Chapter 2
Multimedia Basics and Representation

2.1 Digital multimedia characteristics

2.2 Audio formats and MIDI

2.3 Image formats and color models

2.4 Video formats and color models

2.1 Digital multimedia characteristics

A Multimedia system has four basic characteristics:

- Multimedia systems must be **computer controlled.**
- Multimedia systems are **integrated**.
- The information they handle must be represented **digitally**.
- The interface to the final presentation of media is usually **interactive.**

Computer Controlled

- Producing the content of the information – e.g. by using the authoring tools, image editor, sound and video editor
- Storing the information – providing large and shared capacity for multimedia information.
- Transmitting the information – through the network.
- Presenting the information to the end user – make direct use of computer peripheral such as display device (monitor) or sound generator (speaker).

Integrated

- All **multimedia** components (audio, video, text, graphics) used in the system must be somehow integrated.
- Every device, such as microphone and camera is connected to and controlled by a single computer.
- A single type of digital storage is used for all media type.
- Video sequences are shown on computer screen instead of TV monitor.

Interactivity

- **Level 1**: Interactivity strictly on information delivery. Users select the time at which the presentation starts, the order, the speed and the form of the presentation itself.
- **Level 2**: Users can modify or enrich the content of the information, and this modification is recorded.
- **Level 3**: Actual processing of users input and the computer generate genuine result based on the users input.

Digitally Represented

- Digitization: process involved in transforming an analog signal to digital signal.

2.2 Audio formats and MIDI

2.2.1 Audio Formats

- Digital sound files must be organized and structured so that your media player can read them.
- It’s just like being able to read and understand a different language.
- If the player “speaks” the language that the files are recorded in, it can reproduce the song and make beautiful music. If it can’t speak the language, the numbers of the music don’t add up, and you get an error message — and no music.
- The major file formats discussed here will be WAV, MP3 and MIDI.

WAV

- WAV format is the most detailed and rich of the available formats in Windows XP.
- All the detail is recorded at the chosen bit rate and sampling speed, and it’s all done without compression schemes.
- Unfortunately, it takes up huge amounts of memory in the process. Four or five minutes of WAV sound can consume 40–50MB of memory, making it difficult to store a decent number of files.

MP3

- MP3 single-handedly powered the popularity of digital music.
- It is an audio layer of the larger MPEG file format.
- Because of its small file size, MP3 files are ideal for listening on a computer or a portable player. MP3 is a specific way to make the music file smaller while retaining much of the quality of the original CD or WAV file.
- The other advantage of MP3 is that it’s almost universally recognized. Just about any media player or portable audio player can recognize and play an MP3 song. That makes it popular among users.

MIDI

- MIDI or Musical Instrument Digital Interface, is radically different from any other format.
- Technically, MIDI is not even audio; it’s a set of instructions on how something (like your computer’s sound card) should create music.
- Because it’s just a set of instructions, the MIDI file size is quite small (often measured in kilobytes as opposed to the larger megabytes).
- How those instructions sound can vary depending on the device that is used to play those instructions.

2.2.2 MIDI: Musical Instrument Digital Interface

- A protocol that enables computer, synthesizers, keyboards, and other musical device to communicate with each other.
- This protocol is a language that allows interworking between instruments from different manufacturers by providing a link that is capable of transmitting and receiving digital data.
- Transmits only commands, it does not transmit an audio signal.
- It was created in 1982.

Components of a MIDI System

1. **Synthesizer:**

- ✓ It is a sound generator (various pitch, loudness, tone color).
- ✓ A good (musician’s) synthesizer often has a microprocessor, keyboard, control panels, memory, etc.

2. **Sequencer:**

- ✓ It can be a stand-alone unit or a software program for a personal computer. (It used to be a storage server for MIDI data).
- ✓ Nowadays it is more a software music editor on the computer.)
- ✓ It has one or more MIDI INs and MIDI OUTs.

Basic MIDI Concepts

Track:

✓Track in sequencer is used to organize the recordings.

✓Tracks can be turned on or off on recording or playing back.

Channel:

✓Channels are used to separate information in a MIDI system.

✓There are 16 MIDI channels in one cable.

✓Channel numbers are coded into each MIDI message.

Timbre:

✓The quality of the sound, e.g., flute sound, cello sound, etc.

✓Multitimbral – capable of playing many different sounds at the same time (e.g., piano, brass, drums, etc.)

