On the Determinants of Size-Constancy in a Virtual Environment

Xun Luo [1], Robert V. Kenyon [2], Derek G. Kamper [3], Daniel J. Sandin [2] and Thomas A. DeFanti [4]

[1] *Office of the Chief Scientist R&D, Qualcomm Inc., United States*
[2] *Electronic Visualization Laboratory (EVL), University of Illinois at Chicago (UIC), United States*
[3] *Biomedical Engineering Department, Illinois Institute of Technology, United States*
[4] *California Institute for Telecommunications and Information Technology (Calit2), University of California San Diego (UCSD), United States*

Abstract—**An important aspect of a subject's perception of virtual objects in a virtual environment is whether the size of the object is perceived as it would be in the physical world, which is named size-constancy. The ability of subjects to appreciate size-constancy in an immersive virtual environment was studied while scene complexity, stereovision and motion parallax visual factors were manipulated resulting in twelve different viewing conditions. Under each visual condition, 18 subjects made size judgments of a virtual object displayed at five different distances from them. Responses from the majority of our population demonstrated that scene complexity and stereovision have a significant impact on subjects' ability to appreciate size-constancy. In contrast, motion parallax produced by moving the virtual environment or by the movements of the observer alone proved not to be a significant factor in determining size-constancy performance. Consequently, size-constancy is best obtained when scene complexity and stereovision are components of the viewing conditions.**

Index Terms—**Size-Constancy, Stereovision, Scene Complexity, Motion Parallax.**

I. INTRODUCTION

Virtual Environments (VEs) are used for a variety of research and commercial purposes, such as medical rehabilitation training, scientific data mining and industry manufacturing [16, 20, 9, 11,[9]. The effectiveness of VE applied to such applications relies heavily on its ability to create perceptions within the user that faithfully replicate those experienced in the physical world. However, the limitations of the VE can have an adverse affect on its use and the credibility of the environments that it offers. One significant aspect of this problem is whether users can perceive size-constancy in the VE. That is, does the perceived size of objects rendered in a VE remain constant regardless of its distance from the observer?

The recent work of Kenyon et al. [18] demonstrated size-constancy behavior in subjects using a CAVE® (CAVE Automatic Virtual Environment) [8]. For a majority of their population monocular cues to depth were required to accompany the persistent steoreoptic attribute of the object to reinforce its true size. With only steroeptic cues available in the scene, a majority of their subjects failed to exhibit size-constancy and adopted a visual angle performance, i.e., size of the virtual object was perceived as proportional to its projected size on the CAVE screen. Although subjects were free to move their head or body during the experiments, which would have produced motion parallax, they did not do so. Thus the results of [18] leave open the question of whether motion parallax, an additional monocular cue to depth, could improve performance. In this study we exposed subjects to both active and passive motion parallax conditions in addition to changes in scene complexity and the availability of stereoptic cues. Our results were similar to those performed in the physical world where size-constancy was more prevalent when a rich environmental scene was accompanied by stereoptic cues. When the richness of environment was significantly reduced and stereoptic cues were 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 observers, had a mix effect on the perception of size-constancy. Some subjects benefited from motion parallax while others showed no effects at all.

II. RELATED WORK

Huber et al. [1] studied the effects of stereoscopy and observer-produced motion parallax for distance judgments for tasks under associated with minimal access surgery (MAS). Results indicated that stereoptic cues confer a considerable performance advantage, while providing motion parallax information was not beneficial. Experiments by Beall et al. [2] where subjects judged the size of objects, whose visual dimension varied fourfold, concluded that absolute motion parallax only weakly determined the visual scale of nearby objects. Rondot et al. [3] studied distance perception during a tele-operation task. Their results suggested that stereoptical and motion parallax cues were of equal significance in distance judgment, and users' performances varied widely dependent on whether they used a head mounted display (HMD) or projection-based VE system.

Additional studies showed inconsistent effects of motion parallax. Ikehara et al. [4] compared the results of different experimental methodologies for size-distance perception tests. Their results argued that two experiment configurations: using point light sources or using rods could produce different results

about subjects' performances in size and distance perception, but these differences were not statistically significant. Watt et al. [5] raised the question of whether enhanced motion parallax, i.e. visually magnified motion parallax would alter the result found when using standard motion parallax stimuli. They found no significant improvement when augmented motion parallax was used. Rosen et al. [6] showed using object symmetry as a measure, that subject judgments changed under different VE 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 by Hirose et al. [7]. The authors emphasized the effectiveness of a haptic interface in improving distance perception accuracy, but size-constancy perception was not discussed.

III. METHODS

3.1 Subjects

Eighteen subjects, numbered EC1-EC18, were tested. 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, using standardized Snellen eye chart and Litmus Stereo Fly Test. All subjects had corrected vision of 20/20 and normal stereovision.

