Video Compression
視訊壓縮

Yu-Lun (Alex) Liu
劉育綸

with slides by Wen-Hsiao Peng,
Shao-Yi Chien,
Hsueh-Ming Hang,
and Aggelos K. Katsaggelos

Week	Date	Topic	Assignments
1	2025-09-01		
2	2025-09-08	Introduction to Image and Video Processing	
3	2025-09-16	Signals and Systems	#1 – Color Transform, due: 2025-09-29 1:19pm
4	2025-09-22	Fourier Transform and Sampling	
5	2025-09-29	教師節補假	
6	2025-10-06	中秋節	
7	2025-10-13	Fourier Transform and Sampling	#2 – 2D-DCT, due: 2025-10-27 1:59pm
8	2025-10-20	Motion Estimation	Final project assigned (group together in fours)
9	2025-10-27	Lossless Compression	#3 – MEMC, due: 2025-11-10 1:59pm
10	2025-11-03	Image Compression	
11	2025-11-10	Video Compression	#4 – Entropy coding, due: 2025-11-24 1:59pm
12	2025-11-17	Learning-based Image/Video Compression	
13	2025-11-24	Paper Presentation	
14	2025-12-01	Guest Lecturer – MEDIATEK Meta NETFLIX	
15	2025-12-08	Guest Lecturer – MEDIATEK Meta NETFLIX	
16	2025-12-15	Final Project Presentation	

Introduction to Image and Video Processing

with slides by Wen-Hsiao Peng, Shao-Yi Chien, Hsueh-Ming Hang, and Aggelos K. Katsaggelos

Introduction to Image and Video Processing

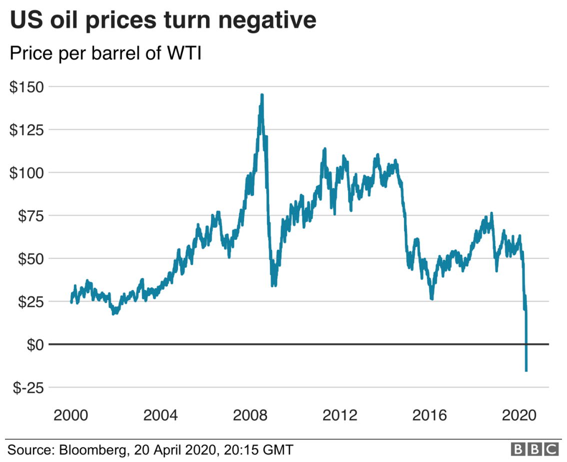
- Analog v.s. Digital Signals
- Image and Video Signals

Introduction to Image and Video Processing

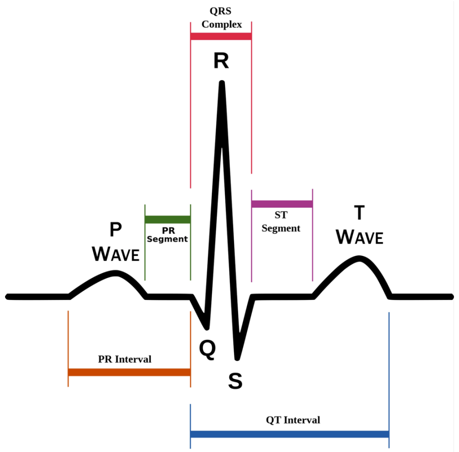
- **Analog v.s. Digital Signals**
- Image and Video Signals

Signals

- A **function** containing **information** about the behavior or the nature of some phenomenon of interest.
  - $X(t)$
  - $X(t_1, t_2)$
- In the physical world, any quantity exhibiting variation in time and/or space is potentially a signal.

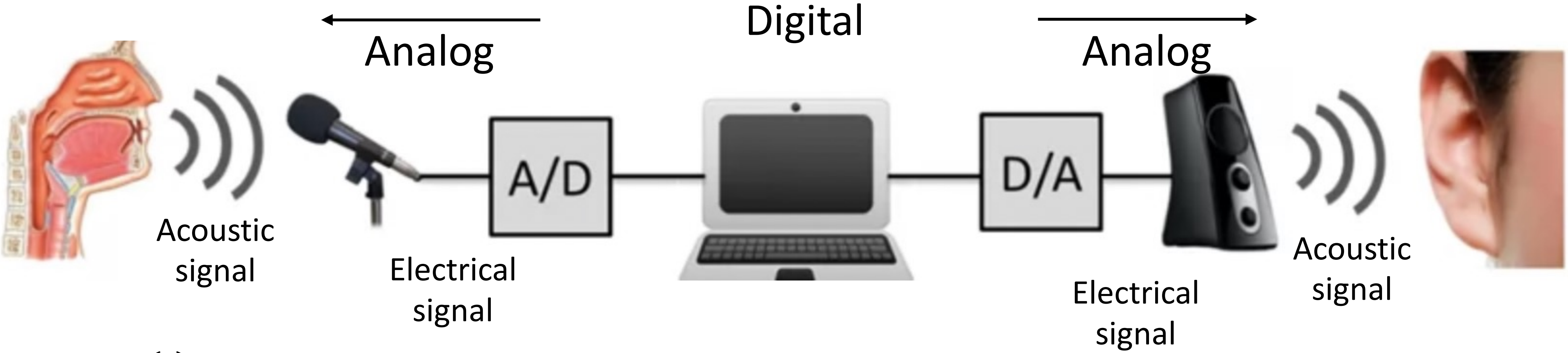


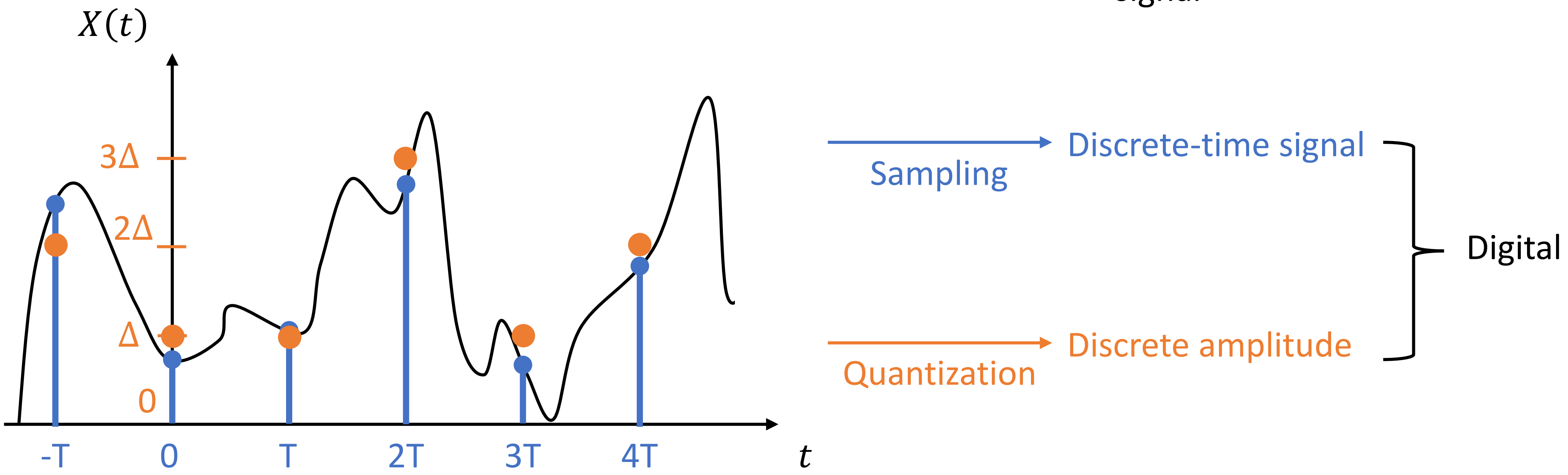
The price of the oil over the years.



