

Aspectus operis and visual attention: from Vitruvius to Virtual Reality

Damien Vurpillot

► **To cite this version:**

Damien Vurpillot. Aspectus operis and visual attention: from Vitruvius to Virtual Reality. 2016.
<hal-01302286>

HAL Id: hal-01302286

<https://hal-bnf.archives-ouvertes.fr/hal-01302286>

Submitted on 13 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Aspectus operis and visual attention : from Vitruvius to Virtual Reality

Short paper: Poster ; Theme: Theory, Methodologies, Preservation and Standards (Cognitive Sciences and Interaction)

Damien Vurpillot

Chrono-Environnement Laboratory (UMR 6249)

University of Franche-Comté

Besançon, France

damien.archeo@gmail.com

Abstract—How does one interpret remains of the past without projecting one’s cultural baggage in the process? This is a daily concern both for historians and archaeologists. In the present study, we investigate the possibility that a careful analysis of visual attention can provide invaluable hints leading us toward a better understanding of symbolic space. Through the examination of low level visual inputs we are able to explore the spatial composition of an archaeological landscape as it was originally intended to be seen and perceived. This provides a possibility for identifying socially meaningful features in a realistic virtual environment.

Keywords—Visual attention; virtual reality; archaeology; sanctuary; human behavior

I. INTRODUCTION

Vitruvius introduces the idea of *decorum*, one of the six principles of architecture, by stating that every monument has a meaning, and this meaning should arise from its appearance and location¹. Although the extent of the application of Vitruvius’s recommendations is open to debate, the idea of *aspectus operis* [1], literally the visual impact conveyed by a monument, remains demonstrably practiced throughout Antiquity. Most of the time we lack the cultural background required to process this dense web and weave of visual symbolism, yet the temptation to drift toward a culturally biased interpretation, based on our own thought patterns which may not be relevant for those ancient populations, is great. To properly study how people engage with past places it is an absolute necessity to look for new approaches which may help us to circumvent our own deeply rooted cultural mores. Studies building on research in neuro and cognitive sciences may open new ways to approach these topics.

A. Visual attention: a brief summary

Specifically, research in neurosciences in recent decades may fortuitously provide scientific framework, allowing us to build a bridge between physical and social surroundings and the interpretation of them. Unconsciously we build this kind of bridge each time we open our eyes. Humans are continually confronted with an overwhelming amount of visual information, so much that a selective process is required to operate within the limited informational capacity of our visual system [2]. This selective process is called attention, and allows us to prioritize which visual information is relevant to behavioral priorities and objectives and must be processed, while ignoring other things. The brain is actively repositioning the center of gaze on regions of interest, referred as the focuses of attention (FOA). FOA appears to be driven by two complementary mechanisms: a “bottom-up” process, which conducts rapid task-independent scans through saliency maps and a slower “top-down” process, guided by task-dependency and volition. We are still far from a global consensus on how visual attention works, but recent progress on to this topic has led to the development of comprehensive theoretical models of a mechanistic characterization of attention, linking perception and cognition [3].

B. From visual attention to *aspectus operis*

Our interest lies in the capacity of such models to predict human attention by detecting salient aspects of scenes and, by extension, to evaluate the hierarchy of meaningful elements in an archaeological environment. Computational models are driven by visual features such as luminance, colors, edges, corners, and orientation. To these we add flicker, motion, and their respective contrasts when assessing dynamic scenes [4]. The major advantage of relying on basic visual properties lies in the disconnect from the cultural background of the contemporary observer, as we are analyzing low level inputs related to a general “bottom-up” process. Returning to the idea of *aspectus operis*, the prediction of human fixations tackles the problem at its roots. Rather than the usual approaches which try to derive the meaning of an environment from descriptions of experience or physical and spatial properties themselves, we suggest that we should begin the enquiry by

¹ Vitruvius, *De Architectura*, I, 2, 5

means of a visual attention study. From a static point of view, the distribution of visual attention provides us with hints about what parts of the archaeological landscape were deliberately highlighted from a given position. Additionally, using a dynamic sequence of views allows the shifts in gaze locations to inform us about the connections between the components present in our visual field and how they work together in order to drive the decision making process [5].

II. VISUAL PERCEPTION AND ARCHAEOLOGY

The combination of neurosciences and archaeology is a recent development. Previous work focused on exceptionally well preserved and documented areas, and relied on highly detailed and precise records of spatial properties, and not on less certain 3D reconstructions [6]. Obviously well preserved remains are relatively rare in archaeological circumstances. The present project investigates the possibility of extending this methodology to less well preserved situations, relying on detailed 3D reconstructions based on excavation records and other documentation.

