Mosaic Sensor Arrays in Digital Photography

Douglas A. Kerr, P.E.

Issue 1
August 4, 2003

ABSTRACT

Many contemporary digital still cameras use a *mosaic sensor array* to develop a digital color image of a scene. In this article we describe this device, the principles of its operation, and its implications on the nature of digital camera images.

INTRODUCTION

Many contemporary digital still cameras use a *mosaic sensor array* to develop a color image of a scene. This device represents a clever compromise from the type of sensor array used in such apparatus as professional video cameras.

COLOR

Before we begin, I need to avert any possible ambiguity in some terminology that will figure heavily in the discussion.

Color, in the formal, technical sense, is a property of visible light that embraces both the properties of *luminance* and *chromaticity*. *Luminance* is the property that describes the brightness of the light. *Chromaticity* is the property that distinguishes, for example, between red and orange (different *hues*) and also between red and pink (different *saturations*).

Unfortunately, *chromaticity* is often called "color", both in general conversation and in many types of technical conversations. One reason is perhaps that it is easier to spell. In technical, conversations, we assume that the context makes it clear whether "color" means *color* or *chromaticity*. Sometimes it doesn't, resulting in tragic misunderstandings.

In this paper, I will generally use each term only with its proper formal meaning. However, I will use "color" in a looser sense when it appears in a well-recognized descriptive phrase.

Color is a "three-dimensional" property (in the mathematical, not geometric, sense)—that is, three different numerical values are needed to describe a particular color.

SAMPLING OF AN OPTICAL IMAGE

In digital imaging, we sample a two-dimensional, spatially-continuous optical image of a scene by capturing its color at a number of sample points, normally evenly spaced across the image in a rectangular array. If the image contains no spatial frequency components (fine detail) at or above half the spatial frequency corresponding to the spacing of the sampling points, and assuming that the capture of the image color at each point is "perfect", then in theory we can, from the set of sample data alone, reconstruct the scene completely and precisely. This is predicted by a two-dimensional application of the Nyquist-Shannon sampling theorem.

THE SENSOR ARRAY

Ideally, we would actually do this by letting the optical image fall on an array of photosensors, one at each sampling point, each of which could determine the color of the light at that point of the image. Typically, each such sensor would be made up of three coincident sub-sensors, each sensitive to a different range of wavelengths within the spectrum of visible light. Often we would characterize those three wavelength ranges as corresponding to the hues *red*, *green*, and *blue*. Thus each three-fold sensor would deliver a set of three values (called R, G, and B) for each point, describing the color of the light there. (Such three-fold sensors are sometimes called *tristimulus* sensors, or "full-color" sensors.)

Later in the process of encoding the image data for recording as a computer file, we change the representation of the color at each sample point. We take the R, B, and B values, multiply each by a certain coefficient (fixed, but different for each of the three), and add the results.[1] This value represents the luminance, or brightness, of the light at the point, and is often given the symbol Y.

We than take the values B and R and subtract Y from them, getting "color difference" values Cb and Cr.[2] The set of three values Y, Cb,

---

[1] We will ignore here the fact that the R, G, and B values are first converted to a non-linear form for various reasons.

[2] Cb and Cr collectively describe a property called *chrominance*.

and Cr describe the color at the point.[3] The three coefficients have been chosen to make the value of Y track well with the human perception of the relative luminance (brightness) of colors of different chromaticities, as has been determined by subjective testing.

THE MOSAIC ARRAY

Until recently arrays of tristimulus sensors have been prohibitively costly for use in even "semiprofessional" digital still cameras. Happily, owing to recent development work, such sensors are now used in some cameras of that class, such as the Foveon sensor used in the Sigma SD9 digital camera. They are expected to emerge in other camera lines in the near future.

