The Effects of Scene Complexity, Stereovision, and Motion Parallax on Size Constancy in a Virtual Environment

Xun Luo[1]
Electronic Visualization Laboratory
University of Illinois at Chicago

Robert Kenyon[2]
Electronic Visualization Laboratory
University of Illinois at Chicago

Derek Kamper[3]
Bioengineering Department
Illinois Institute of Technology

Daniel Sandin[4]
Electronic Visualization Laboratory
University of Illinois at Chicago

Thomas DeFanti[5]
Electronic Visualization Laboratory
University of Illinois at Chicago

Abstract

In this paper, the effects of three visual factors: scene complexity, stereovision and motion parallax on correct perception of a virtual object's size were analyzed in an immersive virtual environment. We designed a controlled experiments set to incorporate visual conditions that reflected all twelve different configuration combinations of the three visual factors. Under each visual condition, subject performed the task of making judgments of the sizes of a virtual object displayed at five different distances from him/her. A total number of eighteen subjects participated in our study. The subjects' judgments and the corresponding actual sizes of the virtual object were recorded. Based on the colleted data, two quantitative measures of subjects' performance were derived and analyzed. The results of our experiments were consistent across the majority of the subject population and suggested that scene complexity and stereovision could have significant impact on the performance of a user of virtual environments to make correct judgments on a virtual object's size. On the contrary, motion parallax, either produced by the virtual environment or by the observer, might not be a significant factor in determining that performance.

Categories and Subject Descriptors

I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – *Virtual reality*.

Index Terms

Measurement, Human Factors

Keywords

Virtual Reality, Size Constancy, Visual Factors, Effects.

[1 2 4 5] 2032 ERF Building, 842 W. Taylor St., Chicago, IL 60607. {xluo1, kenyon, dan, tom} @ uic.edu
[3] 345 E. Superior Street, Chicago, IL 60611. d-kamper @ northwestern.edu

1 Introduction

Virtual Environments (VEs) are nowadays used for a large variety of research and commercial purposes, such as medical diagnosis, scientific data mining and industry manufacturing, just to name a few ([8], [14]). The effectiveness of VE in its applications relies heavily on its ability to create perceptions within the environment that faithfully replicate those in the physical world. However, due to limitations the VE can have a number of flaws that adversely affect its use and the credibility of the environments that it offers. One of the more significant aspects of this problem is whether the perceived size of an object in the VE is equivalent to that perceived in the physical world when object distance from the observer changes.

Many studies of perceived size of objects in the physical world have been performed. Descartes (1637) first described the phenomenon known as "Size Constancy" where an object is perceived as being the same size regardless of its distance from the observer even though the retinal size of the object gets smaller with increasing distance from the observer. Holaday ([9]) showed that removal of various cues would change this behavior to one relying on the physical optics of the situation. He showed that as the number of two-dimensional (2D) cues to depth [eg. Shadows, motion parallax, etc] is reduced performance suffers and subjects adopt a size judgment that is based on the visual size of the object on the retina also know as visual angle (VA) size judgments. Holway and Boring ([15]) confirmed these findings for objects from 10-40ft from the observer. Harvey and Leibowitz ([10]) showed similar results at distances of 1-9ft from the observer. Furthermore, they and Leibowitz and Dato ([11]) showed that removal of 3D cues to depth (i.e. Stereovision) had little to no effect on performance and that performance was only affected by the removal of 2D depth cues.

Unlike other electronic forms of visual display, VE can provide veridical size and distance cues to the user. In VE, both stereovision and 2D cues to depth (i.e., motion parallax, perspective, etc) can be made available. Therefore, one would expect similar size-constancy changes to those reported in the physical world. However, when Eggleston ([12]) reproduced the experiments in [15] using a head mounted display (HMD) their subjects showed no size-constancy but visual angle performance. That is, instead of the actual size of the object remaining the same regardless of its distance, the object size perceived by the subject changed with the distance of the object from the subject. Baitch and Smith ([13]) showed similar results for an object that was approximately 15 inches from the subject using a CAVE ([8])-like system that provided stereovision but few 2D cues to depth. However, we believe that these results are the consequence of either exceeding the visual limits of the VE or using a sparse environment that eliminated the 2D cues to depth that others have shown to be so important in this task in the physical world.

This research was initiated to measure the perception of object size when virtual objects are placed at different distances from the subject within the VE. The visual environment presented to our subjects in this experiment is one where the virtual object is viewed at different distances and then the subject adjusts the size of the virtual object until it becomes the correct size according to the subjects' perception. We studied the effects of three major visual factors on size-constancy: scene complexity, stereovision and motion parallax. Our results were similar to those performed in the physical world where size-constancy was more prevalent when the environment had a rich environment or stereovision was provided. As the richness of environment decreased and stereovision was removed most of the subjects adopted a visual angle performance. Results of our experiments also suggested that motion parallax, either created by the VE or the observer had little effect on size-constancy performance.

2 METHODS

2.1 Subjects

Eighteen subjects were tested (EC1-EC18). Nine were experienced in VE and had a minimum of 6 months of using immersive VEs; for the other inexperienced subjects, this was their first exposure to an immersive VE. All subjects were tested for visual acuity and stereo acuity. Only subjects with a corrected vision of 20/20 and normal stereo vision were included in our results.