Pitch:

✓The musical note that the instrument plays

Voice:

✓Voice is the portion of the synthesizer that produces sound.

✓Synthesizers can have many (12, 20, 24, 36, etc.) voices.

✓Each voice works independently and simultaneously to produce sounds of different timbre and pitch.

Patch:

✓The control settings that define a particular timbre.

2.3 Image formats and color models

- An image could be described as two-dimensional array of points where every point is allocated its own color.
- Every such single point is called pixel, short form of picture element.
- Image is a collection of these points that are colored in such a way that they produce meaningful information/data.
- Pixel (picture element) contains the color or hue and relative brightness of that point in the image. The number of pixels in the image determines the resolution of the image.
  - ✓ A digital image consists of many picture elements, called pixels.
  - ✓ The number of pixels determines the quality of the image

    i.e. image resolution.

- Higher resolution always yields better quality.
- Bitmap resolution, most graphics applications let you create bitmaps up to 300 dots per inch (dpi). Such high resolution is useful for print media, but on the screen most of the information is lost, since monitors usually display around 72 to 96 dpi.
- A bit-map representation stores the graphic/image data in the same manner that the computer monitor contents are stored in video memory.
- Most graphic/image formats incorporate compression because of the large size of the data.

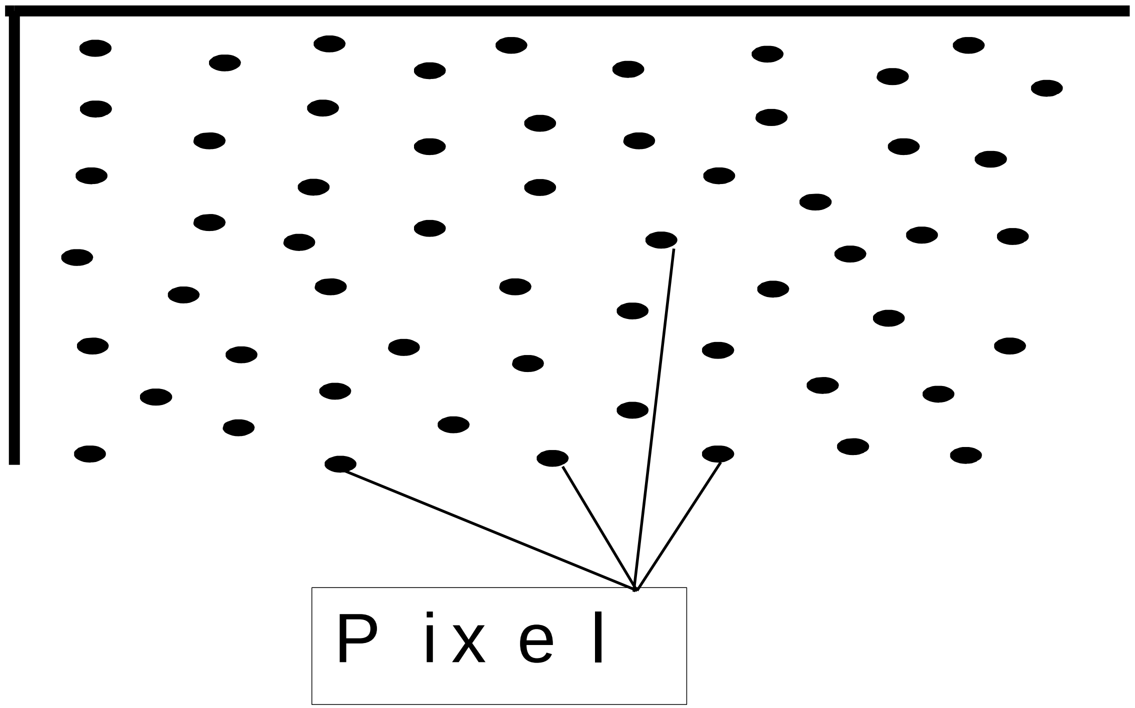

Fig 2.1: A pixel

Types of Images

- There are two basic forms of computer graphics: bit-maps and vector graphics.
- Bitmap formats are the ones used for digital photographs. Vector formats are used only for line drawings.

