INTEGRATION OF MONOCULAR CUES TO CREATE DEPTH EFFECT

Cassandra T. Swain

AT&T Labs
101 Crawfords Corner Road, Room 4B505
Holmdel, NJ 07733
swain@research.att.com

ABSTRACT

This paper presents a novel approach for using monocular cues in a *single* 2D image to improve depth perception. Monocular depth cues–blur, shading, brightness, and occlusion–are applied to 2D images. The contribution of the first three cues to depth perception is additive, each weight being equivalent. Since occlusion cues modify the object geometry, they are applied after the application of other cues. Results show that monocular depth cues can successfully improve depth perception in a single 2D image, creating a *pseudo* 3D image. The advantage of this approach is that it requires a single image, rather than an image pair used in traditional methods. The main limitation is that only depth perception, not precise depth measurements, is possible. This work looks very promising for low bitrate video coding and other applications where bandwidth is limited.

1. INTRODUCTION

Traditionally, the ability to perceive depth from two-dimensional images has been accomplished by binocular methods [1, 2]. Binocular depth cues, such as disparity, have been used. This required using multiple images and matching corresponding points, a computationally complex task. More recently, researchers have developed monocular approaches [3, 4]. Monocular depth cues, such as blur, have been used to perceive depth. In these approaches, point matching was not required; however, multiple images were required for depth perception.

Researchers have begun looking at integration of binocular and monocular cues. Several researchers have studied the result of combining cues to perceive depth from 2D images [5, 6]. Others have studied how these different cues interact in creating a depth effect [7, 8, 9]. In these studies, binocular cues have been considered large contributors to depth perception in 2D images. However, as mentioned previously, multiple images are required to extract these cues.

The question then becomes: can depth be perceived from monocular cues alone and in a single image. The advantage of such an approach is that it requires only a single image rather than an image pair to extract cues.

This paper presents a novel approach for depth perception in 2D images from monocular cues *only*. Monocular cues are extracted from a *single* 2D image and used to create a depth effect. The result is a *pseudo* 3D image, with improved depth perception over the original 2D image. This result is promising for applications, such as low bitrate video coding, where bandwidth is limited. Section 2 describes the depth cues. Section 3 describes the algorithm. Section 4 shows experimental results. Conclusions and future work are presented in Section 5.

2. DISCUSSION

Depth cues are used to perceive depth and dimensionality. The objective of this work is to extract and analyze depth cues in 2D images, then combine these cues to create a 3D effect. The four monocular depth cues used are blur, occlusion, shading, and brightness.

2.1. Blur

Blur is a measure of the loss of detail in an image. The principle of depth from blur is a simple one. When an object in a scene is in focus, other objects are blurred. The amount of this blur depends on the distance from the focused object. A method was developed to measure this amount of blur to get depth [3].

A blurred image is the response of the camera to a single point source. This response is called the camera point spread function (PSF), $h(x, y)$, which is approximated as a Gaussian [3]

$$h(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}}. \quad (1)$$

Copyright 1997 IEEE

A blurred image $g(x, y)$ is generated by convolving a focused image $f(x, y)$ with the PSF $h(x, y)$.

$$g(x, y) = f(x, y) * h(x, y). \quad (2)$$

Our algorithm blurs an image to exaggerate the depth of objects in the image. Simply, the larger the variance of the Gaussian (a lowpass filter), the more blurred the image. So, very different variances are applied to images to mimic a 3D effect.

2.2. Occlusion

Occlusion indicates that the *occluded* object is farther away than the *occluding* object. Our algorithm generates the occlusion effect by enlarging an object to appear closer and to occlude any remaining objects in the scene.

2.3. Shading

Another effect similar to the occlusion effect is shading, where either an object casts a shadow over another object or some regions of an object are in shadow and other regions are not. Our algorithm addresses the former case. The shading effect is generated by extracting the shape of an object, darkening it to grey like a shadow, and then applying it to background objects.

One of the concerns is how much of the shading will appear. This means how much distance between objects is the shading meant to imply. Our answer is as much as possible to improve depth perception.

2.4. Brightness

In the brightness cue, closer objects appear brighter than more distant objects. Our algorithm exaggerates the brightness of the closer objects by increasing their intensities.

3. METHODOLOGY

This approach contains two main steps–depth modification and combination. The original image has been segmented into objects. The inputs to our algorithm are the segmentation map and the original image. The segmentation map is necessary for cues that apply to specific objects, rather than the entire image, e.g. occlusion and shading.

Modifications are made to the original image based on monocular depth cues. A blur image is created when blur is applied to the original image. An occlusion image is created by enlarging the foreground object to appear closer and to occlude even more of other objects. A shadow in the shape of the foreground object is cast onto background objects to create a shading image. A brightness image is generated by brightening foreground objects in the original image.

Next, the depth-enhanced images are combined to create a depth effect. The resulting image shows increased depth perception over the original image. Blur, shading, and brightness are combined by addition, where each cue has equal weight. Occlusion cues are applied later to the added result because they modify object geometry.