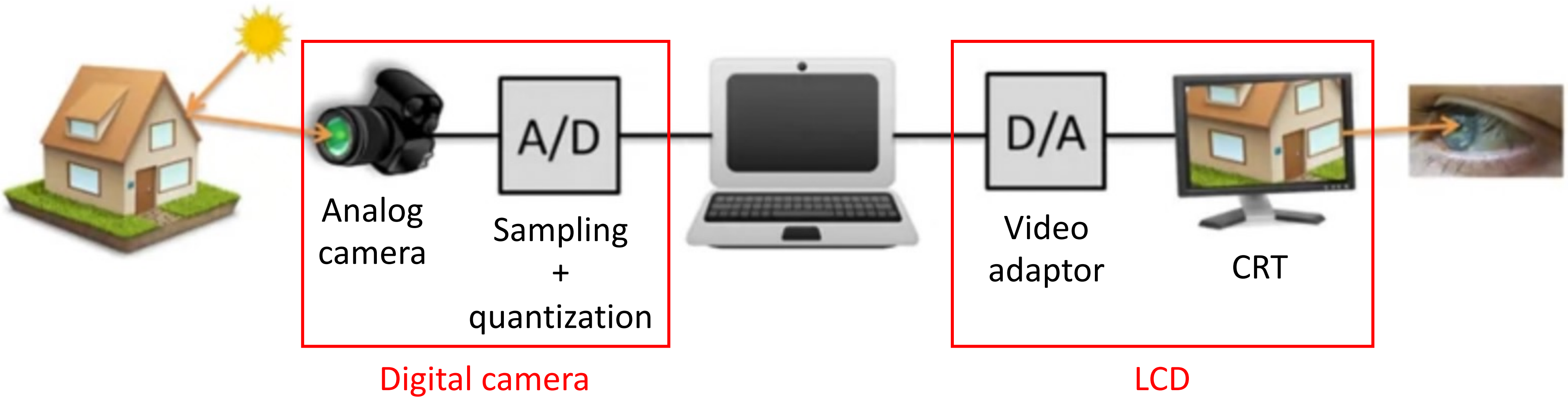
An EKG signal and electrocardiogram.

