Experimental Studies of Stereoscopic Vision for Command and Control Operations

B. Wolfe[1*], B. Kim[1], B. Aeschliman[1] and R. Sedlmeyer[1]

[1] *Department of Computer Science, Indiana University-Purdue University Fort Wayne, Indiana, United States*

Abstract:

3D stereoscopic vision is used in many applications, but the level of benefit to the user differs depending on the particular application. We studied its benefits for command and control applications such as battlefield visualization or disaster response. We conducted experiments where the subjects completed some simple military planning exercises both with and without 3D vision. 3D users had lower error in general, particularly when judging line of sight between two points. Furthermore, survey results show that most subjects preferred 3D. However, 3D took longer for completing the tasks. We also compared two ways of rendering military symbols in the virtual environment, finding that billboard style visualization boosted the subjects' efficiency when compared with draping the symbol on the terrain.

1. Introduction

Command and control (C2) software displays information to commanders about the units under their control, the locations of those units, and other information that may be relevant for the current situation. This information is typically displayed on 2-dimensional screens or monitors. However, the locations of military units, points of interest, and even the terrain of the battlefield itself is naturally 3-dimensional data. One method of displaying 3D information on a 2D device is to project the 3D information into the 2D plane. This is called a 2.5D display. However, the projection into 2D loses depth information, which can make it difficult for the user to estimate depth from the 2.5D display (e.g., determining the slopes of the mountains in Fig. 1).

---

* Corresponding author: Department of Computer Science, Indiana University-Purdue University Fort Wayne (IPFW), 2101 E. Coliseum Blvd., Fort Wayne, IN, 46805, USA, phone: 1-260-481-6867, E-mail: wolfeb@ipfw.edu

In contrast, a 3D vision system renders the data from two slightly different perspective angles so that a user actually perceives objects in three dimensions by using stereoscopic vision. Stereoscopic displays are a natural fit for geospatial information such as battlefield locations. Our work studies the strengths and weaknesses of stereoscopic vision specifically for military planning tasks with C2 systems. This study also evaluates two different ways of presenting military symbols for the C2 operations and surveys users' preferences about visualization techniques.

2. Background and Motivation

3D stereoscopic devices have become increasingly prevalent in the consumer market and benefit users in various ways. They have been utilized in a variety of areas such as simulation, training, entertainment, education, physical sciences, geography, and medicine [1-5].

3D stereoscopic environments provide depth perception advantages, but there are known limitations that prevent users from accurately perceiving depth. The lack of natural depth cues and differences between the user's actual convergence in reality and the viewer's convergence on the screen can make depth perception difficult [6]. Researchers have studied methods for enhancing depth perception and reducing visual fatigue when using 3D vision technology [7]. Previous experimental studies have produced mixed results, showing a general trend of underestimation in depth perception in 3D environments [1,8-10]. However, another study showed 3D stereoscopic visualization to be helpful for egocentric distance estimation during robot teleoperation [11]. Another study showed that users' depth perception accuracy varied not only when using different graphics effects, but also when using different zooming levels of virtual spaces [12].

When compared with a standard 2D interface, 3D interfaces projected into 2D are sometimes beneficial [13] and are sometimes detrimental [14] to the usability of the system [15]. The existing mixed results on 3D and 2.5D interfaces underscore the importance of systematically studying the strengths and weaknesses of 3D and 2.5D displays in the specific case of C2. Conducting usability tests to measure users' preferences and performance with 3D vision technology is an important precursor to exploiting 3D in practical C2 operations.

