Gaze Direction in 3D Virtual Scenes

by

Niall Groarke, HBSc(Mathematical Science / Computing)(Honours)

Dissertation

Presented to the

University of Dublin, Trinity College

in fulfillment

of the requirements

for the Degree of

Master of Science in Computer Science (Interactive Entertainment Technology)

University of Dublin, Trinity College

September 2014

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

Niall Groarke

September 2, 2014

Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

Niall Groarke

September 2, 2014

Acknowledgments

I would like to thank my supervisor John Dingliana for all of his assistance, advice and encouragement throughout the project.

I would also like to thank my Mum, Dad, and sister Una, for their support, interest, and encouragement.

NIALL GROARKE

University of Dublin, Trinity College
September 2014

Gaze Direction in 3D Virtual Scenes

Niall Groarke
University of Dublin, Trinity College, 2014

Supervisor: John Dingliana

Many modern videogames present photo-realistic guided experiences for players. Designers face an increasing challenge in coming up with new ways to direct players through carefully crafted level designs, lighting, etc. . This dissertation proposes an image-based form of gaze direction designed to guide users through photo-realistic, highly-detailed 3D environments in a subtle, sub-conscious manner. Evaluation of this subtle visual guidance method is performed via saliency mappings and live user experiments. Implementation of the portable, user-friendly Eyetribe eye-tracker in the application explores the potential of the proposed 3D subtle guidance method with regards to commercially viable eye tracker solutions.

Contents

Acknowledgments **iv**

Abstract **v**

List of Figures **viii**

Chapter 1 Introduction **1**

1.1 Motivation 1

1.2 Methodology 6

1.3 Contribution 7

1.4 Summary of Chapters 7

Chapter 2 Background **9**

2.1 Saliency-Based Visual Scene Analysis 9

2.2 Subtle Gaze Direction 11

2.3 Other attention guidance 16

2.4 NPR attention guidance 21

2.5 Other papers 22

Chapter 3 Design and Implementation **23**

3.1 Attention Guidance in OpenGL 23

3.1.1 Expanding SGD to 3D 23

3.1.2 Performance Analysis and Improvement 26

3.2 Experiment Procedure 28

3.2.1 Setup 28

3.2.2 Experiment Process and Parameters 32
3.3 Eye Tracker Implementation 34

Chapter 4 Results and Evaluation **36**
4.1 Search Task Performance 36
4.1.1 Search Task Performance in Static Trials 36
4.1.2 Search Task Performance in Dynamic Trials 37
4.2 Level of Distraction 39
4.2.1 Static Scene Distraction 40
4.2.2 Warm-Cool/White-Black Distraction compared 41
4.2.3 All or nothing 42
4.3 Image Saliency 44
4.3.1 Saliency Toolbox 44

Chapter 5 Conclusion/Future Work **49**

Appendices **52**

Bibliography **58**

List of Figures

1.1 Map design of first level from Doom[e1m] 2
1.2 Screenshot from Half-Life 2: Episode Two[Val04] 4
1.3 Screenshot from Crysis[EA07] 5
1.4 Dishonored[Bet12], Deus Ex: Human Revolution[SqE11], Assassin's Creed: Black Flag[Ubi13] 6

2.1 Example construction of the salience map[sal] 10
2.2 The normalization operator $\mathcal{N}(.)$[IKN+98] 11
2.3 A hypothetical image with current fixation region F and predetermined region of interest A. Saccadic masking illustrated in table.[BMSG09] . . 12

3.1 Application of SGD to a particular pixel in an example scene 25
3.2 Demonstration of using important item coordinates from mipmapped buffer texture 27
3.3 Screenshot of initial environment to test experiments in 29
3.4 Screenshot of environment [hou] to test experiments in 30
3.5 Trial Setup 33
3.6 Eyetribe eye-tracker[Eye14] 35

4.1 Average performance per static trial 37
4.2 Average performance per dynamic trial 38
4.3 Average performance in dynamic scenes per image modulation technique 39
4.4 Average Level of Distraction reported per static trial 40
4.5 Average Level of Distraction reported per color modulation type in moderate intensity trials 41

