IMAGE-SIDE PERSPECTIVE AND STEREOSCOPY

John Bercovitz
Lawrence Berkeley National Laboratory
Berkeley, CA

Abstract

Correct perspective is crucial to orthostereoscopy. That is to say, the observer must view from the same points in space, relative to the image, that the stereo camera's lenses had relative to the scene. Errors in placement of the observation points result in distortion of the reconstructed stereo image. Although people adapt easily to visual distortions, they may not do it well enough or quickly enough for critical telepresence and telerobotic applications.

Further, it is difficult for humans to reliably determine by sight the correct observation point relative to an image. A mathematical guide to correct perspective is therefore useful. The mathematical key to perspective is that all images must subtend at the eye the same angles which the objects that generated them subtended at the camera. The center of perspective on the object side of a camera is the entrance pupil, but where is the center of perspective on the image side of an asymmetrical lens? A simple formula simply derived answers that question.

By way of background, pertinent optics and stereoscopic reconstruction errors, including perspective error, are reviewed in this paper. New work begins in the fourth section.

Keywords: ortho*, stereo*, perspective, telepresence, telerobotics

1. Introduction

Geometrically correct stereoscopic reconstruction (orthostereoscopy) has many advantages in telerobotics and telepresence (although in some few particular applications, knowledgeably altered stereo reproduction can actually be more useful than undistorted stereo reconstruction).

The human visual system adapts quickly and well to all manner of distortions in the geometry presented it, but perhaps it never does so perfectly and certainly it never does so instantly. Therefore, in telepresence and in telerobotics, for ease and accuracy of operation and for safety and speed, it is best to present the eyes with an undistorted view of reality, a geometrically correct presentation.

This paper will review the classes of stereo reconstruction errors and their effects, then present the new work, which is a method of finding the correct perspective point from which to view the image, and will conclude with miscellaneous practical matters.

To establish correct perspective when viewing an image, an understanding of the position of the "perspective point" in the camera and in the viewing device is required. In many cases, the focal length of the camera lens is a suitable approximation of the distance from the image plane to the perspective point. However, if a lens is highly asymmetrical (e.g. telephoto or retrofocus) and of long focal length compared to the distance from it to the object, this approximation is no longer valid and will result in significant distortion of the reconstruction presented to the eyes.

One must realize, when investigating stereo reconstruction, that the reconstruction is merely what is presented to the eyes. When the brain is presented with a geometry, it will interpret that geometry based on prior experience (internal models) regardless of whether the geometry presented is distorted or not. As an extreme and well known example, if the eyes are presented with a pseudoscopic (left and right views transposed) stereo pair of the human face, the brain will not see a concave face as it would if it were to perform a rigorous stereo reconstruction. Rather, it will see a normal convex face. Therefore presentation of a distorted geometry does not make seeing a distorted geometry at all certain, particularly for familiar objects. However, if motion is added, the ability to "mend" these distorted presentations is greatly impaired.

Note: In the figures below which show stereoscopic reconstruction, rays are drawn to run from top to bottom rather than the conventional left to right.

2. Stereoscopic reconstruction errors

The basic stereo reconstruction errors are caused by:

• mismatch of stereobases of camera and personnel
• mismatch of perspective of camera and personnel
• toe-in of the cameras causing keystone error

Toe-in results in warped reconstructed space[1]. It is sometimes used to set the stereo window distance because it is often easier to implement this change than it is to shift the lenses relative to the image gates. Other potential problems are vergence error (which is a mismatch between the vergence required to view a reconstructed image point and the vergence required to view the original object point) and offset (which is a result of viewing from a point not directly over the correct point in the image). This last, which isn't generally found in practice and which will not be covered in any depth here, results in a type of keystone error (Fig. 1b).

Vergence

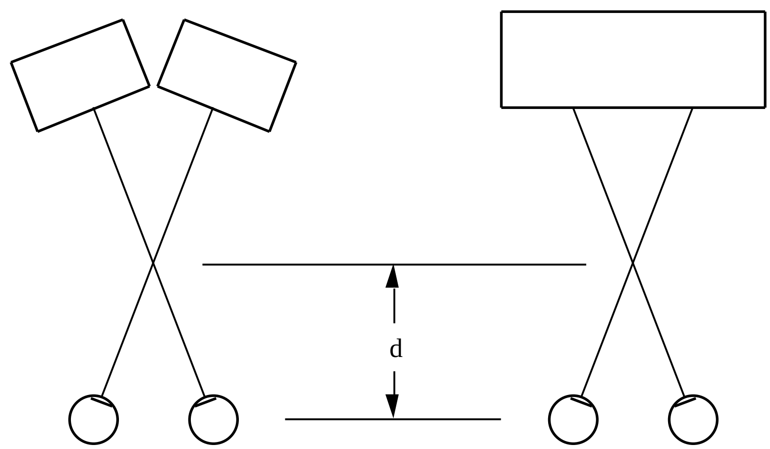


Figure 1a
Vergence error without key-stone error (two monitors)

Figure 1b
Vergence error with key-stone error (one monitor)

Vergence, a weak depth clue, indicates that objects are located at distance d.