3.2 Apparatus

All tests were performed using a single wall CAVE – the C-Wall (Configurable Wall) [19]. The C-Wall is a high-quality, head-tracked, active stereo wall, that displays an image in front of the viewer by means of a 10x10 ft. 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. The field of view available to the subjects was determined by the characteristics of the stereo glasses: 100 °H x 55 °V. A Pentium IV PC created the images 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 to generate the personalized stereo images. 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 ±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 used as the virtual object. The

experimental setup is similar to that used in [18]. Characteristics of VE scene were manipulated in order to test the effects of scene complexity, motion parallax, and stereovision on the perception of size-constancy.

• Scene Complexity

Two environment scenes were used: a "rich" environment (ENV), containing monocular and stereoptical cues to depth and a "sparse" environment (No-ENV) where cues to depth were confined to the bottle displayed in the scene. The ENV scene consisted of a gray-green checkered floor with a wooden textured table; 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 heights (30, 33 and 36 inches). For the No-ENV case, the environment consisted solely of a virtual Coke bottle presented in front of a gray background. The virtual Coke bottle was displayed 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 below. The head was tracked by the Eagle system as described previously.

• Stereovision

Two viewing conditions were examined: monocular vision (MONO) and stereovision (STEREO). For the STEREO condition, disparate images were presented to the two eyes. IPD was measured for each subject, and the images for the two eyes were created to reflect the different vantage points in order to present a stereo image of the scene. For the MONO condition, the IPD was set to zero in the CAVE program therefore the same image was presented to each eye. Consequently, the subjects continued to view the scene through the Stereographics glasses thus imparting STEREO and MONO environments with the same visual conditions except for the parameters changed by the experiment.
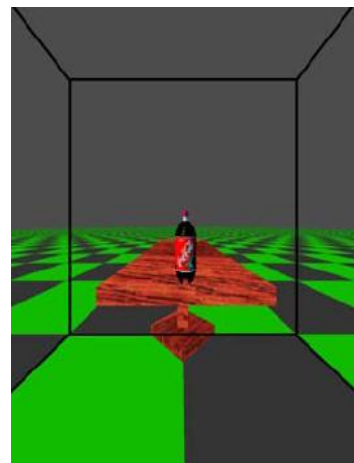
• Motion Parallax

Fig. 1. The virtual coke bottle with rich scene environment.

Three different motion parallax settings were tested: no motion parallax (No-MP), motion parallax generated by the VE (Passive-MP), and motion parallax generated by the lateral movement of the viewer (Active-MP).

For the No-MP condition the subjects were instructed to hold

their head still and look straight ahead with no lateral head movement. To ensure the subject was not moving, the experimenter monitored the lateral head movements from the tracker, and prompted the subject whenever there were head movements greater than 1 inch, the threshold needed to incur motion parallax.

For the Passive-MP condition, the whole scene displayed on the C-Wall moved in the horizontal direction in a sinusoidal fashion at 0.25 Hz. Peak scene displacement was 1 ft each way and peak velocity was 4 ft/s. 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 subjects were instructed to move their head laterally from side to side at 0.25 Hz with a minimum head displacement of 1 ft. An electronic metronome provided an audio cue to keep the subject moving at the appropriate frequency. The experimenter monitored lateral head movement through the tracker and prompted the subject whenever lateral movement amplitude fell below the desired level.

3.3 Experimental Protocol

Subjects were instructed to adjust the size of the virtual object (2-liter Coke bottle) so that they perceived its size as being identical to that of a physical Coke bottle if placed at the 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 3 ft tall wooden stand covered with black cloth. The stand was positioned at the front left side of the C-Wall at a distance of 3.5 ft. from the subject. Both the physical and the virtual Coke bottles were 12 inches tall and 5.5 inches 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 its normal size (12 inches). 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/eyes.

The independent variables of scene complexity, stereovision, and motion parallax had 2, 2, and 3 possible states respectively. Thus there were 12 visual conditions in total. 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 conditions mapped one-to-one to each other. TABLE 1 shows this mapping relationship, with the trials numbered T1-T12. During experiment, the show-up sequence of T1-T12 was randomized for each subject.

Subjects were encouraged to take 5 minute breaks between trials or as often as they needed to avoid fatigue. The total experiment time varied from 45 to 60 minutes across our subject population.

TABLE 1: MAPPINGS BETWEEN TRIAL IDS AND VISUAL CONDITIONS

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

3.4 Data Analysis

Subject performance was evaluated quantitatively using several measures based on the selected size of the virtual bottle. In brief, the metric SizeRatio represented the relative size of the virtual bottle compared to the proper size of the physical bottle:

$$SizeRatio = \frac{Bottle\ Size\ Set\ by\ Subject}{Correct\ Bottle\ Size} \qquad (1)$$

The numerator in Eq. 1 corresponds to the size of the virtual bottle set by the subject in each run and the denominator was fixed at 12 inches (height of the physical 2-liter Coke bottle).

