Perceptual Evaluation of Cardboarding in 3D Content Visualization

Alexandre Chapiro[1,2]*, Olga Diamanti[2]†, Steven Poulakos[1,2]‡, Carol O'Sullivan[3]§, Aljoscha Smolic[1]¶, Markus Gross[1,2]‖
[1] Disney Research Zurich, [2] ETH Zurich, [3] Disney Research Los Angeles

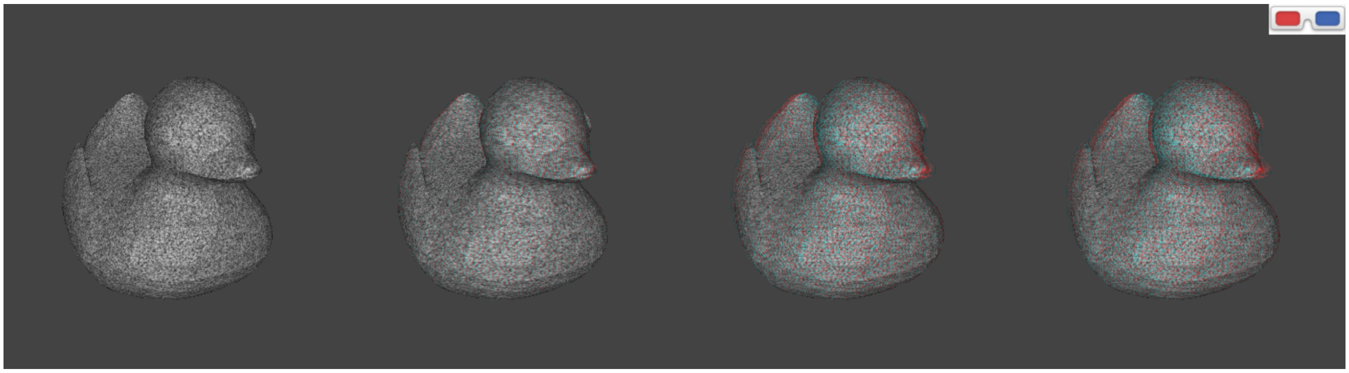

Figure 1: *The cardboarding effect is illustrated in these anaglyphs, with depth compression levels of* $\alpha = 0.0$ *(completely flat),* $\alpha = 0.2$*,* $\alpha = 0.8$*, and* $\alpha = 1.0$ *(fully 3D). In our studies, we found that differences between the left three images were detected significantly often, whereas the right two appeared to be the same and equally acceptable to our participants.*

Abstract

A pervasive artifact that occurs when visualizing 3D content is the so-called "cardboarding" effect, where objects appear flat due to depth compression, with relatively little research conducted to perceptually quantify its effects. Our aim is to shed light on the subjective preferences and practical perceptual limits of stereo vision with respect to cardboarding. We present three experiments that explore the consequences of displaying simple scenes with reduced depths using both subjective ratings and adjustments and objective sensitivity metrics. Our results suggest that compressing depth to 80% or above is likely to be acceptable, whereas sensitivity to the cardboarding artifact below 30% is very high. These values could be used in practice as guidelines for commonplace depth mapping operations in 3D production pipelines.

CR Categories: I.3.1 [Computer Graphics]: Hardware Architecture—Three-dimensional displays I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Stereo;

Keywords: Non-photorealistic rendering, Stereo, Stereo 3D

1 Introduction

Creating high-quality 3D content is a challenging task, with many efforts in academia and industry directed towards the development of an effective pipeline for 3D content production and delivery. Unlike with regular displays, 3D viewing can often be physically uncomfortable when unsuitable depth volumes are displayed. To avoid this discomfort, depth limitations on displayed content for comfortable watching have been determined [Shibata et al. 2011].

However, these inherent limitations of 3D-capable displays in showing depth are not uniform, and may change to a large degree depending on the technology used (for example, auto-stereo screens have much smaller ranges than most displays that use glasses). From a content production perspective, this means that content depths must often be adapted before they can be displayed. When voluminous objects are shown with a reduced depth profile, such as one that could result from depth re-mapping to suit a display's capabilities, the reduced depth profile appears unnaturally flat and results in a disturbing perception of the scene geometry known as "cardboarding" (see Figure 2). Although this perceptual artifact is very common, it has been relatively unexplored.