2.2 Apparatus

All tests were performed using a single wall CAVE – the C-Wall (Configurable Wall). The C-Wall is a high-quality, head-tracked, active stereo wall, that displays an image before the viewer by means of a 10x10ft rear-projection screen. The back projector pointed to a mirror, which reflected the images onto the screen. To create stereoscopic objects, two off–axis perspective images are consecutively displayed; one visible to the right eye, the next to the left eye. The visibility of images by each eye is controlled by the stereo glasses (Stereographics, Inc. Beverly Hills, CA) which rapidly turn each lens on and off in synchrony with the corresponding images on the screen. A Pentium IV PC performed the image processing for the C-Wall. The image resolution was 1024x768 pixels with a refresh rate of 120 Hz and an update rate of 60 stereo images per second. Each subject's interpupillary distance (IPD) was measured (R.H. Burton Digital P.D. Meter, R.H. Burton LLC, Drive Grove City, OH) and incorporated into the CAVE program customize generation of generate the stereo images for each subject. A six-degrees-of-freedom camera tracking system (Eagle Digital System, Motion Analysis Corp., Santa Rosa, CA) provided real-time head position which was used to calculate the correct stereoscopic perspective projections for the C-Wall as the viewer moved his/her head. The head tracking system had a latency of 65 ms and was calibrated to an accuracy of $\pm$ 0.1 inches for the tracking distances used in these experiments. A cordless joystick (RamPad, Logitech Inc., Fremont, CA) held by the viewer provided interaction with the VE.

A virtual coke bottle textured with the image of a physical 2-liter coke bottle was drawn to test size perception. Different configurations of VE were presented in order to test the effects of scene complexity, motion parallax, and stereovision on perception of virtual object size. Figure 1 illustrates one of these configurations.

- Scene Complexity

Two types of environment were provided, either a rich environment (ENV) with many cues to depth or a sparse environment (No-ENV) with minimal cues to depth. The ENV consisted of a gray-green checkered floor with a wooden textured table in the scene; the coke bottle sat on top of the table. The table's height above the floor was randomly set at one of the three possible textures and three possible heights (30, 33 and 36 inches). For the No-ENV case, the environment consisted solely of a gray background. The virtual coke bottle was presented as being suspended in mid air at different heights from the floor (corresponding to the table heights) and at a number of different distances from the user as described in the previous section. The head was tracked identically to that described above.

- Stereovision

The effects of stereo vision on size perception were also tested. Two conditions were examined: monocular vision (MONO) and stereo vision (STEREO). For the MONO condition, the same image was presented to each eye. For the STEREO condition, disparate images were presented to the two eyes. Interpupillary distance was measured for each subject, and the images for the two eyes were created to reflect the different vantage points in order to evoke a stereo image.

- Motion Parallax

Three different motion parallax conditions were tested in this study: no motion parallax (No-MP), motion parallax generated by the VE (Passive-MP), and motion parallax generated by the viewer (Active-MP).

For the No-MP condition the subject was instructed to hold his/her head still and look straight forward with no lateral head movement. To ensure the subject was maintaining a static posture, the experimenter monitored the tracking readings in the lateral direction, and prompted the user whenever there were head movements greater than 1 inch, the minimum value needed to incur motion parallax.

For the Passive-MP condition, the whole scene displayed on the C-Wall moved in a sinusoidal fashion at 0.25 Hz. Peak scene displacement was 1 ft and peak velocity was 4 ft/sec. These parameter values were chosen to conform to natural human lateral movement in order to facilitate comparisons with active motion parallax ([2], [3]).

For the Active-MP condition the subject was instructed to move his/her head laterally from one side to the other at 0.25 Hz to a minimum displacement of 1 ft. The subject was provided with audio cues for proper movement frequency from an electronic metronome. The experimenter monitored lateral head movement through the tracker and prompted the subject whenever lateral movement fell below the desired level.

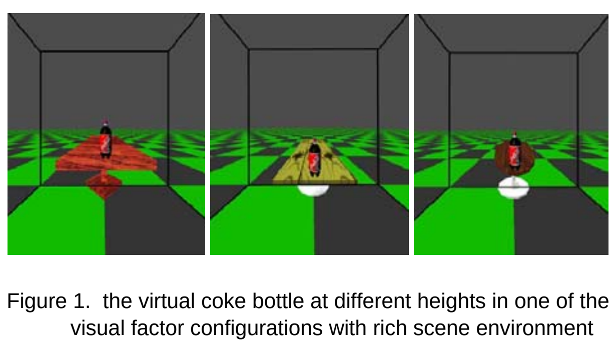

Figure 1. the virtual coke bottle at different heights in one of the visual factor configurations with rich scene environment

2.3 Experimental Protocol

The subjects were instructed to adjust the size of the virtual object (2-liter Coke bottle) so that they perceived the virtual object's size as being identical to that of a physical coke bottle if placed at same distance from the subject. To aid in this task, a physical 2-liter coke bottle was visible to the subjects for comparison to the virtual object. The 2-liter coke bottle was placed on a wooden stand covered with black cloth at a height of 3 ft. The stand was positioned at the front left hand side of the C-Wall at an approximate distance of 3.5 ft. from each subject. Both the physical and virtual coke bottle was 12 inches tall and 5.5 inches (maximum) wide. The physical coke bottle, lit by a standing spotlight, was visible to the subjects by simply turning their head 40° to the left.

The virtual coke bottle was displayed randomly at one of the five distances from the subject: 3.5, 5.0, 6.5, 8 and 9.5 ft. The subject sat 5 ft. from the C-Wall screen; thus, the virtual object could be located in front of, on, or behind the C-Wall screen. The computer randomly set the initial size of the virtual coke bottle from 0.2 to 3.0 times the normal size (12 inches) of the bottle. Subjects used the cordless joystick to increase and decrease the size of the virtual coke bottle to what they perceived to be the appropriate size for each trial. The head was tracked so the scene was updated appropriately to the position of the subject's head.

