A Comparative Study of Virtual Reality and 2D Display Methods in Visual Search in Real Scenes

Juan Carlo M. Figueroa[✉], Raul Alberto B. Arellano,
and Janeen Mikee E. Calinisan

College of Engineering, University of the Philippines Diliman,
1101 Quezon City, Philippines
{jmfigueroa, rbalberto, jecalinisan}@up.edu.ph

Abstract. People perform visual search tasks every day: from trivial tasks to emergencies. Classical research on visual search used artificial stimuli to identify factors that affect search times and accuracy. Recent studies have explored visual search in real scenes by simulating them on two-dimensional displays. The scientific community continues to use new technology to formulate better methods and practices. Virtual reality is a new technology that offers its users immersivity and elicits "real" responses. The purpose of this study is to compare search efficiencies in real scenes on 2-D displays and Virtual Reality. A visual search experiment measuring reaction times and accuracy was conducted to evaluate both methods. Results suggest that visual search in real scenes is significantly faster and more accurate in Virtual Reality than in 2-D Displays. These findings could open up new opportunities for visual search research on real scenes and real life scenarios.

Keywords: Ergonomics · Virtual reality · Visual search

1 Introduction

1.1 Background of the Study and Rationale

Visual search is the study of the human ability to distinguish and identify a target among the presence of distractors. Classical and standard methods of visual search studies include the use of artificial objects, such as arbitrary shapes, as search targets on blank artificial backgrounds. Subjects are tasked to search for the target among a number of distractor figures. These studies identified certain features and factors that make targets more visually detectable than others do. Some of the basic factors include shape, size, color, and spatial frequency [1].

These search scenes are typically more complex than the artificial tests used in the classical methods, which is why there have been studies on visual search in real scenes in recent years [2]. Some of these studies include the study of the role of memory for visual search in real scenes [2], the study of the factors that contribute to the seemingly efficient search in real scenes [3] and the parametric modeling of search efficiency in

real scenes [4]. Most of these studies used two-dimensional (2D) displays for their respective experiments.

Recent advances in technology provide opportunities for better research in the scientific community. A new technology that has captured the attention of both the scientific community and the commercial industry is Virtual Reality. Virtual Reality (VR) is a three-dimensional computer generated environment, usually displayed on a flat screen, a room-based system, or a head-mounted display. A unique advantage that VR offers is stereoscopic depth, which makes the viewer see objects in a virtual space and creates the illusion of reality. VR has also been suggested to elicit the sense of presence, which is the ability to make users feel like they are "there" and as a result evokes users to respond the same way in VR as in reality. One of the main determinants of presence is immersivity, which is characterized by factors such as field of view, field of regard, and display size [5].

Virtual Reality has already been used as an alternative tool for scientific research on the human information processing system. Kober and Neuper propose that presence is characterized by increased attention toward stimuli in the virtual environment and decreased attention to irrelevant stimuli [6]. Another study also concluded that immersive environments are better remembered by subjects [7]. VR is also being used in psychology research such as studies on PTSD treatment [8] and natural human behavior.

Visual search in real scenes could possibly be applied in virtual reality. Its application could provide insight on the ecological validity of "point and click" methods currently used in research on complex human processes such as attention and perception [5].

1.2   Problem Statement

There is currently no comparative study on the accuracy and efficiency of visual search in natural scenes between using Virtual Reality tools and 2-D displays, which may provide insight on the ecological validity of current methods.

2   Review of Related Literature

2.1   Classical Visual Search Methods

A study conducted by Wolfe discussed the standard paradigm for visual search [1]. There are two basic methods used in visual search: the reaction time method and the accuracy method. In the reaction time method, reaction time is the dependent measure used. Subjects look for a target along some distractor objects. The subject gives a response to indicate whether the target object is present or absent. Reaction time is analyzed as a function of set size. The set size is the total number of items in a display. The slopes and the intercepts of these reaction time x set size functions are used to infer the mechanics of the search. In the accuracy method, the display is presented briefly. The accuracy is plotted as a function of stimulus onset asynchrony. Classical methods use artificial objects such as letters or shapes as targets on artificial or blank backgrounds which are displayed on two-dimensional displays.

