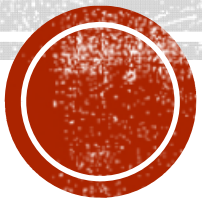
AY 2025/2026 Semester 1

EE6427 Video Signal Processing

Part 1 Introduction and Basics



Dr Yap Kim Hui

Room: S2-B2b-53
Tel: 6790 4339
Email: ekhyap@ntu.edu.sg

Course Outline

- Part 1: Introduction & Basics
- Part 2: Image Compression & JPEG Standard
- Part 3: Video Compression & Standards
- Part 4: AI Models & Architectures
- Part 5: Video Analysis and Understanding
- Part 6: Emerging Topics in Image / Video Processing

References

- Ze-Nian Li, Mark S. Drew, Jiangchuan Liu, Fundamentals of Multimedia, Springer, 2022.
- R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4nd edition, Pearson, 2018.
- Saeed Vaseghi, Multimedia Signal Processing, Wiley, 2007.
- Fred Halsall, Multimedia communications: applications, networks, protocols and standards, Addison-Wesley, 2001.
- Yun Q. Shi and Huifang Sun, Image and video compression for multimedia engineering: fundamentals, algorithms, and standards, CRC Press, 2000.
- Tim Daly, “A Beginner's Guide to Digital Photography,” Quarto Publishing, 2002.
- Julie Adair King, “Digital Photography for Dummies, ” Wiley Publishing, 2005

Copyright Issues

- Some materials in the slides are copyrighted by respective publishers and authors.
- The lecture slides should be used for your own personal studies only.
- No part of these slides may be recorded, reproduced or distributed in any form or any means, without the prior written permission of the author.

Assessment

- Examination: 60%
  - 4 Questions
  - Closed-book

- Continuous Assessments (CAs): 40%
  - CA 1: Homework Assignment (20%)
  - CA 2: Quiz (20%)

Part 1: Introduction & Basics

Part 1 Outline

- Image & Video Basics
- Video Signal Processing & Applications
- Image & Video Standards
- Video Analysis & Understanding Basics
- Emerging Topics

Image & Video Basics

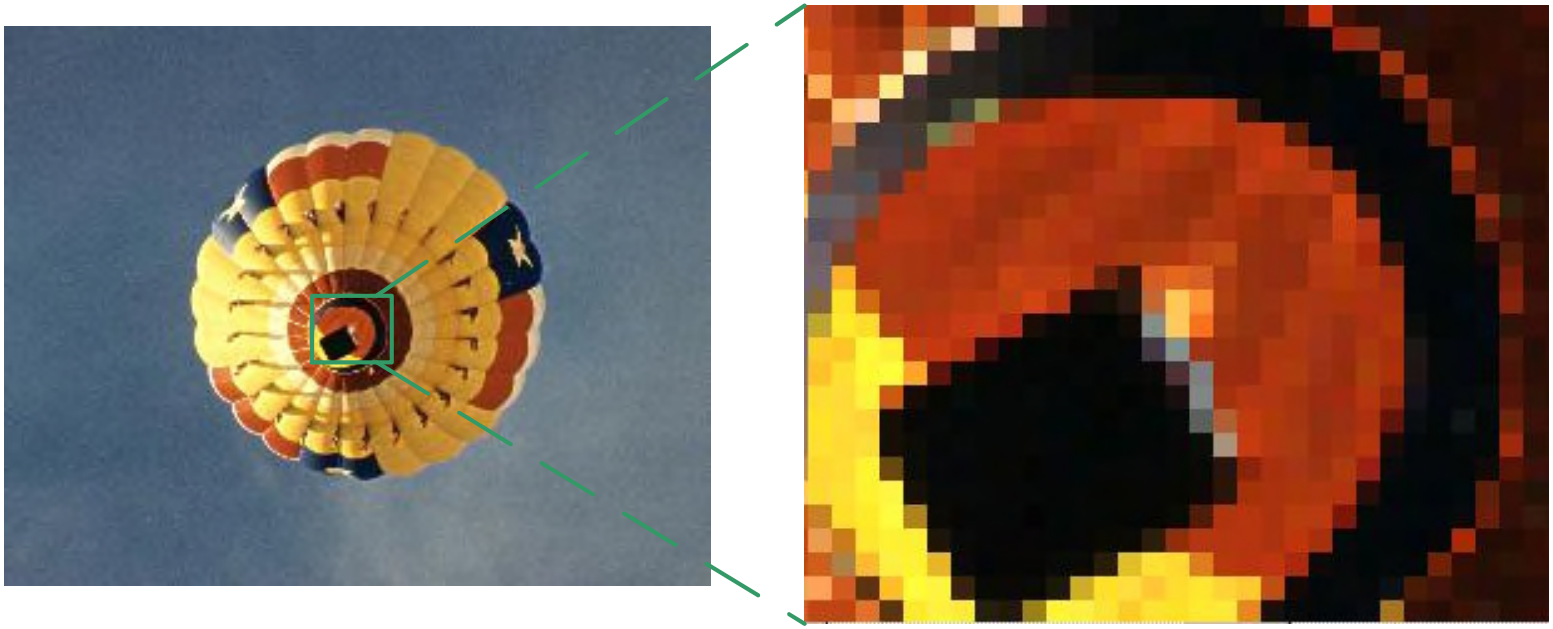
What is an Image?

- An important media type in information storing, processing & presentation.
- Represented by a 2D matrix of rows and columns of pixels.
- Some popular image formats:
  - JPEG (Joint Photographic Expert Group), TIFF (Tagged Image File Format), BMP (Bitmap), etc.

