Chrominance Subsampling in Digital Images

Douglas A. Kerr, P.E.

Issue 1
November 2, 2005

ABSTRACT

The JPEG and TIFF digital still image formats, along with various digital video formats, have provision for recording the chrominance information (which conveys in a special way what the lay person would describe as the "color" of the pixels) in a resolution lower than that of the image being encoded. This concept, followed for over half a century in television broadcasting, takes advantage of the properties of the human perceptual system to reduce the amount of data required to convey an acceptable full-color image. There are various standard "patterns" for performing this "chrominance subsampling", and a curious and confusing notation for indicating them. In this article we discuss the concept of chrominance subsampling and describe this system of notation.

BACKGROUND

The color space

A digital image that is to be encoded using the JPEG image data coding and compression system, one form of the TIFF image coding system, and various digital video formats is first put into what is called a luma-chrominance color space. In that form, the color of a pixel is described by two values, one very nearly (but not quite) describing its luminance (brightness), and one describing what a lay person would think of as its "color". The latter is in the form of chrominance, a slightly different concept from the basic color science concept of chromaticity, but we need not concern ourselves here with the distinction. The metric for chrominance is, as we might expect, two-dimensional in the mathematical sense: two numerical values are actually required to express it (a total of three values for the color).

As mentioned just above, the first value in this scheme does not actually describe the luminance of the pixel's color. As a result, it is sometimes called "luma", a term borrowed from the analog system used for television signals. This term is a clue that the value does not quite describe luminance, and also tips us off to its nonlinear form. And in fact, the value pair giving the chrominance is also sometimes called "chroma", again primarily to tip us off to its nonlinear form. But here we will use the term *chrominance*, as it best matches normal editorial practice for the topic area we are considering.

Thus, for each pixel, there are three numerical values that collectively describe its color. They are identified as Y, Cb, and Cr. Y is the luma value, and Cb and Cr collectively form the chrominance value.

Chrominance subsampling

During the early work on color television systems (analog, of course), note was taken of the fact that the human eye is able to discern finer detail conveyed by differences in luminance than for detail conveyed by differences in chromaticity. The encoding scheme adopted there separately conveys the luminance-related value *luma* and the chromaticity-related value *chroma* (chrominance) over "subchannels" having different bandwidth (and thus supporting different levels of resolution)—the chrominance subchannel having reduced resolution capabilities. The result was a system that well matched human perceptual response, allowing the conveyance of quality images with less overall bandwidth requirement than if equal bandwidth were allocated to luma and chrominance information.

Not surprisingly, the developers of systems for the encoding of digital still images decided to exploit this same consideration to get the "biggest bang for the bit" in digital images being prepared for transmission or storage. There, the process is called *chrominance subsampling*.

Simply stated, here is the principle. We include in the digital data stream to be encoded by the JPEG system the luma value (Y) for each pixel in the image. But we only include a single Cb+Cr pair (a "chrominance value", often described as a *chrominance sample*) for a group of pixels—which in the schemes generally recognized can comprise 2, 4, or even 8 image pixels. Thus the data load for the chrominance information—which otherwise would be twice that for the luma information (Y, Cb, and Cr are all recorded in the same number of bits, usually 8)—is now reduced by a factor of 2, 4, or even 8.

In fact, it is really useful to think of this in terms of the chrominance being given for "chrominance pixels" which are 2, 4, or even 8 times the size of the image pixels.