It follows that when 3D content is compressed in depth, cardboarding should be avoided if possible. Since the effect has not yet been fully explored in the research literature, content creators struggle to make well-informed decisions when implementing mapping methods and must follow heuristic solutions or adjust content manually. The contribution of our paper is a perceptual exploration of preferences and thresholds for cardboarding effects in simple scenes, which can be applied as guidelines to improve existing methods in depth re-mapping. We present three experiments, the results of which each painted a consistent picture of the effects of cardboarding on four models. This methodology can now be used to explore the cardboarding effect further in more complex scenes.

In this text we will present some 3D examples using anaglyph. Such figures are marked with this icon 👓. They can be viewed in 3D using anaglyph glasses (red - left, cyan - right). Please note that to get a better depth perspective you can zoom in on the figures.

---

* e-mail:alexandre.chapiro@disneyresearch.com
† e-mail:olga.diamanti@inf.ethz.ch
‡ e-mail:psteven@disneyresearch.com
§ email:Carol.A.OSullivan@disney.com
¶ email:smolic@disneyresearch.com
‖ grossm@disneyresearch.com

Figure 2: *This anaglyph image showcases cardboarding. The top image has a starkly reduced depth profile, resulting in an unnatural perception of the scene's geometry. The bottom image has a more natural depth profile, and is provided for reference. Notice how the perception of the size of the room changes when looking at the back wall.*

2 Related Work

The cardboarding effect, along with other stereoscopic distortions, is believed to influence both perceived image quality and visual comfort [Meesters et al. 2004; Lambooij et al. 2009]. One factor influencing the perception of cardboarding is the mismatch between perception of object size and object disparity with distance. Howard and Rogers [2002] point out that size sensitivity is inversely proportional to distance, while disparity sensitivity is inversely proportional to the squared distance. This results in a conflict between size and depth scaling.

Another significant factor that influences cardboarding is a geometric mismatch between the stereoscopic capture, display and viewing conditions. These geometric relationships have been well studied [Woods et al. 1993; Jones et al. 2001; Masaoka et al. 2006; Yamanoue et al. 2006; Zilly et al. 2011]. Masaoka et al. [2006] sought to develop a spatial distortion prediction system to determine the extent of the stereoscopic cardboarding effect. However, they developed geometric relations without taking the subjective perception of the artifact into account.

Yamanoue et al [2000] experimentally evaluated perceived cardboarding by exploring several factors including lighting and variation of spatial thickness. They observed a significant effect of spatial thickness in the subjective rating of perceived cardboarding. Only one object with three spatial thickness values was evaluated, thereby making it difficult to draw more general conclusions regarding fine-scale changes in spatial thickness. Yamanoue et al. [2006] later modeled the cardboarding effect as the ratio of size and depth magnification. They observed a good correlation with their previous experimental observations from one object [2000]. Our aim is to further build on this work by introducing more objects and to rigorously observe the effects of cardboarding effect using several experimental paradigms.

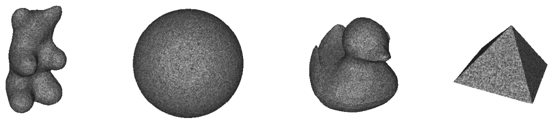

Figure 3: *This figure shows a monoscopic view of the stereo scenes used in the experiment described in section 3.1. The meshes shown were displayed with varying depth profiles on a solid gray background.*

With the goal of staying well within the zone of comfort [Shibata et al. 2011], Siegel and Nagata [2000] proposed the concept of microstereopsis, in which small interocular separation is combined with alignment of interesting content about the zero parallax plane. Their informal experiments demonstrated sensitivity to small disparities and they hypothesize that minimal detectable disparity is sufficient when combined with other visual cues for depth. Didyk et al [2011; 2012b] formulated depth discrimination thresholds and demonstrated an application of minimal stereopsis.

Finally, depth adaptation is often necessary for various applications, with a range reduction being the standard. This means that cardboarding is a significant concern in practice. Previous work such as [2012a; 2012b] re-map depths based on models that take depth perception into account. They do not, however, target cardboarding specifically. Work in [Chapiro et al. 2014] and [Lang et al. 2010] interpret a scene and re-target depth based on the importance of different areas. While [Chapiro et al. 2014] is aimed specifically at avoiding cardboarding when generating content for auto-stereo displays (that have a particularly small depth budget), no quantitative characterization of the effect is provided. These mapping operations could therefore benefit from a better understanding of cardboarding.