Vergence error is usually the least important of the depth clues enumerated here because vergence is not used to judge depth except in the extreme vergence cases. Instead, distance is judged by comparing two images and determining the amount of parallactic disparity in different parts of the image. From this, a depth map is built up. Most commonly the effects of vergence error are seen in "cross-eyed" CRT displays (Fig. 1). If the creator of the images wishes to make the images on the CRT large (larger than human interpupillary distance) in order to obtain more resolution, he will often put the left image on the right and the right image on the left and ask the observer to cross his eyes. This is easier for the observer than diverging his eyes to see large side by side images. The result of cross-eyed viewing is that, although the observer may be looking at something which is supposed to be at infinity, his vergence is correct only for something a short distance from his face. The conflict is resolved when the brain decides that since the image is close by, it must therefore be small (the angle subtended by the image is fixed and known to some degree). Additionally, if the two images are displayed on a single monitor (Fig 1b), keystone error results from the eye not being over the correct point in the image. Surprisingly, with all its technical faults,

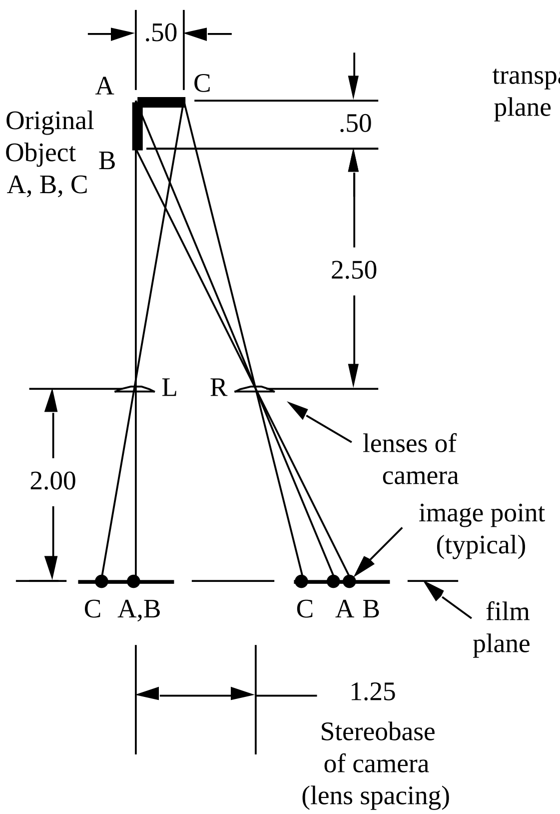
[1] When a rectangle is viewed from off center, it becomes a trapezoid or keystone, hence the name "keystone error". This topic is ably covered in: Andrew Woods, Tom Docherty, Rolf Koch (1993) "Image Distortions in Stereoscopic Video Systems" Stereoscopic Displays and Applications IV, Proc. SPIE vol. 1915, pp. 36-48.

the cross-eyed display method works well enough for casual demonstrations. Conversely, stereo viewers which position near objects' image points at the same spacing as the spacing of the observer's eyes will encourage the observer to interpret these near objects as being far away and large.

Stereobase mismatch

The effect of a mismatch between taking and viewing stereobases.