Two assumptions are made in this approach. First, the input image has been segmented into objects. There is *a priori* knowledge of the image, i.e. which objects are in the foreground and background and which objects occlude other objects. Second, foreground objects are focused and background objects are blurred.

4. EXPERIMENTAL RESULTS

Figure 1 shows examples of three monocular depth cues–blur, shading, and brightness–in a test image. In 1(a), the original image is shown, followed by images affected by blur, shading, and brightness (1(b)-(d)). It should be noted that each cue alone does not render a sufficient 3D effect. Therefore, it is necessary to combine them.

In this example, the resulting images are added to produce the pseudo 3D image. The weights from each effect are equivalent, such that the combined result is an average. Figure 2 shows the result, where the depth perception is improved. Here, the center of the face is perceived to be closer than the ears, which appear closer than the wall.

Figure 3 shows another example of two cues–blur and occlusion–on a test image. In 3(a), the original image is shown, followed by images affected by blur and occlusion(3(b)-(c)). Here, the original image is blurred, followed by enlargement of the foreground object to occlude more of the background. Figure 3(d) shows the result, where the depth perception is improved. Here, the edge of the box appears to be closer and the background object farther away than in the original image.

5. CONCLUSION

This paper presents a new approach to enhance depth perception in a single image using monocular depth cues. Each depth cue was enhanced and/or modified in the original image, then combined to create an image with improved depth perception.

The advantage of this approach is its use of single images to create the depth effect. This is very at-

Copyright 1997 IEEE

(a) (b) (c) (d)

Figure 1: Depth cues. (a) original, (b) blur, (c) shading, (d) brightness.

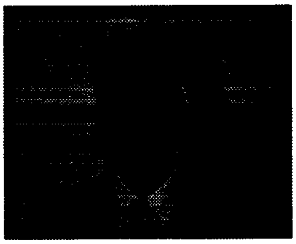
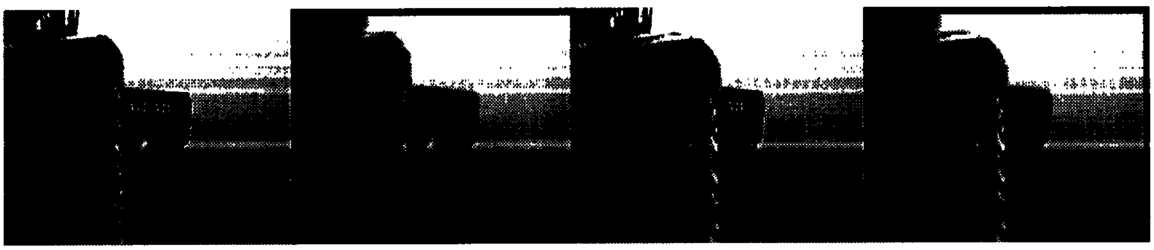


Figure 2: Image with improved depth perception.



(a) (b) (c) (d)

Figure 3: Depth cues. (a) original, (b) blur, (c) occlusion, (d) enhanced depth result.

Copyright 1997 IEEE

tractive for MPEG4 in low bitrate video coding and in video synthesis.

The main limitation is that results yield qualitative depth perception, rather than quantitative depth measurements. In addition, *a priori* knowledge of the scene is necessary.

Future work will address some areas of improvement. Specifically, the integration of cues is additive and successive, which may not be the best ways to combine cues. We will examine different types of integration and the effect on results. In addition, we will look at possible real-time enhancement. We will add perspective as a cue. And, most importantly, we will explore ways to use this method in acquiring depth measurements.

6. REFERENCES

[1] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.

[2] S. Barnard and M. Fischler. Computational stereo. *Computing Surveys*, 14(4):553–572, December 1982.

[3] A. Pentland. A new sense for depth of field. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(4):523–531, July 1987.

[4] G. Surya and M. Subbarao. Depth from defocus by changing camera aperture: a spatial domain approach. *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 61–67, 1993.

[5] N. Gershon. Visualizing 3D PET images. *IEEE Computer Graphics and Applications*, 11(3):11–13, 1991.

[6] S. Marapane and M. Trivedi. An active vision system for depth extraction using multi-primitive hierarchical stereo analysis and multiple depth cues. *SPIE*, 1956:250–262, 1993.

[7] R. Surdick, E. Davis, R. King, G. Corso, A. Shapiro, L. Hodges, and K. Eliot. Relevant cues for the visual perception of depth: is where you see it where it is? *Proc. 38th Mtg. Human Factors and Ergonomics Society*, pages 1305–1309, 1994.

[8] R. Srinivasan, M. Shridhar, and M. Ahmadi. Computing surface information from monocular and binocular cues for vision applications. *Proc. 27th Conf. Decision and Control*, pages 1085–1089, 1988.

[9] S. Das and N. Ahuja. Dynamic integration of visual cues for position estimation. *SPIE*, 1382:341–352, 1990.

Copyright 1997 IEEE