Derivation of the "Cosine Fourth" Law for Falloff of Illuminance Across a Camera Image

Douglas A. Kerr, P.E.

Issue 4
May 1, 2007

ABSTRACT

In a photographic system, for a given object luminance (brightness), the image illuminance on the film or equivalent declines as we move outward from the center of the image as a result of the geometric optics involved. The result is a relative darkening of the image toward its borders. If we consider a lens having certain ideal properties, it can be shown that the decline in relative illuminance goes very nearly as the fourth power of the cosine of the angle by which the object point is off the camera axis.

Here we derive this relationship. We also discuss differing results given by other authors.

INTRODUCTION

In photography, when we set the aperture of the lens to a certain f/number, we are actually controlling the relationship between the luminance at each small area in the scene and the illuminance upon the film (or equivalent) at the corresponding area in the image. That illuminance, in concert with the exposure time (shutter speed), determines the exposure on the film at that point.

Ideally, for any given aperture that relationship would be constant for every part of the image. However, unavoidable matters of geometric optics result in the image illuminance (for a given scene luminance) declining as we move outward from the center of the frame. The decline can be substantial, and is a common cause of "darkening" of the final image near the edges and corners of the frame (sometimes spoken of as *natural vignetting* or "light falloff").

Prediction of this phenomenon for a typical real lens design is highly complex. However, if we consider a lens having certain ideal properties, it can be shown that a simple mathematical function very closely approximates phenomenon. It is often called the "$\cos^4$" or "cosine fourth" law[1], for reasons that will soon become obvious. We will derive it here.

DERIVATION OF THE FALLOFF FUNCTION

The approach

We will attack this issue with a very clean approach, which we will summarize here:

- We consider a small off-axis region of the object ("scene") having a certain luminance.
- We determine the amount of luminous flux from that region that is captured by the entrance pupil of the lens.
- We recognize that all of that luminous flux is inevitably deposited on the film on the region that is the image of the object region.
- We recognize that (by definition) the *illuminance* on that image region is the ratio of that amount of luminous flux to the area of the region.

Note that by using this approach, we have no need to consider what goes on between the lens and the film (other than to determine the area of the image region). That's good, for doing so is very tricky.

Our lens model

Our analysis is predicated on a hypothetical lens whose performance is ideal in the following ways:

- The lens has 100% transmission and no flare (that is, all the light captured is deposited properly onto the image, none being lost by absorption or by "scattering" to unrelated parts of the image).
- The lens is rectilinear (that is, it has no geometric distortion; it gives constant magnification across the entire object).
- The entrance pupil of the lens is located at the first principal point of the lens. (The entrance pupil is the virtual image, from in front of the lens, of the aperture stop. It is the "port" through which the lens appears to collect light from the object.)

---

[1] "Law" is used here in the sense of a known relationship between variables, as in the "E *vs*. I" law of a semiconductor diode, rather than in the sense of a natural law.

- When observed from an off-axis location, the projected area of the entrance pupil relates to its "head on" area just as if the pupil were a physical circular opening in a plate.

We can follow the derivation on Figure 1, which (for simplicity) shows a "thin" lens (although there is no loss of generality from so doing).