2.2 Visual Search in Real Scenes

A study by Wolfe et al. discussed visual search in real scenes [3]. This study determined if the search for arbitrary objects in real scenes are actually efficient, as believed from an introspective view and determined what guides efficient search in these scenes. The study tested the reliability of using the set size as an index for search efficiency in real scenes. Its reliability was already verified in artificial scenes in previous studies. Set size of a real scene was estimated by labelling objects found in the scene. Six different sets of methods were performed. Each differing whether a word cue or picture cue was presented. Some methods required localization of target objects by using a mouse. Some isolated the target object on a white background while another experiment used a black background. The reaction times for each experiment were compared to determine its efficiency. In analyzing the data, trials were removed if their reaction times fall out of the range depending on the scene presented. Statistical tests were performed to determine if the difference in reaction times for each scene are statistically significant. Results indicated that set size was a poor parameter for measuring search efficiencies in real scenes. It was also concluded that object search is not efficient outside of a scene context and efficient in a scene context. Therefore, it follows that the scene makes an important contribution to the efficiency.

Another study focused on establishing a good measurement of search efficiency in real scenes to replace set size. The researchers studied selected factors: set size, visible size (of target), visual crowding, and eccentricity. Visual crowding was measured using a variable, Ds which described the target-flanker separation in real scenes. Fourteen participants were presented with scene images from two datasets and were tasked to locate a target object. Image attributes (set size, visible size, etc.) were noted. Reaction times were recorded and analyzed against the factors to derive correlations. Results indicated that only visible size and target-flanker separation had significant effect on reaction times. Results indicated that reaction times decreased as visible size increased and as Ds increased [4].

2.3 Virtual Reality and Its Application in Human Factors and Behavior Research

A study conducted by Kozhevnikov and Dhond in 2012 assessed and compared visual-spatial processing of three-dimensional stimuli using non-immersive 2D displays and 3D immersive environments. In this study, experiments that focused on participants performing mental rotation tasks using 2D non-immersive, 3D non-immersive (3D Glasses), and 3D immersive (head mounted display) visual presentations were performed. Results of the experiments indicated that cognitive processing in a 3D immersive environment differs greatly from that in a 2D non-immersive and 3D non-immersive environment. Visual-spatial processing was also different in the immersive environment where participants were encouraged to use a viewer-centered frame of reference during the said tasks [9].

Another study by Lee et al. from 2003 investigated the potential of using virtual reality for cue exposure. The authors conducted the study with 22 male smokers half of

which will be tested with an immersive virtual reality, generated using information from a pilot survey on cues to nicotine craving, while the other half will be tested using pictures, the classical method. Participants were asked before and after the test on the level of their cravings. They have concluded that virtual reality is more effective at eliciting craving symptoms in individuals compared to using pictures, which they attributed to the added spatial stimulus that virtual reality has. This study highlights factors that may contribute to the results of the comparative study of the two methods. By observing how spatial stimulus and its interaction with visual stimulus affect results, this can help further distinguish the effectivity of both methods [11].

A study by Zhang et al. from 2016 compared the performance of participants in 3D/2D visual search tasks using artificial stimuli. The authors conducted an experiment with 16 subjects and used 2 different kinds of television, one with 3D polarization and the other having 3D switch, and recorded their performance and search time. The conditions of the television were found out not to be significant but the two visual methods, 2D and 3D, were significantly different, with the former having significantly longer times. It was concluded that the search environment had an impact on search performance. This information can help the researchers in analyzing results from using real scenes instead of artificial stimuli [10].

Another study by Li et al. also from 2016 investigated the relationship of memory to attention allocation in everyday actions. The authors compared the results of searching in a 3D environment and flat images of that same 3D environment. They had participants roam and search in the 3D environment of an apartment in empty rooms for different objects at each trial. They had another group rested their head and kept still while they comb through images of the 3D environment looking for certain objects. The results they have gathered showed that 2D and 3D search methods are almost the same however body movement allowed better use of memory for participants and help become more efficient in allocating attention by ignoring regions deemed insignificant. According to the authors, this is due to the spatial awareness from roaming the 3D environment compared to the minimal movement of the 2D search [12].

3   Methodology

An integrated methodology of the methods used by Wolfe et al. found in the related literature, which is suitable for both the VR method and 2D method, will be used by the researchers. Among the six methods, the method that exhibited the most efficient result became the basis of the methodology in the current study. The researchers planned their methodology accordingly to take into consideration body movements and memory when comparing methods by holding them constant to reduce contributing factors of variation.