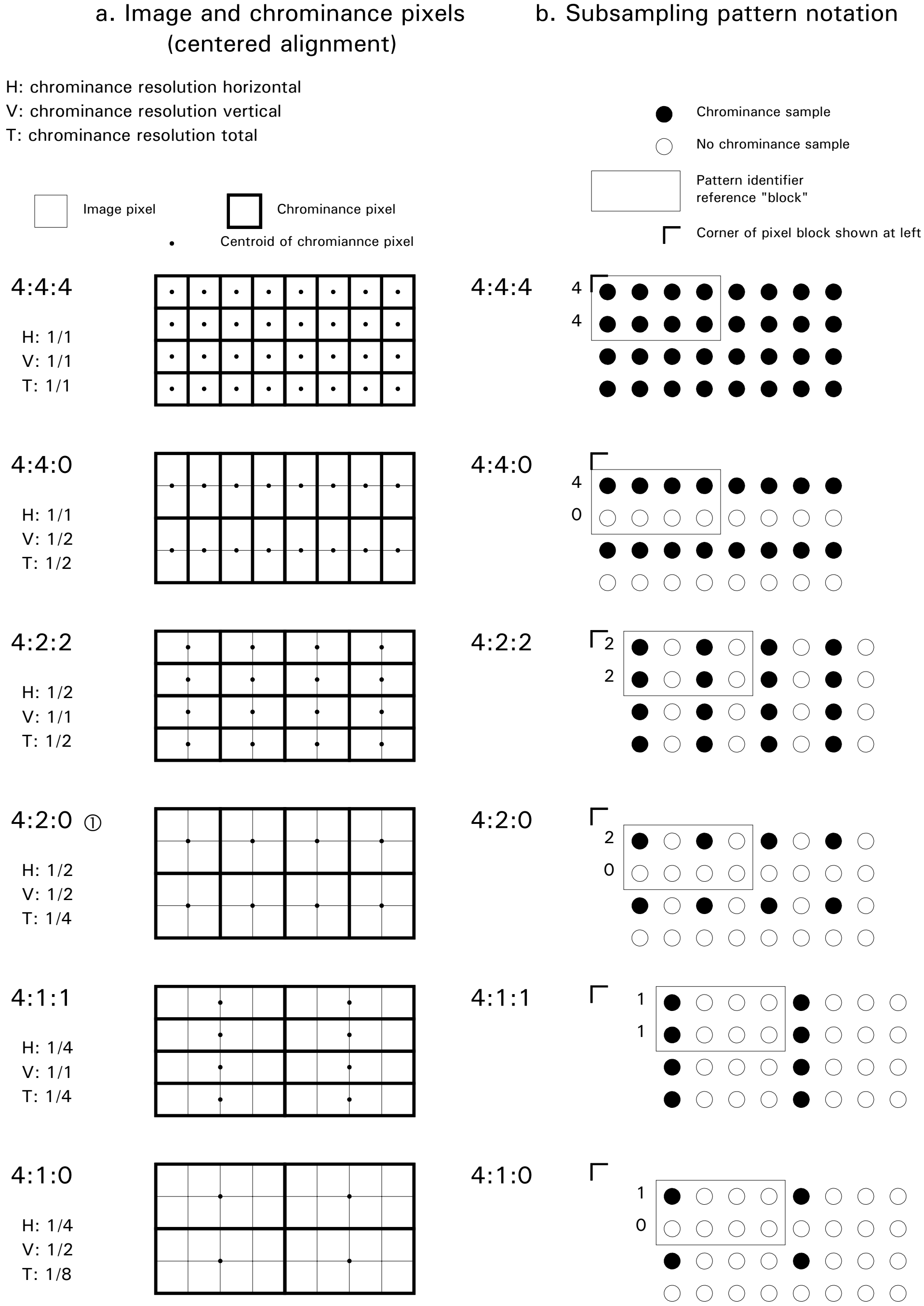
"Siting" of the chrominance samples

This now leads to another issue. Suppose were are using a pattern in which the "chrominance pixel" is twice as wide and twice as high as an image pixel. Should its "centroid" be at the center of an image pixel, or should it be at the center of the group of four image pixels? In fact, there can be advantages to each, and both possibilities are potentially available for each subsampling pattern. We'll hear more about that later.

Subsampling pattern notation

A system of notation is used for describing a family of widely-used chrominance subsampling patterns. It uses indicators that look like this: 4:2:2. It has a very curious basis, and is widely misunderstood (and often misused). We will defer discussing its definition until we have a more thorough understanding of what the patterns are like. But for convenience we nevertheless will use it to identify the patterns to be shown shortly.

Figure 1.



① This is the most common "centered" form for 4:2:2 for still images; others are used in video

COMPARISON OF SUBSAMPLING PATTERNS

Figure 1 shows, in part **a**, six chrominance subsampling patterns (actually, the first one is no subsampling at all), including all the ones widely used in common image encoding schemes (including a popular standard form of image file using the JPEG coding and compression system). These patterns are identified by the notation just described (even though we haven't discussed how it works—that will come shortly).

