Visual cognitive styles in virtual environments: Constructing and evaluating adequate tasks

Pavel Ugwitz, Vojtěch Juřík, Čeněk Šašinka

Department of Psychology, Faculty of Arts MU
HUME Lab, FF MU
Arna Nováka 1
602 00 Brno

172577@mail.muni.cz
jurik.vojtech@mail.muni.cz
cenek.sasinka@mail.muni.cz

Abstrakt

Vizuální kognitivní styly se manifestují způsoby, kterými lidé implicitně zpracovávají vizuální podněty z okolí. Je však otázkou, jak vizuální kognitivní styly měřit. Klasické způsoby měření využívají dotazníků, či jednoduchých obrázkových úloh. Tyto způsoby však ze své podstaty nemohou zachytit komplexitu skutečných vizuálních kompozicí. S nástupem (mobilních) eye-trackingových zařízení se sice možnosti měření kognitivních stylů posouvají, jedná se zde však pouze o receptivní zpracování reálného vizuálního podkladu. Cílem tohoto příspěvku je představit možnosti i úskalí, které v sobě skýtá konstrukce virtuálních prostředí a jejich následné nasazení v úlohách zaměřených na zachycení vizuálních kognitivních stylů, v rámci virtuální reality, při užití eye-trackingu. Konstrukce vlastních virtuálních prostředí sice otevírá nové možnosti, technologické limitace i specifika vizuálního vnímání s sebou nicméně nesou řadu omezení – je tedy třeba klást důraz na metodické podchycení zásad konstrukce virtuálních úloh.

Abstract

Visual cognitive styles manifest in ways humans implicitly process visual cues from their surroundings. A question arises: how are cognitive styles to be measured? Classic approaches make use of questionnaires or figural tasks. However, the nature of these approaches fails at capturing complexity of real visual compositions. With the onset of (mobile) eye-tracking devices, this limitation can be overcome; the paradigm, however, still revolves around receptive processing of existing visual material. The aim of this paper is to consider constructing 3D virtual scenes, followed by utilizing such scenes in visual cognitive styles tasks – using virtual reality and eye-tracking devices. While constructing virtual environments does indeed offer new possibilities, technological constraints and visual perception limitations do, however, introduce some new limitations to the table. Therefore, offering a methodological grasp of virtual scene construction is relevant.

1 Introduction to visual cognitive styles

As characterized by some of the pioneers in the field of cognitive differences research (Ausburn & Ausburn, 1978), cognitive style is: *a psychological dimension that represents consistencies in how an individual acquires and processes information*. A visual cognitive style is a further, finer differentiation of the former concept. However, since this metaphorical terrain of cognitive styles research has grown ever so complex over the years (Kozhevnikov, 2007), it is also worth mentioning that in this proposal, we chose to adhere to the definition of visual cognitive styles formulated by Kozhevnikov & colleagues (Kozhevnikov et al., 2005).

According to Kozehnikov's classification, there are categories of object visualizers and spatial visualizers. Their definition is as follows ("Object-Spatial-Verbal Cognitive Style Model"):

- **Object visualizers** prefer to construct vivid, concrete and detailed images of individual objects; they rely primarily on visual-object strategies; they do better on object imagery tasks.
- **Spatial visualizers** prefer to schematically represent spatial relations of objects and spatial transformations; they rely primarily on visual-spatial strategies; they do better on spatial imagery tasks.

To measure visual cognitive styles, the original authors have been using the *Object-Spatial Imagery questionnaire* (Blajenkova et al., 2006). While taking this into consideration as a baseline, in our research, we intended to propose computerized ways of measuring visual cognitive styles that make use of eye-tracking within 3D virtual environments.

We shall review potential eye-tracking solutions that can be implemented into 3D virtual environments. Some technological constraints of these solutions and the visualization devices in use will be considered. Then, we'll delve into the problematics of constructing 3D scenes salient for eye-tracking stimuli, and go over scene optimization (so as to ensure smoothness of data collection and future data analysis). In conclusion, we'll mention some other approaches to visual cognition testing that can be used to complement the eye-tracking method.

