Pixel Resolution in Digital Photography

Douglas A. Kerr, P.E.

Issue 1
November 9, 2004

ABSTRACT

This article gives insight into a number of aspects of the concept of pixel resolution in digital photographic practice. Topics include: What do we mean by resolution, and what is pixel resolution? What is the resolution indicator in a digital image file, and what does it mean? What are resizing, resampling, and interpolation? What do publishers mean by their resolution requirements for submitted digital photographs? What is the difference between a pixel and a dot? How do we accommodate the resolution appetite of a printer?

BACKGROUND

Pixel-oriented images

The preponderance of the images encountered in digital photography are recorded as an array of discrete *picture elements*, or *pixels*[1]. In the formats of interest to us, this array is rectangular and upright. In almost all cases of interest to us, the pixel spacing is the same in both the x (horizontal) and y (vertical) directions (said to be a "square pixel" situation).

The image data describes the color of each pixel. (Color is used here in the full scientific sense, embracing both *luminance* and *chromaticity*.)

Resolution

The term resolution is a noun derived from the verb, *to resolve*. In the sense of interest to us, "to resolve" means to distinguish between two features of a scene. When we speak of the resolution of an imaging system, we mean its quantitative ability to record, as separate, two different closely spaced scene features.

Often testing of the resolution of an imaging system involves a pattern of alternate thin dark and light areas, or lines. We may express the resolution as the "finest" pattern of such lines which is "adequately captured" by the system.

Of course, the definition of "adequately captured" is a subjective one, one of the limitations of this simplistic approach to quantifying resolution. Still, this concept of resolution is widely used, as it is easily grasped.

---

[1] The "pix" part is drawn from photographers' slang for "pictures".

But in what units can we do this? If the place at which we wish to examine the resolution has a physical size (as in the case of the film frame in a film camera, or the imager of a digital camera), we may quantify the “fineness” of this “finest pattern” in terms of lines per unit distance, for example, lines per millimeter[2].

If the place at which we wish to examine the resolution does not have an explicit physical size (as in the case of a digital camera image—much more about that later—or a video signal) we may instead express the fineness of the pattern in terms of lines per picture height.

In the case of a digital imaging system, the available resolution is a result of the interplay of several factors, including:

1. The spacing of the pixels
2. A phenomenon called the Kell effect.
3. The degree to which the lens transfers the scene to the imager without “spreading” of the image points

If only factor (1) were involved, then an imager (sensor array) with 2000 rows of pixel sensors from top to bottom would yield a resolution of 2000 lines per picture height. I call this number the “pixel resolution” of the system (sometimes I use the term “geometric resolution”).

Regarding factor (2): If we had an imager comprising 2000 rows of pixel sensors from top to bottom, and we projected on it the image of a test pattern of 2000 lines, alternately dark and light, would it be captured?

Well, if the lines happened to fall square on the rows of pixel sensor rows, yes it would. But if the lines happened to fall exactly on the boundaries between the rows of pixel sensors, each one (in any line) would see half of a dark area of the pattern and half of a light area. Each pixel sensor (in every line) would deliver a report of “gray” for its location, and we would have a gray image devoid of any evidence of the lines in the test pattern.

If we consider numerous possible situations of the alignment of repeating patterns on a sensor, we would find that, on the average, we could “adequately” record patterns (in our example) having about 1400 lines per picture height.

---

[2] Note that in this notation, the light and dark lines each count, even though it is tempting to think that the light ones are not lines at all but just the background upon which we see dark lines. Adding to the possibility of confusion here is the fact that in many cases, the fineness of the pattern is expressed in “line pairs per millimeter”. The problem is made worse by the use of badly-thought out abbreviations.

The ratio of this practical resolution to the "geometric" resolution (0.7 in this example) is called the *Kell factor*, after the scientist who first characterized this phenomenon.

Factor (3) is of course of great importance in camera behavior. If we have 2000 rows of pixel detectors from top to bottom of our imager, but the lens, as a result of its imperfections in performance, takes each infinitesimal patch of the scene and presents it to the imager as a blob whose diameter is 1/1000 the height of the imager, we will certainly not receive the benefit of the 2000 lines/picture height resolution the "geometric resolution" value might seem to promise.