3. Experiments

We used a basic C2 application called C2VE that visualizes real geographic regions (Fig. 1) and allows us to compare 2D, 2.5D, and 3D displays [16]. This study investigated users' efficiency and their acceptance of 3D vision over traditional 2D and 2.5D displays. C2VE combines a satellite image with an associated height map to display the terrain of a battlefield. One of the primary features of C2VE is the sense of depth that the user experiences when using the 3D system. Of course, this 3D experience cannot be adequately conveyed in a paper. For example, from the 2.5D images in Fig. 1, it is difficult to discern the slope of the mountains. Moving the camera in a 2.5D environment can help with perceiving the slopes, but excessive movement could be disorienting. Instead, 3D lets users perceive the slope (and other depth information) even without moving the camera.
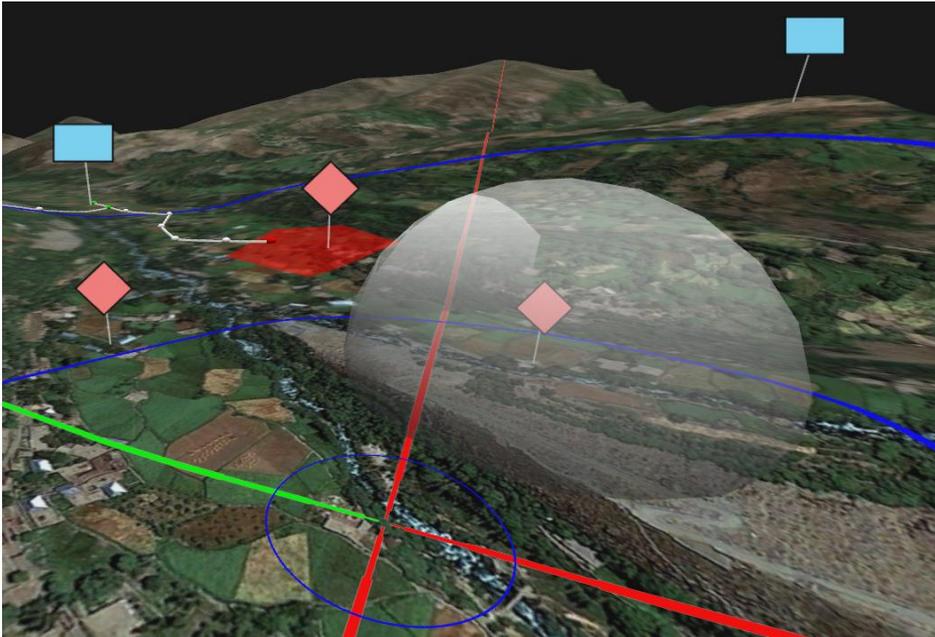
Fig. 1: Screen shot of C2VE software

In 2.5D and 3D modes, the user can fly the camera around the environment using a 3D Space Navigator, a 6-degree of freedom joystick about the size of a computer mouse. The symbols in the environment can be displayed as billboards or draped on the terrain (Fig. 2). The 2D mode is similar to a standard paper map in that the elevation is not perceptible except through contour lines. The user simply sees a bird's eye view of the aerial imagery. The contour lines were generated from the `contourc` function of GNU Octave [17]. The user can still zoom and pan the map using the Space Navigator, but the camera is limited to a bird's eye vantage point.

Our experiments included four kinds of tasks.

1. Training: The subject is asked to navigate the camera to certain locations and to place symbols in certain locations. This gives them a chance to acclimate to the experimental environment.
2. Line of Sight: The subject is asked if certain points on the map are visible from several observation posts that are marked on the map.
3. Check Routes: The subject is shown a map with several routes and asked to determine if each route is viable or not. They also mark down their certainty
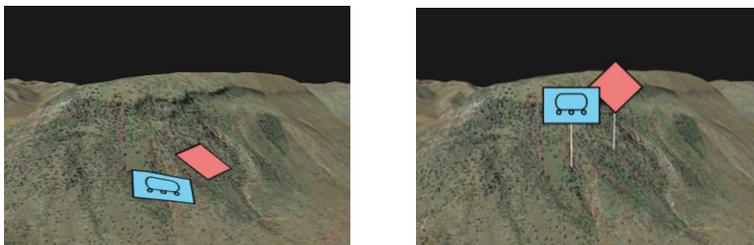

Fig. 2: Draped (left) and billboard (right) symbols

Tab. 1: Subject information

Age →	18–21	22–29	30–39	40–49	50–59	60+	*Total*
Military experience		5	1	1		1	8
No military experience	9	6	1				16
No response	6	1					7
Total	15	12	1	2		1	31

in their answers and highlight parts of the routes that are problematic.
4. Make Routes: The subject is shown pairs of starting and ending points and is asked to lay out the best route between each pair of points.

There were 31 subjects who completed the experiment. We recruited subjects from U.S. Army ROTC cadets, local National Guard personnel, and the campus community. Tab. 1 breaks down the subject pool by age and military experience. Each subject was randomly assigned into one of twelve experimental groups. The groups differ by the two display depths that the subject used (selected from 2D, 2.5D, and 3D), and whether they used draped or billboard symbols throughout the experiments.

The experiment session is organized into 3 sets of tasks, each of which uses a different map. Each set has 1 or more parts, where part 3.2 denotes set 3 part 2. Set 1 was always conducted in 2.5D, whereas sets 2 and 3 used the subject's two randomly assigned display depths. The experiment parts are as follows: 1.1: Training; 2.1: Training; 2.2: Line of Sight; 2.3: Make Urban Route; 2.4: Check Routes; 2.5: Make Routes; 3.1: Training; 3.2: Line of Sight; 3.3: Check Routes; 3.4: Make Routes. Each Line of Sight part consisted of 12 questions, 6 from each of 2 observation posts. The Make Urban Route part asked the subject to build a route through a flat, urban area, with recent improvised explosive device (IED) activity marked on the map. Each Check Routes part included 8 routes, two near a river and the others through mountainous terrain. The Make Routes parts involved one route around a ridge and another to get across a river. The routes tasks were developed in consultation with a captain in the U. S. Army who recently served as a convoy commander in Afghanistan. He provided the correct answers for the Check Routes tasks, and he graded the routes that the subjects built. Furthermore, he provided the training manuscript for advising the subjects about important considerations for viable routes

