The Perception of Depth in Photographic Images

Cathleen Daniels Cerosaletti, Michael E. Miller, and Richard A. Drexel
Eastman Kodak Company
Rochester, New York

Abstract

There is an abundance of literature in the domain of pictorial art that discusses two-dimensional cues to depth; however, this literature does not address how pictorial cues influence the perceived depth of real-world scenes rendered onto a photographic image. To this end, we conducted two studies to better understand the perception of depth in photographic prints. In Experiment 1, we asked a group of naïve observers to view a number of consumer photographs and judge the amount of depth in the main subject and the overall scene. In Experiment 2, naïve observers were asked to compare a number of consumer-like stereoscopic images (print and slide) to photographs and rate the comparative amount of depth per scene. The results of these studies indicate that there are specific photographic image attributes that drive the perception of depth in photographs.

Introduction

In the human visual system, there are a few processes that are critical to the perception of the world. Through the functions of detecting form, motion, color, contrast, and depth, humans are able to identify and react to objects within the visual environment. The detection of motion, color, and contrast, as applied to images, has been studied quite extensively. However, there is little research with regard to the perception of objects and depth in still photographic images.[1-7] Previous research has concentrated on the distortion of scene-distance perception as a result of imaging systems. However, there is some psychological research that explores the connection between pictorial art and visual perception.[8-11] This body of literature addresses the cues that are used by pictorial artists to provide a greater feeling of depth or space within an image. These techniques are used quite extensively in Renaissance art. It is reasonable to believe that the visual techniques used to make pictorial art effective may also be used to understand depth in photographs.

An especially interesting problem is to understand how a photographer captures the image and how scene-depth cues are used to imagine the appearance of a pictorial image during capture. When photographers view the original scene prior to capturing an image, they are using primary depth cues. These primary cues, such as accommodation, vergence, and disparity are used to assess depth and distance from physical characteristics of the environment. During image capture, the viewfinder or digital camera display, reduces perceived depth by enclosing the scene in a frame that tends to reduce the impact of primary depth cues.[12-13] Finally, the same scene is viewed as a two-dimensional representation in the form of a photograph, and depth is inferred through secondary depth cues. The secondary depth or pictorial cues are typically identified as: occlusion, convergence of parallel lines (or obviousness of the vanishing point), shading and shadows (direction and amount), texture gradients (relating to perspective), and atmospheric perspective, among others. Cognitive mechanisms contribute to the mind's representation of the original scene or the remembrance of primary depth cues. Thus, when looking at a photograph, we simultaneously perceive the representation of depth and flatness.[14]

In our research, we hoped to uncover secondary depth cues in photographic images that differentiate scenes in their ability to carry depth information. To better understand depth in photographic images, we conducted two studies. The first study explored the perception of depth in photographic prints, and the second study examined comparative depth in prints and other stereo viewing modes.

Experiment 1: Attributes of Photographic Perceived Depth

The intent of this first study was to empirically determine the image attributes that contribute most significantly to the perception of depth in photographs

Methodology

Observers

Thirty observers participated in this study. All observers were enthusiastic consumer photographers and had experience in both traditional and digital photography.

Scenes

Seventy-eight (78) scenes were gathered from a number of sources to include a variety of subject matter. When selecting images, careful attention was given to balance the following scene attributes:

- Camera angle relative to the main subject.
- Camera-to-subject distance.
- People and non-people scenes.
- Indoor and outdoor scenes.

Experimental Design

Scene content was the primary, independent variable that was not varied systematically, but the image set was balanced with the factors listed above. All images were viewed under D5000 lighting on 20% gray paper.

Observer Evaluations

Initially, the observers were screened for visual acuity, color vision, and depth perception. Next, the observers rated 78 prints for "overall image depth" on a 100-point scale bi-anchored at the extreme ends with "low perceived depth" and "high perceived depth." The same procedure was followed when rating "main subject depth." In the study protocol, depth was defined as follows:

> "We are interested in understanding the aspects of an image that communicate perceived depth. Although an image is only a representation of a real three-dimensional scene, it is possible for it to communicate the amount of depth in the original scene. That is, the size and shape of objects in the image as well as the distance between objects may give the impression of depth. Also, the image may NOT communicate these attributes and the image may look flat."

