Visual search fixation strategies in a 3D image set: an eye tracking study

Article

Accepted Version

Visual Search Fixation Strategies in a 3D Image Set: An Eye Tracking Study

Markos Kyritsis[*a], Stephen R. Gulliver[a], Eva Feredoes[b]

[a]*Henley Business School, Business Informatics Systems and Accounting, Informatics Research Centre, University of Reading, RG6 6UD, United Kingdom*

[b]*School of Psychology and Clinical Language Sciences, University of Reading, Whiteknights Road, RG6 6AL, United Kingdom*

*****Corresponding author

email: m.kyritsis@henley.ac.uk

Abstract

In this study we explore whether inclusion of monocular depth within a pseudo-3D picture gallery negatively affects visual search strategy and performance. Experimental design facilitated control of i) the number of visible depth planes and ii) the presence of semantic sorting. Our results show that increasing the number of visual depth planes facilitates efficiency in search, which in turn results in a decreased response time to target selection, and a reduction in participant average pupil dilation – used for measuring cognitive load. Furthermore, results identified that search strategy is based on sorting, which implies that appropriate management of semantic associations can increase search efficiency by decreasing the number of potential targets.

Keywords: Usability testing, Laboratory experiments, Empirical studies in HCI, Eye tracking

Research Highlights

- Increasing visible depth layers in 3D pictures galleries increases likelihood of users adopting efficient (parallel) search strategies, which reduces time to target selection
- Semantically sorting the environment, without informing the user, can decrease time to target selection, making the search task more efficient.
- Trials with lower time to target selection will in turn reduce average pupil dilation –indicative of lower cognitive load.

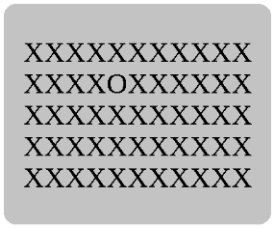
Introduction

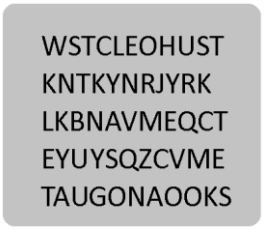
The introduction of a pseudo third dimension (i.e., 2D projections simulating 3D objects on a flat screen) within a Graphical User Interface (GUI) has witnessed sporadic bursts of interests over the years (Boyle et al., 1996; Robertson et al., 1998; Agarawala & Balakrishnan, 2006; Leal et al., 2009; Pakanen et al., 2013; Kyritsis et al., 2015a; Zuev, 2016). Despite this interest, and multiple attempts to introduce mainstream 3D user interfaces (3DUIs) - such as Looking Glass and Bumptop desktop – results have had limited commercial success. The issues that arise with the introduction of 3DUIs include: i) difficulties navigating cluttered or dynamic environments, e.g. objects being occluded or existing outside the user's field of view (Argelaguet & Andujar, 2013); ii) problems integrating available input devices; iii) problems identifying appropriate interaction techniques (see Jankowski & Hachet, 2013 for a review); and iv) a lack of benefits gained, especially when undertaking search tasks (e.g. looking for a specific file in a folder) (Kyritsis et al., 2013).

Visual Search Paradigms

One increasingly significant shared function, within both social media and/or graphical operating systems, is the storage, retrieval, and sharing of photos and images. Finding the right image, however, within an ever-expanding picture library, is a complex visual search task. Interestingly this problem is commonly used by cognitive psychologists to understand visual attention and working memory (see Wolfe, 2014, for a review), as it allows the controlled exploration of the task; allowing researchers to build design recommendations by extracting properties from the environment that can be controlled in order to make the search more efficient.

All pictorial information is made up of low-level features, such as color, size, and orientation of lines, which bind together to create a coherent object representation (conjunctions) that normally carries semantic meaning (e.g., a face, an animal, a building etc.). Studies have shown that a contrast between the features of the target and the distractors (i.e. non-target images) can alter the efficiency of the visual search task. In extreme cases, a target can differ significantly from the distractor set (see figure 1a). In such a case a 'feature search' occurs – see figure 1a – where visual processing is done pre-attentively (Treisman and Gelade 1980); i.e. the search is not susceptible to the number of items, and the inclusion of additional distractors won't affect the response time to target selection in a linear way (Desimone and Duncan 1995).

a) 

b) 

c) 

Figure 1 – a) finding the 'O' amongst the 'X' requires a 'feature search'; b) finding the letter 'G' requires a 'conjunction search'; c) a suitable search strategy allows a decrease in number of distractors.

As the homogeneity of features in the distractor-set decreases, the visual search task becomes less efficient. A 'conjunction search' is therefore more likely to occur in picture galleries where the heterogeneity of item features in the image set is high (Treisman and Gelade 1980) - see figure 1b. During a conjunction search, efficiency strategies are adopted in order to decrease the set size. Egeth et al. (1984) showed that sorting the stimulus can significantly support search efficiency. For example, searching for letters 'G' amongst a set of randomly generated alphabet letters (figure 1b) requires a serial search, however, if half the letters are sorted as capitals, and the other half are sorted as lower-case letters (see figure 1c), then all lower case letters can be ignored; decreasing the set-size by half and making the serial search more efficient.

