Principle of the Split Image Focusing Aid and the Phase Comparison Autofocus Detector in Single Lens Reflex Cameras

Douglas A. Kerr, P.E.

Issue 5
August 27, 2005

ABSTRACT

Many single lens reflex (SLR) cameras are equipped with an arrangement in the viewfinder known as a *split image focusing aid*, intended to facilitate accurate visual determination of the point of proper focus when focusing manually. In this article, we explain the principle by which this arrangement operates. We also describe another related viewfinder manual focusing aid, the *microprism field*, and discuss the application of the split image principle to one type of automatic focus detection system, the *phase comparison* system.

INTRODUCTION

In a single lens reflex (SLR) camera, basic manual focusing is done by observing the image formed on a "ground glass" screen in the reflex viewfinder. However, under many circumstances, it is difficult to accurately determine the point of best focus merely by observing the sharpness of the finder image.

Many SLR viewfinders are equipped with a special feature intended to facilitate precise determination of the point of proper focus, known as a *split image focusing aid*. This involves a circle in the center of the focusing screen, divided into two portions by a horizontal, vertical, or diagonal line. The central part of the scene is seen through this circle, without substantial blurring even if the focus is somewhat incorrect. However, if the focus is incorrect, even by a small amount, the two halves of the image in the circle will not precisely line up. The human eye is very sensitive to such misalignment (one example of the *vernier acuity* of the eye), and thus the photographer can readily bring the focus to the proper point.

The widespread use of automatic focusing mechanisms in both film and digital cameras has led to SLR manufacturers less frequently including such a focusing aid (or even making it available as an option). But in fact many photographic tasks require manual focusing for best results, and thus there remains continued interest in the split image system, often in the context of its "retrofit" to existing cameras.

We will begin our investigation of the split image system by reviewing the way an image is formed in a camera, following which we will look into manual focusing with the traditional ground glass screen. We will then illuminate the underlying principle of the split image system, using a "lecture hall" arrangement not actually utilized in the real system plus a hard-to-believe assumption. We will then rationalize this fanciful demonstration with the way the system is actually implemented.

Image formation and focus in a camera

Figure 1 illustrates the basic principal of the generation of a focused image on the film of a camera.

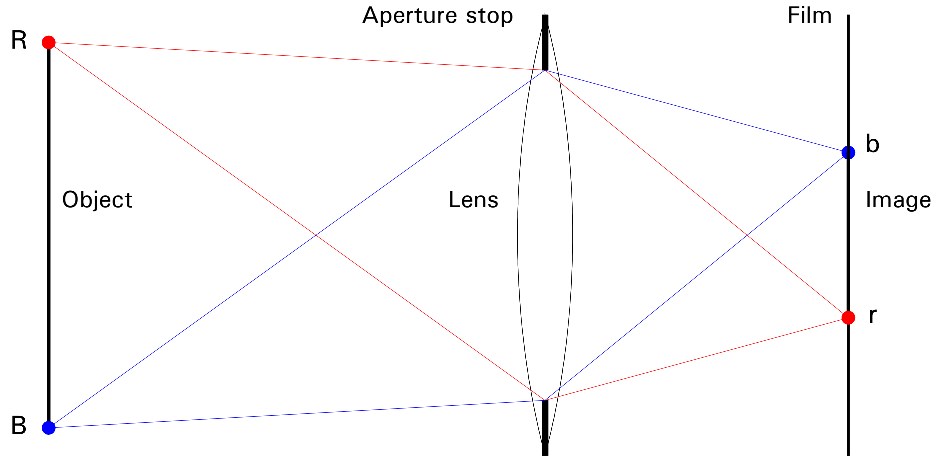

Figure 1

The lens we show is a "simple thin" lens, and we imagine that the aperture stop (whose diameter is varied to control the amount of light passed through the lens as one aspect of controlling exposure) is actually inside the lens. We adopt the fiction that the light passing though the lens is refracted ("bent") at the plane at the center of the lens (rather than both surfaces of the lens, as is actually the case), since this makes the pictures easier for me to draw and less cluttered.

On the left, we see our hypothetical object, a straight line. On it we note two tiny "patches"[1] at its ends, one red ("R") and one blue ("B").

If the position of the lens is proper (that is, the camera is properly focused on this object), the light from patch R is brought to a focus at point r on the film as a sharply-defined "image patch", and the light from patch B is brought to a focus at point b on the film.[2] The same occurs for all other patches of the object, and the resulting array of image patches constitutes a complete, sharp image of the object. (In the figure, we see only the two "outermost" rays from each object point, but the other rays follow suit.)

---

[1] These are assumed to be very tiny. It is tempting to speak of them as "points", but in fact a point has zero area and thus no light can be emitted from it.

[2] Or digital sensor; for conciseness, we will usually just say "film" from here on.

Misfocus

In figure 2, we see the situation in which the lens has intentionally been shifted “forward” to spoil the focus. In that case, the “cone of light” from either of the patches is brought to a focus at a place in front of the film, and continues on to form a finite-sized “blob” (*or circle of confusion*) on the film. The collection of these blobs constitutes a blurred image.