Each example shows a portion of the original image 8 pixels wide and 4 pixels high, and indicates (with heavy lines) the boundaries of the "chrominance pixels". The luminance of all the image pixels covered by each chrominance pixel is averaged and included (as a pair of Cb and Cr values) in the image data for the chrominance pixel. The dots show the centroids of these chrominance pixels, and also help us do a visual "head count" of the chrominance values. Note that all these examples show the "centered" alignment: the chrominance pixels embrace a set of integral image pixels. As a result, the centroids of the chrominance pixels are located in the center of the set of the centroids of the associated luminance pixels.

Under the indicator for the pattern (*e.g.*, 4:4:4) we show how the resolution of the chrominance pixels compares to the resolution of the image itself. The H value is the relative resolution in the horizontal direction, the V value is the relative resolution in the vertical direction, and the T ("total") value is the relative resolution in terms of pixel count (sometimes called the "areal" resolution), all as fractions.

Note that each image pixel gets a luma value (luma sample). In most writings about this matter, resolution comparisons are made between the "chrominance samples" and "luma samples", rather than between the "chrominance pixels" and "image pixels", as we do here. And often the "ratio" is described other-side up as a sampling factor—a sampling factor of "4" in the horizontal or vertical direction means a resolution of 1/4 the image (or luma) resolution.

The first pattern shown (4:4:4) is in fact the case where there is really no chrominance subsampling at all—every image pixel has its chrominance value included.

There are two patterns (4:4:0 and 4:2:2) which have chrominance pixels twice the size of image pixels (T:1/2). In the first of these the (rectangular) chrominance pixels are vertically-oriented, and in the other, horizontally-oriented.

There are two patterns (4:2:0 and 4:1:1) which have chrominance pixels four times the size of image pixels (T:1/4). In the first of these the chrominance pixels are square, and in the other, rectangular and horizontally-oriented.

In the last pattern (4:1:0), the chrominance pixels are eight times the size of the image pixels (T:1/8), and are rectangular and horizontally-oriented.

Note that the specification for the kind of JPEG image file used today by most digital still cameras (the JPEG Exif file), only two of these patterns are allowed: 4:2:2 and 4:2:0.[1]