However, most digital still cameras have to make do with a compromise solution, known as the *mosaic array*. Here, we again have an array of sensors, one for each sampling point of the image, but these do not comprise three sub-sensors. Over the sensors is placed an array of filters, one for each sensor, each only passing light over a particular band of wavelengths within the spectrum of visible light, again generally corresponding to the hues red, green, and blue. [4]

The three filter types appear in a repeating pattern. Different patterns are used by different manufacturers. One popular pattern, called the Bayer pattern after its developer, repeats in groups of four sensors[5]. Two of the sensors have a green filter, one a red filter, and one a blue filter. Each sensor only delivers a single value, corresponding to the relative luminance of the light reaching it (which is only part of the light from the corresponding image point, owing to the filter).

So how does this tell us the color of the light at any sampling point? It doesn't. Instead, the camera fakes that, making reasoned guesses based on the outputs of several of the sensors around the point of interest. The strategy for the guesses is defined by a *demosaic algorithm* (sometimes called a *demosaicing algorithm*). The process is

---

[3] We do not need a "Cg"; only three values are needed to describe a color. The influence of G is felt through its appearance as a component in calculating Y.

[4] It is common to say that the filters are of three different "colors", but of course that doesn't jibe with the formal definition of *color*, so I'll avoid that usage. Nevertheless, the collection of filters is often called a color filter array (CFA).

[5] Because the Bayer pattern is the most widely used (although hardly universal), it is common to speak of any mosaic array as a "Bayer" array.

sometimes described as *Bayer interpolation* (a term which is not completely apt[6]).

How is this even possible? The algorithm takes account of how luminance and chrominance typically vary in most scenes, and matches this with the actual measured values of R, G, and B from sensors in the neighborhood of each point. Because of the nature of human image perception, getting the best estimate of luminance for each individual point is considered more important that getting the chrominance right for each individual point. It is of course important to get the chrominance right over regions larger than one point, which the algorithm is better able to do.[7]

Since a G value is a better bellwether of luminance than an R or B value (G has the largest coefficient when combining R, G, and B to determine Y), the Bayer pattern provides two G-filtered sensors per set of four sensors.

Although this scheme is very ingenious, and, thanks to the clever work of the algorithm designers, produces amazingly-good results in most cases, it is important to note that **the color of no point on the image is captured by the sensor, nor determined with certainty by the demosaic algorithm.**

PIXELS

You may wonder why I have not used the term "pixel" at all in discussing the sensor array and its outputs. "Pixel" is a coined term for "picture element", based on the traditional photographers' slang term for pictures, "pix". It means just what it says: an element of a picture (but only for one made up of discrete elements, such as digital images—there are no pixels to a traditional photographic image).

The sensors on a digital camera sensor array are not pixels. Nor, in the case of a mosaic sensor, do they deliver the color values of pixels

---

[6] *Interpolation* formally means estimating the value for a data point lying between two points for which we have actual values. In this case, we do not have actual values for the color at **any** sample point, so we are going way beyond the concept of interpolation.

[7] Recognizing this, when an image is encoded in JPEG form, the chromaticity aspect of the image may be carried with a lower resolution that the luminance aspect. A similar approach is used in conventional color TV signals, where the chrominance signal (which conveys chromaticity) is allotted greater channel bandwidth than the luminance signal.

in a completed image. They deliver outputs from which the demosaic algorithm estimates pixel values in the completed image. That completed image may quite properly be described as being composed of pixels.

IMPACT OF THE MOSAIC SENSOR COMPROMISE

The following is excerpted from an excellent article by Mike Chaney, publisher of the well-respected image management and printing program, Qimage. The context is the comparison of the Foveon "full-color" sensor of the Sigma SD-9 camera with a mosaic ("Bayer") sensor.

> In conclusion, some things to look for when comparing Bayer and Foveon sensor images:

- Bayer sensors have a very significantly reduced resolution when resolving detail comprised of mainly red/blue primaries, such as a red sports car with black pin stripes, a blue sweater with red lettering, red soda can with black lettering, etc. In these cases, resolution of the Bayer sensor is reduced to less than 1/4 of its "image" resolution! Black and white details will show the highest resolving power on a Bayer sensor, while saturated color detail will vary greatly. A Foveon sensor is much more consistent, resolving near the full resolution of the images for every color combination.

- Bayer sensors will produce images that are softer and less detailed due to the "smoothing" needed to eliminate artifacts and color distortions.