Image Evaluations

All of the images were evaluated for the following variables:

- Camera-to-subject distance (in feet)
- Depth of field (short, intermediate, infinite)
- Location (indoors and outdoors)
- 3-Layer (assessment of the presence of three horizontal layers of sky, mid-ground, and foreground)
- Main subject vs Landscape (assessment of the presence of a clear main subject or a landscape image)
- Repeated objects in perspective (as objects of similar size receding in space)

Results

Comparison of Overall and Main Subject Depth Ratings

The distributions of overall and main subject depth ratings were quite similar (see Fig. 1). In addition, the main subject was predicted reasonably well by the overall image-depth rating (see Fig. 2). The correlation between mean main subject depth and overall image depth was 0.85. The data were skewed to indicate that there is greater perceived depth in the overall scene than the main subject. In fact, in some scenes, such as landscapes in which there is no specific main subject; the entire scene is the main subject of interest. Furthermore, there is evidence to suggest that, when an observer looks at an image containing a clear main subject, they will attend to the main subject and evaluate the image, based on that subject. However, if the image is a scenic landscape, the observer is more likely to evaluate the entire scene. These observations lead us to reason that the overall and main subject ratings are redundant. In addition, the overall depth ratings may be a stronger predictor of pictorial depth than the depth of the main subject. Therefore, we will limit our analyses to overall depth ratings.

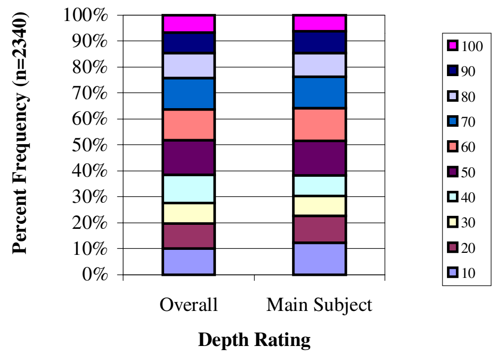

Figure 1. Percent frequency as a function of overall and main subject depth ratings. The legend corresponds top to bottom to information in the plot.

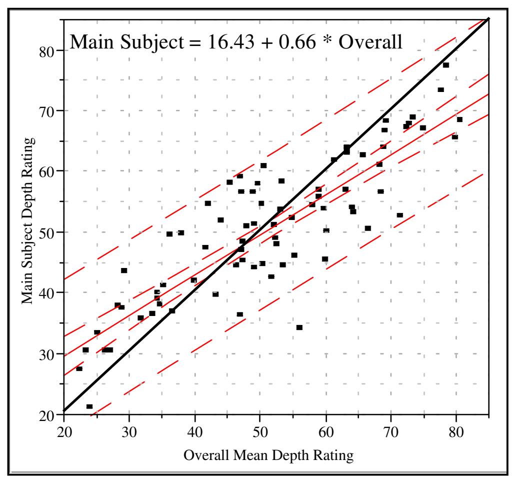

Figure 2. Mean, main subject depth rating as a function of mean, overall depth rating with linear fitted line. The inner dashed lines are the confidence curves for the fitted line. The outer dashed lines are the 95% confidence intervals for the predicted individual values. A 1:1 line is indicated in solid black.

Principle Component and Cluster Analyses

Combined analyses of principal component and cluster analyses were used. Principal component analysis was completed on the correlation structure of the variables to reduce the dimensionality of the set of data so that the data could be represented by as few variables as possible. Cluster analysis is a multivariate technique of grouping items that share similar values. In this case, scenes were grouped according to patterns of responses by observers. Cluster analysis was performed using the hierarchical Ward method on standardized data.

All of the image evaluation variables were projected onto the principal component and cluster values in an effort to understand the characteristics driving each cluster and principal component. These variables were combined with subjective interpretation to derive the attributes that will be described below. For mean, overall depth ratings, the contribution of the third highest order principle component is small; thus, it is most efficient to limit our initial analyses. In these analyses, we will therefore, limit our explanation to the first two principle components. These components account for 58.22% of the variation in the overall depth ratings over observers. The first principal component corresponds to depth rating, such that positive values along the first principle component axis relate to higher overall depth ratings, and negative values relate to lower overall depth ratings. The clusters, as distributed across the first principle component axis, also relate to magnitude of depth rating (see Figs. 3 and 4). Images with high depth ratings seem to have higher camera-to-subject distances, perspective cues, high depth of field, and consist of landscape-type images. Images with lower depth ratings seem to have low camera-to-subject distance, short depth of field, and a close-up, well-defined subject. The second principle component does not appear to be strongly related to a statistic of overall depth ratings. However, there does seem to be some relationship between clusters and polarity of principle component two. To further explore the meaning of the clusters, each cluster was characterized by specific image attributes as well as by mean depth rating per cluster (see Table 1).