4.6 Average Level of Distraction reported per dynamic trial 43
4.7 Saliency model architecture [sal14] 45
4.8 Saliency Toolbox Saliency map results for random screencaps of 3D scene w/ 0.2 intensity. The yellow outline denotes the most conspicuous region reported by the toolbox. The text above each image refers to the feature map with greatest impact on the overall saliency map. For example, the first image has its strongest features detected in the Red-Green feature map[IKN+98]. 46
4.9 Saliency Toolbox Saliency map results for random screencaps of 3D scene w/ 0.1 intensity. The green circle denotes the location of the region of importance. 47
4.10 Saliency Toolbox Saliency map results for random screencaps of 3D scene w/ 0.1 intensity, with toolbox parameters adjusted to weight intensity maps less. Region of importance is notably more successful in its detection as the winner. 48

Chapter 1

Introduction

This dissertation proposes the use of an autonomous system to work in tandem with modern graphics engines and libraries. This system is intended to allow designers to quickly and easily flag "areas of interest" in a 3D environment. The designers should be confident in the knowledge that the methods used to render these areas interesting - be they locations in the environment, objects, or even NPCs - are not only implemented in a manner that is not obvious to the player(so as to ensure the player's immersion in the environment) but are also backed up by extensive experimentation and analysis of their effectiveness.

The system is image-based in nature, and acts on a rendered image using information gained from earlier passes of the scene. Areas of the scene considered important are used as variables to adjust the manner in which the attention guidance is implemented. The adjusted image is rendered to the screen, naturally with the intention of subtly manipulating the user's view.

1.1 Motivation

As computer graphics become more and more capable of displaying environments in greater detail then ever, the need to be able to guide users of a 3D application through a 3D environment becomes ever-more challenging and important. Modern videogames attempt to place a great emphasis on ensuring players can navigate through environments with as little (detected) guidance from the designers as possible. Designers try to encourage this sense of autonomous player navigation by providing directions to the player in as subtle a manner as possible. These directions can be as minor as lighting a particular location in a 3D environment slightly more aggressively than its neighbouring areas, setting textures of higher resolutions to objects of interest, or highlighting them with outlines.

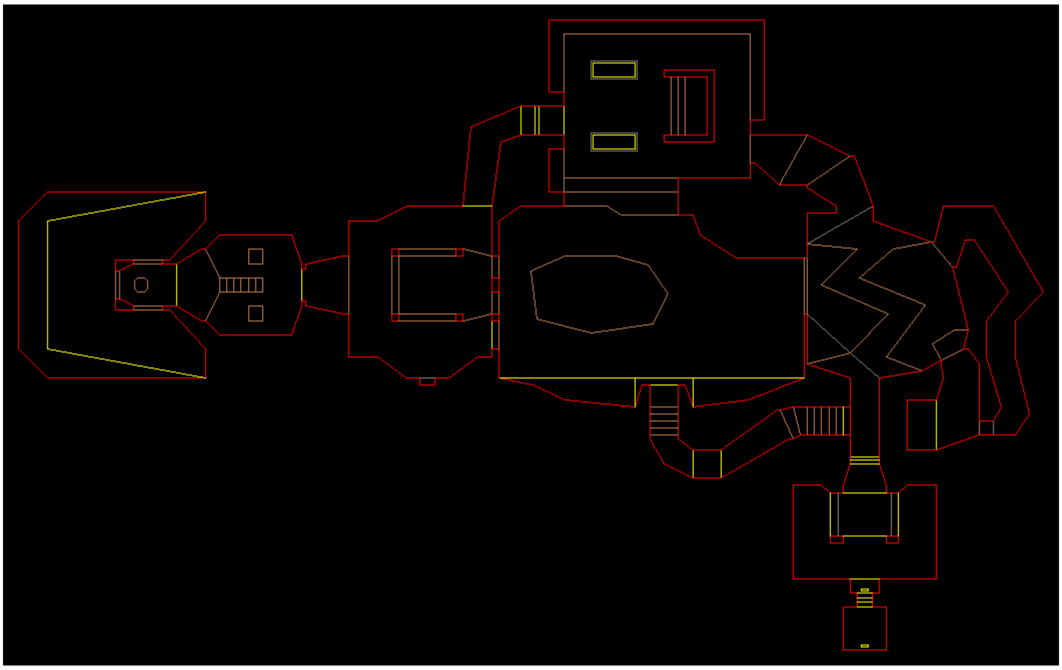

Figure 1.1: Map design of first level from Doom[e1m]