Analog vs Digital Signals

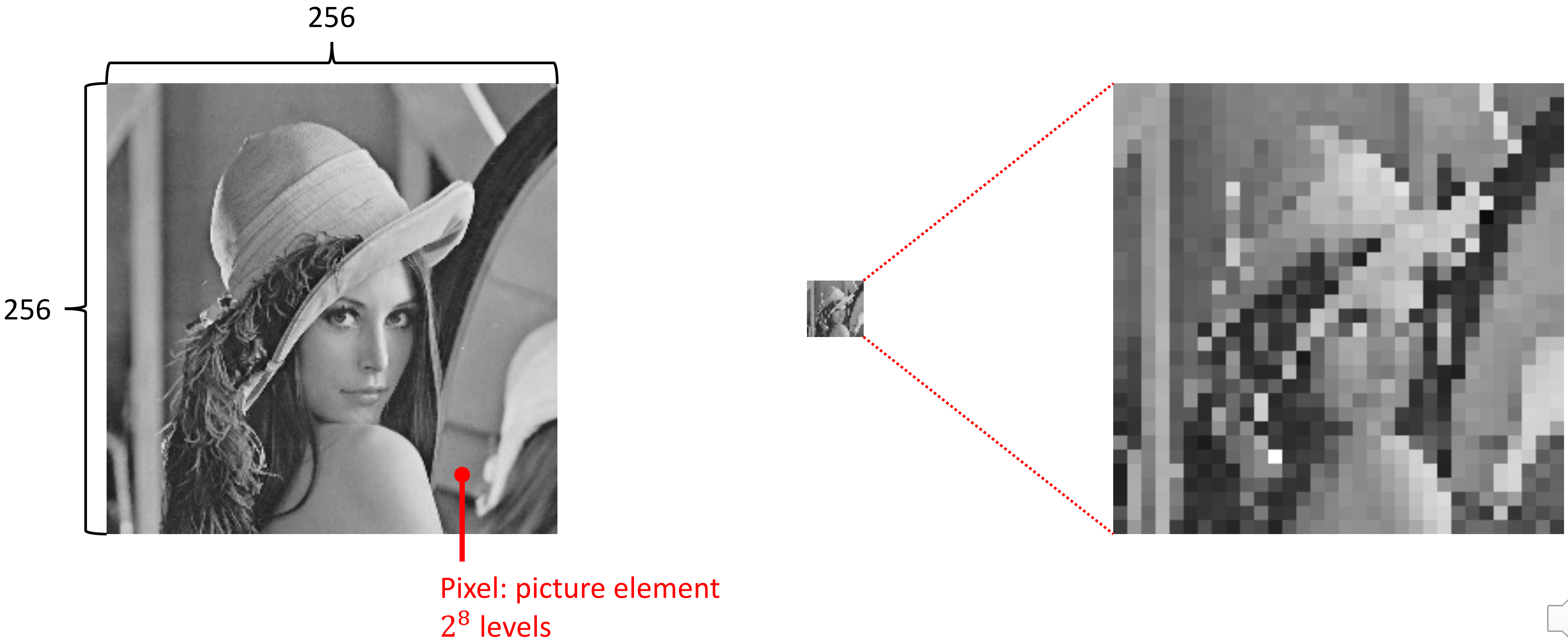




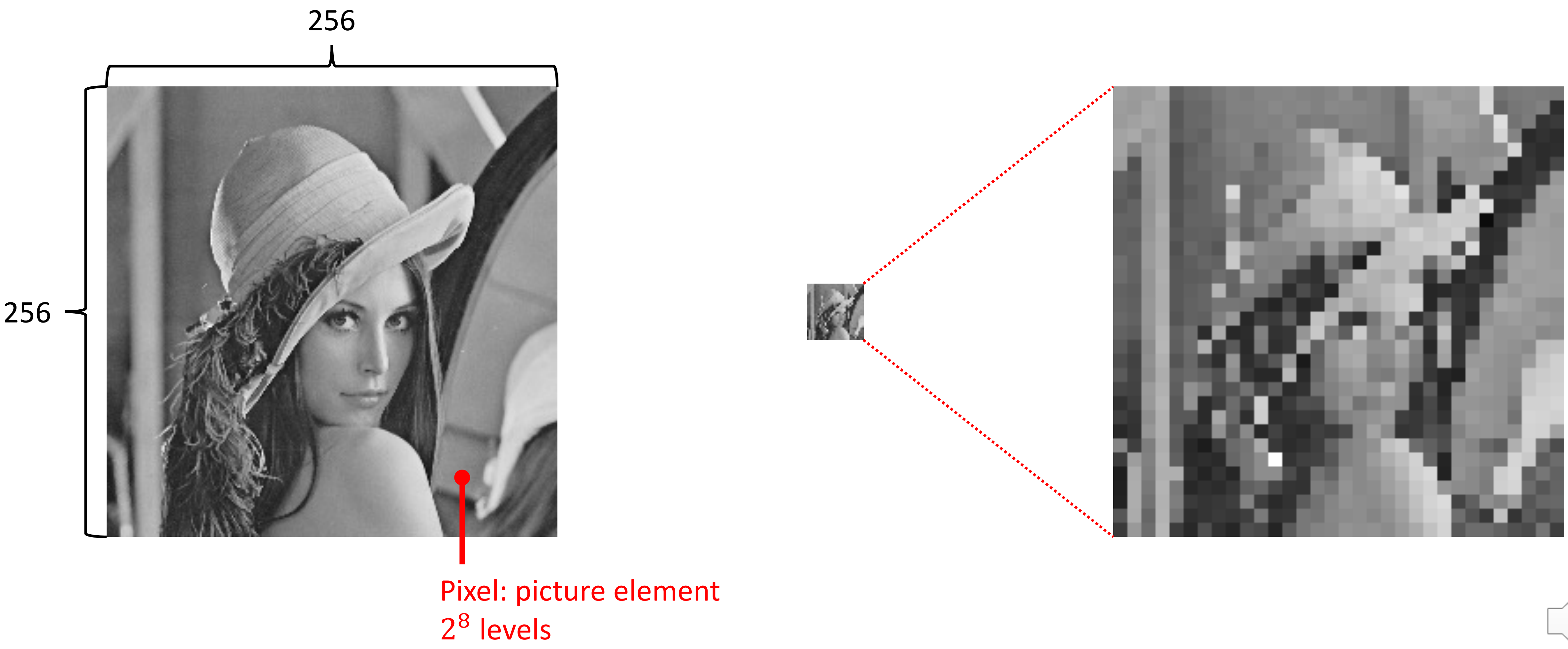
Analog vs Digital Signals



Sampling



Sampling



Quantization

8 bits per pixel
$2^8 = 256$

4 bits per pixel
$2^4 = 16$

2 bits per pixel
$2^2 = 4$

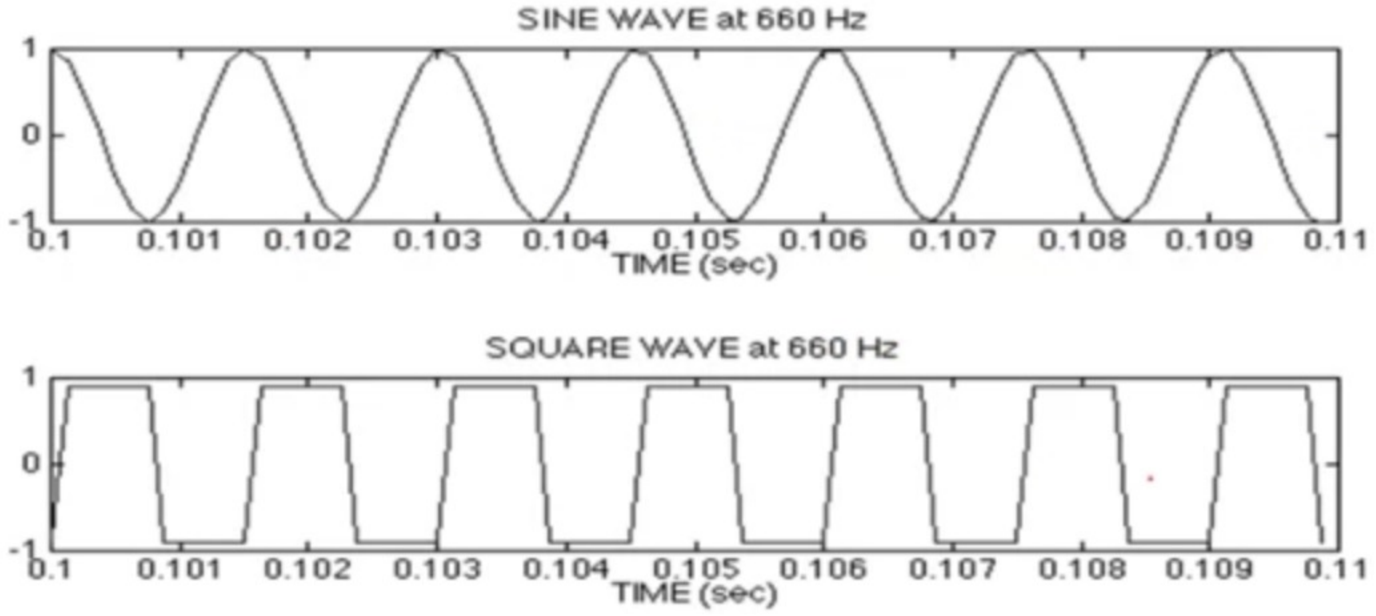
Introduction to Image and Video Processing

