Visual Cues and Pictorial Limitations in Photorealistic Images

Christopher G. Barbour and Gary W. Meyer

CIS-TR-90-27

Department of Computer and Information Science
University of Oregon

* 35 mm slides for the color illustrations labelled Slide 1 to Slide 13 are available from Gary Meyer at the cost of $0.75 each.

Visual Cues and Pictorial Limitations in Photorealistic Images

Christopher G. Barbour
Department of Psychology
University of Oregon
Eugene, OR 97403

Gary W. Meyer
Department of Computer and Information Science
University of Oregon
Eugene, OR 97403

The limitations of two dimensional pictures as representations for reality are discussed. A review is done of the perceptual cues necessary to convey a sense of realism. These cues include, but are not limited to, binocular disparity, field of view, accommodation, vergence, and chromatic adaptation. Examples are given of how the physical characteristics of two dimensional pictures limit the use of these cues in computer graphic images. Techniques developed by artists and photographers to overcome some of these limitations are discussed.

1.0 Introduction

The objective of much work in realistic image synthesis is to create a picture that is optically correct. The focus, therefore, of many synthetic image generation algorithms is to determine the array of electromagnetic energy that passes through a particular plane in space. The sophistication of this approach has continued to increase, drawing on work from areas as diverse as radar scattering theory and radiation heat transfer (Hall, 1989). This work has now reached the point where it is impossible in a controlled experiment to tell the difference between a real scene and a computer generated picture of that scene (Meyer, Rushmeier, Cohen, Greenberg, & Torrance, 1986).

Unfortunately this approach almost completely ignores the fact that two dimensional pictures are severely limited in their ability to provide a perfect optical reproduction. The above comparison experiment between the real scene and the computer generated picture could only be conducted under restricted viewing conditions. Physical constraints such as limited dynamic range and field of view also diminish the ability of pictures to achieve optical identity. In addition there are practical considerations that make a perfect optical reproduction undesirable even if it could be achieved.

The physical limitations of pictures and the practical uses to which images are put makes it questionable whether optical identity is the correct objective in realistic image synthesis (Mills, 1985). However, if the reproduction is not optically indistinguishable from the original environment the viewer will see a different two dimensional array of light and their impression of the scene will be altered. An approach which takes the perceptual experience of the viewer into account is required. To achieve the best realism it is necessary to understand how the percept generated by the reproduction differs from that produced by the original scene and to modify the image synthesis algorithm to account for these distortions. Fortunately, the solutions to many of these problems can be found by studying techniques that have been developed by artists and photographers. In several cases, however, research remains to be done in order to achieve the best result.

This paper reviews the limitations that prohibit pictures from presenting an observer with an optical field identical to that experienced in an actual scene. The focus is on two dimensional static images (see Van de Grind, 1986 for a discussion of dynamic images). Physical limitations like fixed point of view, finite size, and flatness are covered first, followed by lighting problems such as limited dynamic range and the need for a viewing illuminant. For each limitation the relevant psychophysical data are discussed and the impact on the viewer's percept is described. Attempts by artists and photographers to overcome each limitation are also discussed. Wherever possible, examples of the limitations in existing computer generated images are presented.

2.0 Pictures are made from a single station point

The creation of a computer graphic image involves the perspective projection of a three dimensional scene onto a two dimensional image plane. This is accomplished by following projection lines from a single station point back into the scene and then determining the point of intersection between these lines and the image plane. As such, each computer generated image represents the world as seen from a single eyepoint. Because our visual system has two eyes, our brain receives information about the natural world from two slightly different viewpoints. A single computer generated picture therefore lacks one of the cues that our visual system uses to help us decipher the three dimensional world. This cue is known as binocular disparity.

2.1 Binocular disparity

When both eyes fixate on an object, some areas of the object project to corresponding points (same retinal positions) on the two retinas and some areas project to noncorresponding points (Figure 1). This occurs because the eyes view an object from two different locations in space. Furthermore, as a result of the eyes being separated, an imaginary surface (called the horopter) is located in visual space at a distance from the observer that depends on where the eyes are fixated. Any object falling on this surface will project a retinal image that is located on corresponding points of both retinas (Goldstein, 1984). Conversely, any object located off this surface will project to noncorresponding points. The difference in distance between these noncorresponding points on the two retinas is called *disparity* and the degree of disparity between the two points will vary as a function of how far the object is from the horopter (Figure 1). If the disparity is great enough, the percept of the object will be doubled. Otherwise, a single fused image will be perceived. Whether the image is doubled or fused, disparity is a depth cue our visual system uses to perceive a three dimensional object from a two dimensional retinal image (Goldstein, 1984).

According to Graham (1965), the limiting range of binocular disparity as a cue can be considered as the greatest distance at which an object can be placed and still be considered nearer than an object at infinity. When the angular disparity in the separation between the two images is taken as 30 sec of arc and the interocular distance is taken as 65mm, the limiting range has been calculated to be 495 yards. This underscores the large range of distances over which disparity can act as a cue for depth. Another way of measuring the limits of binocular disparity is to determine the threshold angular disparity between two images needed to produce stereopsis. This value has been found to be approximately 2.0 sec of arc, but can vary up to 40 sec depending on the conditions used in the experiment (e.g. adapted state of the eye; see Mueller and Lloyd, 1948).

Pictures represent a three dimensional object in a two dimensional plane. If the viewer fixates on an area of the represented object, all other areas will be in the same plane and fall on corresponding points, thereby eliminating any disparity that would have normally occurred if the object was actually three dimensional (Helmholtz, 1881; Meinel, 1973; Hochberg, 1979, 1980). This lack of disparity not only reduces the number of depth cues available to an observer but also cues the observer that the object is in fact two dimensional (Pirenne, 1970; Hochberg, 1979; Sedgwick, 1980; Goldstein, 1984;). Concerning the double images caused by large disparity, Meinel (1973) proposes four perceptions that are lost as a consequence of pictorial representation: 1) various gradations of transparency of half of the double image in which one half may vary between entirely transparent to entirely opaque; 2) the creation of unique shapes due to the overlapping of dissimilar images; 3) the halves of the double image seen at different vertical positions as a result of slight tilting of the head; and 4) color/brightness effects resulting from the overlapping of two images that vary in color and/or luminance.