- Bayer sensors tend to omit chrominance (color) information when sampling high frequency detail. If you look at a picture of a tree that has many small branches with a brick wall behind it for example, you will see that many of the smaller branches "morph" into the color of the bricks in the background. This is because the branches are not wide enough to cover the multiple pixels needed to derive accurate color information on a Bayer sensor. Full color sensors completely eliminate this problem.

- Bayer sensors tend to produce color moiré on high frequency detail.

The complete Chaney paper may be accessed at:

http://www.ddisoftware.com/reviews/sd9-v-bayer

SOMETHING DIFFERENT FROM FUJIFILM

Fujifilm, Ltd. has recently introduced a sensor array for digital still cameras in which the individual sample point sensors are arranged in a staggered fashion, which can be thought of as a traditional row/column arrangement rotated by 45°. This arrangement adds a new wrinkle to the part of the image-processing chain where the demosaic algorithm is applied.

The standards for the coding of digital still images work in terms of pixels assumed to be arrayed in vertical rows and horizontal columns. Any output from a digital camera intended for interchange must follow this layout.

The sample points in a Fuji sensor of the described don't fall into such a pattern. But no matter. Remember, no sensor on a mosaic sensor array gives the color value for a pixel of the delivered image. Each pixel of the delivered image is derived from a complex estimating process involving the values of many sensors in that neighborhood.

If we want the output image pixels to fall in a neat V/H pattern, we just declare them that way. The value of each one will still need to be derived by this estimation process involving many sensor outputs from the array. But, unless the output grid bears some reasonable geometric relationship to the sensor grid, the process would be very messy—there would have to be a unique algorithm for estimating the value of each output pixel.

For the reason just mentioned, the first complete image generated by the demosaic algorithm follows a pixel pattern constructed by taking the sensor locations as a start and then adding pixel locations in the gaps between those. The result is that the output image produced by the demosaic algorithm has twice as many pixels as the array had sensors. In one common application, the Fuji sensor array has about 3 million sensors; the output image developed by the demosaic algorithm has about 6 million pixels.

Since not every user will wish to store a file as large as needed for a 6 megapixel (6 Mpx) image, the camera then provides the option of "downsampling" the image to an output image with a smaller number of pixels, a popular choice being 3 million (3 Mpx). Since the spacing of pixels in a 3 Mpx image is 0.707 of the spacing of pixels in a 6 Mpx image, this can't be done in any trivial—nor even simple—way.[8]

Thus a resampling algorithm must be used. It bases the value of each of the 3 M pixels in the output image on the values of a number of pixels of the original 6 Mpx image in the vicinity of the output pixel being generated. This downsampling process is often spoken of as

---

[8] For comparison, if we wanted to downsample the 6 Mpx image to a size of 1.5 Mpx (twice the pixel pitch), we could just take each block of four pixels (2x2), average the values, give that value to a new pixel in the center of the four original ones, and discard the four original ones.

"interpolation", and here the term applies in a justifiable, although unusual, way.[9]

The emergence of this Fuji approach sparked extensive discussions among camera fans seeking to properly describe the result. These discussions sometimes involved questions such as the following (for which I supply an answer):

- Q: In the 6 Mpx "first image" how many of those pixels really came from points of the scene as observed by the sensor array? [10]
- A: Not a one
- Q: In the 3 Mpx output image how many of the pixels really came from points of the scene as observed by the sensor array?
- A: Not a one
- Q: in the 3 Mpx output image, how many of the pixels came from pixels of the 6 Mpx first image?
- A: Not a one

The motivation behind these questions is generally an interest in how the resolution of the resulting images could be expected to compare with images with comparable pixel counts derived from "non-rotated" sensor arrays. That matter is beyond the scope of this article.

CONCLUSION

The mosaic array is a clever compromise that makes practical the construction of economical color digital still cameras. It cannot, even theoretically, reproduce both luminance and chrominance aspects of a scene with a resolution consistent with the number of sensors in the array.

#

---

[9] We are in fact determining values lying between values of the original points, even though none of those original points are retained.

[10] Often actually asked, "How many of those pixels are real pixels?"