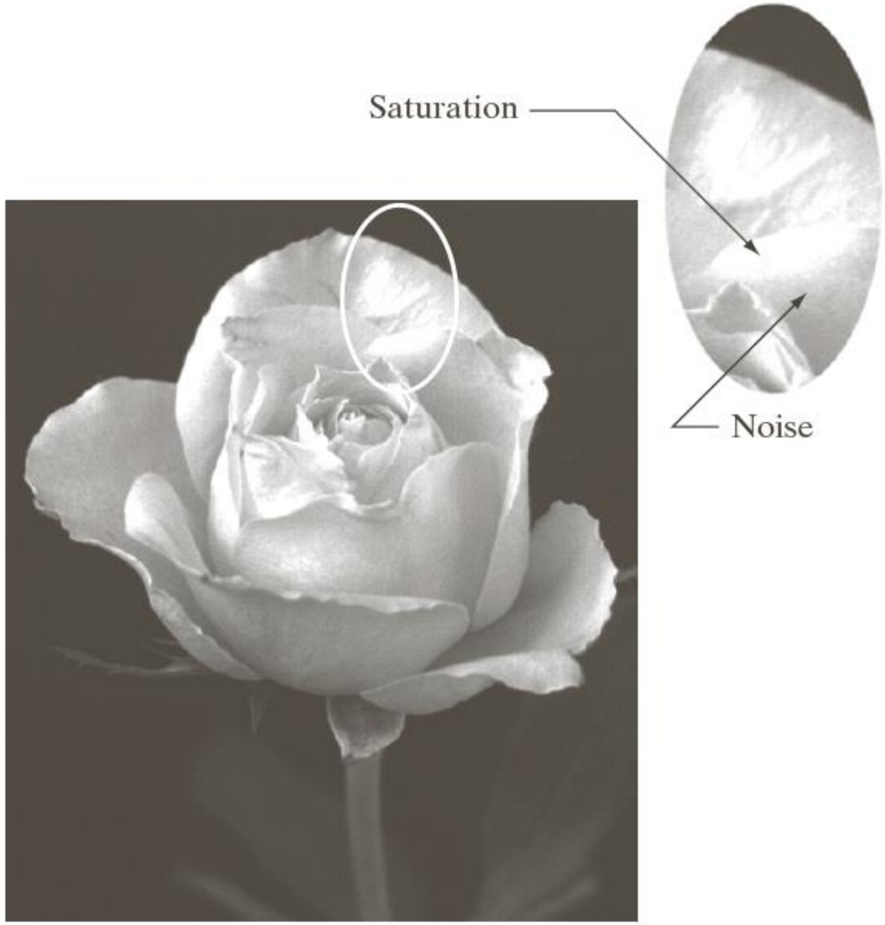
FIGURE 2.21 *(Continued)* (e)–(h) Image displayed in 16, 8, 4, and 2 gray levels. (Original courtesy of Dr. David R. Pickens, Department of Radiology & Radiological Sciences, Vanderbilt University Medical Center.)

TABLE 2.1

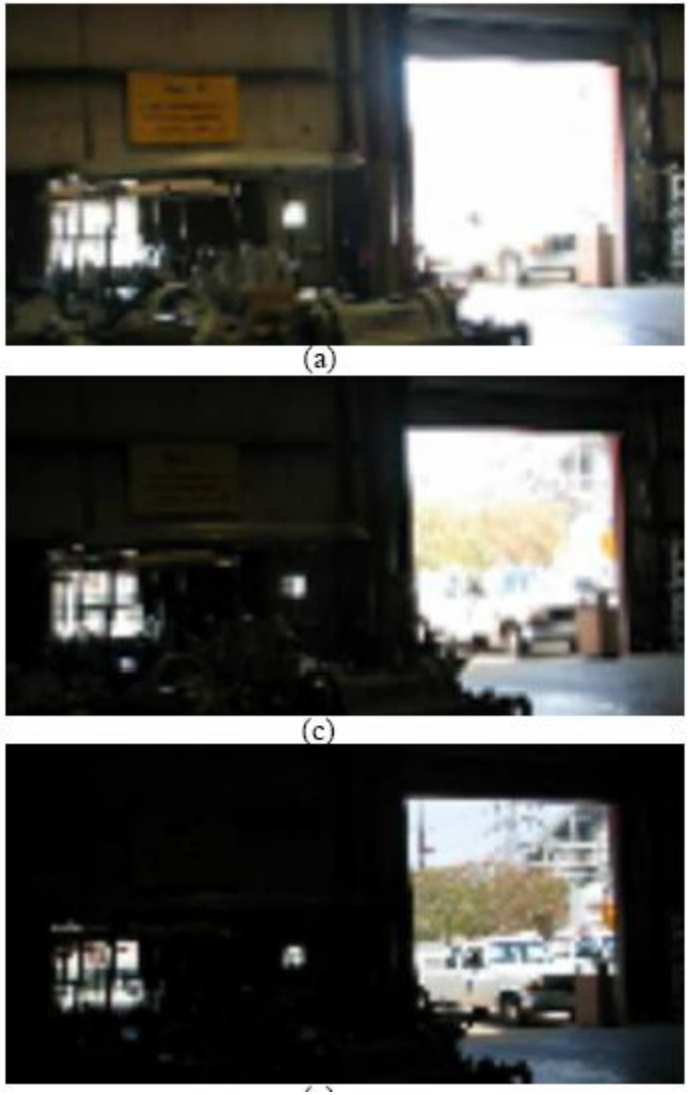
Number of storage bits for various values of N and k.