Source: https://pixabay.com/photos/tree-sunset-clouds-sky-silhouette-736885/

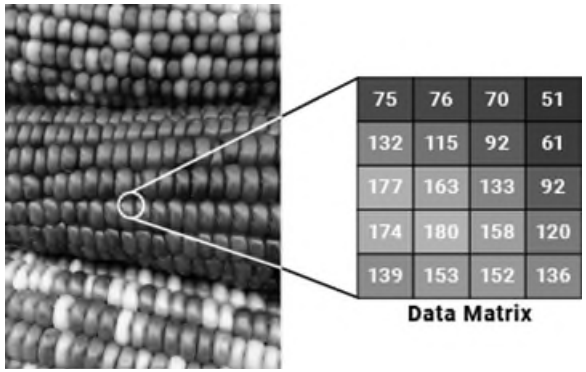
Pixel and Resolution

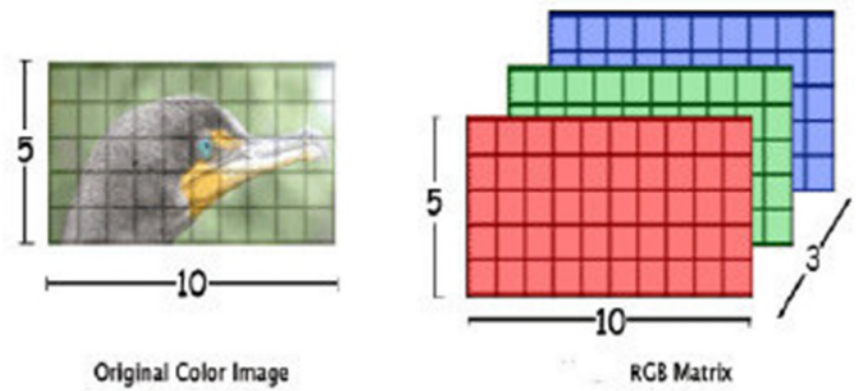
- Pixel (picture element): smallest element of digital images
- Resolution: number of pixels in an image
  - Expressed as total pixel count, width x height or pixels per inch (ppi)
  - E.g., 5 megapixels, 640 x 480 pixels, or 300 ppi
  - Measure the ability to distinguish fine details of images



Bit Depth/Color Depth

- Bit depth/ color depth: number of bits for each pixel.
- Binary images (1 bit), grayscale images (8 bits) and color images (24 bits).

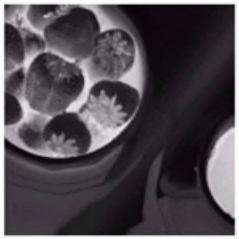




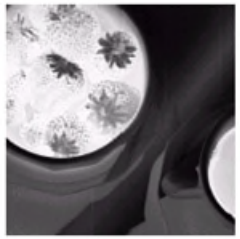
Source: https://www.researchgate.net/figure/Color-image-and-RGB-matrices_fig1_339435251
https://www.mathworks.com/help/matlab/creating_plots/image-types.html