The interaction between bottom-up features and top-down strategies was modelled by Wolfe (1994) in the second 'Guided Search Model', which proposes that perceived stimuli is filtered through pre-attentive 'categorical' channels (e.g., for color, orientation, etc.). These categorical channels produce a mental feature map, which activates in response to feature differences (bottom-up) and/or task demands (top-down). The end result is an activation map that unconsciously guides the visual system towards the target. Wolfe et al. (2011) suggested that the process could be combined with a non-selective visual processing pathway, to help the viewer get 'the gist' of the environment (i.e. low level semantic information) in order to facilitate target selection. It has been hypothesized that searches can be facilitated if stimuli contains: i) differences in contour shapes; ii) semantically different pictures (Levin et al, 2001); iii) higher level of luminance; and/or iv) a low number of edges (Kyritsis et al., 2016). There is also some evidence – although limited – that semantic information can be used to guide visual searches; with some pictorial semantic categories - particularly animal pictures - facilitating response time to target selection (Levin et al, 2001, Kyritsis et al., 2016). Finally, semantic information can increase conjunction search efficiency through the process of category grouping; even when users are not explicitly told that the environment has been sorted (Kyritsis et al., 2016). In other words, if the item set is sorted according to semantic categories, then people will either i) unconsciously avoid irrelevant pictorial categories, i.e. to limit their search to more relevant items, or ii) quickly learn the semantic associations in the item set (for example see figure 2).

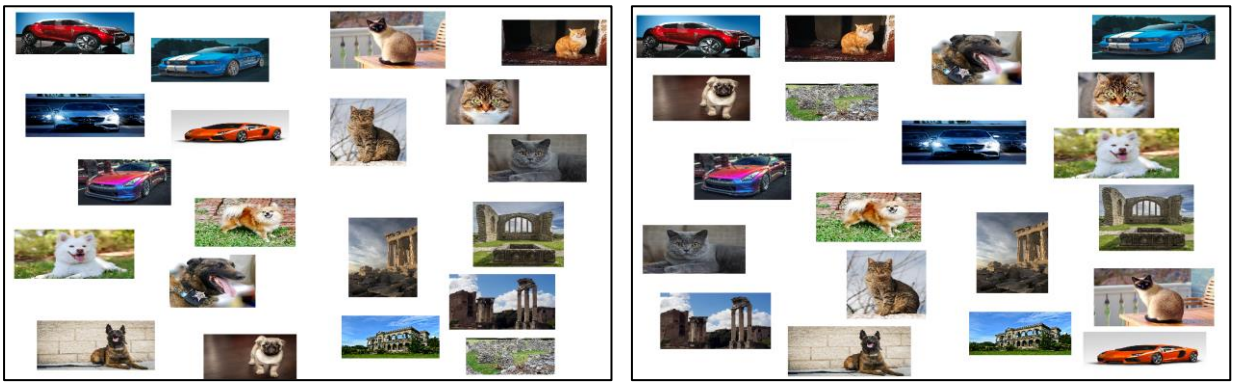


Figure 2 - Grouped semantic pictures in clusters (left), or ungrouped (right). Studies show that semantic grouping facilitates target selection by decreasing the relevant stimulus set size.

Conjunction search On Monocular Depth

Even though there is a substantial body of literature concerning human 2D visual search, there is limited research considering human behavior and performance during a pseudo-3D visual search task. One might theorize that increasing the item (image) set, i.e. by introducing more layers of monocular depth during a conjunction search (i.e. the visible depth window on the z-axis), may at best not affect the response time, yet at worst will create additional clutter and/or complexity that negatively affects the visual search process. Finlayson and Grove (2015) found that increasing binocular depth, as well as increasing the number of depth planes, led to a decrease in search efficiency. In contrast, however, Pomplun et al. (2013) found that participants search differently depending on the task. Participants undertaking a conjunction search start by fixating on the center of the image set; allowing their gaze to be extrafoveally guided from the center by stimulus properties (e.g. color). When participants undertake a feature search, participants start the search in the upper-left corner, and apply left-right and top-down eye pursuits. Results suggested that inclusion of depth did not affect search method or performance.

A proposed experimental platform for studying visual search tasks in a WIMP (Windows Icons Menus and Pointers) styled 3D environment was suggested in Kyritsis et al. (2013) (figure 3). Item sets were placed on virtual depth planes, which activated depth perception via monocular cues; navigating in and out of the environment (on the z-axis) using the mouse scroll wheel. The environment could be rotated by holding down the middle mouse button and moving the mouse of the horizontal axis, this way more pictures in the posterior depth planes were made visible. The response time to target selection was measured for conjunctively defined items of varying perceptual salience; controlled by distinct color differences

between target and distractors. Kyritsis et al. (2013) found that depth facilitates response time to target selection, but only for perceptually salient targets.

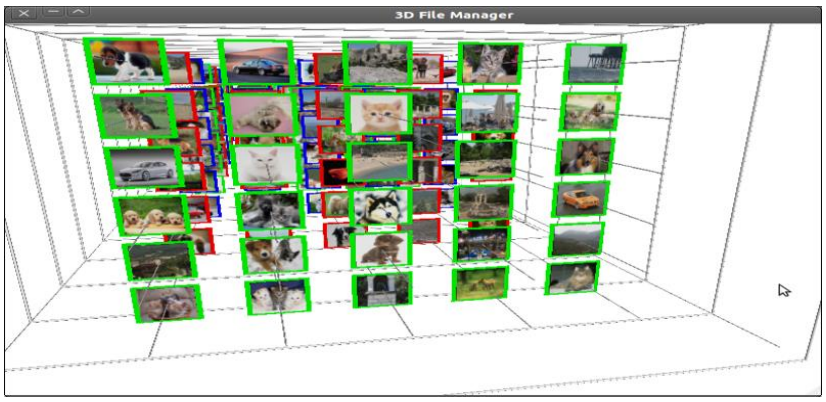


Figure 3 - Experimental platform used by Kyritsis et al., (2013). Notice the lack of culling on visible depth planes and visual depth aides in the form of converging (Ponzo) lines.