Figure 2a
Object and camera geometry



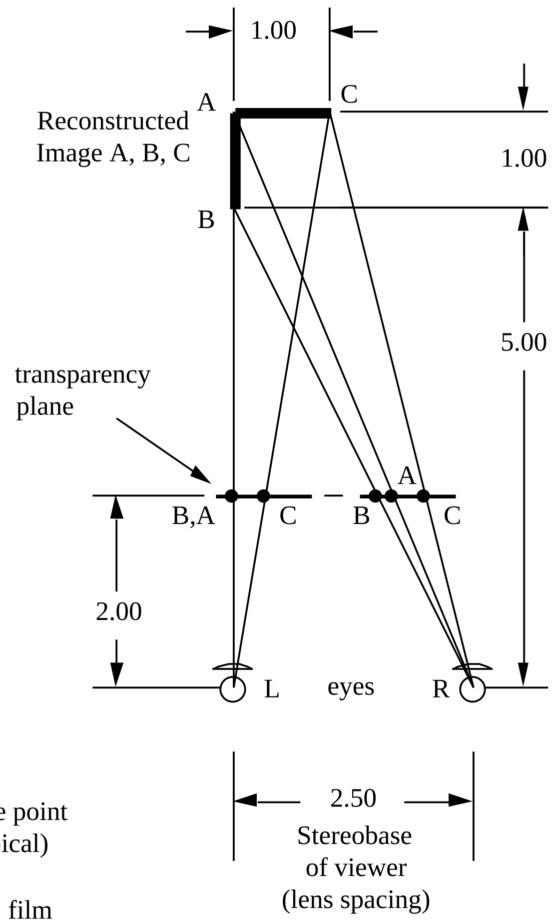


Figure 2b
Reconstructed Image
and Viewer Geometry

Stereobase mismatch occurs when the spacing between the stereo camera's lenses is different from the spacing between the observer's eyes. The image reconstruction changes generated by this mismatch can be determined by examining the optical geometries of the camera and the viewer in use. The result is a change of size, but not shape, of the reconstructed image.

In Fig. 2a an example object is shown being photographed by a stereo camera which has a lens spacing of 1.25 units. The film records the image points A, B and C on the film (as shown by the rays from the object to the film). When the film, now a transparency, is viewed by an observer, the reconstructed image can be found by projecting rays (in reverse) from the eye through the image points on the transparency and noting homologous rays' intersections.

Fig. 2b shows an observer with an eye spacing double that of the camera. The reconstructed image is seen to have all of the original object's dimensions doubled, including the distance from the observer to all points in the reconstructed image. Other ratios of stereobase mismatch yield a similar proportional result.

Note in this example that the distances from the camera's lenses to the film, and from the viewer's eyes to the transparency are the same at 2 units.

The stereobase mismatch shown in Fig. 2 can be useful if telerobotic work is being done on something very small. In the case of telepresence, increasing the stereobase of the camera above normal will allow depth to be seen at a great distance. As an example, depth can easily be seen all the way across the Grand Canyon if the stereobase is set to several hundred meters.

A useful mnemonic for stereobase changes: A giant's eyes are spaced widely. He can see depth to a great distance and scenes appear small to him. A stereo camera with widely spaced lenses gives the giant's view.

Perspective mismatch

Perspective mismatch is the result of viewing from a distance other than the distance from image to center of perspective and causes a distortion of the reconstructed image. The distortion is a change of depth, but no other dimension, of the reconstruction.

In the example shown in Fig. 3 the two stereo transparencies shown are exactly the same, but the distance from the eyes to the transparency differs by a factor of two. The apparent positions of object points A, B, and C are moved proportionally away from the viewer as the viewer's eyes are moved back from the slide transparency. The apparent absolute size of the face of an object stays the same, though the angle it subtends at the eye decreases. As can be seen, changing the perspective changes only the depth dimension.

