Optimal Aperture in Photography. 1. Theory

Phil Service
Flagstaff, Arizona, USA
30 September 2013

Introduction

Camera movement and vibration, subject movement, lens aberrations, defocus, and diffraction can all contribute to blur in photographic images. The subject of this paper is defocus and diffraction, and their combined effects. Defocus blur *decreases* with smaller apertures.[1] On the other hand, diffraction blur *increases* with smaller apertures. Because defocus and diffraction blurs respond in opposite directions to changes in aperture, there can be an optimal setting that minimizes their combined effect. The purpose of this paper is to develop the concept of optimal aperture in photographic applications, and to provide formulas for its determination. This treatment is an expansion of the work of Peterson[2] and Hansma[3]; and the immediate inspiration for the present analysis was provided by several iOS apps developed by George Douvos.[4] In particular, my goal has been to make the mathematics behind optimal aperture calculations explicit, and to extend the analysis to consideration of camera format (sensor size).

1. Blur

1.1 Total Blur

Peterson and Hansma suggest that the combined effects of defocus and diffraction can be calculated by the following formula:

$$c_t = \sqrt{c_d{}^2 + c_a{}^2}$$

1(a)

where c_d is defocus blur, c_a is blur due to diffraction, and c_t is total blur (disregarding other sources such as lens aberrations). Equivalently,

$$c_t{}^2 = c_d{}^2 + c_a{}^2$$

1(b)

---

[1] "Smaller" in relation to aperture means a smaller opening in the aperture diaphragm, which corresponds to a larger f-value: f/11 is a smaller aperture than f/4. Similarly, a "larger" aperture means a smaller numerical f-value.

[2] Peterson, Stephen. Image sharpness and focusing the view camera. Photo Techniques, March/April 1996:51-53.

[3] Hansma, Paul K. View camera focusing in practice. Photo Techniques, March/April 1996:54-57.

[4] You are strongly encouraged to visit the web site of George Douvos and to investigate his iOS apps: including OptimumCS-Pro, TrueDoF-Pro, and FocusStacker. The documentation for these apps covers much of the same material as this paper, and the apps perform most of the calculations that are described here. http://www.georgedouvos.com/douvos/ George_Douvos__Innovative_Apps_for_Apple_Mobile_Devices.html

1.2 Defocus Blur

There are several possible formulas for defocus blur. Standard treatments of depth of field give the following two equations:

$$c_{d(n)} = \frac{f(v_n - v)}{Nv_n} = \frac{f\,\partial v_n}{Nv_n}$$

2

and

$$c_{d(f)} = \frac{f(v - v_f)}{Nv_f} = \frac{f\,\partial v_f}{Nv_f}$$

3

where f is lens focal length and N is aperture number. v is the distance, *behind the lens,* of the image of an in-focus object that is at distance u from the camera. v_n is the (hypothetical) distance behind the lens of the image of a foreground object at distance u_n from the camera; and v_f is the distance behind the lens of the image of a background object at distance u_f from the camera. $c_{d(n)}$ is the diameter of the blur circle of an infinitesimal point on the foreground object; and $c_{d(f)}$ is the blur circle for the background object. In the usual treatment of depth of field, DoF limits are computed for the case in which $c_{d(n)} = c_{d(f)}$, and the "n" and "f" subscripts are omitted (as is the "d" subscript). ∂v_n ("delta-vee sub-en") is simply $(v_n - v)$, and ∂v_f is $(v - v_f)$. In the case of conventional cameras, v, v_n, and v_f must be estimated from object distances by means of the standard lens equation:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$$

As a numerical example, suppose that we are using a 50 mm lens to focus on an object that is $u = 5.7$ m from the camera. Then, v_n for a foreground object at $u_n = 4.0$ m ($= 4{,}000$ mm) is given by:

$$\frac{1}{v_n} = \frac{1}{f} - \frac{1}{u_n} = \frac{1}{50} - \frac{1}{4000} = 0.01975$$

from which $v_n = 50.633$ mm. Similar calculations give $v = 50.442$ mm. Then, from Eq. 2, if aperture N is 5.6, $c_{d(n)} = 0.03368$ mm $= 33.68$ μm.

Peterson derived a third formula for defocus blur:

$$c_{d(P)} = \frac{\partial v}{2N}$$

4

where N is aperture, as before, and ∂v (without subscript) is the *focus spread*, defined as v_n - v_f. Peterson's analysis of defocus blur considers the case where the goal is to jointly minimize blur on two objects at different distances from the camera. In this case, Hansma states that for most purposes the focus distance can be obtained by setting $v = (v_n + v_f)/2$. (The exception is when the ratio of distances of near and far objects is much different than one.) If $v = (v_n + v_f)/2$, then $\partial v_n = \partial v_f = \partial v/2$. If we substitute for ∂v_n in Eq. 2, we get:

$$c_{d(n)} = \frac{f \, \partial v}{2 N v_n}$$

This is similar to Peterson's equation (Eq. 4), but with the added term (f / v_n). If the foreground object distance, u_n, is equal to 20 times the lens focal length, then $(f / v_n) = 0.95$, and becomes closer to 1.0 as the foreground object distance increases. In most cases, then, the approximation incurred by omitting the (f / v_n) term is reasonable. Similar considerations apply to substitution of $\partial v/2$ for ∂v_f in Eq. 3. The result is that Eq. 2, 3 and 4 are equivalent ways of calculating defocus blur.[5] When the particular expression for defocus blur is unimportant, I will use "c_d" without subscript.

1.3 Diffraction Blur

The diffraction blur spot diameter is:

$$c_a = \frac{2.44 \alpha \lambda N}{k}$$

5

where λ is the wavelength of light in nanometers (usually taken to be about 550 nm), N is the aperture, and k is a scaling factor whose value depends on the units that c_a is to be expressed in.[6] The parameter α is the *diffraction blur coefficient*. It is omitted in other treatments of diffraction blur that can be found in the literature. Effectively, it is assumed to equal one, if it is considered at all. If $\alpha = 1.0$, then Eq. 5 is the diameter (to the first minimum) of the central bright disk of the Airy diffraction pattern. For purposes of example calculations in this paper, I will use $\alpha = 1.0$. The next paper in this series will describe an experiment that estimates its actual value for a

---

[5] Before leaving the subject of defocus blur, it is worth noting that the Peterson — Hansma analysis assumes that a view camera is being used. In that case, it is possible to read v_n and v_f directly from a scale on the camera rail. That makes it unnecessary to use the lens equation to calculate image distances (and ∂v) from object distances.

