ECE 634: Digital Video Systems
Formats: 1/12/17

Professor Amy Reibman
MSEE 356
reibman@purdue.edu

http://engineering.purdue.edu/~reibman/ece634/index.html

Applications of digital video

- Entertainment
- Education
- Interactive communication
- Memorabilia, life-logging
- Medical and Scientific Imaging
- Information extraction
  – Surveillance, scene understanding

Items in blue: FOR humans

Items in red:
FOR humans or FOR machines

Items in red:
FOR humans or FOR machines

This course

Items in black:
FOR machines

- Motion models, estimation, and tracking
- Video compression (theory and practice)
- Video transport (error resilience; scalable coding)
- Stereo, 3D video, lightfields and beyond
- Video quality and how we see
- Video enhancement, stabilization
- Scene understanding and video analytics

Items in blue:
FOR humans

Today's outline 1/12/17

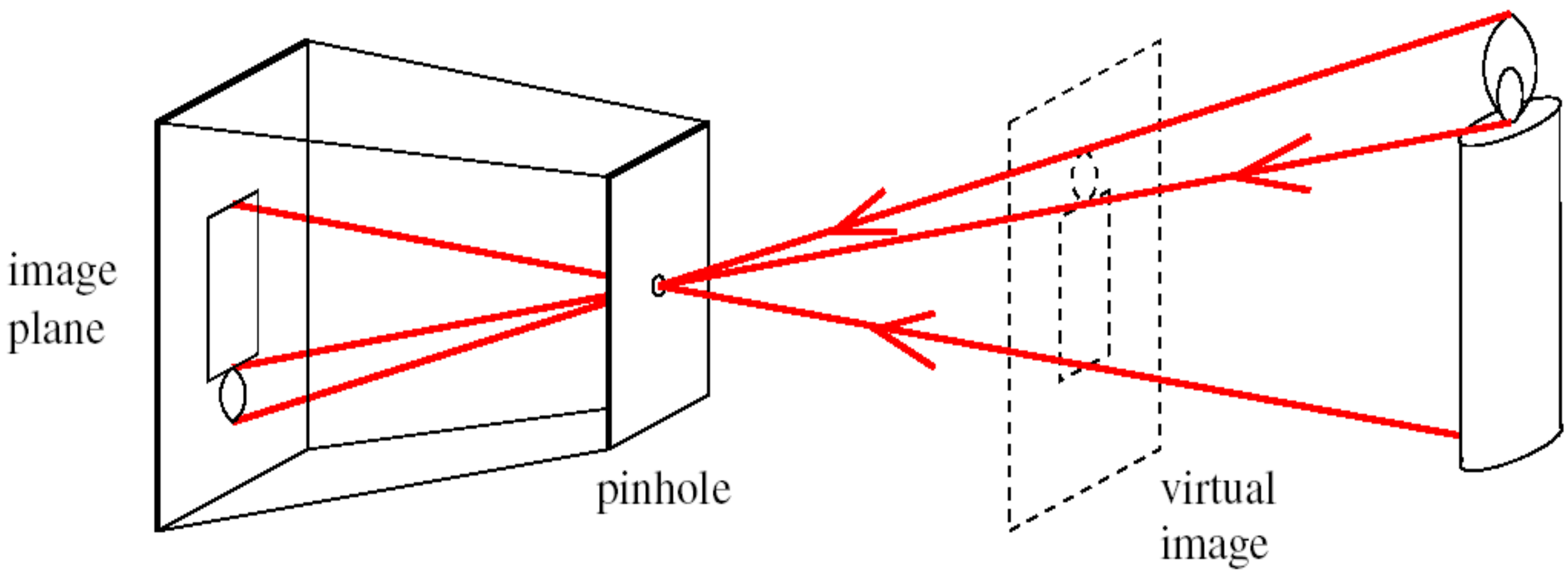
- Video formats
- Sampling the plenoptic function
- Color formats

Plenoptic Function (Adelson '91)

- Measures the intensity of light that passes through a particular point in space
- Every possible viewing position, with any viewing angle, at every moment in time
  - 3 location coordinates
  - 2 angular directions
  - Time
  - Wavelength

Image Formation in a Pinhole Camera

- Light enters a darkened chamber through pinhole opening and forms an image on the back surface



Video signal

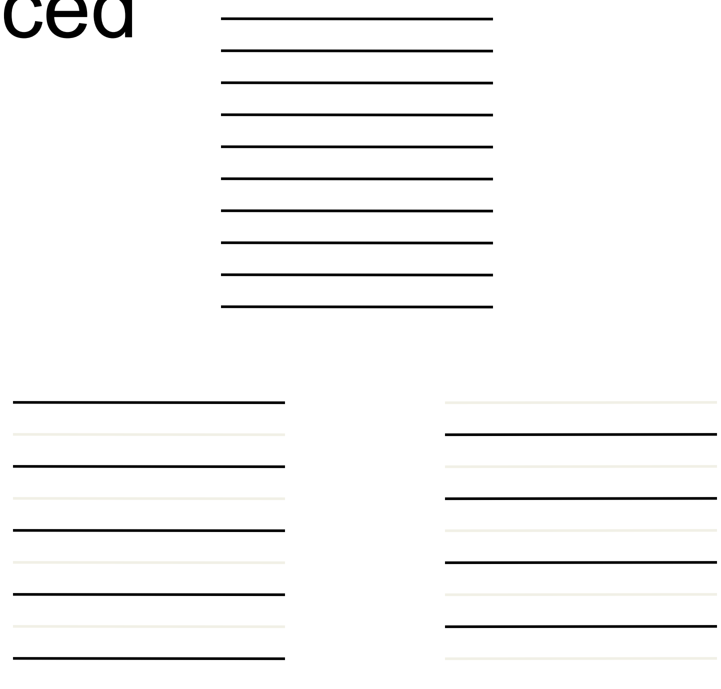
- What enters through the pinhole and projects on the image plane is a continuous 3-D signal (temporal, horizontal, vertical)
- Film records samples in time but continuous in space (typically 24 frames/sec)
- Analog video samples in time and samples vertically; continuous horizontally (about 30 frames/sec)
  - Number of lines controls the maximum vertical frequency that can be displayed for a given viewing distance
  - *Video-raster* = 1-D signal consisting of scan lines from successive frames
- Digital video: samples in time, vertically, and horizontally