The independent variables of scene complexity, motion parallax, and stereovision had 2, 3, and 2 levels, respectively. Each condition was repeated 6 times for each bottle location for a total of 360 repetitions. To avoid ambiguity hereafter, we call each repetition of size judgments that was performed under the same configuration of the independent variables a *run*, and the consecutive block of runs a *trial*. Additionally, subjects performed an initial trial to familiarize themselves with the process. It could be seen that except for the initial trial, trials and visual factor configurations mapped one-to-one to each other. Table 1 shows this mapping relationship between trial IDs and visual factor configurations.

Table 1. mapping between trial IDs and visual factor configurations

Trial ID	Scene Complexity	Stereovision	Motion Parallax
T0	*Initial trial for familiarization*		
T1	No-ENV	MONO	No-MP
T2	No-ENV	MONO	Passive-MP
T3	No-ENV	MONO	Active-MP
T4	No-ENV	STEREO	No-MP
T5	No-ENV	STEREO	Passive-MP
T6	No-ENV	STEREO	Active-MP
T7	ENV	MONO	No-MP
T8	ENV	MONO	Passive-MP
T9	ENV	MONO	Active-MP
T10	ENV	STEREO	No-MP
T11	ENV	STEREO	Passive-MP
T12	ENV	STEREO	Active-MP

Subjects were encouraged to take 5 minute breaks between runs as often as they needed to avoid fatigue. The total experiment time varied among subjects, from 45 to 60 minutes.

2.4 Data Analysis

Subject performance was evaluated quantitatively using several measures based on the selected size of the virtual bottle. One basic measure, which we named as SizeRatio, represented the relative size of the virtual bottle compared to the proper size of the physical bottle:

$$SizeRatio = \frac{\text{Bottle Size Set By Subject}}{\text{Correct Bottle Size}} \quad (1)$$

The numerator in (1) corresponds to the size of the virtual bottle set by the subject in a certain run and the denominator was fixed at 12 inches (height of the physical 2-liter coke bottle). For example, the SizeRatio values would be 1 at each bottle location if the subject sets the bottle size according to size-constancy. If the subject set the bottle size larger than the actual bottle size then the size-ratio would be greater than 1.

After the SizeRatio was calculated at each bottle position in each run, a linear regression of SizeRatio values versus the distances of the virtual bottle from subject was then performed over all the runs in a trial, resulting in the subject's regression slope in that trial. The fitness of the linear regression was verified by the R-Square value of the linear model. Since with projection-based VE everything is drawn on the CAVE wall, we calculated the visual angle (VA) setting that would result if subjects perceived their distance to the bottle as being the distance they were from the CAVE wall regardless of the bottle's intended distance from the subject. If the subjects' performance is purely determined by visual angle, the size-ratios will theoretically form a fixed slope α, which is determined by the following formula:

$$\alpha = \frac{\text{Correct Bottle Size on CAVE Wall}}{\text{Distance to CAVE Wall}} \quad (2)$$

In our experiment setting, the bottle size is 12 inches, the distance between the subject and the CAVE wall is 5 ft., and so α is 0.2. The percentage relationship between the subject's SizeRatio data regression slopes to that of the predicted VA performance was calculated using the equation:

$$\textit{Percent VA slope} = [\frac{FittedSlope}{\alpha}]*100\% \quad (3)$$

While SizeRatio measured subject's performance in a given run, the percentage relationship between the regression slopes and α indicates the consistency of how well the subject performed across all the runs in a given trial. For example, if the regression slopes of the subject's data were identical to α, then the "Percent VA slope" would be 100%, implying that the subject was showing no size-constancy. On the contrary, if the subject regression data showed perfect size-constancy, the regression slope would be zero and the "Percent VA slope" would consequently also be zero.

Absolute error for each run and mean absolute error across a trial were calculated as another indicator to examine the differences between ideal performance and the size-ratio data collected from our population. Absolute error indicates the deviation of a judgment in a run to actual virtual bottle size. Mean absolute error averaged absolute errors within a given trial. They were computed using the following equations:

$$AbsoluteError = |SizeRatio - 1| \quad (4)$$

$$MeanAbsoluteError = \frac{1}{n}\sum_{1}^{n} AbsoluteError(i) \quad (5)$$

Percent VA slope and AbsoluteError were both derived from SizeRatio values and as aforementioned, described these values from two separate perspectives. For the VA slope percentage, we did repeated measures analysis of variance (ANOVA) using SPSS, with the independent variables to be the three visual factors: scene complexity, stereovision and motion parallax. The purpose of using ANOVA was to discover the significance of each visual factor in affecting size-constancy performance. While for AbsoluteError, we investigated its mean and distribution in each trial. Comparison of these indicators was to reveal that in which trials, i.e. under which visual factor configurations did subjects had better size-constancy performance.

3 RESULTS

ANOVA test of Percent VA slopes across different trials revealed that our experiment data was best fitted by linear models. It also implied that scene complexity and stereovision were the significant factors in determining subjects' performance of size-constancy (both had $p < 0.0001$ in single-factor linear models), while motion parallax did not exhibit a significant influence ($p = 0.3963$ in single-factor linear model). Furthermore, analysis of the linear interactions among these three visual factors suggested that there were no significant interactions. The strongest interaction was between scene complexity and stereovision, with p-value of the corresponding model to be 0.1818. All other models that used interactions did not explain the data well and all had $p > 0.70$. Detailed p-values are listed in Table 2

Table 2. P-values of the linear models, '*' denotes interaction

Factors in Model	P-Value
Scene Complexity	< 0.0001
Stereovision	<0.0001
Motion Parallax	0.3963
Scene Complexity * Stereovision	0.1818
Scene Complexity * Motion Parallax	0.7372
Stereovision * Motion Parallax	0.7524
Scene Complexity * Stereovision * Motion Parallax	0.9721

We further looked into both the percent VA slope and absolute error data to find out under which configurations of the significant factors the subjects achieved closer performance to size-constancy. For the scene complexity factor, subjects' performance on size-constancy was better when viewing in the ENV conditions than the NO-ENV conditions. For stereovision factor, subjects performed better under STEREO conditions than MONO conditions. As aforementioned, the tested population did not exhibit significant difference of size-constancy performance under different motion parallax configurations (No-MP, Passive-MP and Active-MP). We go deep to analyze each visual factor separately in below. Before that, we list the statistics of SizeRatio data collected at all five virtual bottle distances across all trials, including means and standard deviations, in Table 3 is referenced when needed in the following text.