All this notwithstanding, for the reminder of this article we will concern ourselves with only the geometric resolution, listed above as factor (1).

Finally note that another outlook on expression of the pixel resolution of an image is to mention both pixel dimensions (rather than a density in pixels per inch or pixels per picture height.

THE RESOLUTION INDICATOR(S)

Introduction

Of great interest to the digital photographic community is a factor I will call here, for the moment, the *resolution indicator*, a numerical value carried by most modern form of digital camera output files, especially those with filetype extensions of JPG (JPEG files). There is great confusion and misunderstanding about this quantity, and I will devote considerable space here to sorting this all out.

JPG file standards

First, we need to look into just what kind of file carries filetype extension JPG. We know, for example, that in such a file, the image data is coded in a "compressed" form, under a scheme denoted JPEG[3].

But there are at least three file format "standards" that have that property, and all three types of file customarily have the filetype extension "JPG[4]". Let us briefly look at these, as the difference can lead to unexpected results regarding the topic of this section.

1. The basic JPEG file. The body of the JPEG standard does not define the file format in which the data set resulting from a JPEG coding/compression is to

---

[3] The designation comes from the Joint Photographic Experts Group, a task force working under the auspices of several international standards bodies which developed this standard coding system.

[4] In certain spheres, the alternative filetype extensions "JPE" and "JPEG" are found.

be conveyed. That data set is just a set of byte values. However, an Annex to the standard does suggest a file format for conveying such a data set. This prescription does not nail down a lot of properties that would need to be known to be able to unambiguously reconstruct the image as intended without benefit of other information.

2. The JFIF file. This is the "JPEG File Interchange Format". It is based on the standard JPEG file format, but nails down certain properties, and provides for some metadata that describes what values of certain properties are involved. From a file of this type, a receiving application can generally reconstruct the image intended to be conveyed, but there may be some uncertainties.

3. The Exif file. This is the "Exchangeable image file format for digital cameras". A prominent feature of this format is its provision for a rich roster of metadata for describing the image and the circumstances of its recording. This includes the information on camera make and model, shutter speed, aperture, and so forth that has become so beloved to digital camera users. Many people believe that "Exif" refers only to this body of metadata. In fact, it refers to the entirety of this specific file format definition. The Exif standard nails down, or provides for the explicit indication of, even more factors critical to the unambiguous reconstruction of the conveyed image.

We also hear of the DCF files. These are defined by the "Design rule for Camera File system". This actually uses the Exif file format, but adds to it a structure for directory and file naming and other important "housekeeping" matters, in the interest of facilitating the distribution and utilization of digital camera output files. It is not really a distinct file format (although there are certain details of the Exif format that are impacted by its use in the DCF context).

Which type of file we are dealing with can be ascertained by examining the file in a "hexadecimal" editor (using its ASCII pane). Nicely enough, a JFIF file will show the string "JFIF" early in the file; an Exif file will show the string "Exif". (The exact locations may vary owing to the tag/pointer nature of the data structure.) A "basic JPEG" file will of course show neither of these strings.

Some image editors[5], having acquired an image from an Exif file, will write that image in a file which is a hybrid of the JFIF and Exif formats. These carry the resolution indicators in both the JFIF and Exif forms, and show both the strings "JFIF" and "Exif" early in the file.

[5] For example, Adobe Photoshop Elements 2.

The resolution indicator(s)

The JFIF and Exif file formats provide (in totally different data structures) for conveying two values describing the resolution of the image in terms of pixels per unit distance (one for each direction). The distance basis of the unit can be either the inch or the centimeter; another field in the file tells which. Unfortunately, the Exif format standard calls the unit "dots per inch (or "dots per centimeter"), thus opening the door to a misunderstanding about the difference between dots and pixels that we will treat shortly.

Here I will speak of "dots per inch" only when referring to the specific language of the standard; I will speak of "pixels per inch" otherwise, to prepare for a later section in which we will speak of real dots, which are not pixels.