Such methods of player guidance typically occur in games where the player is being led down a set path to their goal. Though the environment being navigated may not give the player the impression of being heavily guided through it, there is always a handful of ways (possibly just one) to complete a particular level. Doom [Id93] was a title that pioneered the development of level design being a core component of a videogame structured in this way(Figure 1.1). The player begins in one area of a level and must fight and navigate their way to a clearly marked exit. Levels often provide multiple approaches to take in order to complete the level, but the sequence of events is carefully crafted by the designers of the game. For example, a level may require a player to acquire the red key to open the red door, then the blue key to open the blue door, at which point they can access the level's exit.

Level design such as this necessitates very careful placement of items and enemies, as well as meticulous map layouts for the level. The aim is to challenge the player with regards to navigating their way to the exit, without getting lost or frustrated in the process. To help with this goal, game designers incorporate various design philosophies to help the player find their way through the environment.

Level designers often cite lighting as being one of the most effective ways to guide players. Another method would be providing higher resolution textures to objects of interest in the world, so as to allow the player to sub-consciously interpret higher quality objects as the game's internal language indicating what is and isn't important. These approaches remain heavily utilized in modern games, such as Half-Life 2[Val04] and The Last of Us[ND13].

Though these approaches are regularly found in games containing what is considered good level design, their implementation still necessitates repeated play-testing to ensure they manipulate players' actions as intended. The philosophy of good level design generally has no concrete definition, and tends to vary greatly from one game studio to another, and such a topic hasn't received a great deal of attention in academic papers.

Discussions by developers concerning level design often revolve around finding ways to guide players in a non-intrusive way. The challenge of such a task is considered non-trivial by level and game designers. Figure 1.2, taken from the commentary track of Half-Life 2: Episode Two provides a good demonstration of how some level designs can be simply considered too complex.

For example, Valve, the company behind the Half-Life series, make heavy use of play-testers to tune a particular level to cause the player to approach areas and obstacles in the level exactly as they intend. One detail they often mention - particularly with regards to their first-person puzzle game Portal [VaP04] - is how challenging they often find trying to make players look up or down when examining their environment.

Figure 1.2: Screenshot from Half-Life 2: Episode Two[Val04]

Modern players become accustomed to the typical linear map layout of a set of rooms or even arenas, generally laid out in a horizontal manner. Players are rarely inclined to look up or down to a significant degree, since such an action is often considered unnecessary and rather jarring for most players.

The original Crysis [EA07] featured a level where the player was required to navigate an environment without the presence of gravity(Figure 1.3). The increased freedom of movement resulted in many players becoming lost and disoriented, despite the fact that the level was quite linear and there was usually only one way to enter and exit a particular room, and the player character would also automatically re-orient himself into an upright position whenever he rolled or flipped. Designers alleviated this issue by relying heavily on bright, warm-coloured lighting that stood out from the cool grey/blue architecture of most of the rest of the environment.

Figure 1.3: Screenshot from Crysis[EA07]

There are also cases where developers have opted to simply highlight the player's intended goal in a very noticeable way. Figure 1.4 demonstrates games that opt for icons and high-contrast outlines to indicate important objects or paths to players. Keep in mind that these kinds of games tend to provide the player with more freedom than previous examples, and as such, present designers with a greater challenge when it comes to guiding players to look in certain directions, interact with certain objects, and go down certain paths.

With that said, however, the impediment to the player's immersion into the game world that is brought about by these HUD elements cannot be ignored. Immersing a player into a photo-realistic game world is a high priority in modern games, and this can be hampered when the player's vision is drawn to clear "sign-posting" elements that remind them they are still playing a videogame.