Color Spaces

Full color

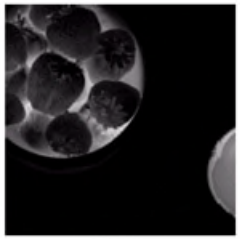


Cyan



Magenta

Yellow

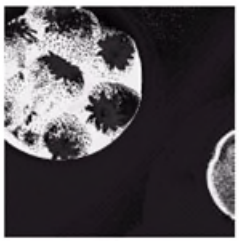


Black

Red

Green

Blue



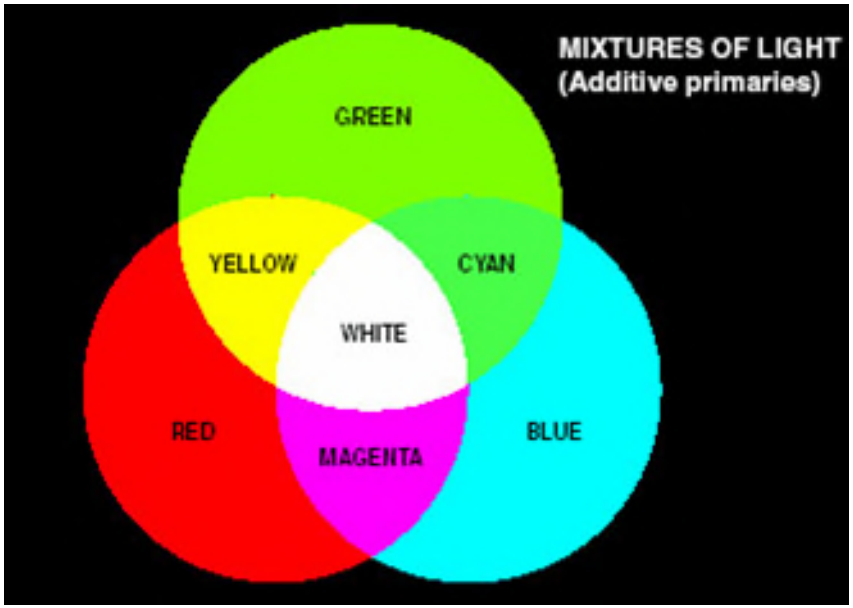
Hue

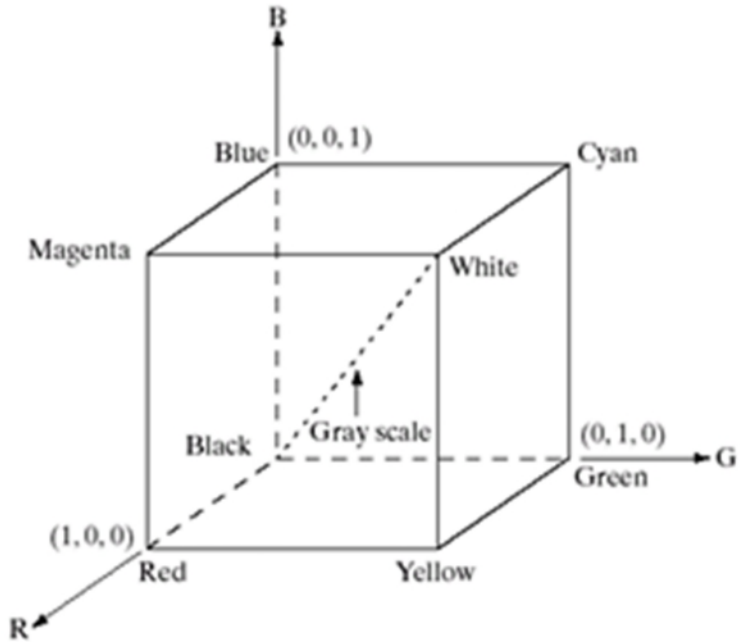
Saturation

Intensity

Color Images

- Color pictures can be expressed in different color spaces.
- Common color space: RGB, HSI, YCbCr, etc.
- RGB: color represented by triplet of red (R) , green (G), and blue (B)





RGB

- 24-bit color image:
  - each color pixel has three components – RGB (Red, Green, Blue)
  - 8 bits for each component, ranging from 0 to 255.
- Black is (0, 0, 0) and White is (255, 255, 255)

Red

Green

Blue

YUV, YIQ, YC_bC_r

- Variation of YUV-like models: YIQ and YC_bC_r.
- Y (luminance) is the brightness (black-and-white part) of video signal, UV (chrominance) is the colour part of video.
- YUV is suitable for video broadcast since it makes efficient use of bandwidth.
- Used in TV industry:
  - YUV (PAL, SECAM), YIQ (NTSC), YC_bC_r (JPEG, MPEG)

RGB -> YIQ

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

YIQ -> RGB

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.0 & 0.956 & 0.620 \\ 1.0 & -0.272 & -0.647 \\ 1.0 & -1.108 & 1.700 \end{bmatrix} \begin{bmatrix} Y \\ I \\ Q \end{bmatrix}$$

A Snapshot into History

BUSINESS | JANUARY 6, 2012

Their Kodak Moments

Career Snapshots: Workers Take a Look Back as Many Struggle With Uncertainties, Finances, Retiree Benefits

Article | Video | Stock Quotes | Comments (80)

Email | Print | Save | +1 11 | Tweet 286 | A A

By DANA MATTIOLI

Former Kodak workers look back with nostalgia on a company that once had a monopoly on film, where job security was once a guarantee and many employees met their spouses. Dana Mattioli has details on The News Hub. Photo: AP

At its peak, Eastman Kodak Co. employed 145,000 people worldwide. The company drew engineers, Ph.D.'s and scientists from around the world to its Rochester, N.Y., headquarters, talented professionals who clamored to work for the imaging pioneer.

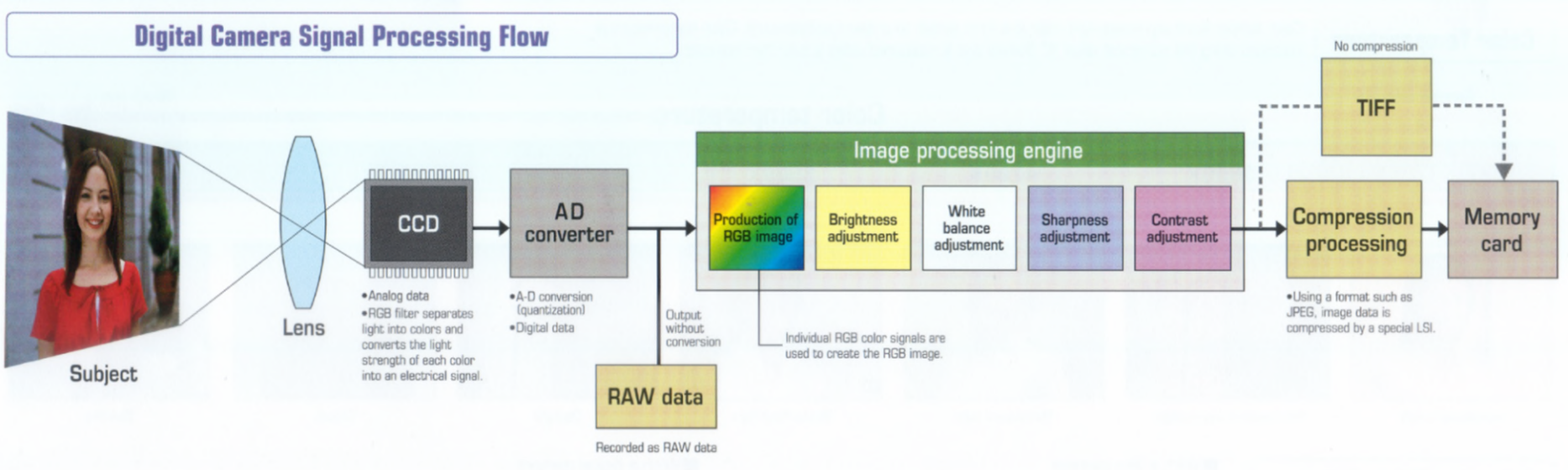
Now, as the 131-year-old former blue chip prepares for a possible bankruptcy protection filing, retirees' enthusiasm has been replaced by concerns about the fate of their benefits. Former workers look back with nostalgia on a company that once dominated the film business, where jobs felt secure, and where it wasn't uncommon for employees to meet their spouses.

