Some Principles of Photographic Optics

Douglas A. Kerr, P.E.

Issue 2
September 5, 2004

ABSTRACT

In this article, we review a number of areas of optics that are especially pertinent to the field of photography, including focal length, focus, magnification, exposure, aperture and f/number, field of view, and depth of field. Basic mathematical formulas for factors of interest are given.

INTRODUCTION

Our approach

In many cases, to facilitate understanding of a principle, we will introduce a topic by stating the principle as it applies to an restricted, idealized situation ("for a point on the lens axis", "for an object at a great distance", and so forth). In some such cases, after the principle has been established, we will then describe how it applies to the more general case.

So as not to slow down the process of introducing concepts, we will often not mention the restrictive conditions at the outset, rather revealing them a little later.

The "patch"

Often we will speak of a very small area on an object (that is, part of a scene) or image, which we will refer to as a "patch"[1]. It may seem that we could just refer to a "point", but in fact, since a point has no size, no light can be emitted from it, and thus a point can't serve as a proper "source" in photometric discussions.

Film

The receptor for the image developed by the lens in a camera may be photographic film, an analog electronic "target" (as in an analog television camera), or a digital sensor array. Most of the principles described here apply equally to any of these. For conciseness, we will just refer to the receptor consistently as the *film*.

[1] The mathematician would speak of this as a "differential area".

FOCAL LENGTH AND FOCUS

Focal length of a lens

An important optical parameter of a lens is its *focal length*. Consider a basic, "thin", single-element lens. Imagine an object at a very great distance (actually, an "infinite" distance). All the light rays from a given patch on that object that are able to pass through the lens are brought to convergence ("focus") at a point behind the lens at a distance *f* from the lens. *f* is said to be the focal length of the lens.

To be precise, this is only exactly true:

- For an object patch that is on the lens axis
- For those rays from the patch that pass through a very small circle in the center of the lens.

With respect to the latter restriction, if we in fact consider *all* the rays from the patch that enter the lens, the ones entering the outermost portion of the lens will be converged at a point closer to the lens than the rays entering the center of the lens. This is the manifestation of *spherical aberration*, one of the classical lens *aberrations* (imperfections in behavior).

The image

In a camera, the collection of patches created behind the lens, each resulting from the convergence of the rays of light emanating from a patch in the scene, is called the *image*. We allow it to fall on the film in order to be recorded.

The focus equation

If we have an object not at an "infinite" distance, but rather at a distance *P* in front of the lens, the rays from a patch on the object will converge into an image patch lying a distance *Q* behind the lens. *P* and *Q* are related by:

$$\frac{1}{P}+\frac{1}{Q}=\frac{1}{f} \qquad \text{Equation 1}$$

where *f* is the focal length of the lens.

(As before, this relationship is only precise for patches lying on the lens axis and for rays entering at the center of the lens.)

We can rewrite this as:

$$Q=\frac{Pf}{P-f} \qquad \text{Equation 2}$$

Note that if *P*, the object distance, becomes infinite, this degenerates into:

$$Q = f \qquad \text{Equation 3}$$

the result we in fact stated earlier for an infinite object distance.

Beyond the thin lens

If the lens of interest is not a "thin" lens, either in that it is a single element of substantial thickness or (as in the case of most photographic lenses) that it is made of a number of separate elements, the equations above still hold. However, the distances *P* and *Q* cannot now just be said to be measured "from the lens". Rather *P* is measured to a point known as the *first principal poin*t of the lens, and *Q* is measured to a point within the lens known as the *second principal point* of the lens.

Note that while both of the principal points are generally within the lens assembly itself, there are some special, widely-used lens designs in which one of the points or the other is outside it.

MAGNIFICATION OF A CAMERA LENS

The magnification of a camera lens is defined as the ratio of the size of some feature on the image to the size of the corresponding feature on the object itself. It is solely a function of the object and image distances, as follows:

$$m = \frac{Q}{P} \qquad \text{Equation 4}$$

where *m* is the magnification and *Q* and *P* are the image and object distances, measured to the appropriate principal point. (Again, the familiar condition of an object on the lens axis applies.)

Substituting from Equation 1, we then get:

$$m = \frac{f}{P - f} \qquad \text{Equation 5}$$

or

$$m = \frac{Q}{f} - 1 \qquad \text{Equation 6}$$

two forms that can be useful in various further work.

