DEFINITIONS OF PERSPECTIVE DIMINUTION FACTOR AND FORESHORTENING FACTOR: APPLICATIONS IN THE ANALYSIS OF PERSPECTIVE DISTORTION

Hongyu Guo, University of Houston–Victoria

Abstract

There has been considerable confusion about the terms perspective diminution and foreshortening in the literature. Foreshortening factor has been defined and investigated in computer graphics literature, but only for the case of parallel projections. The case for perspective projection does not appear to have been investigated. In this paper, these concepts are elucidated through rigorous quantitative definitions of the perspective diminution factor and foreshortening factor. The inverse law and inverse square law that govern the perspective diminution factor and the perspective foreshortening factor in relation to depth were investigated and applied to a quantitative analysis of perspective distortion.

Introduction

Depth Perception and Representation by Humans and Computers

Human perception of the three-dimensional (3D) world is accomplished via a central projection onto the 2D retina. Because the retina has fewer dimensions, after the projection, the depth information of the original 3D scene is lost. This creates two tasks of opposite nature. One is depth perception, which is the inference of a 3D scene out of a 2D image. The other is depth representation, which is to output an image onto 2D media, given a 3D model. Humans have learned to deal with both. Depth perception is acquired through subconscious learning. Depth representation on 2D media is the art of perspective drawing, which was discovered in early Renaissance. Today, computers must deal with these, too, and which represent research topics in computer vision (for depth perception) and computer graphics (for depth representation).

In most of the situations in everyday life, a human can cope with this loss of depth information by using depth cues [1-3]. For example, when one is driving on the highway and sees cars ahead, there is first a process of object recognition; the images are recognized as cars. Next, prior knowledge is called upon; that is, all of the cars have approximately the same size, so prior knowledge must be used as a depth cue. A bigger image of the car indicates a shorter distance in depth, while a smaller image indicates a greater depth. Objects that are farther away look smaller. This phenomenon is known as perspective diminution.

Occasionally, the eyes are deceived, as in the case of trompe l'œil art. One popular form of trompe l'œil art is the chalk art on street pavements, creating 3D illusions. Julian Beever is one of these popular street pavement artists [4]. When viewed from a certain angle, the chalk drawing on the pavement can be perceived as points on certain 3D objects with varying depth; but, in fact, those are just points with colors on the ground plane. This illusion is called depth ambiguity, which is caused by the central projection.

Perspective Diminution and Foreshortening

The perspective diminution was understood even in prehistory art. However, the precise representation of depth on 2D media was not discovered until the Renaissance, when the principles of perspectives were fully understood. Before and during early Renaissance, artists had struggled with perspective drawing and had made various mistakes in inaccurate diminution rate of objects in depth.

The terms perspective diminution and foreshortening have been widely used in fine art, computer graphics and computer vision. However, the definitions of these terms tend to be vague and confusing. These two terms are also often used interchangeably [5]. William Longfellow [6] pointed out:

> The distinction between mere diminution from distance, and foreshortening, which is diminution from obliquity of view, is not to be forgotten here. If we stand in front of a square or circle, its plane being at right angles to our line of vision, it looks like a square or circle at any distance; it may grow larger or smaller, but its proportions do not change. If we look at it obliquely, the farther parts diminish more than the nearer, the lines that are seen obliquely

more than those that are seen squarely, and the shape is distorted. This is foreshortening. (p.2)

However, to the best of the author's knowledge, a quantitative analysis of perspective diminution and foreshortening is still missing in the literature today. Figure 1 is a drawing by Jan Vredeman De Vries, a Dutch artist in the Renaissance, in his book *Perspective* [7], published in 1604. It illustrates perspective diminution and foreshortening. When a person stands vertically but moves back in the depth direction, his size diminishes (diminution) and the different parts of his body diminish proportionally. When a person is lying down on the ground along the depth direction, the different parts diminish disproportionately, with the part in greater depth diminishing much faster (foreshortening).

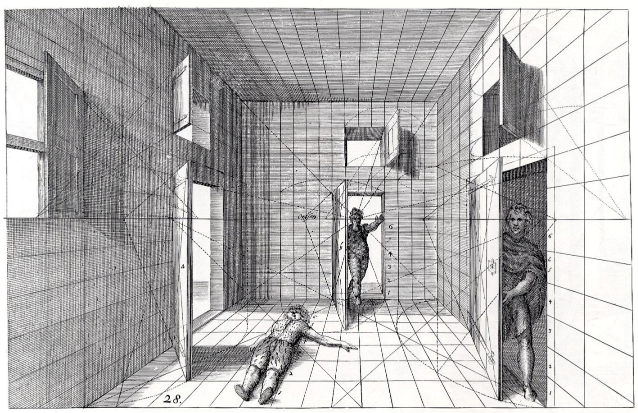


Figure 1. Perspective Diminution and Foreshortening [7]

In computer graphics literature, the foreshortening factor is defined for parallel projections, but not for central projections [8]. In parallel projections, the foreshortening factor is a constant, depending only on the angle of parallel projection lines to the image plane and the orientation of the line in 3D space relative to the image plane, but independent of depth. The computer graphics literature does not distinguish diminution from foreshortening either, because in parallel projection there is no diminution with depth. The diminution in depth in the central projection case has not been discussed.