On sampling temporally

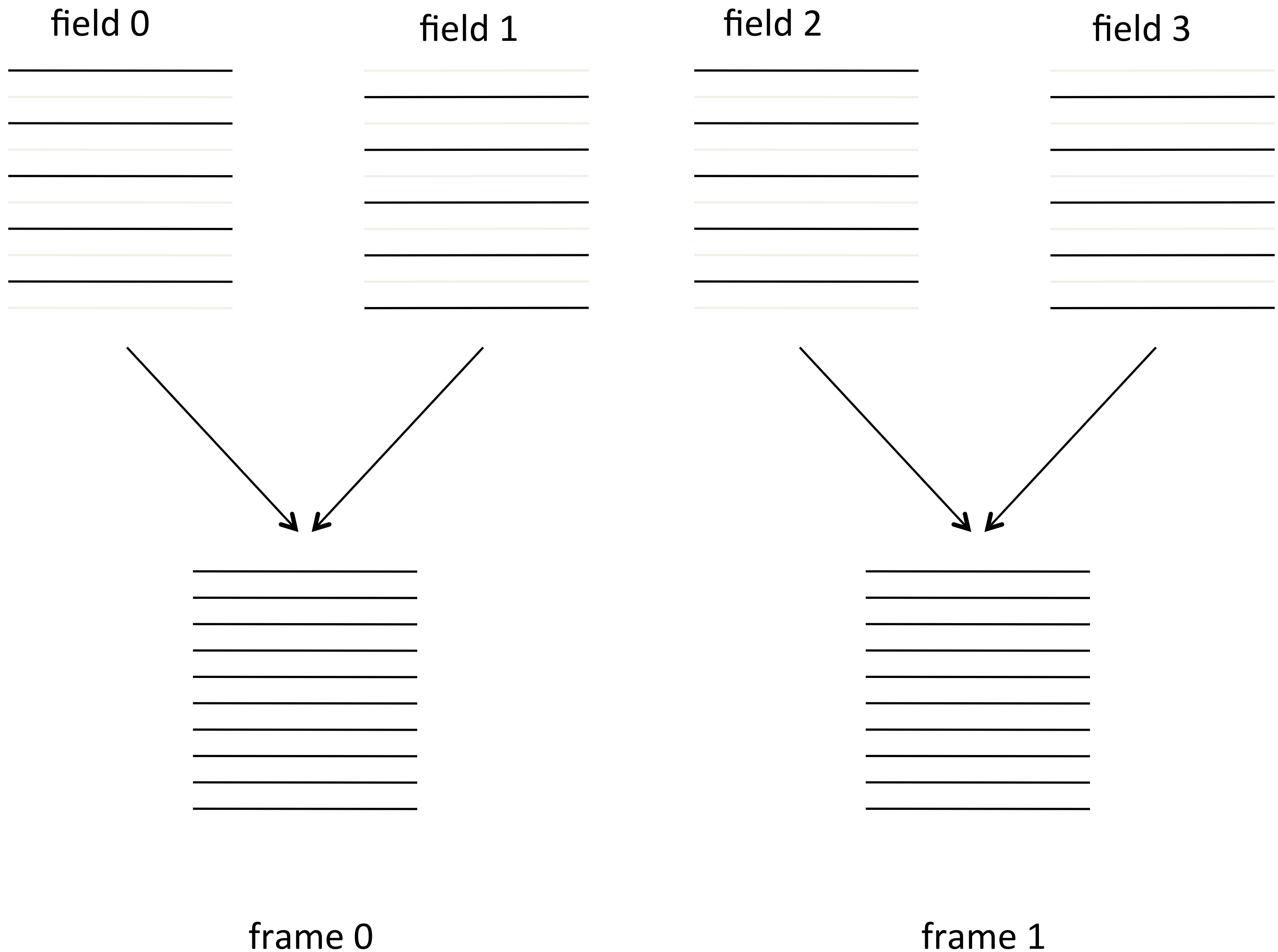
- Several different approaches:
  - Sample entire frame at the exact same time instant
  - Sample in raster scan order (a tracing finger)
  - Rolling shutter
  - Interlace
- We will touch on these more as they become relevant
  - For many applications, the first is a sufficiently accurate approximation