Camera: Data acquisition

How Digital Camera Works?

Source: deconstructed digital camera
http://www.youtube.com/watch?v=nSF3xdIWrPI

Digital Camera Signal Processing



Elements Affecting Digital Image Quality

A digital camera's image quality is determined by three elements: lens performance, CCD image sensor pixel count and performance, and image processing engine performance.

Digital image quality is determined by:

Lens + CCD + Image processing engine

Other Camera Sensors

- RGB-D Cameras
- Near Infrared Cameras
- Thermal Cameras
- Event-based Cameras.
- Etc.

Figure sources:
https://www.mapir.camera/en-gb/products/survey3n-camera-near-infrared-nir
https://www.cetm.com.sg/home/1708-acision-htc7630-handheld-thermal-camera.html
https://www.prophesee.ai/2022/03/17/what-is-an-event-camera/

What is a Video?

- Video is a sequence of frames / images / pictures, with audio signals.
- Presented at a high speed that gives perception of smooth motion.
- Rely on human's persistence of vision.
- Some popular video formats:
  - MPEG (Moving Pictures Expert Group), AVI (Intel), WMV (Microsoft)

Digital Video Organization (1)

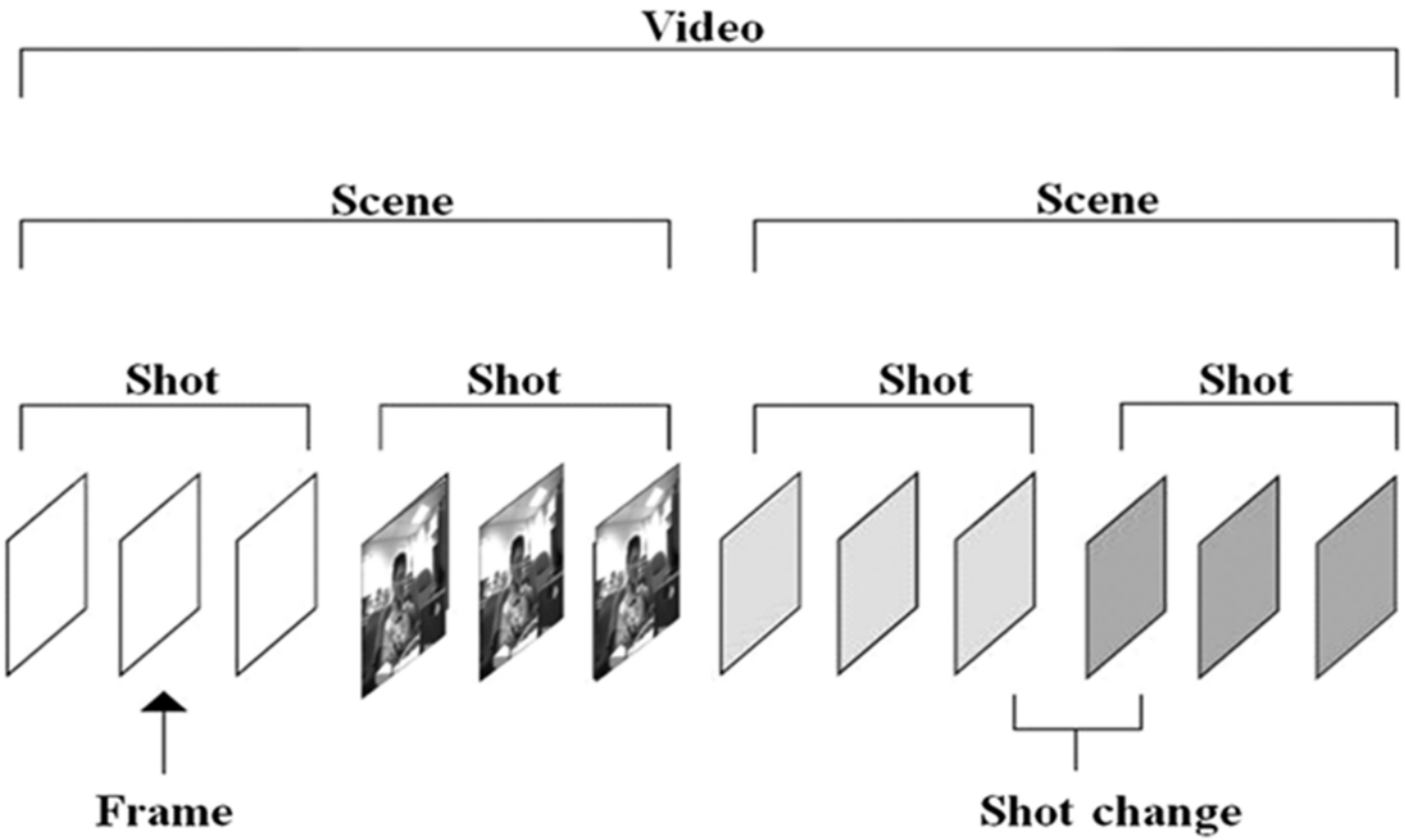
- Shot: all video is made up of shots
- Frame: each single picture in the video
- Composition: layout of content in a frame

VWS (Very Wide Shot)
Shows the subject's environment

WS (Wide Shot)
Shows the whole subject

CU (Close Up)
Shows a feature of the subject

Digital Video Organization (2)