[6] Differences in customary units of length lead to confusion. For example, lens focal length is usually expressed in millimeters, wavelength in nanometers, blur spots (circles of confusion) in millimeters or microns, and so on. For the present calculations, it is crucial that c_a and c_d be expressed in the same units. The scaling factor, k, is the ratio of the scales used for blur spots and wavelength. For example if blur spots are expressed in microns (10^{-6} m) and wavelength is in nanometers (10^{-9} m), $k = 10^{-6} / 10^{-9} = 10^3$. On the other hand, if blur spots are expressed in millimeters and wavelength in nanometers, $k = 10^6$.

specific lens and camera. To illustrate calculation of diffraction blur, if the aperture, $N = 5.6$, and $\lambda = 550$ nm, then $c_a = 7.52$ μm. (In this case $k = 1000$ in order to express c_a in microns.)

1.4 Total Blur - Example

We can combine the values of c_d and c_a from the examples in Sections 1.2 and 1.3 to obtain total blur. By Eq. 1(a), the total blur, $c_t = \sqrt{(33.68)^2 + (7.52)^2} = 34.51$ μm. In this case, the diffraction blur is much smaller than the defocus blur and contributes almost nothing to the total blur. That result immediately suggests that the total blur for the images of foreground and background objects could be reduced by decreasing the aperture: if, by doing so, defocus blur decreases faster than diffraction blur increases. We will return to this example below.

2. Optimal Aperture

I suggest two definitions of optimal aperture. The first assumes that maximizing the sharpness of objects at different known distances from the camera is of primary importance. The optimal aperture is then the one that minimizes the total blur of those necessarily defocused subjects. The second definition assumes that the primary consideration is control of total on-sensor (or film) blur at some predetermined value. The optimal aperture is then the one that maximizes "depth of field" for that amount of total blur. These are two different optimization problems: the first minimizes blur (maximizes sharpness) for a given separation between objects; the second maximizes the distance between sharp planes in the object field in the case where "sharp" is predetermined. A third definition of optimal aperture might be the aperture that minimizes the hyperfocal distance. As will be shown below, the third definition is equivalent to the second.

Given that we have three equations for defocus blur, we have three possible expressions for total blur. To illustrate the derivation of optimal aperture, I will use Eq. 4 for defocus blur. Derivations using Eq. 2 and 3 can be found in the Appendix. Substituting Eq. 4 and 5 into Eq. 1(b), we have:

$$c_t^2 = \frac{(\partial v)^2}{4N^2} + \left(\frac{2.44\alpha\lambda}{k} \right)^2 N^2 \qquad 6$$

It is worth noting at this point that lens focal length and camera format (sensor size) do not directly enter into the calculation of total blur, nor, as we will see below, the calculation of optimal aperture. Camera format can affect our results indirectly, however, in several ways. For example, sensor size may influence the choice of lens focal length which, in turn, will affect v_n and v_f (and ∂v and c_d).

2.1 Maximum Sharpness (Minimum Total Blur) of Near and Far Objects

In this case, ∂v is fixed (because u_n and u_f are fixed) and aperture, N, is thus the only variable on the right-hand-side of Eq. 6. The required optimization is accomplished by differentiating c_t^2 with respect to N, then setting the derivative to zero and solving for N.

$$\frac{dc_t^2}{dN} = -0.5N^{-3}(\partial v)^2 + 2\left(\frac{2.44\alpha\lambda}{k}\right)^2 N$$

If $dc_t^2/dN = 0$, then:

$$0.5N^{-3}(\partial v)^2 = 2\left(\frac{2.44\alpha\lambda}{k}\right)^2 N$$

From which:

$$N_{min(Blur)} = \sqrt{\frac{(\partial v)k}{4.88\alpha\lambda}} \qquad\qquad 7$$

As previously noted, this optimum aperture is indeed independent of camera format (sensor size) and lens focal length (except in so far as focal length will affect the focus spread, ∂v).[7]

Let's return to the example that was introduced in Section 1.2, with the added requirement that we want to jointly minimize the blur of objects at 4 and 10 m from the camera. In addition to $v_n = 50.633$ mm, which was calculated previously, we need $v_f = 50.251$ mm (from the lens equation and $u_f = 10$ m). Then $\partial v = 0.382$ mm $= 382$ μm, and

$$N_{min(Blur)} = \sqrt{\frac{382\times10^3}{4.88\times550}} = 11.93$$

The nearest in-camera aperture that we could actually set would probably be f/11. Taking $N = 11$ for our aperture, $c_{d(P)} = 17.35$ μm, and $c_a = 14.76$ μm by Eq. 4 and 5; and $c_t = 22.78$ μm by Eq. 1. If we had chosen to use f/13 instead, the blurs would have been 14.68, 17.45, and 22.80 μm,

----

[7] I have chosen the subscript "min(Blur)" for this optimal aperture to indicate that the optimization was achieved by minimizing total blur. Eq. 7 is equivalent to Hansma's formula for optimum aperture, N = $\sqrt{(375\,\partial v)}$, where ∂v is expressed in mm (so $k = 10^6$), $\lambda = 546$ nm, and $\alpha = 1.0$.

respectively.[8] The difference in total blur between f/11 and f/13 is negligible, and either of these two apertures is optimal in this case: larger or smaller apertures will produce greater total blur.[9] Note the difference between this result and that obtained by conventional treatments of depth of field. In conventional treatments, sharpness of foreground and background objects always increases with smaller apertures. In our analysis, however, decreasing aperture beyond $N_{min(Blur)}$ will not improve image sharpness because the decrease in defocus blur that will be achieved will be more than offset by an increase in diffraction blur.

It is important to emphasize that this procedure does not necessarily ensure that the near and far objects will be *sufficiently sharp*; only that they are as sharp as can be made given their distances from the camera and the focal length of the lens being used. Whether they are sufficiently sharp will depend on the criterion of sharpness employed, and on the final image size and viewing distance. Greater sharpness can be obtained by using a lens of shorter focal length or by moving the camera farther from the objects. Both alternatives decrease the image sizes of the objects, and the latter also alters their relative sizes (perspective).