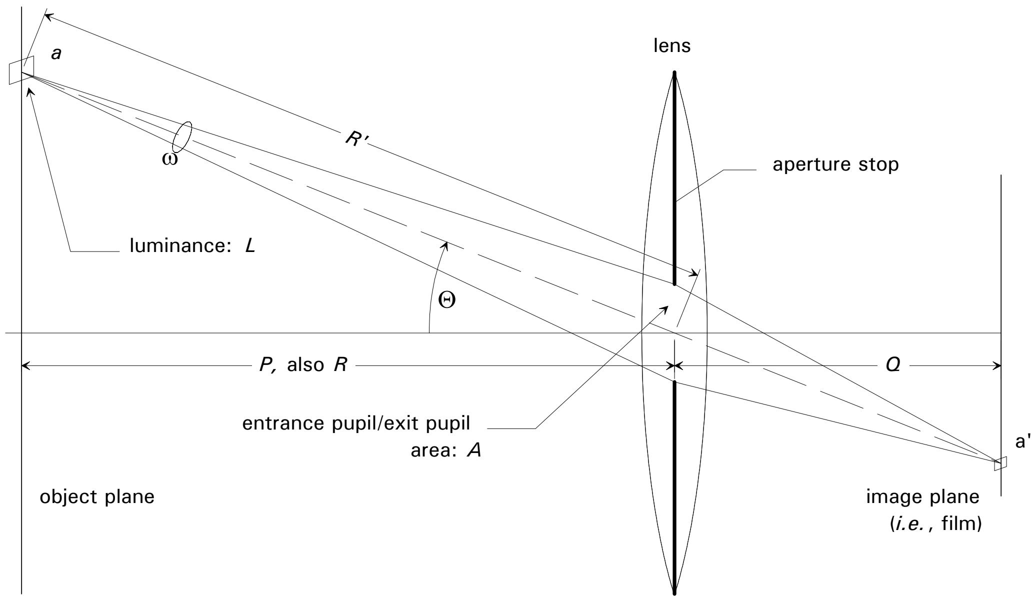

Figure 1.

We consider an object plane at a distance from the lens *P*, and on it a differential (infinitesimal) area of the object, *a*, off the axis of the camera by an angle Θ, and having a luminance *L*. The distance from the object plane to center of the entrance pupil is *R*; in this case, it is equal to *P*, but we label it separately for generality.

The projected area of the differential area toward the lens, a_p, is:

$$a_p = a\cos\Theta \qquad [1]$$

We assume that the object surface is a diffuse Lambertian reflector, and thus exhibits its luminance from any angle of observation. Thus the luminous intensity, I, emitted from the differential area[2], in the direction toward the center of the entrance pupil, is given by:

$$I = La_p \qquad [2]$$

[2] We note that, strictly speaking, only a point source has a luminous intensity. However, we can usefully attribute the property of luminous intensity to an extended (surface) source of infinitesimal size.

Substituting from equation 1:

$$I = La\cos\Theta \qquad [3]$$

Let the area of the entrance pupil be *A*. The projected area of the entrance pupil as seen from the differential object area, A_p, is given by:

$$A_p = A\cos\Theta \qquad [4]$$

The slant distance from the differential area to the center of the entrance pupil, *R′*, is given by:

$$R' = \frac{R}{\cos\Theta} \qquad [5]$$

The solid angle subtended by the projected area of the entrance pupil, ω,, seen from the differential area, is approximately[3] given by:

$$\omega = \frac{A_p}{(R')^2} \qquad [6]$$

Substituting for A_p and *R′* from equations 4 and 6 gives us:

$$\omega = \frac{A\cos^3\Theta}{R^2} \qquad [7]$$

The luminous flux, Φ, from the differential area that is captured by the entrance pupil is given by:

$$\Phi = I\omega \qquad [8]$$

Substituting for I and ω from equations 2 and 7 gives us:

$$\Phi = \frac{LaA\cos^4\Theta}{R^2} \qquad [9]$$

Now we must determine the area of the image of the object area. The magnification of the system, *m*, is given by[4]:

---

[3] The approximation is very close if the distance to the object is large compared to the diameter of the aperture. The precise expression is very painful.

[4] This is in particular the expression for the magnification at the axis. Recall however that we have assumed a constant magnification across the field, and thus this expression applies overall.

$$m = \frac{Q}{P} \qquad [10]$$

The area, *a'*, of the image of the differential area of the object, *a*, is then given by:

$$a' = m^2 a \qquad [11]$$

Substituting for *m* from equation 10 gives us:

$$a' = \frac{aQ^2}{P^2} \qquad [12]$$

All the flux collected by the lens from the object area (Φ in equation 9) is unavoidably deposited on the corresponding image area (since we have assumed no loss by absorption, scattering, etc.) **This concept is central to our approach to deriving the falloff relationship**, and may be spoken of as the "doctrine of conservation of luminous flux".