2 Methods: focus on eye-tracking

There is, without a doubt, a multitude of ways of measuring visual cognitive styles. These may include: *questionnaires, verbal statement analysis, map drawing tasks, path tracking, etc.* Some of these methods may be used as supplementaries to virtual environment-based testing, some of them may be even implemented into the virtual environment-based testing. However, for this paper, we have decided to focus on the VR eye-tracking method – since our context (VR headset + eye-tracker implanted into the headset) makes this a method that cannot exist beyond virtual reality.

By using a virtual reality headset with eye-tracking capability, extracting valid visual cognition data out of a real-time 3D virtual world is entirely possible. There are, however, different possibilities of collecting that eye-tracking data (and the choice of eye-tracking solution influences 3D scene creation, and vice versa). The two key solutions to eye-tracking that are to be considered are:

- **Heatmap tracking**[1]. As known from simple, static, two-dimensional eye-tracking applications, this approach collects the X/Y coordinates of the eye; after the data collection phase is done, a mathematical model plots the most/least frequented gaze areas onto the 2D plane, by adding a colored "heatmap" overlay on top of the plane. Whereas this approach is relatively easily applicable and interpretable in static 2D solutions, with 3D solutions, this gets more complex, as the third dimension adds the Z axis (X/Y/Z), and a non-static movement through the 3D environment introduces an ever-changing viewport. The potential level of measurement accuracy is great; however, both technical implementation and data analysis of this approach are rather complex.

- **Object tracking**. The eye tracking happens on the level of 3D objects. Any object in the scene has two states: they are either being fixated by eye-tracking, or not. This allows for simplicity in data analysis[2] and implementation; on the other hand, when gazing into details of a single object, there is no further data discrimination. In other words, the pros & cons of the object tracking approach are the direct opposite of the heatmap approach.

Due to the aforementioned complexities of 3D dynamic heatmap tracking, we opted for the object tracking approach. This means that the elements of the scene shall be broken into individual 3D objects, so that data can be collected on said objects.

3 Reviewing the technical equipment

For any visual cognition measurements to make sense, let it be known that constructing actual 3D graphics visual scenes (along with a functional data-collecting API[3]) is required. Let's not focus on the technicalities of constructing such scenes here; let's, however, cover the psychological aspects of creating, and experiencing such scenes.

Every visual scene differs from the other; every single one is a unique piece – and from this perspective, one could argue that what we proposed to do here is severely lacking and rather wild (as in considering all the possible variations/interactions of variables), in terms of scientific standards. Such notion holds true indeed – up until to the point where we grasp the ongoing phenomenon of perceiving a visual display of space and objects, and subject it to more rigorous approach. Let's do that in this narrative right now; in fact, let's start from the "lens" that gateways our perception of the visual world – i.e., the visual interface.

The hardware properties of a visualization interface (a virtual reality headset[4]) allow or limit us from using it in certain ways. In essence, this can be compared to the limitations of human sight – e.g., one cannot simply perceive far-away objects in great detail, see clearly in pitch-black darkness, etc. This consideration should go double for the engineers of the VR visualization interfaces, as they should always keep in mind that not only do their devices carry their inherent limitations,

[1] 3D heatmap tracking seems to be a niche field. We are yet to find an open source solution that is freely available and usable within the bounds of the Unity graphics engine we are using. Irregardless, we have encountered a commercial 3D heatmap solution ("EyeSee3D website").

[2] In terms of data output, this approach produces a series of gaze sequences, e.g. *ABDEBABCD*, where each individual example letter represents a sequentially fixated object of the scene. This ties to graph theory; in-depth overview into the possibilities and data interpretation of this approach is covered in the following paper: (Dolezalova, Popelka, 2016). In short, the idea is about discovering similar (sub)patterns of eye movement people incline to – from this, sort of a cluster analysis of visual cognition patterns can be extracted.

[3] API stands for *Application Program Interface*. Where does this API come from? Let there be a 3D graphics engine - Unity (we are using version 5.4). It is worth mentioning that the whole of Unity is written in the C# programming language; due to the engine's massive extendability, one who is able to program in the C# language is also able to implement a system of data collection program routines. This served as our API, a way of obtaining eye-tracking data.

[4] For a recent overview of appropriate devices, i.e. virtual reality headsets, refer to a recent overview article (Lamkin, 2017). To further explain the bold claim of these devices being "appropriate", this is so due to these devices occupying the whole of the visual *field of view* of any participant – as opposed to viewing scenes in other, less immersive visualization interfaces (such as computer screens), where objects off the screen can intervene with visual processing.

these limitations also accumulate on top the constraints of people's own, natural eyesights.