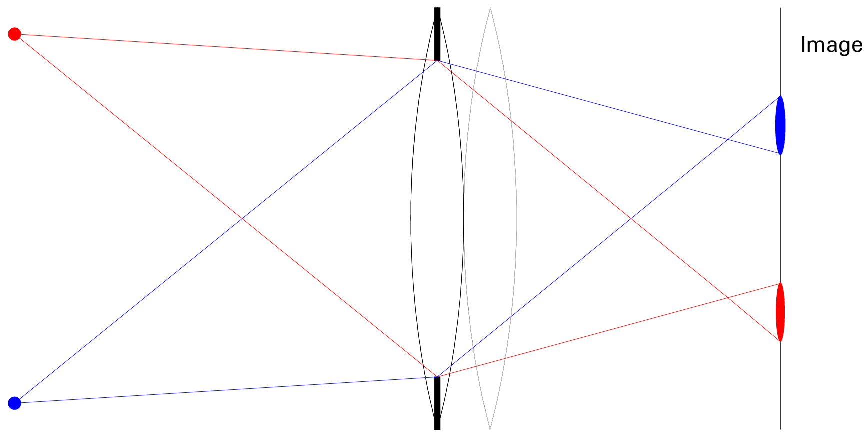

Figure 2

The ground glass focusing screen

In a single-lens reflex (SLR) camera, during the “viewing and focusing” phase, the image is not directed to the film but rather, via the reflex mirror, to a *ground glass* (diffusing) focusing screen, which we view through a special prism and an eyepiece lens. In Figure 3, we see this in straightened-out form, with the mirror, prism, and lenses omitted. The figure shows the lens properly focused on our object. (We don’t show the “shaft” of the object from here on, concentrating only on the two end patches.)

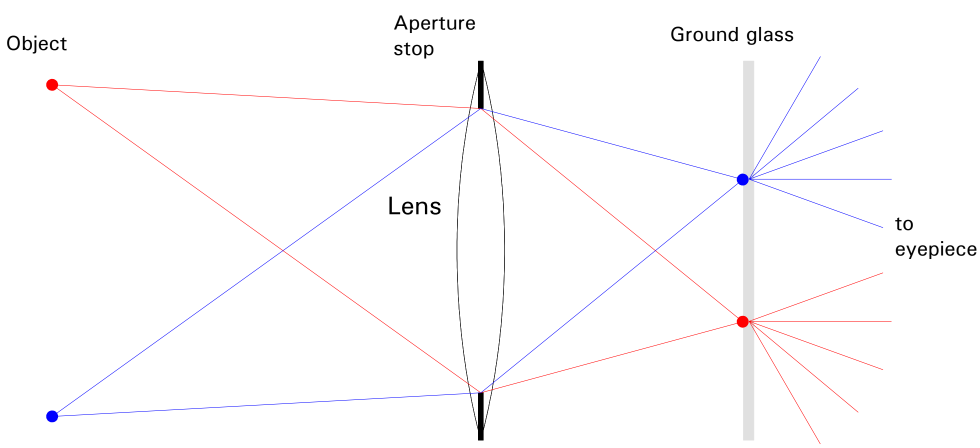

Figure 3

In this case, the patches of light forming the image fall on the face of the screen, which, being translucent, allows light from each to be emitted from the back of the screen, toward the viewfinder eyepiece system (which actually contains some of the lenses we have neglected to draw). Thus the user sees the image of the scene through the viewfinder eyepiece, in the familiar manner.

Note that the light emitted from the back of the screen is spread over a large angle (due to the diffuse nature of the ground glass screen), but its distribution is not centered about the direction toward the eyepiece (as the screen is not a perfect transmissive diffuser). The result is that only a fraction of the light emitted reaches the user's eye, not desirable from the standpoint of viewfinder image brightness.

In reality, there is a lens just behind the screen, called a *field lens*, which concentrates the emitted light from each point of the image and bends it toward the eyepiece.[3] We won't, however, show this.

Misfocus again

In figure 4, we see the situation with the focusing screen for the out-of-focus condition. The blurred image (made up of the "blobs" from each patch of the object) is seen by the photographer through the eyepiece.

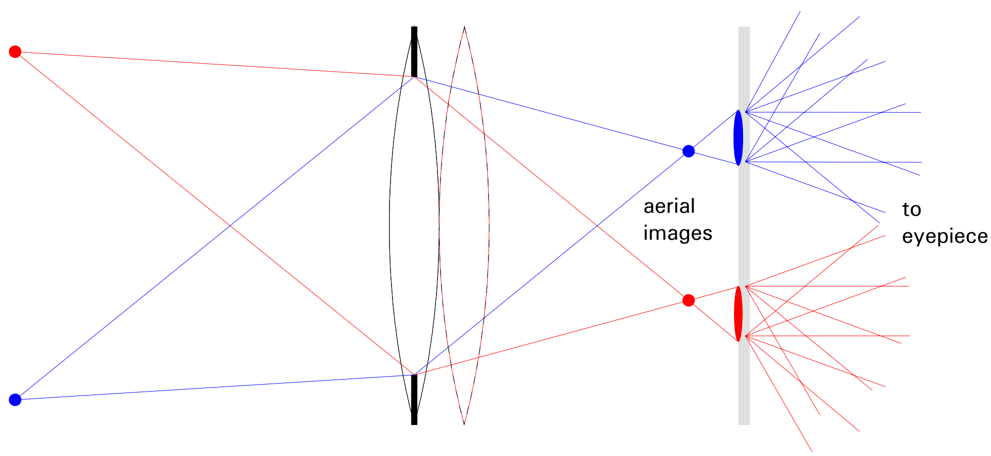

Figure 4

---

[3] This lens is often of the "Fresnel" type, made of numerous tiny concentric circular grooves which are each, in effect, little rings cut from the surface of a conventional lens. This design eliminates the thickness we otherwise find at the center of a lens.

THE SPLIT IMAGE FOCUSING AID

Introduction

Now we will begin to develop the principle of the split image focusing aid itself.

