The Effect of Focus Cues on Separation of Information Layers

Patrick Bader[1,2], **Niels Henze**[1], **Nora Broy**[3], **Katrin Wolf**[4]
[1]University of Stuttgart [2]Stuttgart Media University
[3]BMW Group [4]BTK - University of Art and Design
[1]{firstname.lastname}@vis.uni-stuttgart.de, [3]Nora.NB.Broy@bmw.de, [4]katrin.wolf@acm.org

ABSTRACT

Our eyes use multiple cues to perceive depth. Current 3D displays do not support all depth cues humans can perceive. While they support binocular disparity and convergence, no commercially available 3D display supports focus cues. To use them requires accommodation, i.e. stretching the eye lens when focusing on an individual distance. Previous work proposed multilayer and light field displays that require the eye to accommodate. Such displays enable the user to focus on different depths and blur out content that is out of focus. Thereby, they might ease the separation of content displayed on different depth layers. In this paper, we investigate the effect of focus cues by comparing 3D shutter glasses with a multilayer display. We show that recognizing content displayed on a multilayer display takes less time and results in fewer errors compared to shutter glasses. We further show that separating overlapping content on multilayer displays again takes less time, results in fewer errors, and is less demanding. Hence, we argue that multilayer displays are superior to standard 3D displays if layered 3D content is displayed, and they have the potenal to extend the design space of standard GUI.

ACM Classification Keywords

H.5.2 User Interfaces: GUI.

Author Keywords

Multilayer; 3D; display; S3D; depth cues.

INTRODUCTION AND BACKGROUND

Depth perception arises from a number of cues. These are typically divided in binocular cues that result from perceiving the world with both eyes and monocular cues that can be observed with one eye. Most monocular depth cues, such as motion parallax and perspective, can be rendered by 2D displays. Commercially available 3D displays are essentially stereoscopic (S3D) displays. They add additional depth cues by showing different images to the left and to the right eye. Shutter glasses, for example, show alternating images for the left and the right eye while blocking the view of the other eye. Additional depth cues provided by current S3D displays are binocular disparity (observing the world from two different viewpoints, one from each eye) and convergence (stretching the extraocular muscles to focus on the same object with both eyes). Today, S3D displays are mainly used to present natural and virtual 3D scenes. Previous work, however, also proposed to use 3D displays to add an additional degree of freedom for GUIs. Sunnari et al., for example, presented a 3D menu for S3D displays [18] and Häkkilä et al. presented a phonebook application with a S3D user interface [7]. Broy et al. even provided tools for prototyping 3D layouts [5].

The performance while using S3D displays has been the subject of a body of work. Froner et al., for example, looked into fine depth perceptions using S3D displays and conclude that the display technology is a key factor [6]. In the context of data visualization, Ware and Mitchell found lower error rates on 3D graph exploration tasks with S3D compared to 2D displays [19]. Surveying work that studied the use of S3D displays, McIntire et al. conclude that 3D shows performance benefits on a variety of depth-related tasks [14]. However, the authors state that S3D displays can induce a number of problems including eyestrain, headache, fatigue, disorientation, nausea, and malaise for up to 50% of the population. As the use of extreme parallaxes and the vergence-accommodation conflict are major reasons, observing defined parallax limits solve possible discomfort [12]. Broy et al. identified parallax boundaries for a S3D shutter display allowing the comfortable perception of S3D content [4]. However, these boundaries limit the 3D design space to a certain depth range.

The concept of layered user interfaces was first introduced by Harrison et al. [8]. They state that structuring information on several depth layers enhances both, focused and divided attention. Atchley et al. [1] showed that switching attention between two depth layers compared to switching attention on one depth layer requires longer saccades. This provides evidence for a depth aware attentional spotlight. In general, Mowforth et al. showed that switching attention from a rear layer to a front layer occurs faster than vice versa [15].

Current S3D displays do not provide all perceivable depth cues. They do not support accommodation, requiring the viewer to changing the focal length by stretching or squeezing the eye lens. As a result of the eye's limited depth of focus, accommodation causes defocus blur at distances other than the focal length. This focus cue is needed to overcome the vergence-accommodation conflict. Previous work proposed displays that support this additional depth cue. Multi-layer

CHI'16, May 7–12, 2016, San Jose, CA, USA

http://dx.doi.org/10.1145/2858036.2858312

displays [16, 17, 2] combine transparent displays to a stack of displays. The physical distance between the displays enables content that can be shown at different depths while supporting all depth cues humans can perceive. In contrast to current S3D displays, they also require accommodation and thus, create true defocus blur. Similarly, light field displays [3, 11] can support focus cues. Recent work even demonstrated compact light field displays with a large range depth of field [20, 10].