Since magnification is a ratio, we can express it in various ways, including the following (the examples are for a magnification of one-fifth):

- 1/5
- 1/5X
- 1:5
- 0.2
- 0.2X
- 0.2:1

We are often especially interested in magnification in connection with closeup photography (including **macrophotography**, which is defined as the photography of very small, but not microscopic, objects). There, it is typically the maximum magnification of the lens which is of interest. This occurs with the closest available focusing distance[2].

Our interest there in magnification is due to our desire to have the image of the small object fill a substantial portion of the film frame. If we know the sizes of the object and the film frame, we can determine the magnification required to meet that objective.

Note that it is not usually possible to calculate the maximum magnification of a lens just from knowledge of the focal length and the specified closest focusing distance (perhaps using Equation 5), since the specified closest focusing distance is usually defined from the film plane and not the first principal point of the lens (that is, we do not actually know *P*).

Other uses of the term magnification

The term *magnification* with respect to a lens is sometimes used with other meanings, some of which are ambiguous, some questionable, and some downright invalid. We will not discuss those meanings here.

EXPOSURE AND APERTURE

Exposure

The quantitative phenomenon that characterizes the impact on the film that causes the film to record the image is known as *exposure*. It is defined as the product of

---

[2] For any particular focal length, in the case of a zoom lens. We cannot conclude that the greatest maximum magnification occurs with the longest focal length or the shortest—it could even occur for some focal length in between, depending on the lens design.

the *illuminance* upon the film (luminous flux per unit area) and the length of time that illuminance persists (the "exposure time")[3].

If we set aside a couple of small wrinkles, which we will look into later, the value of exposure on a patch of the image results from the interaction of these three factors:

1. The luminance (brightness) of the corresponding patch of the object ("scene")

2. The exposure time ("shutter speed")

3. The relative aperture of the lens (expressed as an "f/number").

Just to confuse things, we must note that a second, equally-legitimate use of the term *exposure* is to represent the combined effect of only factors 2 and 3.[4] Thus, when we encounter the term, we must carefully consider the context to be certain that we appreciate it with the proper meaning.

Relative aperture

The *relative aperture* of a lens (usually just called *aperture*) is described by the *f/number*, defined as the ratio of the focal length of the lens to the diameter of the *entrance pupil*.

What is the entrance pupil? In most lenses, we have an *aperture sto*p, an opening of adjustable diameter in an opaque plate (the *diaphragm*) someplace in the path of the light through the lens. The diameter of the opening is changed in order to control the amount of light passing through the lens and thus to affect exposure. The entrance pupil is the aperture stop as it appears (in both location and diameter) from in front of the entire lens.

The relative aperture is commonly stated as an "f/number", this way: "f/3.5". This in fact means that the diameter of the entrance pupil is the focal length of the lens (*f*) divided by 3.5.[5]

---

[3] In fact, for an exposure involving extreme values of exposure time, the impact on the film may not be the same as for the same exposure involving a shorter exposure time (and correspondingly greater illuminance). This phenomenon is known as "reciprocity failure".

[4] This is the factor that, in a logarithmic form, as designated "exposure value" (Ev).

[5] The numerical value is a ratio and can also be stated as such. Under the provisions of an international standard, the maximum relative aperture of a lens is marked on the lens this way: "1:3.5" (for an f/3.5 maximum aperture).

People are often mystified as to why the factor involves the focal length of the lens. After all, isn't it merely the diameter of the entrance pupil (and thus its area) that governs how much of the light emitted by a patch on the object is collected by the camera?

Indeed it is, but it is not only the amount of light gathered in which we are interested.[6] Rather, what we are concerned with is how the lens transforms the *luminance* (brightness) of a patch on the object into the *illuminance* delivered to the film to form the corresponding patch of the image. This involves both the amount of light gathered through the entrance pupil from the object patch and the size of the resulting image patch. Because of the latter, the distance from the lens to the film is also involved.

That transformation is described by[7]:

$$E = L\frac{\pi D^2}{4Q^2} \qquad \text{Equation 7}$$

where *E* is the illuminance given to a patch on the image[8], *L* is the luminance of the source object patch, *D* is the diameter of the entrance pupil, and *Q* is the distance from the second principal point of the lens to the film.