That being the case, then, based on the fundamental photometric definition, the illuminance, *E*, on the differential area of the image, *a'*, is given by:

$$E = \frac{\Phi}{a'} \qquad [13]$$

Substituting for Φ and *a'* from equations 9 and 12 gives us:

$$E = \frac{LP^2 A \cos^4 \Theta}{R^2 Q^2} \qquad [14]$$

But, in our model, $P=R$, and so this becomes:

$$E = \frac{LA \cos^4 \Theta}{Q^2} \qquad [15]$$

If we evaluate *E* for an off-axis object area (E_Θ) and for an on-axis area (E_0) (the two areas having equal luminance) and compare them by taking their ratio, *L*, *A,* and Q^2 (which are constant) cancel out, and we get:

$$\boxed{\frac{E_\Theta}{E_0} = \cos^4 \Theta} \qquad [16]$$

This tells us that for a given object luminance and a given lens "setup", the illuminance on the film plane falls off as the fourth power of the cosine of the off-axis angle of the area on the object, where that angle is measured in "object space" from the center of the entrance pupil.

This is the famous "cosine fourth" illuminance falloff function, *quod erat demonstrandum*. We have boxed it in for emphasis.

Countin' the cosines

After that long derivation, it may be helpful to review the source of the four $\cos\Theta$ terms in our result. They come from:

- The variation in the projected area of our differential object area as seen from the entrance pupil of the lens.
- The variation in the projected area of the entrance pupil as seen from the differential object area.
- (2 terms) The variation in the effect of the distance to the entrance pupil on the solid angle subtended by the projected area of the entrance pupil from the differential object area.

LENSES WITH DISPLACED PUPILS

In many lens designs, the entrance and exit pupils do not fall at the locations of the first and second principal points. When this is true, the pupils are said to be *displaced*. In such a case, the pupils also inevitably have different diameters, and the degree of displacement can be quantified by stating the ratio of the pupil diameters (the *pupil magnification*).

As we pointed out above, if the entrance pupil is not located at the first principal point of the lens, the distances P and R are no longer equal, and our simple result is no longer precisely correct.

However, where both P and R are large (as in many cases of interest), the difference is inconsequential, and the result given in equation 16 remains valid.

WHAT WE DIDN'T DO

Especially when we compare the derivation we used above with those put forth by some other authors, what stands out is what we didn't do.

We didn't look at all into what goes on between the lens and the film (other than to determine the area, a', of the image of the differential object area). We didn't look into the diameter, or axial location, of the exit pupil of the lens, or how its projected area varies with angle of observation; nor the distance from the exit pupil to the differential area of the image, or how it varies with where the area is in the frame; nor the angle, Θ_X, by which area *a′* is off axis, measured from the center of the exit pupil. We didn't have to, because once we:

- Determined the amount of luminous flux from a differential area of the object that is captured by the lens (that is, by its entrance pupil),
- Recognized that all this flux must land on the film within the differential area of the image corresponding to the differential area of the object, and
- Recognized that the illuminance on the differential area of the image is the ratio of the flux landing on it to its area (regardless of angle of incidence),

then we were done. What goes on between the lens and the film must work in unavoidable fulfillment of the scenario just summarized, and we need not look into it.

RESULTS OF OTHER AUTHORS

Other authors present slightly different results for the falloff function. They differ from the result given above in equation 16 only for the situation in which the off-axis angle to an image point, measured at the exit pupil (said to be in *image space*, the region between the lens and the film) differs from the off-axis angle of the corresponding object point, measured at the entrance pupil (in *object space*, the region between the object and the lens).

That difference only occurs when the pupils are not collocated with the principal points (that is, where there is pupil displacement). Pupil displacement can, however, be quite substantial for certain classes of lens design. Thus we cannot dismiss the discrepancy between the results of these authors and the result derived here as being insiginifcant or only pertaining to peculiar situations. We must address the discrepancy head on.

For reference, first recall that our result for the falloff function is:

$$\frac{E_\Theta}{E_0} = \cos^4 \Theta_N \qquad [16]$$

where E_0 is the film illuminance (for a given scene luminance) at a point on-axis, E_Θ is the illuminance (for that same scene luminance) at a point off-axis, and Θ_N is the angle by which the object point is off-axis, measured (in object space) from the center of the entrance pupil. The subscript *N* (for entrance), first seen here, prepares us to distinguish this angle from another off-axis angle we will encounter shortly.