Overall, previous work suggests that S3D displays can improve users' objective performance but can also lead to a number of negative effects [14, 12]. It remains to be investigated, however, how displays which provide the full range of perceivable depth cues, affect users' performance while interacting with GUIs. In this paper, we therefore compare a S3D display with shutter glasses with a two-layer display to identify the effects of focus cues on content recognition on multiple depth layers. We hypothesize that:

H1: Recognition of non-overlapping content shown on different depth layers is easier with consistent focus cues.

H2: Switching between depth layers is more demanding and slower when eye accommodation is required.

H3: Visual separation of overlapping content shown on two depth layers is easier with consistent focus cues.

THE EFFECT OF FOCUS CUES

We conducted an experiment to investigate the effect of focus cues on users' performance and task load. In the experiment we displayed content on two depth layers and compared two conditions. In the first condition (*disparity*), we used a S3D display with shutter glasses to make use of binocular disparity as a depth cue. In the second condition (*focus*), we displayed each layer at its correct focal plane to prevent vergence-accommodation conflicts using a multi-layer display. Although we use two discrete depth layers, we expect our results to be generalizable to displays which provide continuous depth and consistent focus cues [20, 10].

Apparatus

We developed an apparatus that realizes both conditions with minimal adjustments to ensure comparable results. It uses two BenQ XL2411 displays, one for each depth layer, which support stereo 3D with shutter glasses. The front screen is located left of a half-silvered mirror and the back screen is located behind the half mirror. Images from the displays are combined with a half-silvered mirror sheet (see Figure 1).

The front layer is 30 cm away from the participant in both conditions. The *disparity* condition is implemented with shutter glasses to show S3D images. The display showing the virtual back layer is physically located 30 cm from the participant. Using binocular disparity a virtual depth layer is shown behind the actual display to let the layer appear to be 50 cm away from the user, see Figure 1 A. In the *focus* condition the physical distance between the participant and the back display is increased to 50 cm. Thus, the focal planes of the two displays are separated although both layers appear to be at the same distances as in the *disparity* condition, see Figure 1B. Participants wore shutter glasses in both conditions, regardless of binocular disparity being enabled or not. A chin rest was used to restrict head movement to prevent the use of further focus cues such as motion parallax. We calibrated the apparatus to ensure that content on the displays had the same brightness and color. For both conditions, the viewports of the displays were carefully aligned.

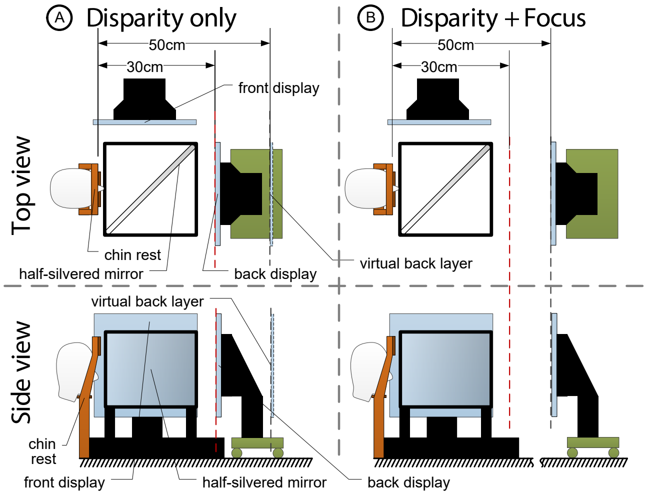

Figure 1. Top and side view of the apparatus in both display conditions. A shows the configuration for binocular disparity. B shows the configuration for disparity with correct focus cues.

Tasks

Task 1: Recognizing Information

In the first task, we investigated the effect of focus cues on recognizing information displayed on a specific depth layer. For each trial, between ten and eleven circles were shown at the same time. The circles were distributed randomly among the two layers with at least one circle per layer. Before each trial, participants were instructed to count the circles on one specific layer and press the space key to start. After 700 ms the circles were shown and a timer started. As soon as participants knew the answer they had to press the space key again so the circles disappeared and time measurement stopped. Participants said their answer aloud and the task continued with the next trial. Participants had to complete 40 trials per condition. Each circle pattern was shown twice for counting on the front layer and on the back layer respectively. Participants were told to perform the trials as fast and accurately as possible. We measured the task completion time (TCT) and error rate.