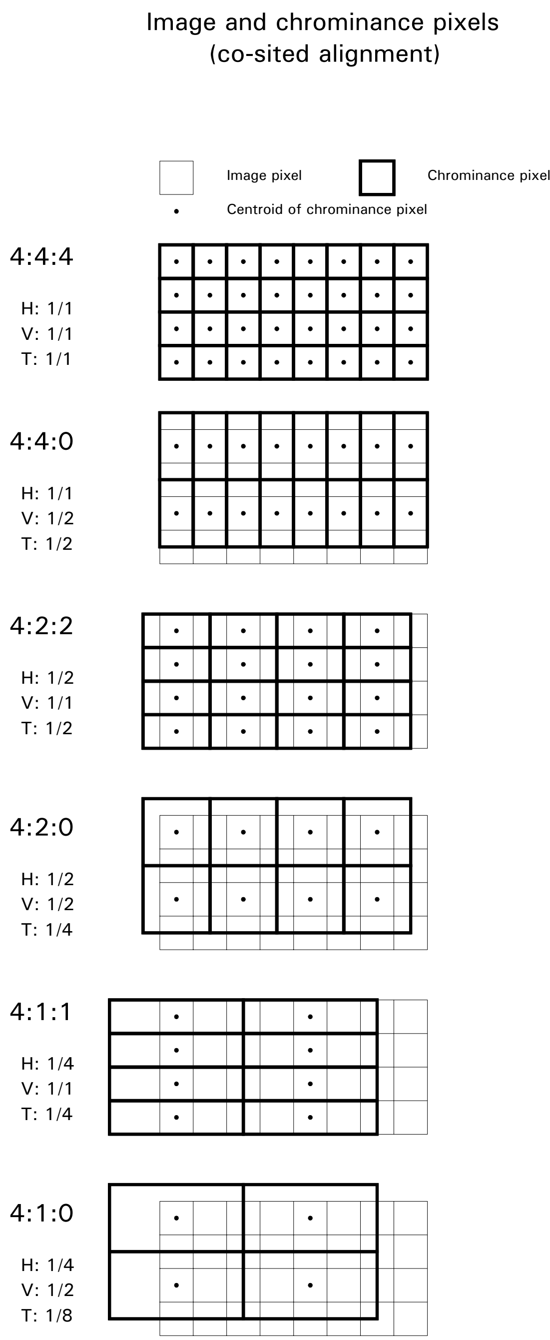
---

[1] The 4:2:0 scheme is often incorrectly identified as "4:1:1". The origin of this widespread error is not known to me.

Chrominance pixel alignment

These examples all show the arrangement when the chrominance pixel actually embraces a number of full image pixels (known as the "centered" alignment). This is the arrangement generally used for digital still camera images. In figure 2, we see one form of the other alternative (the "co-sited" alignment).

Figure 2.



SUBSAMPLING PATTERN NOTATION

Now we are ready to tackle the notation used to indicate these various subsampling patterns. We can follow the action on part b of figure 2. The scheme indicator is of the form J:a:b. The notation revolves around the concept of a "reference block"—a conceptual region J pixel spacings wide and 2 pixel spacings high. (For all schemes we encounter, J is 4.) It is not necessarily aligned with the grid of image pixels. The small chevron at the upper left of each reference block shows the relative location of the upper left corner of the block of image pixels as shown to the left.

The dots in the figure (white and black) represent the chrominance samples (each recorded as a Cb value plus a Cr value) **that would exist if there were no subsampling.** The black dots show the chrominance values that actually exist for this scheme.

Note that, if we consider our reference block, the indicator value **a** shows the number of chrominance samples actually present with respect to the top row of the block; the indicator value **b** shows the number of chrominance subsamples actually present with respect to the bottom row of the block. We see that emphasized by the little figures to the left of the reference block.

Note that there is a one-to-one correspondence between the black dots in part b of the figure and the little black dots indicating the centroids of the chrominance pixels in part a of the figure.

Note that the 4:2:2 pattern could as well have been designated "2:1:1", as its thrust is to convey relative sampling frequencies. However, for patterns where the ratios involve only the numbers 1, 2, and/or 4, it is customary to make J=4. (There are patterns, used in some specialized video systems, in which J is 3, thus accommodating a 1/3 resolution chrominance subsampling.)

Misunderstandings

Not surprisingly, this peculiar system of notation has been subject to some misunderstandings, unfortunately widespread. We will mention three of them here.

The meaning of **a** and **b** in the "J:a:b" notation

Often, especially in the area of digital video work, we hear the subsampling pattern notation system described this way:

> "The first number gives the number of luma samples that we consider. The second number gives the number of Cb values over that span, and the third number gives the number of Cr values over that span."

This is generally followed by something like this:

> "Notations such as 4:2:0 do not follow the rule." (No kidding!)

Note that the erroneous definition does in fact appear true when **a=b.**

4:2:0 vs. 4:1:1

Very commonly, the 4:2:0 pattern is erroneously described as “4:1:1”. The author has not been able to track down the origin of this error.

This error is found in many image editing packages offering the opportunity to select different subsampling patterns when an image is saved in JPEG form.

U and V vs. Cb and Cr

This is not really an error, but a matter of editorial practice. It can however be confusing in following the literature.

Often we will hear the Cb and Cr values described as **U** and **V**.

U and V are the coordinates of the color space YUV color space which underlies the YCbCr color space. Cb and Cr are the quantized digital representations of the U and V values of a color in the YUV color space. Thus it may be reasonable to speak, conceptually, of the chrominance of a pixel itself in terms of U and V, or of a chrominance **sample** as comprising U and V values. However, in a digital image context, it is more useful to make reference to Cb and Cr (which is how the values are designated in the actual digital image data).

Sampling factor notation

There is a different system of expressing a compression pattern in terms of three “sampling factors” for each direction (vertical and horizontal). Unfortunately, a set of three such factors is sometimes written in this form: y:b:r (thus inviting confusion with the J:a:b notation discussed in this article).