Some Terms in Videos

- Frame rate: number of frames per second.
- Spatial resolution: number of pixels in each frame.
- Aspect ratio: ratio of image width to image height.
- Color resolution: bit depth of frames.
- Video data rate & file size.

e.g., 24-bit color video, with 640 by 480 resolution at 30 fps

Uncompressed video rate

= 640 x 480 x 3 x 8 x 30

= 221, 184, 000 bps

Video Resolution

- Standard definition (SD)
  - NTSC (480i): 720 x 480, 640 x 480
  - PAL (576i): 720 x 576
- High definition (HD)
  - 720p: 1280 x 720 progressive
  - 1080i: 1920 x 1080 interlaced
  - 1080p: 1920 x 1080 progressive (Full HD / FHD)
- Ultra high definition (UHD)
  - 4K ~ 4 times FHD
  - 8K ~ 4 times 4K

Video Resolution

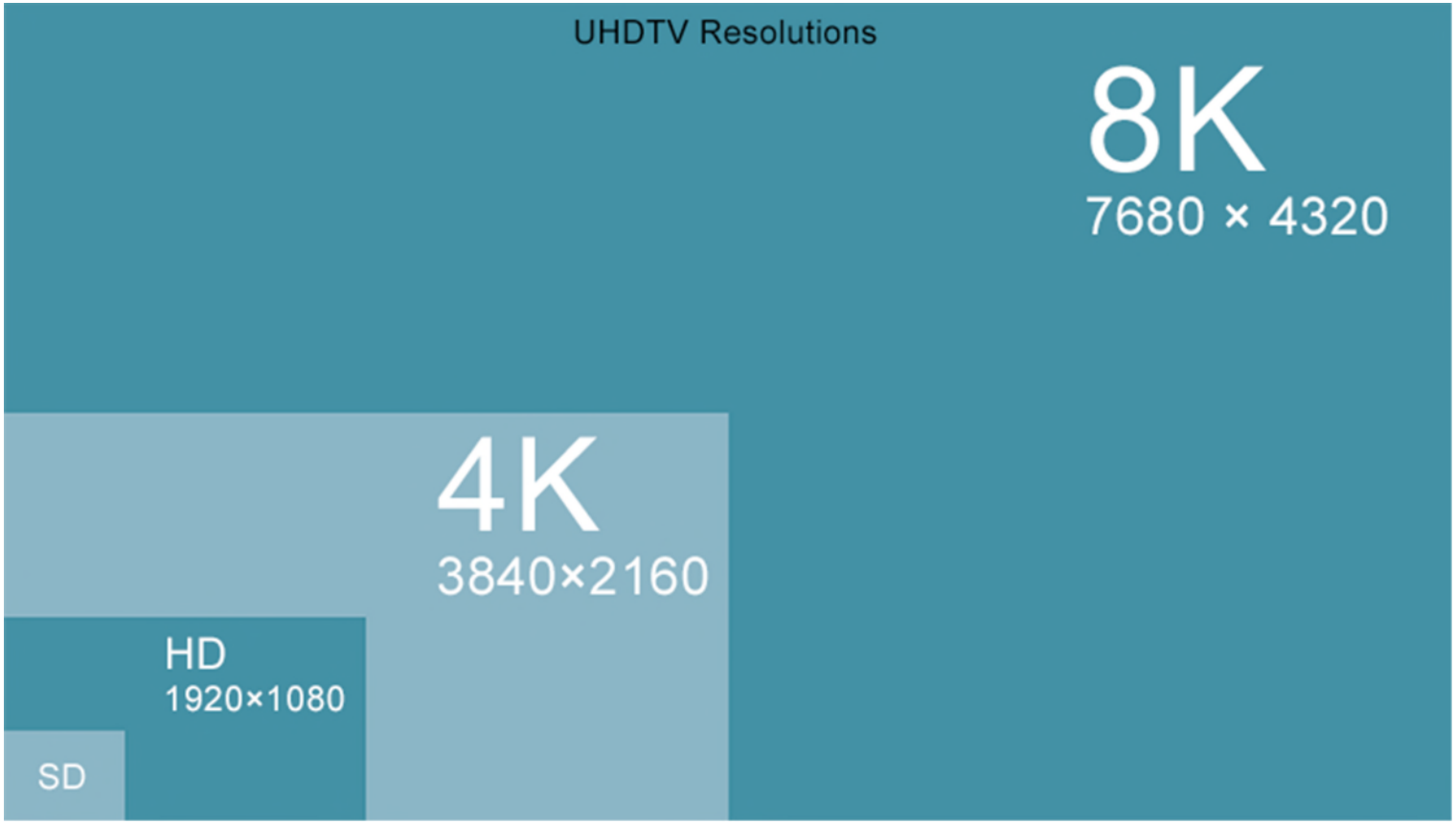
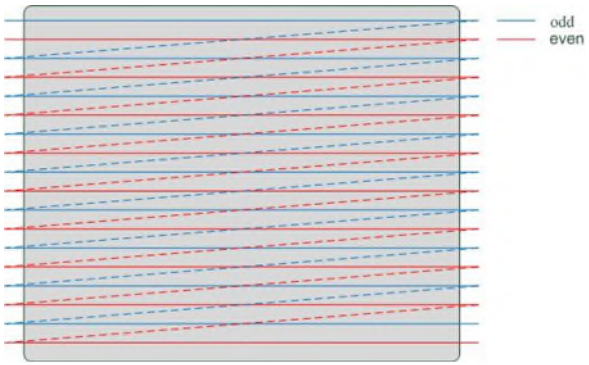
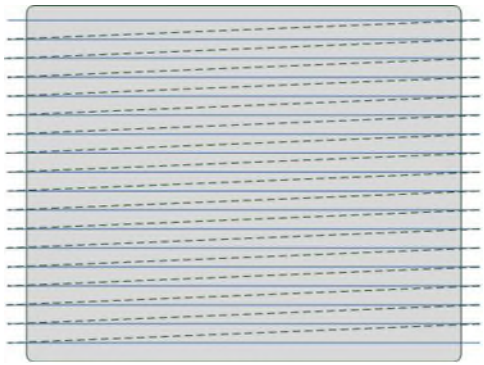


