VISUAL ELEMENTS IN VIRTUAL ENVIRONMENTS: FINDING PARALLELS BETWEEN CGI VISUALIZATIONS AND HUMAN PERCEPTION

Pavel Ugwitz, Zdeněk Stachoň

Mgr. Bc. Pavel Ugwitz
Department of Geography, Faculty of Science, Masaryk University
Kotlářská 2, 61137 Brno
+420 737 022 264, 172577@mail.muni.cz

Mgr. Bc. Zdeněk Stachoň, Ph.D.
Department of Geography, Faculty of Science, Masaryk University
Kotlářská 2, 61137 Brno
+420 549 494 925, zstachon@geogr.muni.cz

Abstract
As outlined by various comparative studies, there's an ongoing question of suitability of certain (cartographic) visualizations & visualization elements for tasks at hand - be it specialized cases, unique visualization interfaces, etc. However it may be, in the end, a recipient always processes such "suitability" through human visual perception and cognition.
In this paper, our approach was to take a step back from the comparative research, and to offer a theoretical overview of visual cognition and graphics pipelines instead. I.e., we put together a working model for each, followed by our comparison of the two. Specifically: we revisited notable cognitive psychology theories on processing visual stimuli, and from these, we extracted a hierarchical working model; as for computer graphics, we reviewed existing 3D graphics pipelines, along with contemporary visualization engine implementations.
The purpose of this theoretical detour was to provide grounding for research that is to follow. Not only do our findings unveil potential experimental cases, they also prove an insight on what is feasible in the context of today's visualization engines and hardware.

Keywords: *Virtual Environments, Computer-Generated Imagery (CGI), Graphics Pipelines, Visualization Engines, Visual Perception, Visual Cognition, Visual Elements*

INTRODUCTION

When visualizing particularly information-dense spatial subjects, esp. in three dimensions, the potential for visual complexity in processing such subjects ensues [1]. And as we breach such problematics, many inter-related questions arise – e.g., the nature of the subject that is to be visualized, the method-of-choice of visualizing said subject, the expertise of the individuals that are to be working with the subject [2], the (steering of the) direction of the cognitive processes in said individuals, etc. One way of handling some of these questions is by knowing and controlling the visual elements a spatial subject consists of.

The focus of this paper involves asking the following question: *within the context of visual elements research (esp. in 3D), how do the theories/methodics/terms of human perception/cognition of imagery and space compare against artificial, computer-made approaches to visualization?* The specificity of chosen visualization method-to-be is irrelevant to us in this context, as we intend on grasping such question holistically; our intention here lies in offering a new perspective on the subject and/or honing the methodics regarding the scope of visual element research when artificial representations of imagery and space are in play.

In other words, to make sense of any visual scene, a scene can be divided into atomic visual elements that make up such a scene; elements, that, if removed from the scene, would cause the existing scene falling apart irretrievably. There are processes, both in human perception and in the field of computer graphics, which utilize visual elements to (re)create the whole of a scene. And since a human perceiving a visual scene and a computer rendering a representation of space are two separate subjects based on different paradigms, we expect there may be a gap when comparing the two – i.e., something worth the scrutiny of being explored further (as in attempting to bridge an inter-relation of the existing terminology).

In yet another words, in an ideal world, our perceptions and/or representations of an outer world would provide 100% accuracy with zero distortion. However, when a real person perceives a visuo-spatial subject, in their perception, they introduce distortions [3]. When existing space is represented as computer-generated imagery, i.e. a virtual environment, distortions occur, too (rasterization, resolution constraints, detail limitation). But we, as researchers, tend to fail to consider situations when these two distortions potent each other – i.e., when a person perceives computer-generated representation of space (see fig. 1 – there can surely be a circular loop from the last icon to the first).

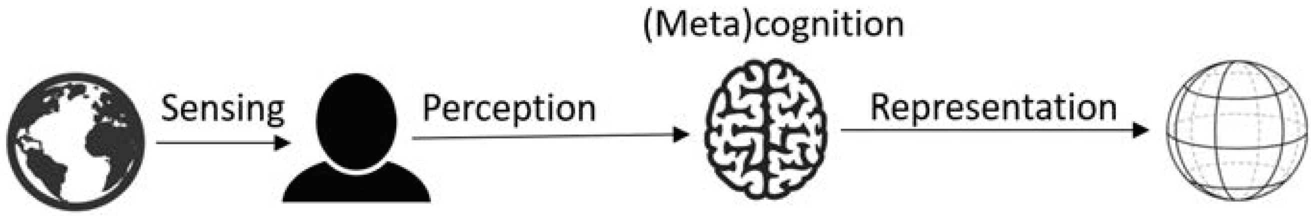

Figure 1: A human senses the surrounding visual stimuli; through perception, they can make sense of the world they see. Then, through visual cognitive process, they can choose their further strategies in relating to the world and space. Based on their understanding of what they see, they can then proceed with creating their own representations of the world.