Table 1. Cluster number, number of scenes, and attributes by mean overall depth rating.

Mean Overall Depth Rating	Cluster Number	Number of Scenes	Attributes
28.07	3	10	Contrast and framing problems and close-up, well-defined subject: nature and people
34.07	6	9	Close-up, well-defined subject: nature and people
45.58	1	18	Close-up, well-defined subject: nature and people
52.81	5	19	Intermediate distance, people, and objects repeated in depth (head size in perspective)
54.89	2	12	Scenes with 3-layers: foreground, horizon, and sky
70.46	4	10	Perspective cues with many depth planes

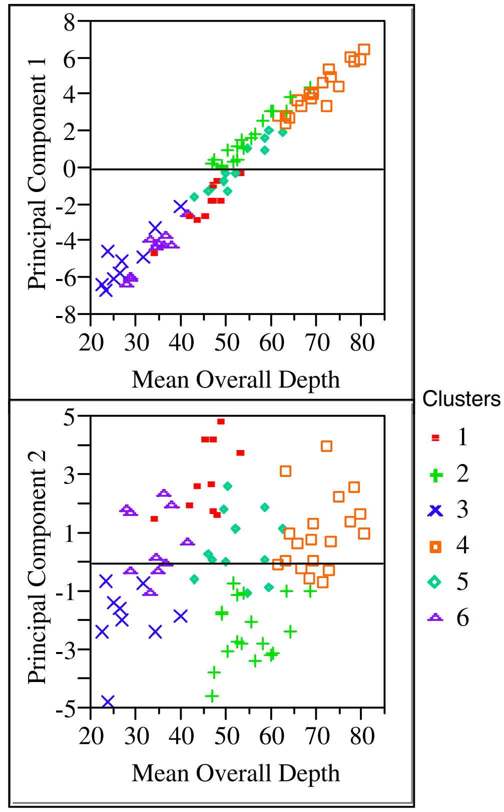

Figure 3. Principal components 1 and 2 as a function of mean overall depth rating and cluster.

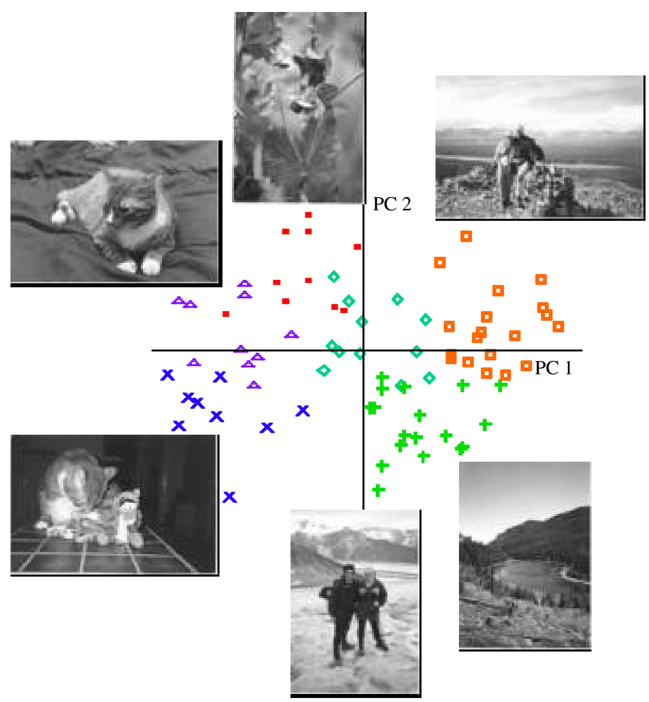

Figure 4. Principal components a) 1 and b) 2 as a function of mean overall depth rating and cluster.

Experiment 2: Comparative Perceived Depth in Prints and Stereo Slides and Stereo Prints

Background

Experiment 1 was an exploration of the factors that contribute to the perception of depth in two-dimensional photographic scenes. In Experiment 2, we attempted to gain further insight into the attributes that predict depth through comparing images with virtual three-dimensional content to two-dimensional photographic images.

Methodology

Observers

Eighteen (18) observers participated in this study and were selected with the same characteristics as observers in Experiment 1.