N/k	1 (L = 2)	2 (L = 4)	3 (L = 8)	4 (L = 16)	5 (L = 32)	6 (L = 64)	7 (L = 128)	8 (L = 256)
32	1,024	2,048	3,072	4,096	5,120	6,144	7,168	8,192
64	4,096	8,192	12,288	16,384	20,480	24,576	28,672	32,768
128	16,384	32,768	49,152	65,536	81,920	98,304	114,688	131,072
256	65,536	131,072	196,608	262,144	327,680	393,216	458,752	524,288
512	262,144	524,288	786,432	1,048,576	1,310,720	1,572,864	1,835,008	2,097,152
1024	1,048,576	2,097,152	3,145,728	4,194,304	5,242,880	6,291,456	7,340,032	8,388,608
2048	4,194,304	8,388,608	12,582,912	16,777,216	20,971,520	25,165,824	29,369,128	33,554,432
4096	16,777,216	33,554,432	50,331,648	67,108,864	83,886,080	100,663,296	117,440,512	134,217,728
8192	67,108,864	134,217,728	201,326,592	268,435,456	335,544,320	402,653,184	469,762,048	536,870,912

Saturation



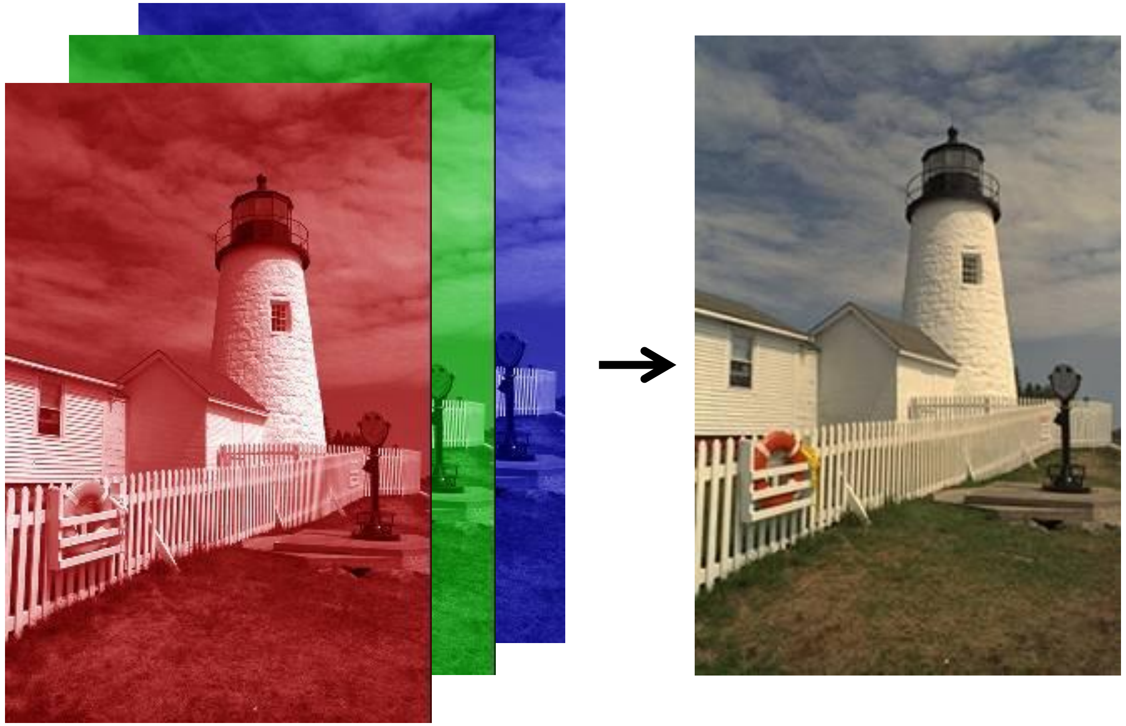
Exposure

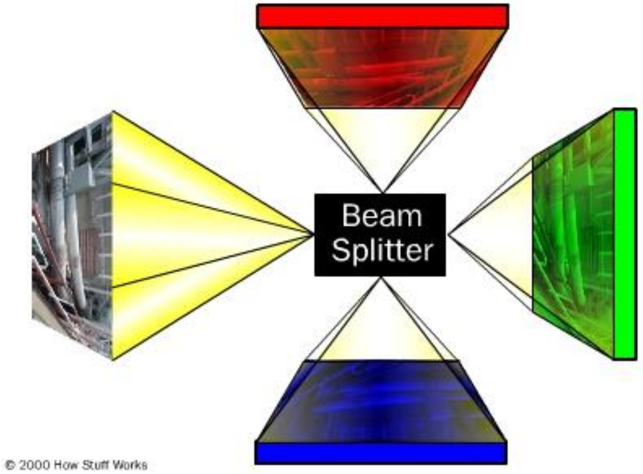


Long exposure time

Short exposure time

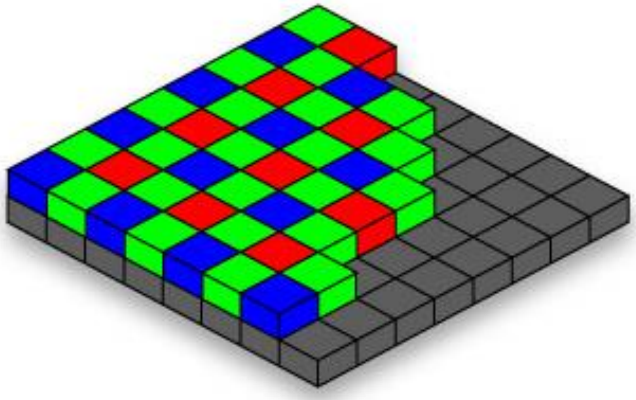
Color Detection

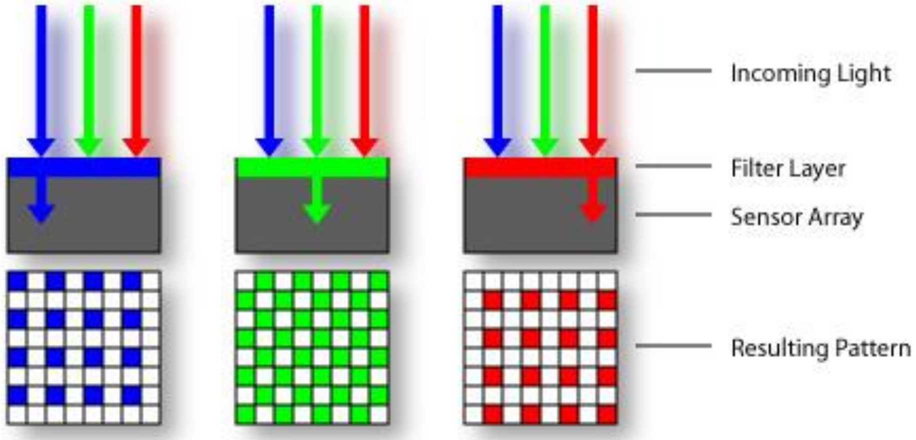




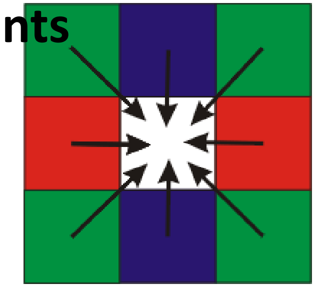
Color Detection

Bayer grid





Estimate missing compo-nents from neighboring values (demosaicing)