Herein lies our criticism of visual elements research done to date – usually, for the ease of evaluation of such research (and for us to drive our point across), there would be an experimental group and a control group; the participants in both groups would be presented with a similar visualization of a virtual environment, with one distinction in mind – the participant in the experimental group being presented with an altered value of a visual element the study follows (i.e., the independent variable – be it a difference in object detail, texture, color, etc.). Beyond that, there would be little to no consideration for the computer visualization engine, i.e. no objective critical consideration of whether the operationalization of the independent variable is proper. Either that, or the visual element in question can be poorly grasped – oftentimes not being an actual visual element, but a complex combination of multiple elements – e.g. something that is hard to operationalize/standardize.

VISUAL ELEMENTS IN HUMAN PERCEPTION

In this chapter and the following one, we are about to review visual elements in human perception and in computer graphics. At this point, it has to be stressed that the human perception portion has to precede any other topic – as this is the basis on which we see external world around us, i.e. a basis from which other concepts follow. Therefore, any potential perceptual limitations/errors that can occur, do occur here or later on.

Unfortunately, human perception does not offer a solid, fully deterministic foundation on which the subject can be expanded upon further. This is so because of human perception originates from biology (esp. neuroscience), psychology, and, to an extent, from cultural learning [4]. As the starting point to visual perception, there is sensing: the visual pathway receives light photons through the eye, and the responsible photon-receptive cells proceed in translating this light into electric potential that then travels to the occipital cortex of the brain; physics (optics) comes to play at this level of visual sensing, and eye pathologies, if present, do translate to vision distortions. Once transported into the visual cortex of the occipital lobe, image reconstruction commences. The occipital cortex is interconnected with the temporal cortex through the ventral stream – which allows for object recognition (and putting things in context, as far as our knowledge lexical learning, value system and cultural appropriation of objects goes); the other connection, dorsal stream, which interconnects with the parietal lobe allows for perceiving the relations between objects, e.g. spatial cues [5]. As it seems, the mammal brain has developed some specialized functionalities in these areas, e.g. depth and distance estimation, face recognition, recognition or pursuit of a rotated object, etc. Perceptual errors on this level vary from object misrecognition to brain lesions that cause perceptual blindness, the inability to refer to an object by name, to the structure of the scene partially or completely falling apart.

Regarding visual elements, let’s focus on the visual cortex itself [6] – as the ongoing research in neurosciences suggests, some proto-representations of visual elements are, in fact, present in the six segments of this cortex. The primary visual cortex (V1) projects imagery obtained from retina onto its cells, in an uneven manner, due to uneven retinal cell distribution (cortical magnification). V1 is also primarily sensitive to edge detection – when two neighboring shapes of different colors are processed, it is not the color itself that is of processing significance, but the radical shift in value upon transition – creating the perception of the edge. Shape recognition does not happen on the level of V1; however, due to edge detection and projection of image, this does give an accurate baseline to recognize shapes in following perception processes. The secondary visual cortex (V2) allow for some rudimentary recognition of orientation, color, size, shape, or spatial frequency, along with integrating some of the more complex processes (e.g. spatial cues or perceiving along the Gestalt principles). The function of V3 is yet somewhat unknown; it is suggested that complete visual representation materializes here. V4 extends the abilities of V2, in addition to processing spatial saliency, i.e. some shape and color recognition, along with the ability to focus attention on a stimulus of choice. The last two segments of the visual cortex, V5 and V6, are not that significant for our subject – the former supposedly specializes in recognizing objects of motion, whilst the latter aids in continuous recognition while the individual themselves is in motion.