The situation for central projection is more complicated because this factor also depends on the depth. The relationship between the length of the object in 3D space and its 2D projection is nonlinear if the object has a span in depth. It is even different along transverse directions and the depth direction. Because of this, two different factors, diminution and foreshortening, are defined in this paper; they are also defined locally in the differential sense for small-sized objects. The laws that govern the perspective diminution and foreshortening factors are investigated and applied to a quantitative analysis of perspective distortion.

Depth Inference from a Single Image in Machine Vision

Even with the loss of depth information in an image, humans can easily make inferences about 3D depth by examining a photograph in most situations, except for some exceptional cases like that of trompe l'œil art. Of course, such depth inference relies on object recognition and prior knowledge of the 3D objects, which are known as depth cues. One example of such depth cues is the known size of familiar objects like human figures or cars. Another example is the recognition of some planes and line segments and prior knowledge of their orientations; for example, the ground plane and the trunks of trees. It is known that the trunks of trees are in the vertical direction and, hence, all of the points on the same trunk have the same depth. The leaves connected to the trunk should have approximately the same depth, too. Since the bottom of the trunk is on the ground plane, this depth cue can be used to figure out the depth of the bottom of the trunk. Hence, the depth of the entire trunk and that of the leaves can be easily solved.

Finding space distance through measurements on the photographs has long been the effort in photogrammetry and machine vision. Accurate and fast depth calculation from images is important in machine vision and applications in unmanned vehicles [9-11] and robotics [12]. It is well understood that at least two photographs taken from different angles are needed in order to recover the depth information. Recently, there have been new efforts in depth inference from a single image [13-19]. Part of the reason is that stereopsis has its own drawbacks. In theory, stereopsis can find depth for any situation, though in practice it depends on the feature points extraction and registration of feature points in two images. When the depth is too large, the difference in the two images is too small, the uncertainty will be too big and in, general, will fail. This leads to an argument that a robot is better off being equipped with one eye than two [20].

However, depth inference from a single image is an ill-posed problem, and it makes sense only with prior knowledge and assumptions about the scene. For example, when a photograph of buildings and trees is given, it is possible that this is a photograph of the 3D scene with varying depth, but it is equally possible that this is a photograph of a photograph hanging on the wall, in which case all of the pixels have the same depth. Unavoidably, all of the depth inference from a single image relies on depth cues. Crankshaw [13] used computer-aided depth inference with applications in computer graphics to translate the photographic view of a building into the plan view. This process is known as reverse perspective analysis. Yu et al. [21] preprocessed the image in order to detect edges, extract straight lines, group the lines to form quadrilaterals and group them as planes. Saxena et al. [18] used supervised learning methods for depth inference. The results of these efforts was a depth map of the image, meaning a depth value was assigned to each pixel in the image. Experiments have been done with controlled scenes, the inferred results from which were compared with ground truth, which is a depth map obtained with a 3D laser scanner. The process is fully automated. However, this goal of fully automated and depth inference for each pixel seems too ambitious and the error rate is likely to be high [18].

A Model of the Camera and the Scene

The projective model for the camera and the scene are described next. The discussion in this paper is based on these assumptions of the camera model and the scene model. In the theory of optics, the subject distance μ (distance from the object to the center of the lens), the image distance v (distance from the image to the center of the lens) and the focal length f of the lens are related by

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

For different subject distance, μ, the image distance v is different. If the scene is non-planar in a 3D space, the image points of the scene do not lie in the same plane. A more precise definition of a scene will be presented shortly.

In practice, the film or image sensor of the camera is a plane. For each 3D scene point, a very small circular disk, called a circle of confusion, is recorded on the image sensor. When the scene point is within the depth of field, the circle of confusion is so small that it is indistinguishable from a point to the human eyes. This leads to the projective model, or geometric model, of a camera. It is also called the pinhole camera model. The image of the 3D world is the central projection on the image plane. The center of projection (COP) is the optical center of the lens.

Practically, with the cameras for photography, the image distance is slightly bigger than the focal length of the camera lens. When optics is not the concern, the term focal length and image distance are often used interchangeably in photography. The image plane (image sensor) of a camera is behind the lens and the image is upside down. With the geometric model of the camera, it does not hurt if the image plane is placed in front of the COP, just like the canvas of the artist. Everything in theory will be the same, but the image will be upright. In computer graphics, animation and video games, a virtual camera is employed. The image plane is the computer screen, which is the viewing window. The COP of the virtual camera is also called the viewpoint and is located in the position of the eye of the user of the computer. The virtual camera looks into this viewing window and sees the 3D virtual world, which is projected onto the computer screen. In the following discussion, the projective camera model is used.