Based on the posit that visual search is efficient in real scenes and the studies on Virtual Reality's ability to elicit real-world responses, the researchers propose that search times and accuracy would be significantly better using a virtual environment rather than two-dimensional displays. To test this hypothesis, the researchers performed a comparative visual search experiment.

3.1   Selection of Test Subjects

Cochran's formula was used to determine the sample size needed for the study.

$$n = \frac{z^2 pq}{e^2}. \tag{1}$$

The confidence level (z) was set to 95% (equal to 1.96), the level of precision (e) was set to 0.1 and the estimated proportion of the attribute (p) was set to 0.5 (q is 0.5 respectively). Given those parameters, the minimum number of randomly selected participants (n) is ninety-six (96).

All subjects were required to have normal or corrected-to-normal, and good color vision.

3.2   Gathering of Quantitative Data

The subjects were divided into two groups depending on the medium used in the visual search. Group A used the VR equipment while Group B used a laptop. The equipment used were the commercially available Samsung Gear VR and a 13-inch MacBook Pro (2011 model).

Each subject was given a short briefing before the task. The subjects were presented with objects as targets. Their task was to locate corresponding targets in five different real scenes and determine whether the target object is present or absent in the scene. The same scenes were shown for both methods to isolate variation due to set size.

For Group A, the subject was asked to adjust the focus of the VR equipment. For Group B, the subject was placed on the apparatus. The apparatus used was an $18 \times 18 \times 24$ inches box with a black background inside to emulate the viewing conditions of a VR. This was to minimize the variation between the two methods.

For each scene, a picture cue was flashed for two seconds before presenting the scene to avoid confusions with the target object. Sceneries are either indoor or urban. The subject responded "present" if they located the target and "absent" if they believe the target is not present in the scene presented. The accuracy and reaction time of the subject were recorded for each scene.

3.3   Gathering of Qualitative Data

Subjects from Group A were exposed to the classical (2D) method while subjects from Group B were exposed to the VR method. The subject was asked to rate how different the two methods were in terms of total visual experience. They were then asked to enumerate the differences or similarities of the two methods.

Each participant was asked in which method they think it would be easier to locate a target and why. This question gave an insight on the factors why the subjects showed a preference on one method over the other.

3.4   Data Analysis of Quantitative Data

The mean accuracy of subjects from both groups were compared. Accuracy was derived as a percentage of total number of correct responses (present or absent) over the number of participants in the method (50). Accuracy was calculated by using the following equation:

$$Accuracy = \frac{\# \, Correct \, Responses}{50} \times 100\% . \tag{2}$$

Times where subjects committed an error (i.e. false positive, false negative) were removed. A boxplot was generated with the errorless data for each scene per group to determine the acceptable range of each subgroup. Outliers were removed from the data afterwards.

Two-Sample T-Test was conducted on the response times of Group A and Group B per scene, testing the following hypothesis:

$$H_o : \mu_{2D} - \mu_{VR} = 0. \tag{3}$$

$$H_1 : \mu_{2D} - \mu_{VR} \neq 0. \tag{4}$$

The null hypothesis (H_o) states that the means of the both groups are equal, while the alternative hypothesis (H_1) states that they are not equal.

Test of Two Variances was next conducted on the response times of Group A and Group B per scene, testing the following hypothesis:

$$H_o : \frac{\sigma_{2D}}{\sigma_{VR}} = 1. \tag{5}$$

$$H_1 : \frac{\sigma_{2D}}{\sigma_{VR}} \neq 1. \tag{6}$$

The null hypothesis (H_o) states that the variance of the both groups are equal, while the alternative hypothesis (H_1) states that they are not equal.

3.5   Data Analysis of Qualitative Data

Data visualization was used to help analyze qualitative data. Subjects' ratings of difference in visual experience were graphed according to frequency and the average rating was obtained. Their responses to preferred method were graphed in a pie chart. Subjects' responses to the questions asked were tabulated accordingly to the category they fall under.

4   Results and Discussion

4.1   Study Demographics

One hundred (100) subjects were randomly selected to be a part of the study. Each group consisted of twenty-five (25) males and twenty-five (25) females.