The "angle in image space" result

One widely-cited alternative relationship is (presented in the same form as for equation 16):

$$\frac{E_\Theta}{E_0} = \cos^4 \Theta_X \qquad [17]$$

where Θ_X is the angle by which the image point is off-axis, this time measured (in image space) from the center of the exit pupil. (The shading is to remind us that this is not part of my derivation)

This result is often presented as more accurate (on theoretical grounds) than the result I derive above.

Discussion

The derivation of this alternative result typically involves the handy premise that the exit pupil of a lens, viewed from a point in the image, behaves as if it were a luminous disk exhibiting across its entire area a luminance which is the same as the object luminance at the associated object point.

Proofs of this premise can be found in standard optical engineering texts. These, however, invariably treat the pupil as viewed from an image point on the optical axis. There is no demonstration that the premise actually holds for viewing from an off-axis point.

In fact, this premise **does** hold for viewing from a point off the axis if the exit pupil is located at the second principal point of the lens. However, it appears that it **does not** hold for viewing from a point off the axis if the exit pupil is not located at the second principal point.[5]

Nevertheless, the proponents of this alternate result for illuminance falloff proceed as if the premise applies to viewing from points off the lens axis even if the exit pupil is not located at the second principal point. This seems to be an unwarranted stretch.

---

[5] The rigorous analytical demonstration of this limitation of the exit pupil model is rather daunting (and I indeed have not yet undertaken it). But if we take the premise at face value, and consider a lens in which the exit pupil is forward of the second principal point, and an off-axis object, we find that the amount of luminous flux deposited on the image of the object is greater than the amount of luminous flux collected from the object by the lens. Thus the premise cannot be valid in that case.

Accordingly, I cannot endorse this alternate result as having a valid theoretical basis.

The COS COS^3 result

Another widely-cited alternate result is:

$$\frac{E_\Theta}{E_0} = \cos\Theta_N \cos^3\Theta_X \qquad [18]$$

where, as before, Θ_N is the angle by which the object point is off-axis, measured (in object space) from the center of the entrance pupil, and Θ_X is the angle by which the image point is off-axis, measured (in image space) from the center of the exit pupil.

The origins of this result are also rather questionable. One author has suggested that it is arbitrary, found empirically to gave the best agreement with actual behavior in many real lenses. (Rarely does a real lens exhibit precisely the $\cos^4\Theta_N$ behavior because of various departures from the ideal assumptions upon which that result is predicated.)

I cannot endorse this result as having a valid theoretical basis.

Wishful thinking

As we mentioned before, if a lens has its pupils displaced (not located at the principal points), then the angles Θ_N and Θ_X will not be equal. If the pupils are displaced forward (toward the object), then Θ_X will be smaller than Θ_N. Incidentally, in most situations (with the object at a substantial distance), we can think of angle Θ_N as indicating how far off axis an object point (and its corresponding image point) lie.

By placing reliance on the result shown in equation 17, some authors assert that a lens with the exit pupil displaced forward will exhibit less falloff (for object points off axis by a certain angle Θ_N) than a lens without pupil displacement. But, in light of the shortcomings of that alternate result, I cannot endorse this conclusion as justified.

MITIGATING THE FALLOFF

The illuminance falloff we discuss here leads to a difference in exposure over the entire frame (often manifesting itself as substantially darkened corners). There are a number of lens design techniques that can mitigate this undesirable falloff. The most important one has to do with the projected area of the entrance pupil from an off-axis object point. Note that in equation 4, in accordance with one of our assumptions about lens behavior, we assume this to be:

$$A' = A\cos\Theta \qquad [4]$$

Recall, though, that the entrance pupil is not a physical hole in a plate, but rather a virtual creation of the lens elements in front of a physical hole in a plate. By proper design of the lens, designers can arrange for the projected area of this virtual hole to decline more slowly than $\cos\Theta$ with off-axis observation, thus reducing the decline in the acceptance of light from off-axis object regions, and reducing the falloff.[6]

Even when such a deviation of the behavior of the entrance pupil from that of a simple disk is not intentional, it often occurs (sometimes in a disadvantageous direction). This is a prominent reason why the actual falloff function of real lenses departs from the "cosine fourth" relationship derived here.

Another phenomenon often encountered is overt obstruction of the light paths for off-axis object points (*obstruction vignetting*).

Both effects will generally vary with the aperture to which the lens is set.