Linear regression of resulting SizeRatio values against the distances of the virtual bottle from subject was then conducted. Since with projection-based VE everything is drawn on the CAVE wall, we calculated the visual angle (VA) value 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 virtual distance from the subject. If the subjects' performance is purely determined by visual angle, the SizeRatios will theoretically form a straight line with a fixed slope α based on the following equation:

$$\alpha = \frac{Correct\,Bottle\,Size\,on\,CAVE\,Wall}{Viewer\,Dis\tan ce\,to\,CAVE\,Wall} \qquad (2)$$

In our experiment, α was set at 0.2 given a bottle size of 12 inches (1ft), and a distance between the subject and the CAVE wall of 5 ft. While SizeRatio measured subject's performance in a given run, the ratio between the regression slopes and α indicated the consistency of how well the subject performed across all the runs in a given trial. This percentage relationship between the subjects' SizeRatio data regression slopes and that of the predicted VA performance was calculated using the following equation:

$$Percent\,VA\,Slope = \left(\frac{Re\,gression\,Slope}{\alpha}\right) * 100\% \qquad (3)$$

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. In contrast, if the subject's regression data showed perfect size-constancy, the regression slope would be zero and the value of Percent VA Slope would consequently be zero as well.

Absolute error for each run and mean absolute error across a trial were calculated as two other metrics to examine the differences between ideal performance and the SizeRatio data collected from the subject population. Absolute error indicates the deviation of a judgment in a run compared to the actual virtual bottle size. Mean absolute error averaged absolute errors within a given trial. They were calculated using the following equations:

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

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

Percent VA Slope and AbsoluteError were both derived from SizeRatio values and as aforementioned, described these values from two separate perspectives.

To uncover the significance of each visual factor affecting size-constancy, an analysis of variance (ANOVA) with repeated measures was performed on percent VA slopes using SPSS (SPSS, Inc). The independent variables were the three visual factors: scene complexity, stereovision and motion parallax. To reveal under which visual conditions our subject population showed better size-constancy performance, we calculated the mean and distribution of the AbsoluteError, in each trial.

IV. RESULTS

For our subject population, size-constancy performance, as measured by percent VA, was better when viewing under the ENV conditions than under the NO-ENV conditions and better under STEREO conditions than under MONO conditions (by single-factor ANOVA results). Our subject population performance across the three motion parallax configurations did not show any statistically significant difference, i.e., the addition of motion parallax had no effect

on our population. Furthermore, there were no significant interactions among the three visual factors of scene complexity, stereovision and motion parallax. All models that used interactions did not explain the data well, with the smallest p value being 0.188.

4.1 Effect of Scene Complexity

Comparing the Percent VA Slopes (Eq. 3) among our subject population, for the ENV vs. No-ENV trials (Fig. 2) that had the same motion parallax and stereovision conditions, i.e. T1 vs. T7, T2 vs.T8, T3 vs.T9, T4 vs.T10, T5 vs.T11 and T6 vs.T12, showed that subjects' size-constancy performance was significantly better under the ENV conditions rather than the No-ENV conditions (p < 0.0001). The Percent VA Slopes obtained under the ENV conditions (20±15%) more closely matched the slopes expected with size-constancy whereas the slopes under the No-ENV viewing conditions (140±20%) more closely matched those associated with visual angle performance.
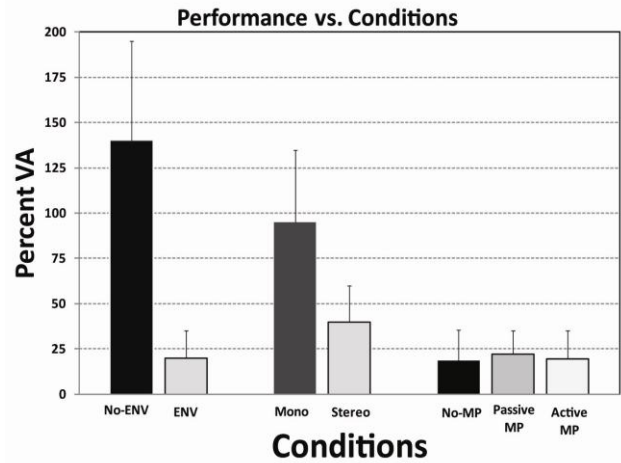

Fig. 2. Percent VA slope means and standard deviations under different conditions. Motion Parallax (MP) results for ENV conditions only.

In addition, subject performance in the ENV condition was more consistent and the task was easier to perform according to subject reports. As seen in Fig. 3, SizeRatio settings were consistently closer to 1 in ENV conditions than in No-ENV conditions for different bottle positions, especially for the bottles farther from the subject. In contrast, the mean SizeRatio for the No-ENV condition increased as the bottle's position receded from the subject. The SizeRatio values also exhibited wider ranges of variance in the No-ENV condition compared to ENV condition.