1.Bit-map images (also called Raster Graphics)

- They are formed from pixels—a matrix of dots with different colors. Bitmap images are defined by their dimension in pixels as well as by the number of colors they represent. For example, a 640X480 image contains 640 pixels and 480 pixels in horizontal and vertical direction respectively.

2. Vector graphics

- They are really just a list of graphical objects such as lines, rectangles, ellipses, arcs, or curves—called primitives.
- Draw programs, also called vector graphics programs, are used to create and edit these vector graphics.

Vector graphics have a number of advantages over raster graphics. These include:

- ✓ Precise control over lines and colors.
- ✓ Ability to skew and rotate objects to see them from different angles or add perspective.
- ✓ Ability to scale objects to any size to fit the available space. Vector graphics always print at the best resolution of the printer you use, no matter what size you make them.
- ✓ Color blends and shadings can be easily changed.
- ✓ Text can be wrapped around objects.

Types of Bitmap Images

1. Monochrome/Bit-Map Images

- Each pixel is stored as a single bit (0 or 1)
- The value of the bit indicates whether it is light or dark
- A 640 x 480 monochrome image requires 37.5 KB of storage

Fig 2.2: Monochrome 1-bit Lena image

2. Grayscale Images

- Each pixel is usually stored as a byte (value between 0 to 255)
- This value indicates the degree of brightness of that point. This brightness goes from black to white
- A 640 x 480 grayscale image requires over 300 KB of storage.

Fig 2.3: Grayscale image of Lena

3. 8-bit Color Images

- One byte for each pixel
- Supports 256 out of the millions possible, acceptable color quality
- Requires Color Look-Up Tables (CLUTs)

  Basically, the image stores not color, but instead just a set of bytes, each of which is actually an index into a table with 3-byte values that specify the color for a pixel with that lookup table index.

- A 640 x 480 8-bit color image requires 300 KB of storage (the same as 8-bit grayscale)

Color Look-up Tables (LUTs)

- The idea used in 8-bit color images is to store only the index, or code value, for each pixel. Then, e.g., if a pixel stores the value 25, the meaning is to go to row 25 in a color look-up table (LUT).

➢For 25, R= 00011110, G=10111110 ,and B= 00111100

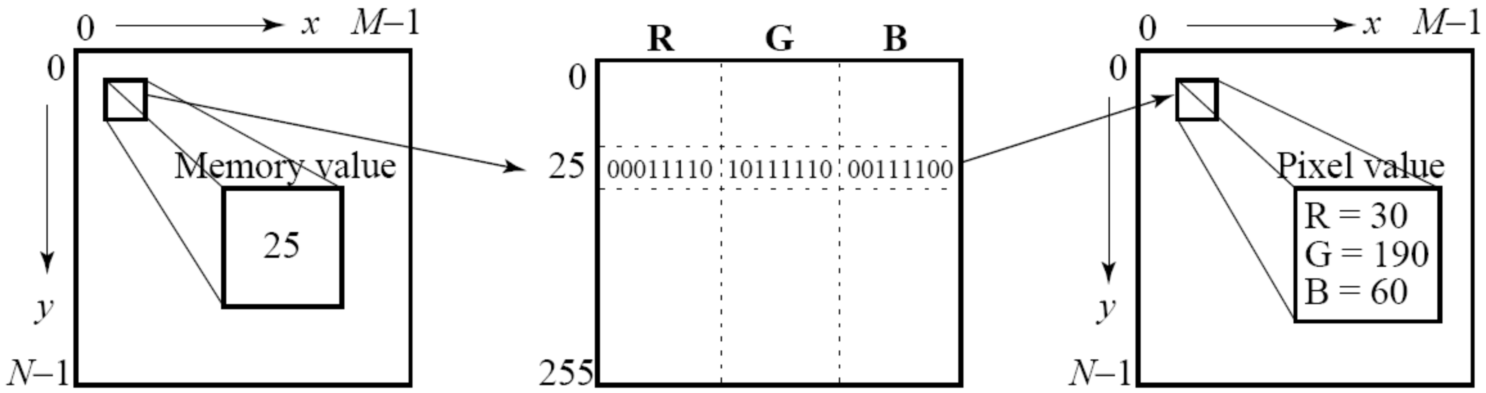

Fig 2.4: Color LUT for 8-bit color images

4. 24-bit Color Images

- Each pixel is represented by three bytes (e.g., RGB)
- Supports 256 x 256 x 256 possible combined colors (16,777,216)
- A 640 x 480 24-bit color image would require 900 KB of storage
- Most 24-bit images are 32-bit images, the extra byte of data for each pixel is used to store an alpha value representing special effect information (e.g., transparency).