The two separate values pertain to the resolution in the X (horizontal) and y (vertical direction, as those could be different. They are not different in the case of any camera files we are likely to encounter, but the provision is there to cater to a rather broader outlook on digital image files.[6] Understandably, only one value is reported by most image-handling software. Nevertheless, I will refer to these values in the plural in this article.

With regard to the details, such as the formal names of the two values, unless I mention otherwise, I will use the conventions in the Exif file format.

In the Exif format standard, the two data items are formally called XResolution ("image resolution in width direction") and YResolution ("image resolution in height direction"). In the JFIF format standard, they are called, Xdensity ("Horizontal pixel density"), and, Ydensity ("Vertical pixel density").

But what do they mean?

Based on the discussion earlier, the fact that these values are in terms of pixels per inch (called "dots per inch" in the Exif standard) would suggest that they apply at the imager itself, which of course has an "inch" size. But they do not. In fact, in the Exif format, there is a totally different pair of values (FocalPlaneXResolution and FocalPlaneYResolution) that give the pixel resolution at the imager.

So clearly the resolution indicators we are discussing here are intended to pertain to the delivered digital image itself. But a resolution in pixels per inch only has meaning if there is an inch size associated with the image. Does a digital camera output image have a fixed size at which it is intended, or even expected, to be displayed or printed? Not ordinarily.

---

[6] For example, in the early days of personal computers, the display systems often had a different pixel pitch in the two directions, and pixel-oriented image files intended for use on them often matched that asymmetry.

Thus, in fact, in most circumstances of any interest with respect to camera output files, the values XResolution and YResolution have **no meaning** whatsoever. The values found in the image files are wholly arbitrary.

Of course, the image seems to get an inch size when we look at its properties in an image editor or certain types of image viewers. For example, if we have a digital camera output image with pixel dimensions of 2160 x 1440 px, and the file (for whatever reason) carries resolution indicators with the value 72 px/in. ("dots/inch"), the editor will report that the image has dimensions of 30" x 20". It of course derives these dimensions by dividing 2160 px and 1440 px by 72 px/in. What does that mean—that it is expected that we will normally display or print this image at a size of 30" x 20"? Hardly.

So in fact, a digital camera output file has no business carrying resolution indicators of the "pixels per unit distance" type.

Why then is there provision for this in the format standard? I don't know. Probably accommodation of some particular situation coming from the distant past of one of the committee members.[7]

Is there any case in which they are meaningful?

We can certainly construct a scenario in which the resolution indicators are meaningful. Here we go.

A manufacturer has commissioned a graphic artist to design a new label for one of their product packages. The design will be initially used only to make prototype labels to put on mocked-up packages to be shown to a focus group.

For expediency, the client plans to print the prototype labels on a photo-type inkjet printer. Accordingly. he asks the artist to deliver the design in pixel form in a JPG file. He wants the image to have a resolution of 300 pixels/inch (px/in.) at finished size (which his printer can accommodate). The label itself will be 2" wide by 3" high.

But if the artist is using a "vector-based" illustration program, no concept of pixels will be involved during the drawing process itself. All work will be in terms of inch dimensions of the object being drawn.

---

[7] In the JFIF file standard, there is provision for not putting actual resolutions in the two data fields but rather putting in them numbers that give the ratio of horizontal to vertical pixel pitch (for the cases where they are not the same) without stating an actual value for either. Those guys knew what they were doing! My two older digital cameras that generate JFIF files wisely use this latter approach, thus not stating a resolution in "dots/pixels per inch".

Then, when the artist instructs his illustration program to output the design as a JPG file, he will need to instruct the program as to the output resolution to be used—otherwise, the "vector to pixel" conversion routine will have no idea how big, in pixels, the resulting image should be, or (for example) how many pixels wide and high should an 0.75" x 0.25" rectangular object in the drawing should be as "pixelated". In this case, he would set the value 300 px/in. (for both directions).

The result will be a 600 x 900 px image, carrying resolution indicator values of 300 "dpi". And in this case, they will mean exactly what they seem to mean.

But this case is far distant from the normal use of a digital camera where, in any event, we have no opportunity to change the values the camera will put in the file for the resolution indicators.