3 Experiments

We conducted three perceptual experiments in order to explore the effects of cardboarding. In the *Dial* experiment, we aimed to determine whether preferences for the appearance of stereo scenes could be self-selected by our participants. We found that this was a difficult task, with much variation in the quality levels selected, even for a single participant. However, the flatness was almost never disturbing above 80% and nearly always noticed below 30%. We followed up with a *Pairs* study, to determine whether this wide range of preferences was due to a lack of sensitivity to the cardboarding artifact. We found that participants were relatively efficient at detecting differences between more flattened images, but less sensitive the fuller i.e., more 3D, the images became. Finally, we ran a subjective *Ratings* experiment, and found that the results were consistent with the previous two studies. In particular, we found that both objective sensitivity performance and subjective preference rating indicate a lack of sensitivity and hence similar ratings for compression to 80% and above of the full, 3D model, whereas cardboarding up to 30% almost always noticeable.

3.1 Method

We recruited 19 naive participants (2F,17M) aged between 23 and 34 with normal or corrected-to-normal vision in both eyes. Of this group, 15 participants performed all three experiments in random order, while four performed only the Pairs and Rating studies. The observers viewed 3D scenes consisting of the simple objects shown

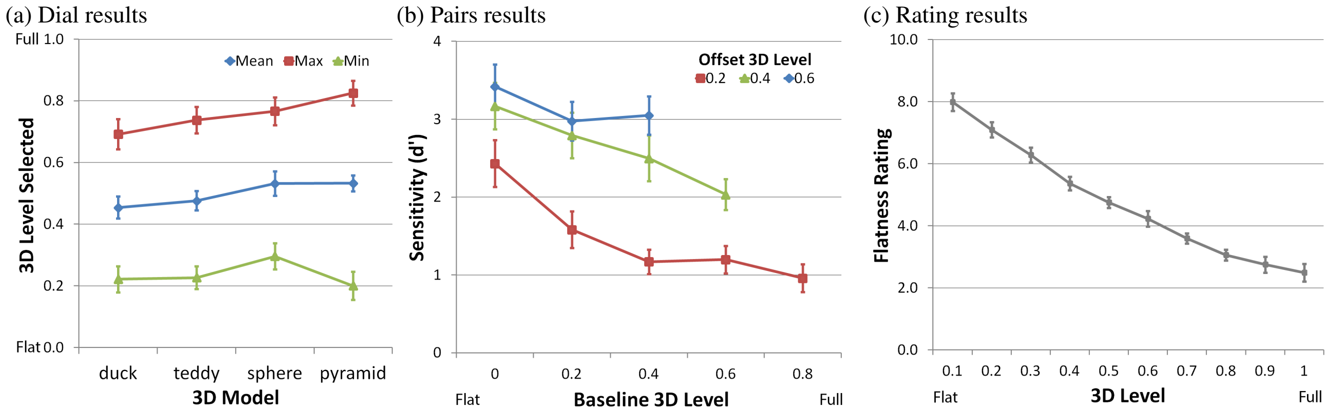

Figure 4: *Results of our three experiments. Standard error bars are shown in each case. Figures (b) and (c) are averaged over all models.*

Base	**Offset**	*PAIR Groups*								**Rating**	*LEVEL Groups*							
0.8	0.2	†								1.0	†							
0.4	0.2	†	†							0.9	†							
0.6	0.2	†	†							0.8	†	†						
0.2	0.2		†	†						0.7		†	†					
0.6	0.4			†	†					0.6			†	†				
0.0	0.2				†	†				0.5				†	†			
0.4	0.4				†	†	†			0.4					†			
0.2	0.4					†	†	†		0.3						†		
0.2	0.6						†	†	†	0.2							†	
0.4	0.6							†	†	0.1								†
0.0	0.4							†	†									
0.0	0.6								†									

Table 1: *Homogeneous groups calculated using Fisher's LSD post-hoc analysis for: Pair effect in the Pairs experiment (l); Level effect in the Rating experiment (r). Each column indicates which pairs, or levels, were found to not be significantly different from each other. The values are graphed in Figure 4.*

in Figure 3 on an Alienware 2310 23" 3D capable monitor with the help of time-multiplexed glasses, and sat approximately 60 centimeters from the screen. A mix of geometric and natural objects was selected, with both angular and round appearance. The standard setup for each experiment mimicked the position of the observer's eyes as cameras in the renderer, which were located 6 centimeters apart and 60 centimeters away from the objects being rendered. In this way, the rendered unmodified 3D scene showed objects with similar 3D characteristics as those of a real-world object at the center of the screen. At the start of each experiment, cardboarding was explained to each participant and they received training on each task, and written instructions were available throughout for reference.