- Analog v.s. Digital Signals
- **Image and Video Signals**

Images and Videos

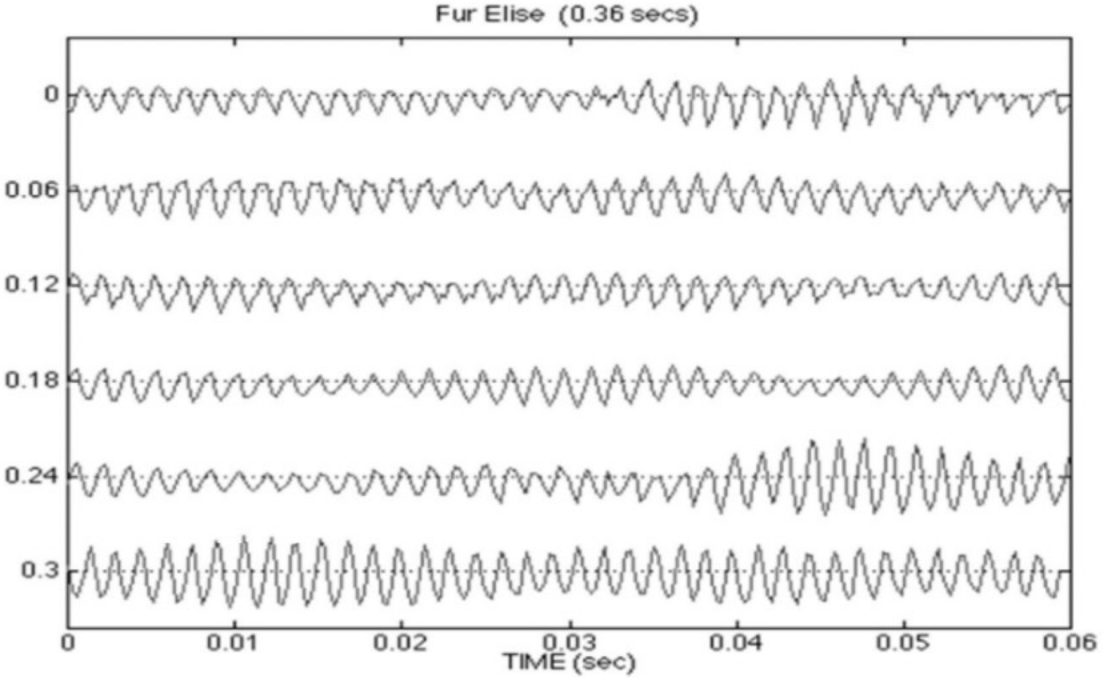
- 1D: tones, speech, audio, biomedical, remote sensing, etc.
  - $S(t)$
  - $S(n)$
- 2D: text, grayscale, color, multispectral, hyperspectral, images, etc.
  - $S(x, y)$
- 3D: video, 3D volume, etc.
  - $S(x, y, t)$
  - $S(x, y, z)$
- MD: video of a volume, etc.
  - $S(x, y, z, t)$

1D Signals: Tones



McClellan, James H., Ronald W. Schafer, and Mark A. Yoder. *Dsp first*. Pearson Education, 2017.

1D Signals: Piano Piece



McClellan, James H., Ronald W. Schafer, and Mark A. Yoder. *Dsp first*. Pearson Education, 2017.