We used NVIDIA 3D Vision glasses (active shutter technology) with a 22 inch 3D-ready computer monitor (1920x1080 resolution). The subjects began the experiment by viewing training videos that described how to use the software and some factors to consider when choosing a convoy route (e.g., steepness of the terrain, sharpness of turns). Each subject could earn a small monetary bonus (up to $10) for correctly completing the tasks, which provided incentive to complete the tasks to the best of their ability. After finishing the tasks, each subject completed a survey about the software.

4. Analysis

We examined the experiment data to determine the effects of display depth (Sections 4.2 and 4.3) and symbol type (Section 4.4). The user survey results are detailed in Section 4.5.

4.1. Metrics

We use three metrics to evaluate the subjects' interaction with the system. The first is the amount of time (in minutes) that it took the subject to complete each part of the experiments.

The second is the error score. For the Line of Sight and Check Routes parts, the user answered yes/no questions and indicated their confidence in each of those answers (none, little, some, or a lot). The error score is 0 for a high-confidence correct answer, medium for a low-confidence incorrect answer, and high for a high-confidence incorrect answer. Specifically, each of the 8 possible responses was mapped to an integer score from –4 to 4. "No" answers were negative and "yes" answers were positive. The level of confidence determines the magnitude of the score, with a lot of confidence being ±4 and no confidence being ±1. The user's error for a question is the distance between the correct score and the user's score. The correct answers had a lot of confidence, except for a few borderline routes where the best answer is to indicate lower confidence. The user's error for an experiment part is their average error for all the questions in that part.

For the Make Routes parts, each route that the user constructed was given an error score from 0 to 10. A former convoy commander in the U. S. Army graded the users' routes based on their efficiency, likelihood of rollover, and susceptibility to enemy attacks. The user's error for each Make Routes part is the average error for the routes constructed in that part.

The third metric is the amount of camera and cursor movement. The overall movement score combines three quantities: linear camera movement, measured by Euclidean distance; camera rotation, measured by the angular movement in pitch and yaw (roll was disabled); and cursor movement, measured by the Euclidean distance, discounting elevation. Each quantity was standardized so that it had mean 0 and standard deviation of 1 across all subjects and tasks. Then the standardized quantities were added together to get the overall movement score. Negative movement scores are good, indicating less movement than average.

4.2. Display Depth: Change in Individuals' Performance

In this section, we analyze how the performance of each subject differed with the two display depths they used (i.e., one depth in set 2 and a different depth in set 3). We found that 3D leads to significantly lower error than either 2D or 2.5D, but it also takes significantly more time than either 2D or 2.5D. The rest of this section details our methods for arriving at these conclusions.

To account for differences in the tasks for set 2 and set 3, we transformed each subject's scores into their empirical cumulative distribution function (ECDF) values. The ECDF value for a score s is the fraction of subjects who scored at or below s on that experiment task. For example, if a subject X has an ECDF score of 0.6 for their time in part 2.3, it means that 60 % of the subjects completed that experiment part in less time or the same time as subject X.