Table 3. SizeRatio statistics at all virtual bottle distances access all trials. Columns list the distance of the virtual bottle from the viewer, rows list trial IDs. Data is means followed by standard deviations.

	3.5ft	5ft	6.5ft	8ft	9.5ft
T1	0.95±0.33	1.12±0.38	1.38±0.44	1.83±0.49	1.93±0.53
T2	0.95±0.32	1.13±0.37	1.35±0.43	1.60±0.50	1.90±0.52
T3	0.94±0.31	1.11±0.38	1.38±0.46	1.63±0.52	1.96±0.57
T4	0.95±0.04	1.06±0.08	1.23±0.12	1.41±0.15	1.69±0.27
T5	1.01±0.13	1.12±0.21	1.27±0.28	1.47±0.34	1.7±0.38
T6	0.99±0.05	1.09±0.10	1.24±0.19	1.43±0.26	1.67±0.33
T7	1.01±0.11	1.29±0.16	1.18±0.23	1.33±0.3	1.43±0.37
T8	1.09±0.05	1.34±0.1	1.24±0.18	1.35±0.23	1.39±0.32
T9	1.07±0.15	1.33±0.2	1.21±0.26	1.38±0.36	1.45±0.44
T10	1.12±0.08	1.28±0.11	1.1±0.14	1.26±0.12	1.25±0.28
T11	1.09±0.04	1.23±0.09	1.04±0.13	1.14±0.18	1.12±0.25
T12	1.12±0.01	1.25±0.01	1.06±0.05	1.19±0.11	1.18±0.17

3.1 Effect of Scene Complexity

Keeping the motion parallax and stereovision factors unchanged, the ability of subjects to set the virtual bottle to the correct size (a size-ratio of 1) was better under the ENV conditions than the No-ENV conditions. Not only was the performance consistent with that for size-constancy but also the task was easier to perform according to subject reports.

Depending on the settings of the motion parallax and stereovision factors, there were totally six pairs of conditions under each of which we could compare the subjects' performance with/without rich environment, i.e. T1 against T7, T2 against T8, T3 against T9, T4 against T10, T5 against T11 and T6 against T12.

The first analysis was to average the size-ratio settings for each bottle position across subjects for the No-ENV and ENV conditions. Due to limitation of space and similarity across all comparison between trials, we plot data from two of the six pairs: T1 against T7 and T4 against T10 in Figure 2 and Figure 3 respectively. Interested readers for other pair of trials should be able to find them in Table 3. Without causing ambiguity, in the two figures T1 and T4 are mentioned as No-ENV conditions and T7 and T10 as ENV conditions.

We found that size-ratio settings were consistently closer to 1 in ENV conditions than in No-ENV conditions. This could be observed in the figures that for the ENV condition subjects produced a mean size-ratio that hovered close to a size-ratio of one for different bottle positions. In contrast, the mean size-ratio for the No-ENV condition increased as the bottle positions receded from the subject. These observations were independent of the setting of stereovision, the other visual factor which also had significant effect.

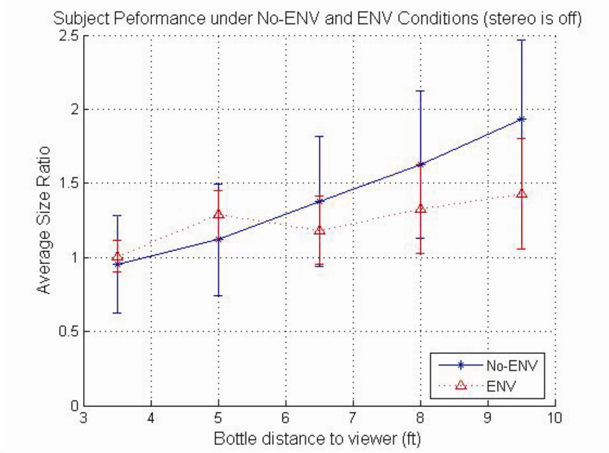

Figure 2. Average subjects' performance in SizeRatio setting in trials T1 and T7, under which stereovision was turned off in VE and there was no motion parallax.

It could be seen that under No-ENV conditions, subjects had a wider range of size-ratio settings as well. The size-ratio settings for the ENV condition when stereovision was turned off in VE ranged between 0.9-1.8 for the bottle distance of 3.5ft- 9.5ft from the subject, for the No-ENV condition under same stereovision configuration the size-ratio settings ranged from 0.62 – 2.46. When stereo vision was turned on, the size-ratio settings under ENV condition ranged from 0.96 – 1.53. Under No-ENV condition, the size ratio ranged from 0.91 – 1.96.