Interlacing

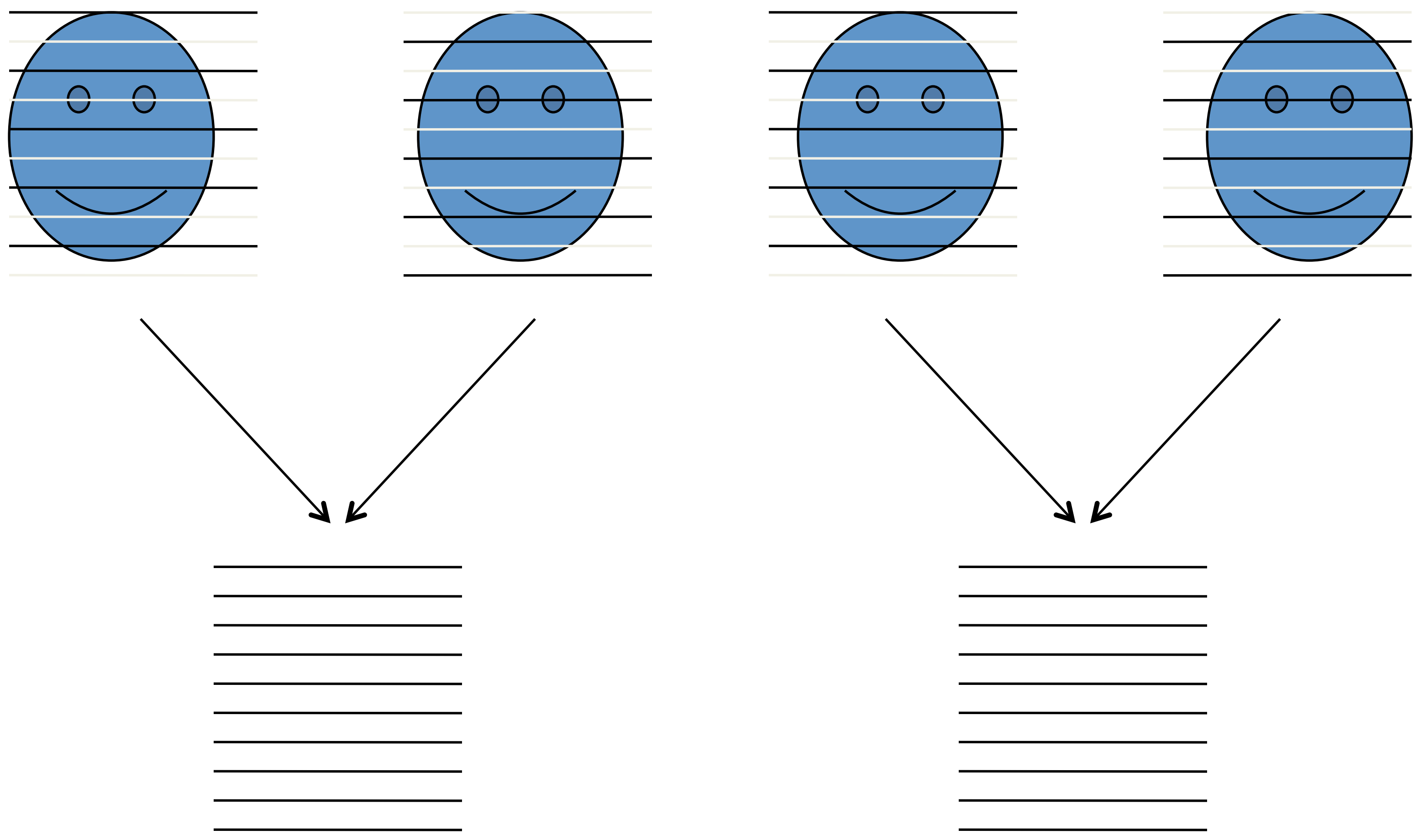
- (Almost) all Standard Definition video is interlaced



Interlaced video

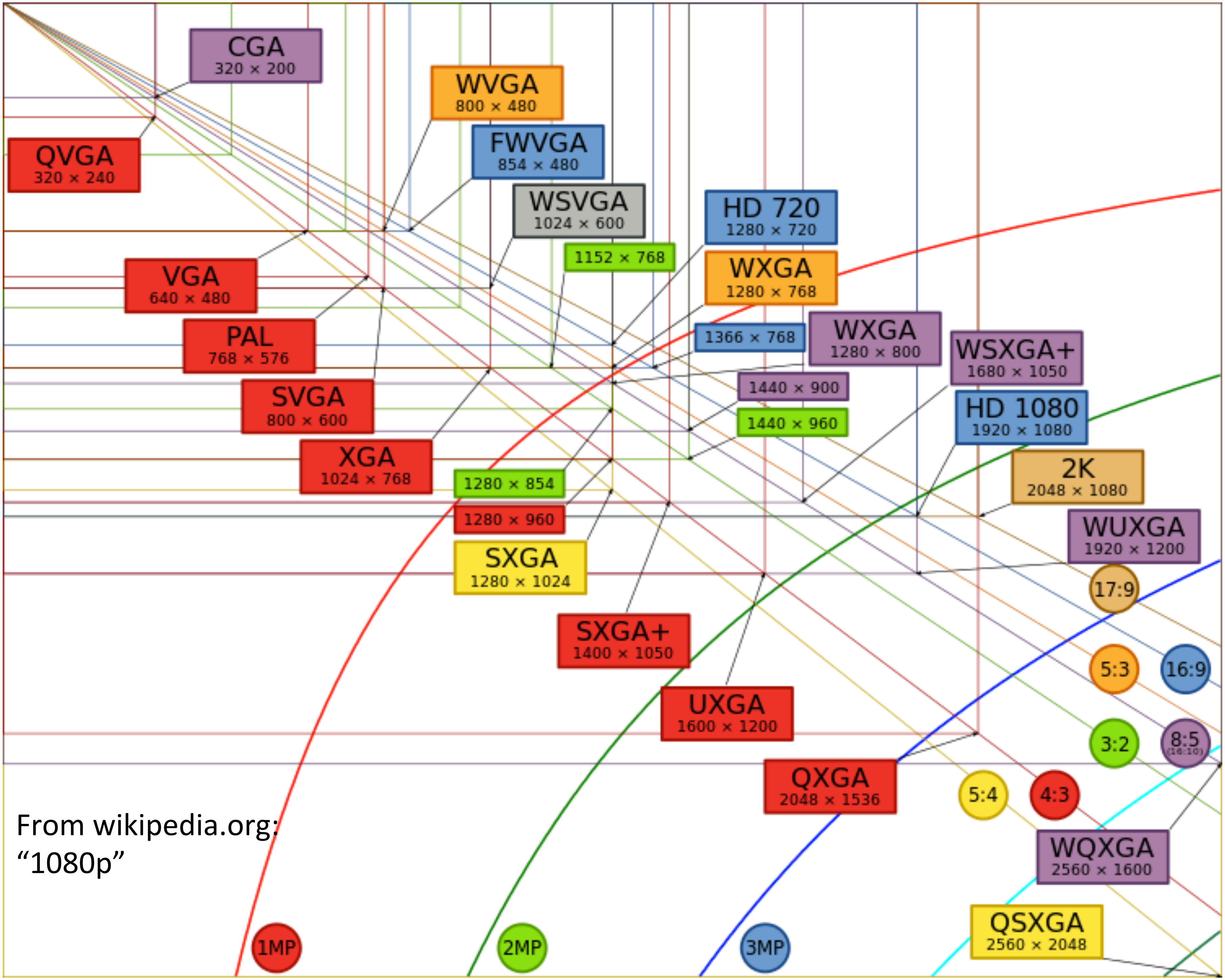


Interlaced video



Next up: sampling “angularly”