A clarification of the scene is needed. A scene is defined to be a subset, S, of points in the 3D space. Some restrictions on subset S need to be imposed in order to be considered a scene. Let the origin of the coordinate system with $(x,y,z) = (0,0,0)$ coincide with the COP. The convention in some computer graphics applications (like DirectX) is to make the positive z pointing into the depth direction of the scene. It is assumed that all of the scene points are located in the half space $z > 0$. Imagine that all projection rays radiate into the half space, $z > 0$, starting from COP. It is further assumed that each ray intersects with S on at least one point. In general, the entire half space is not considered as a scene. It does not hurt but it is just not interesting or useful looking at a photograph of dense fog with a uniform gray color recorded on the image. Occlusion is allowed in the scene for the scene model to be general enough. If a ray intersects with S at more than one point, occlusion occurs. In such a case, the front-most point (with the lowest z value) will be projected onto the image plane, while other points on the same ray will be occluded and will not show up in the image.

A curved smooth surface is an example of a scene, but a scene is not limited to a single smooth surface. Surfaces with sharp edges, like a cube, can be part of a scene, where smoothness of the surface is not assumed. Occlusion in the scene in general is allowed. A sphere in front of a background surface is an example. In such cases, when the projection ray sweeps continuously across the scene, discontinuity in the z direction may occur. Buildings and trees, together with the ground surface, are an example of a scene. If any closed surface (like a cube or a sphere) has the interior as part of a scene, all of the interior points will be occluded and will not show on the image, no matter from what angle the image is taken. The back faces of the closed surface will be occluded as well. Only the front faces will show on the image.

In general, the scene points do not lie on the same plane. The terms 3D scene and non-planar scene are used interchangeably. A special case of a scene is when all of the scene points are located on a plane embedded in the 3D space. Such a scene is called a planar scene or 2D scene.

Perspective Diminution Factor and Perspective Foreshortening Factor

Coordinate Setup for Perspective Projection

A coordinate system (x,y,z) is set up with the origin O at the COP, as shown in Figure 2. The image plane is parallel to the x - y plane with an image distance f. Imagine a person is looking through this virtual camera. It is customary for the viewer to set positive x in the right direction, positive y in the upward direction and positive z in the forward or the depth direction. This way, all of the scene points that a viewer can see have positive z values that increase as a point moves forward, from the viewer's perspective. This convention of coordinate setup results in a left-handed coordinate system. In computer graphics applications, Microsoft DirectX adopts this left-handed coordinate system as its default for this viewer's convention. The x and y directions are called transverse directions, while the z direction is called the longitudinal or depth direction. On the image plane, an X - Y coordinate system is used. X is parallel to x and y is parallel to y, with C at the origin, where C is the principal vanishing point and is the intersection of the image plane and the z axis.

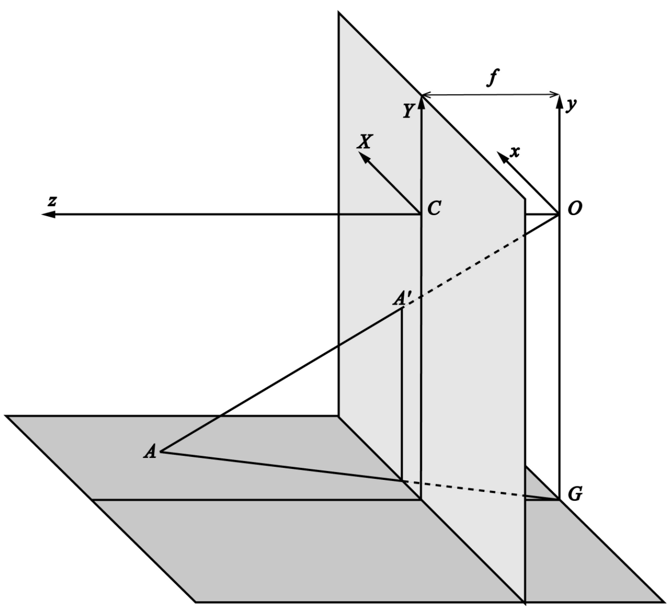


Figure 2. Camera Coordinate System Setup and Perspective Projection

With this coordinate system setup, for any scene point A with coordinates (x,y,z) in 3D space, the coordinates (X,Y) of the image point A' in the image plane are

$$X = f\,\frac{x}{z}$$
$$Y = f\,\frac{y}{z} \qquad (2)$$

Perspective Diminution Factor

For simplicity, a scene point, P , in the y - z plane (x = 0) is considered in Figure 3. However, this restriction of x = 0 is not necessary for the theory to hold; it simply makes the diagram easier to draw and easier to see. The definition of perspective diminution factor, perspective foreshortening factor, as well as the following theorems apply to any scene point with any x,y,z coordinate values.