We varied the total number of circles to prevent participants from counting the circles on one layer and subtracting the result from a fixed total number. All circles were shown in a different color from a set of predefined colors to counter potentially remaining chromatic differences between displays and layers. We varied the circles' sizes to prevent participants from guessing the layer based on circle size. On the front layer circles were 14±0.5 mm large. We randomly placed the circles on a 4x4 grid to prevent overlapping. On the front layer the grid rows and columns were 35 mm apart. We randomly varied the circles' positions slightly (±7 mm on the front layer). All sizes were scaled for the back layer to compensate for perspective projection.

Task 2: Switching Layers

In the second task we investigated the time required to switch between two information layers. A letter (*Q* or *O*) was shown either on the front or the back layer at the center of the screen. Participants had to press the corresponding keyboard key as quickly as possible. Afterwards, the next letter was immediately displayed. Since letters can either be shown on the front or back layer, for this task four transition types are possible:

front-back: Previous letter was shown on the front layer and current letter is shown on the back layer.
back-back: Both previous and current letter are shown on the back layer.
back-front: Previous letter was shown on the back layer and current letter is shown on the front layer.
front-front: Both previous and current letter are shown on the front layer.

Each transition type was shown 50 times. Thus, a total of 200 letters had to be read for each condition. We randomized both letters and transition types. With this design sometimes neither the letter nor the layer changes. Thus, we varied the colors of all letters from a predefined set of ten colors to give users additional feedback about the switch. The letters *O* and *Q* were chosen for their similarity to force participants to precisely focus before being able to identify the correct one. As the letters are relatively far apart on the keyboard and users had to blindly type them, a key mapping had been used during the study. The key *J*, which can easily be recognized blindly by its haptic marker, was mapped to the letter *Q*. Its neighboring key *K* was mapped to *O*. Both key mappings were written on the keyboard. This remapping might have introduced some additional cognitive load. However this applies to all conditions the same and should not affect the results negatively. We measured TCT and error rates for each transition and condition.

Task 3: Separating Overlapping Content

In the third task, we investigated the effect of real focus cues on the ability to separate two depth layers when overlapping content is displayed on two depth layers. Participants had to count the occurrences of specific words on overlapping texts shown on both display layers. The texts were aligned to maximize overlapping of the texts, which represents a worst case scenario for reading overlapping content. For both texts we used the same font (Microsoft Sans Serif) and size (5 mm in height for the front layer and back layer scaled accordingly to correct for perspective projection). We showed 150 words long texts taken from *Grimm's Fairy Tales* on each layer. Participants had to count short common words (e.g. she, each, snow) with three to four letters on both layers.

Similar to the previous task, first text instructions indicated which word to count. After pressing the space key the texts were shown until participants pressed the space key again and the texts disappeared. TCT was measured during this period. Participants were then instructed to type in how many word occurrences they had counted. For this task, participants had to complete 8 trials in total (4 per display condition) and text fragment order was randomized for each participant. Again we measured TCT and error rate.

Participants and Procedure

We recruited 21 (5 female) participants for the study with an average age of 27 years ($SD = 3.46$, $Md = 27$). All except three participants had normal or corrected to normal visual acuity and stereo vision. We excluded the results of these three participants from our analysis. First participants had to fill in a demographic questionnaire and answer questions regarding visual impairments including stereo vision. Afterwards, they proceeded with the three tasks. During each task participants had to place their head on the chinrest. After each task a raw NASA TLX [9] questionnaire was answered. The study took 30 to 45 min per participant. All participants were rewarded with sweets for taking part in the study.

Results

Task 1: Recognizing Information

Paired samples t-tests revealed a statistically significant different task completion times for the disparity condition ($M = 4.515, SD = 2.597$) and the focus condition ($M = 3.167, SD = 1.344$); $t(17) = 3.781, p = 0.001$, see Figure 2 left. The error rate was significantly higher for disparity ($M = 0.186, SD = 0.218$) than for focus ($M = 0.106, SD = 0.140$); $t(17) = 2.240, p = 0.039$.