4.2   Quantitative Data: Accuracy

From Table 1, it is seen that VR has a higher accuracy for each scene than 2D. This shows that the VR group is more accurate compared to 2D.

Table 1. Summary of mean accuracy for each scene per group and the difference between them.

Accuracy	2D	VR	Difference
Scene 1	66%	82%	16%
Scene 2	96%	98%	2%
Scene 3	92%	92%	0%
Scene 4	98%	100%	2%
Scene 5	94%	100%	6%

4.3   Quantitative Data: Time

After removing the reaction times of errors, the following boxplot seen in Fig. 1 was generated along with Table 2 showing the acceptable range:
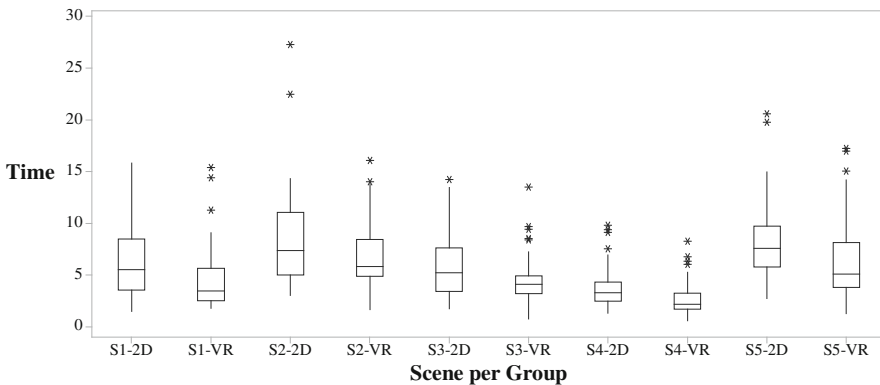

Fig. 1. Boxplot of scene times (in seconds) per group.

Reaction times that exceeded the corresponding limits seen in Table 2 were considered outliers and therefore removed from the analysis.

Table 2. Acceptable range of time (in seconds) for each scene per group.

	2D		VR	
	Lower limit	Upper limit	Lower limit	Upper limit
Scene 1	1.47	15.81	1.77	9.08
Scene 2	3.03	14.30	1.64	13.79
Scene 3	1.75	13.44	0.75	7.24
Scene 4	1.30	6.94	0.60	5.28
Scene 5	2.72	14.98	1.27	14.17

A Two-Sample T-Test for Means was conducted per scene using the adjusted errorless data generating the following results seen in Table 3.

Table 3. Summary of Two-Sample T-Test per scene.

μ	2D	VR	Est. difference	P-Value	Decision
Scene 1	6.38	3.93	2.456	0.002	Reject H_o
Scene 2	7.90	6.29	1.612	0.011	Reject H_o
Scene 3	5.67	3.77	1.895	0.000	Reject H_o
Scene 4	3.39	2.42	0.971	0.000	Reject H_o
Scene 5	7.64	5.95	1.690	0.007	Reject H_o

According to Table 3, the null hypothesis is rejected across all scenes as well indicating that there is significant difference in the mean reaction time of searching an object between 3D and 2D in all of the scenes. It was observed that the mean time of the VR group across all the scenes is smaller compared to the 2D group suggesting they accomplish the task much faster. A one sided Two-Sample T-Test was conducted to test the same null hypothesis and the following alternative hypothesis:

$$H_1 : \mu_{2D} - \mu_{VR} > 0. \tag{7}$$

Results rejected the null hypothesis as well showing the difference of VR and 2D with the performance of the VR group being indeed faster.

A Test of Two Variances was also conducted per scene using the same data generating the following results seen in Table 4.

Table 4. Summary of Test of Two Variances per scene. The lowest P-Value was selected between the Bonett's test and Levene's test results.

σ^2	2D	VR	Ratio	P-Value	Decision
Scene 1	14.235	3.986	3.571	0.012	Reject H_o
Scene 2	10.091	7.819	1.291	0.120	Do not reject H_o
Scene 3	8.093	1.969	4.110	0.000	Reject H_o
Scene 4	1.826	1.227	1.489	0.301	Do not reject H_o
Scene 5	7.644	9.578	0.798	0.520	Do not reject H_o

According to Table 4, the variance of the VR group is smaller in four out of the five scenes (scenes 1, 2, 3, and 4) meaning that the performance of the subjects was nearer to the mean time or more consistent. Two of these scenes concluded to reject the null hypothesis showing a significant difference in their variance.