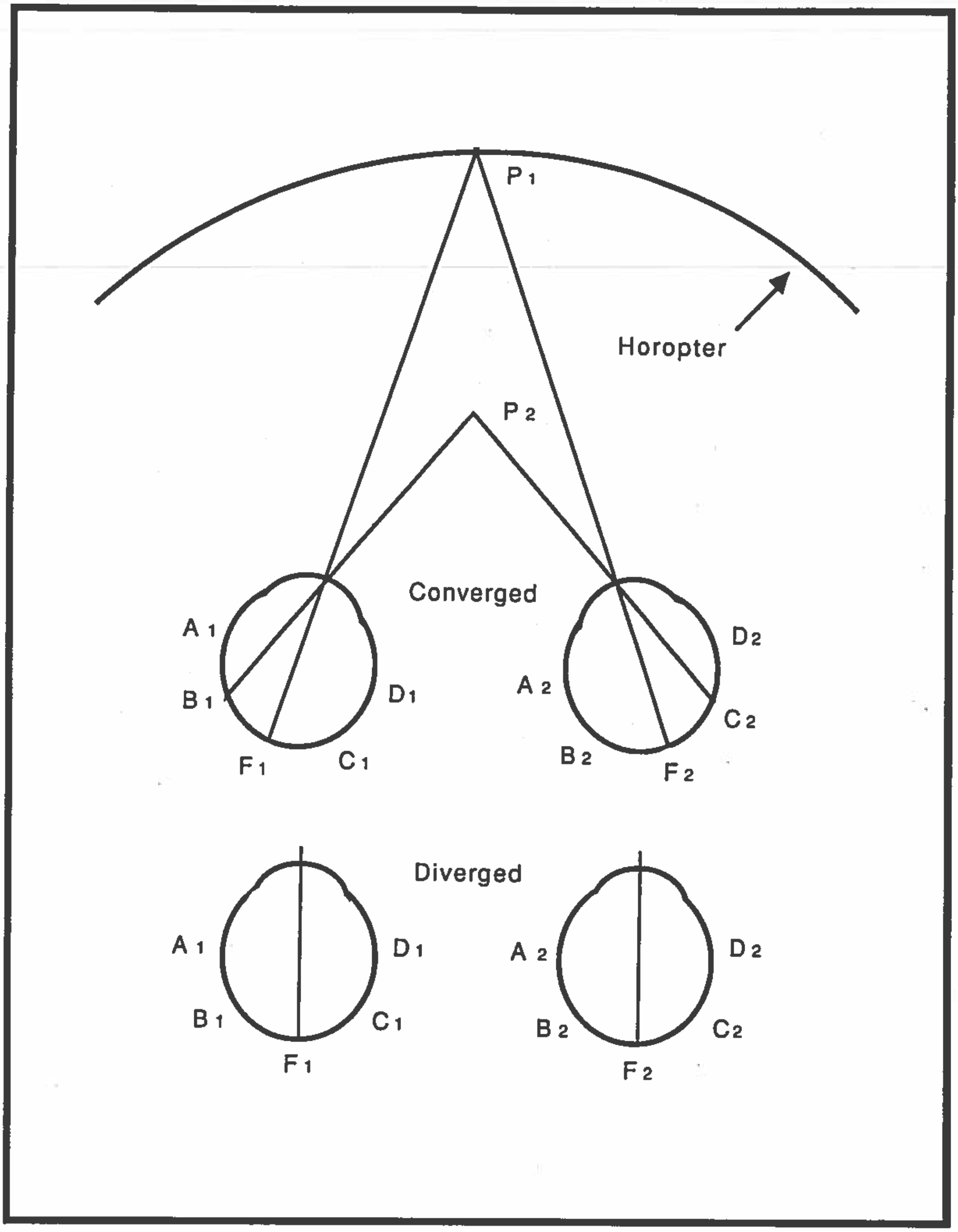

Figure 1: Due to independent movement of each eye, converged eyes (top) image point P_1 to the same points F_1 and F_2 of the retina while diverged eyes (bottom) do not. For the eyes converged to point P_1, all points on an arc (the *horopter*) through P_1 will project to corresponding points of the retina. Point P_2 is not such a point.

2.2 Overcoming limitations of monocular images

Meinel (1973), like Hochberg (1980), suggests that painters are not without possible solutions to offset the limitations caused by a single point of view. He refers to the work of Evan Walters (1894-1951) in which the four lost perceptions mentioned above are actually painted into the picture. The creation of stereo image pairs is the straightforward solution to the absence of the disparity depth cue in monocular images. Stereo image pairs have been created since the earliest days of computer graphic image synthesis. Recent development of liquid crystal shutters that can alternately present images to the left and right eyes has stimulated new interest in this area. The use of head mounted displays as part of the effort to produce an entire *virtual world* for an observer (Chung, Harris, Brooks, Fuchs, Kelley, Hughes, Ouh-young, Cheung, Holloway, & Pique, 1989) will demand that additional attention be paid to this subject.

Despite the evidence supporting the importance of disparity as a depth cue, Rock (1975) points out that this cue is often present with other monocular depth cues (see Section 4.4) that enhance the perception of depth. If these monocular cues are not present, the sensation of depth is more difficult to achieve for some observers, even though the disparity cue is still present. One method of overcoming the loss of binocular disparity in individual computer graphic images is to make sure that these monocular cues are properly rendered and are emphasized wherever possible. The importance of monocular cues in the perception of depth is illustrated by experiments that show a pronounced sensation of depth when information that conflicts with the monocular depth information is not available to the viewer. This is done by viewing the picture monocularly through a peep hole (see Section 3.2).