Smoothness of the experience is a necessary prerequisite of any non-negatively inhibited cognitive processes to take place in the participants. It is necessary to ensure that the perception of the participants is not, in any way, distorted in some of them (or, in consideration of the tech limitation: distorted for all the participants to the same limiting extent, and not less). This translates to technical specifications like refresh rate, or the use of high-persistence display (i.e. to experience an unblurred vision while looking around the virtual environment rendered from within the VR headset interface).

The absence of experiencing simulation sickness (Stanney et al., 2008) does play into this, too. When a man is feeling sickly (Howarth, Costello, 1997), they have a tendency of diverting their attention from cognitive/visual processing to the bodily experience created by the ill-causing stimuli and/or to preoccupying themselves with the ill-causing stimulus itself. An individual under the influence of simulation sickness can become agitated and (nearly) devoid of frustration tolerance or other compensation mechanismi; this, in turn, makes them a poor subject for visual cognition testing. The fact that some individuals are more predisposed to experiencing motion/simulation sickness than others (Basu et al., 2016) calls for experimental interfaces and stimuli being used/designed in a way that respects these highly sensitive participants.

Display acuity of the experience tells the experimenter what sorts of visual data are (not) worth planting/exploring/analyzing. Using very small objects, or objects very far away, in a visualization that does not account for the limitations of detail perceived by the human fovea or displayed by the (limited) subpixel density of a visualization device[5], renders such tiny objects insignificant.

Optical illusions, or any sorts of dissonance-inducing stimuli (such as unbelievable vistas) are most certainly not wanted either, as such subjects have the potential of swaying the participant to investing most of their attention to such unnatural occurrences.

When supplementary devices (e.g. the integrated eye-tracker) are used, it is worth mentioning that these devices, unless wholly integrated into visualization devices in an effective 1:1 variable conversion design, introduce limitations of their own. If their differentiation (of resolution, of measurement frequency) is of greater magnitude than the one of the visual interface, a man need not worry. If, however, they differentiate at a lower rate[6], they become the bottleneck of the research design and the related data output. E.g., eye-tracker resolution matters when targeting small/distant object; considering this, a question arises: what is considered an object small enough, or distant enough not to be implemented into a visual scene that tests for object (re)cognition?

[5]One thing is certain: in terms of VR headset resolutions, there is a huge room for improvement (Orland, 2013). It is one thing to look into a FullHD/ 4K/ 8K display from the distance of one's chair/desk; having the display cover the whole field of view from an up-close distance is something else.

[6]A differentiation rate that is not an exponent of the framerate of the visualization device does introduce some marginal error as well.

4 Constructing the scene

Figure 1 shows an eye-tracked scene we created for the means of our research. The rest of this chapter will be devoted to the methodical approach we adopted to produce such a scene.

Fig. 1: An example scene with eye-tracked objects in it.

To actually have something worthy of measuring in a scene, the intended scene must be purposely designed in way of it being saturated in stimuli that are relevant to the theory we are following. In case of finding ways of differentiating from *spatial visualizers* and *object visualizers*, the scene must contain measurable elements that engage both of these kinds of cognitive styles, differently, and these differences can be recorded. On top of this, knowing that visual cognition is a continuous process, predicting possible recognition patterns these two types of visualizers may undertake has the potential of solidifying our method.

For a predominately object-oriented visualizer, outfitting the scene with objects rich enough in detail is crucial. When constructing and placing such objects into the scene, let them purposefully be:

- **Unique**. It is a common shortcoming of computer-generated 3D scenes that these scenes appear rather bland. This is so due to the creators of such scenes re-using the same 3D objects over and over again. Therefore, the object of focus shall be made of an unique 3D model; on the other hand, reasonable modesty in visual appropriateness should still be considered: when a scene is constructed within the constraints of a visual theme, any object within that scene should still respect that theme. There is such a thing as an object breaking the overall impression of a theme; where the line of such over-intrusion lies, that is, however, unclear (albeit discoverable by pilot testing).

