The effect of stereoscopy and motion cues on 3D interpretation task performance

Boris W. van Schooten
University of Twente
P.O. Box 217
7500 AE Enschede, NL
schooten@ewi.utwente.nl

Elisabeth M.A.G. van Dijk
University of Twente
P.O. Box 217
7500 AE Enschede, NL
bvdijk@ewi.utwente.nl

Elena Zudilova-Seinstra
University of Amsterdam
P.O. Box 19268
1000 GG Amsterdam, NL
e.v.zudilova-seinstra@uva.nl

Avan Suinesiaputra
LUMC, Leiden University
P.O. Box 9600
2300 RC, Leiden, NL
a.suinesiaputra@lumc.nl

Johan H.C. Reiber
LUMC, Leiden University
P.O. Box 9600
2300 RC, Leiden, NL
j.h.c.reiber@lumc.nl

ABSTRACT

We study the effectiveness of stereoscopy and smooth motion as 3D cues for medical interpretation of vascular structures as obtained by 3D medical imaging techniques. We designed a user study where the user has to follow a path in a maze-like solid shaded 3D structure. The user controls rotation of the model. We measure user performance in terms of time taken and error rate. The experiment was executed with 32 (medical and non-medical) users. The results show that motion cue is more important than stereoscopy, and that stereoscopy has no added value when motion is already present, which is not consistent with previous experiments.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User interfaces—*Evaluation/methodology*; I.3.7 [**Computer Graphics**]: Three-dimensional Graphics and Realism

General Terms

Graphics, Human Factors

Keywords

motion cue, stereoscopy, medical visualization, angiography

1. INTRODUCTION

We are studying the importance of different visual cues for visual search tasks in the medical domain. The medical task we are interested in is inspection of 3D angiography data, as is obtained by for example magnetic resonance angiography (MRA) or computed tomography angiography (CTA). Therefore, we study the visualisation of networks or graphs

of interconnected tubes in 3D, rendered with basic shading and occlusion cues. This is similar to angiographic volumes rendered with isosurface or direct volume rendering. We are interested in two additional visual cues which could be added to medical visualisation with the help of recent technological developments. These are motion cue, which can be added when computers are fast enough to produce real-time rendering, and stereoscopy cue, which can be added with an appropriate stereoscopic device.

There have been many experiments studying stereoscopy and motion cue as basic parameters influencing task performance. However, most of them involve tasks which cannot be considered appropriate interpretation tasks. Examples are positioning and distance estimation in 3D [4], tracing a wire (using a 3D input device) [3], manipulating simple objects, or navigating in a virtual environment [5]. Data gathered on both speed and error rate for these "basic" tasks show that both stereoscopy and motion are relevant for many, but not all tasks. In some cases they are additive [10] and in others not [4, 13]. Generally motion is more important.

More relevant for us is a series of experiments involving 3D graph structure interpretation tasks, comparing monoscopy versus stereoscopy and motion versus no motion [10, 1, 11, 12]. In these four papers, two tasks have been proposed: a tree task and a graph task. The tree task was first proposed as an analogy of a vascular interpretation task. Ware et al. primarily studied graph tasks in the context of visualisation of abstract information. In both tasks, the user has to decide whether two nodes are connected to each other or not within a reasonable time limit, so that slower users are not seriously disadvantaged. This is a binary decision that produces 50% error rate for a random guessing strategy. These studies clearly conclude that, for these tasks, both motion and stereoscopy are significant, and usually, performance improvement adds up when both are used.

Unfortunately, all of these experiments have the same major shortcomings. We argue that this warrants further study with different experimental parameters. The main issue follows from the fact that the experimental variable is the users' error rate for binary choice tasks. Necessarily, errors are made only if the structures are designed to be very difficult to interpret. In fact, Ware and Franck [11] admit that some of the structures became ambiguous when viewed as a 2D
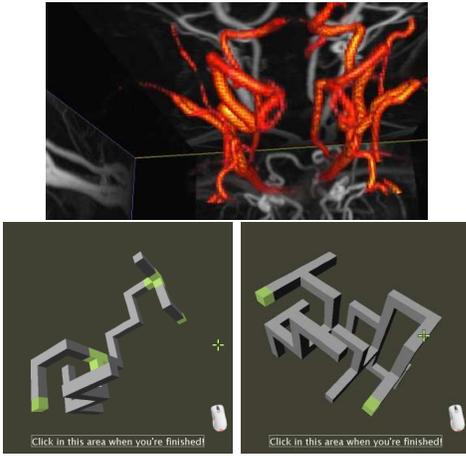
Figure 1: Maze structures used in our experiment, and comparison to real-life vascular structures. Top: typical vascular structure from MRA image in our domain; Bottom left: typical easy maze structure with junctions selected; Bottom right: typical hard maze structure.

projection, and that motion or stereo cues disambiguated the structures rather than making them easier to interpret. This means the experiments tested the *information* benefits of the 3D cues, rather than the *cognitive* ones. Further, there is a limited range of difficulty between zero (too easy) and 50% error rate (too difficult), which means there is a danger of floor and ceiling effects, both of which occurred in the experiments [10, 1]. Additionally, almost all of these experiments use wire (thin line) models, while our medical application domain uses shaded solid volume models. Arguably, the resulting shading and occlusion cues would be a meaningful addition for other domains such as graph visualisation as well.