This can be rewritten as:

$$E = L\frac{\pi}{4}\frac{1}{\left(Q/D\right)^2} \qquad \text{Equation 8}$$

From this form we can clearly see that the ratio *Q*/*D* describes the performance of the lens in turning object luminance into image illuminance.

Now recall that, for a large value of P (that is, for an object whose distance is large compared to the focal length):

$$Q = f \qquad \text{[Equation 3]}$$

Thus, in that situation, the factor of importance becomes *f*/*D*, which is of course the f/number.

---

[6] It is sometimes said that the f/number describes "the light-gathering power" of the lens, but that concept is misleading, as we will see shortly.

[7] The equation is given for SI (metric) units.

[8] Note that the symbol *E* is also used for *exposure* (in the first sense mentioned above).

Since most (but certainly not all) of our photographic work is done with objects at substantial distances, we can in most cases conveniently use the f/number of the lens to characterize its role in affecting exposure. We will later talk about the case where the assumption of great object distance does not hold (in "closeup" photography).

The "stop"

In some early cameras, adjustment of the aperture was achieved by having a metal strip, carrying holes of different diameter, which was slid through the lens. The photographer positioned the strip so the hole of the desired diameter was in line with the lens axis. The different holes were said to be different "stops".

Most commonly, the sizes of the holes were such that the area of adjacent holes differered by the factor 2 (the diameter by the square root of 2), as this provided a sufficiently-fine adjustment for the technique of the time.

As a result, a change in aperture giving twice (or half) the area is said to be a "one-stop" change. This notation is also extended to other factors affecting exposure, including shutter speed, where a change in shutter speed by a factor of two is said to be a "one-stop" change.

Today, apertures and shutter speeds can often be set in the camera with an increment of one-half or one-third stop. For example, in the case of aperture area, a change of one-third stop represents a ratio of the third root of two; for aperture diameter, the sixth root of two. The ratios involved for these different increments are given in this table:

	Shutter speed or aperture area		Aperture diameter	
"Stops"	Increase	Decrease	Increase	Decrease
1	2.00	0.500	1.41	0.707
1/2	1.41	0.707	1.19	0.841
1/3	1.26	0.793	1.12	0.891

Closeup photography

When we work in a regime where the assumption of a great object distance is no longer valid (such as closeup photography), the f/number is no longer a good approximation to Q/D, which as we recall is the actual lens parameter affecting exposure. We usually overcome this not by actually using the ratio *Q*/*D* instead of the f/number, but rather by using a correction factor by which we multiply the f/number to get the "effective f/number" for the situation (which is in fact *Q*/*D*). This correction factor is often called the "bellows factor"[9]. Thus:

$$N' = BN$$

where *N*′ is the "effective" f/number, *B* is the correction factor, and *N* is the actual f/number of the lens.

B can be calculated as:

$$B = \frac{F}{Q} \qquad \text{Equation 9}$$

Of course, we rarely know Q. But often in closeup photography, we may know the magnification involved. Then we can determine the correction factor, *B*, as:

$$B = m + 1 \qquad \text{Equation 10}$$

where *m* is the magnification.

Thus, at a magnification of 1.0 (1:1), *B* becomes 2, and the effective f/number, *N*′, is twice the f/number itself. This represents a "two-stop" decrease in the photometric performance of the lens compared to what the f/number would imply.

Lens transmission

The demonstration above that the f/number is the indicator of lens behavior in "connecting" object luminance to image illuminance—for the case of an object distance that is great compared to the focal length—assumes that all the light gathered by the lens ends up on the image.

In reality, a portion of the light gathered by the lens is redirected by reflections at the various glass-to-air surfaces of the lens, and is lost to the image. The ratio of the light delivered by the lens to the image to the light collected by the lens is described by the factor *T*, the *lens transmission*. We ignore this when we treat the f/number as the indicator of lens photometric performance.

[9] On cameras in which the lens was connected to the body with a bellows in order to accommodate movement of the lens for focusing, values of this correction factor were described as depending on "bellows extension". Values of the factor were sometimes presented on a scale on the rail on which the lens board traveled.