To compare each subject's performance on their set 2 and set 3 tasks, we measure the change in their ECDF values. For example, if subject X had an ECDF score of 0.2 for their error in the route checking task where they used 3D and an ECDF score of 0.4 for their error in the route checking task where they used 2D, then that provides some evidence that 3D leads to lower error than 2D. The key quantity is the difference in the ECDF scores, particularly the sign of that difference. Fig. 3 shows histograms of the
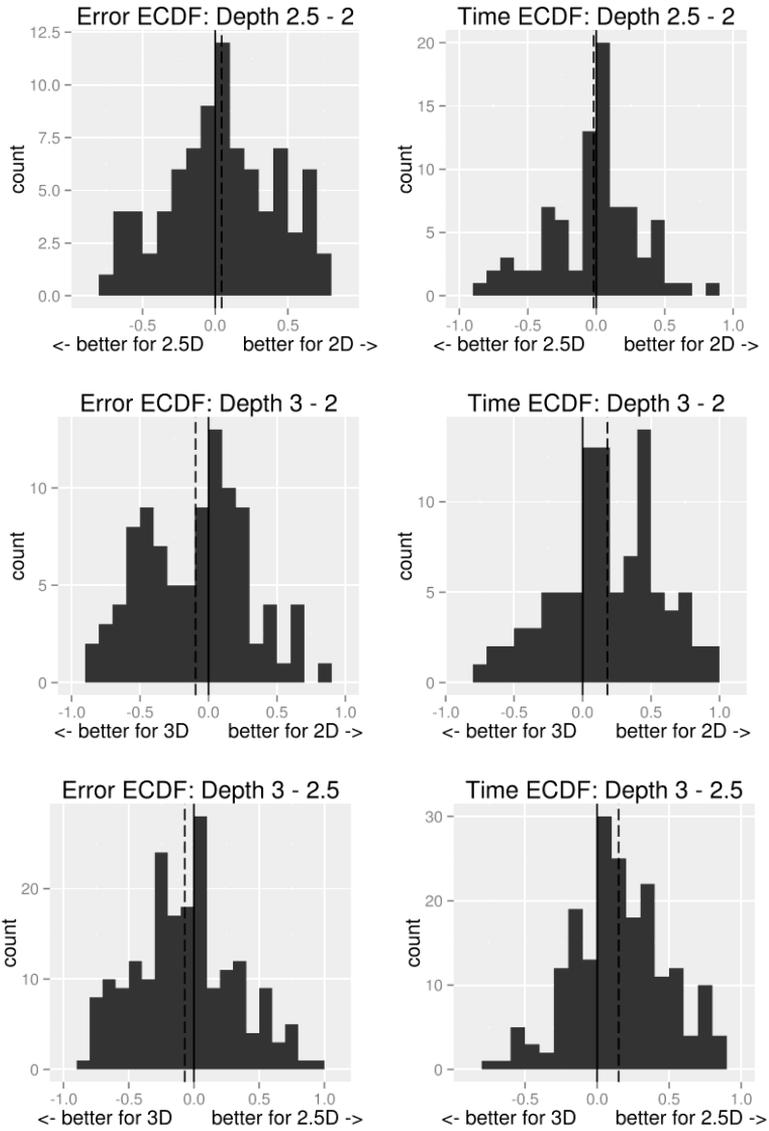
Fig. 3: The distributions of ECDF differences. The solid vertical line is a difference of 0.0 and the dashed vertical line indicates the mean difference in ECDF.

ECDF differences for the error and time metrics and each pair of display depths. Histograms that have a mean near 0.0 and are fairly symmetric indicate little difference between the two display depths. However, when most of the mass of the distribution lies on one side or the other of 0.0, that indicates that the two display depths lead to different results.

Tab. 2: Wilcoxon signed rank test results, comparing each pair of display depths using the differences in ECDF for all task types. D1 and D2 are the two display depths being compared. n is the number of data points used in the comparison. "Frac. D1 Better" is the fraction of data points were D1 was better than D2. "Adj. p" are the p values after applying Holm-Bonferroni to each metric.

Metric	D1	D2	*n*	Frac. D1 Better	Test Stat.	*p*	Adj. *p*
Error	2.5	2	84	0.440	1887.5	0.28619	0.28619
Error	3	2	96	0.542	1587.0	0.02194	**0.04388**
Error	3	2.5	192	0.568	6721.5	0.00992	**0.02976**
Movement	3	2.5	189	0.455	9022.0	0.47954	0.47954
Time	2.5	2	84	0.452	1582.0	0.85720	0.85720
Time	3	2	96	0.271	3503.5	0.00002	**0.00004**
Time	3	2.5	192	0.292	12160.5	0.00000	<**0.00001**

To quantitatively inspect if the ECDF differences were significantly different than 0.0, we ran a Wilcoxon signed rank test [18] for each error metric and pair of display depths. The signed rank test essentially checks for a consistent tendency of the data to be greater than 0.0 or a consistent tendency of the data to be less than 0.0. Statistical tests were performed using the R software package [19]. We corrected for multiple comparisons by using the Holm-Bonferroni method [20] to adjust the *p* values from the Wilcoxon signed rank tests to keep the familywise error rate (FWER) below α=0.05 for each metric. The FWER is the probability of making one or more false discoveries, so limiting the FWER to 0.05 means that there is less than 5 % chance that we incorrectly conclude that display depth significantly affects user error, e.g. A FWER bound of 0.05 is stronger than a 5 % error bound on each individual comparison, because the latter leads to a higher probability of at least one false discovery.

The results are in Tab. 2. For the error metric, 2.5D compared with 2D did not show a significant difference, whereas 3D is significantly different than both 2.5D and 2D. In both cases, 3D leads to less error on average. For the movement metric, only the 2.5D versus 3D comparison is valid (because movement is completely different in 2D), but it is not statistically significant. For the time metric, the 2.5D and 2D comparison did not show a significant difference. However, 3D was significantly different than both 2D and 2.5D. In each case, 3D took more time. This matches with subjects' survey responses (Section 4.5).

