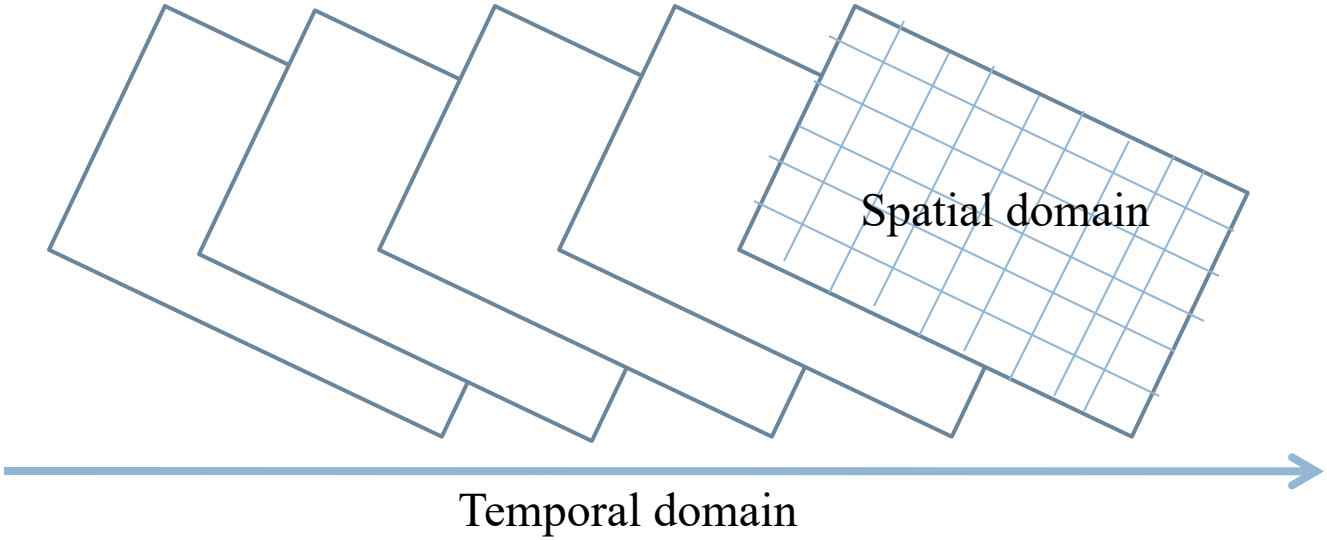
Essence of Video

Wei-Ta Chu

Constitution of Digital Video Data

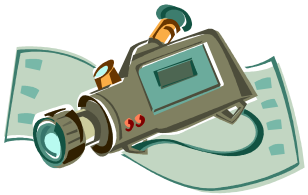
- A natural video stream is continuous in both spatial and temporal domains.
- In order to represent and process a video stream digitally it is necessary to sample spatially and temporally.



Video Stream

Natural scene

Camera

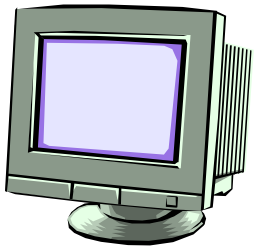


RGB to YC_1C_2

Processing, Storage, Transmission

YC_1C_2 To RGB

Monitor



Video Data Representation

- RGB is not very efficient for representing real-world images, since equal bandwidths are required to describe all the three color components.
  - E.g. 8 bits per component, then 24 bits per pixel
- Human eye is more sensitive to luminance.
- Many image coding standards and broadcast systems use luminance and color difference signals.
- YUV and YIQ for analog television standards, YCbCr for their digital version.

The YUV Color Model

- The U signal is then created by subtracting the Y from the blue signal, and then scaling; V is created by subtracting the Y from the red, and then scaling by a different factor.

