

Introduction to YUV (YCbCr) Format

1. Relationship Between YUV and YCbCr

Just as coordinate spaces describe sets of coordinates in geometry, color spaces use mathematical methods to describe sets of colors. The three fundamental color models are RGB, CMYK, and YUV.

YCbCr was developed as part of the ITU-R BT.601 recommendation during the formulation of world digital video standards. It is essentially a scaled and offset version of YUV. The Y component in YCbCr has the same meaning as in YUV, while Cb and Cr also represent color but differ in their representation method. Within the YUV family, YCbCr is the most widely used member in computer systems, with extensive applications in domains such as JPEG and MPEG, both of which adopt this format. Generally, when people refer to YUV, they mostly mean YCbCr. YCbCr has several sampling formats, such as 4:4:4, 4:2:2, 4:1:1, and 4:2:0.

1.1 YUV

YUV is a color encoding method commonly used in various video processing components. When encoding photos or video, YUV allows for reduced chroma bandwidth by leveraging human perceptual capabilities.

YUV encompasses several types of true-color color spaces. Proprietary terms like Y'UV, YUV, YCbCr, YPbPr can all be referred to as YUV, with overlapping meanings. "Y" represents luminance (or luma), corresponding to the grayscale value. "U" and "V" represent chrominance (or chroma), describing image color and saturation to specify pixel color.

The scopes referred to by Y'UV, YUV, YCbCr, YPbPr are often confused or overlapping. Historically, YUV and Y'UV were typically used to encode analog television signals, while YCbCr described digital video signals suitable for video/image compression and transmission (e.g., MPEG, JPEG). Today, however, YUV is widely used in computer systems.

1.2 YCbCr

In YCbCr, Y is the luminance component, Cb is the blue chrominance component, and Cr is the red chrominance component. The human eye is more sensitive to the Y component of video. Therefore, by subsampling the chroma components to reduce their data, changes in image quality become imperceptible to the human eye. The primary subsampling formats are YCbCr 4:2:0, YCbCr 4:2:2, and YCbCr 4:4:4.

4:2:0: Indicates 4 luminance components and 2 chrominance components (YYYYCbCr) per 4 pixels, sampling only odd scanlines. It is the most common format for portable video devices (MPEG-4) and video conferencing (H.263). 4:2:2: Indicates 4 luminance components and 4 chrominance components (YYYYCbCrCbCr) per 4 pixels. It is the most common format for DVDs, digital TV, HDTV, and other consumer video devices. 4:4:4: Represents full pixel dot matrix (YYYYCbCrCbCrCbCrCbCr), used for high-quality video applications, studios, and professional video products.

2. Main Sampling Formats

The main sampling formats are YCbCr 4:2:0, YCbCr 4:2:2, YCbCr 4:1:1, and YCbCr 4:4:4. Among these, YCbCr 4:1:1 is relatively common. Its meaning is: Each point stores an 8-bit luminance value (Y value), while every 2x2 points store one Cr and one Cb value. The visual perception of the image remains largely unchanged. Originally, using the RGB model (R, G, B all 8-bit unsigned), each point requires $8 \times 3 = 24$ bits. With this method, only $8 + (8/4) + (8/4) = 12$ bits are needed, averaging 12 bits per point. This achieves data compression of the image by half.

The above gives a theoretical example; actual data storage may differ. Below are several specific storage patterns:

1. YUV 4:4:4

The sampling rates for all three YUV channels are identical. Thus, each pixel in the generated image contains complete information for all three components (each component typically 8 bits). After 8-bit quantization, each uncompressed pixel occupies 3 bytes.

Four pixels: [Y0 U0 V0] [Y1 U1 V1] [Y2 U2 V2] [Y3 U3 V3]

Stored bitstream: Y0 U0 V0 Y1 U1 V1 Y2 U2 V2 Y3 U3 V3

2. YUV 4:2:2

The sampling rate for each chrominance channel is half that of the luminance channel. Thus, the horizontal chroma sampling rate is only half of 4:4:4. For uncompressed 8-bit quantized images, each macropixel consisting of two horizontally adjacent pixels requires 4 bytes of memory.

Four pixels: [Y0 U0 V0] [Y1 U1 V1] [Y2 U2 V2] [Y3 U3 V3]

Stored bitstream: Y0 U0 Y1 V1 Y2 U2 Y3 V3

Mapped pixels: [Y0 U0 V1] [Y1 U0 V1] [Y2 U2 V3] [Y3 U2 V3]

3. YUV 4:1:1

4:1:1 chroma sampling performs 4:1 horizontal chroma subsampling. This remains acceptable for low-end users and consumer products. For uncompressed 8-bit quantized video, each macropixel consisting of 4 horizontally adjacent pixels requires 6 bytes of memory.

Four pixels: [Y0 U0 V0] [Y1 U1 V1] [Y2 U2 V2] [Y3 U3 V3]

Stored bitstream: Y0 U0 Y1 Y2 V2 Y3

Mapped pixels: [Y0 U0 V2] [Y1 U0 V2] [Y2 U0 V2] [Y3 U0 V2]

4. YUV 4:2:0

4:2:0 does not mean only Y and Cb without Cr. It means that for each scan line, only one chroma component is stored at a 2:1 subsampling rate. Adjacent scan lines store different chroma components. If one line is 4:2:0, the next line is 4:0:2, the next 4:2:0, and so on. For each chroma component, both horizontal and vertical sampling rates are 2:1, resulting in an effective chroma sampling rate of 4:1. For uncompressed 8-bit quantized video, each macropixel consisting of a 2x2 block (2 rows, 2 columns) of adjacent pixels requires 6 bytes of memory.

Eight pixels: [Y0 U0 V0] [Y1 U1 V1] [Y2 U2 V2] [Y3 U3 V3]

[Y5 U5 V5] [Y6 U6 V6] [Y7 U7 V7] [Y8 U8 V8]

Stored bitstream: Y0 U0 Y1 Y2 U2 Y3

Y5 V5 Y6 Y7 V7 Y8

Mapped pixels: [Y0 U0 V5] [Y1 U0 V5] [Y2 U2 V7] [Y3 U2 V7]

[Y5 U0 V5] [Y6 U0 V5] [Y7 U2 V7] [Y8 U2 V7]