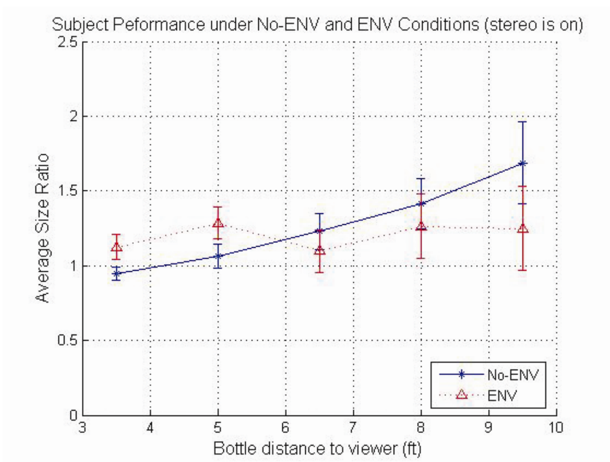

Figure 3. Average subjects' performance in SizeRatio setting in trials T4 and T10, under which stereovision was turned on in VE and there was no motion parallax.

The second analysis was to examine the absolute errors for size judgments made in all the six ENV and six No-ENV conditions among our population. As each of the eighteen subjects did 360 runs, there was 360 * 18 = 6480 runs in total, of which 3240 were performed under ENV conditions and 3240 were performed under No-ENV conditions. Figure 4 shows a clear overall image in difference between ENV and No-ENV performances, by the frequency distribution of absolute errors. Examination of the absolute error for all judgments shows that 66.48% of the errors were 0.2 (or 2.4 inches if measured in the error of size judgment) and below with the ENV condition while only 27.6% of the errors fell within this range with the No-ENV condition. The mean absolute error values calculated using equation (5) were 0.53 for all six No-ENV conditions and 0.26 for all six ENV conditions.

The last analysis was quantified by examining the degree of similarity between regression slopes for their data and those computed for a theoretical visual angle performance.

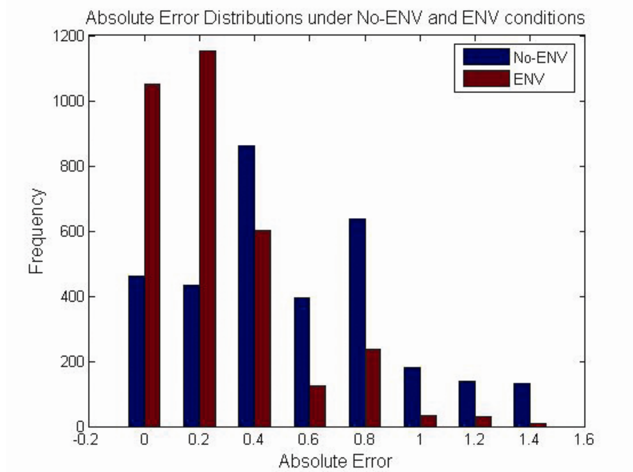

Figure 4. Absolute error value distributions under No-ENV and ENV conditions

Figure 5 illustrates once again that our population's performance in the ENV condition was very different from that in No-ENV condition. We found that the regression slopes obtained in the ENV conditions (0.04±0.03) more closely matched the slopes expected with size-constancy and conversely the slopes in the No-ENV viewing conditions (0.28±0.04) more closely matched those associated with visual angle performance.

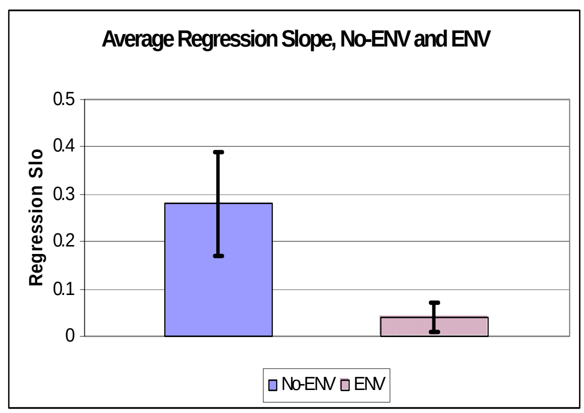

Figure 5. regression slopes mean and standard deviation, under No-ENV and ENV conditions

3.2 Effect of Stereovision

Keeping the motion parallax and stereovision factors unchanged, the ability of subjects to set the virtual bottle to the correct size (a size-ratio of 1) was better under the STEREO conditions than the MONO conditions.

Depending on the settings of the scene complexity and motion parallax factors, there were totally six pairs of conditions under each of which we could compare the subjects' performance with stereovision on/off in VE, i.e. T1 against T4, T2 against T5, T3 against T6, T7 against T10, T8 against T11 and T9 against T12.

The first analysis was to average the size-ratio settings for each bottle position across subjects for the MONO and STEREO conditions. We plot data from two of the six pairs: T1 against T4 and T7 against T10 in Figure 6 and Figure 7 respectively. Data of other pairs of trials are able to be found in Table 3. Without causing ambiguity, in the two figures T1 and T7 are mentioned as MONO conditions and T4 and T10 as STEREO conditions.

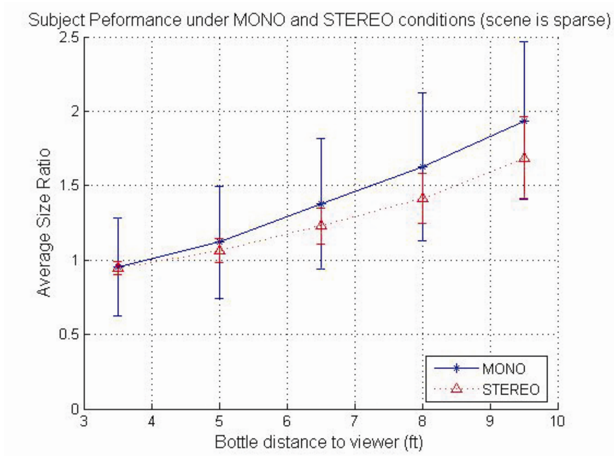

Figure 6. Average subjects' performance in SizeRatio setting in trials T1 and T4, under which scene was sparse in VE and there was no motion parallax.