Another notable regard is that beyond the reconstruction of standard visual stimulus, human visual processing also contains some specialized functions (stemming from the concept of modular brain theory [7]) – these are kinds of superstructures built on top of the general visual perception; these specialized functions range from low-level, somewhat generic ones (e.g. spatial cues recognition, perceptional sensitivity to Gestalt principles), to high-level, specific ones (e.g. face recognition); their use is to allow humans/primates to orient themselves in the environments they live in more rapidly/effectively. For out purposes, let's mention the Gestalt principles and spatial cues – particularly where they fit within the context of other visual processing. When visual input is understood enough so that proto-representations of shapes stand out (we are talking about what would be two-dimensional portions of a two-dimensional imagery), Gestalt principles [8] can come into play. These principles (e.g. proximity, continuity, good shape, illusory contours, figure/background relation etc.), allow us to put shapes into groupings and/or to assume shape continuity in places where this continuity is otherwise obstructed or lost due to the limitations of the current viewpoint – effectively giving us some heuristic interpretation of what would otherwise be confusing visual layout, while also grouping/sidelining the similarities. Beyond the Gestalt principles, once forms of three-dimensional objects are brought into the perception process, spatial cues [9] can be utilized. These cues, some of which are monocular (e.g. relative sizes of objects, object overlaying, perspective, texture gradient, atmospheric perspective, light distribution, elevation), some of which are binocular (eye convergence/divergence, binocular disparity of image perceived in the left/right eye), allow us to determine the three-dimensionality of space and how navigate in it (where the retinal/cortical projection of visuals is merely 2D, these cues allow us to reconstruct the 3D perception of the surrounding world around us).

At this point, let's take a detour into art theory. Why would we want to do that? One reason is that art is a product of visual cognition – i.e. a product of thinking about the visual composition of imagery and spatial layouts; a human being producing art would, therefore, reflect on their own understanding of the physical arrangement and properties of the visual world prior to re-producing such information into a man-made visualization - this, by itself, is a level of metacognition of space; yet another level would be us thinking about how an artist produced their visualization (i.e. from thinking of space, to thinking of how to represent existing space onto a medium, to thinking how another had thought about representing space).

The other reason for mentioning art theory is because the term visual elements actually originates from here. This is so since art precedes scientific theory [10]. Unfortunately, this also introduces further inaccuracies into our effort of thoroughly defining visual elements – the art theory backing the term is merely based in practice and tradition of analyzing and representing man/made visualizations projected onto two/dimensional planes, i.e. mostly paintings. Based on that alone, we can see that if we are to engage in 3D virtual environment research (with the likes of interactive, real-time virtual reality visualizations in mind), the basic *visual elements in art* terminology is expected to be lacking.

The seven visual elements in art, as recurrently mentioned [11] are as follows: line, shape, form, value, color, texture, pattern. These are not independent on each other, as some amount of dependency/hierarchical ordering exists. E.g., a line can be understood as a brush stroke that connect two points (which may or may not be present within the scope of the visualization). A shape can only emerge from a combination of lines (a triangle, the most primitive of shapes, is made of three lines); a form is a three-dimensional extrusion of a shape (and again, more lines are required – to bring out the shape into its third dimension using lead lines of a 3D projection of choice). Value is subset of color (hue, value, intensity), and color is, once more, a subset of texture; what all three have in common is that they fill out some space which may or may not be pre-defined by lines/shapes/forms. Pattern stands by itself as an intended ordering and/or repetition of the other elements, so as to create an art style. All these well-known art styles (e.g. realism, surrealism, expressionism, cubism, etc.) can be thought of as purposeful utilization of certain (clusters of) visual elements, while also potentially suppressing others.

Nevertheless, the art theory of visual elements does not prescribe inter-relations and/or dependability in between its elements; we can, however, extrapolate that knowledge from some general metacognition of space of our own, or even better: from drawing parallels with the aforementioned processes in human visual perception. Onto visualizing all this:

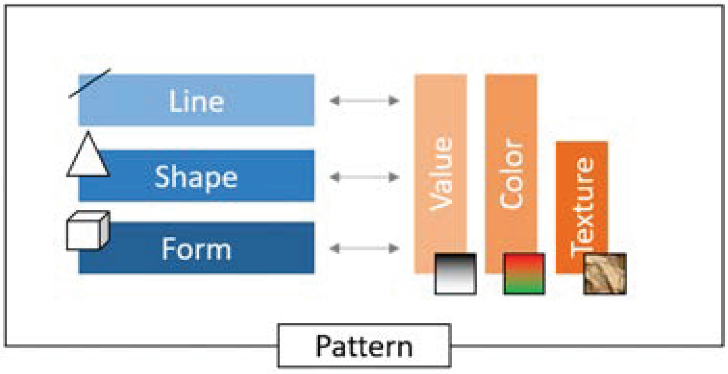

Figure 2.1: Our hierarchical ordering of the visual elements in art

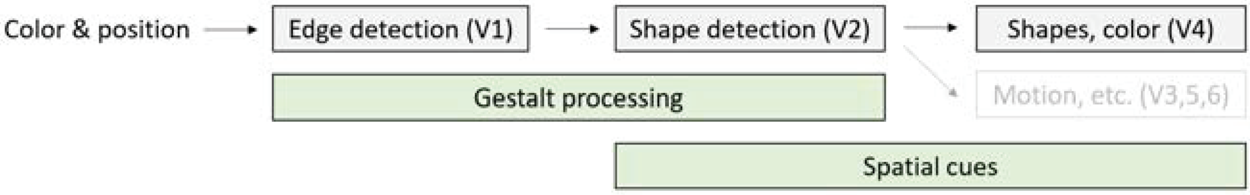

Figure 2.2: human visual perception, as per its stages/ visual cortex interconnection, and functionality

As seen in figure 2.1, this is our hierarchical understanding of visual elements: there is a transition from 1D to 2D to 3D in lines/shapes/forms; all of which can be represented by values/colors; a texture is notably two-dimensional, therefore it requires 2D to be portrayed; all of this can be varied/stylized by patters of choice. Moving on to human visual perception (figure 2.2): due to relative simplicity of both this model and the aforementioned one, the parallels between the two can be easily spotted; a notable difference is that within the context of human perception, there was no mention of textures (however, a texture is, in essence, nothing else than a variation of color across space) or patterns (we could, however, relate Gestalt processing to that).

VISUAL ELEMENTS IN COMPUTER GRAPHICS

As opposed to visual perception (based on evolution) and art theory (based on practice), the approaches in computer graphics are based on mathematical approximations of reality and the use of data structures to store their values (with a bit of practicality in tow as well) [12]. The problem of applied mathematical representations is that they are inherently inaccurate; if they are to be made somewhat accurate, this would take both time to implement (which increases exponentially as we go into further detailing), and time to render (i.e. computational difficulty, which can be concerning especially in real-time visualizations).

Despite the constraints, there is a multitude of ways to represent visuals in CGI. The approach of modeling solids [13] utilizes stacking 3D objects into Boolean relations with each other (e.g. additions, subtractions, intersections); some implementations make use of splines (Bezier, NURBS) [14]; however, the most common approach in real-time rendering employs the use polygonal meshes (typically lots of triangles) [15]. Within the polygonal mesh approach, we can, also, explore the subject of visual elements of such visualizations; furthermore, unlike human perception and/or art, polygonal visual representations are constructed from a pre-defined data structure - in other words, full determinism is ensured.

The basic element of any polygonal structure is a three-dimensional point – a vertex. Multiple vertices are then connected to form a polygon, i.e. a shape (a triangle or a more complex shape). When a set of polygons is put together, an object can be created. Edges between vertices exist, but usually not on their own – they have to be linked to a polygon first (these rules may not hold 100% true for every visualization engine since implementations differ, yet this is the most common practice). From this classification, it is apparent that vertices/edges/polygons are fairly similar to the visual elements of lines/shapes.

Beyond basic geometry, however, things do get more complicated. Where the traditional classification refer to a simple visual element of texture, current-generation 3D visualization engines operate with complex models based on light dispersion/reflection/illumination. Therefore, instead of a simple texture that would represent the color & structural gradient of the material, there is now a multitude of supplementary textures that aid in simulating complex lighting models. Especially when the 3D renderer implements a computational model of global illumination, overall coloration of objects is influenced by the surrounding virtual environment. Some of these extra so called "textures" (we use that term lightly here, as they deviate from the original meaning of the word quite significantly) add in additional geometric detail to 3D forms on demand – while the so called normal map creates an illusion of this across a shape by altering light reflection projected onto a shape's texture, the technique of tessellation actually adds in real detail geometry by dividing the existing polygons and applying deformation to them.