Kyritsis et al. (2016) investigated the effects of monocular depth on a visual search task in a 3D pictures folder. In two separate experiments: a pilot study, which was conducted online, and a lab-based experiment using an eye tracker. The experimental platform resembled the one by Kyritsis et al., (2013), however, rotating the environment was only implemented in the online study. In the eye-tracking study this functionality was disabled in order to reduce confounding variable effects. The authors found that there is a complex interplay between bottom-up features and top-down semantic guidance, which can alter the response time to target selection. For example, a contrast in luminosity, as well as the number of edges (item complexity) between target and surrounding distractors, impacts response time and number of errors - i.e. when a distractor is selected instead of the target. As a result of this interplay, the authors reported that increasing the number of pseudo-depth planes decreases the number of selection errors, and decreases the response time to target selection. These findings were in-line with the model discussed by Wolfe (1994), i.e. where the efficiency of a conjunction search varies depending on low-level (features) and high-level (task-demands) attributes. Although used attentional resource cannot be reliably measured via eye tracking, the focus of attention via fixations is indicative of target detection. Accordingly, Kyritsis et al. (2016) incorporated eye tracking within their study to better understand the focus of attention in varying levels of monocular depth. Results showed that an increase in the number of visible depth planes - from two to three - increased the ratio between the posterior layers (i.e. items in all background layers) and the anterior layer (the items on the current / active / visible layer). The reason for why an increase in fixations facilitates target selection was not explored.

In summary, there is a lack of clarity as to whether the introduction of pseudo-depth planes into a pictorial gallery can benefit the visual search task. There is evidence, however, in both theoretical and experimental literature, of a complex interplay between conjunctive features of targets and distractors, however whether this is due to an increase in the number of fixations, or simply limited to extrafoveally guided smooth pursuits is as yet to be explored. We hypothesize that even though an increase in visible monocular depth will lead to an increase in set size, certain item features will have a larger likelihood of attracting the focus of attention, and therefore result in a shorter response time to target selection.

Current Study

In this study we aim to investigate: a) whether reduced response time occurs in semantically sorted environments due to participants unconsciously learning semantic associations; b) whether an increase in the number of monocular depth planes can alter visual search strategies by introducing a larger disparity between the number of fixations on the anterior layer (i.e. items on the current / active / fully visible layer) and the posterior layers (items on lower z-axis / background layers); c) if the ratio of time spent fixating on items in posterior planes over the total number of fixations on all depth planes can reliably predict the response time to target selection; and d) whether an increase in posterior layer fixations and semantic sorting leads to a decrease in cognitive load. The experiment received ethical approval by the School of Psychology and Clinical Language Science at the University of Reading.

Method

Participants

We recruited 18 postgraduate students (15 female, 3 male) in the age range of 21-27 to take part in our experimental study. All participants had normal or corrected to normal vision. Color blindness was not reported by any of the participants. Information sheets and consent forms were provided to the participants, and informed consent was taken prior to starting the eye tracking experiment. Participants were made aware that no compensation would be provided, and that they could stop the experiment at any time without need of justification. None of our participants reported any nausea or fatigue by the end of the experimental study.

Apparatus

Capture of gaze location, fixation duration, and cognitive load was required in order to support the study outcome. There are several instruments described in literature for measuring cognitive load. Paas et al., (2003), for example, validated use of pupillary response as a very sensitive instrument for measuring fluctuations in cognitive load. Accordingly, since our experimental design already incorporates eye

tracking, we captured pupillometry as the instrument for measuring cognitive load. We used an Eyelink 1000 eye tracker (SR Research, Montreal) to record eye movements and pupil size. A chin rest was placed 70cm away from a 28" CRT colour monitor with a 16:10 aspect ratio, the default colour and luminance settings, and a refresh rate of 75Hz. The screen resolution was kept at 1920 x 1200 pixels, and the sampling rate for the eye tracker was set to 500Hz. Standard 9-point grid calibration was used prior to the start of all trials, with an average error of less than 0.5 degrees as the threshold for an acceptable calibration. The room itself was built to control for fluctuations of light from external sources and was dedicated for eye tracking studies only. However, due to the nature of the experiment, controlling fluctuations of luminance from the experimental platform itself was not possible.

Materials

Four diverse semantic groups (i.e. animal; person; groups of people; and landscapes/landmarks) were chosen to simulate the diversity of content observed on social media sites. Each semantic group contained 76 image items – 304 image items in total, which were randomly scrambled for each trial. The 3D environment was made up of image layers, each containing 4 x 4 images (see figure 4). In each trial there were 19 stacked image layers, however the participant was only able to view the anterior (i.e. the top / visible) layer, and up to three posterior (i.e. background) layers - depending on the experimental condition.

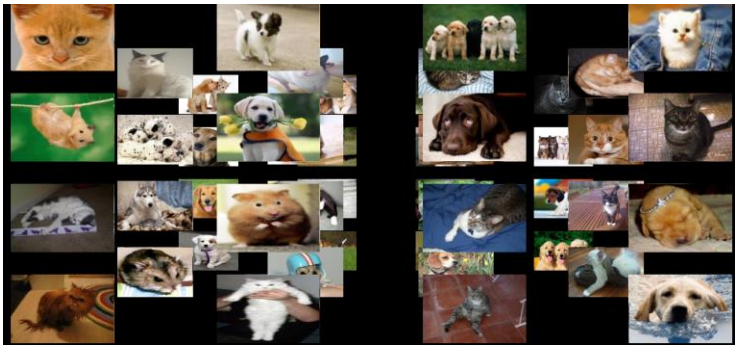


Figure 4 - Environment used for the experiments. The condition shown is three layers of depth (4x4 items in the anterior layer), semantic sorting, and no visual aides (Ponzo lines).