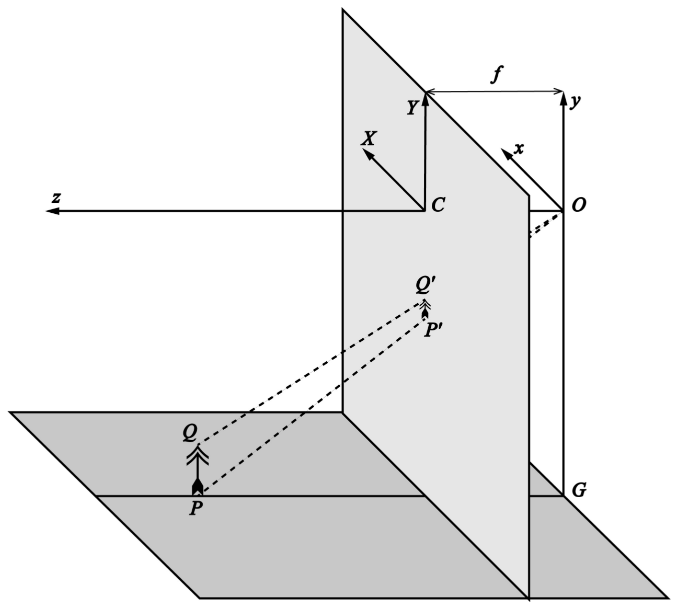


Figure 3. Perspective Diminution Factor

In perspective projection, the length in 3D space and the length of its projected image on the image plane have a nonlinear relationship. Locally, this nonlinear relationship can be approximated linearly using differential calculus. It is assumed that the object is a small line segment, PQ. Assume the line segment PQ is in the transverse y direction. After perspective projection with the center at O, the images of P and Q are P' and Q', respectively. If you imagine that PQ is a small tree on the ground, then $P'Q'$ is its image on the film of the camera.

Suppose point P has depth z and vertical coordinate y. Note that y is the vertical coordinate of point P, instead of the size of PQ. The size of the object is denoted by $\Delta y = PQ$. The size of the image is denoted by $\Delta Y = P'Q'$. From Equation (2), when z is kept constant but y is varied by a small amount (Δy), the image Y will change by ΔY:

$$\Delta Y = P'Q' = \frac{f}{z} PQ = \frac{f}{z} \Delta y \tag{3}$$

In the limit of $PQ \to 0$, the ratio

$$\lim_{PQ \to 0} \frac{P'Q'}{PQ}$$

is a constant, which shall be defined as the perspective diminution factor.

Definition 1. *Suppose a small object has coordinates* (x,y,z) *in 3D space and its image has coordinates* (X,Y) *in the image plane. The quantity*

$$M_{\parallel} = \left| \frac{\partial Y}{\partial y} \right| \tag{4}$$

is called the ***perspective diminution factor*** *at depth z.*

The symbol $M_{\parallel}$ is used because the transverse direction is parallel to the image plane. The diminution factor defined above is similar to the transverse magnification of optical systems, but here the projective model of the camera is used. It can easily be found that the diminution factor is completely determined by the depth z of the object in 3D space, and has nothing to do with optics.

It should be mentioned that it was assumed earlier that the object line segment PQ is oriented along the vertical y direction to intuitively motivate the definition but, in fact, Definition 1 applies to objects of any orientation (other than exactly parallel to the x - z plane, which has no variation in y). In this definition of the diminution factor, the key is to take the partial derivative of Y with respect to y. The object can be oriented in any direction, as long as there is a variation in y.

Theorem 1. *The perspective diminution factor* $M_{\parallel}$ *of an object located at* (x,y,z) *is inversely proportional to its depth z, as shown in Equation (5)*

$$M_{\parallel} = \frac{f}{z} \tag{5}$$

where f is the image distance, which is an intrinsic parameter of the camera.

The proof is straightforward by taking the partial derivative of Y with respect to y in Equation (2).

Meaning of the diminution factor. Suppose that PQ has unit length. When it is placed at a depth z and parallel to the y direction, $M_{\parallel}$ represents the size (height) of the image $P'Q'$. $M_{\parallel}$ is a function of depth z. The diminution factor in the x direction is similar to the y direction.

Corollary 1. *The perspective diminution factor can be equivalently defined as the magnification in the x direction. That is,*

$$M_{\parallel} = \left| \frac{\partial X}{\partial x} \right| = \frac{f}{z} \tag{6}$$

In most of the photographs in which the objects are far enough away (z big enough), $M_{\parallel} < 1$ holds, which means that the image is smaller in size than the object; although, under certain circumstances, when the depth z is small compared to the image distance, it is possible to have $M_{\parallel} > 1$, as in close-up photography and microscopic photography.