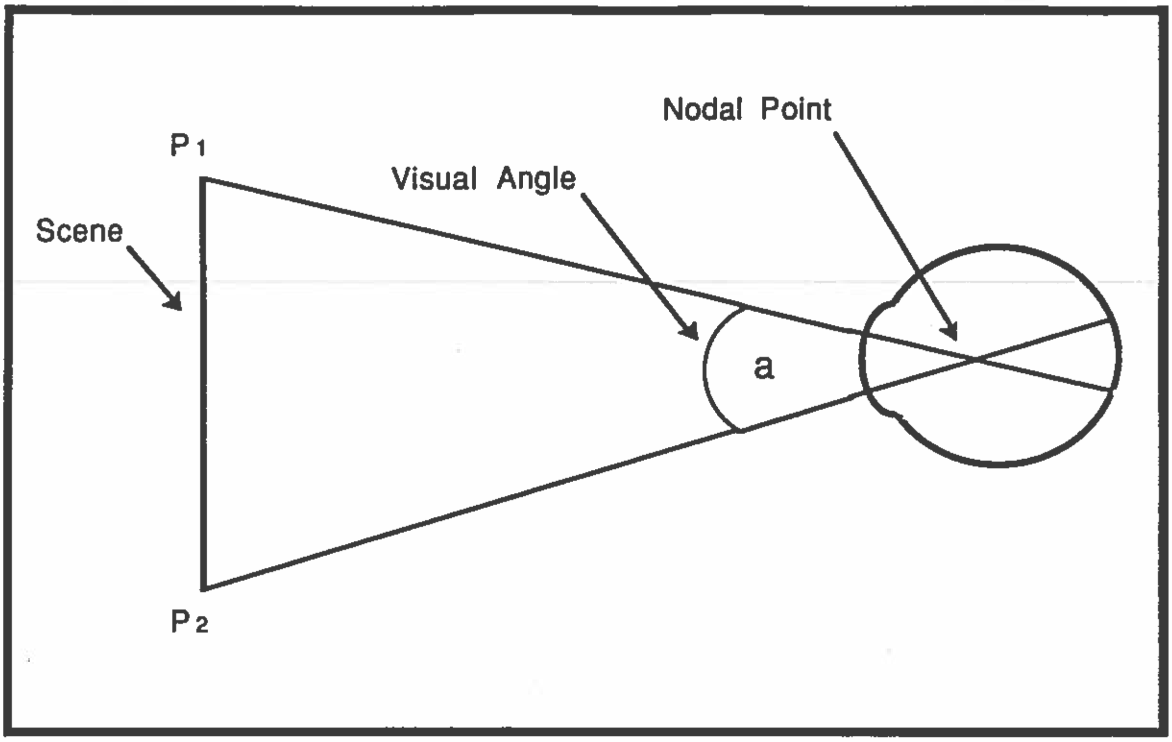

Figure 2: The scene is delineated by P_1 and P_2. Its expanse determines visual angle α with respect to the nodal point of the eye.

3.0 Pictures have a finite size

Most computer graphic images are displayed using color television monitors or color hardcopy devices. While there are projection systems such as IMAX which present very large images, these devices are impractical for the day to day working environment in which most computer generated pictures are used. Therefore a typical computer graphic image is limited in its ability to convey a sense of realism because it has a clearly identifiable boundary and it does not occupy our entire field of view.

3.1 Field of view

An object at some finite distance from an observer projects an image onto the observer's retina. Because the size of the retinal image is directly related to both the size of the object and the distance of the object from the retina, it is useful to describe retinal image size in terms of a single parameter: the visual angle the object subtends with respect to the nodal point of the eye (De Valois & De Valois, 1989; see Figure 2). The entire scene viewed by an observer also projects an image to the retina, and the size of this retinal image is directly related to the expanse of the scene. The visual angle subtended by the scene is therefore a measurement of the expanse, and it is called field of view.

The maximum field of view for an individual using both eyes is approximately 200 degrees visual angle (Darley, Glucksberg, & Kinchla, 1986; Van de Grind, 1986) and in most cases, the field of view of a real scene will approximate this. However, a picture will rarely possess such a large field of view and will almost always be much smaller (Helmholtz, 1881; Evans, 1959; Pirenne, 1970; and Gibson, 1982). Gibson (1982) points out that this "narrowing" of the scene in a picture will not provide information to the periphery of the eye that is normally present during real scene viewing and pictures are therefore limited in this regard.

There is specific evidence as to how field of view affects perception of two dimensional displays. For example, an experiment was conducted in which the "sensation of reality" was shown to vary as a function of field of view. The measured psychological effects were: continuity and unification between display space and observer space; feeling of expanse; naturalness; feeling of depth; and impressive powerfulness. (Hatada, Sakata, & Kusaka, 1980). The results indicated that visual displays with horizontal visual angles between 30 and 100 degrees and vertical visual angles between 20 and 80 degrees produce psychological effects that increase the sensation of reality. Larger fields produced saturated responses. Another related study has shown that the subjective quality of a display, rated on a 1 to 100 scale, increases linearly with the logarithm of picture angle (Westerink & Roufs, 1989).

In addition to a limited field of view, the frame of a picture also minimizes the effectiveness of a picture as a high fidelity surrogate because the scenes that are represented do not have a margin or frame surrounding them (Helmholtz, 1881; Evans, 1959; Gombrich, 1956; Pirenne, 1970, 1975; Hochberg, 1978a; Haber, 1980a; and Reed & Jones, 1982). Moreover, the presence of a frame has been shown to affect visual perception of the image the frame surrounds. For example, the perceived orientation of a rod positioned inside a frame has been shown, among other things, to be a function of the degree of tilt of the frame (Witkin, 1959). Furthermore, because perceived shape and form is affected by the simultaneous perception of other shapes and forms (see Evans, 1959), a picture of a tree will look taller in a long skinny frame compared with a short squat frame.[1]

3.2 Overcoming limitations of image size

Occluding the appearance of the frame has been shown to affect the percept of a picture. This style of viewing helps to eliminate those cues (e.g. frame of the picture, surface of the picture) that inform the viewer that what is being looked at is a picture and not a real scene (Smith and Gruber, 1958; Evans, 1959; Pirenne, 1970; Gibson, 1982;). Helmholtz (1881), Hochberg (1978b), and Pirenne (1970, 1975) report that viewing a picture through a peephole such that the frame is occluded from view, will often increase the sensation of depth in a picture. This increased percept of depth can be so compelling that it has been likened to the 3-dimensional effect that is experienced when looking through a stereoscope (Pirenne, 1970). In fact, this "peephole technique" has been used in art galleries and

---

[1] It is also believed that the frame will provide information to the observer concerning picture surface attributes. This information is thought to affect perception of the picture which further limits the picture as a high fidelity representation of real scenes (see Section 4.3).

museums in order to enhance the sense of depth thereby increasing the visual impact of the picture (Pirenne, 1970)

A view camera is a device which allows the photographer to position the film plane so that it is not perpendicular to the axis of the lens. This provides additional freedom in composition but it does introduce perspective distortion into the picture. One common use of the view camera is to adjust the perspective of tall objects such as buildings so that all verticals in the image are parallel. Without this modification the frame of the picture tends to exaggerate the perspective convergence of vertical lines in a way that does not appear correct (Evans, 1959; Stone, 1987).