2.2 Maximum Depth of Field for Specified Total Blur

In this case, c_t is fixed and the optimization problem requires maximizing ∂v, which is equivalent to maximizing "depth of field" (u_f - u_n). First, rearrange Eq. 6 to solve for $(\partial v)^2$:

$$\frac{(\partial v)^2}{4N^2} = c_t^2 - \left(\frac{2.44\alpha\lambda}{k}\right)^2 N^2$$

$$(\partial v)^2 = 4c_t^2N^2 - 4\left(\frac{2.44\alpha\lambda}{k}\right)^2 N^4$$

Then differentiate $(\partial v)^2$ with respect to N and set the derivative to zero in order to find the value of N that maximizes $(\partial v)^2$.

$$\frac{d(\partial v)^2}{dN} = 8c_t^2N - 16\left(\frac{2.44\alpha\lambda}{k}\right)^2 N^3$$

---

[8] The similarity in magnitudes of c_d and c_a is not accidental. In fact, if we could actually have set the camera aperture to $N = 11.93$, c_d and c_a would have been equal. Given the way in which defocus and diffraction blurs are combined to produce total blur [Eq. 1(a)], it can be proved that they will be equal whenever total blur is minimized. (That said, c_t(min) = 2.0 is a special case, unlikely to be encountered in practice, that allows $c_d \neq c_a$.)

[9] In Sec. 1.4, we calculated $c_t = 34.51$ μm when $N = 5.6$.

Setting this derivative to zero:

$$8c_t^2 N = 16\left(\frac{2.44\alpha\lambda}{k}\right)^2 N^3$$

From which:

$$N_{max(DoF)} = \sqrt{0.5}\,\frac{c_t k}{2.44\alpha\lambda} = 0.29\frac{c_t k}{\alpha\lambda} \qquad\qquad 8$$

Once again, the optimal aperture is independent of focal length and camera format (sensor size), although camera format may influence the degree of image magnification for final display and thus potentially affect the value chosen for c_t.[10]

As a numerical example, suppose our criterion for acceptable sharpness is a total blur circle diameter of 15 µm. If λ = 550 nm, then $N_{max(DoF)}$ = 7.91 ≈ 8. For N = 8, diffraction blur is 10.7 µm (Eq. 5) and defocus blur must be 10.5 µm in order to satisfy the constraint that total blur is 15 µm.[11] The actual depth of field limits with an aperture of f/8 will depend on the lens focal length and the distance to which it is focused, and are determined using only the defocus component of total blur. Let us return to the example just discussed in Section 2.1. In order to jointly minimize blur on the near and far objects in that example, the focus distance should have been 5.7 m.[12] At f/8, with a 50 mm lens focused at 5.7 m and an acceptable defocus circle of 10.5 µm, the near and far limits of the depth of field are 4.8 and 7.1 m.[13] Consistent with our earlier analysis (Sec. 2.1), it is not possible to have subjects at 4 and 10 m from the camera simultaneously acceptably sharp with a 50 mm lens, if our criterion of acceptable sharpness is a total blur of 15 µm. As we saw previously, the best that we could obtain with this setup was a total subject blur of 22.8 µm. Fig. 1 shows how depth of field varies with aperture when total blur is constrained. Note that depth of field falls to zero when all of the available total blur is taken up by diffraction, leaving nothing for defocus blur.

---

[10] I have chosen the subscript "max(DoF)" for this optimal aperture because the optimization maximizes depth of field for a specified total blur at the near and far depth of field limits. Again, the value of the scaling constant, k, depends on the ratio of the units used for c_t and λ (see footnote 6).

[11] If N = 7.91, c_a = 10.6 μm, by Eq. 5. Then, by Eq. 1, $c_d = \sqrt{(c_t^2 - c_a^2)} = \sqrt{(225 - 112.8)} = 10.6\ \mu m$. This is a second example of the rule that optimal aperture, either $N_{max(DoF)}$ or $N_{min(Blur)}$, is obtained when defocus and diffraction blurs are equal (see also footnote 8).

[12] Obtained by setting $v = (v_f + v_n)/2$ = 50.444 mm, and using the lens equation to solve for u.

[13] The near and far depth of field limits are computed with the usual formulas: $D_{near} \approx Hs / (H + s)$, and $D_{far} \approx Hs / (H - s)$ for $s < H$, where s = subject distance and H is hyperfocal distance. The hyperfocal distance is also calculated in the usual way except that the circle of confusion parameter now refers specifically to defocus blur: $H \approx f^2 / Nc_d$, where f is focal length and N is aperture.
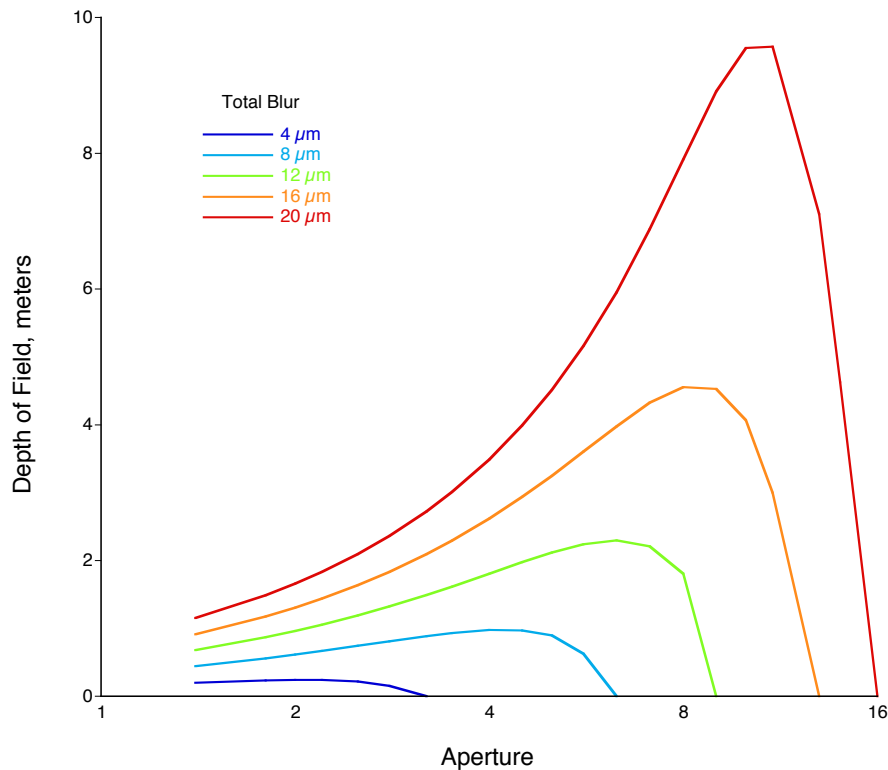
Fig. 1. Depth of field vs. aperture for different values of acceptable total blur. 35 mm lens with subject distance = 5 meters. Note that depth of field falls to zero when all of the available total blur is taken up by diffraction, leaving nothing for defocus blur.