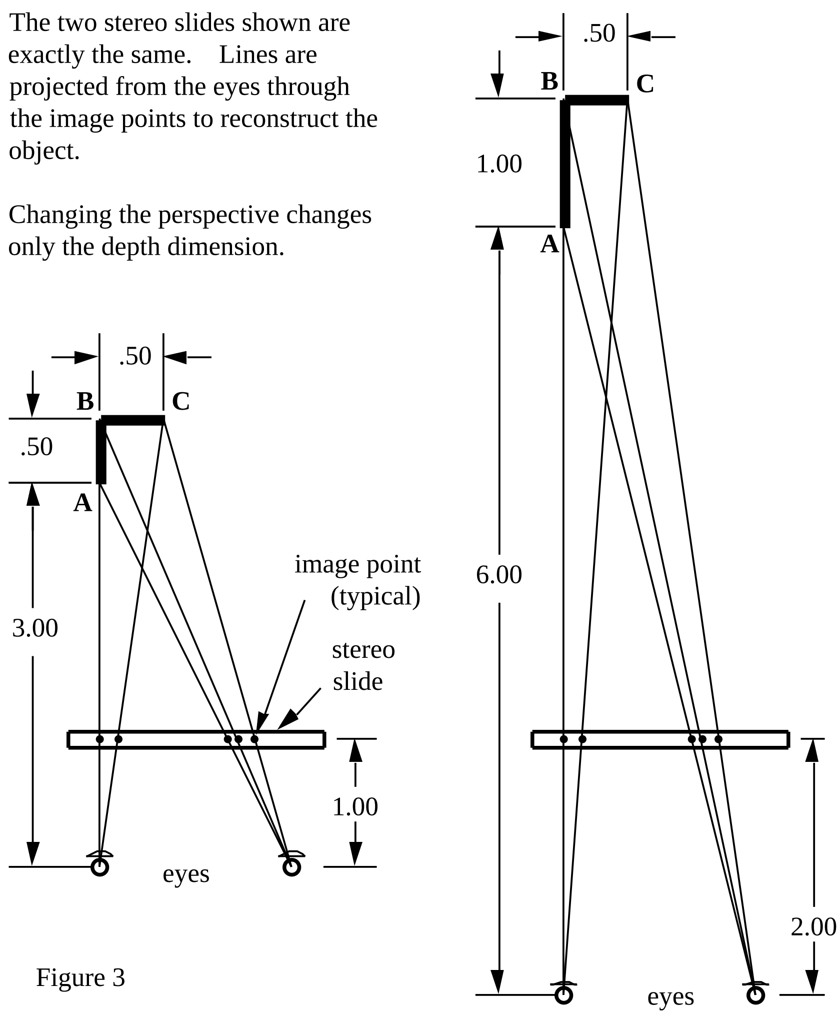


Figure 3

Scaling

It is not uncommon for an image to be enlarged from its size as it was made by the camera lens. To maintain perspective, the image must still subtend the same angle at the eye. Therefore the distance to the perspective point must be scaled with the enlargement. A 20X image enlargement results in the correct perspective point being 20 times as far away as calculated below. The important thing is that the height of the image divided by the distance to it be held constant at the proper, original, object-space figure.

3. Image-side perspective point

The perspective point is not just a distance from the image since it is an actual point in space and so has three co-ordinates which define its position. A perpendicular dropped from the camera lens to the image plane intersects the image at a point which could be called the “sub-lens” point. The perspective point lies a certain distance above the sub-lens point along a line perpendicular to the image. It is this distance which will be determined below.

Optics review

The Gaussian form of the lens equation is more familiar, but the Newtonian form of the equation will also be useful in this work and so by way of review, the former is derived from the latter below.

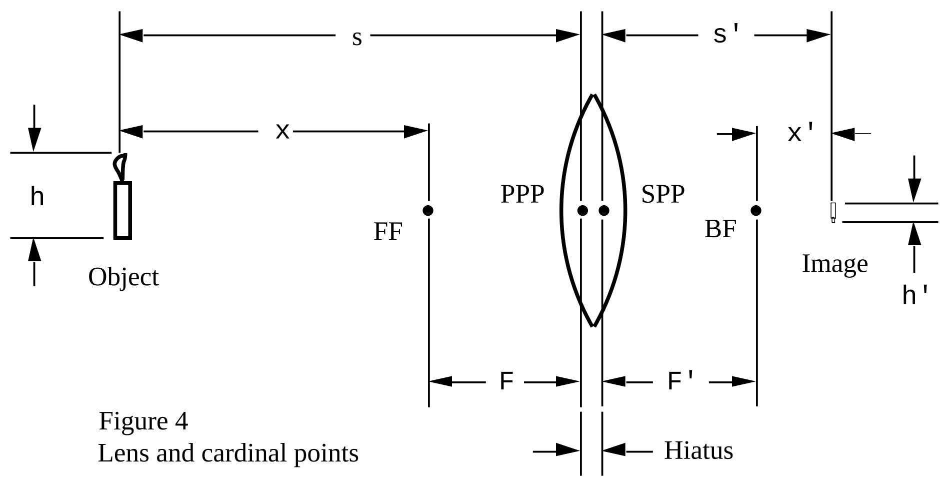


Figure 4
Lens and cardinal points

In Fig. 4 PPP and SPP are the primary and secondary principal points, which are the points at which the lens is effectively located when calculating either conjugate planes or the magnification of the in-focus object.

FF and BF are the front and back foci for rays parallel to the axis of the lens.

F and F ' are the front and back focal lengths of the lens, and they are numerically equal if the same refractive medium (air for instance) is found on both the image and object sides of the lens. This condition will be assumed in deriving the Gaussian lens equation from the Newtonian lens equation below.

Newtonian form: $x*x' = F^2$ (1)

from diagram: $x = s - F$ (2)

from diagram: $x' = s' - F$ (3)

substituting (2) and (3) into (1): $(s-F)(s'-F) = F^2$

$$ss' - Fs' - Fs + F^2 = F^2$$

$$ss' - F(s' + s) = 0$$

$$ss' = F(s' + s)$$

$$1/F = (s' + s) / ss'$$

Gaussian form: $1/F = 1/s + 1/s'$ (4)

Magnification is computed by dividing the height of the focused image by the height of the object it represents. Magnification is usually less than one since the image is usually smaller than the object. Magnification can be expressed in Gaussian variables or in Newtonian variables. Below, the Newtonian form of the magnification equation is derived from the Gaussian form.