The experimental platform allowed the manipulation of: the number of visible layers (two, three, or four visible layers), semantic sorting (sorted/unsorted), and visual depth aides (i.e. converging 'Ponzo' lines to reinforce monocular depth or no converging 'Ponzo' lines); however impact of 'Ponzo' lines is not included in this experiment. For all trials, the starting camera position was on the first layer (i.e. the top layer of the stack). Rotation of the environment was not allowed to ensure control consistency and ease of

use. Semantically sorted trials were grouped together to form blocks (see figure 5 top). Unsorted trials were randomly positioned (see figure 5 bottom).

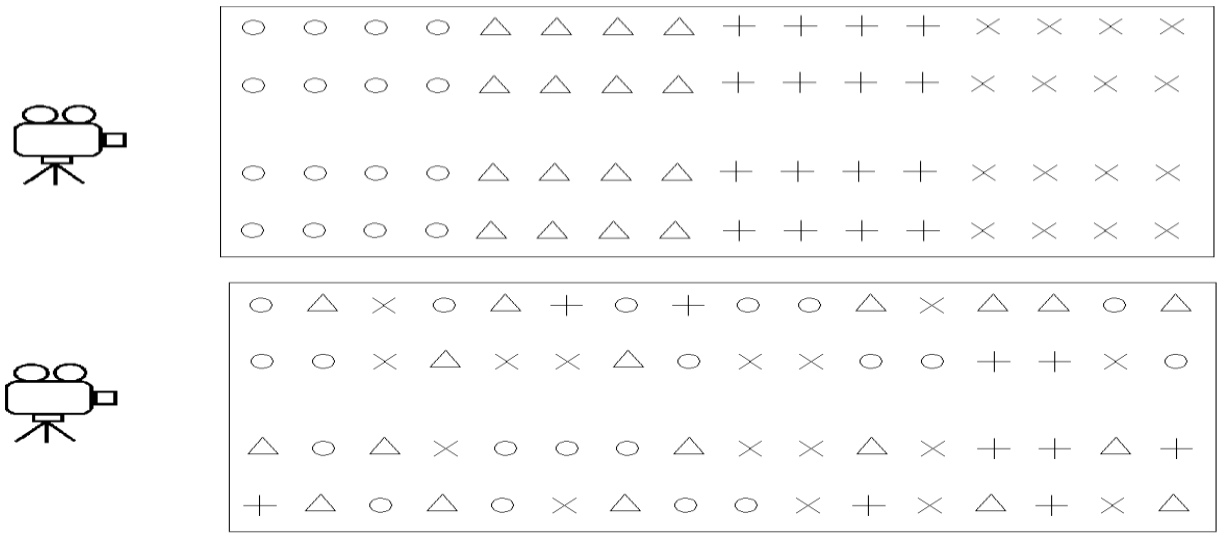


Figure 5 - Conceptual top-down view of semantically sorted trials (top), vs unsorted trials (bottom). The shapes represent semantic categories (i.e. animals; person; groups of people; landmarks/landscapes).

Procedure

Our experimental study applies a 3x2 design, with a total of six conditions. These conditions relate to inclusion of: two, three, or four layers of monocular visual depth, and use of semantically sorted/unsorted image items. The order that the conditions were presented was completely randomized for each participant in order to control for any learning effects. Before undertaking the calibration process, participants were verbally briefed about the nature of the experiment. Participants were not, however, explicitly made aware that some trials were semantically sorted.

At the start of each trial, the target picture appeared on the screen until the participant pressed the spacebar. Once the trial began, participants were able to navigate the environment using the mouse scroll in order to move in and out of the image stack, i.e. moving the camera - using the mouse wheel - along the z-axis. For each trial, the top layer in the stack was defined as the anterior layer. As the participant shifted the shifting camera window on the z-axis, lower layers on the stack become visible as higher levels on the stack disappeared behind the camera view. Although the visible window (i.e. 2 to 4 visible layers depending on the experimental condition) shifted on the z-axis stack, the top (i.e. most visible) layer was always defined as the anterior layer. Lower visible layers were defined as ‘the posterior’. Movement on the mouse wheel shifted the camera forward on the z-axis, however the maximum number of visible layers within the visible window was always limited by the experimental condition. Upon target detection, the participant selected the target by clicking it with the mouse. Wrong selections, i.e. where participants click on a

distractor item instead of the target item, were highlighted with a red hue; thus providing feedback to the participant that repeated clicks were not required. After 60 seconds the participant was free to fail the trial by hitting the 'f' key. The full experiment ended i) after 64 trials were completed, or ii) after 60 minutes - whichever came first. The average amount of trials completed by participants was 54.33 (sd = 11.31, min = 32, max = 64). If required, participants could rest every 10 minutes. If participants rested, the time was paused and the eye tracker was re-calibrated prior to resuming the study. The process has been summarized in figure 6.