With or without stereovision, subject performance under the ENV conditions was consistently better than the No-ENV conditions. With no stereoptical cues (Fig. 3 top), SizeRatio settings for the ENV conditions ranged between 0.9-1.8 for bottle distances of 3.5ft- 9.5ft from the subject, while the No-ENV conditions produced SizeRatio values that covered twice the range, i.e., 0.62 – 2.46. The introduction of stereovision caused the range of values to shrink in both ENV and No-ENV conditions. With stereoptical cues (Fig. 3 bottom) the SizeRatio settings under ENV conditions ranged

from 0.96 – 1.53. Under No-ENV conditions, the SizeRatio range was also smaller, i.e., 0.91 – 1.96, than the stereo-off trials.
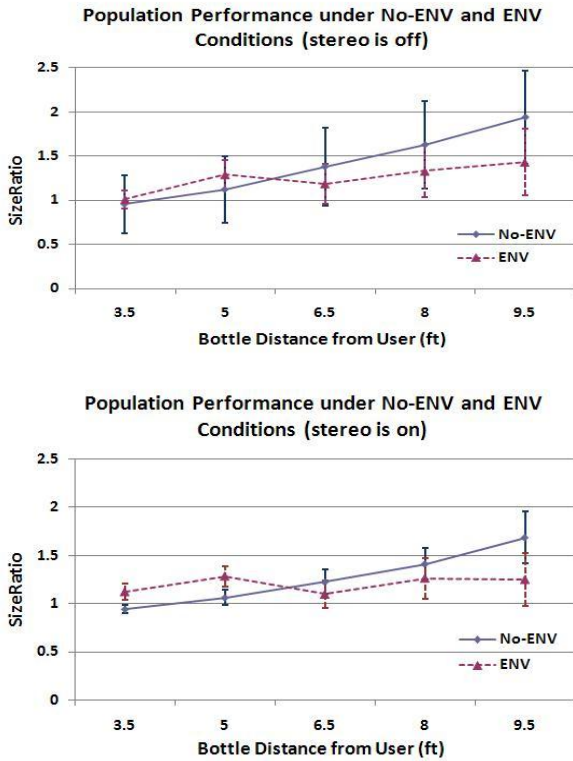

Fig. 3. Population performance under No-ENV and ENV conditions with stereovision off (top) and on (bottom), without motion parallax.

The AbsoluteError values for all six ENV and No-ENV conditions (Fig. 4) showed a difference between ENV and No-ENV performances. The frequency distribution of AbsoluteError values showed that 66% of the errors were 20% of the Coke bottle height (or 2.4 inches) and below in the ENV conditions while only 28% of the errors fell within this range in the No-ENV conditions. The MeanAbsoluteError values calculated using Eq. 5 were 0.26 for all six ENV conditions and 0.53 for all six No-ENV conditions.
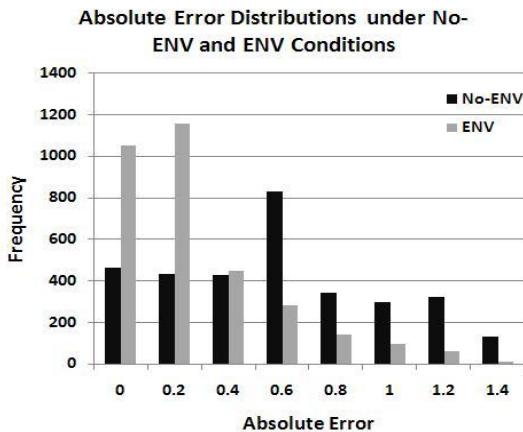

Fig. 4.   Absolute error distributions under No-ENV and ENV conditions.

4.2 Effect of Stereovision

Subjects' performance was more comparable to size-constancy under the STEREO conditions than under the MONO conditions (p < 0.05) given the same configurations of scene complexity and motion parallax, i.e., T1 vs.T4, T2 vs.T5, T3 vs.T6, T7 vs.T10, T8 vs.T11 and T9 vs.T12. Comparing the Percent VA Slope values from our subject population, for the STEREO vs. MONO trials showed that the Percent VA Slopes obtained under the STEREO conditions (40±20%) more closely matched the slopes expected with size-constancy and conversely the slopes in the MONO viewing conditions (95±40%) more closely matched those associated with visual angle performance. This result is shown by the middle bar pair in Fig. 2.

This improved performance can be observed in Fig. 5 as well, where the mean SizeRatio under the MONO conditions increased as the bottle's position receded from the subject. In contrast, for the STEREO conditions although the mean SizeRatio also increased with bottle distance from viewer, it increased at a much lower rate. These observations were independent of scene complexity.
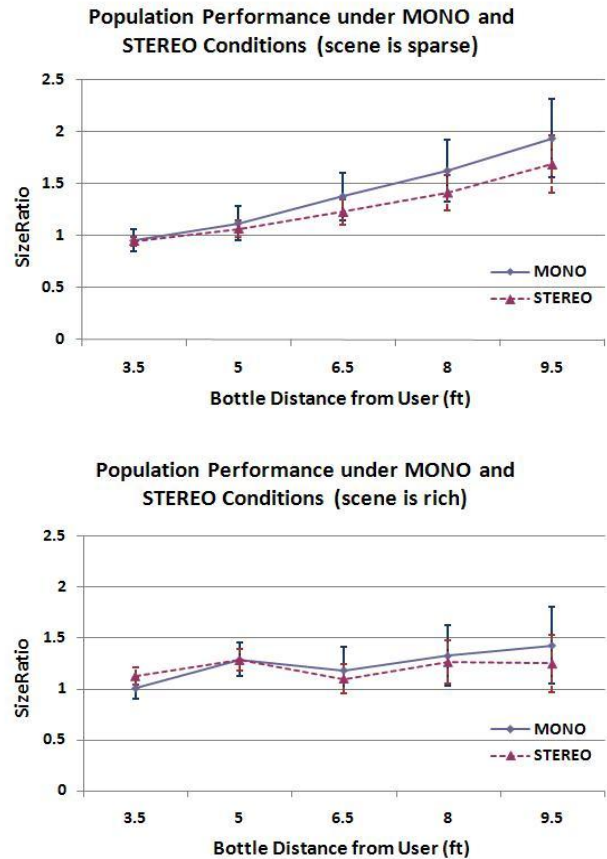

Fig. 5. Population performance under MONO and STEREO conditions, with No-ENV (top) and ENV (bottom) conditions, without motion parallax.