We found that size-ratio settings were consistently closer to 1 in STEREO conditions than in MONO conditions. This could be observed in the figures that the mean size-ratio for the MONO condition increased as the bottle positions receded from the subject. In contrast, for the STEREO condition although the mean size-ratio also increased with bottle distance from viewer, it increased at a much lower rate. These observations were independent of the setting of stereovision, the other visual factor which also had significant effect.

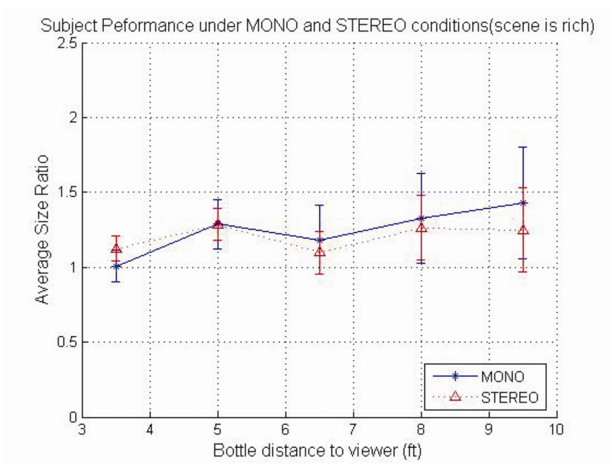

Figure 7. Average subjects' performance in SizeRatio setting in trials T7 and T10, under which scene was rich in VE and there was no motion parallax.

Under MONO conditions, subjects had a wider range of size-ratio settings as well. The size-ratio settings for the STEREO condition when scene was sparse in VE ranged between 0.91-1.96 for the bottle distance of 3.5ft- 9.5ft from the subject, for the MONO condition under same scene complexity configuration the size-ratio settings ranged from 0.62 – 2.46. When scene was rich, the size-ratio settings under STEREO condition ranged from 0.96 – 1.53. Under MONO condition, the size ratio ranged from 0.91 – 1.96.

The second analysis was to examine the absolute errors for size judgments made in all the six MONO and six STEREO conditions among our population. 3240 runs were performed under MONO conditions and 3240 were performed under STEREO conditions. Figure 8 shows the overall image in difference between MONO STEREO performances, by the frequency distribution of absolute errors.

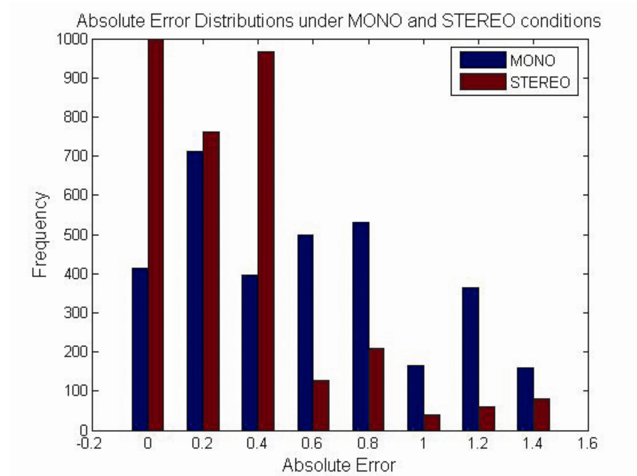

Figure 8. Absolute error value distributions under MONO and STEREO conditions

Examination of the absolute error for all judgments shows that 54.32% of the errors were 0.2 (or 2.4 inches if measured in the error of size judgment) and below with the STEREO condition while only 34.75% of the errors fell within this range with the MONO condition. The mean absolute error values calculated using equation (5) were 0.46 for all six MONO conditions and 0.32 for all six STEREO conditions.

The last analysis was quantified by examining the degree of similarity between regression slopes for their data and those computed for a theoretical visual angle performance.

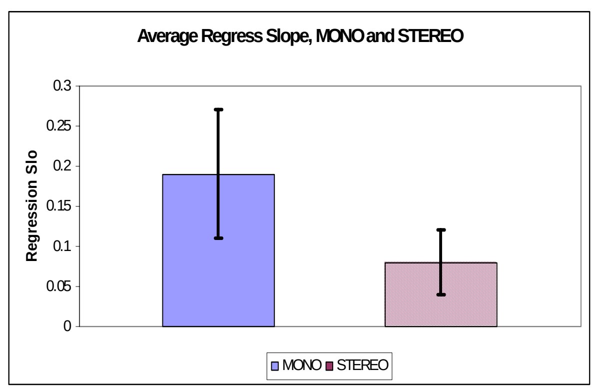

Figure 9. regression slopes mean and standard deviation, under MONO and STEREO conditions

Figure 9 illustrates once again that our population's performance in the STEREO condition was different from that in MONO condition. We found that the regression slopes obtained in the STEREO conditions (0.08±0.04) more closely matched the slopes expected with size-constancy and conversely the slopes in the MONO viewing conditions (0.19±0.08) more closely matched those associated with visual angle performance.

3.3 Effect of Motion Parallax

Keeping the motion parallax and stereovision factors unchanged, the ability of subjects to set the virtual bottle to the correct size (a size-ratio of 1) had no statistically difference under different motion parallax settings, including no-motion parallax, observer-generated motion parallax and VE-generated motion parallax.