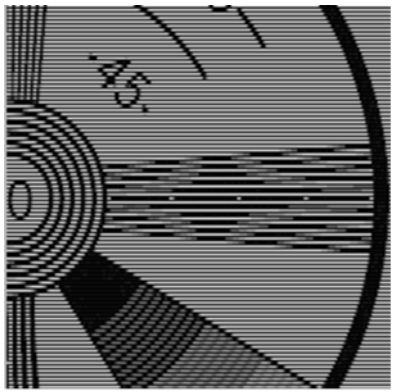
Figure source: https://www.mediacollege.com/video/resolution/

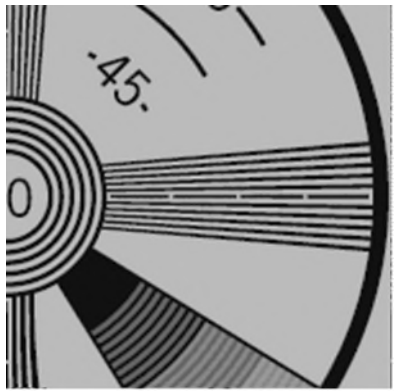
Scan Type

- Interlaced: 2 fields (odd and even lines) are scanned alternately.
- Progressive: all the lines are scanned in sequence.









Frame Rate

- How often a frame is played per second, or frame per second (fps).
- Standards:
  - PAL: 25 fps, used in Australia, Singapore
  - NTSC: 30 fps, used in USA, Japan

Video Signal Processing & Applications

What is Video Signal Processing?

- Video Signal Processing deals with different aspects of digital video signal including presentation, compression, storage, transmission, processing, analysis and understanding.
- Two key focuses of this course:
  - Image & Video Compression
  - Video Analysis & Understanding

Video Signal Processing Applications

- Business
  - Streaming e-commerce, etc.
- Entertainment
  - Video streaming, video on demand, etc.
- Education
  - Technology enabled learning, online lectures, etc.
- Smart city, transportation, etc.
  - Autonomous vehicles, video analytics, etc.

Live-streaming E-commerce

Video on Demand

Video Analytics

Source: AxxonSoft AI Video Analytics

Video Signal Processing Trends

- New AI and deep learning algorithm.
- Big data.
- Faster processors, larger-capacity storage devices.
- Emerging consumer behaviors.
- Better compression algorithms.
- Stronger video analytic techniques.

Technical Challenges

- Coding & Communications
  - Higher quality.
  - Lower bitrate.
  - Quality of service.
- Content Analysis and Understanding
  - Robust performance.
  - Domain gap / shift / generalization.
  - Expensive data annotation.
  - Etc.