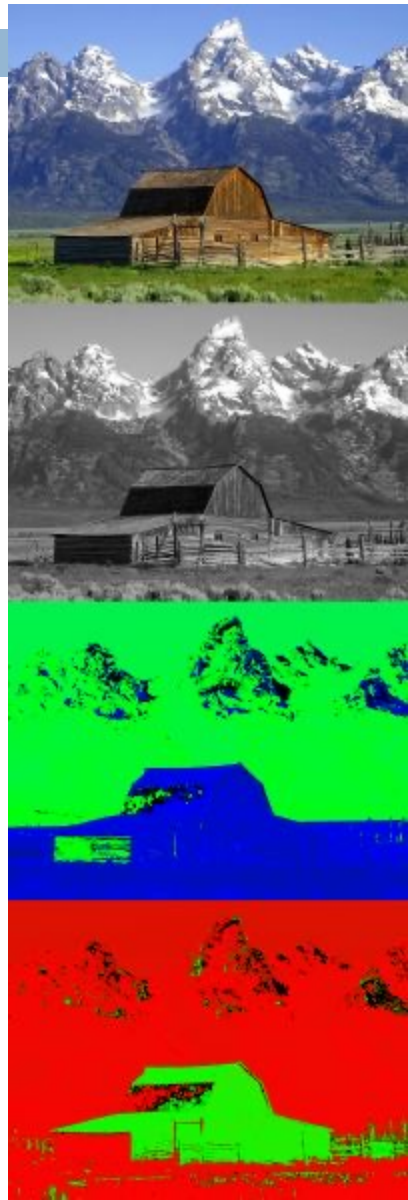
$$
\begin{aligned}
Y &= 0.299 * R + 0.587 * G + 0.114 * B \\
U &= 0.436 * (B - Y)/(1 - 0.114) \\
V &= 0.615 * (R - Y)/(1 - 0.299)
\end{aligned}
$$

$$
\begin{aligned}
Y &= 0.299 * R + 0.587 * G + 0.114 * B \\
U &= -0.14713 * R - 0.28886 * G + 0.436 * B \\
V &= 0.615 * R - 0.51499 * G - 0.10001 * B
\end{aligned}
$$