2.3 Minimum Hyperfocal Distance

Minimum hyperfocal distance is desirable because it increases depth of field — in the sense that the nearest part of the object field that will be acceptably sharp will be closer to the camera when focused at the hyperfocal distance. In conventional treatments that consider only defocus blur, hyperfocal distance decreases with smaller apertures. However, as we know, decreasing aperture increases diffraction blur. Therefore, for a given amount of acceptable total blur, there must be an optimal aperture that will minimize the hyperfocal distance. This optimum is $N_{max(DoF)}$ (Eq. 8). This result follows from the fact that minimizing hyperfocal distance is simply a specific instance of maximizing the focus spread, ∂v. The hyperfocal distance is the focus distance at which objects at infinity are acceptably sharp. For objects at $u_f = \infty$, $v_f = f$ (the focal length). Maximizing focus spread then amounts to maximizing v_n (or, equivalently, minimizing u_n and H).

Consider the case where our acceptable total blur is 20 μm. From Eq. 8, the optimal aperture is $N = 10.5 \approx$ f/11 (assuming $\lambda = 550$ nm). From Eq. 5, when $N = 11$, diffraction blur is 14.8 μm. This leaves $\sqrt{(20^2 - 14.8^2)} = 13.5$ μm for defocus blur, c_d. The hyperfocal distance is then 16.8 m, by the following equation:

$$H = \frac{f^2}{Nc_d}$$

where, as previously, f is focal length (50 mm) and N is aperture. Essentially identical results are obtained by setting aperture to f/10. To see that we cannot decrease the hyperfocal distance further by decreasing the aperture, to say f/13, while maintaining an acceptable total blur of 20 μm, consider the following. At f/13, diffraction blur is 17.4 μm. That leaves 9.9 μm for defocus blur, and the hyperfocal distance, by the above equation, would be 19.4 m. This is a substantial penalty to pay for a 1/3 stop change in aperture: the acceptably sharp near object-field plane recedes from 8.4 m to 9.7 m from the camera. Fig. 2 illustrates how hyperfocal distance depends on aperture and total blur.
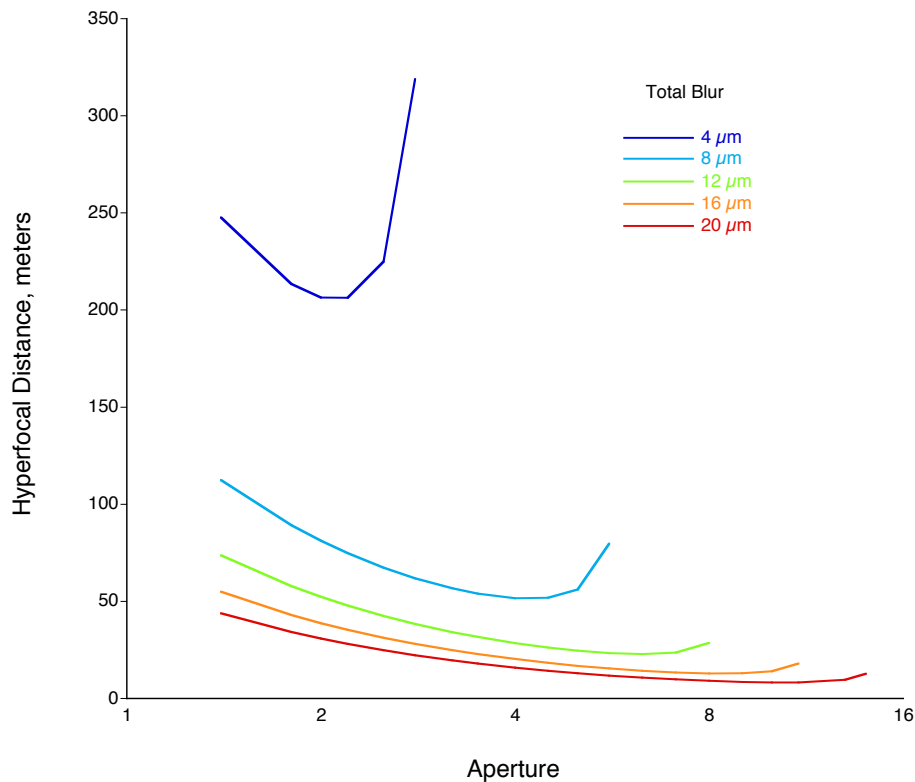

Fig. 2. Hyperfocal distance vs. aperture for different values of acceptable total blur. 35 mm lens. See text (Section 3) for additional discussion.

3. Choosing a Value for Acceptable Total Blur

We now know how to maximize depth of field, given a specified total blur; and also how to determine if subjects at different distances from the camera will be imaged with sufficient sharpness. The question then is: how do we decide on a value for acceptable total blur? I am interested in producing high quality prints. Suppose I intend to make prints with a maximum dimension of 20 inches (50.8 cm). If I am using an APS-C format camera, the maximum dimension of the sensor will be about 24 mm. The image magnification of the print is therefore about 21-fold. Let's further suppose that I want the smallest discernible object on the print to be 0.2 mm — corresponding to one line-pair printed at 5 lp/mm — when the print is scrutinized closely. A 0.2 mm blur circle on the print would then correspond to a 9.45 µm blur spot on the sensor.[14] By Eq. 8, the optimal aperture would be almost exactly f/5. If an on-sensor blur spot of 9.45 µm seems overly stringent, consider that for a 24MP APS-C sensor, the pixel pitch will be about 4 µm. Thus a 9.45 µm blur spot is more than twice the sensor pixel pitch, and objects at the depth of field limits set by that blur spot will be perceptibly less sharp than objects in the plane of focus. I typically shoot in aperture-priority mode with the default aperture set to f/5 or f/5.6. However, I will change the aperture as needed. For example, if I want to minimize depth of field, or I need a different shutter speed, or it's more important to have a low ISO, or if corner sharpness is important and can be significantly improved with a smaller aperture.