Image & Video Standards

Video Standards

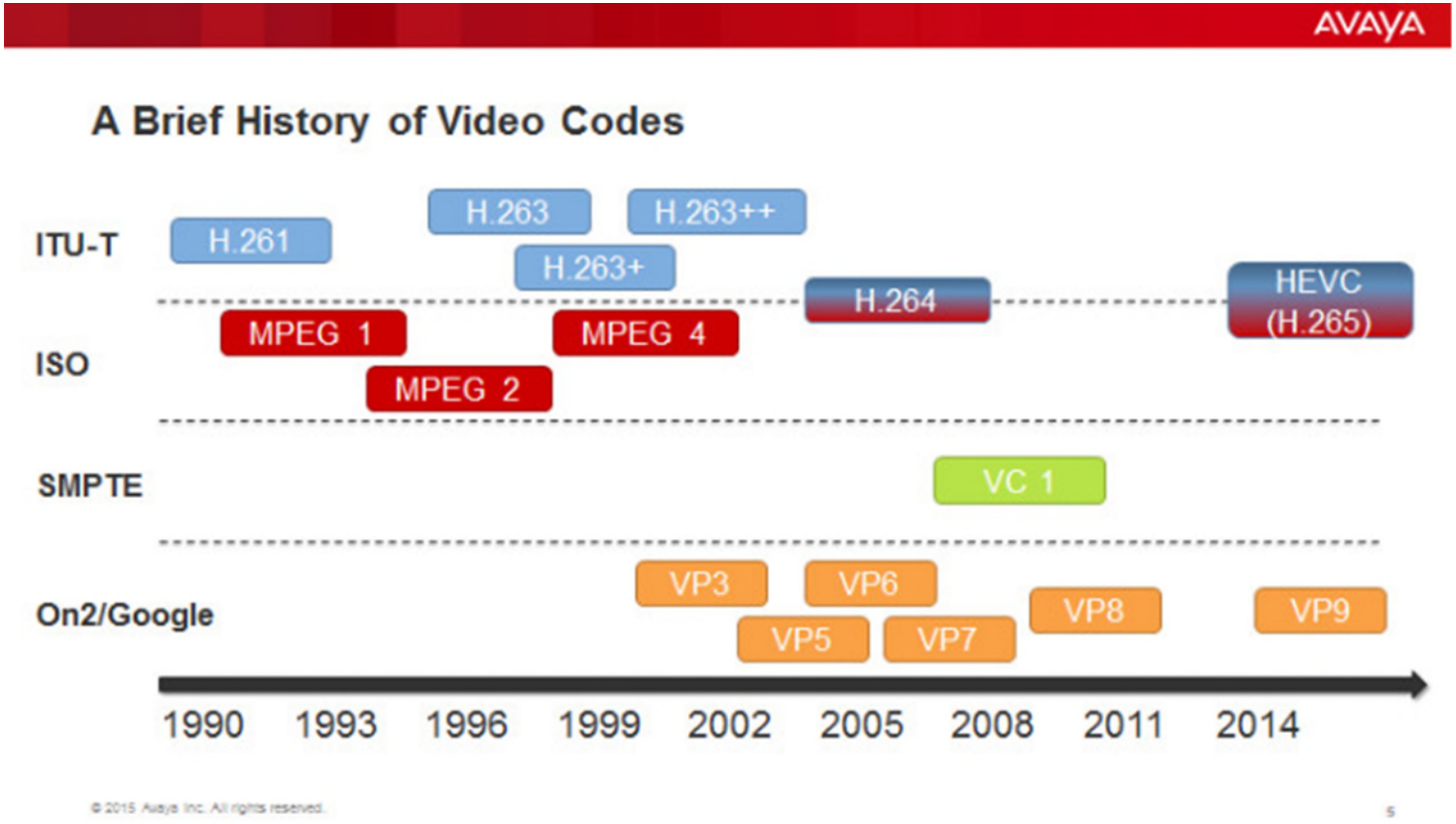


Figure source: https://www.linkedin.com/pulse/brief-history-video-codecs-yoav-nativ/

JPEG

- A popular image compression standard.
- JPEG is known as ISO/IEC international standard 10918 or the ITU-T Recommendation T.81.
- Developed by an international body known as Joint Photographic Expert Group (JPEG).

MPEG Overview

- MPEG-1 (1992)
  - Video and audio coding (CD-ROM, 1.5Mbps)
  - Relevant product: VCD
- MPEG-2 / H.262 (1994)
  - Video and audio coding with different profiles (2-80Mbps)
  - Relevant product: DVD
- MPEG-4 Advanced Video Coding (AVC) / H.264
  - Content-based video coding
  - Coding of natural and synthetic media objects

H.26x Overview

- H.261
  - Developed by CCITT in 1990. (Note: International Telegraph and Telephone Consultative Committee (CCITT) was later renamed as International Telecommunication Union International Telecommunications Standardization Sector (ITU-T)).
  - DCT-based video compression scheme.
  - Many similar features with MPEG-1 video coding.
  - Target application: videoconferencing.
  - Bit-rate is *p* x 64 Kbps, where *p* ranges from 1 to 30.
- H.263
  - ITU-T Recommendation H.263 v1 in 1995.
  - Superior to H.261, standard for videoconferencing.
  - H.263 v2 (H.263+, 1998)
  - H.263 v3 (H.263++, 2000)

H.26x Overview

- H.264 / Advanced Video Coding (AVC) / MPEG-4 Part 10
  - Standardized in 2003.
  - Up to 30% better compression efficiency than H.263.
  - Application: Internet video, computers, HDTV broadcast, Blu-ray discs, mobile and portable devices.
- H.265 / High Efficiency Video Coding (HEVC)
  - Basic structure similar to H.264.
  - Support up to 8K resolution and 120 fps.
  - More efficient in bitrate-quality trade-off than H.264.

Other Standards

- VC-1
  - Standardised by Society of Motion Picture and Television Engineers (SMPTE)
  - Implemented by Microsoft Windows Media Video (WMV)
- VP8
  - Target for environment such as web video
  - Support for Web video format and resolution up to 4K
- VP9
  - Basic structure similar to VP8
  - Support resolution up to 8K and 120 fps

Video Analysis and Understanding

Video Analysis and Understanding

- Perform video content analysis and understanding
- Some important tasks include:
  - Object Detection & Tracking
  - Pose Estimation
  - Human Action Recognition
  - Etc.

Object Detection

Source: https://www.youtube.com/watch?v=MyvOfDFZvgE

Object Tracking

Source: https://www.youtube.com/watch?v=nvoEV_q3bsA

Pose Estimation

Source: https://www.youtube.com/watch?v=98Wic3jUS0E

Action Recognition

Source: https://www.youtube.com/watch?v=PEziTgHx4cA

Egocentric Human-Object Interaction

Source: https://www.youtube.com/watch?v=VzoaKsDvv1o&t=16s

Emerging Topics

Generative AI

- A class of AI models that can create new contents, such as texts, images, audios, videos, etc.
- Examples:
  - Text generation: Generative Pre-trained Transformers (GPT).
  - Image generation: Stable Diffusion.
- Support diverse applications such as chatbots, content creation, virtual assistants, etc.

Foundation Models (FMs)

- A class of models that are trained on large-scale broad data that can be adapted (finetuned) to a wide range of downstream tasks / applications.
- Examples: Large Language Models (e.g., GPT), Vision-Language Models (e.g., CLIP).