- **Breaking the pattern**. It is known that patterned visual stimuli, such as those found in nature, are not very demanding in terms of visual processing/recognition, or worthy of one's prolonged attention (Taylor et al., 2005). Therefore, an object of focus shall aim for coming up with a reasonably unique external shape (silhouette) and internal patterns. Let these be (somewhat) non-repetitive/non-conventional, too.

- **Detailed**. An object-oriented participant is inclined to gaze into the details of individual objects. To facilitate this, let the object have unique features of varying details. As it is with art, bland, visually simplistic areas shall complement detailed, somewhat unexpected parts. Such dynamics should be contained within an object.

Figure 2 shows some of the object-oriented elements placed into the scene. In red: a detailed flag. In blue: a unique object of an atypical chair. In yellow: a pattern-breaking set of bricks.

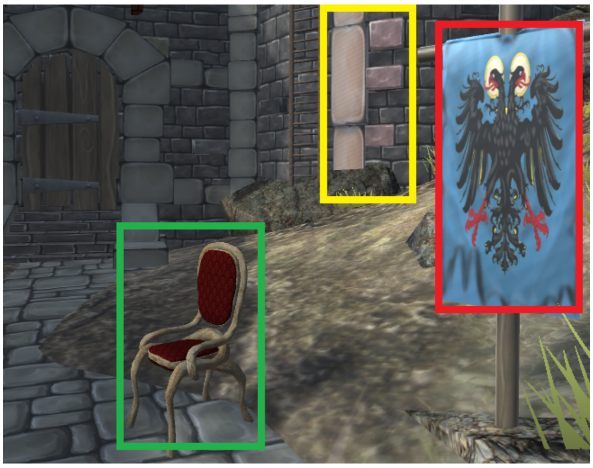

Fig. 2: Some of the object-oriented stimuli in the scene.

As for spatial visualizers, enriching the scene with relationships is the key. This globalistic/holistic individual does not spend much time on individual objects; instead, they search for rules that bound the elements of the scene into patterns. Therefore, let us consider these suggestions while making the scene stimulating for the spatial visualizer:

- **Lead lines**. In visual art, a piece of imagery conveys its message in a flow of lines (three-by-three division, two-by-two-division, one/two/three point perspective, horizon line, convergence of lines to an "action center", etc.[7]) It can be expected for the participant to follow these (guide)lines, to flow across the scene, or to proceed on to an area of focus (ideally, any scene should have more than one of these areas of focus, so as to prevent pro-object visualizer interpretation).

- **Scene boundaries**. These distinctive elements, if in place, should indicate where an intended scene ends. Let them take the form of massive and/or close-up monolithic objects situated on the fringes.

- **Similarities**. If there are multiple objects in the scene with distinctively similar features, these will form a logical group (despite potentially being intentionally displaced out of proximity). The spatial visualizer, being the one who scans across the whole scene faster than the object visualizer, is the prime candidate to notice the pattern of similarities, and to follow these groupings of objects.

- **Encompassion**. This can be achieved by placing a large object that spans across a significant area of the scene, followed by inserting supplementary objects into the body of the large one – all in a way that respects, and complements, the shape and the flow of the large object.

Figure 3 shows the spatial layout of the scene. In blue: prominent building being divided into three portions by lead lines, along with the horizon line. In violet: a surrounding valley that represents scene boundaries.

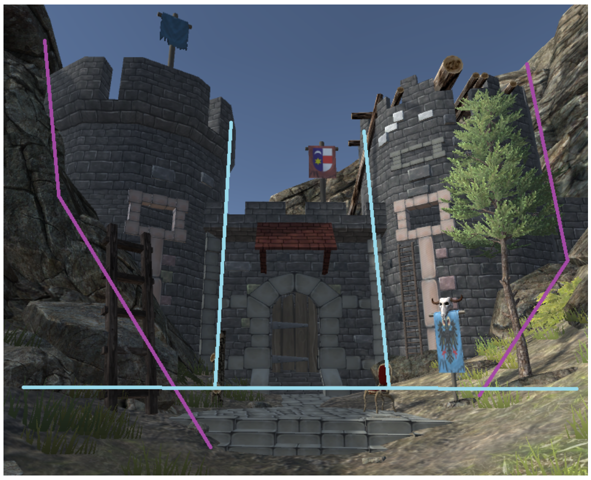

Fig. 3: The prominent spatial cues in the scene.