Analyzing the sub-scales of the NASA TLX revealed that participants rated their *performance* significantly worse using disparity ($M = 9.056, SD = 4.569$) than using focus ($M = 7.167, SD = 3.808$); $t(17) = 2.239, p = 0.039$. Participants rated their *frustration* significantly higher using disparity ($M = 7.444, SD = 4.382$) than using focus ($M = 5.722, SD = 3.908$); $t(17) = 2.550, p = 0.021$. We found no significant differences for *mental demand* ($p = 0.707$), *physical demand* ($p = 0.212$), *temporal demand* ($p = 0.775$), or *effort* ($p = 0.677$).

Task 2: Switching layers

Analyzing the task completion time, Mauchly's test indicated that the assumption of sphericity has been violated for transition type ($p = 0.019$) and the interaction between focus cue and transition type ($p = 0.028$). A Greenhouse-Geisser corrected two-way repeated measures ANOVA revealed that the focus cue ($F(1, 17) = 26.133, p < 0.001$) and the transition type ($F(2.010, 34.173) = 95.002, p < 0.001$) had a significant effect on the task completion time. ($F(2.010, 34.173) = 95.002, p < 0.001$). Furthermore, an ANOVA revealed a significant interaction effect ($F(1.946, 33.081) = 33.116, p < 0.001$). Bonferroni corrected post-hoc tests revealed significant differences ($p < 0.001$) between the layer changing transitions (front-back, back-front) and for transitions on the same layer (front-front, back-back), see also Figure 2 right. Thus, TCT for layer changing is significantly higher when real focus cues are provided.

Analyzing error rates, a two-way repeated measures ANOVA revealed a significant main effect for focus cue ($F(1, 17) = 10.783, p = 0.004$) and for the interaction between focus cue and transition type ($F(3, 51) = 2.916, p = 0.043$) but not for transition type ($p = 0.380$). Error rate for disparity was significantly lower ($M = 0.014, SE = 0.003$) than for focus ($M = 0.025, SE = 0.004$). The subjective answers to the NASA TLX showed significantly better results

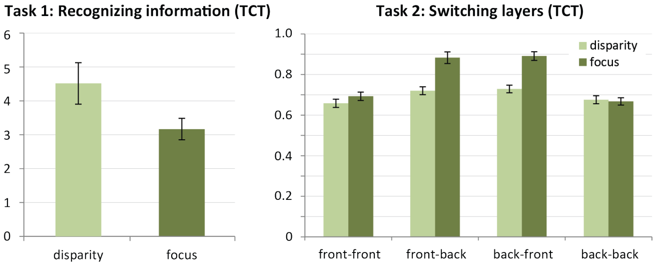

Figure 2. Mean trial completion times in seconds for task one and two, error bars show standard error.

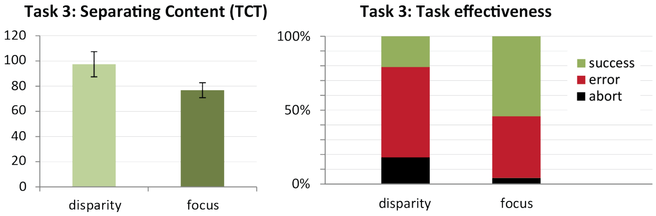

Figure 3. Mean trial completion times in seconds, error bars show standard error (left) and task effectiveness in percent (right) for task three.

for *performance* with disparity ($M = 6.889, SD = 4.114$) compared to focus ($M = 9.000, SD = 3.8957$); $t(17) = -2.133, p = 0.048$. *Effort* was also significantly lower with disparity ($M = 11.222, SD = 5.663$) than with focus ($M = 12.389, SD = 5.237$); $t(17) = -2.333, p = 0.032$ and *frustration* was lower with disparity ($M = 8.333, SD = 5.029$) compared to focus ($M = 10.278, SD = 5.454$); $t(17) = -3.229, p = 0.005$. However, we found no significant differences for *mental* ($p = 0.421$), *physical* ($p = 0.078$) and *temporal demand* ($p = 0.545$).

Task 3: Separating Overlapping Content

For each condition and participant we computed the average TCT and error rate. The task was aborted 18 times for disparity and 3 times for focus. We removed aborted trials and excluded one participant from the analysis who aborted all trials with disparity. Figure 3 shows the percentage of trials which were excluded from the analysis for each condition.