Perspective Foreshortening Factor

The magnification along the depth direction z (which is called foreshortening when the magnification is less than one) is quite different from the transverse magnification (diminution in y direction). In Figure 4, suppose that the segment PR is in the z direction with size $\Delta z = PR$, and the size of its image is $\Delta Y = P'R'$.

When y is kept constant but z is changed by a small amount (Δz), the image position Y will change by ΔY. In the limit of $PR \to 0$, the ratio

$$\lim_{PR \to 0} \frac{P'R'}{PR}$$

is a constant, which shall be defined as the perspective foreshortening factor.

Definition 2. *Suppose a small object has coordinates* (x,y,z) *in 3D space and its image has coordinates* (X,Y) *in the image plane. The quantity*

$$M_{\perp} = \left| \frac{\partial Y}{\partial z} \right| \tag{7}$$

is called the ***perspective foreshortening factor*** *at depth z.*

The symbol $M_{\perp}$ is used because the depth direction is perpendicular to the image plane. The perspective foreshortening factor defined above is similar to the longitudinal magnification of optical systems, but again there are no optics involved here; it is completely determined by the physical location (*x,y,z*) of the object.

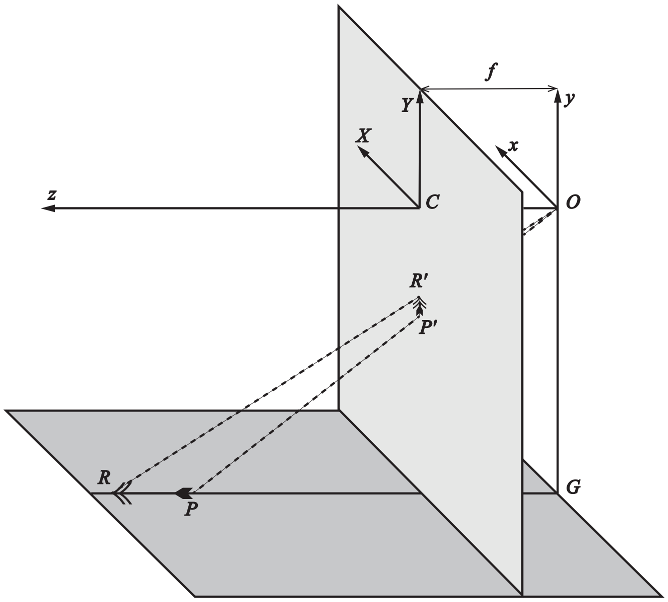


Figure 4. Perspective Foreshortening Factor

Theorem 2. *The perspective foreshortening factor* $M_{\perp}$ *of an object located at* (*x,y,z*) *is inversely proportional to the square of its depth z, but directly proportional to its vertical position* y*, as shown in Equation (8):*

$$M_{\perp} = \left| \frac{fy}{z^2} \right| \tag{8}$$

The proof is also straightforward by taking the partial derivative of *Y* with respect to *z* in Equation (2). The effect that the apparent size of an object in the depth direction diminishes according to the inverse square law in Equation (8) is called perspective foreshortening. The perspective foreshortening factor is also proportional to *y*. When *y* is smaller, $M_{\perp}$ is smaller, meaning that the foreshortening is more dramatic. Notice that $y/z = \tan \alpha$, where $\alpha = \angle COP$ is the view angle of the segment *PR*. Equation (8) can also be written as

$$M_{\perp} = \left| \frac{f \tan \alpha}{z} \right| \tag{9}$$

This means that when *z* is fixed, the foreshortening factor is proportional to tan *α*. The closer the object is to the line of sight (*z*-axis), or the smaller the viewing angle *α* is, the more dramatic is the perspective foreshortening.