2.3.1 Image formats

GIF

- **Graphics Interchange Forma**t (GIF) devised by CompuServe initially for transmitting graphical images over phone lines via modems.
- Limited to only 8-bit (256) color images
- Supports animation
- GIF actually comes in two flavors:

1. **GIF87a**: The original specification.
2. **GIF89a**: The later version. Supports simple animation via a Graphics Control Extension block in the data, provides simple control over delay time, a transparency index, etc.

GIF 87a vs GIF89a

- GIF87a is the original format for indexed color images. It uses LZW compression and has the option for being **interlaced**.
- GIF89a is the same, but also includes transparency and animation capabilities.
- Interlaced (of video image) means scanned in such a way that alternate lines form one sequence which is followed by the other lines in a second sequence.

PNG

- Stands for **Portable Network Graphics**
- It is intended as a replacement for GIF in the WWW and image editing tools.
- PNG uses unpatented zip technology for compression
- PNG-24 is another version of PNG, with 24-bit color support, allowing ranges of color to a high color JPG

JPEG/JPG

- A standard for photographic image compression
- Created by the **Joint Photographic Experts Group**
- Intended for encoding and compression of photographs and similar images
- Takes advantage of limitations in the human vision system to achieve high rates of compression
- Uses complex lossy compression which allows user to set the desired level of quality (compression). A compression setting of about 60% will result in the optimum balance of quality and file size.
- Though JPGs can be interlaced, they do not support animation and transparency unlike GIF

TIFF

- Stands for **Tagged Image File Format**.
- The support for attachment of additional information (referred to as “tags”) provides a great deal of flexibility.

1. The most important tag is a format signifier: what type of compression etc. is in use in the stored image.
2. TIFF can store many different types of image: 1-bit, grayscale, 8-bit color, 24-bit RGB, etc.
3. TIFF was originally a lossless format but now a new JPEG tag allows one to opt for JPEG compression.
4. The TIFF format was developed by the Aldus Corporation in the 1980's and was later supported by Microsoft.

EXIF

- **Exchange Image File** is an image format for digital cameras:
  1. Compressed EXIF files use the baseline JPEG format.
  2. A variety of tags (many more than in TIFF) are available to facilitate **higher quality printing,** since information about the camera and picture-taking conditions (flash, exposure, light source, white balance, type of scene, etc.) can be stored and used by printers for possible color correction algorithms.
  3. The EXIF standard also includes specification of file format for audio that accompanies digital images. As well, it also supports tags for information needed for conversion to FlashPix (initially developed by Kodak).

2.3.2 Color model in images

• Colors models and spaces used for stored, displayed, and printed images.

RGB Color Model for CRT Displays

- We expect to be able to use 8 bits per color channel for color that is accurate enough.
- However, in fact we have to use about 12 bits per channel to avoid an aliasing effect in dark image areas — contour bands that result from gamma correction.
- RGB color model is an additive color model in which red, green and blue light are added together in various ways to reproduce a broad array of colors.(3 additive primary colors: Red, Green and Blue)

Subtractive color: CMY color Model

- So far, we have effectively been dealing only with **additive color**. Namely, when two light beams impinge on a target, their colors add; when two phosphors on a CRT screen are turned on, their colors add.
- But for ink deposited on paper, the opposite situation holds: yellow ink *subtracts* blue from white illumination, but reflects red and green; it appears yellow.
- Remember that Cyan = Green + Blue, so light reflected from a cyan pigment has no red component, i.e., the red is absorbed by cyan. Similarly magenta subtracts green and yellow subtracts blue.

1. Instead of red, green, and blue primaries, we need primaries that amount to -red, -green, and -blue. I.e., we need to *subtract* R, or G, or B.
2. These subtractive color primaries are Cyan (*C*), Magenta (*M*) and Yellow (*Y*) inks.