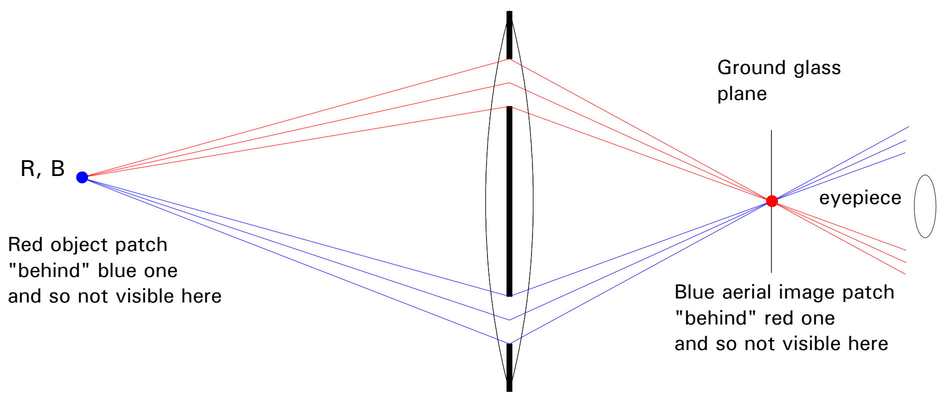

Figure 5

In figure 5, we have a new object, this one a much shorter line, and oriented horizontally—toward and away from us. We again concentrate only on a tiny patch at each end, one red and one blue. The red one is at the "away from us" end of the object so we cannot actually see it from our vantage point.

We have introduced some artifices, not part of the real arrangement, which will help the basic principle to be explained clearly:

- We replace the conventional aperture stop having a single central opening with a fanciful one having an opening at the top and one at the bottom. (We will later find out that in real life this "two-holer" is not necessary, but its presence makes the story unfold more clearly.)

- We assume that, in some magical way, only light from the red patch on the object will pass through the upper opening, and only light from the blue patch will pass through the lower one. (We will later find that something equivalent to this in fact occurs naturally in real life, but it is too early to explain how!)

We have also removed the ground glass, noting the plane where it was.

We again assume proper focus of the lens. Then, as we saw earlier, each of the two bundles of rays we are following is brought to a focus at the plane of the ground glass. But there is no ground glass for the image to fall on.

Nevertheless, that convergence of the ray bundles "in thin air" creates an image—an *aerial image* ("aerial" meaning that it doesn't fall on film or a screen). If the eye intercepts the light that has formed an aerial image (after it passes the point of convergence), the aerial image is seen just as if it were a physical object.

However, in this situation there is a problem. As we can see from the figure, these ray bundles continue on past the point of convergence at such an angle that none of the light in them would actually reach the eyepiece.

In figure 6, we see the solution to this problem.

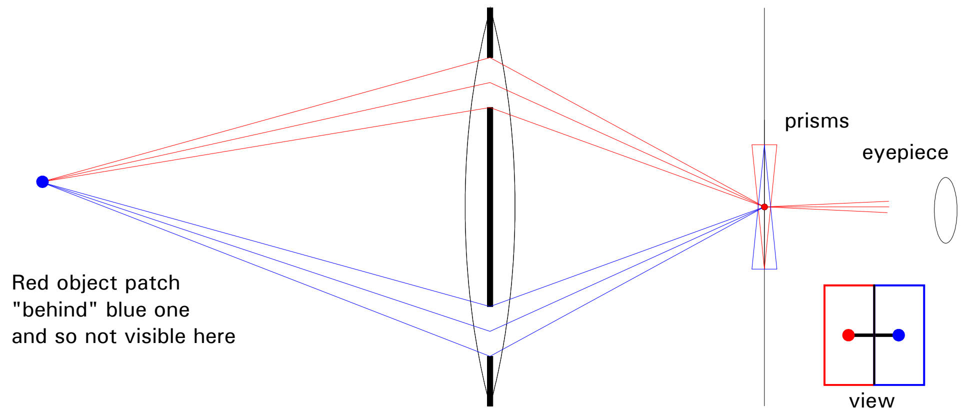

Figure 6

Here, we have mounted two little wedge-shaped glass prisms in the ground glass plane, one closer to us (shown as red) and one farther from us (shown in blue). Note that, because of the relative horizontal positions of the two object patches, the aerial image of the red patch is closer to us than the aerial image of the blue patch. (The image is rotated 180° from the orientation of the object.) Each prism is in fact located so that the only the rays from the correspondingly-colored object patch will pass through it. We'll see how that happens a little later.

In this case, since the lens is properly focused, the rays are actually brought to a focus, creating aerial images right at the prisms.

The prisms redirect the rays so they head toward the eyepiece, allowing those aerial images to be seen by the photographer. (We don't get to see this happening to the "blue" rays in the figure as they are behind the red rays, from our vantage point.)

Note that the aerial images of the red and blue object patches are at the same height. They will thus appear to the photographer to be "in line" horizontally, just as the patches on the object actually are. This in fact is an indication that focus is proper. If we consider the "shaft" of the object (we've been neglecting it for a while), it will appear as an unbroken line across the prism pair.

Next, in figure 7, we shift the lens "forward" again, spoiling the focus. Now, the aerial images of the two object patches are formed in front of the ground glass plane. As before, the aerial image of the red patch is closest to us, and both are at the same height.