The above discussion points out the heavy reliance on play-testing or obvious sign-posting that occurs when crafting a modern guided gameplay experience. Rigorous play-testing requires a large amount of time that could be better spent elsewhere on a project.

Figure 1.4: Dishonored[Bet12], Deus Ex: Human Revolution[SqE11], Assassin's Creed: Black Flag[Ubi13]

1.2 Methodology

The goal is to construct a way of guiding users through 3D virtual scenes using various image adjustment and rendering techniques. The guidance method is intended to operate by gaining the attention of the user in a subtle way. This is done by constructing implementations of these techniques, then testing the effectiveness of such a system as well as the underlying requirements of evaluating such a system via experiments and performance testing.

The methods presented here that are intended to guide players are heavily based in the field of computer vision. As such, the viability of eye trackers in assisting with the user guidance methods is also investigated, though this is in addition to the main goal of developing a system that operates without the need for an eye-tracker.

In addition, performance of the system is also evaluated using existing tools for testing how particular regions of an image are considered conspicuous based on architecture emulating early primate vision.

1.3 Contribution

We present a gaze direction system to guide user attention within a 3D virtual environment in a subtle manner, so as to avoid impeding their immersion into the potential interactive experience.

Through user experiments and performance analysis, we have gained new insight into the advantages offered by such a system. We have shown how the attention guidance methods presented here can be used to successfully improve user performance when it comes to search tasks in a 3D scene, thereby displaying their viability in an interactive 3D application. We also perform an analysis of user perception on the subtlety of the attention guidance method, where we gain new understanding of the restraints that can be imposed by such a system, as well as valuable data that can be used to develop the attention guidance system further.

1.4 Summary of Chapters

This dissertation is structured as follows:

- **Chapter 2** gives an overview of previous work done related to attention guidance, user attention models, and user game behaviour. A detailed overview is given to the main works relating to this project, subtle gaze direction and salience mapping.

- **Chapter 3** presents the design and implementation of the attention guidance method for use in an OpenGL application. Technical performance is discussed, and the methodology behind the setup of the user experiments is also outlined.

- **Chapter 4** details the results gained from the user experiment, as well as an evaluation of said results. A discussion relating to the results and evaluation acquired from the afore-mentioned image conspicuity tool in relation to the system is also present.

- **Chapter 5** summarizes the project discussed here, and provides a discussion of future work.

Chapter 2

Background

2.1 Saliency-Based Visual Scene Analysis

The work of [IKN$^+$98] presents a good starting point for considering proven image modulation techniques for a gaze direction system. The paper describes a visual attenuation system which is designed with the neurological architecture of the early primate visual system in mind. The system is described as being built upon a second biologically plausible architecture and is related to the feature integration theory.

The system is intended to model a saliency map for an image. Salience is a term used in neuroscience to describe the visual prominence of an item, be it an object, person, pixel, etc. . Essentially, the salience of an item describes the state or quality by which it stands out from its neighbours. An example of the construction of a salience map can be seen in Figure 2.1. The saliency map described in this paper topographically codes for local conspicuity over the entire visual scene, a map which is believed to be located in the parietal cortex[GKG98], and various visual maps of in the pulvinar nuclei of the thalamus[RP92] of primates. This information is of course relevant to the overall understanding of the human brain with relation to visual stimuli, since that is the area of highest priority in a system for gaze direction.

In [IKN$^+$98], input images are presented in the form of static color images. The images