A. Identifying Suitable Archaeological Data

As highly anticipated as contributions of visual attention studies may be, we face a practical concern: how can we determine if a body of archaeological evidence is complete enough to be well suited for this kind of study? We have already stated that the possibility of escaping our own culturally based preconceptions is what makes these models so attractive. On the other hand, it is abundantly clear that the more 3D reconstruction is involved, the more preconceptions are introduced. This concern, while relevant, doesn't mean that we must restrict the use of visual attention studies to preserved remains. Rather, it should be acknowledged that the questions posed have to be adjusted to the data quality; one can't expect to expose laser scan data and handmade 3D reconstructions to the same level of scrutiny. We can ask very precise spatial questions of the former, i.e. "When does this carved inscription become visible while the observer is walking along a specific path, does the height of the observer have consequences, etc." The latter is restricted to more general assumptions viewed through a veil of uncertainty, i.e. "Is this structure meant to be more relevant or attractive than its surroundings? Is this true from all points of view?" In this regard, for the most schematic visual reconstructions, visual attention analysis can be used in combination with the more usual visibility analyses. At the end of the day the only way to address concerns about the quality of archaeological data is to set reasonable goals, and to establish how much uncertainty one is ready to accept.

B. Posing Archaeological Questions

We are focusing on a single variable, the visual experience, by investigating low level inputs responses in a controlled environment. The raw results are generally a saliency map for each frame of a video, based on a weighted sum of the saliency

values extracted from each scene (e.g. luminance, colors, edges, corners, orientation maps). Based on these we can compute expected visual fixations using various algorithms [7-9]. As an isolated piece of information, these raw results don't mean much - they only begin to make sense through comparative studies. We can ask how the distribution of visual attention evolves when the physical environment or social activities performed within it are altered. For example, what is the effect of changing the location of the observer, the ambient luminosity, or adding color or features? This allows us to ask, in turn, how attention might be intentionally manipulated, and to begin a more structured exploration of experience in the past.

C. *The sanctuary of Hercules at Deneuvre (Meurthe-et-Moselle, France)*

For this project a Gallo-Roman sanctuary, the sanctuary of Hercules at Deneuvre (Meurthe-et-Moselle), located in eastern France, was reconstructed in 3D. This sanctuary is suitable as we have reliable and extensive archaeological data [10], and the area is thought to have a high visual symbolic value, illustrated by terrace systems and *stelae* alignments. The sanctuary is best-known in its fourth-century AD phase, which was reconstructed. Numerous monuments were still in place at this phase, while others, which had been torn down, could be traced back to their original locations. The digital elevation model (DEM) of the area was reproduced from excavation data, and then superimposed on a modern 25m DEM. Due to challenging conditions, only a few *stelae* and monuments were documented through photogrammetry. Remaining monuments were then carefully modeled by hand from excavation data.

III. EXPERIMENTAL METHODOLOGY

A. Participants

21 participants (13 males, 8 females) were recruited for an experiment whose aim is measuring visual attention. Ages ranged from 18 to 49. They were also divided between people with a limited or no knowledge of archaeology and/or history of ancient religions (16), and people familiar with those themes (5). All provided informed consent, and had normal or corrected-to-normal vision.

B. Setup

Six interactive scenarios were produced with the Unreal Engine 4, in which aspects of a core environment were varied. The first two scenarios act as a baseline, as they only include essential features of the sanctuary. Most of vegetation and remains of offerings were absent, and we expect that this first visual encounter would focus on the monuments, as the only objects in the scene. One scenario features the rising sun shedding its light from the east (A1), while the other one is set in the middle of the afternoon (A2). The next two scenarios test the potential impact of vegetation (trees, high grass and flowers) and offerings, while keeping the time of day

variations (B1 and B2). Finally the last two scenarios add a new prominent feature. The morning light unveils colorful *stela* (C1), while the afternoon sun shines through a hypothetical *lucus* – a sacred wood (C2). Each scenario includes a random starting point.