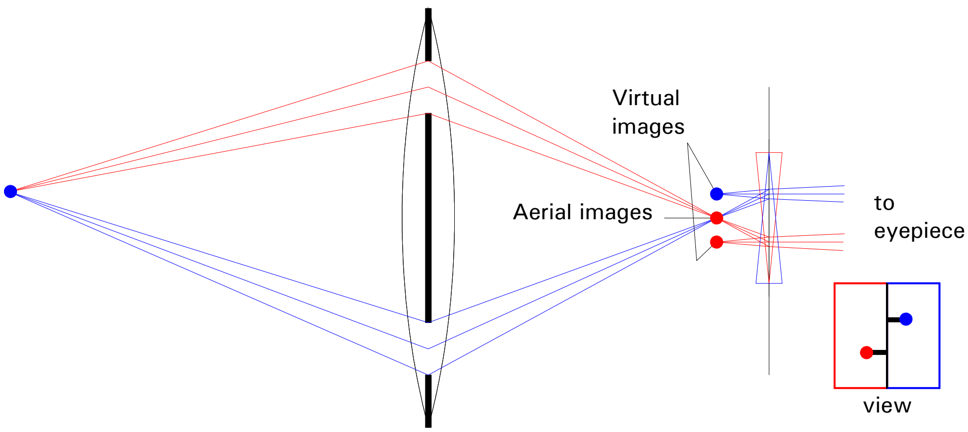

Figure 7

The prisms still allow the photographer to see the two aerial image patches. To be precise, he doesn't really see the aerial images themselves, but rather their "virtual images" as seen through the prism. On the figure, we can see where those are by tracing back the "eyepiece-bound" rays to their apparent origins—the locations of the virtual images.

And guess what? The apparent position of the blue patch image is above the apparent position of the red patch image—the two are no longer horizontally aligned. At the lower right we see the photographer's view of the prism pair, showing the misaligned virtual images of the red and blue object patches.

More importantly, we also see here the "shaft" of the object, on which the misalignment can be most readily seen. In actual use, of course, it is the dislocation of lines crossing the prism "split" through which we observe the misalignment of the two image portions. The eye is very sensitive to even a very small such dislocation (again, as a result of the eye's *vernier acuity*).

It is this dislocation that alerts the photographer to the fact that focus is not proper.

Getting real

Now this is all well and good, but not very realistic. After all, we still have the artifice of the aperture stop with two holes and the fanciful notion that magically only the red rays go through the top hole and only the blue rays go through the bottom holes. In figure 8, we will get rid of the two-holer.

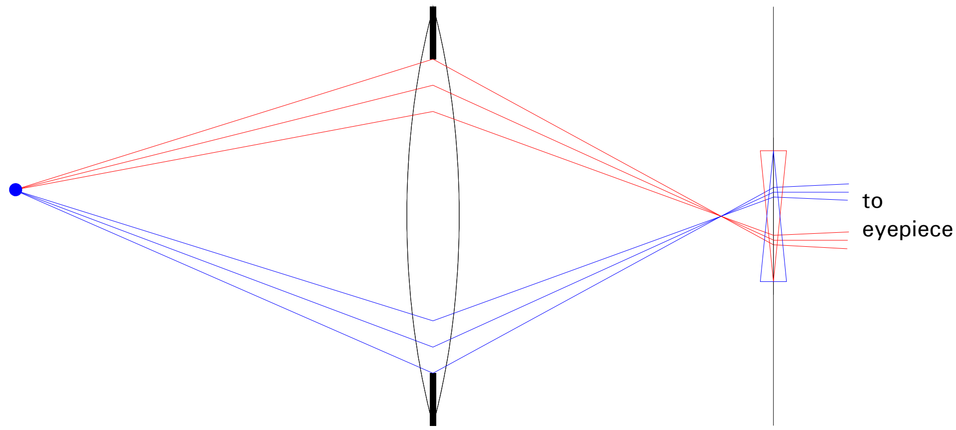

Figure 8

In this more realistic picture of the actual system, the two-hole aperture is gone. But nevertheless the photographer can still only see, through a given prism, light coming from a certain portion of the aperture, just as when we had the holes. The prism, like a “see around corners” device, has in fact directed the photographer’s view to only that area—near the top of the aperture through the “red” prism, and near the bottom through the “blue” prism.

We see this more explicitly in figure 9.

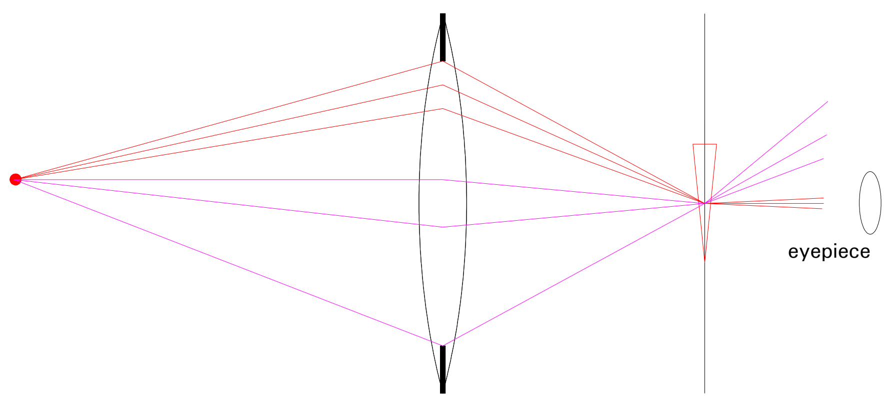

Figure 9

We concentrate here on only “red” rays, and we have left out all the blue stuff to get a clearer view. We have shown some additional “red” rays, ones that we previously ignored (with justification, as we will see). We’ve actually made them magenta here, to allow us to more clearly follow the destinies of the two groups of "red" rays.

Note that these newly-shown rays (which would not properly participate in the scenario of how the focusing aid works, not having come through the "top hole" in the imaginary special aperture) in fact are deflected by the prism so they never reach the eyepiece. So we see that the real arrangement works just like the one with the two holes.

Now what about the requirement that, somehow, only red rays go through the red prism and only blue rays go through the blue prism? In figure 10, we see what makes this happen—nothing supernatural, only geometry.