A second approach to choosing a total blur spot diameter is to base it on the photosite size of the sensor. If we consider for a moment just the part of the image that corresponds to the object-field plane of focus, then the only source of blur is diffraction (assuming that lens aberrations are fully corrected). For me, careful examination of images made under controlled conditions suggests that the effects of diffraction become just perceptible *in unsharpened images* when the diffraction blur spot is about equal to the pixel pitch. Initially, the effect is to reduce detail contrast very slightly, although resolution *per se* is not affected. So, if you want the objects at your depth of field limits to be nearly as sharp as the objects in the plane of focus, set your total blur to between 1✕ and 2✕ the sensor pixel pitch. In the case of the 24MP APS-C sensor mentioned in the previous paragraph, the total blur spot diameter should be 8 µm (or less), which corresponds to an optimal aperture of f/4.2 (f/4 or f/4.5 given available camera settings).[15]

One aspect of specifying acceptable total blur is that it sets a limit on the smallest permissible aperture. As aperture is decreased, diffraction blur increases. At some point the diffraction blur alone will exceed the specified total blur, as which point nothing is left for defocus blur, hyperfocal distance becomes undefined, and there is no depth of field — that is, no

---

[14] The "standard" treatment of depth of field assumes enlargement to 8 x 10 inches, a viewing distance of 25 cm, and visual acuity of about 5 lp/mm at that viewing distance. The difference between my requirements and the standard treatment is that I assume that 20-inch prints will be viewed as closely as 8 x 10's. Therefore, I am not willing to tolerate a larger on-sensor blur on the assumption that my 20-inch prints will be viewed from some distance greater than 25 cm.

[15] It is fortunate in regard to this discussion of total blur and optimal aperture that many, if not most, lenses perform best in the range of f/4 to f/5.6, at least in the image center. At larger apertures, image quality is typically degraded by uncorrected lens aberrations. And at smaller apertures diffraction will limit sharpness and resolution, although the onset of diffraction effects will depend on sensor photosite size; and for sensors with larger photosites, smaller apertures may be optimal.

part of the object field is acceptably sharp (Fig. 1). In the example of Section 2.3, at f/14 the diffraction blur, c_a, is 18.8 µm (out of 20 µm total). In fact, f/14 is the smallest usable aperture, at which setting the hyperfocal distance has increased to 26 m. Also, it should be noted that the smallest usable aperture is independent of lens focal length and sensor size because diffraction blur depends only on aperture and wavelength, (Eq. 5). The right-hand ends of the curves in Fig. 2 indicate the minimum permissible apertures. As can be seen, small total blur places severe limits on the range of usable apertures. The minimum permissible aperture for any acceptable total blur can be calculated from Eq. 5 by setting c_a equal to total blur and rearranging to solve for N. In the present example, the minimum aperture would be f/14.9 ($\approx$ f/14).
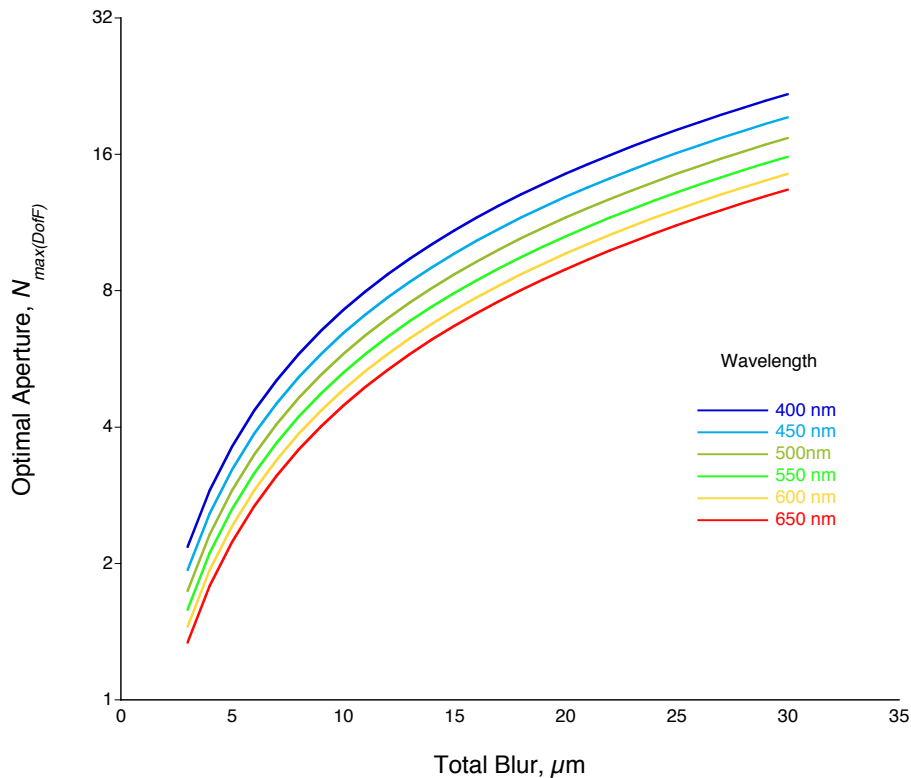

Fig. 3. Optimal aperture (Eq. 8) as a function of total blur for different wavelengths.