Participants were placed in the virtual world by means of a head mounted device (HMD), in this case a Oculus Rift DK2 with 960 x 1080@75 Hz resolution per eye, thus allowing a complete visual immersion as shown in Fig. 1. A PS4 controller was used to control movement and interaction. Every scenario was recorded through OpenBroadcaster² with a 1280 x 720 resolution at 30 frames per second, and processed through Unwarpv³ in order to reverse the distortion and chromatic aberration introduced by the Oculus Rift software to compensate for lens distortion. The capture was then converted another time to a 1280 x 720 video at 5 frames per second, because, due to visual acuity limitations, our eyes can shift (saccade) up to five times every second as the light from our FOA is projected onto the fovea [11]. Each frame is then processed by the Graph-Based Visual Saliency (GBVS) model [12] in order to obtain detailed saliency maps like the one depicted in Fig. 2. Finally it is processed another time with a Matlab model, “Visual scanpaths via Constrained Levy Exploration of a saliency landscape” [8, 9] to pinpoint the most relevant FOA. In the end, we obtain a 1280 x 720 video at 5 frames per second, including both a saliency map and the main FOA in each frame.

C. Procedure

Participants were informed that they would have to interact with three game scenarios, and that the level of interaction would progressively increase. They had no prior knowledge of the sanctuary, but were allowed to ask basic questions while in the game. It was decided that participants would be divided in two groups. The first group was introduced consecutively to scenarios A1-B2-C1, while the other group played through scenarios A2-B1-C2. The first scenario, either A1 or A2, was the Exploration Phase. Participants were asked to get familiar with their environment by walking freely around the game for a few minutes. They were asked to stop the experiment when they felt accustomed to the area. They then took a short break outside the Oculus Rift, while what they should do in the Navigation Phase scenario was explained. A map with two specific locations that they should pass through in no particular order was provided. At the end of the second scenario participants took another short break, while instructed to go through the same two locations as in the last scenario, in a Task driven Phase. In the final phase they would have to make an offering at designated locations by means of a dedicated in game interface. Additionally participants filled out a form asking them to list at least five objects they saw in the scenarios, and to indicate the most important one on the list. Half of them were asked to do so after the first scenario, and half after the third scenario.

The advantages of technologies like the Oculus Rift for research on perception are significant, but there are procedural challenges. Foremost is the discomfort felt by those immersed in a VR environment. In this experiment most people were able to play through the six short scenarios with little to no discomfort reported, as immersed interactions were kept short and followed by breaks and the game was scrupulously optimized. The use of a standing or a sitting position is likewise a choice. While the most powerful sense of presence is achieved by standing up, it’s also the most disconcerting experience. Most people required a short period of adaptation in a transitional scene prior to each experimental scenario, and so we must choose between minimizing immersed time and allowing users to adjust to the VR experience.

IV. DISCUSSION

The experiment was designed to address a series of questions. First, how does the addition and variation of environmental factors impact the distribution of visual attention? To address this we varied luminosity (time of day), prominent features (vegetation and the remains of offerings) and color (painted *stela*). Each of these elements relates to archaeological topics: What is impact of performing religious activities at dawn? How do built and natural environments jointly communicate implicit meanings to visitors? Does the well-known but understudied use of color, both in buildings and monuments, act an integral part of this encoded and nonverbal message?

Second, how do computed attentional features compare with what participants named when asked to describe what they saw [13]? By using measurements of visual attention we can begin to study visual entities that might have been important but for which no appropriate word exists, allowing us to think outside the constraints of our own vocabulary.

Third, does prior “experience” of the sanctuary affect our visual attention? Playing either an exploration scenario or a task driven scenario seems to have a tremendous effect on how people behave in the virtual landscape. While the former could be conceived of as a contemplative state, the latter seems to be driven by the desire to efficiently undertake the task at hand. This difference is reflected especially in the amount of time spent playing out the scenario. Archaeologists have long been interested in discrepant experience and this approach is a real opportunity to scrutinize the differences between people with different backgrounds and motivations.

Finally, can the different paths taken inside the sanctuary by participants be used to link visual features and human movement decision-making? Influenced by processual approaches in archaeology, many models have been developed with the goal of explaining how people are moving through space [14] and visual perception appears to be a relevant addition to further explore the “visual experience”. The data collected is currently under study, and this will certainly lead to new and more subtle questions, and further elaboration of these preliminary interpretations.

² <https://obsproject.com/>

³ <https://github.com/eVRydayVR/ffmpeg-unwarpv>

ACKNOWLEDGMENT

The support of MSHE C. N. Ledoux, University of Franche-Comté and University of Arkansas is gratefully acknowledged.

I would also like to thank L. Nuninger (Chrono-Environnement Laboratory), R. Opitz (Center for Advanced Spatial Technologies) and David Fredrick (Tesseract Interactive Humanities & Game Design) for their help and feedback.