As in the first task we used paired samples t-tests to analyze task completion time and error rate. We took the absolute differences of word occurrences and user's answers as error metric: $error = |\#_{total} - \#_{user}|$. A t-test revealed that participants were significantly slower using disparity ($M = 97.343, SD = 41.208$) compared to focus ($M = 76.766, SD = 24.559$); $t(16) = 2.722, p = 0.015$), see Figure 3 left. We found no significant difference in error rate ($p = 0.109$), yet success rate was significantly better for focus ($M = 0.54, SD = 0.324$) than for disparity ($M = 0.208, SD = 0.214$); $t(17) = 4.000, p = 0.001$.

Analyzing the sub-scales of the NASA TLX revealed that participants rated their *performance* significantly worse for disparity ($M = 15.722, SD = 2.845$) than for focus ($M = 10.222, SD = 3.859$); $t(17) = 5.229, p < 0.001$. *Effort* was also perceived higher for disparity ($M = 16.667, SD = 2.612$) compared to focus ($M = 12.778, SD = 3.904$); $t(17) = 5.160, p = 0.000$. Similarly, *frustration* was significantly higher for disparity ($M = 14.667, SD = 3.498$) than for focus ($M = 10.056, SD = 3.171$); $t(17) = 3.953, p = 0.001$. No significant differences were found for *temporal* ($p = 0.889$) and *physical demand* ($p = 0.087$), yet *mental demand* was perceived significantly higher for disparity ($M = 16.000, SD = 3.181$) than for focus ($M = 13.667, SD = 4.472$); $t(17) = 2.752, p = 0.014$.

Discussion

Analysis of the first task revealed that enabling accommodation results in shorter task completion times and in fewer errors when recognizing content on a specific layer. So we could accept the hypothesis on content recognition (H1). For the third task, participants also needed less time to separate two depth layers and made less errors with correct focus cues. Therefore our hypothesis on visual separation (H3) was also accepted. Subjective feedback for both tasks is consistent with objective results. They were less frustrating and performance was perceived better.

In the second task, participants, however, needed more time to switch between layers and made more errors with correct focus cues. A potential reason is that accommodation takes additional time which is not required for the S3D display. Subjective feedback from our participants supports these findings. Our initial hypothesis on switching depth layers (H2) was therefore accepted. Furthermore participants of the study were mainly young. Previous work showed that accommodation time increases with age [13]. Older users might therefore need even more time to switch between information layers if they have to accommodate.

CONCLUSION

In this paper, we investigated addition of consistent focus cues to 3D displays when showing content on two depth layers. We showed that recognition and visual separation of content on two layers takes less time, results in fewer errors and is perceived better with consistent focus cues compared to S3D displays with shutter glasses.

We envision using multilayer displays to extend the design space of standard GUI. A separate information layer could show annotations in a text viewer or tool windows in image editing software above the main content. Moreover, context labels could be placed on maps or a video player timeline. However, our results show that accommodation takes time and makes switching between depth layers slower. This has to be taken into account when designing interfaces for multilayer displays to avoid canceling out their advantages.

In our study, we used a two-layer display. While we assume similar results, the effect on recently proposed compact light field displays [20, 10] still has to be investigated. Furthermore, a combination with motion parallax that is naturally provided by multilayer displays if the user is free to move, needs further research.

ACKNOWLEDGEMENTS

This work is partially supported by the graduate program Digital Media of the Universities of Stuttgart and Tübingen, and the Stuttgart Media University (HdM).