2. EXPERIMENT

2.1 Task Design

In our experiment, users have to interpret maze-like structures. We chose these because they are self-avoiding branching structures which are easy to render and look structurally similar to vascular networks, see figure 1. We matched the complexity of the structures to typical complexity of the vascular images in our domain. The maze corridors were depicted as shaded rectangular boxes of significant thickness, so shading and occlusion provide clear 3D cues, and pixel resolution is not so important. Mazes are generated by running a standard maze digging algorithm for a 4x4x4 grid, and then selecting only those mazes that conform to a number of parameters deemed relevant to difficulty level, in particular total number of corridors and total number of junctions.

The actual task consists of following one particular path through the maze and pointing out the junctions along the path. This is a relatively easy task that requires minimal interaction, but it does require the user to perform 3D interpretation. It is analogous to visual inspection of a particular vessel in a vascular network. The two endpoints of the path are highlighted as blinking green. The user can select junc-

tions by clicking on them with the right mouse button. The selected junctions are highlighted as bright green. When the user thinks s/he is finished, s/he clicks at the bottom of the screen. The system reports the time performance and number of errors made. Errors are highlighted in blinking purple.

The users can rotate the structure freely. This ensures that no structures are ambiguous because of accidentally inconvenient view angles. Note that occlusion cues require user rotation to bring forward the occluded parts for both viewing and selection. This is another reason for having user-controlled rotation. Rotation is done with the left mouse button using the common "two-axis valuator" method [2].

For displaying a stereoscopic image, we follow some common guidelines [8]. Stereo is best achieved using: (1) parallel eye directions (to avoid keystoning error), with (2) depth (binocular disparity) not too exaggerated, but just enough to make out the necessary differences, and (3) the middle of the figure at zero parallax, and the front of the figure not sticking too far out of the screen 2D plane. The binocular disparity we chose amounted to slightly flatter than would be natural when viewed from a normal distance (60-70 cm from the screen). The 3D structures were presented so that some blank space was left on all sides. This way, there is no negative effect from one view being cropped differently than the other in stereo mode.

Displaying the mouse cursor in stereo was done using a variant of the "depth cursor" [6], which prevents interference of the 2D mouse pointer with the stereoscopic effect. We presented the mouse cursor stereoscopically, but with the depth adapted automatically so that it hovers just above the frontmost 3D features of the structure (which is the one that it selects if you click). Mouse depth is unchanged when the mouse moves to an area without 3D features behind it.

We wanted to test the presence versus absence of motion cues without resorting to lowering the frame rate, which would result in an unacceptably slow interactive feedback cycle. When displaying complex graphics, a lot of 3D software simplifies the rendered graphic when it is being moved or manipulated, to ensure that the resulting animation has a sufficiently high frame rate. In our experiment, we simulated such a simplified view of the object by showing only the centerpoints of the maze grid during rotation, and not the connecting corridors. This representation enables the user to see how the model is rotated, and thus provides the minimum necessary interaction feedback, but does not contain the most important information required to interpret the maze.

2.2 Experimental Setup

The experiment is a repeated-measures within-subject design, with the 4 user interface (UI) conditions mono-no motion, stereo-no motion, mono-motion, and stereo-motion. We also tested bimanual (mouse-keyboard) control by adding a fifth UI condition, in which rotation is done using the cursor keys. Though the results of this condition are not discussed in this paper, this means the experiment requires the user to switch input devices. We tried to reduce the effects of any possible confusion by presenting a salient warning to the user before the trial where an input device switch is about to occur.

The experiment involves a total of 30 trials per subject:

10 easy trials, 10 medium trials, and 10 hard ones. It is designed to take only a short time (15-25 minutes), which facilitates the recruitment of unpaid subjects. Training consisted of a 5-minute interactive tutorial, and presenting the easy trials at the beginning. Users were presented with an estimate of how much time a trial is expected to take, and were told to expect to make a few errors, but nevertheless "go as fast as you think you can go without making errors".