Y

U

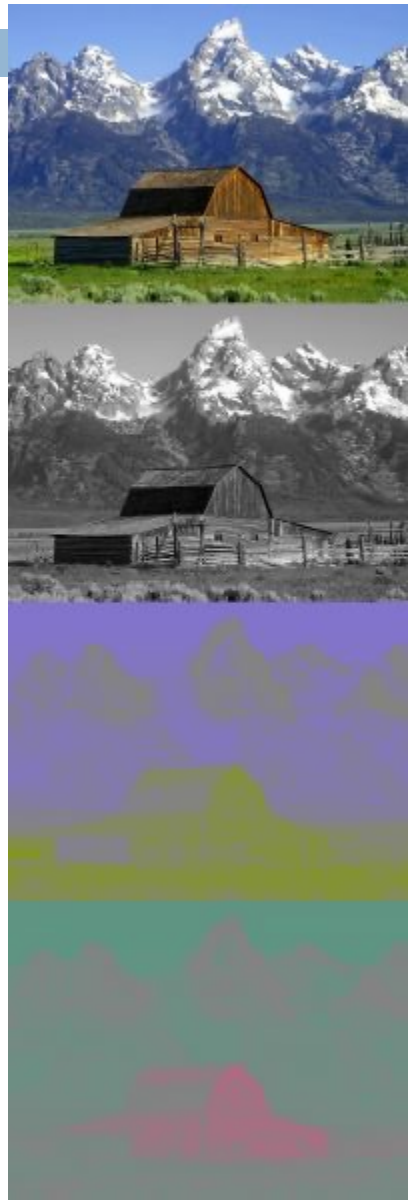
V



The YCbCr Color Model

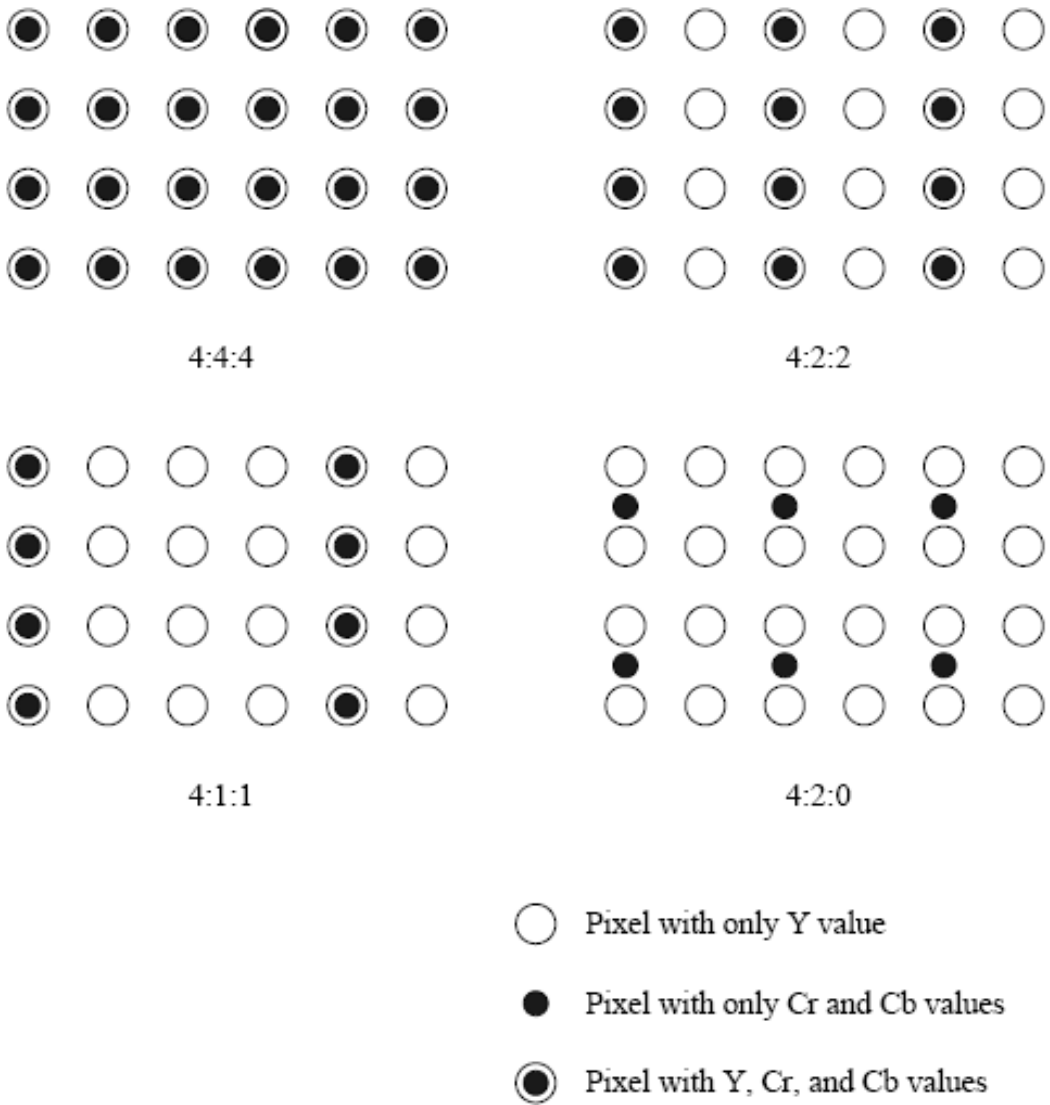
- YCbCr is a family of color spaces used in video and digital photography systems. Y is the luma component and Cb and Cr are the blue and red chroma components.
- Recommendation 601 specifies 8-bit coding:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$



Chroma Subsampling

- "4:2:2" indicates horizontal subsampling of the Cb, Cr signals by a factor of 2.
- Of four pixels labeled as 0 to 3, all four Ys are sent, and every two Cb's the two Cr's are sent.
- (Y0,Cb0) (Y1,Cr0) (Y2,Cb2) (Y3,Cr2) …
- "4:2:0" subsamples in both the horizontal and vertical dimensions by a factor of 2.