REFERENCES

1. Paul Atchley, ArthurF. Kramer, GeorgeJ. Andersen, and Jan Theeuwes. 1997. Spatial cuing in a stereoscopic display: Evidence for a “depth-aware” attentional focus. *Psychonomic Bulletin & Review* 4, 4 (1997), 524–529. DOI: http://dx.doi.org/10.3758/BF03214343
2. Patrick Bader, Valentin Schwind, Niels Henze, Stefan Schneegass, Nora Broy, and Albrecht Schmidt. 2014. Design and Evaluation of a Layered Handheld 3D Display with Touch-sensitive Front and Back. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI ’14)*. ACM, New York, NY, USA, 315–318. DOI: http://dx.doi.org/10.1145/2639189.2639257
3. Robert G. Batchko. 1994. Three-hundred-sixty degree electroholographic stereogram and volumetric display system. *Proc. SPIE* 2176 (1994), 30–41. DOI: http://dx.doi.org/10.1117/12.172650
4. Nora Broy, Florian Alt, Stefan Schneegass, Niels Henze, and Albrecht Schmidt. 2013. Perceiving Layered Information on 3D Displays Using Binocular Disparity. In *Proceedings of the 2Nd ACM International Symposium on Pervasive Displays (PerDis ’13)*. ACM, New York, NY, USA, 61–66. DOI: http://dx.doi.org/10.1145/2491568.2491582
5. Nora Broy, Stefan Schneegass, Florian Alt, and Albrecht Schmidt. 2014. FrameBox and MirrorBox: Tools and Guidelines to Support Designers in Prototyping Interfaces for 3D Displays. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI ’14)*. ACM, New York, NY, USA, 2037–2046. DOI: http://dx.doi.org/10.1145/2556288.2557183
6. Barbara Froner, Nicolas S. Holliman, and Simon P. Liversedge. 2008. A comparative study of fine depth perception on two-view 3D displays. *Displays* 29, 5 (2008), 440 – 450. DOI: http://dx.doi.org/10.1016/j.displa.2008.03.001
7. Jonna Häkkilä, Maaret Posti, Olli Koskenranta, and Leena Ventä-Olkkonen. 2013. Design and Evaluation of Mobile Phonebook Application with Stereoscopic 3D User Interface. In *CHI ’13 Extended Abstracts on Human Factors in Computing Systems (CHI EA ’13)*. ACM, New York, NY, USA, 1389–1394. DOI: http://dx.doi.org/10.1145/2468356.2468604
8. Beverly L. Harrison, Hiroshi Ishii, Kim J. Vicente, and William A. S. Buxton. 1995. Transparent Layered User Interfaces: An Evaluation of a Display Design to Enhance Focused and Divided Attention. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’95)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 317–324. DOI: http://dx.doi.org/10.1145/223904.223945
9. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9
10. Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. 2015. The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-eye Light Field Displays with Focus Cues. *ACM Trans. Graph.* 34, 4, Article 60 (July 2015), 12 pages. DOI: http://dx.doi.org/10.1145/2766922
11. Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec. 2007. Rendering for an Interactive 360&Deg; Light Field Display. , Article 40 (2007). DOI: http://dx.doi.org/10.1145/1275808.1276427
12. Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. 2009. Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology* 53, 3 (2009), 30201–1–30201–14. DOI: http://dx.doi.org/doi:10.2352/J.ImagingSci.Technol.2009.53.3.030201
13. Thurmon E Lockhart and Wen Shi. 2010. Effects of age on dynamic accommodation. *Ergonomics* 53, 7 (2010), 892–903. DOI: http://dx.doi.org/10.1080/00140139.2010.489968
14. John P. McIntire, Paul R. Havig, and Eric E. Geiselman. 2012. What is 3D good for? A review of human performance on stereoscopic 3D displays. *Proc. SPIE* 8383 (2012), 83830X–83830X–13. DOI: http://dx.doi.org/10.1117/12.920017
15. Peter Mowforth, John E. Mayhew, and John P. Frisby. 1981. Vergence eye movements made in response to spatial-frequency-filtered random-dot stereograms. *Perception* 10, 3 (1981), 299–304.
16. Vijay Prema, Gary Roberts, and BC Wuensche. 2006. 3D visualisation techniques for multi-layer display technology. In *IVCNZ*, Vol. 6. 27–29.
17. Jong-Wook Seo and Taeho Kim. 2008. Double-Layer Projection Display System Using Scattering Polarizer Film. *Japanese Journal of Applied Physics* 47, 3R (2008), 1602. http://stacks.iop.org/1347-4065/47/i=3R/a=1602
18. Meiju Sunnari, Leena Arhippainen, Minna Pakanen, and Seamus Hickey. 2012. Studying User Experiences of Autostereoscopic 3D Menu on Touch Screen Mobile Device. In *Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI ’12)*. ACM, New York, NY, USA, 558–561. DOI: http://dx.doi.org/10.1145/2414536.2414622

19. Colin Ware and Peter Mitchell. 2008. Visualizing Graphs in Three Dimensions. *ACM Trans. Appl. Percept.* 5, 1, Article 2 (Jan. 2008), 15 pages. `DOI`: **http://dx.doi.org/10.1145/1279640.1279642**

20. Gordon Wetzstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar. 2012. Tensor Displays: Compressive Light Field Synthesis Using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph.* 31, 4, Article 80 (July 2012), 11 pages. `DOI`: **http://dx.doi.org/10.1145/2185520.2185576**