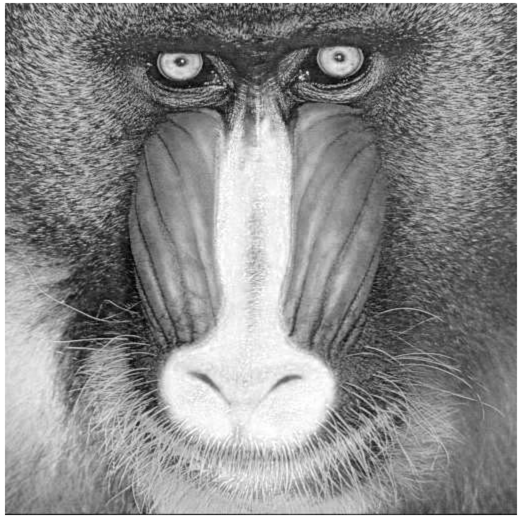
Images

1 bit per pixel

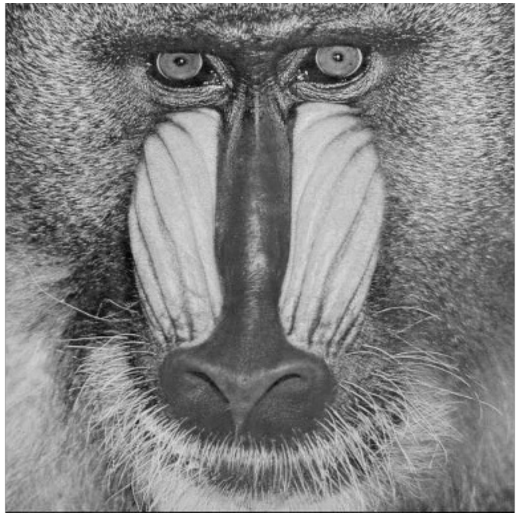
8 bits per pixel

24 bits per pixel

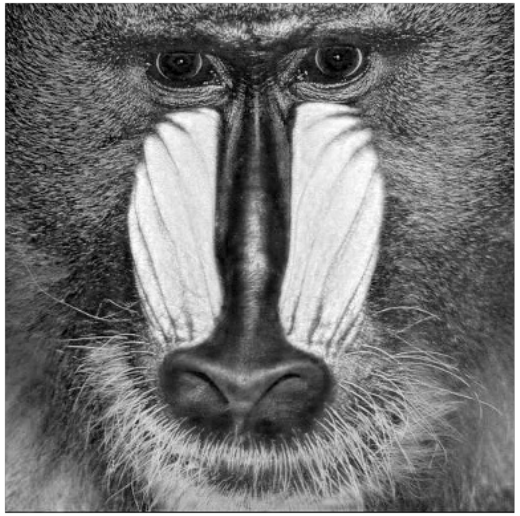
RGB Color Image



R

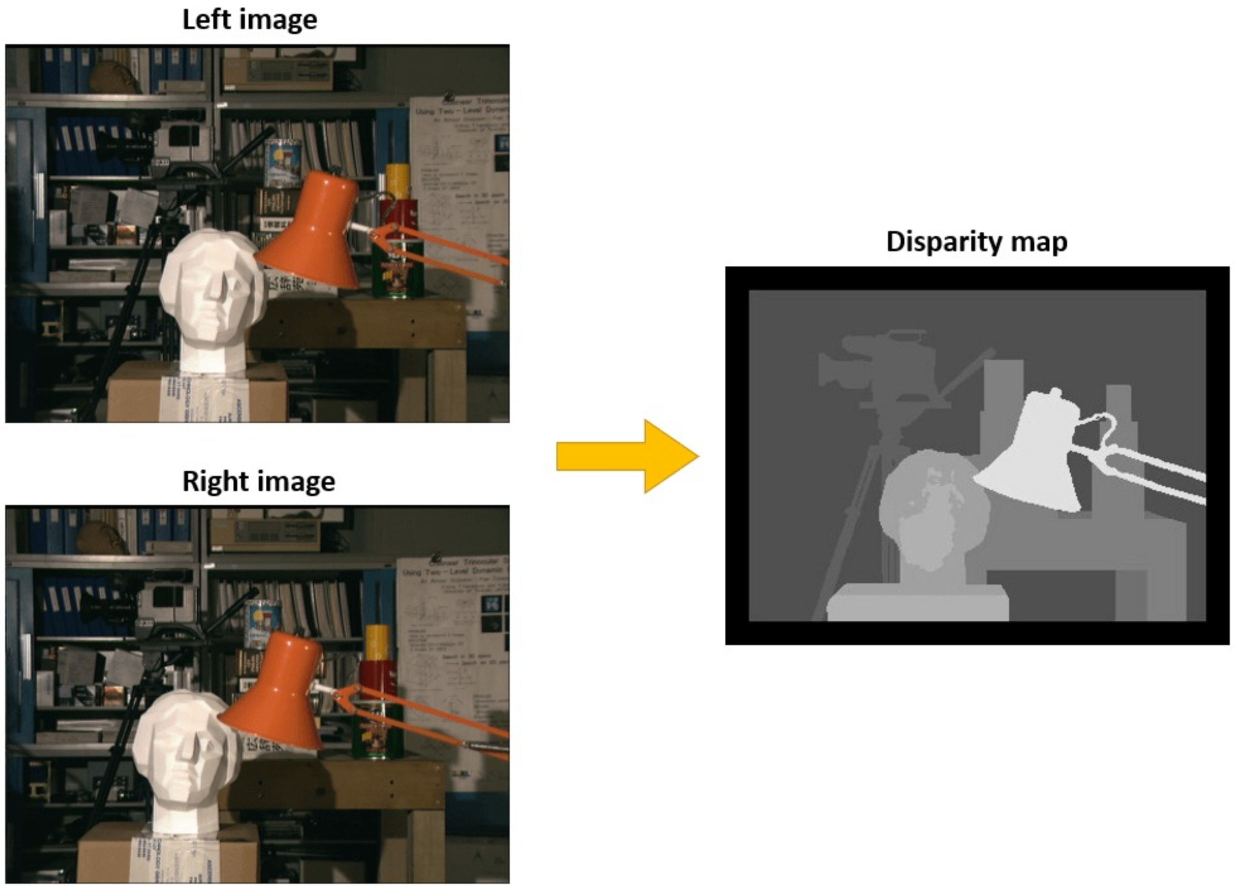


G



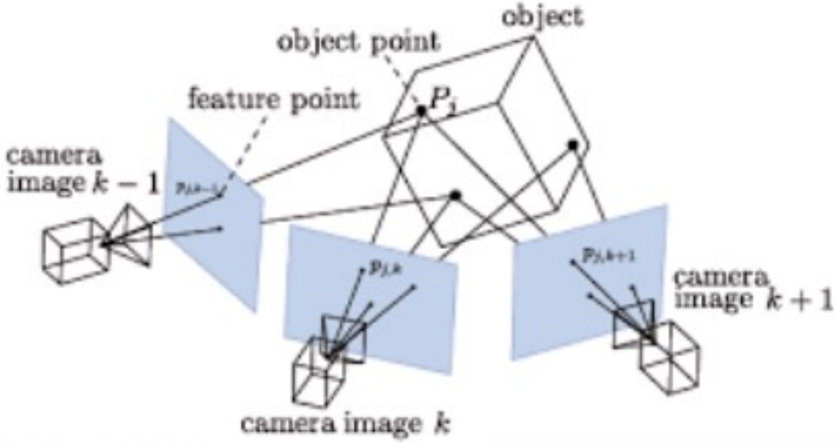
B

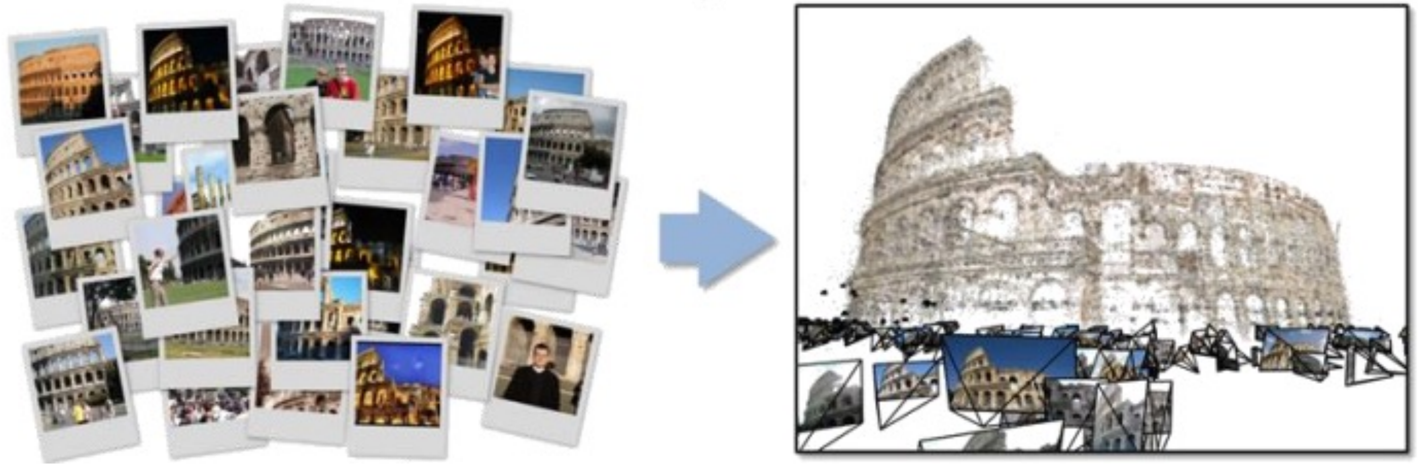
Stereo Images and Disparity



Kinect Images