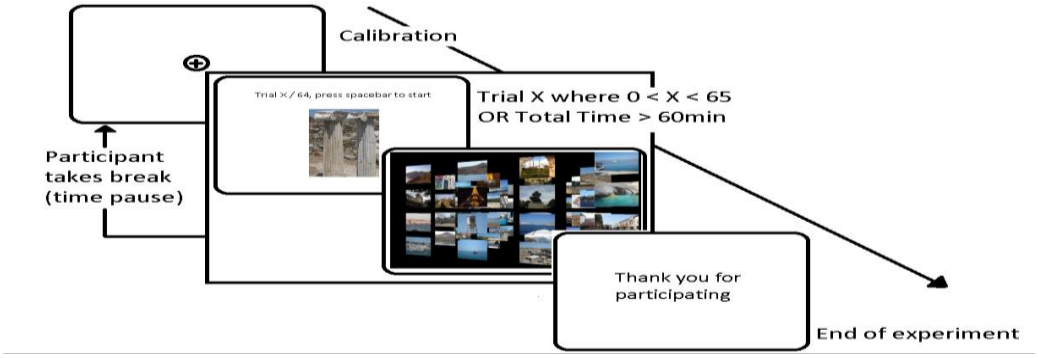


Figure 6 - Illustrative summary of the experimental procedure.

Extracted Environmental Properties

Graphical properties were extracted from both the target, and from surrounding distractors for use as dependent variables. These properties included: item fixations, which were extracted by using a ray-casting algorithm to detect when participant fixations 'collided' with an image item; the depth layer of all collided items (defined as either anterior or posterior); the average pupil size throughout the trial, which was used as a measure of the cognitive load in each trial; and the position of the target in the item stack - starting from 1 (top left corner of top-most layer) to 304 (bottom-right corner of final layer in the stack). The amount of time spent fixating on the posterior layers vs the anterior layer was evaluated by taking the ratio of the posterior fixations over the total fixations, i.e., $FixRatio = fix_{Pos} / (fix_{Ant} + fix_{Post})$

FixRatio was used to support simple analysis of the participant search strategy. If participants undertake a conjunction search (i.e. a linear / serial search considering the features of each anterior item in turn) then there will be a reduction in the number of fixations in the posterior layers and *FixRatio* will be small. If a participant is primarily undertaking a feature search (i.e. where participants are drawn by pre-attentive image features - independent of layer) then there will be an increased number of fixations in the posterior layers and *FixRatio* will be larger.

Results

With the help of the 'lme4' library (Bates et al., 2014), linear mixed effects models were built to support analysis. Marginal R-squared (variance explained by the fixed variables) and conditional R-squared (variance explained by the fixed and random variables) were retrieved using the 'MuMIn' library (Barton, 2018). The plots were drawn with the help of the 'plot_model()' function found in the 'sjPlot' library (Lüdecke, 2018). The presence of Ponzo (depth) lines was shown to have no impact on the measurements (Kyritsis et al., 2016), therefore the condition was not considered within this analysis.

Effect of Posterior Item layer Fixations on Response Time to Target Selection

We used linear mixed effects modeling (LME) with participant ID as the random variable, RT (i.e. response time to target selection) as the dependent variable, and *FixRatio*, i.e. $fix_{Pos} / (fix_{Ant} + fix_{Post})$, as the independent variable; allowing us to explore the impact of *FixRatio* – representing search strategy - on RT. The results of our model suggests that *FixRatio* has a significant effect on RT (b0 = 45356, b1 = -26495, t(873) = -3.59, p < 0.001, R_m^2 = 0.02, R_c^2 = 0.16) (figure 7). Accordingly, our results suggest that an increase in posterior layer fixations reduced response time to target selection, independent of semantic sorting. However, the amount of variance explained by the fixed effect was very small (~2%), with random effects explaining a much larger amount of variance in the model (total variance explained by the model being 16%). Therefore, we suggest that variation in participant performance is only partially explained by visual search strategies in this particular type of 3D environment.

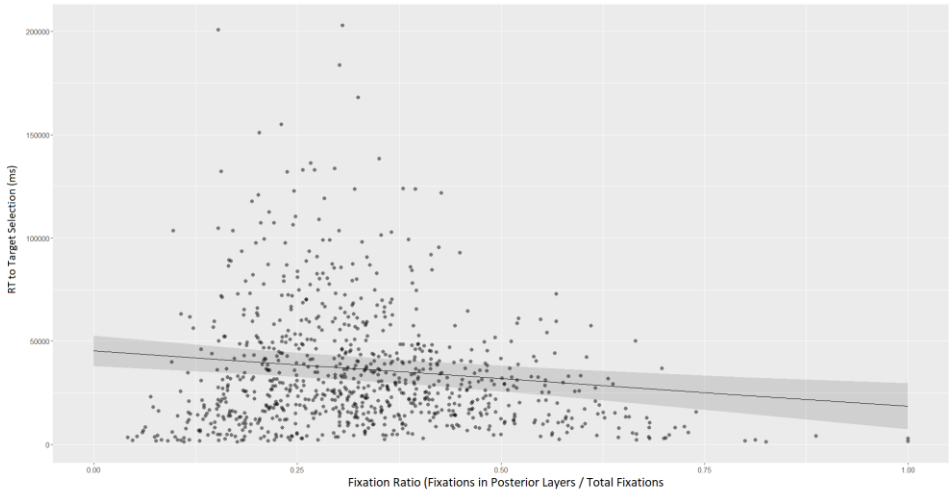


Figure 7 –Increased Fixation Ratio results in decreased response time (RT), indicating that more fixations on posterior layers is indicative of a parallel search. Only the fixed effect is shown for the sake of clarity.