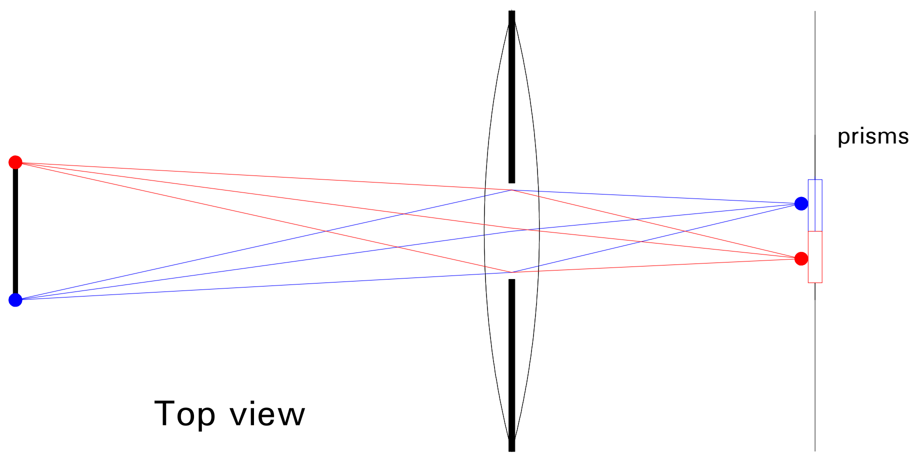

Figure 10

This time we see the system from the top. The rays from the red patch, regardless of the portion of the aperture they go through, form the red patch of the aerial image, which is in front of the "red" prism. The rays from the blue patch, regardless of the portion of the aperture they go through, form the blue patch of the aerial image, which is in front of the "blue" prism. It is just as we had "required" by assumption in our earlier description.

We saw in figure 9 that only rays passing through the top hole can pass through the "red" prism to the eyepiece, and only rays passing through the bottom hole can pass through the "blue" prism to the eyepiece. Therefore, only red rays passing through the upper hole, and blue rays passing through the lower hole, are effective, each through their respective prism. Thus all our requirements in this regard have been fulfilled in the real world without any supernatural devices or forces.

In figure 11, we fill in some more details of the actual physical arrangement.

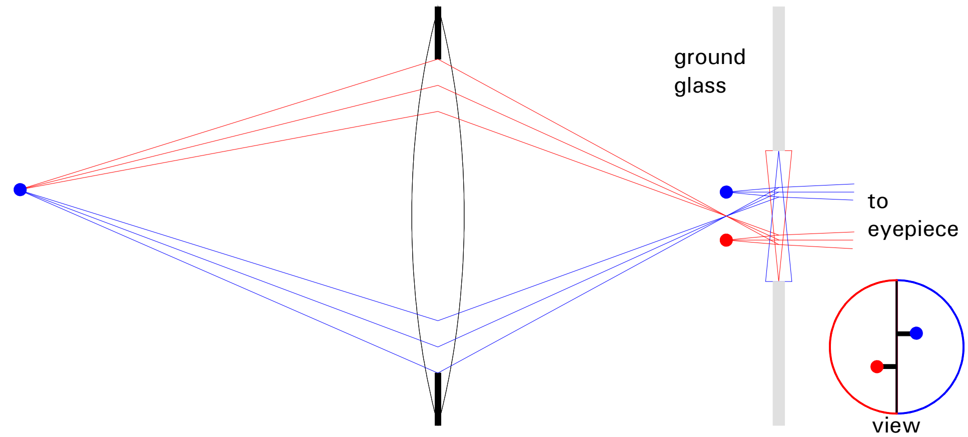

Figure 11

In several earlier figures, we had eliminated the ground glass, totally replacing it with our pair of prisms, which we implied were rectangular in shape. In the actual use of this device in a camera, the two prisms, as seen from the eyepiece location, are semicircular, and they are set, adjacent to one another, into an area in the center of a normal ground glass screen. That way, the entire image is seen during focusing, the part not falling on the prism circle appearing blurred if focus is substantially incorrect.

At the lower right we see the photographer's view of the prism circle, showing the misaligned images of the red and blue object patches, and, more importantly, the dislocation in the object's "shaft", resulting from the misfocus situation of this figure.

Our particular example has the dividing line between the two prisms vertical. This turns out to work best when there are horizontal lines in the central portion of the image. In some cameras, the dividing line is made horizontal (to cater to vertical lines, which are perhaps more common in many scenes), or diagonal (to allow us to exploit either horizontal or vertical lines in the image).

When it doesn't work

The previous figures showed the lens set at essentially its maximum aperture, and in fact that is the normal situation when in the viewing/focusing mode. The proportions also suggest that the maximum aperture of this particular lens is fairly large (fairly small f/number).

Suppose that we have a lens with a substantially smaller maximum aperture. Figure 12 shows the result.

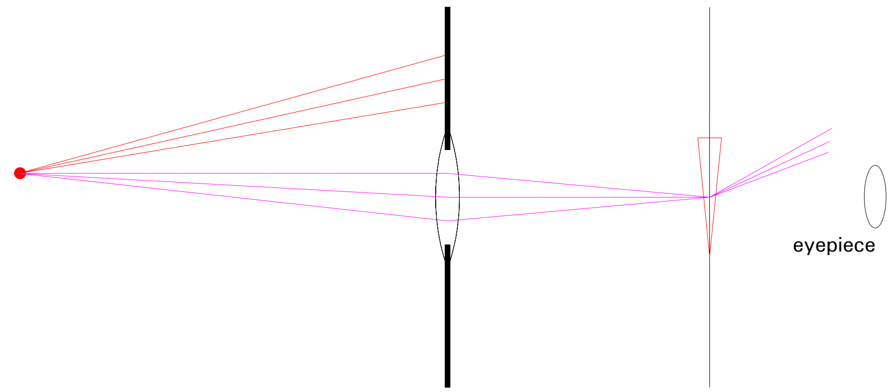

Figure 12

Again, this shows only the "red" aspect of the scenario. The rays actually shown in red are those that, had they been able to pass through the lens, would have arrived at the red prism at such an angle that they would have been deflected toward the eyepiece. They thus would have contributed to the virtual image of the red object patch. But in fact those rays can't pass through the smaller aperture of this lens.