Under the MONO conditions, subjects had a wider range of SizeRatio values as well. The SizeRatio values for the STEREO condition using the No-ENV scene ranged between 0.91-1.96 for the bottle distance of 3.5-9.5ft from the subject. For the same configuration of the other two visual factors,

values under the MONO conditions ranged from 0.62 – 2.46. When the ENV scene was used, the SizeRatio settings under STEREO condition ranged from 0.96 – 1.53, while under MONO condition the SizeRatio values ranged from 0.91 – 1.96.

The AbsoluteErrors under the six MONO and STEREO conditions (Fig. 6) show that under the STEREO conditions 54% of the errors were no greater than 20% of the Coke bottle size (or 2.4 inches) while 34% of the errors fell within this range under the MONO conditions. The MeanAbsoluteError values calculated were 0.46 for all six MONO conditions and 0.32 for all six STEREO conditions.
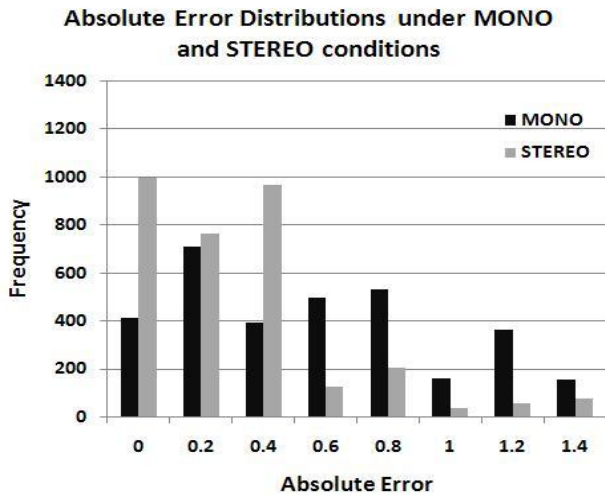

Fig. 6. Absolute error distributions under MONO and STEREO conditions.

4.3 Effect of Motion Parallax

The introduction of motion parallax using the same scene complexity and stereo conditions produced no statistical difference in performance for our population (conditions: T1, T2 and T3; T4, T5 and T6, T7, T8 and T9, T10, T11 and T12). The means and standard deviations of the Percent VA Slope values for all three motion parallax settings overlapped regardless of the scene complexity and stereovision settings. Using a No-ENV scene and stereoptical cues turned off, subjects' SizeRatio values showed a visual-angle performance. In contrast, when viewing the ENV scene and stereovision was turned on, subjects showed a uniform performance consistent with size-constancy, as shown by the right group of bars in Fig. 2. Finally with the ENV scene and stereoptical cues turned off, the subjects' performances laid between those under the above two groups of conditions.

There was no statistically significant difference in the range of SizeRatio values. Viewing the No-ENV scene with the stereoptical cues turned off, the range of SizeRatio values 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. Viewing the ENV scene with stereoptical cues turned off, the range of SizeRatio values under NO-MP was 0.91-2.0, under Passive-MP was 0.9-1.8 and under Active-MP was 1.04-1.71. Viewing the ENV scene with stereovision turned on, the range of SizeRatio values 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.