The order in which the conditions were presented was randomised and counterbalanced. Each condition was presented twice in a row to minimise switching between conditions. After all conditions were cycled through, the cycle was repeated with the next difficulty level. If we call the five UI conditions A,B,C,D,E, a typical session looks like this:

tutorial	— easy —	— medium —	— hard —
abcde	**cceeddaabb**	**ddaabbccee**	**bbcceeddaa**

(tutorial is fixed)	(randomly shuffled)	(counter-balanced)	(counter-balanced)

Unpaid subjects were recruited through mailbox or personal invitations. They were recruited from university staff members of the medical faculty of Leiden University Medical Center (LUMC), and of the computer science faculties of University of Twente (UT) and Amsterdam (UVA).

The experiment is conducted through a Java applet started from a webpage, and stereoscopy was achieved using cardboard anaglyph (red/cyan) glasses, which we found to compare favourably to available alternatives. In order to avoid bias due to possible negative effects of looking through coloured glasses, we asked users to keep the glasses on during the entire experiment including the non-stereo conditions.

Users were asked to perform the experiment at their own desks at the office, at any time that suited them. This enabled us to recruit a large number of subjects from different locations more easily. To compensate for lack of experimental control, a number of measures were taken, following [9] and [7]. To compensate for the lack of observers, we recorded the user interactions at a high frame rate (40 frames per second) so that each entire session could be played back. We analysed the sessions and discarded any obviously problematic trials. Users could be interrupted more easily than in a lab setting, but interruptions could be easily detected due to the high frame rate. As regards the variability of hardware and physical environment, we should note that the within-subjects design mitigates between-subject variability. Additionally, the experiment was designed so that it would run at more than 20 frames per second on almost any machine. As regards the red/cyan glasses, we tested their effectiveness by means of a random dot test at the beginning of each session. Prior to the experiment, we tested the glasses on typical office screens found at the faculties. We found that the glasses had very little ghosting on most standard desktop screens. We also found that many laptop screens did have significant ghosting, so we asked users not to use a laptop, or otherwise test for ghosting first, using the images on the introductory page. We conducted a subjective survey after the experiment. We asked users whether they saw any ghosting (if they could identify what it is) and whether they found the stereo effect convincing or experienced any ill effects, such as eye strain or headache.

3. RESULTS

In total, 32 users completed the experiment. 3 users had not filled in their subjective surveys, and one had not filled in 3 of the questions. The missing results were treated as missing values in the statistical analysis. 11 of the users were from the LUMC medical faculty. Most users were higher education students (18.8%) or had finished their studies (68.8%, mostly PhD students). Age varied from 21 to 65 years old, with 25-30 the largest age group (50%), followed by 30-35 (19%). Only 25% of the users was female. To test the comparability of the medical and non-medical users, we correlated user performance against the presence or absence of a medical background, that is, whether users were from the LUMC or not. Correlation using Kendall's τ_b revealed no significant correlations with either time performance or error rate in general or under any specific UI condition ($p > 0.20$).

The user self-reports regarding the interaction were encouraging. Overall, only 4 users (14%) had one or more serious complaints that indicate stereo effect may be significantly compromised. As regards the animation, only 1 user found the animation seriously lacking in smoothness, while 23 found the animation quite smooth.

After the data collection phase was closed, all interaction recordings were checked for interaction failures that may warrant removal of particular trials. Some minor problems were identified, some of which warranted deletion of some trials. All in all, only 25 trials (3%) had to be deleted because of serious problems. The most common cause for deletion of invalid trials was the user accidentally clicking twice on the "continue" button so the trial was over before it really began. While this is a usability problem that could have been solved easily enough, only 10 trials (1%) had to be deleted because of this. As regards interruptions and frame rate problems, we found little. We deleted only one trial due to a possible interruption, and no trials due to hiccups or low frame rate.

We will now discuss the effects of UI condition on performance over all users. We use $\alpha = 0.05$ as significance level unless otherwise stated. We will first examine the error rates. One or more errors were made in 26.9% of all trials. The difference in error rate between users was huge, varying from 2 trials with errors (7%) to 25 (83%). If we take into account that multiple decisions were made in one trial (one trial contains an average of 3.57 junctions along the path from one endpoint to the other; and users mostly just missed some of them), the percentage of errors per decision is approximately 9.0%.

We tried to obtain a high statistical sensitivity by making use of chi-squared distribution. If we assume that trials are independent events, we obtain a contingency table by categorising each trial within the entire experiment as either being with or without errors. We obtain the following table:

condition	trials w/ errors	trials w/o errors
mono-nomotion	51	132
mono-motion	40	150
stereo-nomotion	44	140
stereo-motion	54	134

The corresponding χ^2 is 3.8, so $p = 0.284$, so there is no significant relationship of condition with error rate.