The "red" rays shown in magenta can indeed pass through the smaller aperture, but they arrive at the red prism at such an angle that they are deflected to fall outside the eyepiece. The result is that the view in the prism circle is completely dark, and the split prism focusing aid is useless. (The magenta "red" rays wouldn't contribute much to the prism effect anyway, because of their geometry.)

Thus, the split prism focusing aid fails with lenses having an insufficiently-large maximum aperture (we say maximum aperture since that's the aperture that is ordinarily in effect during the "viewing and focusing" phase of operation).

By changing the angle of the prisms, we can arrange for the red and blue rays to be picked up from portions of the exit pupil that are closer together—within the diameter of a smaller aperture. (This is referred to as "reducing the baseline distance" of the system, a concept that is discussed more thoroughly later when we describe the application of the split prism concept to automatic focus systems.) This allows the "blackout" effect to be averted down to a smaller aperture. However, doing so reduces the "sensitivity" of the split prism effect: the amount of visible misalignment of the image is less for any degree of focus error.

There are special modifications to the prisms (not involving a change in angle) that can mitigate this "blackout" effect down to smaller apertures than usual while still maintaining a desirable large baseline distance (in the interest of retaining greater "sensitivity" when a large aperture lens is in place). These modifications are used in some split prism focusing screens available today. The principle involved is proprietary, and is beyond the scope of this article.

The split image system in summary

Through this long presentation, we have seen that the classical split image arrangement gives a direct, sensitive indication of imperfect focus of the camera by visible misalignment of the two adjacent image portions—in particular, by causing an apparent dislocation in the image of any line crossing from one side of the circle to the other. We have also seen that, if the aperture of the lens (in the viewing and focusing mode) is too small, the arrangement will not be usable.

THE "MICROPRISM FIELD" FOCUSING AID

Another focusing aid sometimes provided on SLR cameras (either in place of the split image prism circle or in addition to it) is often described as the *microprism field* system. Simplistically, imagine that we had a portion of the focusing screen not in ground glass form but rather covered with tiny prism pairs, conceptually like the split image system we discussed here at length. If the focus is correct, the two prisms of each tiny pair produce to the viewer two aligned virtual images of adjacent parts of the scene. The result is that a clear virtual image is seen over the entire prism field.

If the focus is not correct, the two image fragments from each pair of prisms are not aligned. (In fact, they are not images of the same part of the object at all.) The visual impact of this is that the image seems "exploded", an effect may be much more visible than the blurring the same focus error would produce on the ground glass. (It is at least usually more apparent to the less-trained observer.)

A common arrangement is to provide a split image prism circle in the center of the focusing screen, surrounded by a ring of the microprism arrangement (often called in this case a "microprism collar"), the rest of the screen then being of the basic ground glass form.

In reality, the elements of this field are not pairs of prisms similar to those discussed earlier but rather tiny pyramids. The effect, however, is as we just described.

THE PRINCIPLE APPLIED TO AUTOFOCUS SYSTEMS

This same principle is used by an important class of camera autofocus system, the "phase comparison" class. The task here is to automatically determine if the lens is in proper focus for the object of interest, and if not, in which direction the focus is "off" and by how much. This information is used to direct the refocusing of the lens to what should be the position for ideal focus of the object.

Here, instead of having the two "subimages" (the ones we have spoken of metaphorically as "red" and "blue") displayed side by side and examined jointly by the eye, each subimage is instead led to a separate pixel detector array, something like that used in a digital camera for actually capturing the taken image, but typically only one or a few pixels wide and a modest number of pixels high.

The relative positions of the subimages on these two detector arrays is determined by cross-correlation processing of the two array outputs, leading to essentially the same determination that is made visually with the split image system by visually noting the relative alignment of the two adjacent subimages.

The two subimage detectors need not be placed side by side (since the comparison is made mathematically), and in fact we generally place them one above the other (with a gap between them).[4]

An advantage is that the two subimages need not be of adjacent regions of the scene (as with the visual system). They can be of the identical regions of the scene, thus facilitating comparison of their relative positions on their respective detector arrays.

A typical arrangement of such a focus detector is shown in simplified form on figure 13. To enhance the clarity of presentation, the figure is not to scale. It is believed that this essentially follows the system used in the Canon EOS 20D digital SLR camera.