In Slides 1 and 2 the computer generated still life picture of a book and a ball have been created by using the ray tracing method. In Slide 1 a simple perspective projection was produced by keeping the view plane perpendicular to the rays that were cast into the scene. The verticals of the book are not parallel to the frame of the picture causing the book to appear as if it is tipping. In Slide 2 the picture has been redone with the projection plane held parallel to the front of the book. This eliminates the problem with the book but introduces some distortion of the sphere. While the effectiveness of these modifications are subject to individual interpretation they are representative of the type intentional distortions that photographers introduce into their work by using the view camera.

4.0 Pictures are flat

There are a number of consequences that result from using a perspective projection algorithm in realistic image synthesis to create a two dimensional representation of a three dimensional scene. One of the most important is that the picture becomes an identifiable object in our environment rather than a window into a scene that lies beyond it. This is due to the surface perception capabilities of our visual system. Other inputs from the visual system are also missing because a picture lacks physical depth. These include proprioceptive feedback due to the independent movement of each eye while fixating on objects at different distances (vergence) and focussing on objects at various depths (accommodation).

4.1 Accommodation

Accommodation is the process by which the eye changes the shape of the lens in order to focus on objects that are at different distances from the viewer. It has been shown that, under normal monocular conditions, vergence (see Section 4.2) will occur during accommodation. This is called accommodative vergence. Experiments that control for vergence indicate that accommodation and accommodative vergence are inaccurate sources of depth information for distances past a few feet (Pirenne, 1970; Rock, 1975). However, for shorter distances, these cues become more reliable (Rock, 1975).[2] Information from accommodation should therefore be considered as relevant for those pictures depicting short viewer-object distances.[3]

Because pictures are flat, information from accommodation is completely lost and a viewer will not be able to accommodate to objects of different represented distances (Hochberg, 1979). The missing information has been proposed to be in the form of proprioceptive[4]

---

[2] Rock (1975) does point out that for even these short distances, accommodation and accommodative vergence are not very reliable compared to other depth cues.

[3] Pirenne (1970, 1975) proposes that the information from accommodation is unimportant in pictorial representation. However, he does not address the issue of pictures representing objects that are at short distances from the viewer.

[4] A proprioceptive cue refers to a possible neural signal arising from sensory fibers in the muscles of the eye. These muscles are responsible for converging or diverging the eyes and the sensory fibers in them would be able to provide a neural signal as to the state of degree of contraction of the eyes. This signal could then provide depth information concerning the fixated object (Rock, 1975).

feedback from sensory fibers in the ciliary muscles responsible for changing the shape of the lens (Rock, 1975). Another form of lost information could be the absence of blurred non-accommodated objects. This blurring of objects in front of and behind the accommodated object could be used as a cue for depth (Evans, 1959). Therefore, unless depth of field is provided during picture viewing (see Evans, 1959), this potential source of depth information will not be present.

4.2 Vergence

An observer can view an object with both eyes and maintain a perception of a single image. This is because both eyes can be converged to the angle that causes the image of the object to fall on corresponding retinal points (same retinal positions; see Section 2.1) of both retinas (Figure 1). If an object is positioned at a greater or a lesser distance, a single image can still be maintained because the eyes diverge or converge to a new angle that keeps the image positioned on corresponding retinal points (Figure 1). During vergence, the eye reflexively accommodates. However, experiments studying the effect this reflex has on vergence show accommodation to be an unimportant factor. These studies also find vergence to be an effective depth cue, independent of accommodation, up to distances of six feet and inconsequential for distances past 30 feet (Rock, 1975). Furthermore, within the range where it is effective, vergence appears to be a more effective depth cue than accommodation (Swenson, 1932).

Because vergence provides information about depth for short viewer-object distances, the loss of this information during picture perception is an important factor to consider when creating a pictorial representation. Meinel (1973), Hochberg (1980) and Sedgwick (1980) point out that pictures do not require the observer to converge or diverge the eyes to objects of different represented distances and therefore, any cue that vergence provides towards the perception of depth will be lost. This lost cue is most likely to be feedback from sensory fibers in the eye muscles that code for the degree of muscle contraction (Rock, 1975). Other types of lost information have been proposed by Hochberg (1980) and Meinel (1973). They point out that not only will proprioceptive feedback be lost but, because vergence is the same for all represented distances, the normal doubling and blurring of images (those objects projecting images to noncorresponding retinal points; see Section 2.1) will be missing during picture viewing, and should therefore provide information to the viewer that the image is flat (Hochberg, 1980).[5]

4.3 Surface perception

A picture can be created using different kinds of media. Most of these media (e.g. paint and canvas, photographic emulsion, computer monitor screen) possess surface properties that can be distinguished from the image represented in the picture. For example, a painted picture will have a distinguishable surface, in part, because it has the property of texture. Texture is present because of the uneven application of the paint on the canvas or wood. Moreover, the canvas or wood itself also possesses a texture that will contribute to the overall surface texture of the picture. To the extent that this surface is perceivable, it will be possible to distinguish it from the percept of the represented scene.

There are numerous properties of a surface that are important in determining its visual appearance (Gibson, 1982). They include: whether the picture radiates or reflects light; the quality and quantity of illumination from an outside source; diffuse reflectivity; specular reflectivity; texture; and opacity (Gibson, 1982; Foley, van Dam, Feiner, and Hughes, 1990). The visual response to these various properties has been studied (for example, see

---

[5] This "doubling" occurs because objects that are at different relative distances than the object being fixated upon will create images that fall on non-corresponding points between the two retinas.

Bartlett, 1965; Brown and Mueller, 1965; Graham and Brown, 1965; Gibson, 1982). However, the visual psychophysics of these properties are quite extensive and beyond the scope of this paper. What is important is whether the perception of the picture surface affects the perception of the represented scene and if so, in what way. In most cases, the surface of a picture will be perceived along with the scene the picture is representing. To the extent that this is true, the picture will be limited. Therefore, the effects of surface perception (subsidiary surface awareness, see Pirenne, 1970) on the perception of the represented image should be considered important in the context of realistic pictorial representation.