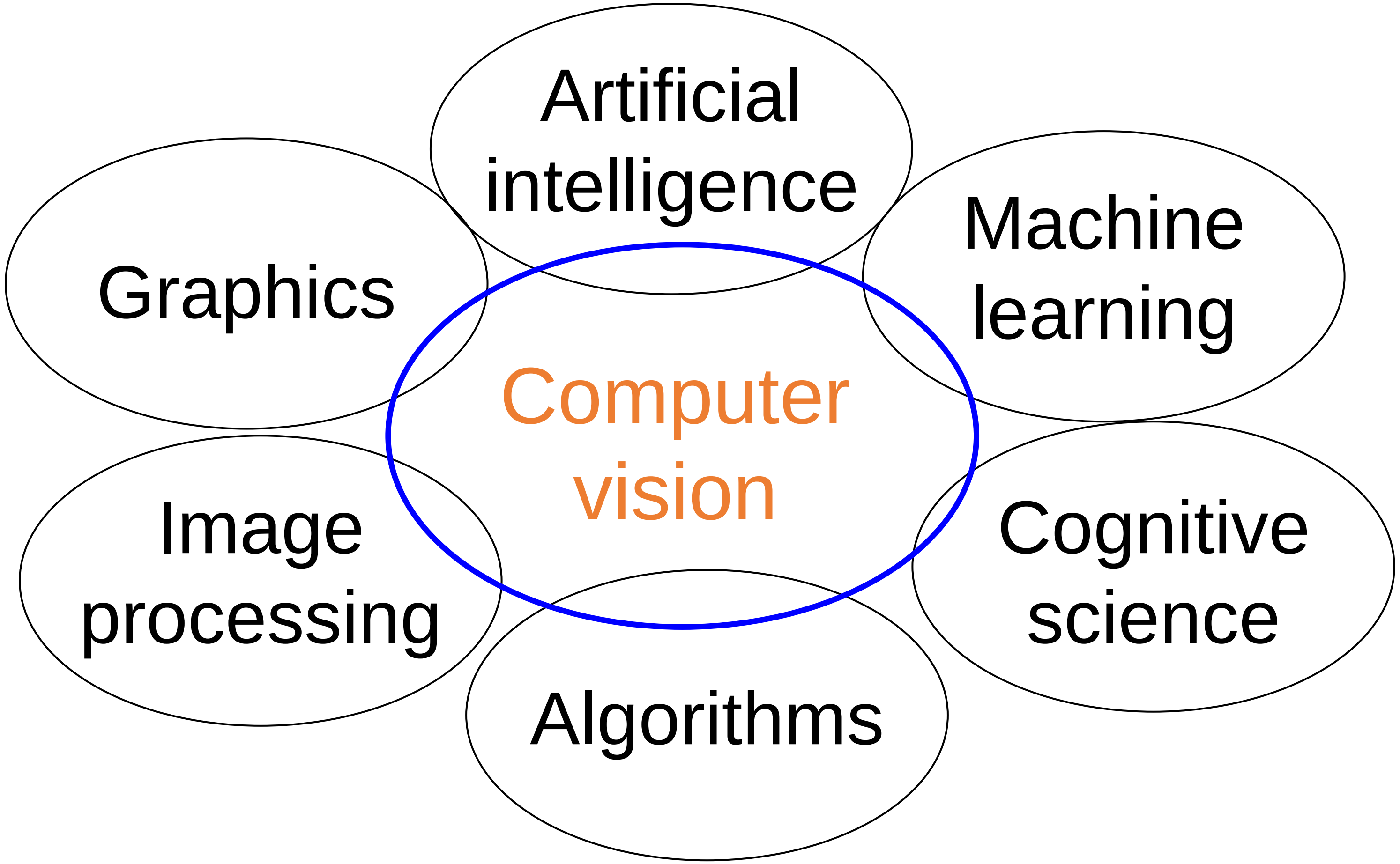
Digital Processing

- Processing done by a digital system (usually a computer).
- Low-level:
  - Processing: images->images
  - Overcoming effects of device and environment
  - Reduce noise, contrast enhancement, sharpening
- Mid-level:
  - Vision: images->attributes
  - Extracting information
  - Segmentation, Description, Tracking, Classification
- High-level
  - Graphics/AI
  - Extracting meaningful information and information fusing
  - Image analysis: images->knowledge, descriptions
  - Multiple images-> Super resolution, stereo vision, 3D reconstruction
  - Knowledge base

Video

- Time varying signal: images on frames, multiple frames/sec
- Video FPS: 30 (NTSC), 25 (Euro).
- Interlacing: video frame divided into lines
  - Lines is usually displayed in 2 fields
  - Displays 60 fields per seconds

Related Fields



Why is CV Difficult?

- Loss of information in projection from 3D to 2D
- Real world is too complex
  - Difficult to formalize the notion of object, scene, event
    - Appearance: photometric and geometric properties
    - Functionality
  - Difficult to formalize all spatiotemporal relations between objects in the scene
- How to evaluate image interpretation?
  - Some quantitative measures are not suitable
  - Using only qualitative evaluation is not satisfying

CV Applications

- Medicine
- Defense
- Meteorology
- Environmental science
- Manufacture
- Surveillance
- Crime investigation