Although as a group our subject population showed no significant change in performance with the addition of motion parallax, examining the performance of individual subject's under different motion parallax conditions did reveal changes in an individual's performance. In Table 2, we grouped the twelve trials into four triples of trials based on the different conditions for scene complexity and stereovision. We rank ordered them based on a decreasing level of visual cues: ENV:STEREO, ENV:MONO, No-ENV:STEREO and No-ENV:MONO. We investigated how each individual subject performed at each of the three configurations of motion parallax. If the Percentage VA Slope value under a particular motion parallax configuration was more than 10% of another then a greater than (>) symbol was used. Less than 10% was represented by an equal symbol (=) was used. Finally, if under all three conditions VA slope showed a less than 10% difference then "same" was used. The abbreviations N, P and A represented No-MP, Passive-MP and Active-MP conditions respectively. For instance, the notation for Fig. 7(top) would be N>A=P which means that the Percentage VA Slope value for No-MP condition was 10 % greater than the values under both Passive-MP and Active-MP conditions. Percentage VA Slope values under Passive-MP and Active-MP conditions, however, were within a 10% difference from each other.
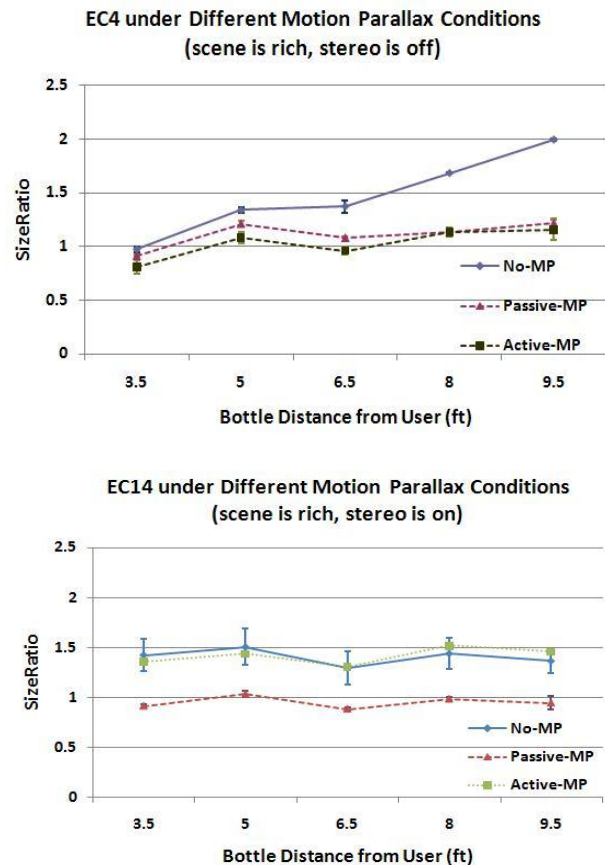

Fig. 7. Individual SizeRatio settings for trials using a rich scene, without (top) and with (bottom) stereo.

These results revealed that the eighteen subjects could be

categorized into four groups, based on their consistency in size-constancy performance across the scene-richness groups. Eight subjects (EC1, 2, 5, 10, 13, 15, 16, and 18) exhibit no significant difference in size-constancy across all three motion parallax conditions, regardless of the variation in scene-richness. Ten subjects showed a change in performance when exposed to motion parallax, but the results were mixed and could not be explained by a uniform model. Among these ten subjects, four of them (EC3, 4, 7 and 12) performed relatively better under Passive-MP configuration rather than Active-MP configuration. Three subjects (EC6, 9 and 11) performed relatively better under Active-MP configuration rather than Passive-MP configuration. Two subjects (EC8 and 17) actually performed worse under both Active-MP and Passive-MP compared to the No-MP conditions. An instance of improved slope with motion parallax is shown in Fig. 7(top) where there is a significant change in the slope under both Passive-MP and Active-MP conditions.

There were some subjects where the slope alone did not give the entire picture of their performance. As shown in Fig. 7(bottom), this subject showed the same slope for all conditions but the Passive-MP condition showed an improvement in accuracy of the size setting behavior since the SizeRatio settings were clearly lower than those under No-MP and Active-MP conditions, and around the correct value of 1.

TABLE 2: INDIVIDUAL SUBJECTS' PERFORMANCE ACROSS SCENE RICHNESS GROUPS, WITH REGARD TO MOTION PARALLAX SETTINGS

Triple Group Subject ID	ENV:STEREO	ENV:MONO	NOENV:STEREO	NOENV:MONO
EC1	same	same	same	same
EC2	same	same	same	same
EC3	A>N>P	A=N>P	same	same
EC4	A>N>P	N>A=P	same	same
EC5	same	same	same	same
EC6	P>N=A	P=N>A	same	same
EC7	A>N=P	same	same	same
EC8	P>A>N	A=P>N	same	same
EC9	P>N=A	same	same	same
EC10	same	same	same	same
EC11	P>N=A	same	same	same
EC12	A>N=P	same	same	same
EC13	same	same	same	same
EC14	same	same	N>A=P	same
EC15	same	same	same	same
EC16	same	same	same	same
EC17	A=P>N	same	same	same
EC18	same	same	same	same

V.   CONCLUSIONS AND DISCUSSION

Our results illuminate several important issues regarding the perception of size-constancy in projection based VE systems (the C-Wall is a CAVE variation). Our work shows, in agreement with a previous study [18], that users can appreciate size-constancy in an immersive projection-based VE, at view distances and screen resolutions that represent mainstream VE systems (10x10 ft. screen, 1024x768 pixels resolution each screen). In addition to scene complexity, we found our subject population's best performance (i.e., size-constancy) occurred when stereovision was made available to subjects. The monocular cues to depth that comprised our complex visual scene were necessary but not sufficient by themselves to equal the benefit afforded subjects when stereovision is added to the mix. Given that the effective range of stereopsis extends beyond the distance at which our virtual objects were displayed (3.5-9.5ft), we find that for targets that are within a space of the size of the CAVE, stereovision is an important visual component in producing the size-constancy perception. Had we used more distant targets our results may have been different[14, 15].