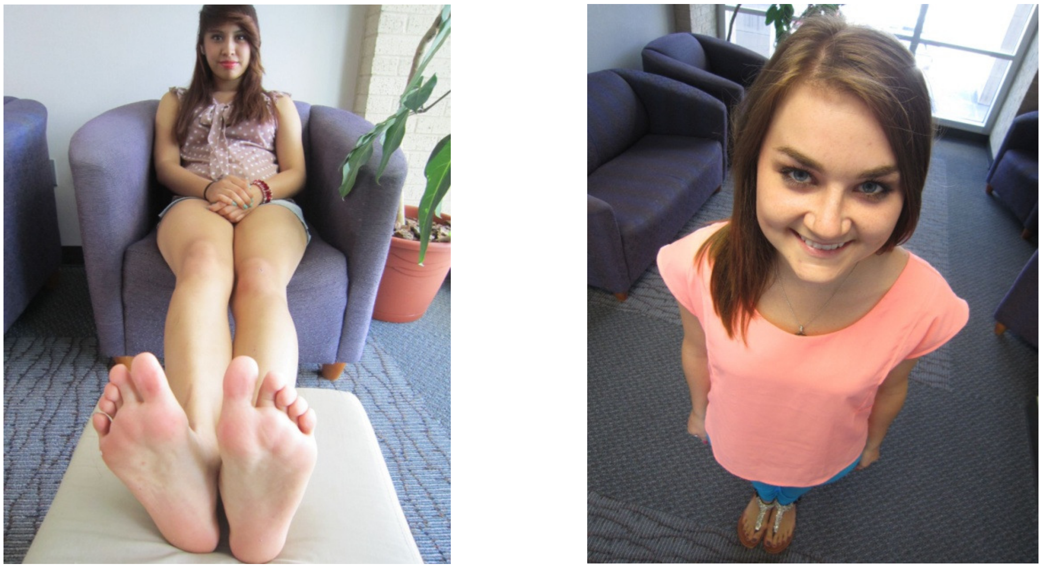
Similar to diminution, when the depth increases, the size of the image of object *PR* gets smaller as the effect of foreshortening. However, when the object *PR* is in the depth direction, the image diminishes, obeying an inverse square law, a rate much faster than diminution. The effect of foreshortening is more dramatic than diminution. Namely, the depth direction is shortened much faster than the transverse directions. When the depth is doubled, the apparent size of objects with transverse orientation shrinks to one half, while the apparent size of objects with longitudinal orientation shrinks to one quarter. This is also the reason that the photographs of 3D objects taken from different distances are not related by a similar transformation.

From Equation (8), when *z* is very small, $M_{\perp}$ and the image size can be very big, especially compared to the size of the objects in greater depth. Figure 5 shows two photographs of the foreshortening effect, where the length of the person is in the depth direction of the camera. The front part of the object (feet in the left picture, head in the right picture) is exaggeratedly magnified, while the far side is foreshortened dramatically. This effect is often associated with photographs shot with wide-angle lenses because one can shoot at very short subject distances with wide-angle lenses.

Quantitative Analysis of Perspective Distortion

Photographs are used in courtrooms and many other places as evidence. There is a saying “The camera does not lie.” However, everyone has some experience of seeing quite different photographs of the same scene. Some photographs can be very misleading; they can cheat the eyes. For example, Figure 6 shows two photographs of the same scene. The depth perception from the two images is quite different. The foreground arch and pillars are about the same size in both photographs. Using this foreground as a reference, the arch and pillars in the background of Figure 6(a) look much smaller than those in Figure 6(b). This makes the background objects in Figure 6(a) look much farther away, while those in Figure 6(b) look much closer. The brick walkway in Figure 6(a) seems about twice as long as that in Figure 6 (b). This phenomenon is known as perspective distortion. They have the same cause as for perspective diminution and perspective foreshortening.

There is a common misconception held by some laypeople, including some photographers, about this perspective distortion: a wide-angle lens stretches distance, while a telephoto lens compresses distance. Some photographers have offered explanations: perspective distortion is not caused by the focal length of a camera, either wide angle or telephoto;



(a) **(b)**

Figure 5. Foreshortening Effects

rather, it is caused by subject distance (subject to camera distance). It is a coincidence that wide-angle lenses are often used for short subject distances, while telephoto lenses are often used for long subject distances. This explanation is correct. In fact, Figure 6(a) was taken with a wide-angle lens from five meters away. Figure 6(b) was taken with a telephoto lens from ten meters away.

Figure 7(a) is a duplicate of Figure 6(a), for easier comparison. If two photographs of the same scene are taken at the same position (same subject distance), using two different cameras, one telephoto and one wide angle, then the two photographs are related by a similar transformation. The objects appear bigger in the photograph shot with a telephoto lens but the field of view of the telephoto lens is smaller. Figures 6(b) and 7(b) were shot at the same position (10 meters); 6(b) with a telephoto lens and 7(b) with a wide-angle lens. It can easily be seen that Figure 6(b) is exactly the same as the central part of Figure 7(b) (in the rectangle) being cropped and enlarged. Telephoto lenses perform this cropping and enlargement mechanically and automatically.

In fact, resizing (proportionally enlarging or shrinking) of a photograph is the true culprit of perspective distortion. The pictures in Figures 5(a) and 5(b) are examples of perspective foreshortening. They are also examples of perspective distortion. The human body's proportion seems to be distorted in Figures 5(a) and 5(b). In Figure 5(a), the feet seem to be disproportionately big, the legs long and the head small. In Figure 5 (b), the head seems to be disproportionately big, while the legs are short and the feet are small.

Resizing of the photograph causes perspective distortion. If the major objects are oriented parallel to the image plane, resizing the photograph will not alter the proportion of the objects, but rather change the perception of distance – an enlarged picture causes a perception that the objects are closer, while a shrunken picture causes a perception that the objects are farther away. Also, resizing the image can alter the perception of spacing in depth direction between the objects. This is illustrated by comparing Figure 7(b) with Figure 6(b).