4.3. Display Depth: Differences in Group Performance

Like the previous section, this section examines the impact of display depth on users' performance. However, this section explores the data from a different perspective, providing a complementary analysis. One of the differences is that the previous section examined the differences between each individual's performance on their two display depths, while this section examines differences among the groups who used the different display depths (i.e., a "within subjects" versus "between subjects" comparison). The individual analysis is stronger, since it accounts for differences in the subjects' abilities. However, when looking at groups (i.e., in this section), there is enough data to examine the results of each experiment part individually. This lets us determine what types of tasks gain the most benefit from 3D.
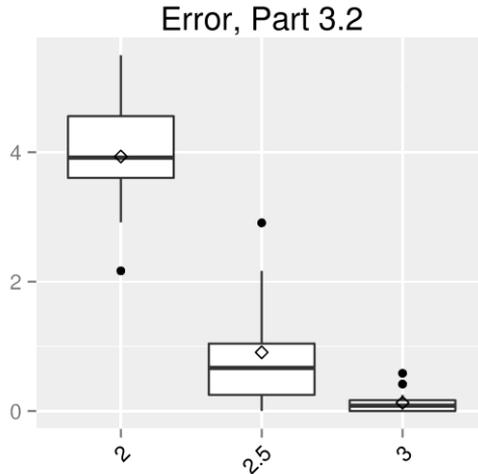
Fig. 4: Error for the set 3 line of sight tasks, broken down by display depth.
3D leads to less error than 2.5D, which is less than 2D. The diamond
indicates the mean value.

Tab. 3: Mann-Whitney U test results for differences in the part 3.2 error based on display depth.
"Adj. p" are the p values after applying Holm-Bonferroni. All differences are significant.

Mode 1	Mode 2	n	n: Mode 1	n: Mode 2	Test stat. (W)	p	Adj. p
2	3	21	8	13	104.00	0.000157	**0.000472**
2	2.5	18	8	10	78.50	0.000702	**0.001405**
2.5	3	23	10	13	112.00	0.003352	**0.003352**

For most metrics and experiment parts, the 3D group of users was neither worse nor better than the 2D group or the 2.5D group. However, for the latter Line of Sight tasks (i.e., part 3.2), the 3D group was significantly better than the 2.5D group, which was significantly better than the 2D group (Fig. 4). The rest of this section describes the methods we used to come to that conclusion.

For each metric and each experiment part, we ran a Kruskal-Wallis test [21] to determine if the different display depths produced significantly different results. Kruskal-Wallis is an alternative to a one-way ANOVA test that makes fewer assumptions about the data than ANOVA. As before, we use Holm-Bonferroni to keep the familywise error rate below α=0.05 for each metric. When looking at the adjusted p values, only the error for experiment part 3.2 shows a statistically significant difference among the display depths at the 0.05 significance level. That comparison included 8, 10, and 13 subjects for 2D, 2.5D, and 3D, respectively, resulting in a χ^2 value of 21.9 and an adjusted p value of 0.000123.

While the Kruskal-Wallis test indicates that display depth significantly affects the error on part 3.2, it does not specify which particular depths are better. Thus, we ran
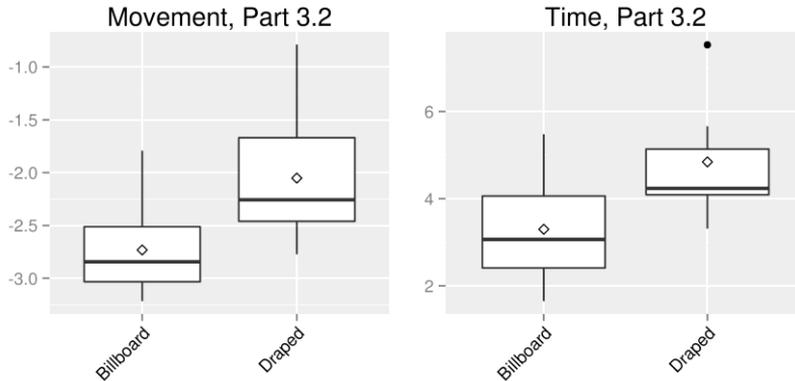
Fig. 5: Movement and time for the set 3 line of sight tasks, broken down by symbol type. Billboard symbols result in significantly less time and less movement. The diamond indicates the mean value.

Tab. 4: Mann-Whitney U test results for the impact of symbol type on subjects' performance. "Adj. p" are the p values after applying Holm-Bonferroni. Significant p values are shown in bold.

Metric	Part	n	n: Billboard	n: Draped	Test stat. (W)	p	Adj. p
Error	2.2	24	12	12	94.50	0.200564	0.401128
Error	3.2	23	12	11	72.00	0.730594	0.730594
Movement	2.2	23	12	11	61.00	0.785949	0.785949
Movement	3.2	23	12	11	20.00	0.003560	**0.007121**
Time	2.2	24	12	12	64.00	0.670660	0.670660
Time	3.2	23	12	11	28.00	0.018795	**0.037590**

pairwise Mann-Whitney U tests [22] to check for significant differences between each pair of display depths: 2D vs. 3D, 2D vs. 2.5D, and 2.5D vs. 3D. We used Holm-Bonferroni to keep the FWER below $\alpha=0.05$. The results are in Tab. 3. All of the pairwise differences are statistically significant. This matches the intuition from looking at the boxplot (Fig. 4), which clearly shows that 2D error is higher than 2.5D, which is higher than 3D.