- On the next slide, each point in the 2D plane is a common format for video
- The diagonal lines indicate different aspect ratio
  - The rectangles are color-coded with the diagonal lines
- The arcs indicate the number of pixels per frame
  - Pixel === “picture element”



From wikipedia.org:
“1080p”

Digital formats

- Digital video is a sequence of frames (x,y,t)
- Often denoted {lines}{i,p} or {lines}{i,p}{fps}
  - 1080i, 720p, 1080p60
- Temporal resolutions
  - Video
    - 25, 30, 60 frames per second (fps)
    - 50, 60, 120 fields per second
  - Film: 24, 48 fps
  - Animation: often lower
- Why use more FPS?

Spatial resolutions

- High-definition TV (HDTV)
  - 1920 x 1080 (1080p or 1080i)
  - 1280 x 720 (720p)
- Standard-definition TV (SDTV or TV)
  - 720 x 480; 480 x 480;
  - D1: 720 x 486, 720 x 576
- Common Intermediate Format (CIF)
  - 352 x 288, 30 frames per second
  - Required for H.261 compression

Spatial resolutions (cont)

- Source Image Format (SIF)
  - 352 x 240; 352 x 288 (various frame rates!)
- Quarter CIF (QCIF)
  - 176 x 120; 176 x 144
- 4CIF

Aspect Ratio

- Picture width relative to picture height
- Display aspect ratios
  - NTSC 4:3
  - HDTV 16:9
- Pixel aspect ratios
  - Computers: square
  - TV: not square

Aspect ratio accommodations: fitting HD into SD, or SD into HD

- Squeeze video to fit
  - Tall skinny people; short wide people
- Letterboxing (Pillarboxing)
  - Fill top and bottom (left and right) with black
- Pan and scan
  - Show only a subset of the full content
  - Change viewing window over time if desired

HDTV formats

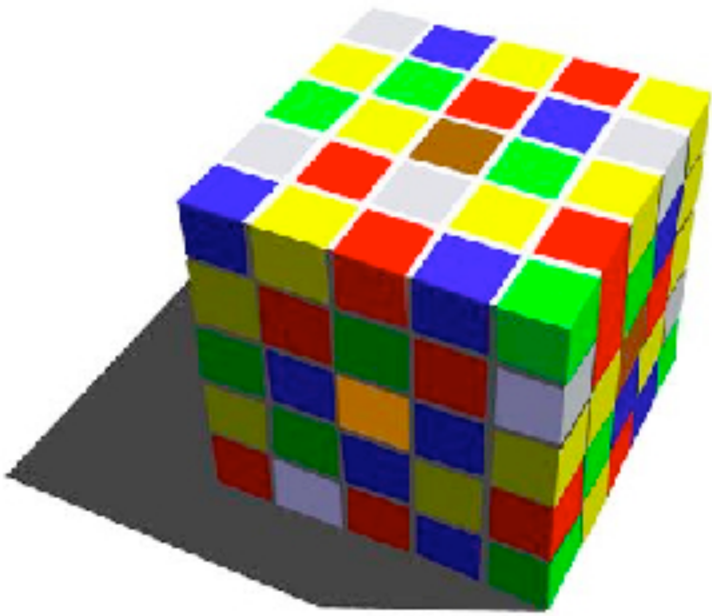
- 1080i
  - CBS, NBC
  - Improved spatial resolution
- 720p
  - Fox, ABC, ESPN, A&E, History Channel
  - Improved motion rendition

- Artifacts are prevalent when switching from one to another (if you're trained to see it)
  - Jagged edges (particularly during motion)

How many bits per pixel?

- Quantization transforms the continuous value at each pixel location into a digital number that can be represented by a fixed number of bits.
- Most video today is 8 bits per pixel (for luminance)
- Emerging High Dynamic Range (HDR) images and video are 10 or 12 or 16 bits per pixel

On sampling the wavelengths (Color)



What color are the central squares?

http://serendip.brynmawr.edu/~laurac/cube/cube.jpg

On sampling the wavelengths (Color)

What color are the central squares?

http://serendip.brynmawr.edu/~laurac/cube/cube.jpg

Colorimetry

- Color itself is a perceptual property
- NOT an attribute of an object, but of how our eyes and our brain perceive it

- We often talk about color of an image in terms of the wavelength of light emitted or reflected from objects in the image
- I'll use the shortcut "color" here for the latter case