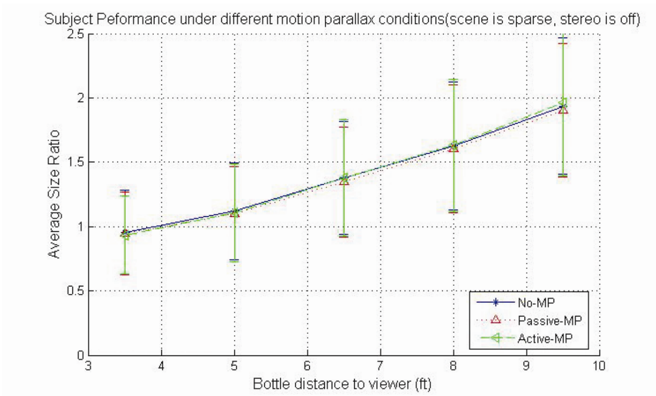

Figure 10. Average subjects' performance in SizeRatio setting across trials T1, T2 and T3, under which stereovision was turned off in VE and scene was sparse.

Depending on the settings of the scene complexity and stereovision factors, there were totally four triples of conditions under each of which we could compare the subjects' performance with different motion parallax settings in VE, i.e. T1, T2 and T3; T4, T5 and T6; T7, T8 and T9; T10, T11 and T12. We plotted data from two of the four triples: T1-T2-T3 and T10-T11-T12 in Figure 10 and Figure 11 respectively. Data of other triples of trials are able to be found in Table 3. Without causing ambiguity, in the two figures T1 and T10 are mentioned as No-MP conditions, T2 and T11 are mentioned as Passive-MP conditions and T3 and T12 as Active-MP conditions.

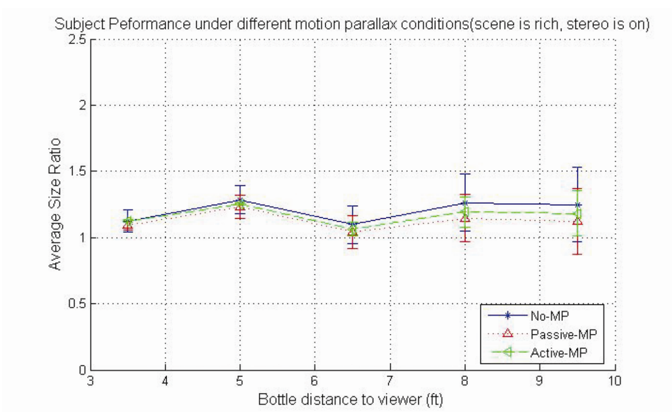

Figure 11. Average subjects' performance in SizeRatio setting across trials T10, T11 and T12, under which stereovision was turned on in VE and scene is rich.

The size-ratio settings across all three motion parallax settings were consistently overlapping with each other. Not only in the mean value, but standard deviations as well. These observations were independent of the setting of scene complexity and stereovision visual factor. When scene was sparse and stereovision was turned off in VE, subjects had the trend to set bottle size in the visual-angle manner. While when scene was rich and stereovision was turn on in VE, they showed up the uniformed performance towards size-constancy.

There was no significant different in the range of size-ratio settings. When scene was sparse and stereovision was turned off in VE, range of size-ratio settings under NO-MP was 0.62-2.46, under Passive-MP was 0.62-2.42 and under Active-MP was 0.63-2.53. When scene was rich and stereovision was turned on in VE, range of size-ratio settings under NO-MP was 0.96-1.53, under Passive-MP was 0.96-1.37 and under Active-MP was 1.01-1.35.

Figure 12 illustrates that our population's performance under the motion parallax conditions were not different from each other. The regression slopes obtained in the No-MP conditions, Passive-MP conditions and Active-MP conditions were 0.15±0.06, 0.15±0.06, and 0.15±0.06 respectively. These values were not statistically different from each other.

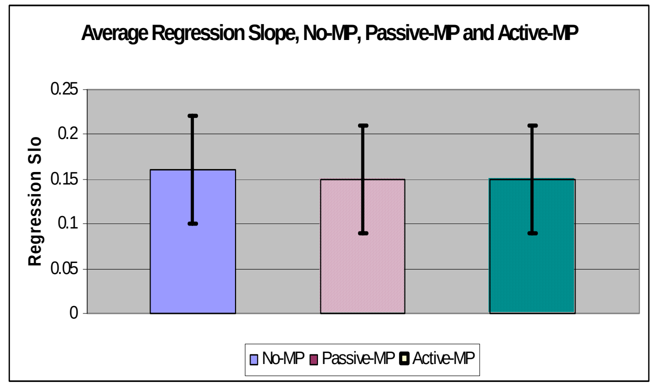

Figure 12. regression slopes mean and standard deviation, under No-MP, Passive-MP and Active-MP conditions

4 CONCLUSIONS AND DISCUSSION

Our experiment first verifies that users could obtain satisfying size constancy performance in an immersive VE, at a view distance range and screen resolution that represent mainstream VR systems (1-9 ft., 1024x768 pixels screen). This verification supports wider deployment of VR system in size and distance perception sensitive applications, such as visual scientific data analysis and virtual metropolitan building planning.

We have found that in the CAVE the ability of subjects to use size-constancy is significantly predicated on the inclusion of rich 2D cues to depth, as well as stereoscopy. The results of our experiments were consistent across the majority of the subject population and suggested that scene complexity and stereovision could have significant impact on the performance of a user of virtual environments to make correct judgments on a virtual object's size. On the contrary, motion parallax, either produced by the virtual environment or by the observer, might not be a significant factor in determining that performance. Our results are similar to results from the majority of previous experiments, either in physical world and VE; despite of the differences in methodologies (a brief summary of related work is given in the following section). These conclusions could be helpful in decision making, for VR system designers who build the systems and for users who utilize the systems for specific applications.