DIGITAL CAMERAS

Some special considerations apply to the matter of exposure falloff in digital cameras.

Background

The *illuminance* on a surface is defined as the amount of luminous flux impacting the surface per unit area of the surface. It is the product of the arriving *luminous flux density*[7] and the cosine of the angle of incidence.

We did not encounter this head-on in our derivation: as a result of our reliance on the doctrine of conservation of luminous flux, we did not have to look into any details of what happened in image space! But the concept is nevertheless in fact involved in the overall physical chain.

---

[6] In effect, these lens designs cause the entrance pupil to "tilt" toward an off-axis point of observation, diminishing the decline in its projected area.

[7] We don't often hear much about *luminous flux density*; it is usually the result of an unshown intermediate step on the road from *luminous intensity* to *illuminance*.

The presence of the cosine factor is not the result of anything supernatural. It merely recognizes the fact that, for a "beam" of light arriving at other than right angles to the surface, the amount of luminous flux traveling within a certain cross-sectional area of the beam lands over an area of the surface larger than that cross-sectional area (thus, by definition, reducing the resulting illuminance).

Imagine a beam 1.0 cm square (cross-sectional area 1 cm^2), carrying a total luminous flux of 100 lumens. The luminous flux density of the beam is thus 100 lumens per cm^2.

If the beam arrives at an angle of 45° to the "normal" (the direction perpendicular to the surface), it will illuminate an area 1.0 cm x 1.414 cm. [8] Since the 100 lumens of flux lands spread out over 1.414 cm^2 of surface, the resulting illuminance on the surface is approximately 70.7 lumens per cm^2. (The cosine of 45° is approximately 0.707.)

The difference between film and a digital sensor array

For photographic film, the phenomenon which affects the film (called *exposure*, in one of two senses of the term), resulting in chemical changes that cause an image to be recorded, is quantified as the product of illuminance and exposure time (shutter speed).

In a digital sensor array, the phenomenon which affects each individual sensor element is essentially the product of the luminous flux impacting the sensor element itself and the exposure time. If we imagine each sensor element having a certain cross-sectional area, it would seem that this would be equivalent to the situation for film. But it is not that simple.

Because each sensor element proper must be surrounded by a "non-receptive" area (perhaps containing various semiconductor gates needed for the overall operation of the array), not all of the luminous flux landing on the array is fruitful. This reduces the sensitivity of the array from its potential.

To give the sensor elements the benefit of all the luminous flux "to which they are entitled", their equivalent acceptance area is increased by mounting above each element a small lens, whose entrance pupil is nearly square and almost as large as the pitch between elements.

But there is unavoidable obstruction vignetting in this system of "microlenses". As the angle of incidence of an arriving "beam" of light

---

[8] Imagine cutting a 1cm x 1cm stick at a 45° angle, and think of the shape and size of the cut end.

increases (that is, the beam becomes more oblique), the amount of luminous flux falling on the face of each lens falls as the cosine of the angle of incidence, just as we would expect. However, not all of that flux actually hits the sensor element itself, a decreasing fraction surviving as the angle of incidence increases.

Thus, for a beam of a certain luminous flux density, the amount of flux striking each sensor element falls faster than the cosine of the angle of incidence. As a result, the falloff in exposure may in fact be faster than would be indicated by the "$\cos^4 \Theta_N$" relationship we derive in this article.

Mitigation of the excess falloff

Because the "excess falloff" resulting from the behavior of the microlenses is a function (among other things) of the angle of incidence of the luminous flux on the surface of the sensor array, which is equal to angle Θ_X, lens designs having the exit pupil displaced substantially forward (in which Θ_X is substantially smaller for a given off-axis angle of an object point) will in fact exhibit less "excess falloff" in a camera of this type.

Such lenses are often characterized as having "telecentric" design (although "quasi-telecentric" would be more apt). A true telecentric lens, often used in optical measurement systems, has its exit pupil located at an infinite distance in front of the lens, and its entrance pupil an infinite distance behind the lens.

Recall that, as I discussed earlier under "Wishful thinking", this design approach does not mitigate the basic phenomenon of illuminance falloff, just the "excess falloff" that exacerbates the phenomenon in digital cameras.

#