The finding that 3D error is lower than 2D or 2.5D on the Line of Sight tasks is consistent with the results from the individual analysis (Section 4.2), where we found that, when looking at all the tasks in aggregate, 3D error is lower than 2D or 2.5D.

4.4. Symbol Type

In addition to looking at the impact of display depth, we also examined how the symbol type—billboard or draped—impacts each of the three metrics. We only consider the Line of Sight tasks with 2.5D and 3D depths, because they were the only tasks that relied upon the symbol type. For each metric and each experiment part, we

ran a Mann-Whitney U test [22] to determine if the symbol type had a significant impacted on the results. As before, we used the Holm-Bonferroni method [20] to keep the FWER below $\alpha=0.05$ for each metric. Tab. 4 shows the results of those tests.

When looking at the adjusted p values, one can see that both time and movement were significantly impacted by the symbol type during the set 3 tasks. The boxplots for that data (Fig. 5) illustrate that the draped symbols take significantly more time and require more camera/cursor movement by the user.

4.5. User Survey

At the end of the experiment session, each user filled out a survey with several questions about the software, including questions about the different display depths and about side effects during 3D usage.

4.5.1. Display Depth Questions

One of the survey questions asked the user to indicate which of the two display depths that they used was better for completing their tasks. A follow-up question asked them to explain the reasoning behind their choice. For those qualitative explanations, we identified the themes that occurred across several subjects' responses.

All 15 of the subjects who used 2D said that it was the worse option. The overwhelming theme in their explanations was that the other mode (2.5D or 3D) let them see the elevation and the terrain. In fact, 14 of the 15 subjects (93 %) explicitly mentioned something about seeing the terrain better or "more realistically" in 2.5D or 3D.

The remaining 16 subjects used 2.5D and 3D, with 12 out of 16 subjects (75 %) saying that 3D was better. Breaking down those subjects by age, 8 out of 10 subjects (80 %) age 18-21 preferred 3D, while 3 out of 4 subjects (75 %) age 22-29 preferred 3D, and 1 out of 2 subjects (50 %) 30 or older preferred 3D.

When looking at the explanations for why subjects preferred 3D versus 2.5D, a few themes emerged. Out of the 12 subjects who preferred 3D over 2.5D, 11 of them (92 %) listed a better understanding of the environment or the terrain as a reason for preferring 3D. The other subject's explanation was similar, but very brief: "because we can see the depth and it is more real." In addition to generally being able to understand the environment better, 5 of the 12 subjects specifically mentioned that 3D was beneficial for route-related tasks, while 4 of the 12 subjects specifically mentioned that 3D was beneficial for line of sight judgments.

Here is an example response from a subject with 9.5 years service in the army, explaining why he preferred 3D to 2.5D: "One of the most difficult aspects of map reconnaissance, even using a 2.5D terrain viewer with a map overlay for a map reconnaissance, is getting an appreciation for the scope and scale of your surroundings when actually on the ground. Both the 2.5D and 3D give a hint of what the ground will actually be like, but the 3D view gives your imagination extra variables to assist in visualizing details of scale (scrub brush, boulders, crags, etc.) that can't be portrayed in this setting."

Here is another example response where the subject touches on both the strengths and weaknesses of 3D: "The 3D view is extremely helpful when compared to the 2.5D in terms of determining depth and steepness of travel routes. The 3D is also slightly better in helping determine the point at which a mountain begins to obstruct view as in

the case when determining whether or not you could see places from the posts. 2.5D is better at viewing places close to the camera, though."

Two other subjects who preferred 3D also mentioned that it was difficult to see things clearly near the camera. One interesting direction for future work is to investigate this issue, studying the relationship between users' perceptual comfort level and convergence, camera separation, and the viewing volume along the screen depth.

The explanations of the four subjects (out of 16) who preferred 2.5D over 3D are noteworthy, so their complete responses about why they preferred 2.5D are listed here:

1. Although depth and perception is better in 3D, I find it faster and easier to understand and compare in 2.5D. It is easier to compare to real life.
2. It does provide enhanced perception of distance and field of view, I don't think much is gained when considering what you can put together by using 2.5D view.
3. Lower learning curve. More forgiving of oversteer from the controller.
4. Less flashing of the glasses. I didn't need to adjust my eyes and they didn't have any pain.