Gaussian: $$m = \frac{h'}{h} = \frac{s'}{s} \quad (5)$$

$$m = \frac{x' + F}{x + F} \quad (5a)$$

from above: $$x*x' = F^2 \quad (1)$$

from which: $$x' = \frac{F^2}{x} \quad (1a)$$

substituting (1a) into (5a):

$$m = \left(\frac{F^2/x + F}{x + F}\right)$$

$$m = \left(\frac{F^2 + Fx}{x^2 + Fx}\right)$$

$$m = \frac{F\,(F + x)}{x\,(F + x)}$$

From which the Newtonian form: $$m = \frac{F}{x} \quad (6)$$

from above: $$x*x' = F^2 \quad (1)$$

from which: $$\frac{F}{x} = \frac{x'}{F} \quad (1b)$$

substituting (1b) into (6):

Alternate Newtonian form: $$m = \frac{x'}{F} \quad (6a)$$

Center of perspective

Generally, the entrance pupil is the image of the iris of the lens as seen through the elements of the lens which lie on the object side of the iris, although lenses have been constructed in which the iris lies in front of the lens on the object side.

The entrance pupil acts as a mask or filter and selects which rays will create the image while the lens determines where those rays will land on the image plane, within the constraints of rectilinearity if the lens is distortionless. All rays from objects meet at the entrance pupil thereby establishing it as the center of perspective. In the same manner, the pinhole becomes the center of perspective of a pinhole camera.

In Fig. 5, corresponding images and objects subtend the same angle at the pinhole of a pinhole camera. Principal points (see Fig. 4) are useful for finding which object and image planes are conjugates, and for finding the image magnification of the in-focus object, but the primary principal point does not necessarily lie at the entrance pupil and so the primary principal point is not necessarily at the center of perspective.

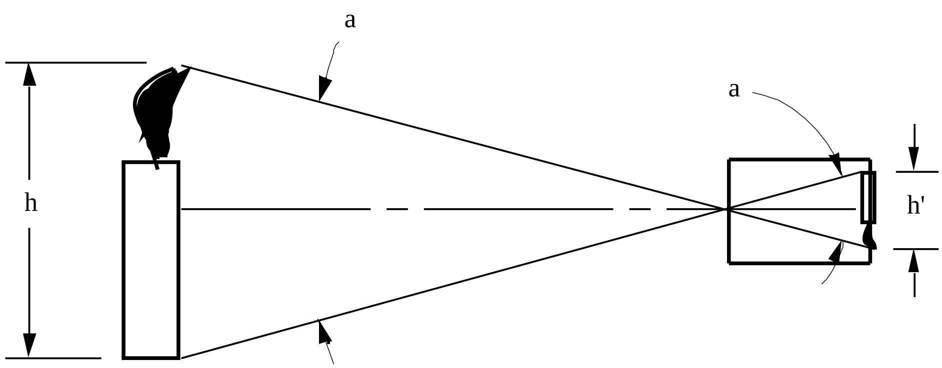


Figure 5   An object and its image in a pinhole camera.

A lens in which the primary principal point and the entrance pupil (center of perspective) are widely separated is depicted in Fig. 6. This highly asymmetrical 4x5 format lens is a 15" f/5.6 Tele-Optar made by Wollensak for Graflex. As can be seen from the drawing, rays from most off-axis object points, after passing through the primary principal point, completely miss the entrance pupil which is 9.9" away. Pupils are shown at nearly maximum diameter, although this is not the normal condition when exposures are being made.