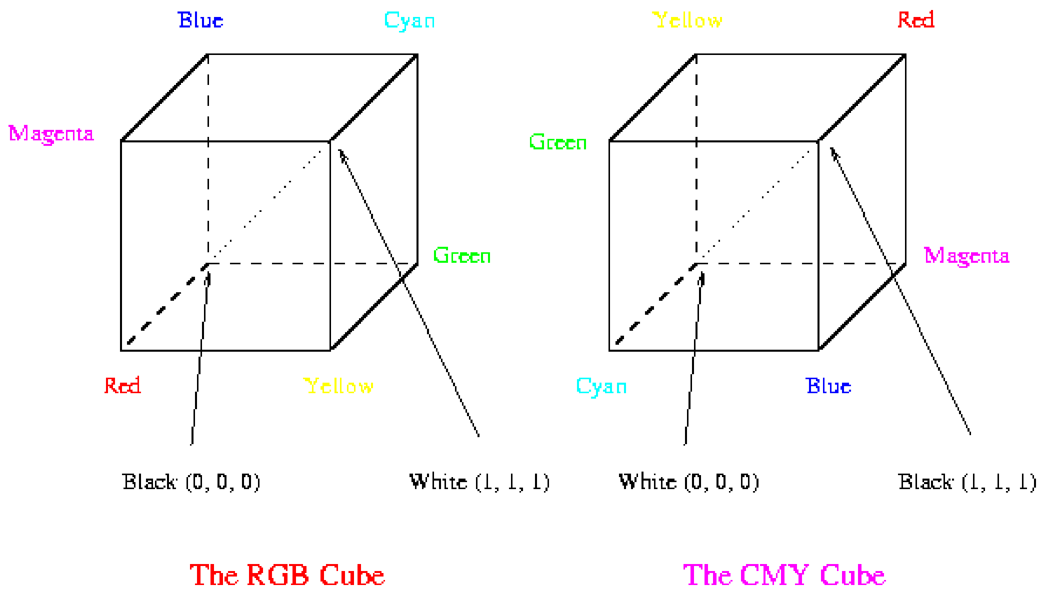

Fig 2.5: RGB and CMY color cubes

Undercolor Removal: CMYK System

- Printers generally use four colors: cyan, yellow, magenta and black. This is because Cyan, Yellow and Magenta produce a dark gray rather than a true black.
- **Undercolor removal**: calculate that part of the CMY mix that would be black, remove it from the color proportions, and add it back as real black.
- The new specification of inks is thus:

$$K \equiv min\{C, M, Y\}$$

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} \Rightarrow \begin{bmatrix} C-K \\ M-K \\ Y-K \end{bmatrix}$$