REFERENCES

- [1] M. Courrént, “*Tenuitas cum bona fama* : éthique et architecture dans le *De architectura* de Vitruve”, Cahiers des études anciennes, vol. XLVIII, pp. 219-263, 2011
- [2] M. Carrasco, “Visual attention: The past 25 years”, Vision Research, 51(13), pp. 1484-1525, July 2011
- [3] P. Pärnamets, P. Johansson, L. Hall, C. Balkenius, M. J. Spivey, and D. C. Richardson, “Biasing moral decisions by exploiting the dynamics of eye gaze”, Proc. Natl. Acad. Sci. USA, vol. 112, n.13, pp. 4170-4175, March 31 2015
- [4] P. K. Mital, T. J. Smith, R. L. Hill, and J. M. Henderson, “Clustering of gaze during dynamic scene viewing is predicted by motion”, Cognitive Computation, vol. 3, Issue 1, pp. 5–24, March 2011
- [5] R. B. Towal, M. Mormann, and C. Koch, “Simultaneous modeling of visual saliency and value computation improves predictions of economic choice”, Proc. Natl. Acad. Sci. USA, vol. 110, n. 40, pp. 3858-3867, October 1 2013
- [6] R. Opitz, “Exploring Digital Landscapes at the Human Scale”, Digital Domains : remote sensing of past human landscapes (20-22 Mars 2014 Dartmouth College), unpublished.
- [7] L. Itti and C. Koch, “Computational modelling of visual attention”, Nat. Rev. Neurosci., Vol. 2, pp. 194-203, March 2001
- [8] G. Boccignone and M. Ferraro, “Modelling gaze shift as a constrained random walk”, Physica A: Statistical Mechanics and its Applications, Vol. 331, Issues 1–2, pp. 207-218, January 2004.
- [9] G. Boccignone and M. Ferraro, “Modelling eye-movement control via a constrained search approach”, Visual Information Processing (EUVIP), pp. 235-240, July 2011
- [10] G. Moitrioux, *Hercules Salutaris* : Hercule au sanctuaire de Deneuvre, Meurthe-et-Moselle, Nancy, Presses universitaires de Nancy, 1992
- [11] J. M. Findlay and I. D. Gilchrist, Active vision: the psychology of looking and seeing, Oxford: University Press, 2003
- [12] J. Harel, C. Koch, and P. Perona, “Graph-Based Visual Saliency”, Proceedings of Neural Information Processing Systems (NIPS), pp. 545-552, 2006
- [13] A. D. F. Clarke, M. I. Coco, F. Keller, “The Impact of Attentional, Linguistic and Visual Features during Object Naming”, Frontiers in Psychology, vol. 4, 12 p., Decembre 13 2013
- [14] K. D. Fisher, "Investigating monumental social space in Late Bronze Age Cyprus: an integrative approach", in: E. Paliou, U. Lieberwirth and S. Polla (Eds.), Spatial analysis and social spaces. Interdisciplinary approaches to the interpretation of prehistoric and historic built environments, Berlin, Boston: De Gruyter, 2014, pp. 167-202



Fig. 1. One of the participant experiencing scenario A2 with the Oculus Rift, while his actions are recorded and monitored on the instructor screen.

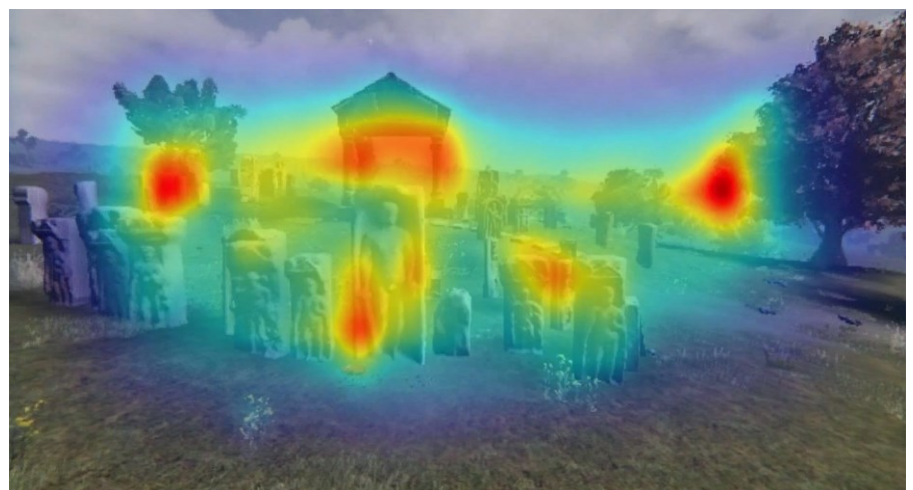


Fig. 2. A saliency map overlaid as a heatmap over one of the scenario capture.