The three factors tell the relative frequency of occurrence (resolution) of Y, Cb, and Cr samples/values in the particular direction. For example, for the 4:2:0 pattern, the factor set would be:

Horizontal: 2,1,1 (sometimes written 2:1:1)

Vertical: 2,1,1

The relative frequency for each component (Y, Cb, or Cr) is a fraction with the number given as the numerator and the largest of all the numbers as the denominator. (The set is in fact scaled so that the largest number is in fact the least common denominator of all the fractions.) In most cases, the “Y” number will be the largest, as the fraction for “Y” is ordinarily expected to be “1”.

Thus, the above set of values would mean:

> In the horizontal direction, the Y samples occur at “full frequency” (one per pixel) and the Cb and Cr samples occur at “half frequency” (one per two pixels; this is of course the same as two per 4 pixels in the model for the J,a,b notation for this pattern, 4:2:0).

In the vertical direction, the Y samples occur at "full frequency" (one per pixel "row") and the Cb and Cr samples occur at "half frequency" (one per two pixel "rows").

Sometimes this scheme of notation is written this way,

2/2 1/1 1/1

where the three pairs refer to the Y, Cb, and Cr relative frequencies, respectively, and the first number in the pair pertains to the horizontal axis and the second to the vertical axis. (The pairs are thus not fractions, even though they at first look as if they might be.)

For the 4:2:2 pattern, this system of notation would look like this,

Horizontal: 2,1,1

Vertical: 1,1,1

or 2/1 1/1 1/1

For the 4:1:1 pattern, this system of notation would look like this,

Horizontal: 4,1,1

Vertical: 1,1,1

or 4/1 1/1 1/1

DATA PACKING

Although it is not part of the real topic of this article, an interesting related matter is the way in which the Y, Cb, and Cr values for an image are arranged as a data stream, perhaps for presentation to the software routines that encode the ensemble of data into JPEG or TIFF form (a matter often called *data packing*). For each subsampling pattern, there may be several standardized data packing arrangements. Just to give some insight into this, we show on figure 3 a common data packing arrangement for the 4:2:0 subsampling pattern.

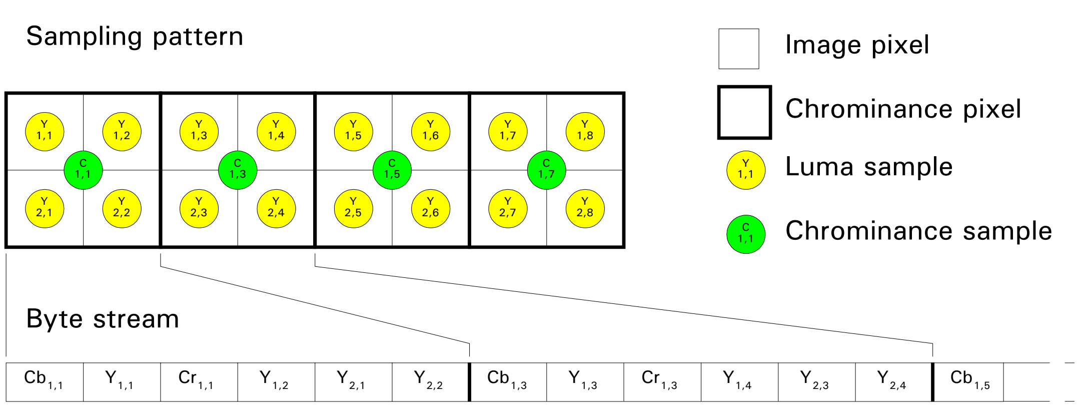


Figure 3. Data packing for 4:2:0 subsampling

Here we see a block of image pixels 8 pixels wide and two pixels high. We show it divided into chrominance pixels 2 x 2 image pixels in size, in the way intimated by the “centered” form of the 4:2:0 subsampling pattern. The yellow dots show the centroids for the luma samples, the green dots for the chrominance samples. The indexes for the chrominance samples (and their Cb and Cr values) are those of the nearest luminance sample above and to the left.

The data packing arrangement operates on an entire chrominance pixel at a time and then moves to the next chrominance pixel; it does not operate on rows of image pixels. The four Y values (one for each image pixel) and the Cb and Cr values (for the chrominance pixel) are placed in the byte stream as shown.