Effect of Semantic Sorting and Depth on Anterior vs Posterior Fixations

LME modelling, using participant ID as the random variable, suggests that there is a main effect of semantic sorting [b0 = 0.2, b1 = 0.02, t(854) = 3.02, p < 0.01] and depth [b2 = 0.04, t(854) = 8.2, p < 0.001, R_m^2 = 0.06, R_c^2 = 0.32] on *FixRatio,* but no significant interaction between the two. Our results suggest that (a) participants spend (on average) more time fixating on items in the posterior planes within semantically sorted trials (figure 8), and (b) participants spend more time fixating on posterior planes in three and four layers of depth trials; i.e. when compared to two layers of depth trials (figure 9). Confidence interval plots indicate that there was no difference in the amount of posterior plane fixations between three vs four plane trials (see figure 9). The authors suggest this is due to increased level of visual obstructions, which implies that use of more than three image layers has limited value when 3D rotations is restricted.

Trials with two visible depth planes is less efficient, as *FixRatio* is smaller. A smaller *FixRatio* suggests that, for two visible layers, participants prefer a conjunction search strategy; resulting in an increase in anterior layer fixations, and a decrease in search efficiency. This increase in anterior layer fixations diminishes when three or four visible planes is made available. We hypothesize that a switch occurs in the participant search strategy when there are more than two visual layers. Increased posterior distractions, which distracts participants from a systematic conjunctive search, actually results in a more efficient search outcome.

Figure 8 - Semantically sorting the item set led to an increase in posterior layer fixations (larger ratio of fixations on posterior layers over total number of fixations).

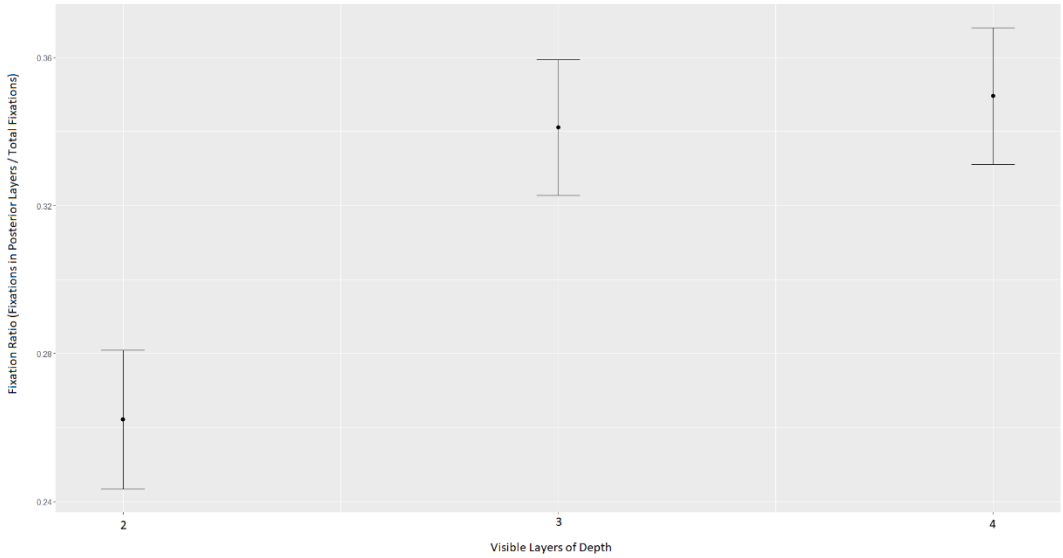


Figure 9 - Increasing depth from two to three layers led to an increase in posterior layer fixations.

Interaction Between Semantic Sorting and Target Position

A linear mixed effects model, with participant ID as the random variable and RT as the dependent variable, showed a significant interaction between Semantic sorting and target position in the set (b1*b2 = -65, t(853) = -3.37, p < 0.001, R_m^2 = 0.19, R_c^2 = 0.33). The interaction between the two variables indicates that as distance of the target (from the starting position) increases, RT increases; however the gradient impact on time to target is greater for unsorted trials (figure 10). This result implies that i) searching in a sorted image set is easier, and ii) there is a learning effect - represented by the reduced gradient in the sorted sample.