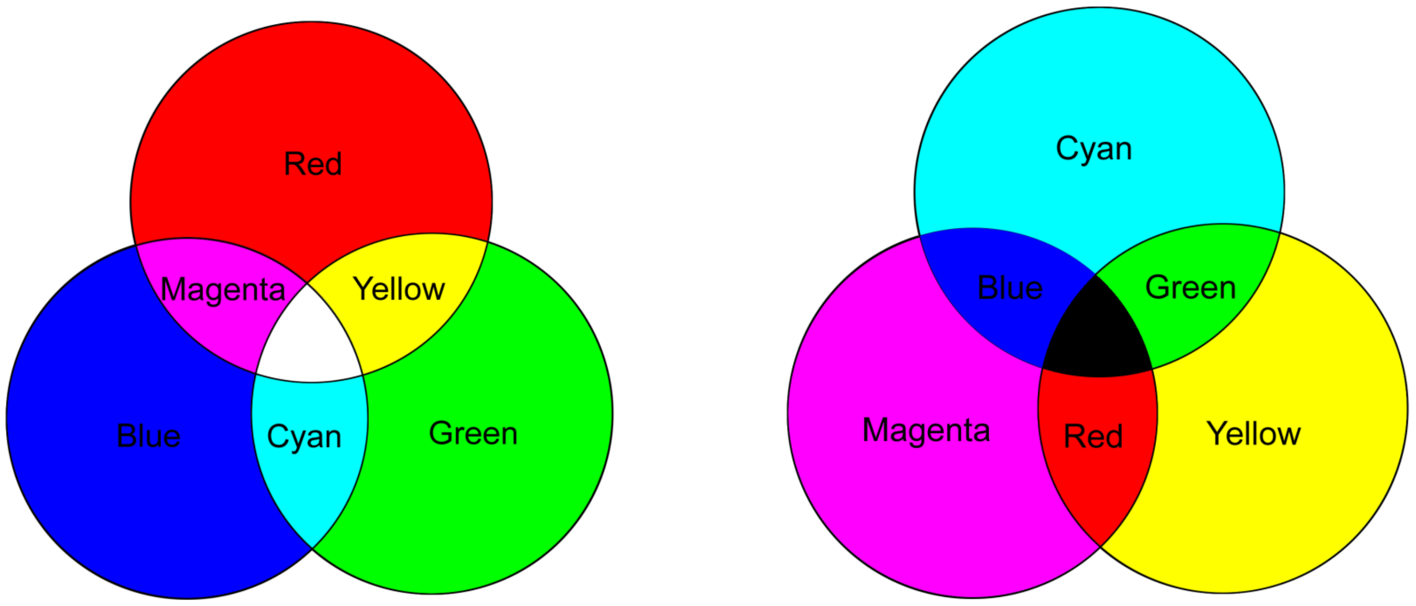

Fig 2.6: Additive and subtractive color.
(a): RGB is used to specify additive color.
(b): CMY is used to specify subtractive color

2.4 Video formats and color models

2.4.1 Video formats

The AVI Format

- The AVI (Audio Video Interleave) format was developed by Microsoft.
- The AVI format is supported by all computers running Windows, and by all the most popular web browsers. It is a very common format on the Internet, but not always possible to play on non-Windows computers.
- Videos stored in the AVI format have the extension .avi.

The Windows Media Format

- The Windows Media format is developed by Microsoft.
- Windows Media is a common format on the Internet, but Windows Media movies cannot be played on non-Windows computer without an extra (free) component installed. Some later Windows Media movies cannot play at all on non-Windows computers because no player is available.
- Videos stored in the Windows Media format have the extension .wmv.

The MPEG Format

- The MPEG (Moving Pictures Expert Group) format is the most popular format on the Internet. It is cross-platform, and supported by all the most popular web browsers.
- Videos stored in the MPEG format have the extension .mpg or .mpeg.

The QuickTime Format

- The QuickTime format is developed by Apple.
- QuickTime is a common format on the Internet, but QuickTime movies cannot be played on a Windows computer without an extra (free) component installed.
- Videos stored in the QuickTime format have the extension .mov.

The RealVideo Format

- The RealVideo format was developed for the Internet by Real Media.
- The format allows streaming of video (on-line video, Internet TV) with low bandwidths. Because of the low bandwidth priority, quality is often reduced.
- Videos stored in the RealVideo format have the extension .rm or .ram.

The Shockwave (Flash) Format

- The Shockwave format was developed by Macromedia.
- The Shockwave format requires an extra component to play. This component comes preinstalled with the latest versions of Netscape and Internet Explorer.
- Videos stored in the Shockwave format have the extension .swf.

2.4.2 Color models in video

- **Video Color Transforms**
  a) Largely derive from older analog methods of coding color for TV. Luminance is separated from color information.
  b) For example, a matrix transform method similar to called YIQ is used to transmit TV signals in North America and Japan.
  c) This coding also makes its way into VHS video tape coding in these countries since video tape technologies also use YIQ.
  d) In Europe, video tape uses the PAL or SECAM coding, which are based on TV that uses a matrix transform called YUV.
  e) Finally, digital video mostly uses a matrix transform called YCbCr that is closely related to YUV

YUV color model

- YUV Color Model represents the human perception of color more closely than the standard RGB model used in computer graphics hardware. In YUV, Y is the luminance(brightness) component while U and V are the chrominance(color) components.
    a) YUV codes a luminance signal (for gamma-corrected signals) equal to Y ′ .
    b) **Chrominance** refers to the difference between a color and a reference white at the same luminance. → use color differences *U, V*:

$$U = B' - Y' , \qquad V = R' - Y'$$

$$\begin{bmatrix} Y' \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

    c) For gray, $R' = G' = B'$, the luminance Y' equals to that gray, since 0.299+0.587+0.114 = 1.0. And for a gray (“black and white”) image, the chrominance (U, V) is zero.