Where do the "dpi" values different cameras use come from?

The Exif specification prescribes that, if the actual resolution values are "not available", the value 72 dpi should be placed in those fields.

Now in the case of a digital camera, "not available" is a euphemism for "not meaningful". And many camera manufacturers do arrange their digital cameras to present the value 72 dpi for Xresolution and Xresolution.

Other manufacturers, or the same manufacturer for different camera families, will choose a different value. Sometimes we see 180 dpi, 96 dpi, or 300 dpi. What do these different values tell us about these cameras, or the images they deliver?

Absolutely nothing. Does a camera with an output image of 3000 x 2000 pixels, whose file carries 180 dpi for the resolution indicators, have a greater resolution (in any sense) than a camera with an output image of 3000 x 2000 pixels, whose file carries 72 dpi for the resolution indicators? No.

Then why the difference? I have no idea. My suspicion is that the manufacturers are concerned that the inch size of the image implied by the combination of the pixel dimensions of the image and the arbitrary resolution indicator values (as reported in image editors) be "reasonable".

Where does the default "72" value come from? Well, often computer displays operate at a resolution of somewhere in the area of 72 px/in. That means that if the resolution indicator is 72 px/in. ("dpi"), and we ask an image editor application to put the image up on the screen on a "pixel for pixel" basis (often called, curiously, a "100% display"—more about this later when we talk about editors), the physical size of the image (in inches) will be about the same as the size implied by the combination of the pixel dimensions of the image and the resolution indicators.

That was a very tidy notion when most images had the same pixel dimensions as the display area of most monitors ("send me that picture in VGA size, please"), and "100% display" was the automatic norm when an image was put on the screen. Today, it just doesn't work, unless (for the camera we described above) one had a 30" x 20" display on the computer.

So do they do anything?

Just because the resolution indicators have no meaning doesn't (unfortunately) mean that they don't do anything.

For one thing, most image editing software packages report the size of an image in both pixel dimensions and inch dimensions. Where do they get the inch dimensions? By dividing the pixel dimensions by the resolution indicators found in the file. So, for the example above, the editor would report the image to have a size of 2160 x 1440 pixels, and 30" x 20". What does that mean? Nothing.

Now how big is an image being displayed going to be shown on the screen? Most editors use a "percent zoom" approach—but in two different ways.

In some cases, 100% zoom means the image just fills the available display space. If we ask for 200% zoom, we see half the height and half the width[8] of the image, and exactly what part of the image we see is controlled by the scroll bars. If we ask for a 50% zoom, we see the whole image at a scale such that it occupies half the available display height and half the available display width.

But in most editors, "100% zoom" means that each pixel of the file matches a pixel of the display layout. (A curious notation!) In those, if we ask for a 100% display of the example image (2160 x 1440 px), and if our display system is such that the available space for display is 1200 x 800 px, then we of necessity only can see 0.55 of the height and width of the image.

In some editors of the latter category, we can advise the editor of the actual resolution in pixels per inch of our monitor setup. This result is that an image being displayed at "25%" will be, on the screen (assuming it all fits), exactly one quarter as wide and high as the "inch size" reported for the image—which still doesn't mean anything.

Some anomalies

The fact that the resolution indicators are encoded in a wholly different way in the JFIF and Exif file formats can lead to some peculiar reports by image editors and

[8] Actually, if the available display space doesn't have the same proportions as the image, then it doesn't quite work out that way.

viewers. Some of them can only read the indicators in one or the other of those two formats.

Many such applications, if they are unable to find the resolution indicators, display and reckon the inch size of the image based on a certain default value. In some cases that is the well-known default 72 px/in. In other cases, it is the value the user has set into the application's preferences for the resolution value to be associated with a new, blank image. In yet other cases, a different, arbitrary value is used.[9] Thus, we may see that default value displayed (and used for "inch size" reckoning) even though there are actual values in the file.

RESIZING, RESAMPLING, AND INTERPOLATION

We may wish to prepare, from an image file, a derivative file carrying all the image material (that is, not a "crop" of the image) but having greater or lesser pixel dimensions than the original file. Why might we want that?