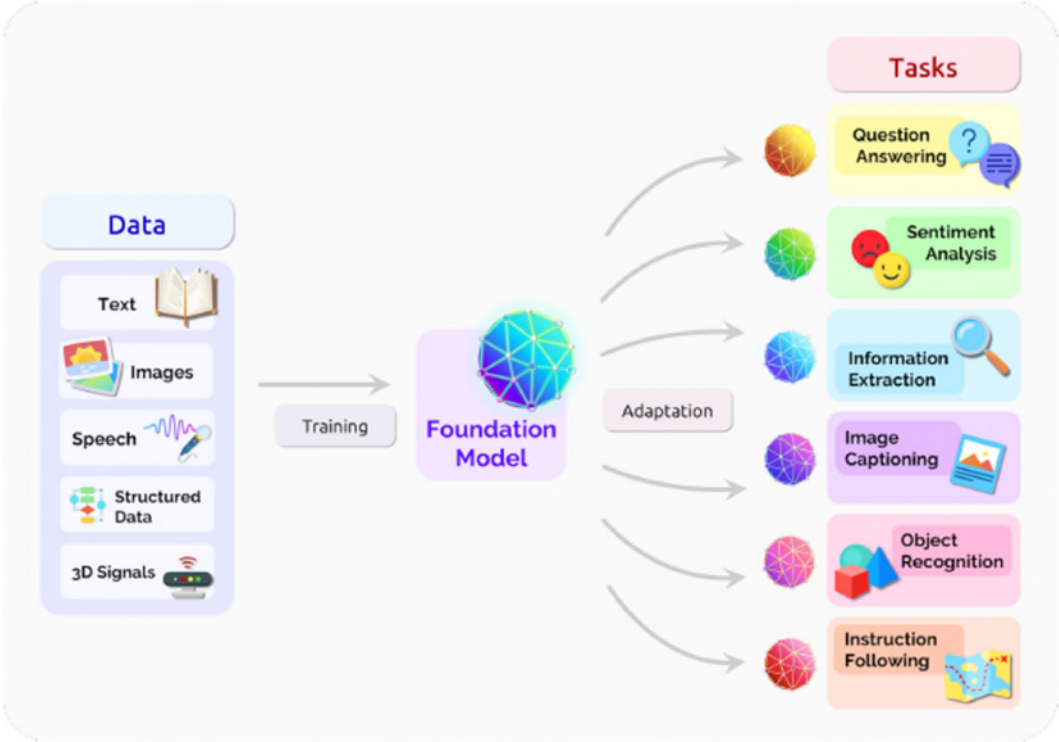
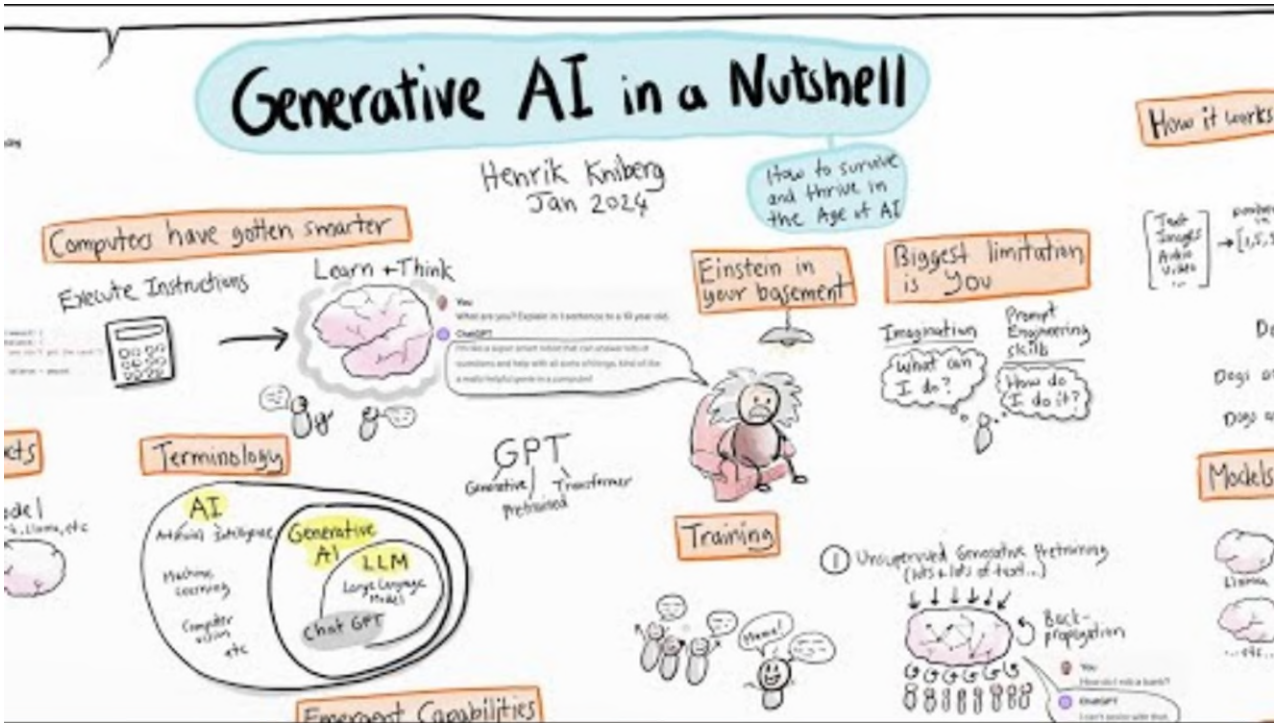


Fig. 2. A foundation model can centralize the information from all the data from various modalities. This one model can then be adapted to a wide range of downstream tasks.

Source: Rishi Bommasani, On the Opportunities and Risks of Foundation Models.

Generative AI & FMs



https://www.youtube.com/watch?v=2IK3DFHRFfw

Part 1 Summary

- This part covers the followings:
  - Image & Video Basics
  - Video Signal Processing & Applications
  - Image & Video Standards
  - Video Analysis & Understanding Basics
  - Emerging Topics