If the major objects are oriented along the depth direction of the lens, as in the case of Figures 5(a) and 5(b), resizing the pictures will cause the distortion of the relative size of the different parts of the same object along the depth direction. Figure 5(a) looks like a picture of a person with abnormal body proportions (extremely long legs). If fact, if the picture is enlarged enough, it does not seem to be abnormal any more. If the photographer looks with his own eyes at the model in the scene in Figure 5(a), when his eyes are very close to the feet of the model, the same position where the camera was placed when the photograph was shot, what he sees is the same as the enlarged photograph of Figure 5 (a). That is, if the human eye is placed at the same position as the camera, the eye sees exactly what the camera sees, which is the photograph. When the photograph is post-processed by shrinking it to a smaller size, as in Figure 5(a), the human brain interprets the model as being farther away. The human brain has been trained and has learned the laws of foreshortening factor $M_{\perp}$ in Equation (8) unconscious-

(a) Wide-Angle Lens, From a Short Distance (5 meters) **(b) Telephoto Lens, From a Long Distance (10 meters)**

Figure 6. Same Scene Shot with a Wide Angle Lens and a Telephoto Lens from Different Distances

(a) Wide-Angle Lens, From a Short Distance (5 meters) **(b) Wide-Angle Lens, From a Long Distance (10 meters)**

Figure 7. Same Scene Shot with the Same Lens (Wide Angle) from Different Distances

ly. When the foreshortening factor $M_\perp$ is applied to objects at larger depth, it results in an interpretation in the brain that the person is out of proportion and, hence, a perception of perspective distortion. Figures 5(a) and 5(b) are examples of perspective distortions caused by shrinking the images (taking close-up photographs with wide-angle lens, then shrinking the image), where distance in the depth direction appears to be longer than it actually is. Figure 6(b) is an example of perspective distortion caused by enlarging the image (taking far-away photographs, then enlarging the images), where distance in the depth direction (length of the walkway) appears to be shorter than it actually is. This enlargement can be done by cropping the central part of a wide-angle photograph (Figure 7b) and editing it with software, or by a telephoto lens mechanically and automatically at the time the picture is being shot (Figure 6b).

The perspective diminution factor and perspective foreshortening factor, which were defined earlier, will now be used to give a quantitative analysis of perspective distortion. The perspective foreshortening factor $M_\perp$ in Equation (8) is inversely proportional to z^2. $M_\perp$ is equal to the length of the image in the photograph corresponding to unit length in the depth direction in 3D space. The inverse of $M_\perp$,

$$\frac{1}{M_{\perp}} = \left| \frac{z^2}{fy} \right| \tag{10}$$

represents the real distance in depth in three dimensions corresponding to unit length measured on the photograph; for example, 1 mm. This is a magnifying process that our brains are trained for when interpreting the depth represented in photographs. This magnification is nonlinear in z. The closer to the center in the photograph (larger depth z), the greater this magnification is. Although the human brain does not use formulas to do the calculation, it is trained through experience. It uses relative size (transverse diminution, like height and width) as a depth cue to calculate depth subconsciously. When the photograph is enlarged, either manually or by a telephoto lens, the brain is cheated, thinking it is closer than it really is. The brain then uses this wrong distance to estimate the spacing in the depth dimension in space. As in the example from Figure 6(b), the photograph was shot with a telephoto lens; thus, the background objects are perceived as closer in depth than they actually are.

The telephoto lens works the same way as telescopes. They are all long focal length optical systems. The focal length of a telescope should be even longer (infinite in the ideal case). The only difference between a telephoto camera lens and a telescope is that a telephoto lens forms a real image on the image sensor, while a telescope forms a virtual image, which is observed by the eye. Some astronomical telescopes form real images and record the images as photographs, too. If you have the experience of watching a far-away person walking in the direction of the line of sight through a pair of binoculars, you should notice that although the person's arms and legs seem to be swinging, the person appears to be marking time without any apparent displacement in the depth direction. Often times, you are not able to see whether the person is facing you or away from you in binoculars. Because the depth does not seem to change, you cannot even tell whether the person is walking toward you or away from you. To explain this, look at how the diminution factor $M_{\parallel}$ changes with depth z in Equation (5). Taking the derivative of Equation (5) with respect to z, one can find

$$\frac{dM_{\parallel}}{dz} = -\frac{f}{z^2} \tag{11}$$

When z is very large, $dM_{\parallel} / dz$ is close to zero. This means that the height of the image of the person is almost constant when z changes. This also explains the two pictures of the same scene shot with a wide-angle lens and a telephoto lens from different distances, as in Figures 6(a) and 6(b). The distance Δz in 3D space between the foreground objects and background objects is the same in the two pictures. However, because the wide-angle photograph (Figure 6a) was shot at a closer distance (smaller z), the difference in image size between the foreground objects and the background objects is big (bigger $dM_{\parallel} / dz$), while in the photograph shot with the telephoto lens (Figure 6b), the difference in image size is smaller because of the larger depth z (smaller $dM_{\parallel} / dz$). Hence, Figure 6(b) gives the illusion that the walkway is shorter and the illusion that the telephoto lens compresses depth.