3.1 The Effect of Wavelength

The wavelength of light affects our calculations of optimal aperture through its effect on diffraction blur. For the examples in this paper, I have chosen 550 nm as the value for wavelength. The choice is somewhat arbitrary. The visible spectrum for humans is about 390 to 700 nm. Values between about 500 and 550 nm, corresponding to the green portion of the spectrum, are generally chosen for calculation of diffraction effects. That choice is justified largely by the fact the such wavelengths fall approximately in the middle of the visible spectrum. The difference in optimal apertures, $N_{max(DoF)}$, between $\lambda = 500$ nm and $\lambda = 550$ nm is slightly

less than 1/3 stop — a probably insignificant effect. However, optimal apertures calculated for λ = 650 nm are more than one full f-stop larger than those for λ = 450 nm (Fig. 3). So, for photography at "non-standard" wavelengths, infrared for example, it would be advisable to make appropriate adjustments for calculations of optimal aperture.

4. Optimal Aperture and Camera Equivalence

Results obtained with optimal apertures agree exactly with expectations based on established principles of camera equivalence. To illustrate, consider the 24MP APS-C format camera discussed in Section 3. Let us compare it to a 24MP "full-frame" camera, so that we are dealing with a sensor-size "crop factor" of 1.5x. If the APS-C camera has a 50 mm lens, then a 75 mm lens will provide an equivalent field of view with the full-frame camera. We also know from camera equivalence theory that the two cameras will have the same depth of field when the f-number of the full-frame camera is 1.5 times the f-number of the APS-C camera (i.e., the aperture on the full-frame camera is approximately 1-stop smaller than on the APS-C camera).

Table 1. Camera Equivalence When Optimal Aperture is $N_{max(DoF)}$ [*]		
	APS-C	**Full-frame**
Millions of pixels	24 (6000 x 4000)	24 (6000 x 4000)
Sensor dimensions (mm)	24 x 16	36 x 24
Sensor pixel pitch (μm)	4	6
Lens focal length (mm)	50	75
Horizontal field of view (º)	27	27
Final print length (inches @ ppi)	20 @ 300 ppi	20 @ 300 ppi
On-print blur spot size (mm)	0.2	0.2
Print magnification	21.167	14.111
Total on-sensor blur, c_t (μm)	9.449	14.173
Optimal aperture, $N_{max(DoF)}$	4.982	7.473
Diffraction blur, c_a (μm)	6.686	10.029
Defocus blur, c_d (μm)	6.686	10.029
Hyperfocal distance (m)**	75.1	75.1

 * Calculations assume λ= 550 nm
 ** Because hyperfocal distances are the same when both cameras are used at their optimal apertures, depth of field will also be the same (Fig. 4).

In this comparison, assume again, that our goal is to make 20-inch prints with an acceptable blur spot diameter of 0.2 mm (as in the discussion above in Section 3). The optimal apertures and equivalence results are shown in Table 1. Note that the f-number for the optimal aperture with the full-frame camera is 1.5 times the optimal f-number for the APS-C camera. In other words, the ratio of optimal apertures is the same as the "crop factor". Consistent, then, with camera equivalence theory, we see that the two cameras have identical hyperfocal distances and depths of field when used at their optimal apertures.

To reinforce this point, let us return to the problem of simultaneously minimizing the blur of two subjects at different distances from the camera. Suppose that in the example discussed in Section 2.1, we were using an APS-C format camera. The equivalent full-frame camera would have a 75 mm lens. The results are presented in Table 2. Once again, optimal apertures obey the "rules" of camera equivalence. In particular, the optimal f-value for the full-frame camera is 1.5 times the value for the APS-C camera. The total blur spots also differ by a factor of 1.5. However, blur spots as ratios of pixel pitches are the same, as are the sizes of blur spots on 20-inch prints.
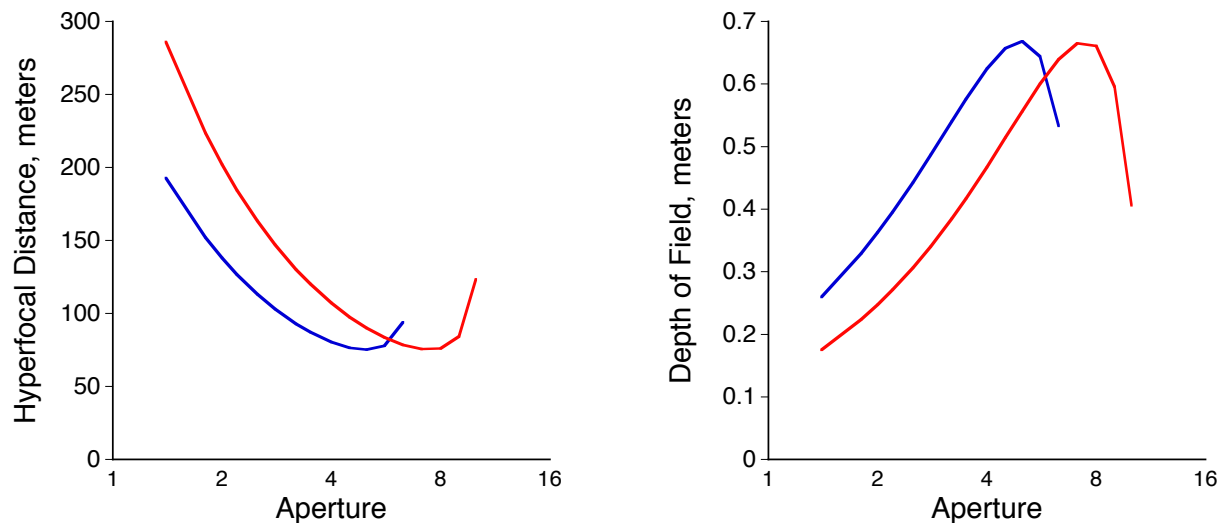

Fig. 4. Hyperfocal distance and depth of field as functions of aperture for equivalent cameras. Blue line, APS-C; red line, full-frame. Camera and lens descriptions, and total on-sensor blur spot diameters are given in Table 1. Depth of field is calculated for a subject at 5 meters. Note the superior depth of field for the APS-C camera at apertures larger than about f/5.6.