The ability to detect the picture's surface appears to degrade the percept of depth in the picture (Evans, 1959; Pirenne, 1970; Haber, 1980a, 1980b). For example, research has shown that when a real scene is given an artificial surface (sheet of cellophane placed between observer and scene) and other depths cues are eliminated (e.g. motion parallax), observers are unable to distinguish the scene from a 2-dimensional representation of the scene (Hochberg, 1962). Other studies have shown that when surface information is absent in the represented scene, it appears more three dimensional. In fact, the percept of depth is likened to that which is experienced under stereoscopic viewing (Helmholtz, 1881; Pirenne, 1970). Haber (1980a) points out that the surface of a picture "contributes massive amounts of information for flatness. The texture of the canvas or photographic paper can usually be seen, and it projects a zero gradient over the surface for the observer standing directly in front of the picture. If the picture is viewed from an angle, the gradient from the slant exactly matches that of the wall surface. Thus the surface-perspective scale of space provides information about the spatial relations within the scene depicted by a picture and at the same time it specifies that the picture surface itself is flat" (Haber, 1980a; p. 374). Surface awareness also appears to affect the appearance of shape of the represented scene.

It has been known for some time that a monocularly viewed picture that is viewed from the wrong perspective will not appear distorted, although the theory of perspective drawing dictates that it should (Pirenne, 1970; Cutting, 1986). This paradox has been termed "La Gournerie's paradox" (it was first discussed by La Gournerie in 1859; Cutting, 1986). The viewer's awareness of the picture surface has been suggested as the critical factor in maintaining the stable perception of the represented scene (Pirenne, 1970). This hypothesis is supported by the fact that a photograph of a picture taken at the wrong viewing point will result in a representation of that picture with perceptual deformations, presumably because the shape and position of the picture surface is no longer available to the viewer (Pirenne, 1970).

4.4 Overcoming image flatness

The visual system uses static monocular depth cues to construct a percept of depth. These cues are interposition, relative size, relative height, atmospheric perspective, familiar size, texture gradient, shadowing, and linear perspective (see Goldstein, 1984). Artists and photographers intentionally introduce these cues into their pictures in order to enhance depth and overcome the inherent flatness of the image. For example, consider the collection of random blocks in the computer generated image in Slide 3. This picture was created using a simple illumination model and scan line hidden surface algorithm. A photographer might first arrange the blocks in this picture so that they occlude one another (Slide 4). This provides the cue of interposition, which gives the observer knowledge about the relative ordering of the blocks in depth. Next the photographer could place the larger blocks in front and smaller ones in the rear (Slide 5). This amounts to a trick because the actual size of the objects is unknown, but it is something that photographers do even when the objects can be identified (Evans, 1959). This rearrangement exploits the cue of relative size, which is the effect perspective has on the size of distant objects. Finally, the blocks might be arranged so as to recede to a vanishing point as in Slide 6. This exaggerates the cue of linear perspective.

In addition to using static monocular depth cues, other techniques have been developed to help overcome image flatness (see Meinel, 1973; Hochberg, 1979). For example, Hochberg (1979, 1980) suggests that painters (e.g. Rembrandt) have developed a solution to offset the vergence problem by selecting a few areas in the real scene that have little depth as "focal regions". These regions are then painted with high detail. Areas outside these regions contain a higher amount of depth information and are painted with large swatches of paint that provide little detail. Objects in these areas will only look normal and recognizable when viewed with the periphery of the eye because of the periphery's low spatial acuity. If viewed with the fovea they will appear as blurred, sketchy, blobs faintly resembling the objects they are meant to represent. As a consequence of this style of painting, the objects depicted will only look normal when the viewer maintains his fovea on the focal regions and periphery on the non-focal regions. The restriction of the viewer's gaze to only one or two areas in which the image appears normal and recognizable should therefore limit the flatness information to the observer (Hochberg, 1980).

5.0 Pictures have a limited dynamic range

There is a limit to the maximum amount of light that the phosphors of a television monitor can emit or to the minimum density that the dyes of a photographic film can obtain. For each reproduction medium the difference between the amount of light emitted (or reflected) by the blackest possible black and the whitest possible white is therefore bounded. This difference in light intensity is referred to as the dynamic range of the device. The human visual system is capable of operating over a much wider dynamic range than can be reproduced on any currently available display device. It accomplishes this by adapting to the average brightness level present in the scene. This means that no one reproduction medium can create the full range of light intensities over which the visual system operates, from the bright light of the midday sun to the dim light of the moonlit sky. Fortunately, brightness adaptation and brightness perception allow our visual system to adjust to the dynamic range available with a particular display device. There are differences that remain, however, between the percept created by the original range of intensities in the scene and the range that are available on the display device.

5.1 Brightness adaptation and perception

Before discussing luminance range, some terminology will first be established. A physical measure of light is radiance. Radiance is the amount of absolute electromagnetic energy emitted from or reflected off of an object. If the light is providing a visual stimulus, this measure can be misleading because the visual system is differentially sensitive to each of the wavelengths of light. Luminance is a perceptual measure of light that takes this differential sensitivity into consideration and thereby indicates the effectiveness of light as a visual stimulus. Luminance is therefore used to measure the intensity of light independent of wavelength composition as experienced by a viewer under real and represented scene viewing (Riggs, 1965).

Visual behaviors will change depending on the luminance value of the scene. For example, acuity, contrast sensitivity and hue discrimination are dependent on the adapted state of the eye which in turn is dependent on the level of luminance in the scene (other factors such as time of exposure and pre-adapted state of the eye are also critical factors in affecting these behaviors; see Bartlett, 1965). Differences in brightness magnitude estimation are also dependent on the level of stimulus luminance (Goldstein, 1984). The potential effect that luminance can have on visual response becomes apparent when one considers the vast range of luminance values that are found for typical visual stimuli. According to Riggs (1965), visual stimuli in the real scene can vary from 10^9 millilamberts (sun's surface at noon) to 10^{-5} millilamberts (white paper in starlight).

Based on the above studies (see Brown and Mueller, 1965; De Valois & De Valois, 1988), the average level of luminance of the real scene and the individual level of luminances for objects in the real scene is an important factor in determining visual responses. Therefore, if pictures are to illicit the same visual response as the scene they represent, it will be important to consider the extent to which pictures are limited in replicating identical luminance levels as the scene and to consider what changes occur in the visual response as a result of this limitation. Moreover, these changes will be of particular importance with those pictures that attempt to represent scenes that possess luminance levels that are far outside the luminance range of the picture (e.g. bright daylight scenes and dark night scenes).

The range of brightnesses in a real scene can potentially vary by a factor of many hundreds to one (Helmholtz, 1881; Hochberg, 1979). However, the maximum luminance range of surface pigments is approximately forty to one (Hochberg, 1979). Other media such as photographs have higher ranges but not to the extent of ranges found in the real scene. (Evans calculated the maximum range of a photograph to be approximately 300 to 1; see Evans, 1959). The inability of these mediums to meet the range of luminances in a given scene is further compounded by their inability to match the overall luminance of the scene (Helmholtz, 1881; Gombrich, 1956). As a result of these restrictions in luminance range and overall luminance level, pictures are limited in their capacity to elicit the same visual response as the scene they represent (Helmholtz, 1881; Evans, 1959; Gombrich, 1956; Pirenne, 1970, 1975; Hochberg, 1978a; Hochberg, 1979; Haber, 1980b; Reed & Jones, 1982; Mills, 1985; Cutting, 1986;).