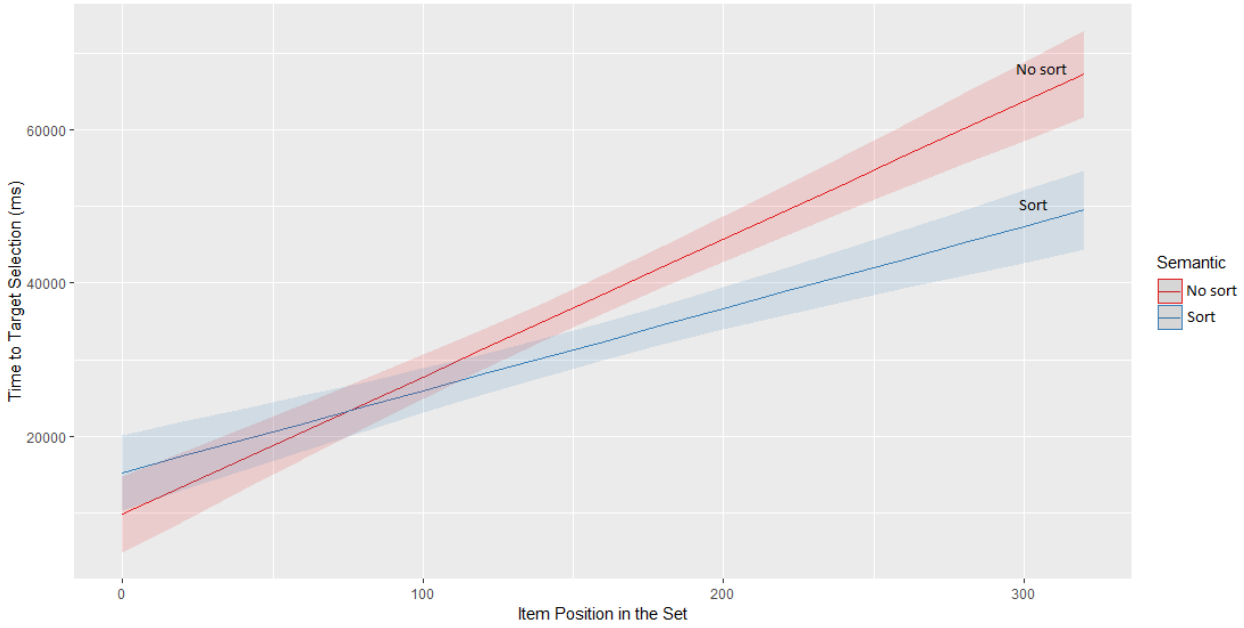


Figure 10 - Interaction between semantic sorting and position indicates an unguided learning effect, i.e. rather than an unconscious attribute-guided effect.

Variables that Impact Pupil Dilation (Cognitive Load)

Mixed effects modelling showed that neither semantic sorting nor number of visible depth planes had an effect on pupil dilation. There were, however, weak but significant main effects i) of *FixRatio* on average pupil dilation (b0 = 0.35, b1 < -0.01, t(854) = -2.2, p = 0.05), and ii) target position (b2 < 0.01, t(854) = 3.57, p < 0.001, R_m^2 = 0.02, R_c^2 = 0.28); but no interaction between the two. These findings suggest that the more time a participant spends on the anterior layer the larger the average pupil dilation, which we hypothesize could potentially correlate to increased levels of cognitive load (Paas et al., 2003) as a result of undertaking a serial search. As distance between the target and the starting position increases, so does pupil dilation. We hypothesize that this result indicates either i) increased cognitive load, ii) increased search frustration - as discussed by McCuaig et al. (2010), or iii) an unsystematic effect due to uncontrolled physiological or experimental artefacts. Additional research is required to qualitatively investigate this condition.

Interaction between semantic sorting and position did not impact pupil dilations. Moreover, there was no significant difference in pupil dilation between semantically sorted and unsorted trials. These results indicate that despite the general decrease in response time (due to semantic sorting), the lack of explicit instructions – i.e. the categories used to sort the environment – did not impact pupil dilation (i.e. a possible measure of cognitive load). However, indirectly, the overall effect of cognitive load is reflected in the time

to target selection (RT), which was a significant predictor of average pupil size during the trial (b0 = 2068.71, b1 = 0.0018, t(855) = 5.26, p < 0.0001, $R_m^2 = 0.01$, R_c^2 = 0.84) (figure 11).

Figure 11 – Total time spent searching for the target had a causal effect on average pupil size during the trial (shown in arbitrary units).

Discussion

In this study we aimed to investigate (a) whether reduced response time in semantically sorted environments was due to conscious participant learning of semantic associations, (b) whether an increase in the number of monocular depth planes can alter visual search strategies by introducing a larger disparity between anterior layer and posterior layer item fixations, (c) if the ratio of fixations on items in posterior depth planes over total fixations can reliably predict response time to target selection, and (d) whether an increase in posterior layer fixations and semantic sorting leads to a decrease in pupil dilation - a possible measure of cognitive load.

Our results suggest that semantically sorted trials had a lower response time to target selection, however, there was an interaction with the target position. This interaction indicates that, despite participants not being made explicitly aware that some trials would be sorted, mental associations were still formed, suggesting a learning effect - i.e. targets in the posterior layers of the stack were more efficient in the semantic trials than in non-semantic trials.

Our study showed that in order to facilitate posterior fixations, increase the number of visible depth planes to greater than two. Our results, however, failed to show a benefit of increasing the visible depth past three layers – which we assume to be due to obstruction effects brought on (in part) by the removal of rotation camera controls.

Our analysis indicates that an increase in the number of anterior layer fixations is a significant predictor of response time to target selection, i.e. posterior layer fixations are essential to achieving more efficient visual search strategies. It seems that increasing the number of monocular depth planes does not negatively affect visual search, i.e. by introducing negative clustering effects (at least not in the defined picture gallery paradigm). Alternatively, an increasing the number of planes in the visual window can be used – when combined with the correct content - to facilitate visual search efficiency.

Finally, our analysis suggests that a decrease in response time to target selection is accompanied by a decrease in average pupil dilation; evidence linking search strategy/efficiency with cognitive load. With two visual planes, participants are more likely to undertake systematic conjunction searches, which increases levels of cognitive processing. With three or more visual planes, participants seem to become more distracted by item features – possibly due to the increased number of items in view and/or the increased level of visual complexity of the interface resulting in a switch from use of a conjunction search to a more feature led search. We showed that, in context of a layer stack, feature search is more efficient, however since feature processing is pre-attentive, it also requires less cognitive processing; evidenced by a reduction in pupil dilation.

To summarize, we provide compelling evidence to suggest that time to target selection in picture folders can be optimized by (a) introducing monocular depth and, (b) introducing semantic sorting to the pseudo-3D layout.