The rendering cameras were oriented parallel to each other along the z axis and the resulting stereo images were re-converged around the center of coordinates, i.e., the center of coordinates always had zero disparity and appeared to be at the screen's depth. For the camera baseline β and point $p = (x_p, y_p, z_p)$, the disparity between the rendered views is d_p. If our camera baseline was changed to be $\alpha * \beta$ with $\alpha \in [0, 1]$, the disparity of p would become $\alpha * d_p$. This effectively gives us the freedom to linearly control the overall disparity compression of our scene by changing the baseline by the factor α. Figure 1 shows an example of a mesh mapped with $\alpha = 0.0$, $\alpha = 0.2$, $\alpha = 0.8$ and $\alpha = 1.0$.

In the *Dial* experiment, a method-of-adjustment process was performed where the same object was displayed twice, once with $\alpha = 0$ and the other with $\alpha = 1$. Pressing one button increased α by 0.02 and another decreased it by the same amount. This gave a total of 16 stimuli (4 models X 2 directions X 2 repetitions). When the object began flat, the task was to select the point when cardboarding stopped being disturbing; when the object began full, the point where cardboarding started to become disturbing was selected. In the *Pairs* study, users were shown two versions of the same model side by side. The shapes were shown with either the same or different levels of fullness (i.e. the same α value), with random left-right placement. One object was known as the baseline level, with possible values of $\alpha \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$ and was compared with another with one of the offsets: $0.0, 0.2, 0.4, 0.6$ added, with three repetitions of each pair. The task was to answer yes or no to the question: "Are the objects the same in terms of cardboarding?" Finally, for the *Ratings* experiment, a single object was displayed in the center of the screen with a random α baseline factor from 0 to 1 with a 0.1 step. Each stimulus was repeated twice, totaling 20 stimuli for each of the four models. The task was to rate the scene on a scale of 1 to 10 with respect to cardboarding, with $1 =$ "Not disturbing at all" and $10 =$ "Very disturbing, completely flat".

3.2 Results

We performed Repeated Measures Analysis of Variance (ANOVA) on participant responses to test for statistically significant effects, and performed post-hoc analysis using Fisher's LSD (Least Significant Difference) test for pair-wise comparisons of means. Effects are considered to be significant at the 95% level ($p < 0.05$). The results are summarized in Figure 4 and Table 1.

For the *Dial* experiment (Figure 4(a)), we performed single factor (4 Model) repeated measures ANOVAs on the Min, Max and Mean of each participant's selected levels, averaged over all participants. There was a main effect of Model for the means ($F(3, 42) = 3.05, p < 0.05$), where the duck was set to a significantly lower level on average than the sphere or pyramid, but not the teddy. This is probably due to the beak of the duck where the change in 3D was much more obvious, in that participants reported that it "came out" of the screen more and contrasted more with the tail in the background. We can see that each participant selected a wide range of acceptable levels, indicating that the decision was a difficult one for them. However, the Max values rarely exceeded 80%, indicating that compression to that level and above was not found to be disturbing. The averages were around 50% and the lowest Min values were around 20%, meaning that in some cases, they accepted very high compression levels for some stimuli.

The task in the *Pairs* experiment (Figure 4(b)) is a signal detection one, so we calculated the sensitivity of each participant to a difference between the two images. The d-prime (d') metric is commonly used in psychophysics to reliably measure sensitivity to a signal, as it takes response bias into account (i.e., the tendency to be over-conservative or over-discriminative) by considering both the Hit Rate (e.g., percentage of time a difference is correctly reported) and the False Alarm Rate (e.g., percentage of time the images are incorrectly reported to be different when they are the same). High values indicate that participants are very sensitive to a difference being present between the stimuli, whereas values of 1 and below are considered to be guessing. We performed a two-way (4 Model x 12 Pair) repeated measures ANOVA on the d' values. A main effect of Model ($F(3, 54) = 7.05, p < 0.0005$) was found, where differences for the sphere were most easily detected, and of Pair ($F(11, 198) = 25.5, p \approx 0.0$), where the same differences between fuller stimuli were far less detectable than between those that were very compressed. This result is expected, as low α values incurred a larger relative change. Again, when compression was to 80% or above, sensitivity was at its lowest, whereas when compression was to 20% or below, performance was above chance. These results are consistent with our findings in the Dial experiment. Please see the homogeneous groups in Table 1(left).