Although stereovision was a necessary addition to the static monocular cues to depth to achieve the best size-constancy, we expected substituting motion parallax for stereovision would have produced subjects' performances equal to that found using stereovision with a complex scene. Unexpectedly, our results showed that motion parallax, produced by either the virtual environment or the observer alone, did not significantly affect the production of size-constancy for our subject population as a whole. However, when we examined individual subjects' performances, we found that the effect of motion parallax varied from one subject to the next. Since motion parallax depends on the richness of the scene and the movement of objects at different distances, it may be that our visual scene or the magnitude of motion used was not ideal to show an effect in most subjects. As expected, the small amount of relative movement that occurs using a sparse scene was generally not sufficient to improve performance. The largest effect can be seen in the ENV-STEREO condition followed by the ENV-MONO condition. As seen in

Fig. 7, we found that some subjects either increased the distance at which they could perceive size-constancy (i.e., a shallower regression slope) or perceived more veridical bottle sizes (SizeRatio ≈ 1). Thus we can see that some subjects were affected by the introduction of motion parallax.

Our results also compare well with experiments performed in the physical world [12, 13, 17]. These studies have shown that a subject's performance lies on a continuum between size-constancy and visual-angle and that this performance is a function of the cues that are present in the scene. In Fig. 8 we show that our subject population's performance moved from size-constancy to VA performance as a function of the cues presented[2] and is similar to Fig. 22 in [17] where they plot their subjects' performance as the visual field-of-view was narrowed

thus reducing the visual cues available[3]. Similarly, one might expect the performance from our subject population to follow a similar course as the cues in the visual field are manipulated. Our Fig. 8 shows just this predicted behavior. We find that the dominant condition for size constancy is a rich scene with stereovision (ENV: STEREO). Reduction in the cues reduction in the cues of the rich scene to a monocular condition (ENV: MONO) reduced the prevalence of size-constancy. Further reduction in cue availability shows an increase in VA performance where No-ENV: STEREO condition is further deteriorated and only modestly improved performance compared to the condition with the least number of cues (No-ENV: MONO).
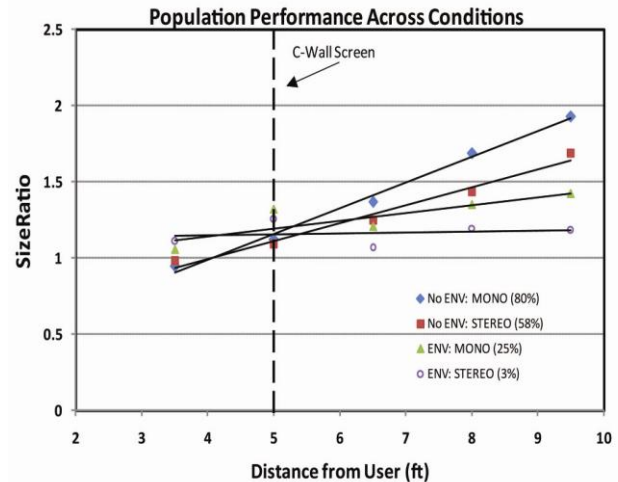

Fig. 8. Averaged fitted slopes across four combinations of scene complexity and stereovision conditions with their associated percent VA slopes for each condition.
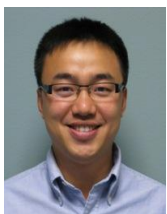
In our experiment we examined three major visual factors influencing size-constancy. However with the enrichment of VE, multi-modal interaction between the user and VE is getting more popular and it might become important to examine the effect of other factors, e.g. display resolution, haptics, 3D audio to name only a few. Additional experiments could help us understand whether other sensory inputs play a significant role in perceiving virtual objects' size. Additional sensory information may be available in other applications, such as visual scientific data analysis, VE-aided physical therapy and virtual metropolitan building planning which may improve size-constancy perception.

Our results could be helpful for VR system designers and for users who utilize such systems for specific applications. As VE matures an increasing number of sensory inputs will become available to the user. However, the addition of such aspects will still increase the cost and complexity of environment generation. Consequently, we will still need to understand the relationships that exist between the physical and virtual environments so as to help us better utilize this extraordinary technology by supplying the most important information to the user.

---

[2] Since motion parallax was not a significant factor in our population's results, we grouped our subjects' performance into categories: No-ENV:MONO, No-ENV:STEREO, ENV:MONO and ENV:STEREO and averaged their regression slope values within each category.

[3] In their figure, size-constancy is represented by a diagonal line and visual angle a flat line. In our figure the opposite is used: size-constancy is a flat line and visual angle is a diagonal line. This is due the differences in the two protocols used.
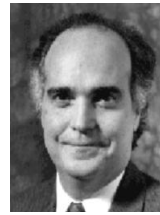
REFERENCES

[1] J. Huber, N. Stringer, I. Davies and D. Field. Only stereo information improves performance in surgical tasks, *SPIE*, 5372, pp.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, pp.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, pp.151-159, 1995.

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

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

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

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

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

[9] E. J. Wegman and J. Symanzik. Immersive projection technology for visual data mining, *Journal of Computational and Graphical Statistics*, 11, pp.163-188, 2002.

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

[11] A. Banerjee, P. Banerjee, N. Ye and F. Dech. Assembly Planning EffectivenessUsing Virtual Reality, *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 2, pp. 204-217, 1999.