Multi-Camera Imaging





Video

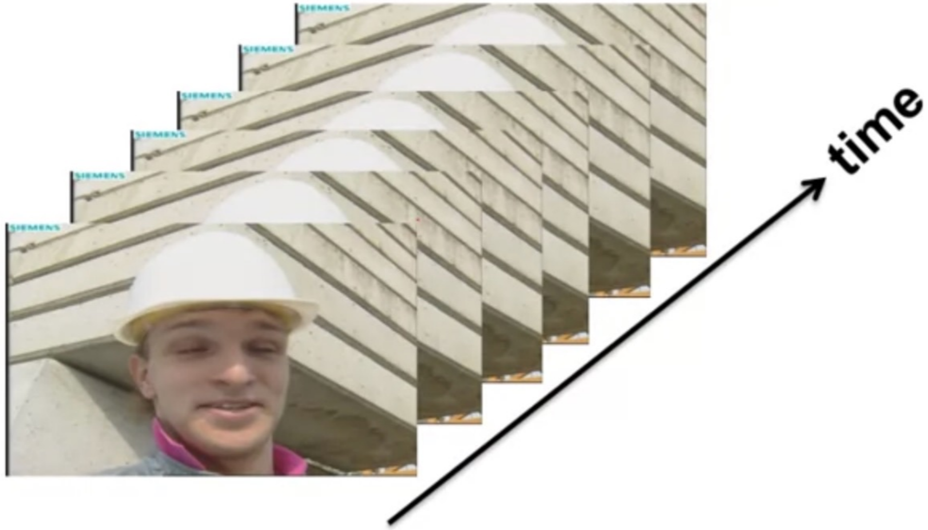
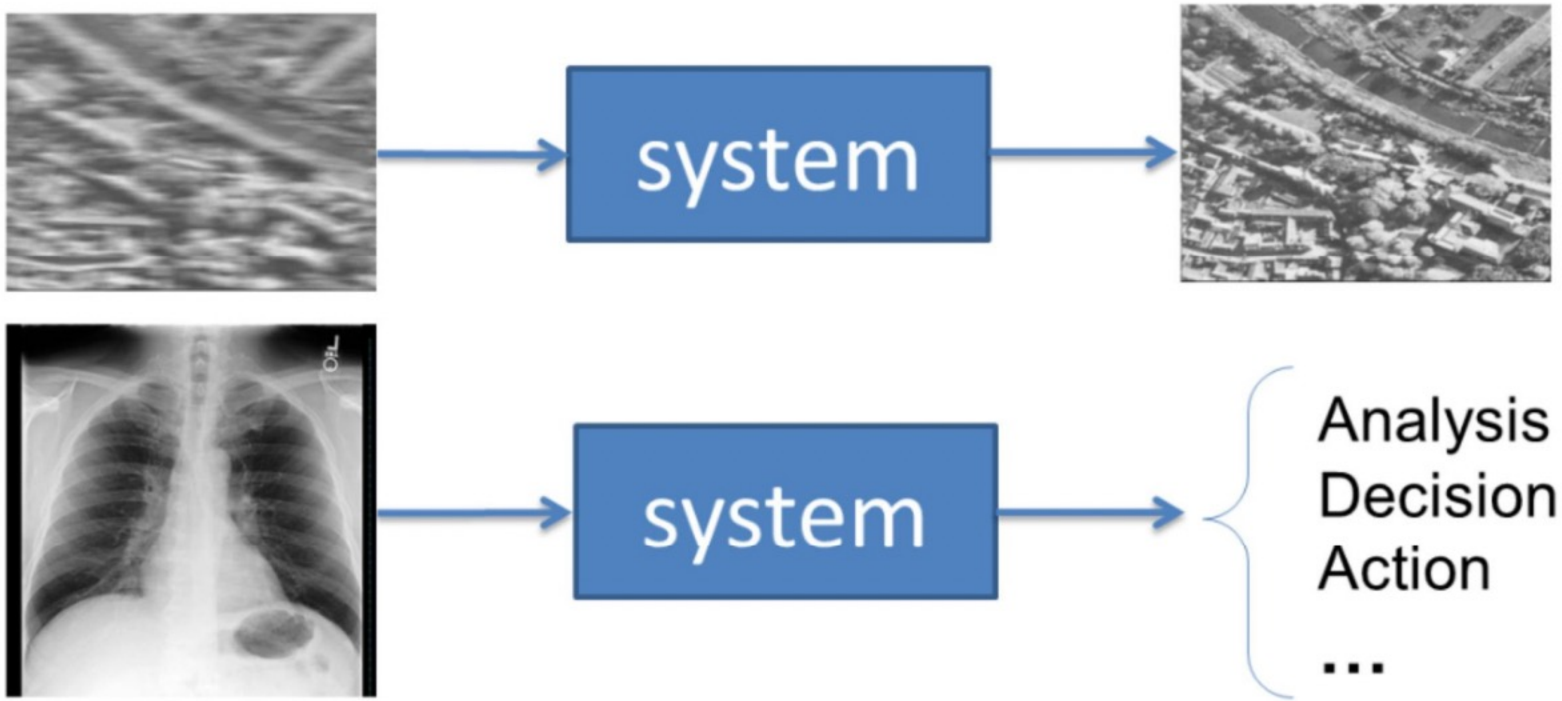


Image and Video **Processing**



Introduction to Image and Video Processing

- Analog v.s. Digital Signals
- Image and Video Signals
- **Color Space**

Video Signals

- A sequence of 2-D images captured from the projection of a 3-D scene onto an image plane