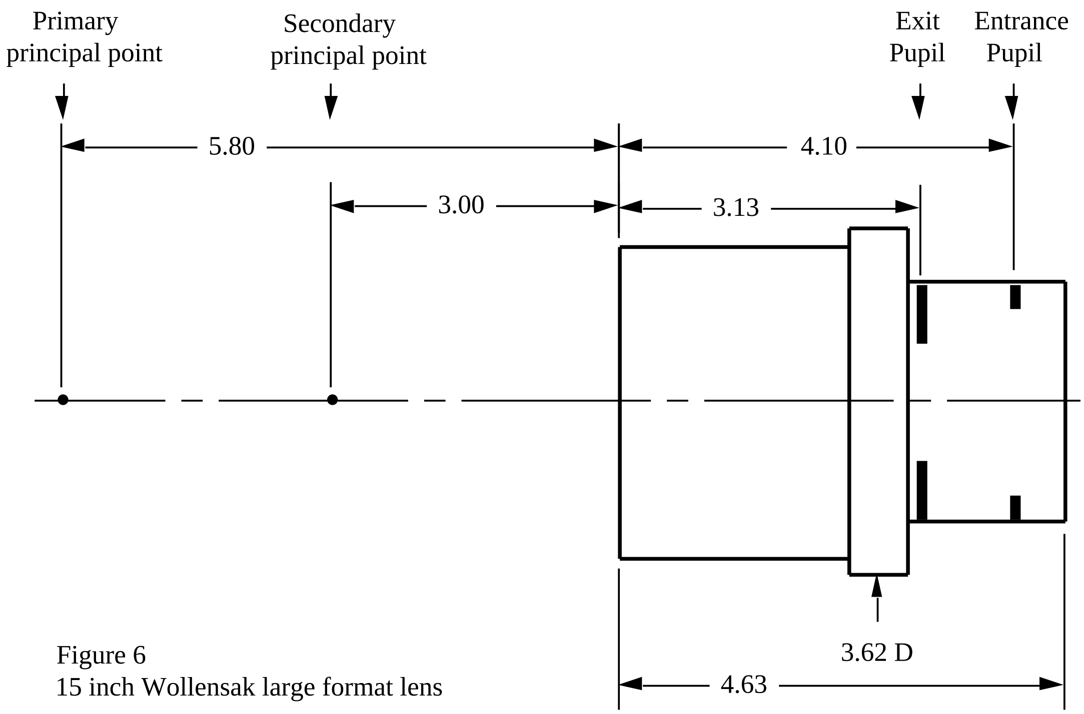


Figure 6
15 inch Wollensak large format lens

4. New work, finding the location of the image-side center of perspective

From above:
•The entrance pupil is the center of perspective.
•The Gaussian primary and secondary principal points and the Newtonian front and back foci are benchmarks for the relationship of the in-focus object and its image.
Fig. 7 shows an asymmetrical lens with some of its cardinal points and also shows the Newtonian and Gaussian variables.

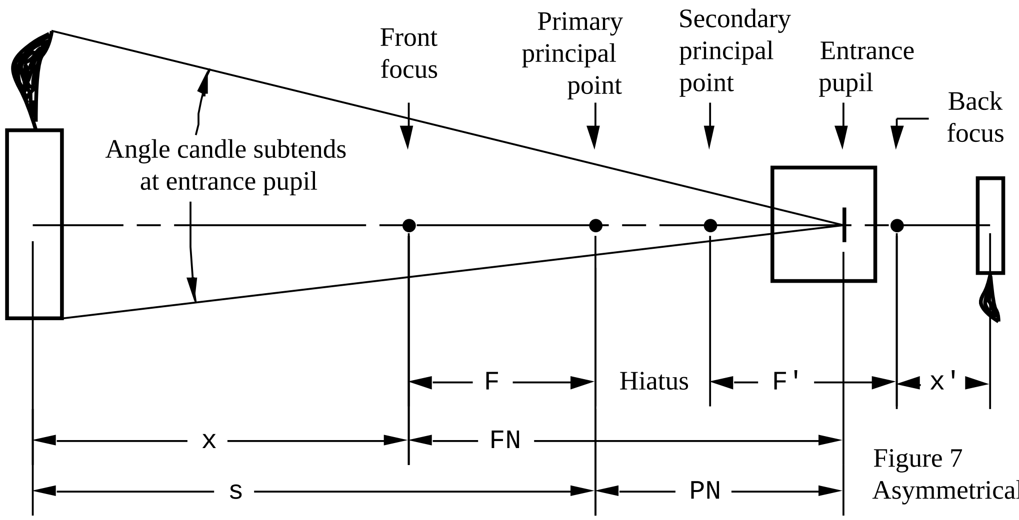


Figure 7
Asymmetrical Lens

To be viewed correctly, the candle's image must subtend the same angle at the eye that the candle itself subtended at the center of perspective of the camera which is its entrance pupil.

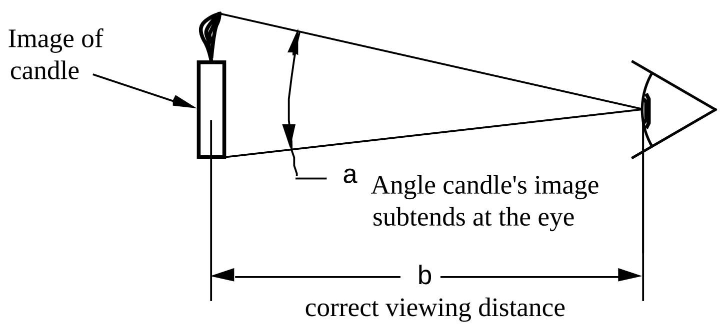


Figure 8

The formulae given above deal only with the in-focus image and its object so the formula for the center of perspective of all the images will be derived from the known relationship of this pair of conjugates. It can be seen from the preceding figures that the angle which the in-focus image subtends at the eye will be equal to the angle the in-focus object subtended at the center of perspective (entrance pupil) of the camera lens if the viewing distance, b, is the distance from the original object to the entrance pupil of the camera lens multiplied by the image magnification of the camera. Below are two derivations based on this statement, one using Newtonian variables and one using Gaussian variables.