Scene 5, however, showed that the VR group had a larger variance but was not statistically proven. Only two scenes rejected the null hypothesis suggesting that overall, there may not be sufficient evidence that proves that the VR group yielded more consistent reaction times than the 2-D group.

4.4   Qualitative Data: Visual Experience

The responses from the qualitative data gathering were summarized in Fig. 2 and Table 5.

According to Fig. 2 and Table 5, most subjects rated that the VR method and the 2D method are different from each other (with an average rating of 7.61). Subjects attributed this difference greatly to the immersivity of VR, which includes the realistic feeling, the depth perception, and the clarity of view to name a few of the subjects' comments.
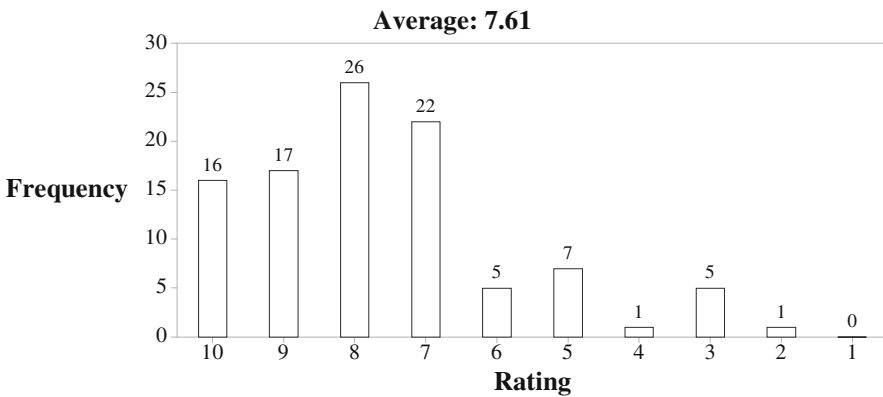

Fig. 2. Frequency of ratings for the difference in visual experience given by the participants.

Table 5. Frequency of most common differences between the VR method and 2D method cited by the participants

Differences	Frequency
Realistic	28
Clarity of view	27
Depth perception	21
Peripheral view	18
Immersivity	12
Spatial awareness	8
Focus	6

4.5   Qualitative Data: Preferred Method

The responses for the preferred method were summarized in Fig. 3, and Tables 6 and 7.

According to Fig. 3, majority of the subjects' preference is the VR method. Similar with the difference in visual experience, immersivity was a major factor in their choice, as seen in Table 6. Subjects who preferred the 2D method chose because of the advantage of seeing everything immediately, as seen in Table 7. Three subjects said that both methods are equally capable and did not have much difference when it comes to visual search tasks. The other two said it would depend on what object they are looking for and in what scene.
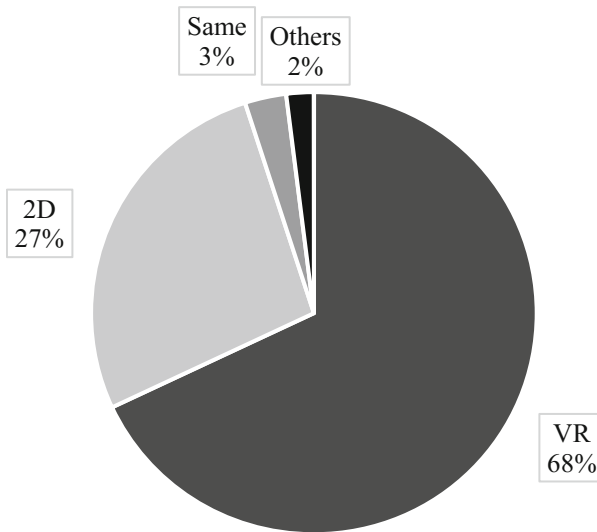

Fig. 3. Pie chart of proportion of preferred method.