We might want an image with smaller pixel dimensions so as to have a smaller file size, which may be more practical when transmission is involved. We might want a smaller pixel size for uses where the image will be displayed pixel-per-pixel (as in some cases of images embedded in e-mail or forum messages) and we wish the on-screen size to be convenient for the recipient. Note that in doing so, we sacrifice image geometric resolution, at least in the sense of pixels/picture height. But looking at geometric spatial resolution (in px/in.), if we know the image will be actually displayed at a small size, that resolution may be perfectly adequate. (Note that this resolution value has nothing to do with the value of the resolution indicator that might be in the file).

We might also want to either increase or decrease the pixel dimensions in preparation for printing in order to match the input resolution of the printer chain for the print size we are commanding (see more on this later). Note that in the case of an increase in pixel dimensions, there is no increase in the actual geometric resolution of the image (as no more true detail information is available than there was in the original image).

Whatever our motivation, most image editors will be glad to do this for us. Depending on how we manipulate the controls for this process, one of two things may happen:

1. The new image will have the desired new pixel dimensions and will carry the original value of the resolution indicator(s). As a result, the "inch size" reported by the editor (or by other editors that may later have a chance to

---

[9] My own image editor of choice, Corel Picture Publisher 8,. uses 150 px/in. in such cases.

regard the file) will be correspondingly different from what we saw before. Of course, that inch size is meaningless, before and after the change.

2. The new image will have the desired pixel dimensions and will carry a value for the resolution indicators that is changed in the same proportion as the pixel dimensions. As a result, the "inch size" reported by the editor (or by other editors that may later have a chance to regard the file) will be the same as was reported for the original file. Of course, that inch size is meaningless.

What is different between the actual images in these two cases? Nothing.

This process is often described as "resizing" and often as "resampling". The latter term relates to the fact that any pixel-form digital image represents the original image developed by the lens in terms of "samples": the color values at a finite set of points, where the pixels are located.

If we thus transform the image to have a different array of pixel locations (usually, more or less than before), the result looks as if we had now sampled the lens image on a different pixel pattern—thus the term, "resampling".

But that name is overoptimistic. Our result is not really the same as if we actually resampled the original lens image on the new pixel pattern—we don't have the original lens image to do that to. All we can do is to make an estimate of **what those pixel values would most likely have been** based on crafty analysis of the values of the pixels we do have in our "first" image. This process is described as *interpolation*, and there are many ways to do it. Some ways produce a result that is more visually satisfactory (whatever that means!) than others. The cost is that those generally require more processing time to execute.

There is a third operation that we can perform, If we set the controls properly, the new image will not have different pixel dimensions than the original image, but the resolution indicator(s) will carry a new value. The way we cause this result in many editors is to set a new resolution and instruct the program, "do not resample".

PUBLISHERS' REQUIREMENTS

We may have inquired about submitting a digital photograph to some magazine, and have received information that we should submit the file at "at least 300 dpi" resolution. What does that mean?

Well, stated that way, it doesn't say anything. The publisher either:

- Has no idea what they mean but just saw some other publication say it, or
- Wants something meaningful but doesn't know how to say it.

Now a useful thing to want, and to ask for, would be this: "Please submit only files with a resolution of at least 300 px/in. for the size at which we will print it".

Now of course, you probably have no idea at what size they will print it. So then a useful thing for them to add might be, "In this publication most photographs for the My Favorite Birdhouse section are usually not printed larger than 3 in. x 2 in." (or whatever).

So if they had done that (and assuming they will not seriously crop your image), if your image has pixel dimensions of at least 900 x 600 px, you will be fine.

A horrifying part of this is that some of these publishers expect the resolution indicator(s) in the file to read at least 300 dpi. (That may in fact be the only thing they expect!)

These poor souls are easily accommodated. As we mentioned previously, most photo editors can make a copy of a file with the resolution indicators set to any value you want. It need not have any effect whatsoever on the pixel dimensions of the image or anything else about it.

PRINTING ISSUES

Pixels and dots

As I mentioned, the resolution indicators in an Exif file are said by the format standard to be in terms of "dots per inch" (or dots per centimeter, but we less often encounter that).