$$b = m\,(x + FN) \qquad (7a)$$

$$m = \frac{F}{x} \qquad (6)$$

$$b = \left(\frac{F}{x}\right)(x + FN)$$

$$b = F\left(\frac{x + FN}{x}\right) \qquad (8a)$$

$$b = m(s + PN) \qquad (7b)$$

$$m = \frac{F}{s - F}$$

$$b = \left(\frac{F}{s-F}\right)(s + PN)$$

$$b = F\left(\frac{s + PN}{s - F}\right) \qquad (8b)$$

Examination of equation 8a shows that F is a good approximation of the distance from the image to the center of perspective only when x is large compared to FN. It is interesting to note that the image-side center of perspective does not have a fixed position relative to the lens as does a cardinal point or a pupil.

The basis of this formula was experimentally verified by: 1) using the Wollensak lens shown in Fig. 6 to take closeup pictures of objects lying at various distances and then measuring the images, 2) taking multiple pictures with a slide bar allowing the back of the camera to slide in its own plane and then measuring the images, and 3) ray tracing an asymmetrical lens. A closer look at experimental verification test 1) will illustrate the principle behind all of the tests and perhaps further explain the general problem. If the image-side and object-side centers of perspective are temporarily drawn coincident in a diagram (Fig. 9), the diagram will look exactly like that of a pinhole camera. (In the diagram, object and image points will retain their correct distances from their centers of perspective.) Now by setting objects of known size at known distances apart in depth, the position of the center of perspective on the object side may be determined from the measured sizes of their images.

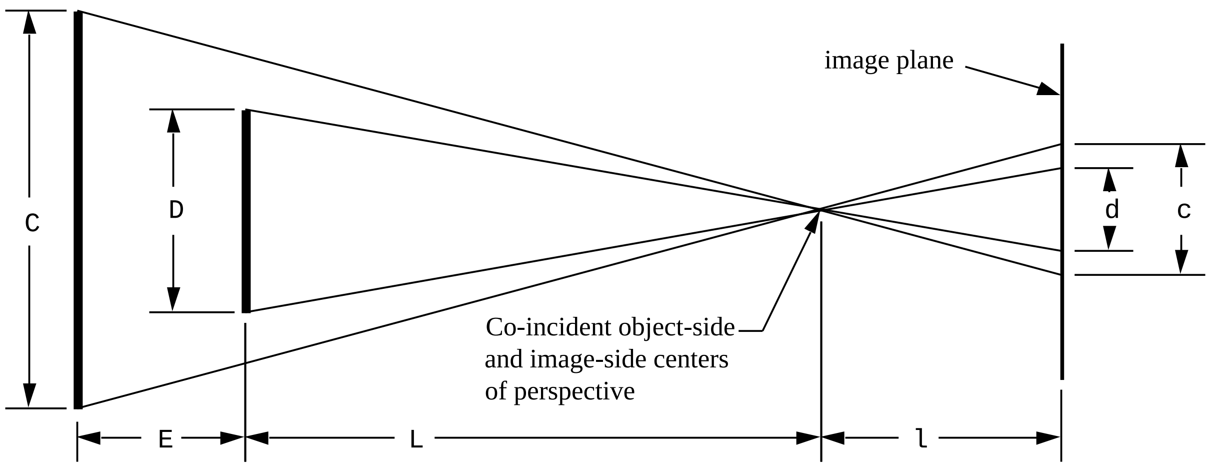


Figure 9 Test setup using Wollensak lens and objects of known size (C & D) separated by a known distance (E).

In Fig. 9, C and D are the sizes of the known-size objects and E is their known separation. The measurements c and d are made at the image plane. The distance L gives the position of center of perspective relative to a known position, that of D. L is found from equa. 15 which is derived as follows.

$$\frac{D}{d} = \frac{L}{l} \qquad \text{or: } l = \left(\frac{d}{D}\right) L \qquad (9,\ 9a)$$

$$\frac{C}{c} = \frac{E + L}{l} = \frac{E}{l} + \frac{L}{l} \qquad (10)$$

Substituting 9a and 9 into 10:

$$\frac{C}{c} = \frac{E}{\left(\frac{d}{D}\right) L} + \frac{D}{d} \qquad (11)$$

$$L = \frac{cDE}{Cd - cD} \qquad (12)$$

Note that the entrance pupil and the primary principal point are 9.9 inches apart in the Wollensak lens. All tests with this lens put the center of perspective within 0.1 inches of the entrance pupil. Similar results were obtained from tests 2) and 3). Having verified experimentally that the entrance pupil is the center of perspective, equation 8 ensues.