Scenes

The study scene set consisted of 15 consumer-like scenes. All images were captured with the RBT Stereo Camera on Kodak Professional reversal film E100S. The lens separation of the camera was 6.4 cm. All 15 scenes were captured outdoors as a result of the light-limiting nature of the dual lens of the camera and the inability to properly synchronize the flash for stereo capture. The scenes were all captured with a 35 mm focal length lens. The same criteria as in Experiment 1 were used for selecting scene content.

Image Processing

All 15 scenes were scanned using the Nikon LS4000 at 2000 dpi. The scenes were reproduced as a single print, a print stereo pair, and a slide stereo pair. Single prints (3.8 x 5.4 inch) were printed on a digital photographic printer at 250 dpi. An ICC-profile-managed system was used to obtain good quality prints from slides. One of the images in each stereo pair was chosen arbitrarily for the single print versions. Print stereo pairs (2.7 x 2.0 inch) were created in Kodak Photoshop software by montaging the scanned image pairs into a 20% gray background (3.3 x 7.0 inch) at a 0.3125-inch separation. Slide stereo pairs were mounted into Gape 24 x 32 mm slides using the near disparity rule.

Viewing Apparatus

Single prints were viewed under D5000 lighting on 20% gray paper. The stereo slides were viewed using a stereo slide viewer, which observers directed at a light source, and the stereo prints were viewed using an antique stereo viewer that was produced for the 1900 World's Fair.

Observer Evaluations

Initially, the observers were screened for visual acuity, color vision, and depth perception. In this protocol, depth was defined in an identical manner to Experiment 1. The photographic print was compared to stereo card and stereo slide viewing modes by rating each scene for all 3 viewing modes on the 100-point scale used in Experiment 1.

Image Evaluations

All of images were evaluated for the same variables as described in Experiment 1.

Results

Significant effects of scene ($F(19,323) = 5.64$, $p < 0.0001$), mode ($F(2,34) = 90.61$, $p < 0.0001$.), and a 2-way scene by viewing mode interaction ($F(38,646) = 2.57$, $p < 0.0001$). We will focus on the 2-way scene by mode interaction by reporting the results of the cluster analysis. For all scenes, the stereo slide was rated highest, the stereo print was rated intermediately, and the print was rated the lowest for depth. However, the relationship of the ratings differed by scene. Again, all of the image evaluation variables were projected onto the principal component and cluster values, in an effort to understand the characteristics driving each cluster and principal component. These variables were combined with subjective interpretation to derive the attributes that will be described below. The images were defined by the characteristics in Table 2.

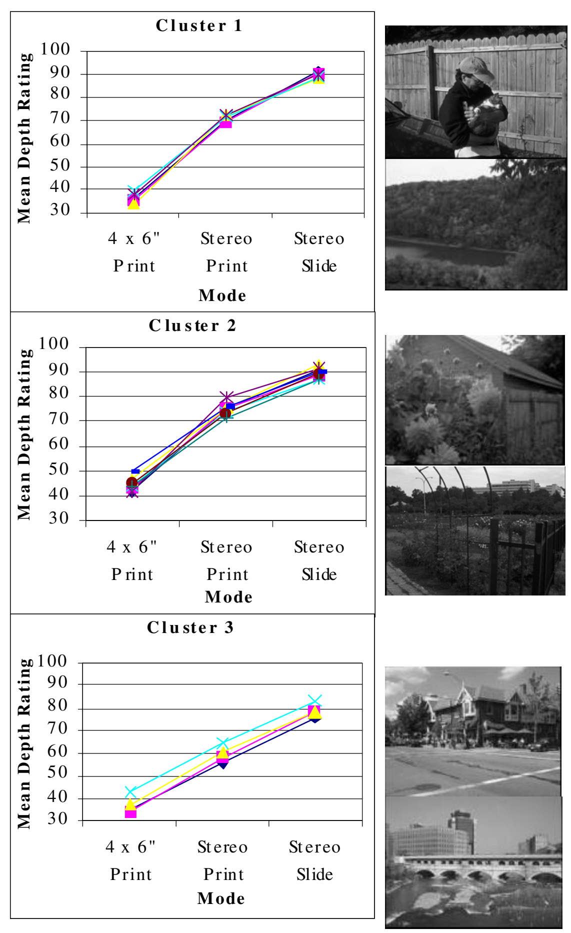

Figure 5. Clusters 1-3 for mean, depth rating as a function of viewing mode.

Table 2. Cluster number, number of scenes, and attributes by mean overall depth rating for single print, stereo print, and stereo slide.