One consequence of this limitation is that most pictures reflect light at a luminance level that keeps the eye modestly light adapted (Helmholtz, 1881). Therefore, most pictures that attempt to replicate outdoor scenes (especially bright, outdoor, daylight scenes or dark, moonlit, night scenes) will not be viewed with the same adapted state of the eye found during real scene viewing (Helmholtz, 1881; Evans, 1959; Pirenne, 1970; Hochberg, 1979; Mills, 1985). For example, Helmholtz (1881) points out that a painter attempting to paint a white object illuminated by the sun compared to a white object illuminated by the moon will usually need to use a pigment that has approximately the same reflectance for representing both objects. Furthermore, both represented objects will often be viewed under the same light level. Therefore, the adapted state of the eye will be approximately constant across both represented scenes. However, in the real scene, the white object in sunlight is approximately 100 million times brighter than in moonlight (Helmholtz, 1881; Riggs, 1965). During real scene viewing then, the eye will be extremely light adapted in the sunlit scene and extremely dark adapted in the moonlit scene. This difference in behavior between the light adapted and dark adapted eye has been accommodated in the design of photographic films that produce pictures to be viewed in either bright (reflection prints) or dark (transmission slides) environments (Bartelson and Breneman, 1967).

Another consequence of a limited luminance range in pictures is the lack of color assimilation effects occurring under picture viewing (see Graham and Brown, 1965). During real scene viewing, if factors such as contrast, saturation and light level are sufficient, and the spatial frequency of surround and target is high, then the target will tend to appear the same color as the surround (see Graham and Brown, 1965; Goldstein, 1984). For example, a small, dark shadow surrounded by bright yellow sunlight should appear slightly yellowish in the real scene even though the shadow is not projecting long wavelength light. However, a picture that represents bright, yellow sunlight surrounding a dark shadow would not possess as high a luminance range and would therefore be unable to induce as great an assimilation effect (Graham and Brown, 1965).

5.2 Overcoming limited dynamic range

Helmholtz (1881) points out that a painter needs to consider the different physiological conditions of the eye present during real scene viewing (e.g. low visual acuity due to a dark adapted eye) and then "translate" these subjective phenomenon into the painting itself. Gombrich (1956) also agrees that the artist or photographer must attempt to suggest the presence of light (or the lack of it) in the picture by painting in the physiological reactions the observer naturally experiences under real scene viewing. An example of this can be seen in Monet's attempts to mimic the visual response of looking at a church with a light insensitive eye (Monet: Rouen Cathedral, West Facade, Sunlight; see Mills, 1985).

Hochberg (1979) suggests how simultaneous contrast can be used to construct pictures that simulate the perceptual response of the visual system. For example, he points out that the early Impressionists such as Corot painted in color contrast effects in order to simulate the effects of saturated colors on a light adapted eye in a brightly lit scene. Furthermore, he suggests that artists such as Rembrandt, Eakins, and Seurat attempted to offset the limited luminance range in pictures by representing objects that have luminance levels outside this range (e.g. bright, shiny highlights from light reflected off of a gold braid) using large swatches of light and dark in the outside regions of the picture. This was done in order to take advantage of simultaneous contrast effects (large dark regions surrounding bright regions cause the bright regions to appear brighter) that occur with low spatial frequency stimuli in the periphery of the eye.

Hochberg (1979) also proposes that the large swatches of light and dark employed by Rembrandt, Eakins, and Seurat take advantage of successive contrast effects that not only increase the percept of brightness but the percept of saturation as well. He points out that an area of the retina that is stimulated by a dark region in the painting will be somewhat dark adapted. Therefore, when a light region of the painting falls on this dark adapted area as a result of minor eye movements, it will appear brighter due to the somewhat increased sensitivity of that area of the eye (Hochberg, 1979). In the case of increasing saturation, when a colored region stimulates a specific area of the retina, an afterimage is produced that is the complementary color of that region. If another colored region that is the complement of the first region were then to stimulate this same area of retina (as a result of minor eye movements), then the color of this second region should appear more saturated (Hochberg, 1979; Goldstein, 1984).

Photographers have also developed techniques to help them overcome the limited dynamic range available with their medium. Shadows are a particular problem because our visual system is capable of seeing detail in real shadows but a photograph doesn't have sufficient dynamic range to reproduce this detail. Photographers therefore try to flatten the lighting in a scene to eliminate deep shadows. The result is that the lighting on a television or a movie set does not appear correct when viewed on location, but looks correct when seen on a television monitor or in a movie theater.

The same problem exists in computer graphics as is illustrated by the synthetic images in Slides 7 to 9. These pictures were produced using a radiometrically correct illumination model (Ward, Rubinstein, and Clear, 1988) In Slide 7 the rear wall appears much darker than it would in the original scene (if we could be there). Trying to fix this by changing the exposure results in the overexposed picture in Slide 8. The correct solution is to "distort" the original lighting by adding some additional illumination to the back wall so as to create a version that looks correct when observed in the final picture (Slide 9).

6.0 Pictures are seen in a viewing illuminant

The effect of viewing illuminant on brightness adaptation has already been discussed in the preceding section about the limited dynamic range of pictures. In addition to affecting the level of illumination to which the visual system is adjusted, the viewing illuminant also has an impact on the color of the light to which the visual system is adapted. As in the case of brightness adaptation, it is possible for the color of the viewing illuminant to be different between the original scene and the reproduction. This process by which our visual system is able to discount the color of the illuminant and see the true color of objects in the environment is known as chromatic adaptation.

6.1 Chromatic adaptation

When the eye is moderately illuminated by a uniform colored light source the true color of the light will initially be perceived. However in a short time, the eye will adapt to the color and the light will be accepted as white (as long as saturation is not too great; Evans, 1959). This is called chromatic adaptation. Therefore, when an observer views an object that is illuminated by colored light, it will at first appear colored differently than its appearance under white light. However, as the eye adapts to the color of the illuminant, the object color will appear more like that found under white light illumination. For example, Evans (1959) points out that, for the adapted observer, a white piece of paper viewed under yellow incandescent light will appear approximately the same color as an identical piece of paper viewed under white light. However, it should be noted that a trained observer, if asked to do so, can often perceive the color of the illuminant following adaptation (e.g. they can see the yellowishness of a paper illuminated by a yellow incandescent light or the bluishness when illuminated by skylight; Evans, 1959).