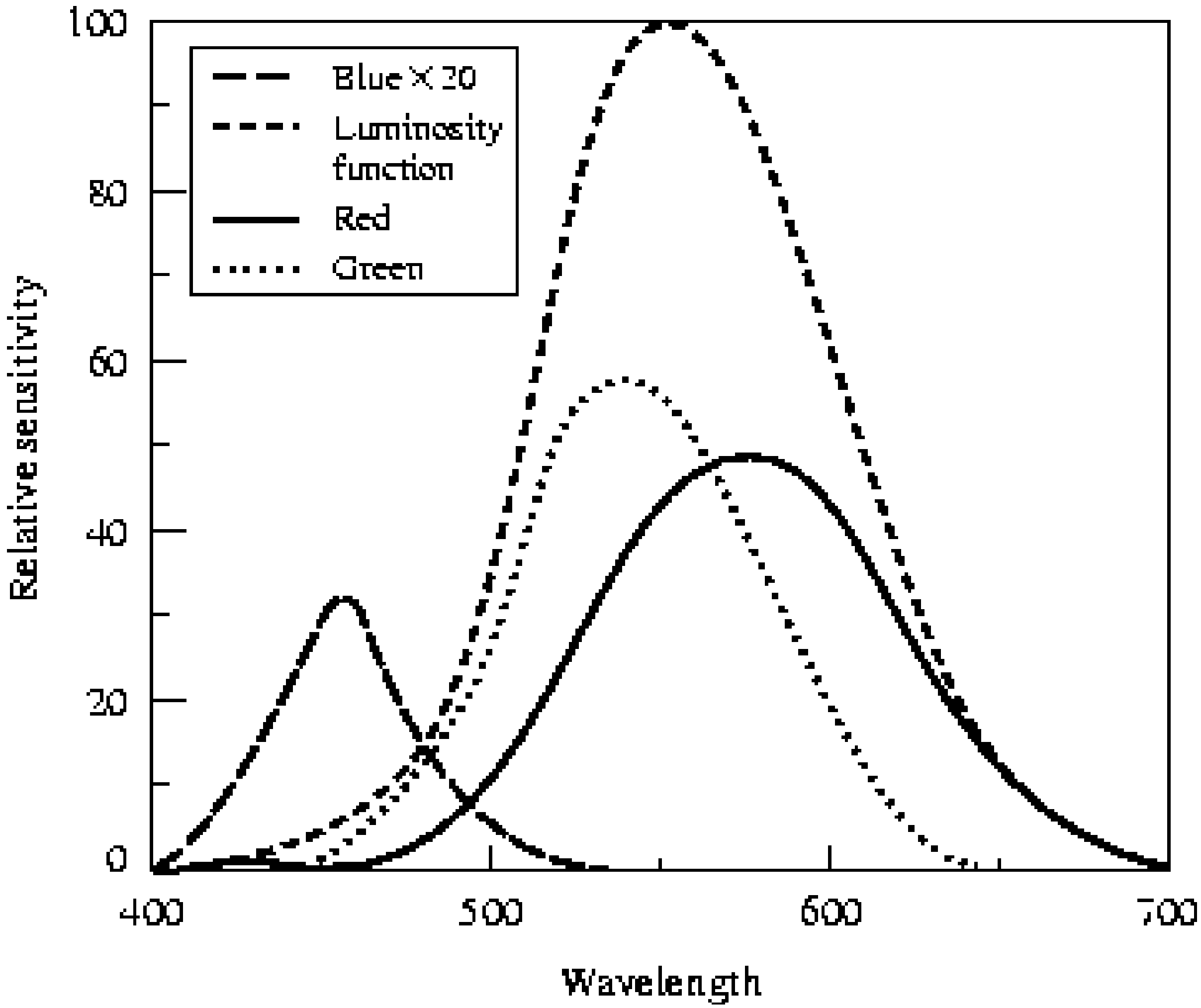
Illuminating and Reflecting Light

- Illuminating sources:
  - emit light (e.g. the sun, light bulb, TV monitors)
  - follows additive rule
    - R+G+B=White
- Reflecting sources:
  - reflect an incoming light (e.g. the color dye, matte surface, cloth)
  - Reflected frequencies are the emitted frequencies minus any absorbed frequencies
  - follows subtractive rule
    - R+G+B=Black

Human Perception of Color

- Retina contains photo receptors
  - Cones: day vision, can perceive color tone
    - Red, green, and blue cones
    - Different cones have different frequency responses
    - Tri-receptor theory of color vision [Young1802]
  - Rods: night vision, perceive brightness only
- Color sensation is characterized by
  - Luminance (perceived brightness)
  - Chrominance (perceived color tone)
    - Hue (color tone or peak wavelength)
    - Saturation (color purity)

Frequency Responses of Cones and the Luminous Efficiency Function



Trichromatic Color Mixing

- Trichromatic color mixing theory
  - Any color can be obtained by mixing three primary colors in the right proportion

$$C = \sum_{k=1,2,3} T_k C_k, \quad T_k : \text{Tristimulus values}$$

- Primary colors for illuminating sources (i.e., monitors):
  - Red, Green, Blue (RGB)
- Primary colors for reflecting sources (i.e. printed papers)
  - (also known as secondary colors)
  - Cyan, Magenta, Yellow (CMY)

Color Representation Models

- Tristimulus values associated with the three primary colors
  - RGB or CMY
- Amplitude specification:
  - 8 bits for each color component; 24 bits total for each pixel
  - 16 million colors
- Luminance and chrominance
  - HSI (Hue, saturation, intensity)
  - YIQ (used in NTSC color TV)
  - YCbCr (used in digital color TV)

Many color spaces

- Conversion between primary and XYZ/YIQ/YUV are linear (3x3 matrix)

$$Y' = \ \ 0.299\ R + 0.587\ G + 0.114\ B$$
$$U = -\ 0.147\ R - 0.289\ G + 0.436\ B$$
$$V = \ \ 0.615\ R - 0.515\ G - 0.100\ B$$

$$U = 0.436\ (B-Y')/(1-0.114)$$
$$V = 0.615\ (R-Y')/(1-0.299)$$

* For BT.601 and SDTV. Matrix for BT.709 and HDTV differs!