Cluster	Scenes	Single Print	Stereo Print	Stereo Slide	Attributes
3	3	35.59	58.19	77.69	Large foreground leads to lower depth
1	5	36.71	71.34	89.82	Hidden depth in stereo
2	8	44.26	74.74	89.53	Perspective cues

Looking at Fig. 5 and Table 2, one can see that the addition of disparity cues produces a dramatic increase in perceived depth for clusters 1 and 2, while the enhanced contrast, sharpness, and other pictorial cues that occur as the scenes are rendered to slides, produce a less dramatic change. However, for cluster 3, the change in perceived depth caused by the addition of disparity information is approximately equal to the enhancement in perceived depth that occurs with the enhancement of pictorial cues, which occurs as the images are rendered to slides. This finding might have been expected because cluster 3 contains scenes with large foregrounds and primary scene elements at larger distances where disparity provides less information.

Conclusions

Experiment 1 clearly shows that there are image attributes unique to photography that impact the perception of depth in images. These attributes are not typical to non-illustrative artistic depiction. They are: head size in perspective, presence of a clear, main subject vs a scenic landscape, and contrast and framing problems. Furthermore, from Experiment 2, the visual cue that greatly differentiates prints from stereo images is hidden depth. Hidden depth occurs when information is lost from an original scene and objects of nearly equal color and contrast are superimposed in depth. These findings may give us insight into image rendering techniques that can add a greater degree of realism to photographs.

References

1. A. Ames, The Illusion of Depth from Single Pictures, J. Opt. Soc. Am., 10, 137 (1925).
2. H. Schlosberg, Stereoscopic Depth from Single Pictures, Am. J. Psychol., 54, 601(1941).
3. H. W. Leibowitz, T. Bussey, and P. McGuire, Shape and Size Constancy in Photographic Reproductions, J. Opt. Soc. Am., 47, 658 (1957).
4. O. W. Smith, H. Gruber, Perception of Depth in Photographs, Percept. Motor Skills, 8, 307 (1958).
5. R. N. Kraft, J. F. Patterson, N. B. Mitchell, Distance Perception in Photographic Displays of Natural Settings, Percept. Motor Skills, 62, 179 (1986).
6. M. A. Pirenne, Optics, Painting, and Photography, Cambridge University Press, London, (1970).
7. P. Petzold, Effects and Experiments in Photography, Focal Press, Burlington, MA, (1973).
8. E. H. Gombrich, Art and Illusion: A Study in the Psychology of Pictorial Representation, Princeton University Press, Princeton, (1969).
9. R. Arnheim, Art and Visual Perception, University of California Press, Berkeley, CA, (1974).
10. M.A. Hagen, Ed. The Perception of Pictures, Volumes 1 and 2, Academic Press, NY, 1980.
11. M. Kubovy, The Psychology of Perspective and Renaissance Art, Cambridge University Press, NY, (1986).
12. J. W. Meehan, Apparent Minification in an Imaging Display Under Reduced Viewing Conditions, Perception, 22, 1075 (1993).
13. D. W. Eby and M. L. Braunstein, The Perceptual Flattening of Three-Dimensional Scenes Enclosed by a Frame, Perception, 24, 981 (1995)
14. J. E. Cutting, How the Eye Measures Reality and Virtual Reality, Behav. Res. Meth. Ins. C., 29, 27 (1997).

Biographies

Cathleen Daniels Cerosaletti is a Senior Research Scientist in the Image Quality Group, Electronic Imaging Products, R&D, at Eastman Kodak Company. Cathy is a graduate of Cornell University and holds an MS in Experimental Psychology from Penn State, as well as an MS in Color Science from the Center for Imaging Science at Rochester Institute of Technology. She has interests in image perception and quality, color appearance, and statistical data analysis.

Michael E. Miller leads a team of psychologists and engineers who apply their knowledge of human visual perception and cognition to improve imaging systems. Michael’s education includes a BS and MS in Industrial and Systems Engineering from Ohio University and a Ph.D. in Industrial in Systems Engineering from Virginia Tech. His current research explores human perception of organic light-emitting displays and large field-of-view projection displays.

Richard A. Drexel is an Image Quality Technician with over ten years experience in testing and development of imaging systems. Formal education includes an AAS in Quality Control Technology and Certificate of Completion in Optical Systems Technology from Monroe Community College. Current research projects involve developing a better understanding of digital photography.