Experiments have addressed questions concerning the appearance of achromatic or chromatic stimuli as a function of the wavelength of the illuminant, of the test and background reflectance, and of the size of the background (Graham and Brown, 1965). The general principles of chromatic adaptation for illumination of achromatic stimuli were: "samples (stimuli) of high reflectance had the illuminant hue, samples of intermediate reflectance were achromatic, and samples of low reflectance had the hue of the afterimage complement of the illuminant." Furthermore, "samples above the adaptation reflectance (background reflectance) take the hue of the illuminant color; samples below it, the hue complementary to the illuminant hue; while samples near the adaptation reflectance are either achromatic or greatly reduced in saturation" (see Graham and Brown, 1965). The above principles also hold for a chromatic stimulus. However, it was found that chromatic stimuli have greater constancy than achromatic stimuli. If even a small component of the illuminant contains the dominant wavelength of the chromatic stimulus, the stimulus will tend to appear as it appears under neutral illumination. Furthermore, size of the inducing field also dictates color appearance: induced color change increases with increase in field size (Graham and Brown, 1965).

As a result of chromatic adaptation, objects illuminated with colored light will appear similar to that found under white light illumination. However, other factors discussed above will influence chromatic adaptation affects. These factors will vary depending on the type of scene in which the object is viewed. As a result, the color of the object will often appear different under real scene viewing compared with its appearance under picture viewing. Therefore, to the extent that chromatic adaptation differs between these two types of viewing, pictures will be limited as realistic representations of real scenes (Helmholtz, 1881; Gombrich, 1956).

Evans (1959) agrees that different chromatic adaptation affects are a limitation for pictures and discusses these limitations in terms of photographs. For example, he discusses a situation in which a photograph is taken of two adjacent pieces of paper illuminated by a yellow, incandescent light. One piece is white with high reflectance, the other is grey with low reflectance. Chromatic adaptation to the yellow illuminant would cause the two pieces of paper to lose most of their yellowish appearance. However, the differential reflectance between the two pieces of paper would dictate the grey paper to appear less yellowish than the white paper. A photograph of these two pieces of paper could be made such that the white paper is reproduced as yellowish however the grey would need to be reproduced as more neutral. Evans (1959) points out that this is not ordinarily possible.

According to Evans (1959) "The requirement for satisfactory reproduction of a scene is that the photograph, under the condition of viewing, shows the proper hue and saturation differences from the *adaptation state of the observers.*" Therefore, the hue and saturation of colors in a picture should be dependent, among other things, on the chromatic adapted state of the observer. According to this then, alterations in colors of objects represented in the picture will be needed for each type of viewing condition (at least those that alter the chromatic adapted state of the observer) if one is to maintain color response fidelity.

6.2 Overcoming effects of viewing illuminant

Slides 10 and 11 illustrate the difficulties that can be encountered because of differences between the illumination simulated in the computer generated picture and the illumination under which the synthetic image is viewed. The torus in Slide 10 appears to have a shiny surface while the torus in Slide 11 looks as if it has a matte finish. In actuality, the same coefficients are used in the illumination model governing the amount of diffuse and specular reflection for each of the tori. It is the difference in illumination between the two pictures that causes the change in surface appearance. We are unable to tell that there is only one light source in Slide 10 and four light sources in Slide 11 because our viewing illumination is constant and different from that in either picture. We are also at a disadvantage because none of the illumination from the lights in the picture spills out into the room as it would if the picture plane were a real window.

A solution to the chromatic adaptation problem has been developed in photography. Two different types of film are employed: one for use outdoors in daylight and the other for use indoors with tungsten light. Slides 12 and 13 are a computer graphic simulation of the chromatic adaptation problem and how tungsten balanced film solves the problem for indoor scenes. In Slide 12 a picture has been made by a straightforward application of the laws of color science (Meyer, 1989). Note the very yellow appearance that this image has. In Slide 13 the effect of tungsten balanced film is simulated. The overall yellow color of the image has disappeared, and white objects reproduce as white. Thus the film concentrates on recording the correct visual percept instead of trying to capture an optical identity.

7.0 Summary and Conclusions

While it is possible to synthesize a computer graphic image that is optically identical to a real scene, it is impossible to display this picture given the color reproduction technology that is available today. As has been shown in this article, pictures have limitations in terms of their physical size, their fixed point perspective, their dynamic range, and the circumstances under which they are observed. These limitations restrict the extent to which pictures can reproduce the cues that our visual system uses to interpret three dimensional scenes. These cues include the differences between the images projected onto each retina, the amount by which each eye must be adjusted to focus on the center of attention, and the level of adaptation that is necessary to adjust to the level and color of illumination present in the scene.

Artists and photographers have developed many techniques for overcoming the limitations of the pictorial media. In the absence of binocular cues to depth they have made sure that monocular cues are properly rendered. They have employed view cameras to adjust perspective in a way that limits the effect of the frame surrounding the picture. Composition has been carefully considered to emphasize such things as object occlusion and the effect of perspective on object size. They have intentionally painted color contrasts into pictures in order to overcome the limited dynamic range provided by pigments. Individuals that are involved in the synthesis of realistic images can learn from the techniques developed by artists and photographers. An optically identical representation for a scene cannot be displayed using existing color reproduction technologies. The reproduction must be manipulated and, in many cases, the original scene must be changed so that the correct percept is produced for the observer.

While employing artistic and photographic techniques can improve the quality of today's photorealistic computer graphic images, future image synthesis algorithms should incorporate the characteristics of the human visual system directly into the image creation process. With the flexibility that computer graphics provides it makes no sense to slavishly simulate a camera and continue to reproduce imaging problems that the camera creates or is unable by itself to eliminate. By moving to an initial representation that is perceptual in nature, efficiency can be improved because computational effort is directed toward those things that are perceptually important and device independence can be achieved because the image has not been created for a particular reproduction device. Perhaps most importantly, this approach concentrates attention on the human being for whom the picture is being made and away from the computational physics which has recently dominated computer graphics. This change of focus is critical if we are to eventually synthesize pictures that are not just realistic but also communicate information to people.