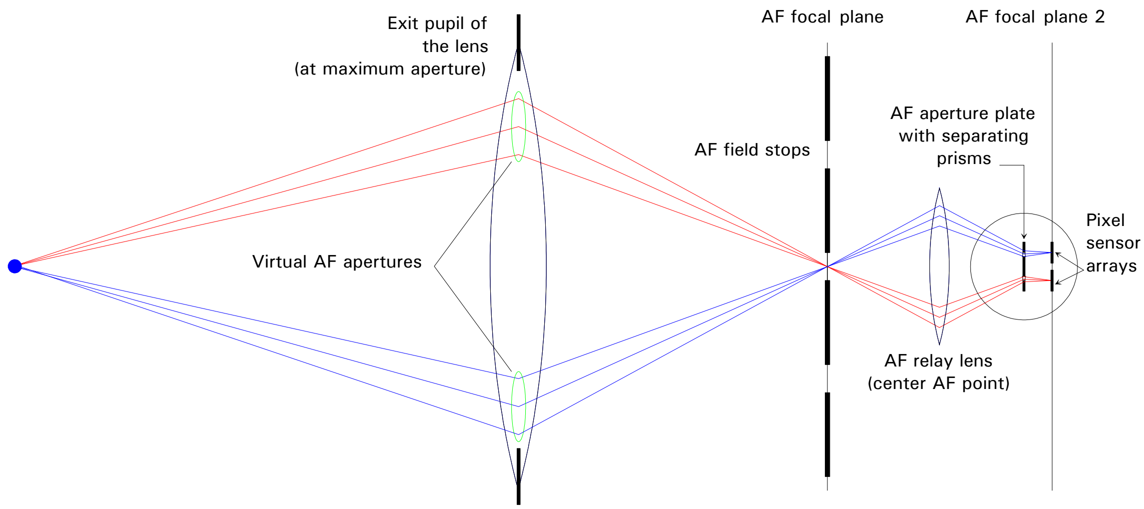

Figure 13.

In reality, in that camera, the system is "bent" in two places by mirrors, forming a sort of periscope arrangement by which the light from the scene eventually reaches the apparatus seen on the right-hand side of the figure. This however does not change the principles of operation we will describe here.

---

[4] This description assumes a focus detector system intended to be responsive to vertical contrast profiles in the scene (essentially, responsive to horizontal edges), the orientation that will be shown in the figure to follow. Of course, systems of the other orientation are also often used.

Again, we see the “thin lens” model. In that model, the exit pupil of the lens is identical to the physical aperture stop. We label it “exit pupil” here because that is what is pertinent in the more general case.

In the camera I mentioned, there are 7 sets of autofocus detection systems, each associated with a different location in the frame, and limiting its attention to a relatively-small region there. In the figure, we show only one of these, in particular the one associated with the center of the frame. We’ll catch just a fleeting glimpse of the provisions for other “autofocus points”.

As mentioned before, the orientation of the system in the figure is one that would be “sensitive to horizontal lines” (just as was the case for the split image prism system we discussed earlier).

The AF (autofocus) focal plane is the plane where, with the lens perfectly focused, the image for the autofocus system would be created with the camera in its “viewing and focusing” configuration. In the camera mentioned, the path to this plane would be by way of the first of two mirrors. This plane corresponds, in effective distance from the lens, to the film or digital sensor focal plane when the camera is in its “shooting” configuration. The image is an aerial image: it is not cast on any surface.

The AF field stop is an aperture which assures that only rays from objects at the appropriate part of the scene can enter the system of this particular detector. We see the field stop for the sensor we will be examining (assigned to the center of the field of view) as well as the AF field stops for two other sensors, assigned to locations in the field of view above and below the center.

The AF relay lens serves to create a replica of the image formed at AF focal plane at an “alter ego” focal plane (AF focal plane 2). This allows for a “clear space” (clear of the AF field stops) for the optical components we see next. In the camera mentioned, this path (after the relay lens) is through the second of the two mirrors.

In this figure, the red and blue ray colors identify portions of the entire “cone” of rays from a certain scene point which have passed through small areas at the upper and lower extremes of the of the exit pupil, known as the “virtual AF apertures”. (Recall the description of the fanciful “two-hole” aperture used earlier in this article.) We will see shortly what “generates” these virtual apertures.

In figure 14, we “zoom in” on the portion of the figure enclosed by the circle, so we can see the smaller detail in that area.

In front of AF focal plane 2 there is an “aperture plate” in which (for each AF sensor) is embedded a pair of prisms, which I call “separating prisms”.

Behind each of these prisms is a pixel sensor array, one or two pixels wide and a number of pixels high.

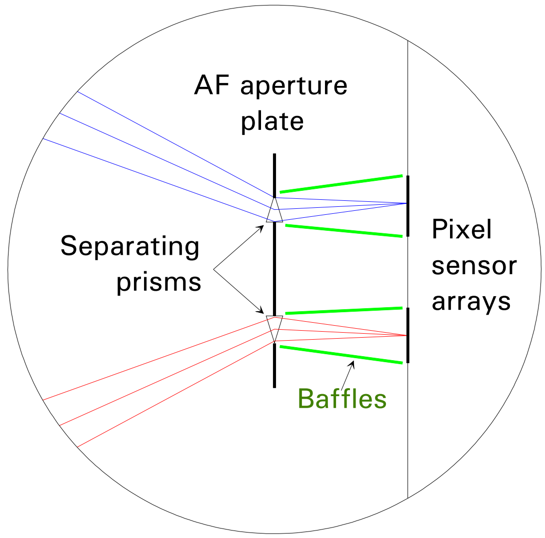

Figure 14.

The figure shows the system with the lens in perfect focus. As a result, were it not for the aperture plate and prisms, all the rays from the object point shown would converge at a point on AF focal plane 2. With the aperture plate in place, but with openings rather than prisms, the red and blue rays (but not any others) would converge at that same point.

But we have put little prisms into those two “apertures”. As a result, the blue rays are deflected upward, so thy will all converge at a point on the upper pixel sensor array; the red rays are deflected downward, so thy will all converge at a point on the lower pixel sensor array.

The “baffles” shown in green prevent light from other prisms (including those for other detectors) from falling on the sensors. It is not known whether these actually exist in the Canon implementation.

The points of convergence of the two sets of rays, given the stipulated condition of perfect focus, fall at locations on the two sensors that are the same with respect to a standard reference point on each sensor. In any case, the relative positions of the images on the two sensors are determined by performing cross-correlation on the luminance patterns reported by the two pixel arrays. It is this process that is responsible for the name of this system: *phase comparison focus detection*.