Considering "dot" to be a synonym for "pixel" is unfortunate, as there is a situation in which we use both terms for different things.

In an ink jet or laser printer, the engine usually cannot really lay down different "densities" of ink. In order to get a range of different gray values (on a black and white printer), or to get different densities of the various primary inks (on a color printer) the printer usually employs a form of *halftone technique*, similar to that used in printing books and newspapers by letterpress or offset processes.

In this technique, the image is printed as a set of dots, usually located on a fixed grid. Normally, there are several dot positions used to form each pixel of the image. The print head can be made to print or not any dot from each pixel's "dot cluster", and in many ink jet printers can vary the diameters of the dots. Even though the ink that is deposited has a fixed "density", the effective density over any area (even over the area of one pixel) can be changed over a wide range by this technique.

As a matter of interest, for many consumer inkjet printers, we never actually hear what their pixel resolution is. [10] [11] What we hear about is their dot density. (There is usually a range, with the specific value selectable by the user, often subject to constraints based on the type of paper to be used.)

Why do we hear about the dot density and not the pixel density? Probably because the dot densities are higher numbers, and this is supposed to be more impressive to the prospective purchaser. ("Now! 2880 DPI in the JetKing 2000.")

For example, my Epson Stylus Color 980 printer (in most modes) has a pixel resolution of 720 x 720 pixels per inch. But the printer driver control panel typically offers me a choice of 1440 and 2880 dots per inch—the difference being the number of dots in a pixel cluster. With more dots per pixel cluster, the gradations of color can be more refined. Is there more "detail" in the printed image at 2880 dpi than 1440 dpi (assuming that there is that much detail in the image itself)? No. The detail in the image is limited by the pixel resolution, 720 x 720 px/in.

At the "2880 dpi" setting, is the dot density 2880 x 2880 dpi (giving 16 dots per pixel)? No. On the printer mentioned, that setting produces a dot pattern of 2880 x 720 dpi (giving 4 dots per pixel).

Resolution compatibility in printing

The resolution I mentioned above (720 x 720 px/in. for my Epson 980 printer) is often referred to as the input resolution of the printer-driver chain, as it is the resolution in which the driver would prefer to receive the image.

But suppose that we wish to print an image at a certain inch size, and the pixel dimensions of the image do not match that size considering the input resolution of the printer driver?

In actual practice, there are three ways that this matter can be taken care of:

1. The printing application (perhaps an image editor or viewer, or perhaps a special "printing management" application) will just feed the image to the printer driver with a specification of the desired output size (as set into the editor by the user), and if the pixel dimensions of the image and the print size don't jibe with the input resolution of the driver-printer chain, the driver

---

[10] This is sometimes referred to as the *input resolution* of the printer driver, as it is the resolution in which the driver would prefer to receive the image.

[11] For example, for many of my printers, the only way I know their input resolution is that the print management program **Qimage** reports it for the selected printer. Qimage finds out from the printer driver.

(of its own accord) internally transforms the image by "interpolation"[12] to a new image having the necessary pixel dimensions. The interpolation algorithm is built into the printer driver.

2. The user, using an image editor, constructs a new image having the exact pixel dimensions that will match the desired print size at the input resolution of the driver-printer chain. Doing so involves interpolation, using an algorithm built into the editor.

3. A print control program is used that, "on-the-fly", constructs in memory a new image having the exact pixel dimensions that will match the desired print size at the input resolution of the driver-printer chain. Doing so involves interpolation, using an algorithm built into the program.

One issue in choosing among these is the matter of how effective the particular interpolation algorithm is at producing a visually-realistic derivative image. Often the interpolation algorithms built into printer drivers (used in scenario 1) are considered less effective in this regard than those typically available in the applications doing the interpolation in scenarios 2 and 3.

Further discussion of this complex issue is beyond the scope of this article.

#

[12] *Interpolation* in this case refers to a process in which appropriate color value for the pixels at the pixel locations in the new image are derived by consideration of the color values of nearby pixels in the original image. Many different algorithms are known, differing (among other things) in the degree of visual realism of the new image as compared to the original image.