However, in fields such as professional motion picture photography, where it is perhaps more critical to calculate the exposure factors precisely, there is a system called the "T-stop" system that does reflect the effect of lens transmission.[10]

The T-stop rating of a lens is essentially the "effective f/number" of a lens, taking transmission into account. It can be used in place of the f/number in precise exposure calculation. It is often expressed this way: T/3.5 (by parallel to the f/number system), or sometimes as T3.5 or T-3.5.

The T-stop value, N_T, is defined as:

$$N_T = \frac{N}{\sqrt{T}}$$ Equation 11

where N is the f/number of the lens and T is the transmission.

Note that the T-stop doesn't wholly replace the f/number in the cinematographer's concern. It is still the f/number that controls such matters as depth of field.

FIELD OF VIEW

Field of view refers to the amount of "space" taken in by the camera in forming the image. It is properly described in terms of the angle(s) subtended by the view. If the image is rectangular (as in most cameras), we may choose to describe the size of the field of view in terms of its width, height, and/or diagonal size (angle).

If the camera is focused at infinity, the angular field of view is closely given by:

$$\theta = 2\arctan\frac{x}{2F}$$ Equation 12

where θ (Greek letter *theta*) is the angular size of the dimension of interest of the field of view, x is the size of that dimension of the camera film frame, F is the focal length of the lens, and *arctan* represents the trigonometric function *arc tangent* (inverse tangent). (The arc tangent of x is the angle whose tangent is x.)

The field of view angle can also be expressed in terms of the size of the field at a stated distance, as "353 feet wide at a distance of 1000 feet" (corresponding to a horizontal field of view angle of 20º). This was formerly the practice for stating the field of view of binoculars, but has today been replaced by the angle in degrees.

---

[10] The Bell & Howell *Foton* 35 mm still camera, however, did have its lenses' apertures marked in T-stop terms, likely a result of the firm's heavy involvement in professional motion picture photography.

Most photographers are not accustomed to thinking of field of view in terms of angle[11]. Rather, they learn what the photographic effect is of the field of view afforded by lenses of various focal length. Of course this relationship varies with the frame size of the camera. Since for many decades the most common type of still film camera used by advanced amateurs (and by many professional photographers) was the full-frame 35 mm camera, it is widely considered today that a useful way to describe a field of view is in terms of the focal length lens that would produce that field of view on a 35 mm camera.

Thus, when dealing with a camera having a frame size different than that of the 35 mm camera (usually smaller, as for many digital cameras), and considering a lens of a certain focal length, we often (in effect) ask the question, "what focal length lens used on a 35 mm camera would give the same field of view as **this** lens will give on **this** camera?" That focal length is often called the "35 mm equivalent focal length" of the lens of interest when used on the camera of interest. It may be calculated thus:

$$f_{35} = \frac{f}{K}$$ Equation 13

where f_{35} is the "35 mm equivalent focal length", f is the (actual) focal length of the lens of interest, and K is the ratio of some dimension of the image frame of the camera of interest to the corresponding dimension of the film frame of a 35 mm camera[12].

More commonly, we define a factor J as the reciprocal of K, so that:

$$f_{35} = Jf$$ Equation 14

Thus, for a camera whose frame is 62.5% the size of a 35 mm camera frame (in linear dimensions), the 35 mm equivalent focal length of any lens used on that camera is 1.6 times the (actual) focal length of the lens[13].

---

[11] An exception is the case of lenses having a very large angular field of view, such as "fisheye" lenses, for which the field of view is commonly in fact expressed in degrees.

[12] Note that if the frame of the camera of interest does not have the same aspect ratio (ratio of horizontal to vertical size) as the frame of a 35 mm camera (3:2), a unique value of this ratio does not exist. In such case, we often nevertheless still use the concept, based on the ratio of the diagonal dimensions of the respective frames.

[13] The factor we call here J is called by some the "field of view crop factor". The rationale is that the difference between the fields of view exhibited by any given focal length lens on a 35 mm camera and a smaller-frame camera is a result of the fact that the image that would have been captured by the 35 mm camera is "cropped" by the smaller frame of the camera of interest. We do not find that term attractive, and discourage its use.

Note that this does not mean that the focal length of a lens is dependent on the frame size or any other parameter of the camera on which it is used. The focal length is a property of the lens itself. The “35 mm equivalent focal length” is **not** a focal length of the lens of interest. It is merely a number that can be used to allow appreciation of the field of view given by the lens on a particular camera in terms of familiar 35 mm camera experience (which of course many users of smaller-frame cameras don’t have!).