8.0 Acknowledgements

Chet Haase wrote the software and created the images shown in Slides 1 and 2. Darren Anderson generated Slides 3 to 13. He used the Radiance software package to produce some of them (Ward, Rubinstein, and Clear, 1988). This work was funded by the National Science Foundation under grant number CCR 90-08445 and by the National Institutes of Health Systems Physiology Training Program grant number GM070257.

9.0 References

Bartelson CJ, Breneman EJ (1967) Brightness reproduction in the photographic process. Photogr. Sci. Eng. 11:254-262

Bartlett N R (1965) Dark adaptation and light adaptation. in: Vision and Visual Perception. Graham C H, Bartlett N R, Brown J L, Hsia Y, Mueller C G, Riggs L A, eds., John Wiley & Sons, Inc., New York

Brown J L, Mueller C G (1965) Brightness discrimination and brightness contrast. in: Vision and Visual Perception. Graham C H, Bartlett N R, Brown J L, Hsia Y, Mueller C G, Riggs L A, eds., John Wiley & Sons, Inc., New York

Chung J C, Harris MR, Brooks FP, Fuchs H, Kelley MT, Hughes J, Ouh-young M, Cheung C, Holloway RL, Pique M (1989) Exploring virtual worlds with head-mounted displays. Three-Dimensional Visualization and Display Technologies, WE Robbins, SS Fisher, Editors, Proc. SPIE 1083, pp 42-52

Cutting J E (1986) Perception with an eye for motion. The MIT Press, pp 1-40

Darley J M, Glucksberg S, Kinchla R A (1986) Psychology. Prentice-Hall Inc., New Jersey, p 86

De Valois R L, De Valois K K (1988) Spatial vision. Oxford University Press, New York, pp 27-31

Evans, R M (1948) An introduction to color. John Wiley and Sons, New York

Evans, R M (1959) Eye, film and camera in color photography. John Wiley and Sons, New York

Foley J D, van Dam A, Feiner S K, Hughes J F (1990) Computer graphics: principles and practice. Addison-Wesley, Massachusetts

Gibson J J (1982) Reasons for realism: selected essays of James J. Gibson. Reed E, Jones R, eds., Lawerence Erlbaum Associates, London

Goldstein B (1984) Sensation and perception. Wadsworth Inc., California, pp 203-287

Gombrich E H (1956) Art and illusion. Princeton Press, New Jersey

Graham C H (1965) Visual space perception. in: Vision and Visual Perception. Graham C H, Bartlett N R, Brown J L, Hsia Y, Mueller C G, Riggs L A, eds., John Wiley & Sons, Inc., New York

Graham C H, Brown J L (1965) Color contrast and color appearances: brightness constancy and color constancy. in: Vision and Visual Perception. Graham C H, Bartlett N R, Brown J L, Hsia Y, Mueller C G, Riggs L A, eds., John Wiley & Sons, Inc., New York

Haber R N (1980a) How we perceive depth from flat pictures. American Scientist 68: 370-380

Haber Ralph (1980b) Perceiving space from pictures: a theoretical analysis. in: The Perception of Pictures. Hagen M, ed., Springer-Verlag, New York, 11

Hall R (1989) Illumination and color in computer generated imagery. Springer-Verlag, New York.

Hatada T, Sakata H, Kusak H (1980) Psychophysical analysis of the "sensation of reality" induced by a visual wide-field display. SMPTE Journal 89: 560-569

Helmholtz H von (1881) On the relation of optics to painting. in: Popular Scientific Lectures. translated by Atkinson E, Appleton, New York

Hochberg J E (1962) The psychophysics of pictorial perception. Audio-Visual Communication Review 10: 22-54.

Hochberg J (1978a) Art and perception. in: Handbook of Perception. Carterette E C, Friedman M P, eds., Academic Press, New York, 10: 225-259.

Hochberg J E (1978b) Perception. 2d ed. Prentice Hall, New York

Hochberg J (1979) Some of the things that paintings are. in: Perception and Pictorial Representation. Nodine C F, Fisher D F, eds., Praeger, New York

Hochberg J (1980) Pictorial functions and perceptual structures. in: The Perception of Pictures. Hagen M, ed., Springer-Verlag, New York, 12

Meinel E (1973) Peripheral vision and painting. British Journal of Asthetics 13 (3): 287-297

Meyer G W, Rushmeier H E, Cohen M F, Greenberg D P, Torrance K E (1986) An experimental evaluation of computer graphics imagery. ACM Transactions on Graphics 5: 30-50

Meyer G W (1989) Reproducing and synthesizing colour in computer graphics. Displays: Technology and Applications 10: 161-170

Mills M I (1985) Image synthesis, optical identity or pictorial communication. in: Computer Generated Images, The State of the Art. Magnenat-Thalmann N, Thalmann D, eds., Springer-Verlag, New York

Mueller C G, Lloyd V V (1948) Stereoscope acuity for various levels of illumination. Proc. Nat. Acad 34: 223-227

Pirenne M H (1970) Optics, panting and photography. Cambridge University Press, London

Pirenne M H (1975) Vision and art. in: Handbook of Perception. Carterette E C, Friedman M P, eds., Academic Press, New York, 5

Riggs L A (1965) Light as a stimulus for vision. in: Vision and Visual Perception. Graham C H, Bartlett N R, Brown J L, Hsia Y, Mueller C G, Riggs L A, eds., John Wiley & Sons, Inc., New York

Rock Irvin (1975) An introduction to perception. Macmillan Publishing Inc, New York, pp 79-153

Sedgwick H (1980) The geometry of spatial layout in pictorial representation. in: The Perception of Pictures. Hagen M. ed., Springer-Verlag, New York, 2

Smith O W, Gruber H (1958) Perception of depth in photographs. American Journal of Psychology 8: 307-313

Stone J (1987) A user's guide to the view camera. Little, Brown & Co., Boston

Swenson H A (1932) The relative influence of accommodation and convergence in the judgement of distance. Journal of General Psychology 7: 360-380

Upton B L, Upton J (1981) Photography. Little, Brown & Co., Boston

Van de Grind W A (1986) Vision and the graphical simulation of spatial structure. Proceedings of the 1986 ACM Workshop on Interactive 3D Graphics. ACM, New York, pp 197-235

Westerink J H D M, Roufs J A J (1989) Subjective image quality as a function of viewing distance, resolution, and picture size. SMPTE Journal 98 (2): 113-119

Witkin H A (1959) The perception of the upright. Scientific American 200 (2): 50-56