It is also worth considering that some differences between the two cognitive styles might not manifest outright; they may come to light only after pilot testing our scene (be it from self-reports of trial participants, or from analyzing their data).

[7] All these terms tie to art fundamentals theory. To explain these further, refer to any serious art textbook – such as this one: (Beloeil et al., 2013).

5 Optimizing the scene

The fact that we use object tracking means that if we, somewhat counter-intuitively, wish to gain extra data from an object in the scene, this object has to be broken into multiple sub-objects (while retaining its appearance of a whole object - the difference being merely functional, happening at a data-collection level).

To prevent inaccuracies and/or data noise, it is also recommended not to overlay eye-tracked 3D objects onto other 3D objects too much; especially with the added depth axis (present in the 3D image, obviously not present in eye-tracking data) it could be difficult to determine the object of focus, should edges of multiple objects frequently cross, especially along the depth axis. In fact, eliminating the depth axis to an extent is considered preferable; this can be achieved by placing the objects into the scene, followed by placing a wall, or some other monolithic barrier of constant depth, beyond these eye-tracked objects.

Here are some further considerations regarding constructing scenes for eye-tracking:

- **Long distance** object placement avoidance. Usually, some 30m may be considered maximum (depending on the overall size of the object).

- **Object size**. To ensure the objects in the scene are big enough, project the 3D scene onto a 2D plane; on this 2D plane, an object should take up at least 5% screen width and height to be considered accurately trackable and displayable within the bounds of current VR headset resolutions.

- **Proper scene lighting** should complement object recognizability.

- **Objects on the edge** of the scene will be implicitly harder to notice than objects in the middle; this goes double for objects that are away from the default field of view – if there are any (i.e. those objects only accessible by initial head movement up/down/to sides).

- **Static camera**. For increased control over the experimental design, do not allow the participants to move through the environment while viewing the eye-tracked scene. However, this solution is a trade-off that throws away movement data and possible interpretations related to it (i.e. participants actively positioning themselves in certain way to make sense of the scene).

- **The intended field of view of the scene**, and the potential need for camera rotation to view a widespread scene are to be considered. A scene may span across the default camera field of view (110 degrees, for most virtual reality headsets), or it can go further. A 360-degree scene is usually the most demanding one for the participants to make sense of, remember and recall – because more object/relations can be stored in broader scenes.

Prior to an actual, in-depth data analysis (which is not the scope of this paper), data cleaning happens. In this case, adhering to the aforementioned suggestions does help with reducing the potential for data noise. When boundaries of few to none eye-tracked objects are neighboring, a number of quickly altering fixations (caused by inaccuracies in eye-tracking) are few as well. Furthermore, by adding some tolerance to object focus boundaries and defining minimum time intervals for eye-tracking fixations (as opposed to discardable eye saccades), data clarity can be achieved.

To analyze our data, we have no specialized software package to use (a situation unlike the conventional 2D eye-tracking). For most of our data analysis, we'll have to develop a solution of our own; to achieve this, a solution based on the Processing data visualization programming language may have to be created.

6 Supplementing our approach

While experiencing the 3D visualization, the participants can verbalize their thought processes; the verbalization data can be then coded into object/spatial meaning units. Another approach is verifying their remembrance post-test, by a means of presenting the participants with multiple-choice imagery of some cut-out portions of the scene, with one variant being correct (for an example of this, see Figure 4).

Asking an open question of *"What is that you noticed in the environment?"* does not prime the participants to answering under the influence of suggestive inquiry – their visual cognitive style preference may unveil naturally this way. As for object visualization disposition, the researcher can ask the participants to describe objects in the scene that they remember, to describe a specific object, or to probe for a specific feature or existence of an object (*e.g. "What did the coat of arms look like?"*, *"Was there a wheel in the scene?"*, etc.) The spatial-disposition questioning researcher asks for relations. They can pick an object in the scene and ask for associations to other objects (e.g. *"Was there a log right to the ladder?"*, *"What was the item highest up in the scene?"*, etc.). This sort of questioning can transcend even beyond the predetermined scenes, across the whole 3D environment.