The first two subjects acknowledge that perception is better with 3D, but mention that 2.5D is either faster (subject 1) or close enough to 3D (subject 2). Subject 3 mentioned that it was harder to adjust to 3D; he was our oldest subject, over 60 years old, with 40 years experience in the army. Subject 4 mentioned discomfort from the 3D, an important consideration that we examined in separate survey questions (Section 4.5.2). Some of the preference toward 2.5D is simply due to familiarity with that technology. In fact, 2.5D displays are standard in computing today, whereas many users have had limited experience with 3D. This makes the 75 % preference rate for 3D even more noteworthy.

In another survey question, we asked the 3D users to rate the degree to which 3D helped them perceive the height and depth of the environment. A large majority (19 out of 24, 79 %) agreed that 3D helped with both height and depth, a few subjects expressed neutral opinions, and only one user disagreed that 3D helped with distance estimation (Fig. 6).

4.5.2. Side Effects from 3D
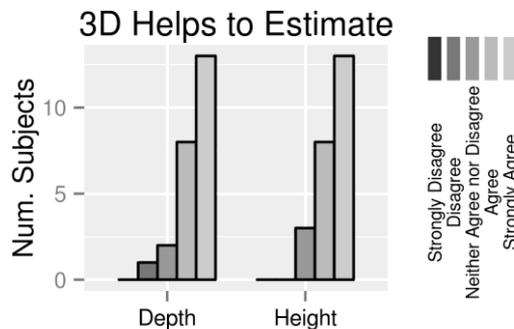The 24 subjects who used 3D reported their levels of dizziness, eye strain, and
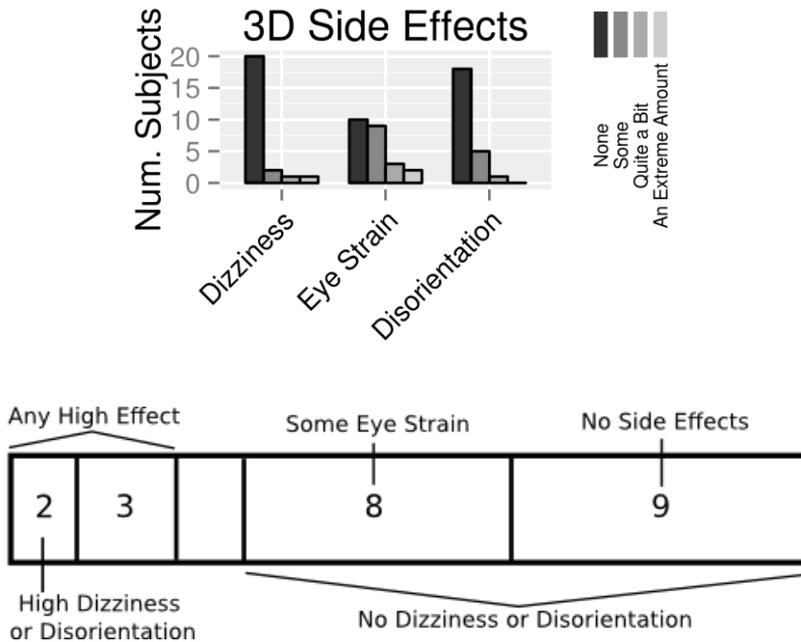

Fig. 6: Survey results

Fig. 7: Side effects: survey results from the 24 subjects who used 3D.

disorientation. Fig. 7 plots the responses for each side effect separately (top) and partitions the 24 subjects into several groups based on the side effects they reported (bottom). Only 2 subjects (8 %) reported "Quite a Bit" or "An Extreme Amount" of dizziness or disorientation. An additional 3 subjects reported "Quite a Bit" or "An Extreme Amount" of eye strain. Those 5 subjects (21 %) were the only ones to report more than "Some" of any side effect. In fact, most of the subjects (17 out of 24, 71 %) reported no dizziness or disorientation. Of those, 8 subjects reported some eye strain, with the other 9 (38 % overall) reporting no side effects.

Only 3 of the 24 subjects who used 3D were age 30 or above, so there is not enough data to draw strong conclusions about a correlation between age and side effects. Of those 3 subjects age 30 or above, only one reported any side effects (an extreme amount of eye strain and some disorientation).

5. Conclusions

We studied two factors that could impact users' performance on a variety of small military tasks in a command and control environment: the display depth—2D, 2.5D, or 3D—and the type of symbol—billboard or draped. We examined three performance measures: error on the task answers, amount of camera and cursor movement, and the time to complete the task.

Billboard style symbols led to more efficient decisions than draped symbols, both in terms of camera/cursor movement and time spent on task, without sacrificing accuracy.