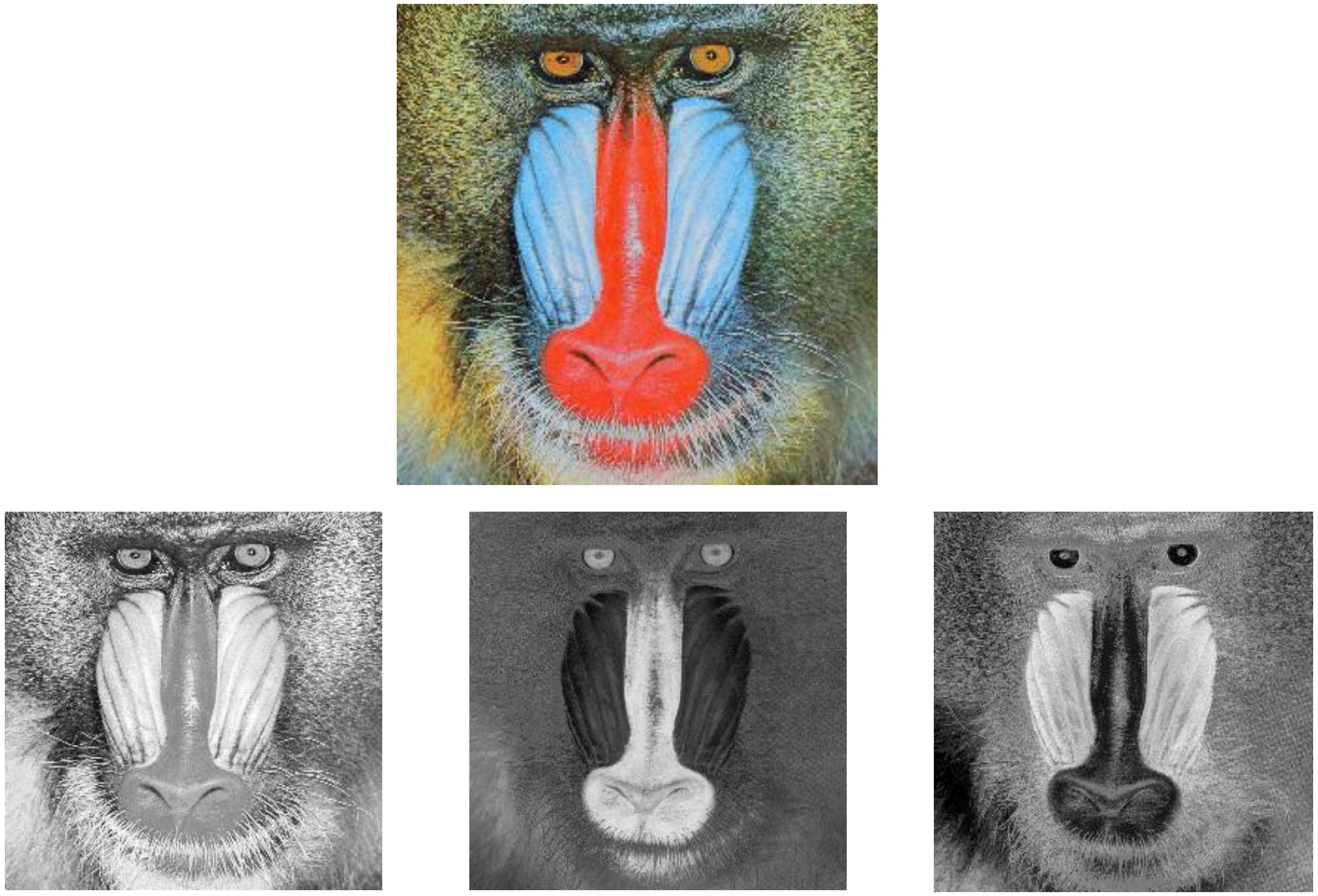

Fig 2.7: Y ′UV decomposition of color image.
Top image (a) is original color image;
(b) is Y ′; (c,d) are (U, V)

YIQ color model

- YIQ is used in NTSC color TV broadcasting. Again, gray pixels generate zero (I, Q) chrominance signal.
  - (a) I and Q are a rotated version of U and V .
  - (b) Y' in YIQ is the same as in YUV; U and V are rotated by 33°:

$$I = 0.492111(R' - Y') \cos 33° - 0.877283(B' - Y') \sin 33°$$

$$Q = 0.492111(R' - Y') \sin 33° + 0.877283(B' - Y') \cos 33°$$

  - (c) This leads to the following matrix transform:

$$\begin{bmatrix} Y' \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.595879 & -0.274133 & -0.321746 \\ 0.211205 & -0.523083 & 0.311878 \end{bmatrix} = \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