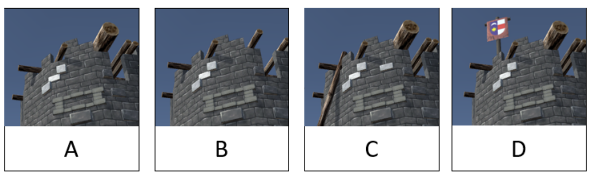

Fig. 4: Post-test questioning (the layout of a tower).

Another way of post-test questioning can be achieved by using maps. To test for spatial preference, the participants will be asked to draw/place objects into the map in accurate proportions/coordinates; as for object-oriented preferences, the participants may be presented with multiple variations of one item to place on the map (one correct variation, the rest of them incorrect).

To make relevant comparison of our findings, our participants were asked to complete the *Object-Spatial Imagery questionnaire* on the side; the intention was to compare the results of this questionnaire side-to-side with our our data.

7 Conclusion

Utilizing the existing technology and some scripting of our own, we have developed a platform that allows for collecting behavioral data in virtual environments; within the scope of such environments, we have proposed a way of creating eye-trackable 3D scenes that can be used in visual cognitive styles testing. The proposed process of creation was intended to respect the visual cognitive styles theory. Also, potential technological shortcomings were covered – so that the processes of data collection and data cleaning are not negatively affected.

We are yet to analyze data and produce results based on an eye-tracking task we have implemented. The approach we have proposed in this paper can, however, serve as a research design baseline for future studies in visual cognitive styles (or in similar subjects).

Acknowledgment

This submission was made possible due to the following project: Vliv kartografické vizualizace na úspěšnost řešení praktických a výukových prostorových úloh (MUNI/M/0846/2015).

References

Ausburn, L. J., & Ausburn, F. B. (1978). Cognitive styles: Some information and implications for instructional design. *Educational Communications & Technology Journal*, 26, pp. 337-354.

Basu, A., Ball, C., Manning. B., Johnsen, K. (2016). Effects of User Physical Fitness on Performance in Virtual Reality. *2016 IEEE Symposium on 3D User Interfaces*, pp. 233-234.

Beloeil, G., Riabovitchev, A., Castro, R. F. (2013). Art Fundamentals: Color, Light, Composition, Anatomy, Perspective, and Depth. 3DTotal Publishing.

Blajenkova, O., Kozhevnikov, M., Motes, M. A. (2006). Object-Spatial Imagery: A New Self-Report Imagery Questionnaire. *Applied Cognitive Psychology*, 20, pp. 239–263.

Dolezalova, J., Popelka, S. (2016). ScanGraph: A novel scanpath comparison method using graph cliques visualization. *Journal of Eye Movement Research*, 9 (4), pp. 1-13.

EyeSee3D website. Retrieved from http://eyesee3d.eyemovementresearch.com/

Kozhevnikov, M. (2007). Cognitive Styles in the Context of Modern Psychology: Toward an Integrated Framework of Cognitive Style. *Psychological Bulletin*, 133 (3), pp. 464–481.

Kozhevnikov, M., Kosslyn. S., Shephard, J.(2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, 33 (4), pp. 710-726.

Howarth, P. A., Costello, P. J. (1997). The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays*, 18 (2), pp. 107-116.

Lamkin, P. (2017, March 14). Best VR headsets 2017: HTC Vive, Oculus, PlayStation VR compared. Retrieved April 07, 2017, from https://www.wareable.com/vr/best-vr-headsets-2017

Object-Spatial-Verbal Cognitive Style Model. Retrieved from http://www.nmr.mgh.harvard.edu/mkozhevnlab/?tag=object-spatial-verbal

Orland, K. (2013, September 11). Virtual Perfection: Why 8K resolution per eye isn't enough for perfect VR. Retrieved April 07, 2017, from https://arstechnica.com/gaming/2013/09/virtual-perfection-why-8k-resolution-per-eye-isnt-enough-for-perfect-vr/

Stanney, K. M., Kennedy, R. S. (2008). Simulation Sickness. *Human Factors in Simulation and Training*, pp. 117-127.

Taylor, R. P., Spehar, B., Wise, J. A. et al. (2005). Perceptual and Physiological Responses to the Visual Complexity of Fractal Patterns. *Nonlinear Dynamics, Psychology, and Life Sciences*, 9 (1), pp. 89-114.