Artist Albrecht Durer’s Perspective Projection

Color Video Camera

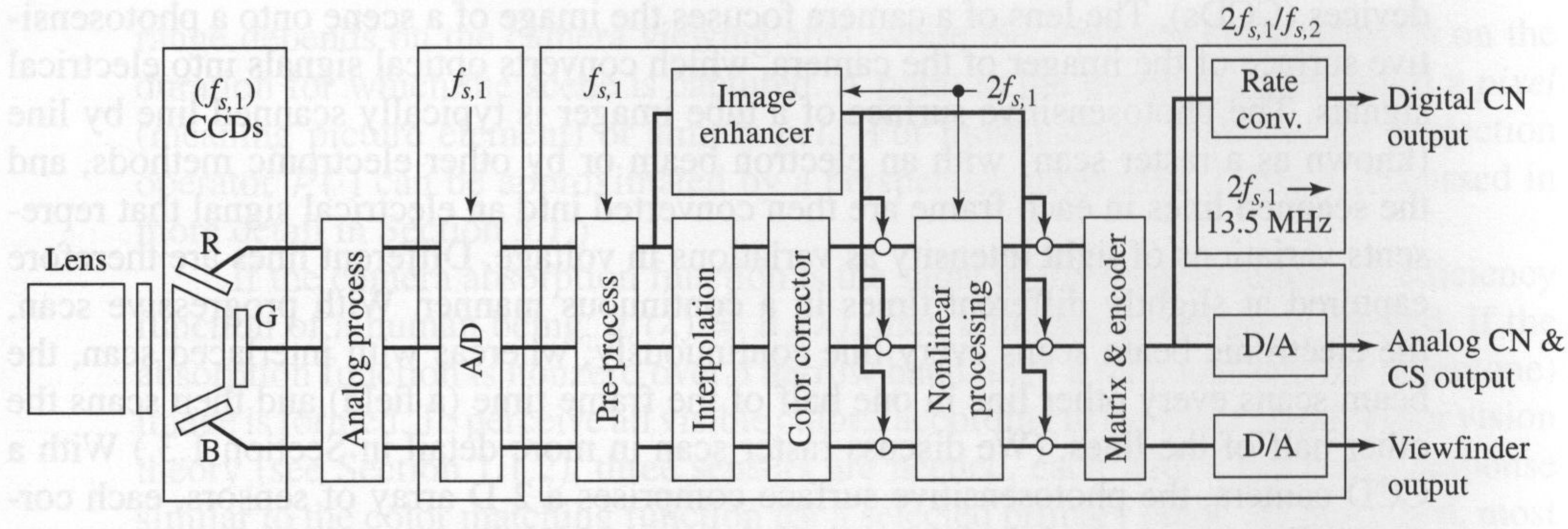
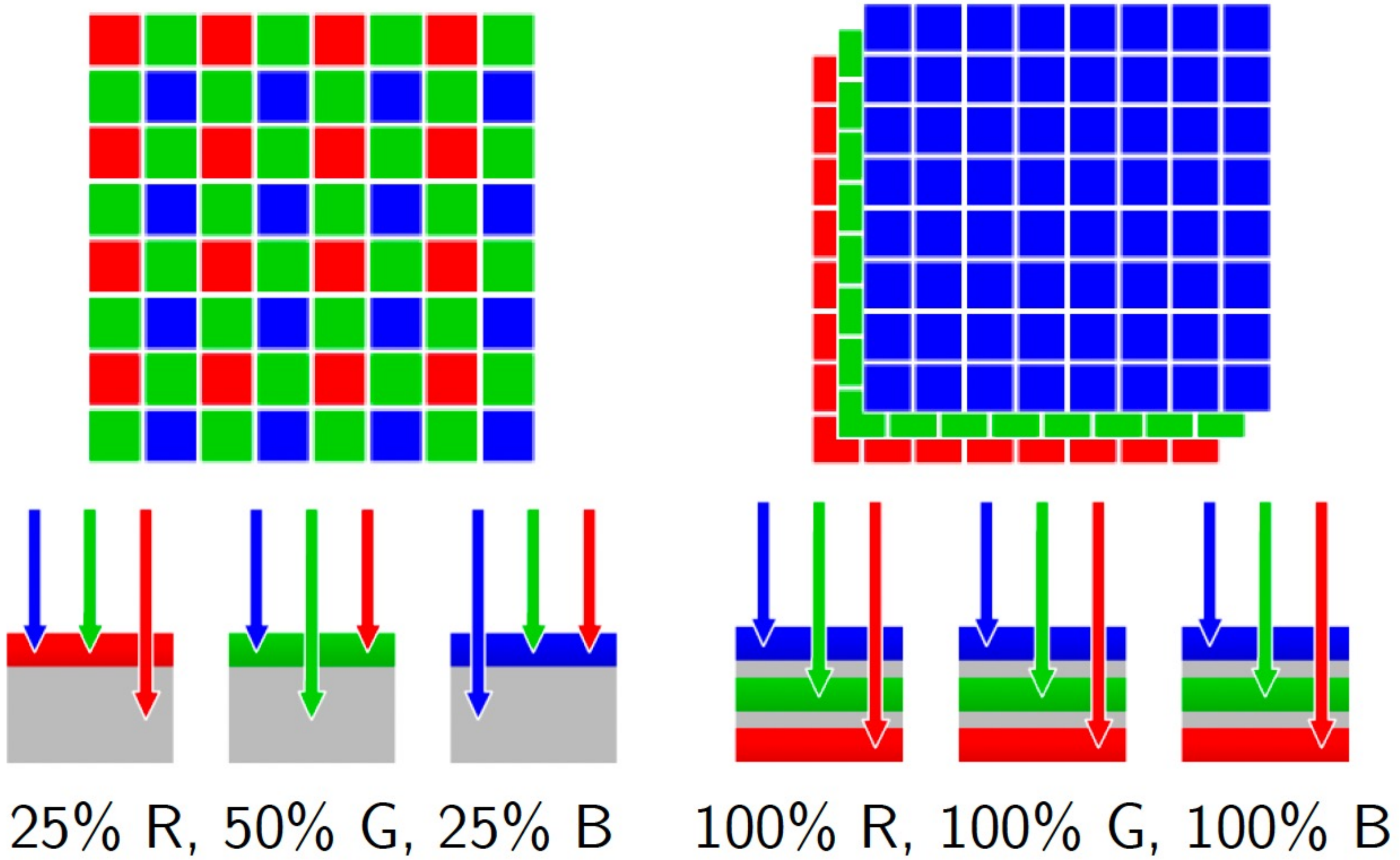