DEPTH OF FIELD

Strictly speaking, when the lens is set at a certain focus position, only an object patch at precisely the corresponding distance will be truly focused on the film plane.

Of course, in almost all real-life photography, we are interested in capturing scene elements lying at varying distances from the lens. We are able to do so only by accepting the fact that the degree of imperfect focus afforded objects at other distances than the ideal one is “acceptable”.

The range of object distances over which misfocus is considered acceptable is known as the *depth of field* of the camera.[14]

To be able to objectively predict the depth of field we will obtain under any given situation, we must establish some objective criterion for how much misfocus we will consider acceptable.

We define our choice of this criterion on the concept of the *circle of confusion*. When focus is imperfect, the image of an infinitesimal patch of the object is not an infinitesimal patch on the image, but rather a roughly-circular pattern of finite diameter. This pattern is known as the circle of confusion. We express our adopted criterion of acceptable misfocus by stating a maximum acceptable diameter of the circle of confusion.

The actual diameter of the circle of confusion (not our criterion for its maximum acceptable diameter) depends on four parameters of the optical system:

- The distance to the object patch of interest
- The distance to the plane of perfect focus (the “focus distance”)
- The focal length of the lens

---

[14] Depth of field is an extremely complex topic, and we will only skim the surface here. A more extensive treatment of the topic is given in the companion article, *Depth of Field in Film and Digital Cameras*, by the same author.

- The actual diameter of the aperture, or, if we prefer, the aperture as an f/number[15]

The selection of a maximum acceptable diameter of the circle of confusion is not a simple one, and does not flow automatically from any simple combination of technical properties. The choice, for one thing, must be based upon some assumptions about how the image is to be viewed, and against what norms are we to judge "acceptable" misfocus.

Under one set of such guidelines, a maximum acceptable diameter of the circle of confusion is selected based on a fixed fraction of the diagonal size of the camera format (film frame or digital sensor size). Often a fraction of 1/1400 is used.

With the various factors in hand, the depth of field can be calculated approximately as:

$$D_d = \frac{Sf^2}{f^2 - SNc} - \frac{Sf^2}{f^2 + SNc}$$ Equation 15

where D_d is the depth of field, *S* is the distance to the plane of perfect focus, *f* is the focal length of the lens, *N* is the lens aperture as an f/number, and *c* is the adopted maximum circle of confusion diameter, D_d, *S*, *f*, and *c* in the same unit.

The approximation is closely valid for values of *S* which are many times the focal length, *f*.

Although it is difficult to see from this equation the effect of changes in the various parameters, perhaps most important is the fact that, for any given values of *S*, *f*, and *c*, the depth of field increases as the f/number (*N*) increases; that is, the smaller relative aperture gives greater depth of field.

The hyperfocal distance

For given values of *f*, *N*, and *c*, there is a focus distance *S* such that the far limit of the depth of field reaches just to infinity. That value of *S* is called the hyperfocal distance for that camera setup, S_h. With the camera focused at distance S_h, the near limit of the depth of field is at $S_h/2$.

Thus, in situations in which it is not possible to focus the camera (perhaps even in a "fixed-focus" camera), setting the focus distance to the hyperfocal distance yields the greatest possible field of view, which hopefully will accommodate most of the photographic needs of the user.

---

[15] Since focal length is a parameter anyway, we can recast the defining equation to accept aperture as an f/number.

The hyperfocal distance is given approximately by:

$$D_h = \frac{f^2}{Nc}$$ Equation 16

Depth of focus

A related property, *depth of focus*, is often confused with depth of field.

If we have an object lying in only one plane, and move the film forward or backward from its position that gives perfect focus, we find that the image becomes misfocused. In effect, moving the film changes the object distance for perfect focus so it no longer corresponds to the actual distance to our object.

Depth of focus is the range of positions of the film plane over which acceptable focus is maintained, for an object at a given distance.

This is reckoned in a way parallel to the concept of reckoning depth of field, and involves the now familiar concept of an adopted criterion for the maximum acceptable diameter of the circle of confusion resulting from imperfect focus.

Depth of focus is of greatest interest in considering such things as the impact of accidental shift in the position of the film plane owing to imperfect film guidance.

#