How exactly is the process of rendering visuals handled in 3D visualization engines? There are a couple low-level graphics programming languages (DirectX, OpenGL, Vulkan) that provide access to the functionality visualization hardware is capable of; actual visualization engines (e.g. Unity) then build their functionality on top of these. Essentially, what any current-generation graphics pipeline does can be divided into two simplified steps: re-creation of geometry (the aforementioned vertices/polygons), and application of color values on top of these. However, things do get complicated with the extendibility/programmability of these steps: various (programmer-made) vertex shaders can significantly modify the input geometry; fragment shaders then serve the purpose of applying complex color/light/texture functions to colorize the existing polygons – so that a finished, colorized geometry can be outputted as a finalized visualization. And there are complex mathematical functions beyond all this.

E.g. the approach of physically based rendering (PBR) [16] is a collection of good fragment shaders practices that can be implemented using current-generation graphics hardware to output a scene of high-quality lighting & texturing. This is achieved by creating 3D objects with PBR materials applied to it. Conventionally, we were previously referring to a shape/form having a texture; this time, there is a multitude of these [17]: diffuse/albedo (to represent the non-directional light the object reflects), specular (reflexivity of directional light), smoothness (the amount of detailing on various parts of the object), occlusion (exposition of parts of the object to the surrounding environment), and others. These are not textures per se – they represent different physical features in regards to light reflectivity/dispersion.

Figure 3.1: an albedo, specular and smoothness texture of a shield.

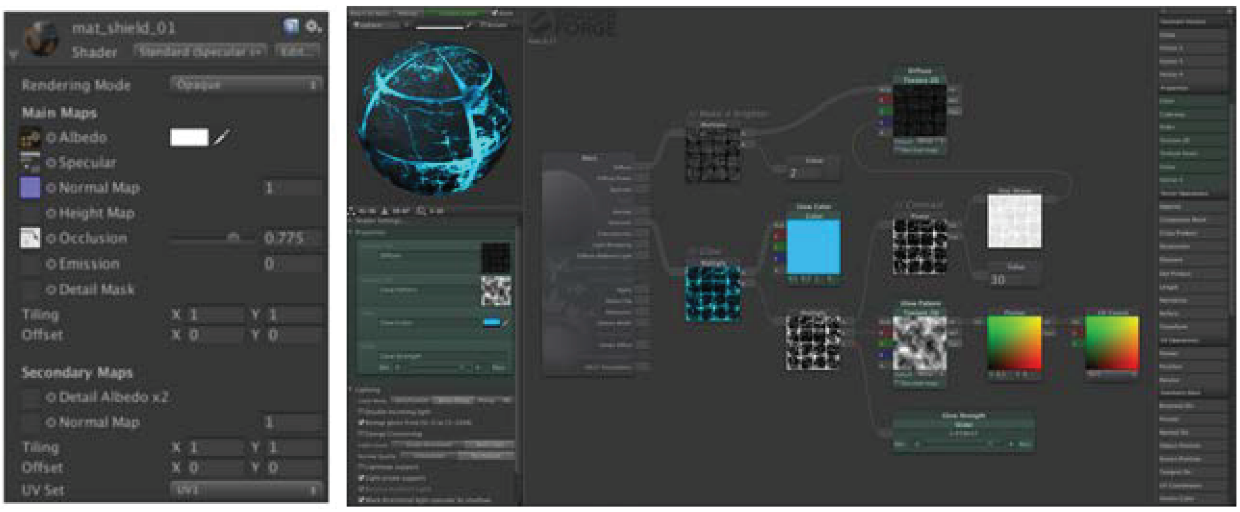

Figure 3.2: A standard PBR shader (left) vs. a custom programmable fragment shader (right).

In this regard, the terminological exodus is apparent. In real-time computer-generated visualizations, we no longer make do with a term like "texture". Furthermore, what we described was a standard PBR shader (fig. 3.1, fig. 3.2 on the left). There are actual shader editors (fig. 3.2) [18] which allow for even more complex, custom materials that behave acc. to various prescriptions of mathematical functions.

CONCLUSION

We have shown that some of the aforementioned visual elements are, in essence, easily transferrable and/or interchangeable across various theories – this applies to basic geometry (lines/edges, shapes/polygons, forms/objects). The paper also explored inter-relations between the elements – how they potentially stack on top of each other; this should, in theory, allow us to be able to manipulate them correctly, to an extent, if an experimental scenario ensues. However, we have also shown that in regards to textures and supplemental values/lighting models, the former classifications are no longer enough.