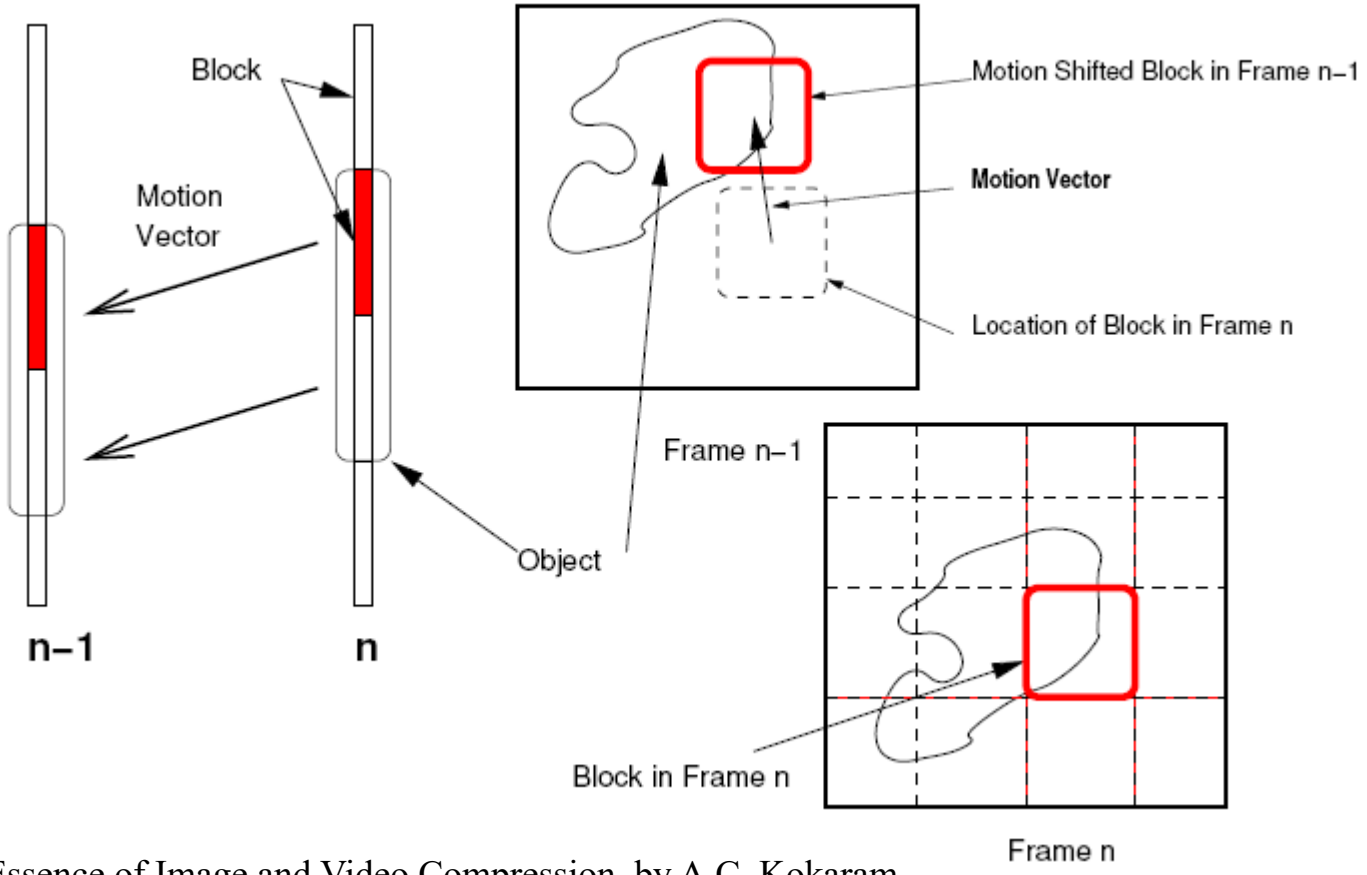
Examples

- Given image resolution of 720x576 pixels represented with 8 bits each component, the bit rate required is:
  - 4:4:4 resolution: 720x576x8x3 = 10 Mbits/frame
  - 4:2:0 resolution: (720x576x8) + (360x288x8)x2 = 5 Mbits/frame