A word of caution: especially for other subsampling patterns, there are data packing arrangements which seem to follow a similar principle regarding the placement of Cb and Cr but in which their order is opposite that shown here, perhaps the idea being to more closely match the sequence R, (G), B.

The calculation of the analog quantities U and V underlying Cb and Cr involve B and R, respectively, thus the notation Cb and Cr. The reason the color space is called YCbCr (rather than YCrCb) is because of the natural order of U and V.

A UNIQUE VARIANT

The “DV” digital video standard, in its “European” (PAL-compatible) version, uses a unique form of the 4:2:0 subsampling pattern. It is shown in figure 4.

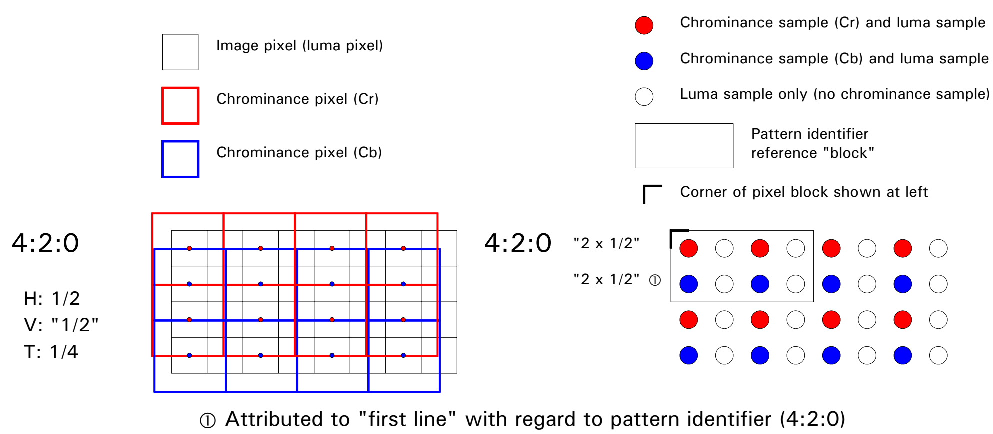


Figure 4. DV-PAL subsampling pattern

The unique feature of this pattern is that the Cb and Cr values are not associated with the same location on the image; that is, to use our notation, with the same chrominance pixel.

If in fact the chrominance values are derived from true chrominance pixels (that is, as an average of the chrominance over several image pixels), it probably has to be done as a weighted average over nine image pixels (all of which fall, at least in part, within the chrominance pixel). The figure shows the chrominance pixels based on that concept.

However, evidently the standard for this subsampling pattern does not prescribe just how that is to be done.

Of course, associating a J:a:b identifier with this subsampling pattern requires a little creativity; the notation system doesn't really apply cleanly there. Officially, it is given the identifier 4:2:0. The right hand part of the figure offers a fanciful rationale for that.

A DOSE OF REALITY

In order to most clearly illustrate the concepts and principles involved, I have spoken in terms if "chrominance pixels" and have intimated that the chrominance values are in fact determined over these (by some appropriate type of averaging of the their chrominance values.

But that is not always done. In some case, a more primitive means of determining what chrominance to "send" is used. In the worst case, the chrominance of one image pixel is snagged and transmitted on behalf of the chrominance pixel.

In any event, what happens at the "receiving" end? There, decoding the YCbCr data stream (which does not contain Cb and Cr values for every pixel) is expected to produce a Y, Cb, and Cr value for **every** image pixel. From those values, we derive an RGB representation of every pixel for further handling.

Ideally, this would be done by interpolation between the transmitted chrominance samples. But that's not always done. For example, in many video systems (especially those using a co-sited arrangement of chrominance pixel centroids), the value of a received chrominance sample (one Cb,Cr pair) is used for the reconstruction of several image pixels (four pixels if we imagine a 4:1:1 subsampling pattern).

This typically results in the following:

- The chromaticity of the resulting image will seem to be applied in "blobs", rather than changing smoothly as we move across an object.

- The chromaticity will seem to be shifted to pixels to the right compared to the luminance (by two image pixels in the example of 4:1:1).

#