Perceptually uniform color spaces

- Perceptually uniform:
  - A small perturbation to a value is approximately equally perceptible across the range of that value

- XYZ, RGB tristimulus values are not perceptually uniform
- L*u*v* (CIELUV) and L*a*b* (CIELAB) are closer
  - Involves a cube-root
  - Computationally complicated

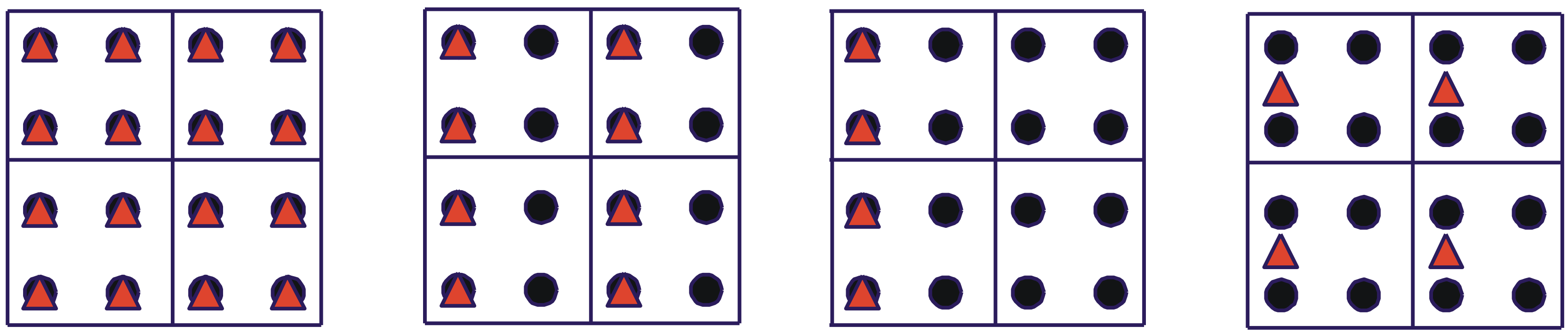
Choosing color coordinates

- For display or printing: RGB or CMY, to produce more colors
- For analyzing color differences: HSI, for linear relationship.
- For processing perceptually meaningful color: L*a*b*
- For transmission or storage: YIQ or YUV, for a less redundant representation

Color in images and videos and OpenCV

- Images are commonly RGB, and each pixel location has 3 colors
  - (this is ignoring Bayer color sampling)
  - BE CAREFUL!!! OpenCV loads images as BGR
- Videos are commonly YUV or YCbCr, and there are fewer color pixels than luminance pixels
  - OpenCV will automatically convert videos in YUV into consecutive images of RGB, upsampling the color information

Chrominance Subsampling Formats



4:4:4
For every 2x2 Y Pixels
4 Cb & 4 Cr Pixel
(No subsampling)

4:2:2
For every 2x2 Y Pixels
2 Cb & 2 Cr Pixel
(Subsampling by 2:1
horizontally only)

4:1:1
For every 4x1 Y Pixels
1 Cb & 1 Cr Pixel
(Subsampling by 4:1
horizontally only)

4:2:0
For every 2x2 Y Pixels
1 Cb & 1 Cr Pixel
(Subsampling by 2:1 both
horizontally and vertically)