[12] L. Harvey and H. Leibowitz. Effects of exposure duration, cue reduction and temporary monocularity on size matching at short distances, *J. Opt. Soc. Am.*, vol.57, no.2, pp. 249-253, 1967.

[13] H. LeibowitzA and R. Dato. Visual size-constancy as a function of distance for temporarily and permanently monocular observers, *American Journal of Psychology*, vol. 79, pp.279, 1966.

[14] R. Eggleston, W. Janson and K. Aldrich. Virtual reality system effects on size-distance judgments in a virtual environment, in *Proceedings of IEEE Virtual Reality Annual International Symposium '96*, pp. 139-146, 1996.

[15] L. Baitch and R.C. Smith. Physiological correlates of spatial perceptual discordance in a virtual environment, in *Proceedings of the Fourth International Projection Technology Workshop*, Ames, Iowa, 2000.

[16] X. Luo, R.V. Kenyon, T. Kline, H.C. Waldinger and D.G. Kamper. An augmented reality environment for post-stroke finger extension rehabilitation, in *Proceedings of the International Conference on Rehabilitation Robotics '05*, Chicago, IL, June 2005.

[17] A. Holway, E. Boring. Determinants of apparent visual size with distance variant, *American Journal of Psychology*, 54, pp. 121-151, 1941.

[18] R.V. Kenyon, D.J. Sandin, R. Smith, R. Pawlicki and T.A. DeFanti. Size-constancy in the CAVE, *Presence*, vol.16, no.2, 2007.

[19] E.A. Keshner, R.V. Kenyon and J. Langston. Postural responses exhibit multisensory dependencies with discordant visual and support surface motion, *J. Vest Res.*, 14, pp.307-319, 2004.

[20] H.C. Fischer, K. Stubblefield, T. Kline, X. Luo, R.V. Kenyon and D.G. Kamper. Hand rehabilitation following stroke: a pilot study of assisted finger extension training in a virtual environment, *Topics in Stroke Rehabilitation*, vol.14, no. 1, pp.1-12.

Xun Luo, Ph.D., is a senior research engineer at Office of the Chief Scientist R&D, Qualcomm Inc. He earned his B.S. and M.S. degrees from the University of Electronic Science and Technology of China, and doctorate degree from the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago (UIC), all in computer science. During 2006 to 2008, he had been a researcher at Motorola Labs with a focus on HCI in pervasive environments. His research interests include pervasive computing, virtual reality, ultra-high-resolution displays and wearable systems. He has published over a dozen conference and journal papers and is the innovator for 5 pending U.S. patents. He authored the book *Personal Augmented Computing Environment: a Framework for Personalized Visualization and Scalable Human Computer Interaction* (VDM-Verlag, 2008).

Robert V. Kenyon, Ph.D., received his B.S. in Electrical Engineering from the University of Rhode Island, in 1970, an M.S. in Bioengineering from the University of Illinois, Chicago, in 1972, and a Ph.D. in Physiological Optics from the University of California, Berkeley, in 1978. From 1979 to 1986, he was a faculty member of the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology, Cambridge. He is currently a Professor of Computer Science at the University of Illinois at Chicago. A majority of his research has examined how biological systems adapt to unusual environments and the contribution of visual information to the control of motor coordination. He has participated in several space shuttle experiments that studied the effects of micro-gravity on human/animal orientation: Spacelab-1, German Space-lab (D-1), and STS-29 (Chix in Space).

Derek G. Kamper, Ph.D, is an Assistant Professor in the Department of Biomedical Engineering at the Illinois Institute of Technology and a Research Scientist in the Sensory Motor Performance Program at the Rehabilitation Institute of Chicago. He received a B.E. degree in electrical engineering from Dartmouth College in 1989 and M.S. and Ph.D. degrees in biomedical engineering from Ohio State University in 1992 and 1997, respectively. His research is focused on hand neuromechanics and rehabilitation following neuromuscular injury, with an interest in using VR for retraining after stroke.

Daniel J. Sandin is director emeritus of the Electronic Visualization Lab (EVL) and a professor emeritus in the School of Art and Design at the University of Illinois at Chicago (UIC). Currently Sandin is a researcher at EVL at UIC and at CALIT2 part of the University of California at San Diego. Sandin's latest VR display system is Varrier, a large scale, very high resolution head tracked barrierstrip autostereoscopic display system that produces a VR immersive experience without requiring the user to wear any glasses. In its largest form it is a semi-cylindrical array of 60 LCD panels.

Thomas A. DeFanti, Ph.D., at the University of California, San Diego, is a research scientist at the California Institute for Telecommunications and Information Technology (Calit2). At the University of Illinois at Chicago, DeFanti is director of the Electronic Visualization Laboratory (EVL), and a distinguished professor emeritus in the department of Computer Science. He has researched computer graphics since the early 1970s. His credits include: use of EVL hardware and software for the computer animation produced for the 1977 "Star Wars" movie; contributor and co-editor of the 1987 NSF sponsored report "Visualization in Scientific Computing;" recipient of the 1988 ACM Outstanding Contribution Award; he became an ACM Fellow in 1994.