It is worth mentioning that in the physical world 2D cues to depth are natural and straightforward. In fact, it takes effort to arrange a situation that would diminish these cues to the subject. In VE, displaying less complex scenes is easier than showing more complex ones. A VE that has numerous cues to depth (2D and stereovision) takes time to program and computer-time to generate. Thus, it is more expensive to generate a complex world compared to a sparse world in terms of cost, programming time, and display time. By understanding the relationships that exist between the physical and virtual environments will help us better utilize this extraordinary technology by supplying the most important information to the user.

In our experiment we only analyzed three major visual factors due to the thought that they might be of most importance in determining size constancy performance. However with the enrichment of VE, multi-modal interaction between the user and VE is getting more popular and it could be interesting to examine the effect of other factors, e.g. display resolution, haptics, 3D audio etc. Additional experiments could help us understand whether these effects play significant roles in perceiving virtual objects' size and distance.

5 RELATED WORK

[1] did experiments under the applied contexts of minimal access surgery (MAS) tasks, and studied the effects of stereoscopy and observer-produced motion parallax for distant judgment. Results indicated that stereoscopy confers a considerable performance advantage, while providing motion parallax information was not beneficial. The experiments in [2] was to judge visual objects' size which varied fourfold range among trials, the authors concluded that absolute motion parallax only weakly determined the visual scale of nearby objects. Distance perception was studied in [3], for users' performance in tele-operation. The paper suggested that stereoscopy and motion parallax were of equal significance in distance judgment, and users' performance varied largely between HMD and projected screen settings.

The studies of [4][5][6][7] were from different perspectives. [4] compared the results of different experimental methodologies for size-distance perception tests. It argued that for size and distance perception studies, point light sources and rods set experiment apparatus could bring different results from each other, but the different was not significant to change the conclusions. [5] raised the question of whether enhanced motion parallax, i.e. visually magnified motion parallax would alter a visual study conclusion. The answer was there was no significant effect of augmentation on motion parallax effect. [6] presented the experiment result that subject made symmetry judgments in VE under different view conditions, and argued that motion parallax was not a significant factor in determining such capabilities. Effects of multi-modal interaction factors in determining size and distance perception were analyzed in [7], and the authors emphasized the effectiveness of haptic interface in improving distance perception accuracy.

6 ACKNOWLEDGMENTS

Major funding is provided by the National Science Foundation (CDA-9303433.) The virtual reality research, collaborations, and outreach programs at EVL are made possible through major funding from the National Science Foundation, the Defense Advanced Research Projects Agency, and the US Department of Energy; specifically NSF awards CDA-9303433, CDA-9512272, NCR-9712283, CDA-9720351, and the NSF ASC Partnerships for Advanced Computational Infrastructure program. The CAVE is trademark of the Board of Trustees of the University of Illinois.

REFERERNCE

[1] J. Huber, N. Stringer, I. Davies, and D. Field, *Only stereo information improves performance in surgical tasks*, SPIE 5372: 463-470, 2004

[2] A. Beall, J. Loomis, J. Philbeck, and T. Fikes, Absolute motion parallax weakly determines visual scale in real and virtual environments, SPIE 2411:288-297, 1995

[3] P. Rondot, J. Lessard, and J. Robert, Study of motion parallax in depth perception with a helmet-mounted display system used in teleoperation, SPIE 2590:151-159, 1995

[4] C. Ikehara, R. Cole, and J. Merritt, *Effects of test structure on depth perception measurement tasks*, SPIE 1669:135-141, 1992

[5] S. Watt, et al, Can observers exploit enhanced motion parallax to control reaching movements within telepresence environments? SPIE 4299:429-438, 2001

[6] P. Rosen, Z. Pizlo, C. Hoffmann, and V. Popescu, *Perception of 3D spatial relations for 3D displays*, SPIE 5291:9-16, 2004

[7] M. Hirose, K. Hirota, and R. Kijima, *Human behavior in virtual environments*, SPIE 1666: 548-553, 1992.

[8] C. Cruz, D. Sandin, T. Defanti, R. Kenyon, and J. Hart, *The CAVE Audio-Visual Environment*. ACM Trans. on Graphics, 35: 65-72, 1992.

[9] B. Holaday, Die Grössenkonstanz der Sehdinge bei Variation der innerenund äusseren Wahrnehmungsbedingungen. Arch. ges. Psychol., 88: 419-486, 1933.

[10] L. Harvey, H. Leibowitz, Effects of exposure duration, cue reduction, and temporary monocularity on size matching at short distances, J. Opt. Soc. Am., 57: 2, pp. 249-253, 1967.

[11] H. Leibowitz, R. Dato, Visual size-constancy as a function of distance for temporarily and permanently monocular observers. Am. J. Psychol. 79: 279, 1966.

[12] R. Eggleston, W. Janson and K. Aldrich, Virtual reality system effects on size-distance judgments in a virtual environment. Proc. VRAIS, pp 139-146, 1996

[13] Baitch, L., Smith, R.C., *Physiological Correlates of Spatial Perceptual Discordance in a Virtual Environment*, in Proc. Fourth International Projection Technology Workshop, Ames, Iowa, 2000.

[14] X. Luo, R. Kenyon, T. line, H.Waldinger and D.Kamper, *An Augmented Reality Environment for Post-Stroke Finger Extension Rehabilitation*. Proc. ICORR 2005, Chicago, IL, June 2005.

[15] A. Holway, E. Boring, *Determinants of Apparent Visual Size with Distance Variant.* American Journal of Psychology, 54, pp. 121-151., 1941.