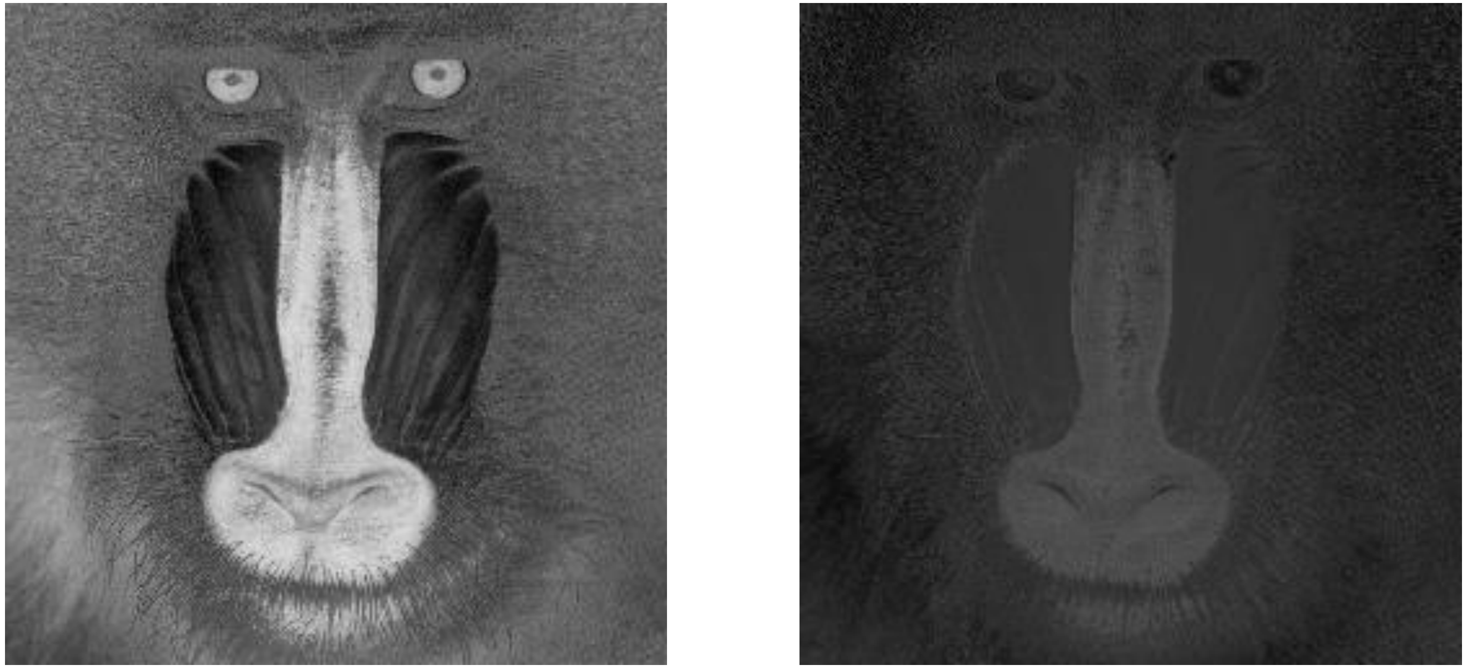

Fig 2.8: *I* and *Q* components of color image

YCbCr Color Model

- The Rec. 601 standard for digital video uses another color space, YC_bC_r, often simply written YCbCr — closely related to the YUV transform.
  a) YUV is changed by scaling such that C_b is U, but with a coefficient of 0.5 multiplying B'. In some software systems, C_b and C_r are also shifted such that values are between 0 and 1.
  b) This makes the equations as follows:

$$C_b = ((B' - Y')/1.772)+0.5$$
$$C_r = ((R' - Y')/1.402)+0.5$$

  c) Written out:

$$\begin{bmatrix} Y' \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.168736 & -0.331264 & 0.5 \\ 0.5 & -0.418688 & -0.081312 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} + \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix}$$

d) In practice, however, Recommendation 601 specifies 8-bit coding, with a maximum Y' value of only 219, and a minimum of +16. Cb and Cr have a range of ±112 and offset of +128. If R', G', B' are floats in [0.. + 1], then we obtain Y', Cb, Cr in [0..255] via the transform:

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

e) The YCbCr transform is used in JPEG image compression and MPEG video compression.