Conclusions

Perspective analysis and depth calculation and inference are very important in machine vision, unmanned vehicles and robotics. Precise definitions of perspective diminution factor and foreshortening factor were given in this paper. With these definitions, various perspective imaging situations and perspective distortions were analyzed. Applying perspective analysis in machine vision will be future research directions.

Image Credits

The drawing in Figure 1, by Jan Vredeman De Vries (1604), is in the public domain. The remaining photographs were taken by the author.

References

[1] Howard, I. (2012). *Perceiving in Depth, Volume 1: Basic Mechanisms*. Oxford University Press, New York.

[2] Mon-Williams, M., Tresilian, J. R., McIntosh, R. D., & Milner, A. D. (2001). Monocular and binocular distance cues: insights from visual form agnosia. *Experimental Brain Research*, *139*(2), 127-36.

[3] Schwartz, S. H. (2009). *Visual Perception*. (4th ed.). McGraw-Hill Medical.

[4] Beever, J. (2010). *Pavement Chalk Artist: The Three-Dimensional Drawings of Julian Beever*. Firefly Books.

[5] Forsyth, D., & Ponce, J. (2011). *Computer Vision: A Modern Approach*. (2nd ed.). Prentice Hall.

[6] Longfellow, W. P. P. (1901). *Applied Perspective for Architects and Painters*. Houghton Mifflin Company, The Riverside Press, Cambridge.

[7] Vredeman de Vries, J. (1968). *Perspective*. Dover Publications.

[8] Angel, E., & Shreiner, D. (2011). *Interactive Computer Graphics: A Top-down Approach with Shader-based OpenGL*. (6th ed.). Addison-Wesley.

[9] Thamma, R., Kesireddy, L.M., & Wang, H. (2012).

Centralized Vision-Based Controller for Unmanned Guided Vehicle Detection. *International Journal of Modern Engineering*, *12*(2), 58-65.

[10] Vargas-Clara, A., & Redkar, S. (2011). Dynamics of a Vertical Takeoff and Landing (VTOL) Unmanned Aerial Vehicle (UAV). *International Journal of Engineering Research Innovation*, *3*(1), 52-58.

[11] Eydgahi, A., Godaliyadda, D., & Falase, S. (2007). A Simulation Software for Autonomous Navigation of Unmanned Surface Vehicles Using MATLAB Environment. *International Journal of Modern Engineering*, *7*(2).

[12] Zhang, B., & Skaar, S.B. (2010). Laser-Assisted Uncalibrated Vision Guide Robotic De-palletizing. *International Journal of Engineering Research Innovation*, *2*(2), 17-24.

[13] Crankshaw, N. M. (1990). CARPA: Computer Aided Reverse Perspective Analysis. *APT Bulletin*, *22*(1-2), 117-129.

[14] Gao, X., Hou, X., Tang J., & Cheng, H. (2003). Complete Solution Classification for the Perspect-Three-Point Problem. *IEEE Trans PAMI*, *25*(8):930-943.

[15] Wu, B., Ooi, T. L., & He, Z. J. (2004). Perceiving Distance Accurately by a Directional Process of Integrating Ground Information. *Nature*, *428*(6978):73-77.

[16] Cao, Y., Xia, Y., & Wang, Z. (2010). A Close-Form Iterative Algorithm for Depth Inferring from a Single Image. *Proc. Perspectives in Neural Computing*, 729 -742.

[17] Saxena, A., Chung, S. H., & Ng, A. Y. (2005). Learning Depth from Single Monocular Images. *Neural Information Processing System*, *18*.

[18] Saxena, A., Sun, M., & Mg, A. Y. (2008). Make3d: Learning 3-d Scene Structure from a Single Still Image. *IEEE Tran. PAMI*, 31, 824-840.

[19] Saxena, A, Schulte, J., & Mg, A. Y. (2007). Depth Estimation Using Monocular and Stereo Cues. *International Joint Conference on Artificial Intelligence*.

[20] Biever, C. (2005). Why a robot is better with one eye than two. *New Scientist*, *188*(2530):30.

[21] Yu, S. X., Zhang H., & Malik, J. (2008). Inferring Spatial Layout form a Single Image via Depth-ordered Grouping. *Workshop on Perceptual Organization in Computer Vision*.

Biography

HONGYU GUO is an Assistant Professor of Computer Science at the University of Houston – Victoria. He received his Ph.D. degree in Computer Science from the University of Florida. His research interests are in computer graphics, computer vision, image processing and artificial intelligence. He can be reached at guoh@uhv.edu.