Our results provide statistically significant evidence that 3D stereoscopic vision is beneficial for reducing users' error when compared to either 2D or 2.5D. However, using 3D also took significantly more time than either 2D or 2.5D. These results from the data are corroborated by the subjects' responses to the survey questions, where 75 % of the subjects who used both 3D and 2.5D preferred 3D, and everyone who used 3D and 2D preferred 3D. When looking at specific types of tasks, we found that 3D produced significantly lower error on line of sight tasks when compared with 2D or 2.5D, and 2.5D was significantly lower than 2D. On specific types of tasks, none of the performance measures was significantly worse when using 3D compared to 2D or 2.5D. Only 8 % of 3D users reported a moderate or high level of dizziness or disorientation.

References

[1] GRECHKIN, T.Y; NGUYEN, T.D.; PLUMERT, J.M.; CREMER, J.F. & KEARNEY, J.K. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Trans. on Applied Perception*, 2010, vol. 7, Article No. 26.

[2] YANG, M.; MCMULLEN, D.P.; SCHWARTZ-BLOOM, R.D. & BRADY, R. Dive into alcohol: A biochemical immersive experience. *In* IEEE Virtual Reality Conf., 2009, p. 281–282.

[3] CHITTARO, L.; RANON, R. & IERONUTTI, L. Vu-flow: A visualization tool for analyzing navigation in virtual environments. *IEEE Trans. on Visualization and Computer Graphics*, 2006, vol. 12, p. 1475–1485.

[4] SCHWARTZ, R.J. & FLEMING, G.A. Real-time aerodynamic flow and data visualization in an interactive virtual environment. *In* IEEE Instrumentation and Measurement Technology Conf., 2005, p. 2210–2215.

[5] ZELLE, J.M. & FIGURA C. Simple, low-cost stereographics: VR for everyone. *In* The 35th SIGCSE Technical Symp. on Computer Science Education, 2004, p. 348–352.

[6] ELMQVIST, N. & TSIGAS, P. A taxonomy of 3D occlusion management for visualization. *IEEE Trans. on Visualization and Computer Graphics*, 2008, vol. 14, no. 5, p. 1095–1109.

[7] CIPILOGLU, Z.; BULBUL, A. & CAPIN, T. A framework for enhancing depth perception in computer graphics. *In* 7th Symp. on Applied Perception in Graphics and Visualization, 2010, p. 141–148.

[8] JONES, J.A.; SWAN II, J.E.; SINGH, G.; KOLSTAD, E. & ELLIS, S.R. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. *In* 7th Symp. on Applied Perception in Graphics and Visualization, 2008, p. 9–14.

[9] WARTELL, Z.; HODGES, L.F. & RIBARSKY, W. A geometric comparison of algorithms for fusion control in stereoscopic HTDs. *IEEE Trans. on Visualization and Computer Graphics*, 2002, vol. 8, p. 129–143.

[10] LIVINGSTON, M.A.; ZHUMING, A.; SWAN, J.E. & SMALLMAN, H.S. Indoor vs. outdoor depth perception of mobile augmented reality. *In* IEEE Virtual Reality Conf., 2009, p. 55–62.

[11] LIVATINO, S. & PRIVITERA, F. 3D visualization technologies for teleguided

robots. *In* ACM Symp. on Virtual Reality Software and Technology, 2006, p. 240–243.

[12] HARTZELL, T.; THOMPSON, M. & KIM, B. The influence of graphics effects on perceiving depth in minimized virtual environments. *Communications in Computer and Information Science*, 2012, p. 235–242.

[13] SUN, Y.; DING, N.; HAO, G. & SHI, X. The research and application of 2D and 3D interactive system. *In* 2nd International Conf. on Information and Computing Science, 2009, p. 252–254.

[14] COCKBURN A. & MCKENZIE, B.J. 3D or not 3D? Evaluating the effect of the third dimension in a document management system. *In* SIGCHI Conf. on Human Factors in Computing Systems, 2001, p. 434–441.

[15] ST. JOHN, M; COWEN, M.B.; SMALLMAN, H.S. & OONK, H.M. The use of 2D and 3D displays for shape understanding versus relative position tasks. *Human Factors*, 2001, vol. 43, p. 79–98.

[16] WOLFE, B.; PODGORNIY, D.; KIM, B. & SEDLMEYER, R. C2VE: A software platform for evaluating the use of 3D vision technology for C2 operations. *In* Proc. of the 17th International Command and Control Research and Technology Symp., 2012.

[17] OCTAVE COMMUNITY. GNU/Octave, 2012.

[18] WILCOXON, F. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1945, vol. 1, no. 6, p. 80–83.

[19] R CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[20] HOLM, S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 1979, vol. 6, no. 2, p. 65–70.

[21] KRUSKAL, W.H. & WALLIS, W.A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 1952, vol. 47, no. 260, p. 583–621.

[22] MANN, H.B.; WHITNEY, D.R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 1947, vol. 18, no. 1, p. 50–60.

Acknowledgements