5. Miscellanea

Finding the position of the entrance pupil

A simple but satisfactory way of finding the depth or location at which an entrance pupil lies in a lens is to use an SLR as a rangefinder. First set up the lens in front of the SLR and then focus on the pupil. Next, select a fiducial point on the lens. Measure the distance from the SLR to the fiducial point. Remove the lens and without changing the focus setting of the

SLR, move a target into focus. Measure the distance from the SLR to the target. Subtract the one distance from the other to obtain the distance of the entrance pupil from the fiducial point on the lens.

Finding the position of the front focus

The front focus may be found by locating the image formed by rays from an infinitely-distant object coming through the lens in the reverse direction from the usual. The principal point is located one focal length toward the lens from the focus.

Finding the focal length

A collimator with targets at known angles is very useful, but one may also make an image of stars and measure the distances from the center of the image to the stars in the resultant image. Very precise positions of stars are well known and widely published. Do be sure to consider the geometry of the situation carefully before embarking on calculations.

Viewers

There are two basic types of viewers, the Holmes (lensed) and the Wheatstone (mirrored). The infinity vergence of a Holmes viewer can be set once and be right for everyone, but the Wheatstone viewer would have to be adjusted for each person, to achieve perfection. However, Wheatstone viewers are usually used in situations where the image is far from the eye and this minimizes vergence error. The lensed viewer has a different problem and that is that the eyes of a person whose interpupillary distance (IPD) is far different from the lens spacing will look through outer zones of the lenses which will usually give poor definition. Sometimes lens spacing is made adjustable, but unless the image spacing tracks the lens spacing, vergence will be changed when lens spacing is changed.

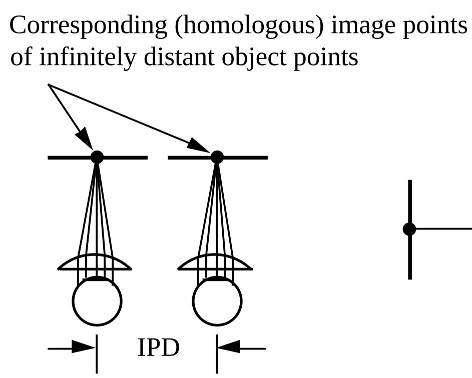


Figure 10a
The Holmes (lensed) viewer

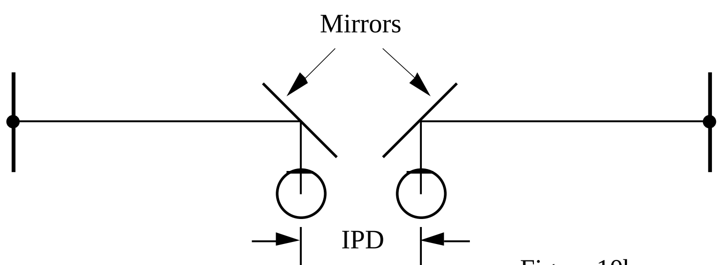


Figure 10b
The Wheatstone
(mirrored) viewer

Cautionary note regarding “PePax”

A few decades ago a writer in the field of stereoscopy said that perspective mismatch and stereobase mismatch could be used to counteract each other, that by doubling one and halving the other, distortion of reconstructed space would be corrected. He called this “PePax” for “perspective/parallax”, and his statement has been uncritically quoted by many noted writers for many years. An examination of Fig's. 2 and 3 shows that what was suggested is impossible.

Conclusion

The distance from an image to its correct perspective point is numerically equal to the magnification of the in-focus object at the image plane times the distance from the entrance pupil of the lens to the in-focus object. Establishment of the correct viewing distance will result in correct stereo reconstruction and that will result in a better man-machine interface. If alteration to reconstructed space is to be made to enhance ease of use in certain applications, that alteration should be made by changing the stereobase of the camera and not by changing the perspective because the latter distorts the shape of the reconstructed image.

Acknowledgments

The author is grateful to Rudolf Kingslake, former Director of Optical Design at Kodak and Emeritus Professor of the University of Rochester, for his educational and inspirational books and to Stephen Spicer of Telstra Research Laboratories (Australia) who, in the capacity of technical editor, reviewed this paper and suggested a great many improvements.

Comments are welcome. e-mail JHBercovitz@lbl.gov

Bercovitz, J. 1998, " IMAGE-SIDE PERSPECTIVE AND STEREOSCOPY "
in Stereoscopic Displays and Virtual Reality Systems V, S. Fisher, J. Merritt,
M. Bolas, Editors, Proceedings of the SPIE vol. 3295.