Finally, we performed a two-way (4 Model x 10 Level) repeated measures ANOVA on the results of the *Rating* experiment (Figure 4(c)) and found a main effect of the preference Level ($F(9, 171) = 64.3, p \approx 0.0$). From the homogeneous groups shown in Table 1(right), we can see that the flattest levels 0.1-0.3 are all significantly different, whereas differences between the fuller 0.8-1 stimuli are much smaller, and not statistically significant, indicating a plateauing effect at about 80%. Compression to 40% was rated on average just above 5, indicating that this is the point after which the flatness became noticeable more often than not. From 30% it was clearly rated flat far more often. From these and the results of the other two experiments, we can conclude that depth compression to 80% fullness or above is likely to be acceptable, whereas below 30% it is probably never going to be. Of course, we cannot generalize from the four simple scenes we presented to more complex scenes, though it seems possible that we have presented a worst-case scenario, and more complexity might mask cardboarding artifacts further, allowing higher compression rates below the conservative 80% limit than we found here, as several mid-range levels were acceptable at least some of the time.

4 Conclusions

We have shown that, at least for the simple scenes depicted in our experiments, depth can be safely compressed by up to 20% without significantly affecting perceived cardboarding. It may be possible to compress at much higher rates, as there appears to be a wide range of compression ratios that appear acceptable to some viewers at least some of the time. However, it appears that below 30% of the natural depth, cardboarding is significantly disturbing. The results obtained in this paper could be directly applied to guide existing depth remapping methods. Further studies are needed to examine the effects of many other factors (e.g., lighting effects, scene complexity, motion) and also to determine more subjective preference measures, in addition to the simple ratings we recorded here. Our findings may provide information that could be used to map depth into a smaller range while avoiding as much as possible the introduction of disturbing cardboarding artifacts. Previous approaches such as [Lang et al. 2010] and [Chapiro et al. 2014] could use the cluster boundaries we have found as targets for the depth budget given to a salient region.

References

CHAPIRO, A., HEINZLE, S., AYDIN, T., POULAKOS, S., ZWICKER, M., SMOLIC, A., AND GROSS, M. 2014. Optimizing stereo-to-multiview conversion for autostereoscopic displays. *Comp. Graph. Forum 33*, 2.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2011. A perceptual model for disparity. *ACM Trans. Graph.*.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2012. Apparent stereo: The cornsweet illusion can enhance perceived depth. In *HVEI, IS&T SPIE's Symp. Electronic Imaging*.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., SEIDEL, H.-P., AND MATUSIK, W. 2012. A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph. 31*, 6.

HOWARD, I. P., AND ROGERS, B. J. 2002. *Seeing in Depth: Volume 2: Depth perception*. Oxford University Press, NY, USA.

JONES, G., LEE, D., HOLLIMAN, N., AND EZRA, D. 2001. Controlling perceived depth in stereoscopic images. In *Proc. SPIE*, vol. 4297.

LAMBOOIJ, M., IJSSELSTEIJN, W. A., FORTUIN, M., AND HEYNDERICKX, I. 2009. Visual discomfort and visual fatigue of stereoscopic displays: A review. *J. Imaging Science and Technology 53*, 5.

LANG, M., HORNUNG, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3d. *ACM Trans. Graph. 29*, 3.

MASAOKA, K., HANAZATO, A., EMOTO, M., YAMANOUE, H., NOJIRI, Y., AND OKANO, F. 2006. Spatial distortion prediction system for stereoscopic images. *J. Electronic Imaging 15*, 1.

MEESTERS, L. M. J., IJSSELSTEIJN, W. A., AND SEUNTIÃNS, P. J. H. 2004. A survey of perceptual evaluations and requirements of three-dimensional tv. *IEEE Trans. Circuits and Systems for Video Technology 14*, 3.

SHIBATA, T., KIM, J., HOFFMAN, D. M., AND BANKS, M. S. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *J. of Vision 11*, 8.

SIEGEL, M. W., AND NAGATA, S. 2000. Just enough reality: Comfortable 3-d viewing via microstereopsis. *IEEE Transs on Circuits and Systems for Video Technology 10*, 3.

WOODS, A. J., DOCHERTY, T., AND KOCH, R. 1993. Image distortions in stereoscopic video systems. *Proc SPIE 1915*.

YAMANOUE, H., OKUI, M., AND YUYAMA, I. 2000. A study on the relationship between shooting conditions and cardboard effect of stereoscopic images. *IEEE Trans.Circuits and Systems for Video Technology 10*, 3 (Apr).

YAMANOUE, H., OKUI, M., AND OKANO, F. 2006. Geometrical analysis of puppet-theater and cardboard effects in stereoscopic hdtv images. *Circuits and Systems for Video Technology, IEEE Transactions on 16*, 6.

ZILLY, F., KLUGER, J., AND KAUFF, P. 2011. Production rules for stereo acquisition. *Proc. of the IEEE 99*, 4 (April).