● Y Pixel

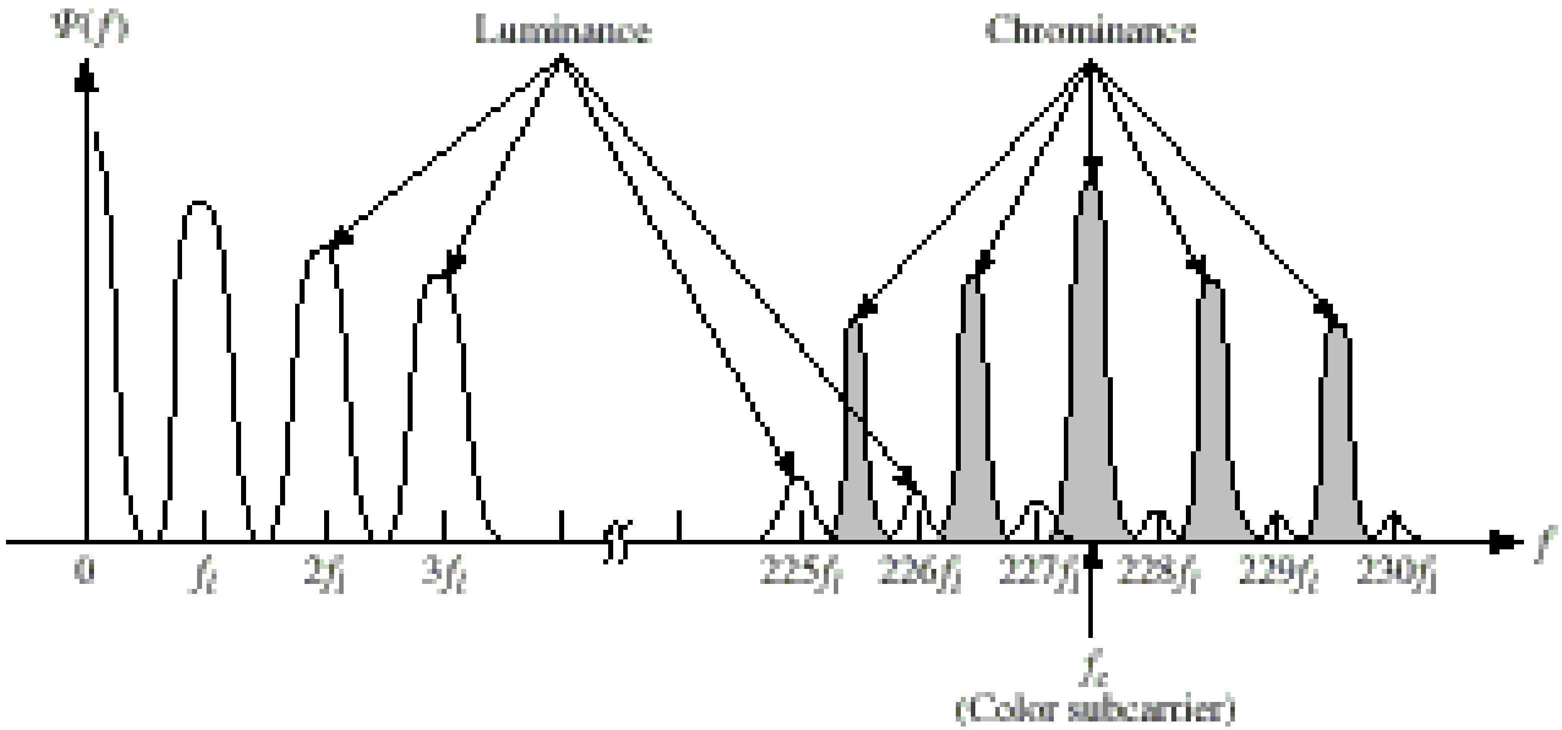
▲ Cb and Cr Pixel

History

History of TV in US

- 1941: First NTSC broadcast, monochrome
  - 4:3 aspect ratio; Interlacing
  - 60 Hz (60 fields per second)
  - 525 lines but only 480 active lines
- 1953: Color NTSC
  - Backwards compatible with black and white TVs
- 1993: Grand Alliance forms to design HDTV
- 1996: First public broadcast of HDTV
- 2000: First HDTV Superbowl transmission
- 2009: Last analog transmission

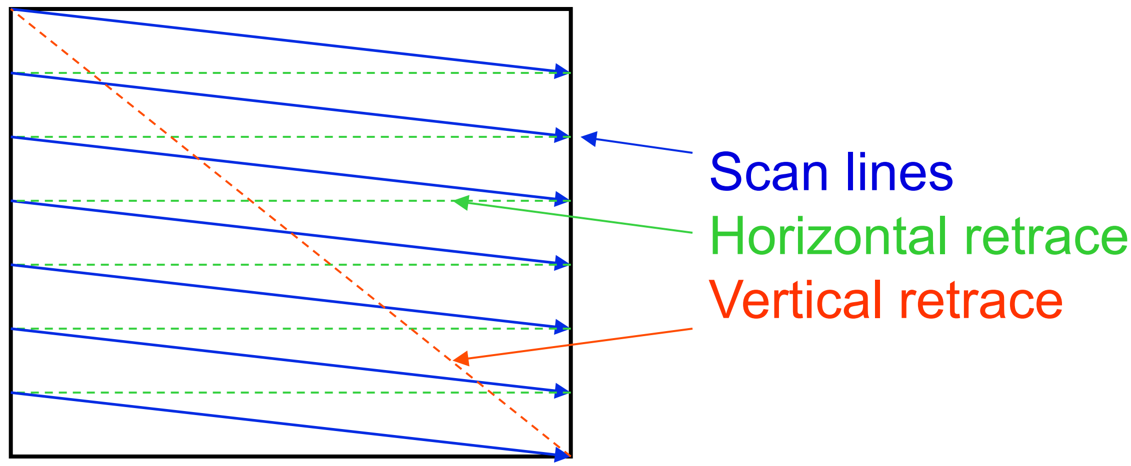
NTSC Spectrum Illustration



Analog video

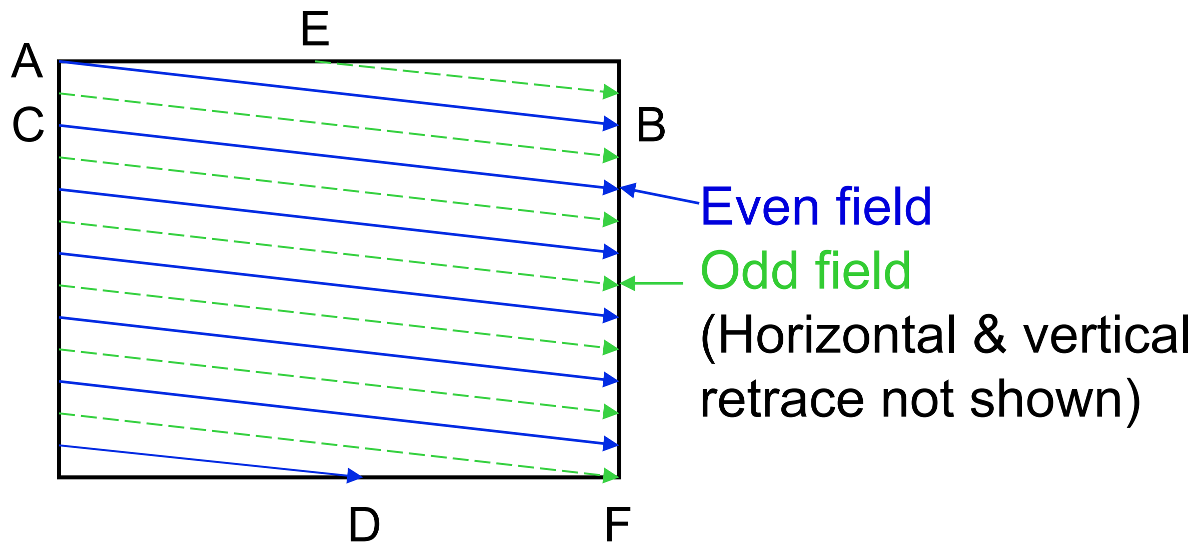
- NTSC (North America + Japan)
  - 59.95 fields per second; 525 lines; YIQ
- PAL (most of Western Europe, India, Australia)
  - 50 fields per second; 625 lines; YUV
- SECAM (most of the rest of Asia; eastern Europe)
  - 50 fields per second; 625 lines; YDbDr

Progressive Scanning



- Progressive scan:
  - Captures consecutive lines
  - Captures a complete frame every Δt sec
  - Also referred to as sequential or non-interlaced
  - Used by computers

Interlaced Scanning



- Interlaced scan:
  - Captures alternate lines (each frame split into two fields)
    - Odd lines are captured (odd field), then even lines (even field)
  - Captures a complete frame every Δt sec
  - Used in analog television

Why interlaced?