Since UI condition was not significantly related to error rate, we exclude errors in the UI condition analysis from this point, concentrating on time performance only. The most careful approach is deleting all trials in which errors

were made. This means finding a way to replace the missing 26.9% of the data. The safest replacement is replacing a missing value by the overall mean performance of the user. However, we found that replacing all of the missing data in this way decreases statistical power unacceptably. The alternative we used is that we also use the performance times of all trials in which only one error is made. We consider this a valid approach because we found that these one-error cases all involved missing a junction (apparently, they were just slips), while most of the task was completed successfully. Additionally, we deleted 2 users who made more than one error in more than 33% of the trials. The remaining missing values (only 5.3% of all remaining data) were replaced by the user mean performances.

We used a 3x5 (difficulty x UI condition) repeated-measures ANOVA. The data was found to be non-spherical, so standard Greenhouse-Geisser sphericity adjustments were used. Post-hoc analyses were done using Bonferroni-corrected pairwise comparisons. The ANOVA revealed strong significance for both difficulty ($F(4, 112) = 19.9, \epsilon = 0.725, p < 0.001$) and UI condition ($F(2, 56) = 42, \epsilon = 0.404, p < 0.001$). The post-hoc analysis of UI conditions divided them into three groups: monoscopic, no motion (slowest), stereoscopic, no motion (in-between), and (mono/stereo) motion (fastest). The differences were highly significant ($p < 0.001$), except the stereoscopic improvement under the no-motion condition, which was significant with $p = 0.023$. The average time performance over all users also shows a very small difference between mono and stereo when motion is present. See the table of averages below.

Condition:	mono nomotion	stereo nomotion	mono motion	stereo motion
zero errors, missing values not counted:				
	16.15 sec.	15.00	12.47	12.20
zero or one error, missing values replaced:				
	16.81 sec.	14.16	11.79	11.68

We conclude that motion cue is more important than stereoscopy, and stereoscopy has added value only if motion is not present.

4. CONCLUSIONS

We executed a 3D interpretation experiment involving solid shaded maze-like structures. We examined the relative user performance (time taken and error rate) under the presence and absence of motion and stereoscopy cues. Our main conclusion is that both motion and stereo cues are effective at improving time performance. Motion has significantly more benefit, and adding stereo when motion is present does not have added value. This result is not consistent with previous 3D interpretation experiments, which claim an additive effect of stereo and motion. This inconsistency may be explained by the extra available 3D cues (shading and occlusion), the lower difficulty level of the task, the fact that time was measured rather than error rate, or the relatively large amount of control given to the users. We believe this outcome indicates limitations of the studies made so far, and warrants further study of the visual cues with a more systematic regard for experimental parameters.

5. REFERENCES

[1] K. W. Arthur, K. S. Booth, and C. Ware. Evaluating 3d task performance for fish tank virtual worlds. *ACM Trans. Inf. Syst.*, 11(3):239–265, 1993.

[2] R. Bade, F. Ritter, and B. Preim. Usability comparison of mouse-based interaction techniques for predictable 3d rotation. In *5th international symposium on smart graphics: SG 2005*, pages 138–150, Heidelberg, Germany, 2005. Springer.

[3] W. Barfield, C. Hendrix, and K.-E. Bystrom. Effects of stereopsis and head tracking on performance using desktop virtual environment displays. *Presence: Teleoper. Virtual Environ.*, 8(2):237–240, 1999.

[4] M. F. Bradshaw, A. D. Parton, and A. Glennerster. The task-dependent use of binocular disparity and motion parallax information. *Vision Research*, 40(27):3725–3734, 2000.

[5] K. S. Hale and K. M. Stanney. Effects of low stereo acuity on performance, presence and sickness within a virtual environment. *Applied Ergonomics*, 37(3):329–339, 2006.

[6] A. Hill. *Withindows: A Unified Framework for the Development of Desktop and Immersive User Interfaces.* PhD thesis, University of Illinois at Chicago, 2007.

[7] K. O. McGraw, M. D. Tew, and J. E. Williams. The integrity of web-delivered experiments: Can you trust the data? *Psychological Science*, 11(6):502–506, 2000.

[8] F. Mikšícek. Causes of visual fatigue and its improvements in stereoscopy. Technical Report DCSE/TR-2006-04, University of West Bohemia, 2006.

[9] U.-D. Reips. Standards for internet-based experimenting. *Experimental Psychology*, 49(4):243–256, 2002.

[10] R. L. Sollenberger and P. Milgram. Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors*, 35(3):483–499, 1993.

[11] C. Ware and G. Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Trans. Graph.*, 15(2):121–140, 1996.

[12] C. Ware and P. Mitchell. Visualizing graphs in three dimensions. *ACM Trans. Appl. Percept.*, 5(1):1–15, 2008.

[13] C. D. Wickens, S. Todd, and K. Seidler. Three-dimensional displays: Perception, implementation, and applications. Technical Report CSERIAC 89–001, Wright-Patterson Air Force Base, 1989.