Table 6. Frequency of most common differences between the VR method and 2D method cited by the participants

Strength	Frequency
Realistic	18
Depth perception	11
Clarity of view	11
Spatial awareness	9
Immersivity	7
Detail	6
Focus	6
Scaling	6
Peripheral view	5

Table 7. Frequency of most common differences between the VR method and 2D method cited by the participants

Strength	Frequency
Immediate visual	11
Flat	3
Bigger view	2
Stable	2
Clarity	2
Familiarity	2

5   Conclusion

Overall, the test subjects who used the VR method showed a better performance in visual search. The statistical tests indicated that their reaction times are significantly faster (by an average estimate of 28.62%) and more accurate as compared to those who used the 2D method in all scenes. It also indicated that the VR method results are more consistent and stable. Furthermore, majority of the subjects preferred using VR in visual search due to its immersivity, field of depth, and clarity of the visual experience.

6   Areas of Further Study

6.1   Localization of Target Objects

Guessing strategies based on the typicality of the target object in a scene are common for visual search especially for indoor scenes [3]. For instance, a subject will normally search for a remote control either on the couch or on table of a living room. To minimize possible guessing, localization of the target object can be done. This may strengthen the accuracy of results presented in the current study. Further studies can be done using better Virtual Reality equipment (one which allows localization or interaction with objects in virtual space). Although such tools are already available in the market, logistical and financial limitations prevented the researchers from using these tools.

6.2   Wider Range of Scenes and Targets

One of the limitations of the methodology used in this study is the restricted range of scenes. A study conducted by Zhang did not address the natural scenes (e.g. forests) due to the different context of these scenes, such as spatial knowledge [9]. Due to lack of incorporation of more natural scenes in the previous literatures, it would be interesting to include these scenes in future research.

6.3   Incorporation of Movement, Interaction, and Auditory Features

The current study was limited to exploring Virtual Reality only as a visual medium. Incorporating movement and interaction with the environment, also being key features

of VR, into the experiment could provide more insight on visual search efficiencies in real-life settings. Including auditory features that mimic real life scenarios could also maximize the level of presence and immersivity experienced by the users and increase the ecological validity of the method.

References

1. Wolfe, J.M.: Visual search. In: Pashler, H. (ed.) Attention. University College London Press, London (1998)
2. Le-Hoa Võ, M., Wolfe, J.: The role of memory for visual search in scenes. Ann. N. Y. Acad. Sci. **1339**(1), 72–81 (2015)
3. Wolfe, J.M., et al.: Visual search for arbitrary objects in real scenes. Atten. Percept. Psychophys. **73**(6), 1650–1671 (2011). Web. 2 Dec 2016
4. Zhang, X., et al.: Parametric modeling of visual search efficiency in real scenes. PLoS One **10**(6), e0128545 (2015). Web. 2 Dec 2016
5. Wilson, C.J., Soranzo, A.: The use of virtual reality in psychology a case study in visual perception. Comput. Math. Methods Med. **2015**, 1–7 (2015). Web. 2 Dec 2016
6. Kober, S.E., Kurzmann, J., Neuper, C.: Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: an EEG study. Int. J. Psychophys. **83**(3), 365–374 (2012). Web. 2 Dec 2016
7. Sutcliffe, A., Gault, B., Shin, J.-E.: Presence, memory and interaction in virtual environments. Int. J. Hum.-Comput. Stud. **62**(3), 307–327 (2005). Web. 2 Dec 2016
8. Nelson, R.J.: Is virtual reality exposure therapy effective for service members and veterans experiencing combat-related PTSD? Traumatology **19**(3), 171–178 (2013). Web. 2 Dec 2016
9. Kozhevnikov, M., Dhond, R.P.: Understanding immersivity: image generation and transformation processes in 3D immersive environments. Front. Psychol. **3**, n. pag. (2012). Web. 2 Dec 2016
10. Lee, J.H., et al.: Experimental application of virtual reality for nicotine craving through cue exposure. CyberPsychol. Behav. **6**(3), 275–280 (2003). Web. 2 Dec 2016
11. Zhang, Y., et al.: A comparative study on 3D/2D visual search performance on different visual display terminal. In: Hale, K., Stanney, K. (eds.) Advances in Neuroergonomics and Cognitive Engineering, pp. 233–242. Springer, Cham (2016). Web. 2 Dec 2016
12. Li, C.-L., et al.: Memory and visual search in naturalistic 2D and 3D environments. J. Vis. **16**(8), 9 (2016). Web. 2 Dec 2016