Image Sensors

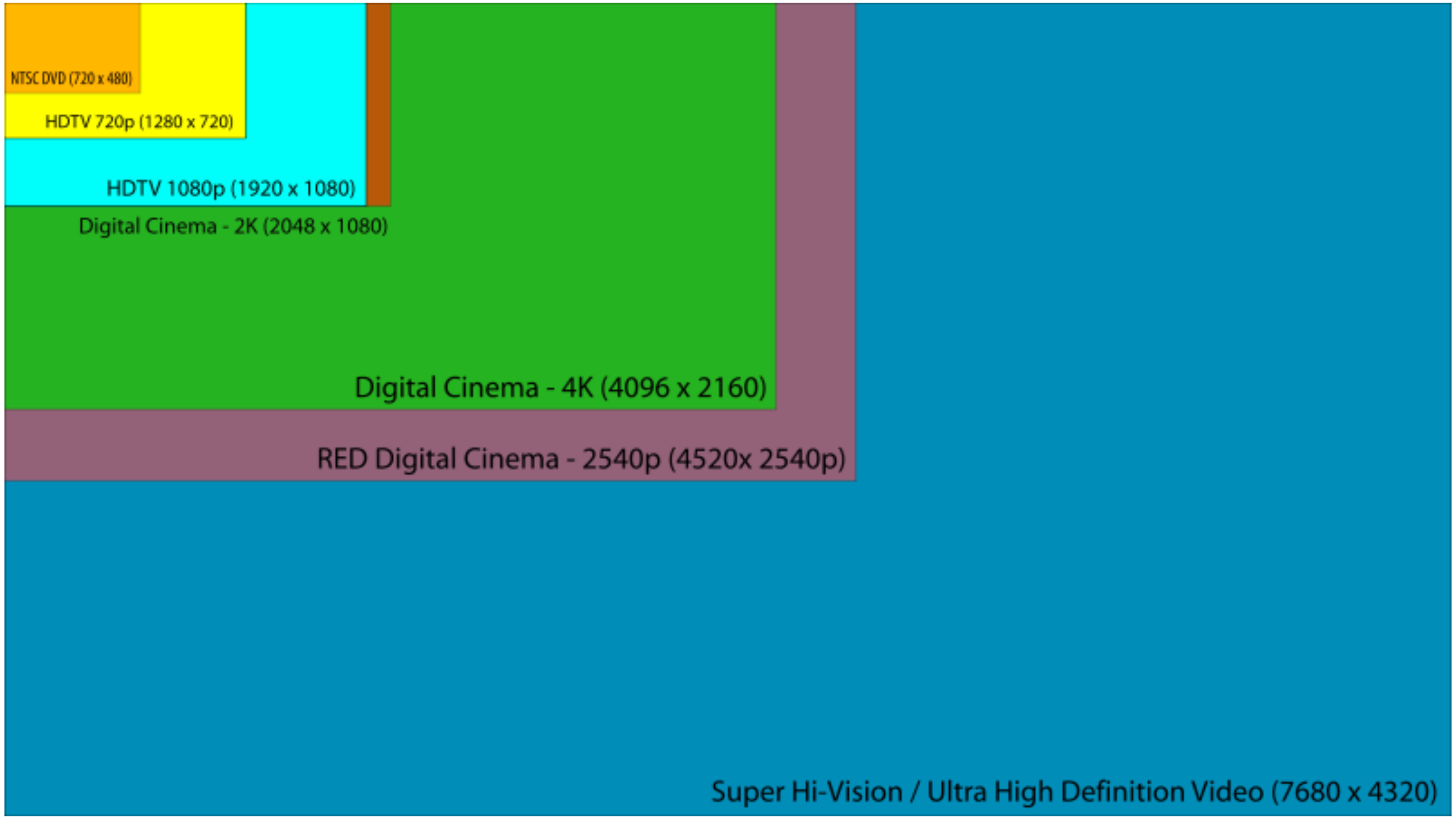
- CCD/CMOS Sensors



Digital Video Format

Video format	Y size	Color sampling	Frame rate	Raw data (mbps)
HDTV over air, cable, satellite, MPEG-2 video 20–45 mbps				
SMPTE 296M	1280 × 720	4:2:0	24P/30P/60P	265/332/664
SMPTE 295M	1920 × 1080	4:2:0	24P/30P/60I	597/746/746
Video production, MPEG-2, 15–50 mbps				
BT.601	720 × 480/576	4:4:4	60I/50I	249
BT.601	720 × 480/576	4:2:2	60I/50I	166
High-quality video distribution (DVD, SDTV), MPEG-2, 4–8 mbps				
BT.601	720 × 480/576	4:2:0	60I/50I	124
Intermediate-quality video distribution (VCD, WWW), MPEG-1, 1.5 mbps				
SIF	352 × 240/288	4:2:0	30P/25P	30
Videoconferencing over ISDN/Internet, H.261/H.263, 128–384 kbps				
CIF	352 × 288	4:2:0	30P	37
Video telephony over wired/wireless modem, H.263, 20–64 kbps				
QCIF	176 × 144	4:2:0	30P	9.1

High Definition and Ultra High Definition



Color Space

- Mixture of luminance and chrominance
  - RGB & CMY primaries
- Separation of luminance and chrominance
  - XYZ - fundamental measurements
  - YUV - PAL
  - YIQ - NTSC
  - YCbCr - digital video
  - HSI - hue, saturation, brightness
- Color space conversion – better representation for processing or data reduction

Color Space Conversion

- $(R, G, B) \rightarrow (Y, U, V)$
  - Y: brightness
  - $U = 0.492(B - Y), V = 0.877(R - Y)$: color differences

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & \boxed{0.587} & 0.114 \\ -0.147 & -0.287 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- Brightness contribution ($G > R > B$)

$$Y = 0.299R + 0.587G + 0.114B$$

- Coefficients of the first row add up to 1, those of the other rows to 0
  - $R = G = B$ -> gray images ($U = V = 0$)

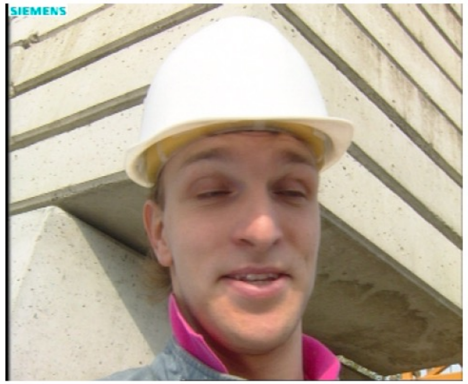
Color Space Conversion

- Y, Cb, Cr: scaled, shifted version of Y, U, V

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

  - R, G, B take values in 0-255
  - Y takes values in 16-235, Cb, Cr in 16-240

RGB vs. YCbCr



Original

R

G

B

Y

Cb

Cr

Objective Quality Measure

- Mean Squared Error (MSE)

$$MSE = \sigma_e^2 = \frac{1}{N}\sum_{k}\sum_{m,n}(\Psi_1(m,n,k) - \Psi_2(m,n,k))^2$$

  - m, n pixel coordinates; k frame index; N total number of pixels.
- Peak Signal-to-Noise Ratio (PSNR)

$$PSNR = 10\log_{10}\frac{\psi_{\max}^2}{\sigma_e^2}\ (\mathrm{dB})$$

  - $\psi_{\max}$ the maximum (or peak) signal value; 255 for 8-bit video.
- Average per-frame PSNR

$$\overline{PSNR} = \frac{1}{K}\sum_{K} PSNR_K$$

  - k frame index; K number of frame compared.

PSNR & MSE

- Excellent (>40), Good (30-40), Poor (<30)

Original

42.6

35.5

28.7

PSNR & MSE

- Both do not correlate well with human perception

MSE = 27.10

MSE = 21.26

Subjective Quality Evaluation

- Mean opinion score
  - Human observers’ judgement of visual quality
- Methods (defined in ITU-R BT.500-11)
  - Double Stimulus Continuous Quality Scale (DSCQS)
    - Reference and impaired sequences graded in a randomized order.
    - Scores converted into a number indicating the relative quality.
  - Double Stimulus Impairment Scale (DSIS)
    - Reference first and then the impaired sequence.
    - Grade the impaired sequence relative to its reference.
  - Single Stimulus Continuous Quality Evaluation (SSCQE)
    - Grade the impaired sequence in a continuous manner without a reference.

Perceptual Similarity Metric

- The Unreasonable Effectiveness of Deep Features as a Perceptual Metric [CVPR 2018]