Which format is better? Unless shallow depth of field is the goal, these comparisons suggest a possible advantage of the APS-C format over full-frame. Specifically, the smaller sensor has a 1-stop greater light-gathering capacity when both cameras are used at their optimum apertures. Either a slower shutter speed or higher ISO must be used with the full-frame sensor in order to match exposure value. On the other hand, all other things being equal, we expect the larger sensor to have better noise characteristics than the smaller one, so the need to use a higher ISO (assuming equal shutter speeds) might not be a noticeable penalty. Furthermore, any argument in favor of the APS-C camera assumes that the resolving power of the *sensor + lens*

Table 2. Camera Equivalence When Optimal Aperture is $N_{min(Blur)}$ *		
	APS-C	**Full-frame**
Millions of pixels	24 (6000 x 4000)	24 (6000 x 4000)
Sensor dimensions (mm)	24 x 16	36 x 24
Sensor pixel pitch (μm)	4	6
Lens focal length (mm)	50	75
Subject distances: u_n, u_f (m)	4, 10	4, 10
Image distances: v_n, v_f (mm)	50.6329, 50.2513	76.4331, 75.5668
Focus spread, ∂v (μm)	381.6	866.3
Optimal aperture, $N_{min(Blur)}$	11.92	17.97
Focus distance (m)	5.7	5.7
Diffraction blur, c_a (μm)	16.0	24.1
Defocus blur, c_d (μm)	16.0	24.1
Total blur, c_t (μm)	22.6	34.1
Total blur / pixel pitch	5.7	5.7
Total blur spot size for near and far subjects on 20 inch print (mm)	0.48	0.48

  * Calculations assume λ= 550 nm

combinations are equivalent. For the APS-C format, a 5 lp/mm on-print resolution would require 106 lp/mm resolution on the sensor. The equivalent full-frame resolution would be 71 lp/mm. So, the choice between formats for equivalent cameras might well be decided by the quality of available lenses — or lens + sensor combinations.[16]

5. Optimal Aperture and Non-equivalent Cameras

What happens if we apply optimal aperture theory to non-equivalent cameras? Given the availability of 36 MP full-frame cameras (as of late 2013), how would a 36 MP full-frame camera compare to a 24 MP APS-C format camera when both are used at their optimal apertures? Table 3 considers the case where the extra resolution of the full-frame sensor enables printing at a larger size while still maintaining a print resolution of 300 ppi. Table 4 illustrates what happens when the the two cameras are used to print at the same size, but the greater

---

[16] In fact, I have been unable to find any test of a lens on a 24 MP APS-C camera in which the reported resolution was greater than about 90-94 lp/mm, which is about 70-73 % of the theoretical maximum linear resolution of such a sensor. It may well be that no current APS-C camera and lens are adequate if the required resolution on a 20-inch print is 5 lp/mm.

resolution of the full-frame camera is used to achieve a smaller on-print blur spot (0.163 mm instead of 0.2 mm). Note that the optimal aperture for the full-frame camera is the same in both cases. The clear message of these comparisons is that the APS-C camera has a distinct depth of field advantage when both cameras are used at their optimal apertures — compare hyperfocal distances. In fact, this depth of field advantage extends to all apertures permissible for both cameras (data not shown). In effect, the greater print size (at equal resolution and sharpness), or

Table 3. Camera Non-Equivalence When Optimal Aperture is $N_{max(DoF)}$ — Constant Print Resolution and Constant Print Blur Spot Size*		
	APS-C	**Full-frame**
Millions of pixels	24 (6000 x 4000)	36 (7350 x 4900)
Sensor dimensions (mm)	24 x 16	36 x 24
Sensor pixel pitch (μm)	4	4.9
Lens focal length (mm)	50	75
Horizontal field of view (º)	27	27
Final print length (inches @ ppi)	20 @ 300 ppi	24.5 @ 300 ppi
On-print blur spot size (mm)	0.2	0.2
Print magnification	21.167	17.286
Total on-sensor blur, c_t (μm)	9.449	11.570
Optimal aperture, $N_{max(DoF)}$	4.982	6.101
Diffraction blur, c_a (μm)	6.686	8.188
Defocus blur, c_d (μm)	6.686	8.188
Hyperfocal distance (m)**	75.1	112.6

 * Calculations assume λ= 550 nm
 ** The APS-C camera will have greater depth of field when both cameras are used at their optimal apertures.

equal print size (at greater resolution and sharpness) that are possible with the full-frame camera must be "paid for" with decreased depth of field.[17] This "cost" is avoided only if the nearest plane of the object field that is required to be sharp is at least 56.3 meters from the camera (112.6 / 2). Note also that the on-print resolution with the full-frame system will now require 86

---

[17] This should not come as too great a surprise. Consider, for example, the comparison in Table 3, in which we are making larger prints from the full-frame camera but **maintaining the same on-print resolution.**

lp/mm on the sensor; so that the demands on the full-frame sensor + lens combination, while still not as great as those on the APS-C combination, are greater than for 24 MP full-frame system.

Table 4. Camera Non-Equivalence When Optimal Aperture is $N_{max(DoF)}$ — Constant Print Size*		
	APS-C	**Full-frame**
Millions of pixels	24 (6000 x 4000)	36 (7350 x 4900)
Sensor dimensions (mm)	24 x 16	36 x 24
Sensor pixel pitch (μm)	4	4.9
Lens focal length (mm)	50	75
Horizontal field of view (º)	27	27
Final print length (inches @ ppi)	20 @ 300 ppi	20 @ 367.5 ppi
On-print blur spot size (mm)	0.2	0.163
Print magnification	21.167	14.111
Total on-sensor blur, c_t (μm)	9.449	11.57
Optimal aperture, $N_{max(DoF)}$	4.982	6.101
Diffraction blur, c_a (μm)	6.686	8.188
Defocus blur, c_d (μm)	6.686	8.188
Hyperfocal distance (m)**	75.1	112.6

* Calculations assume λ= 550 nm
** The APS-C camera will have greater depth of field when both cameras are used at their optimal apertures.

6. A Caveat and Final Thoughts on Sensor Size

At this point, it seems appropriate to ask: how confident are we in this analysis of optimal aperture? As I will describe below, a controlled test of the theory provides good evidence that the theoretical analysis is substantially correct, at least for one type of camera. At the same time, it is worth remembering that in practice, calculated optimal apertures are likely to be approximate. Although in several instances, I have calculated apertures to two or three decimal places, that was done only to avoid rounding errors in subsequent calculations. It is a good thing that today's cameras and lenses do not permit adjustments of aperture finer than 1/2 or 1/3 f-stop, because our formulas for optimal aperture are almost certainly not any more precise than that, and may be less so.