A matter for discussion would revolve around ways to bridge this terminological gap. There are, however, some rather complex prerequisites that needs to be mapped-out prior to such discussion taking place: knowing what current 3D visualization engines can do, fragment-shader-wise, and knowing the range of possibilities with such tools – since there are some optimal and some not-so-optimal practices that can be employed when producing a visualization of a 3D scene. E.g., one can, if they intend to, encompass lines, shapes and lighting values within a texture – and depending on possible experimental design that would take place based on a scene produced like that, the question is whether this would be suitable (also depending on the method of data-collection/evaluation, etc.) Regarding all this, further research is required.

Let us also mention that the available forms of visualizations have expanded drastically over the past few decades. Due to newly gained computational capacity of graphics processors (experimentally from the early 1970s, universally from the mid-1990s), we now have real-time dynamic visualizations available [19]. The dynamicity of it dictates for a need to re-calculate viewpoints and lighting on the go. Therefore, the effective approach would involve having some somewhat-universal model that can be applied across various scenes/alignments (thus the PBR, etc.) Painters of the old did not have such problems or a need for an approach like that – they simply chose a static viewpoint of theirs, along with a pre-set lighting, and then they portrayed a scene of their choice as it appeared to them – without the need for (excessive) knowledge about material properties and/or light propagation.

REFERENCES

[1] Oliva, A., Mack, M. L., Shrestha, M., Peeper, A. (2004). Identifying the Perceptual Dimensions of Visual Complexity of Scenes. In: Proceeding of the 26th Annual Meeting of the Cognitive Science Society, Chicago.

[2] Albert G., Ilyés V., Szigeti C., Kis D., Várkonyi D. (2017) How Hard Is It to Design Maps for Beginners, Intermediates and Experts?. In: Peterson M. (eds) Advances in Cartography and GIScience. ICACI 2017. Lecture Notes in Geoinformation and Cartography. Springer, Cham

[3] Crane, T., French, C. (2015). The Problem of Perception. Available online: https://plato.stanford.edu/entries/perception-problem/

[4] Goldstein, E. B. (2013). Sensation and perception, 9th ed. Wadsworth Cengage Learning, Belmont.

[5] Goodale, M. A., Milner, A. D. (1992). Separate visual pathways for perception and action. Trends Neurosci. 15 (1), pp 20–25.

[6] Hubel, D. (1995). Eye, Brain, and Vision. Available online: http://hubel.med.harvard.edu/book/bcontex.htm

[7] Fodor, J. A. (1983). Modularity of Mind: An Essay on Faculty Psychology. MIT Press, Cambridge.

[8] Todorovic, D. (2008). Gestalt Principles. Scholarpedia 3 (12). Available online: http://www.scholarpedia.org/article/Gestalt_principles

[9] Howard, I. P., Rogers, B. J. (2012). Perceiving in Depth. Oxford University Press, New York.

[10] Lewis-Williams, D. (2002). The Mind in the Cave: Consciousness and the Origins of Art. Thames and Hudson, London.

[11] Available online: http://artyfactory.com/art_appreciation/visual-elements/visual-elements.html

[12] Foley, D. J., van Dam, A., Feiner, S. K., Hughes, J. F. (1996). Computer Graphics. Principles and practive. 2nd ed. Addison-Wesley, New York.

[13] As aforementioned, pp. 533-562.

[14] Becker, M., Golay, P. (1999). Rhino NURBS 3D Modeling. New Riders, San Francisco.

[15] Vaughan, W. (2011). Digitla Modeling. New Riders, San Francisco.

[16] Pharr, M., Jakob. W., Humpreys, G. (2016). Physically Based Rendering: From Theory to Implementation. Morgan Kaufmann, Burlington.

[17] Körner, E. (2015). Working with Physically-Based Shading: a Practical Approach. Available online, as a part of Unity engine documentation: https://blogs.unity3d.com/2015/02/18/working-with-physically-based-shading-a-practical-approach/

[18] Shader Forge documentation. Available online: http://acegikmo.com/shaderforge/nodes/

[19] C. McClanahan. (2010). History and Evolution of GPU Architecture: A Paper Survey. Available online: http://mcclanahoochie.com/blog/wpcontent/uploads/2011/03/gpu-hist-paper.pdf