Conclusion

As the amount of pictorial information increases, the time spent browsing for pictures to share with others on personal devices also increases. The need to decrease the total time, and cognitive load, spent looking for photos is important for a more positive user experience. But is this achievable by introducing monocular depth in the face of risking an increase in clutter and distraction effects? This study shows that the inclusion of monocular depth in pseudo-3D image galleries does not decrease efficiency of visual search. Results showed that use of increased depth actually facilitates increased fixations in the posterior (back) layers, which in turn changes participant search strategy, thus decreasing the amount of time, and cognitive load (measured through use of pupillometry), taken to locate/select the target. Moreover, by sorting pictorial information – i.e. by semantic meaning – it is possible to improve search performance by decreasing time and resources used processing irrelevant picture distractors. This facilitation effect was identified even

though participants were not explicitly informed that the sample was sorted; evidencing the presence of a learning effect.

We would like to acknowledge some the limitations of our study. For example, the experiment was limited to using monocular depth cues for the sake of eye tracking. However, we aim to replicate the experiment in the future using a VR headset in order to confirm the results using binocular depth cues. Furthermore, our sample size was small (18 participants), resulting in a low statistical power. This effect was mitigated by using linear mixed effects modelling, rather than aggregating data, which lead to a sample size of 854 trials, yet use of a more extensive sample would allow consideration of difference as a result of participant demographic and individual variables. Finally, our sample was strongly biased by including more female participants, therefore limiting the generalizability of our findings, and leading us to present our results with caution. Despite these limitations, our results provide evidence to suggest that semantic sorting makes visual search more efficient, yet the current study is not able to indicate whether this was due to bottom-up conjunctive features or top-down activation from the non-selective visual pathway (see Knudson, 2007 for a discussion on visual pathways). Even though additional research in this area is encouraged, results suggest that allowing users to sort items by semantic category can greatly facilitate visual search – by decreasing the set size of the conjunctively defined stimuli. We therefore suggest that designers include monocular depth planes to pictorial galleries, along with semantic sorting to facilitate visual search tasks conducted by users.

References

Argelaguet, F., & Andujar, C. (2013). A survey of 3D object selection techniques for virtual environments. Computers & Graphics, 37(3), 121-136.

Agarawala, A., & Balakrishnan, R. (2006, April). Keepin'it real: pushing the desktop metaphor with physics, piles and the pen. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 1283-1292). ACM.

Barton, K. (2018). MuMIn: multi-model inference. http://r-forge. r-project. org/projects/mumin/.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Boyle, J., Leishman, S., & Gray, P. M. (1996). From WIMPS to 3D: The development of AMAZE. Journal of Visual Languages & Computing, 7(3), 291-319.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual review of neuroscience, 18(1), 193-222.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. Journal of Experimental Psychology: Human Perception and Performance, 10(1), 32.

Finlayson, N. J., & Grove, P. M. (2015). Visual search is influenced by 3D spatial layout. Attention, Perception, & Psychophysics, 77(7), 2322-2330.

IDC Device Market Trends (2018). AR & VR Headset Market Share. Retrieved from https://www.idc.com/promo/arvr

Jankowski, J., & Hachet, M. (2013, May). A survey of interaction techniques for interactive 3D environments. In Eurographics 2013-STAR.

Jeffreys, H. (1961). Theory of probability, Clarendon.

Knudsen, E. I. (2007). Fundamental components of attention. Annu. Rev. Neurosci., 30, 57-78.

Kothari, J. (2019, May 20). Glass Enterprise Edition 2: faster and more helpful. Retrieved from https://www.blog.google/products/hardware/glass-enterprise-edition-2/

Kyritsis, M., Gulliver, S. R., Morar, S., & Stevens, R. (2013, October). Issues and benefits of using 3D interfaces: visual and verbal tasks. In Proceedings of the Fifth International Conference on Management of Emergent Digital EcoSystems (pp. 241-245). ACM.

Kyritsis, M., Gulliver, S. R., & Feredoes, E. (2016). Environmental factors and features that influence visual search in a 3D WIMP interface. International Journal of Human-Computer Studies, 92, 30-43.

Leal, A., Wingrave, C. A., & LaViola Jr, J. J. (2009, September). Initial explorations into the user experience of 3D file browsing. In Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology (pp. 339-344). British Computer Society.

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. Perception & Psychophysics, 63(4), 676-697.

Lüdecke, D. (2018). sjStats: statistical functions for regression models. R package version 0.14, 3.

McCuaig, J., Pearlstein, M., & Judd, A. (2010, June). Detecting learner frustration: towards mainstream use cases. In International Conference on Intelligent Tutoring Systems (pp. 21-30). Springer, Berlin, Heidelberg.

Pakanen, M., Arhippainen, L., & Hickey, S. (2013). Studying four 3D GUI metaphors in virtual environment in tablet context. In 6th International Conference on Advances in Computer-Human Interactions (pp. 41-46).

Pomplun, M., Garaas, T. W., & Carrasco, M. (2013). The effects of task difficulty on visual search strategy in virtual 3D displays. Journal of vision, 13(3), 24-24.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. Educational psychologist, 38(1), 63-71.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., & Van Dantzich, M. (1998, November). Data mountain: using spatial memory for document management. In Proceedings of the 11th annual ACM symposium on User interface software and technology (pp. 153-162). ACM.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive psychology, 12(1), 97-136.

Wiley, J. (2016, Mar 31). Putting the "Real" in "Virtual Reality". Retrieved from https://www.blog.google/products/cardboard/cardboard-plastic/

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. Psychonomic bulletin & review, 1(2), 202-238.

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. Trends in cognitive sciences, 15(2), 77-84.

Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided search. Oxford handbook of attention, 11-55.

Zuev, A. S. (2016). Virtual four-dimensional environments for human-computer interaction. Automation and Remote Control, 77(3), 533-550.