There are several reasons for caution. First, the analysis treats defocus and diffraction blurs as circles with uniform intensity and sharp, clearly demarcated edges. That is certainly not the case for diffraction blur. It is not at all clear, *a priori*, how much of the Airy diffraction

pattern should be included in the calculation of diffraction blur. The usual practice of taking the diameter of the central bright disk to the first minimum seems reasonable. However, it is essentially arbitrary. For example, lens aperture openings are not circular and may not be symmetrical. Deviations from circularity might have considerable effects. In the absence of any contrary evidence, I have used $\alpha = 1.0$ for the value of the diffraction blur coefficient in sample calculations throughout this paper. In the next paper in this series, I will describe an experimental estimate of its true value. Secondly, I am unaware of any rigorous theoretical basis for calculating total blur as the root sum of squares of the equally-weighted defocus and diffraction blurs (Eq. 1(a)). Peterson and Hansma use that equation without justification. Certainly, it has intuitive appeal. For instance, if one blur circle is much larger than the other, it seems reasonable that the smaller source of blur will be "swamped" by the larger, and there will be little, if any, increase in total blur due to the smaller source. Total blur, by Equation 1, behaves approximately that way, as we saw in Section 1.4. Similarly, if two blur circles of equal size are laid directly on top of one another, we would probably not expect the total blur to be twice the individual blurs. Again, Equation 1 conforms to expectation: total blur in that case would be only $\sqrt{2}$ times the diameter of the individual blur spots. Thus, Equation 1 exhibits properties that we intuitively feel are reasonable, but that is no guarantee that it is in any way exact. That said, Hansma does provide data obtained from tests with view cameras that agree remarkably well with the predictions of the theory outlined above, specifically with respect to $N_{min(Blur)}$.

Although the discussion of equivalent cameras indicates a possible advantage for cameras with smaller sensors (assuming shallow depth of field is not the goal), that possible advantage must be weighed against the less favorable noise characteristics of smaller sensors, and the greater resolution required for lenses on smaller sensors. If images will be printed, the degree of magnification will affect acceptable total blur on the sensor. For equal print sizes, larger format cameras will require less magnification, which will permit larger on-sensor blur spots. In addition to making less stringent demands on lens + sensor resolving power, larger blur spots permit smaller optimal apertures (Eq. 8, Fig. 3). Smaller optimal apertures are more likely to fall within the range for which lenses are well-corrected across their entire fields of view — usually f/4 or f/5.6 and smaller. Optimal apertures larger than f/4 should probably be avoided unless a lens is known to perform well at those apertures on the camera being used.

In the final analysis, optimal apertures given by the formulas here are most probably good starting points. They may also be the "best" in most cases, but that remains to be verified by experience, with due consideration given to the resolving power of lenses and cameras.

APPENDIX

To obtain the expressions for optimal aperture when:

$$c_{d(n)} = \frac{f(v_n - v)}{Nv_n} = \frac{f \, \partial v_n}{Nv_n}$$

$$2$$

and

$$c_{d(f)} = \frac{f(v - v_f)}{Nv_f} = \frac{f \, \partial v_f}{Nv_f}$$

$$3$$

re-write Eq. 6, substituting $c_{d(n)}^2$ for $c_{d(P)}^2$ as the first term on the right-hand side:

$$c_t^2 = \left(\frac{f}{v_n} \right)^2 \frac{(\partial v_n)^2}{N^2} + \left(\frac{2.44\alpha\lambda}{k} \right)^2 N^2$$

$$9$$

Then $N_{min(Blur)}$ is obtained by differentiating c_t^2 with respect to N, setting the result to zero and solving for N:

$$\frac{dc_t^2}{dN} = -2\left(\frac{f}{v_n} \right)^2 N^{-3}(\partial v_n)^2 + 2\left(\frac{2.44\alpha\lambda}{k} \right)^2 N$$

$$2\left(\frac{f}{v_n} \right)^2 N^{-3}(\partial v_n)^2 = 2\left(\frac{2.44\alpha\lambda}{k} \right)^2 N$$

$$N_{min(Blur)} = \sqrt{\left(\frac{f}{v_n} \right) \frac{(\partial v_n)k}{2.44\alpha\lambda}} \approx \sqrt{\frac{(\partial v_n)k}{2.44\alpha\lambda}}$$

$$10$$

Note that if $\partial v_n = 1/2 \, \partial v$, Eq. 10 and 7 give identical results. The derivation in the case where defocus blur is $c_{d(f)}$ is omitted. The expression for $N_{min(Blur)}$ is the same as Eq. 10, substituting v_f for v_n, and ∂v_f for ∂v_n.

$$N_{min(Blur)} = \sqrt{\left(\frac{f}{v_f} \right) \frac{(\partial v_f)k}{2.44\alpha\lambda}} \approx \sqrt{\frac{(\partial v_f)k}{2.44\alpha\lambda}}$$

$$11$$

Last revised: 7 May 2014

Note that these equations assume that ∂v_f and ∂v_n are expressed in microns.

In the case of $N_{max(DoF)}$, Eq. 9 is first solved for $(\partial v_n)^2$ and then differentiated with respect to N. First, let $(f / v_n) = b$. Then:

$$b^2 \frac{(\partial v_n)^2}{N^2} = c_t^2 - \left(\frac{2.44\alpha\lambda}{k} \right)^2 N^2$$

from which

$$(\partial v_n)^2 = \frac{c_t^2}{b^2} N^2 - \left(\frac{2.44\alpha\lambda}{bk} \right)^2 N^4$$

Therefore:

$$\frac{d(\partial v_n)^2}{dN} = 2 \frac{c_t^2}{b^2} N - 4 \left(\frac{2.44\alpha\lambda}{bk} \right)^2 N^3$$

Setting this derivative to zero and dividing by N, we have:

$$2 \frac{c_t^2}{b^2} = 4 \left(\frac{2.44\alpha\lambda}{bk} \right)^2 N^2$$

from which:

$$N_{max(DoF)} = 0.29 \frac{c_t k}{\alpha\lambda} \qquad\qquad 12$$

Which is exactly the same as Eq. 8. It follows that substituting $c_{d(f)}$ for $c_{d(n)}$ in Eq. 9 will also lead to the same result for $N_{max(DoF)}$.