If the lens is not in perfect focus, then the rays in either set are not brought to exact focus at a point on the pixel sensors but rather in front of them or “behind them” (or rather, would if there were no prisms). Figure 15 shows this for the case in which convergence would be in front of AF focal plane 2; the magenta “construction line” prolongations of an illustrative blue and red ray show where this convergence would occur.

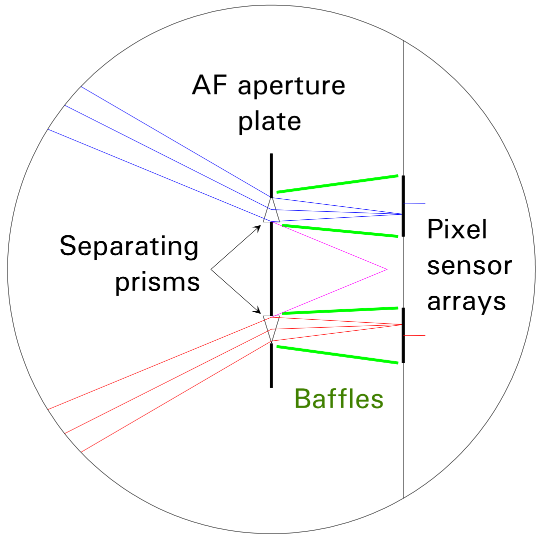

Figure 15.

Because of the small angle of the cone, the “circle of confusion” (blur figure) created on the sensor itself is of very small diameter. Thus, images will still be fairly “sharp” on the sensors, and their relative positions can still be easily recognized by the cross-correlation process. However, that position will be shifted upward on one sensor and downward on the other. (I have placed little “tick marks” on the figure showing the original locations of the point images on the two sensors, for ease ofcomparison.) We see this shift again on figure 16.

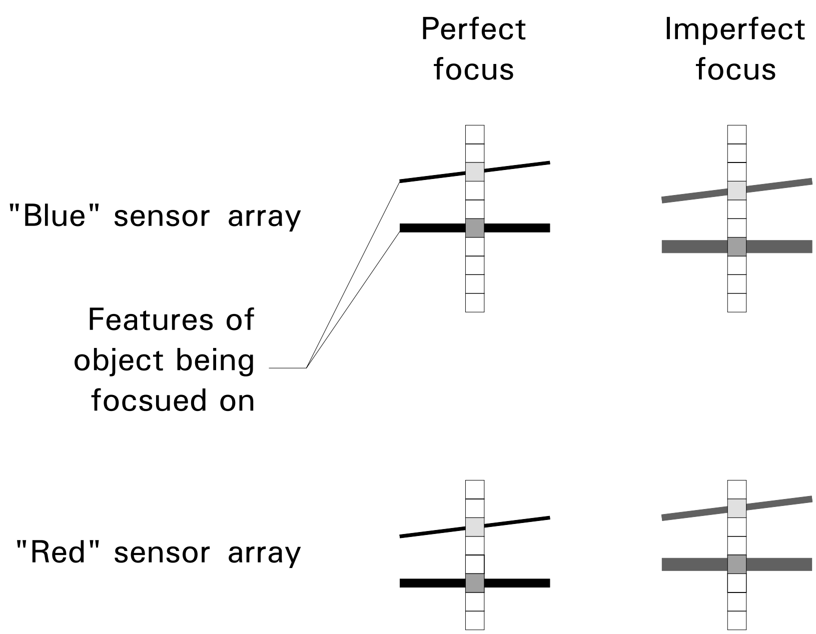

Figure 16.

The magnitude and sign of the discrepancy will indicate the amount, and the direction, of the misfocus. Based on this, the overall autofocus system can adjust the position of the lens as required to bring focus to the proper state.

Returning to figures 13 and 14, note that the size of the AF apertures (even though they are filled with prisms) determine how large a cone of light is admitted by the "red" and "blue" aspects of the system. The spacing between the apertures (between the prisms) determines how far apart are the little regions of the exit pupil from which "red" and "blue" rays are taken. Collectively, these two considerations define the virtual AF apertures within the exit pupil.

Autofocus precision

There will always be some uncertainty in the determination of the relative positions of the two images, leading ultimately to a residual focus error after the AF system has made the correction indicated by the detector system. Because of geometric considerations, the greater the separation of the two virtual AF apertures in the exit pupil, the greater will be the displacement of the two images for a given focusing error; conversely, the greater that separation, the smaller will be the residual error for any given uncertainty in determination of the relative positions of the two images.

The distance of this separation between the two virtual AF apertures is often spoken of as the "baseline" of the detector system, by analogy with the somewhat-corresponding parameter of a rangefinder.

Aperture dependency

If the two "virtual AF apertures" are spaced further apart than the full diameter of the exit pupil (as a result of our mounting a lens having a rather small maximum aperture), there are no blue or red rays that can pass through to the detectors, and focus detection will fail.

To avert this, the focus detectors are usually designed to have virtual AF apertures near the edges of the exit pupil for a modest camera aperture—an aperture likely to be available on most lenses to be used on the camera. Of course, this compromise (a "reduced baseline") results in a reduction in the precision of the focus detection system.

In some cameras (such as the Canon EOS 20D), at one or more particularly-important AF point locations across the frame (typically the one at the center), two autofocus detectors are provided, one workable for modest aperture lenses and one (with a "larger baseline") workable only for larger aperture lenses. The appropriate one is put into action based on the maximum aperture reported by the lens in place. This arrangement exploits the greater precision in focus determination available with a larger aperture lens in place.

#