- Provides a trade-off between temporal and vertical resolution, for a given, fixed data rate

Hold overs from analog video

- Temporal scanning: Interlace and progressive
- Digital video standards included capability for interlace – up until H.265
  (High Efficiency Video Coding HEVC)

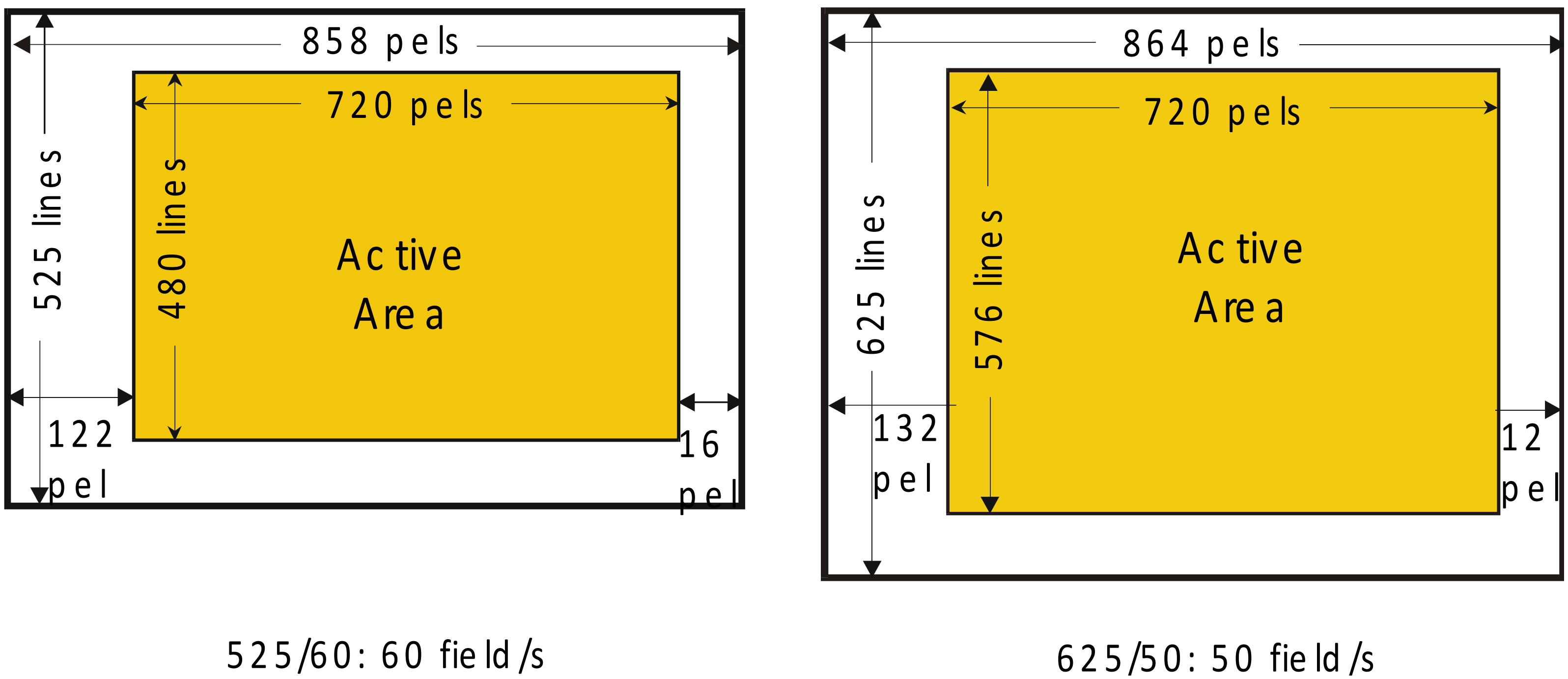
- Different spatial samplings: NTSC and PAL
  - 525 lines at 30 fps; or 625 lines at 25 fps

Spatial resolutions (revisited)

- Common Intermediate Format (CIF)
  - 352 x 288, 30 frames per second
  - Required for H.261 compression
- Source Image Format (SIF)
  - 352 x 240; 352 x 288 (various frame rates!)
- Quarter CIF (QCIF)
  - 176 x 120; 176 x 144
- 4CIF

BT.601* Video Format

Digital encoding of analog video



* BT.601 is formerly known as CCIR601

Over the air HDTV in US

- Flexible picture formats
- Progressive scan or interlace scan
- Video compression based on MPEG-2

Digital Video Formats

Video Format	Y Size	Color Sampling	Frame Rate (Hz)	Raw Data Rate (Mbps)
HDTV Over air. cable, satellite, MPEG2 video, 20-45 Mbps				
SMPTE296M	1280x720	4:2:0	24P/30P/60P	265/332/664
SMPTE295M	1920x1080	4:2:0	24P/30P/60I	597/746/746
Video production, MPEG2, 15-50 Mbps				
BT.601	720x480/576	4:4:4	60I/50I	249
BT.601	720x480/576	4:2:2	60I/50I	166
High quality video distribution (DVD, SDTV), MPEG2, 4-10 Mbps				
BT.601	720x480/576	4:2:0	60I/50I	124
Intermediate quality video distribution (VCD, WWW), MPEG1, 1.5 Mbps				
SIF	352x240/288	4:2:0	30P/25P	30
Video conferencing over ISDN/Internet, H.261/H.263, 128-384 Kbps				
CIF	352x288	4:2:0	30P	37
Video telephony over wired/wireless modem, H.263, 20-64 Kbps				
QCIF	176x144	4:2:0	30P	9.1