Motion Estimation

- Successive video frames may contain the same objects (still or moving).
- Motion estimation examines the movement of objects in an image sequence to try to obtain vectors representing the estimated motion.

Motion Estimation



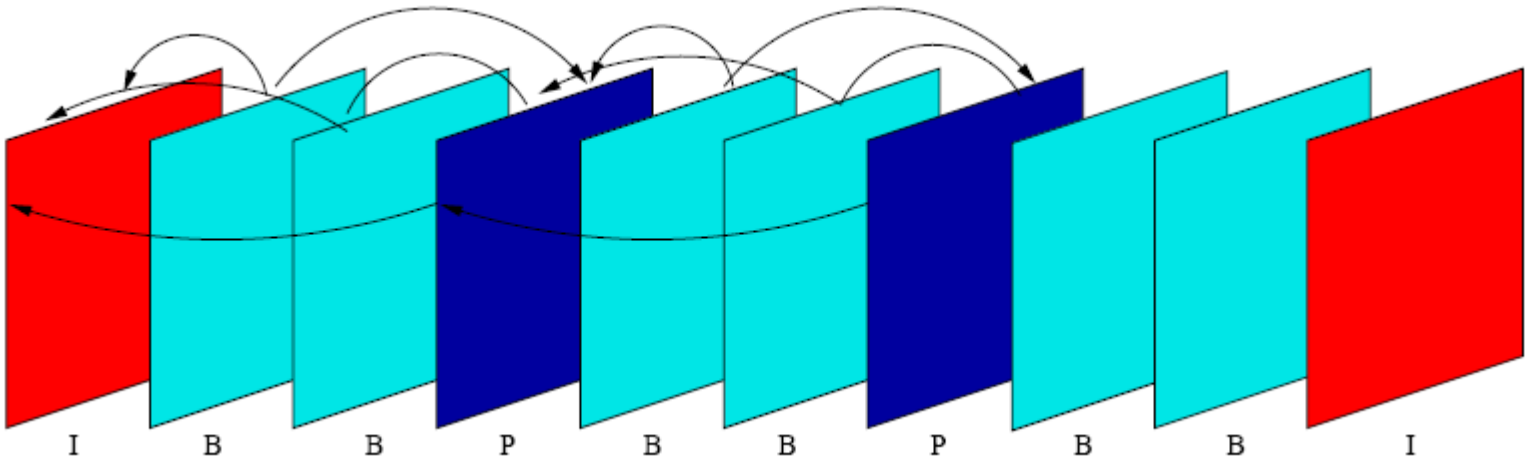
The Essence of Image and Video Compression, by A.C. Kokaram
http://www.mee.tcd.ie/~ack/teaching/1e8/lecture3.pdf

Three Typical Types of Coded Picture

- I frame (intraframe)
  - Intraframe encoded without any temporal prediction
- P frame (forward predicted frame)
  - Interframe encoded using motion predition from the previous I or P frame
- B frame (bidirectionally predicted frame)
  - Interframe encoded using interpolated motion prediction between the previous I or P frames and the next I or P frames.

Motion Prediction

- A typical Group of Picture (GOP) in MPEG-2



Short Introduction to Video Features

- Motion-based features
  - Camera motion, object motion
  - Motion activity/magnitude
  - Moving object detection
- Shot-based features
  - Average shot length/shot change frequency
- Scene-based features

Motion Type

- Camera motion (global motion)
  - Zoom-in/Zoom-out
  - Pan
  - Tilt
- Object motion