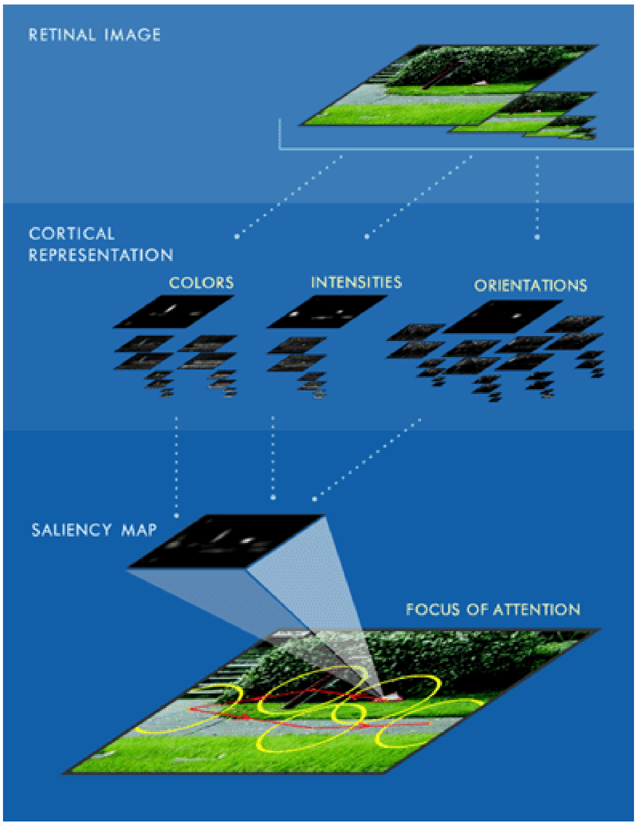

Figure 2.1: Example construction of the salience map[sal]

are initially linear filtered to extract early visual features(ie. color, intensity, orientation). Several forms of these features maps are constructed, then formed together to into three conspicuity maps for the three visual feature channels previously mentioned. A map normalization operator $\mathcal{N}(.)$ is used to deal with the varying ranges of the feature maps, avoid the loss of salient objects in certain maps due to noise from others, applying - by means of multiplication - the difference between a map's maximum and the average maximum across the map. Figure 2.2 shows a visualization of this process. The final saliency map is defined as:

$$S = 1/3(\mathcal{N}(\hat{\mathcal{I}}) + \mathcal{N}(\hat{\mathcal{C}}) + \mathcal{N}(\hat{\mathcal{O}}))$$

The paper goes into detail about the effectiveness of the salience map, as well as its apparent superiority to spatial frequency content models, but the examination of what in an image would actually be considered salient is of more importance here.

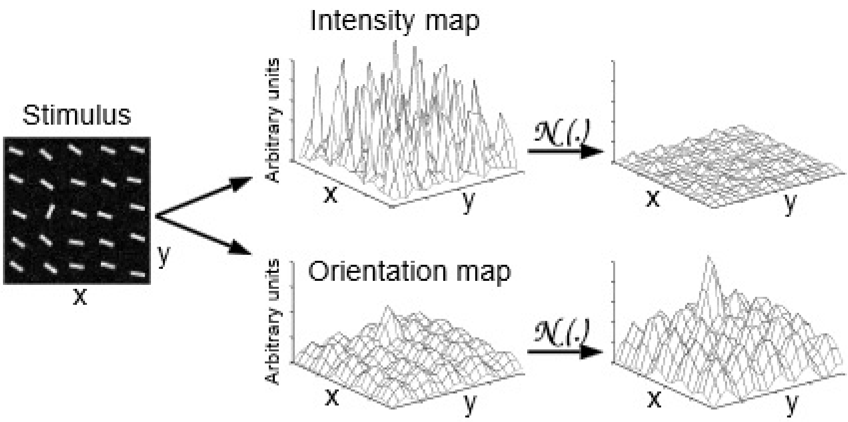

Figure 2.2: The normalization operator $\mathscr{N}(.)$[IKN+98]

Ultimately, the conclusion of the paper presents what is described as a robust model that emulates primate vision behaviour to a degree that the authors consider robust. They note that the question of whether the model is sufficiently robust is somewhat difficult to prove, since the decision of what one would consider salient in an image is rather subjective. They justify their model in its ability to detect objects and faces, even within images where noise is a notable issue.

The paper is most interesting in its methods undertaken to construct the salience map. The automation of constructing what would be considered a map of "interesting objects" within an image is a particularly important aspect of the paper.

2.2 Subtle Gaze Direction

There are various published approaches to adjusting a viewer's gaze when observing an image, as well as studies of the characteristics of the human brain in relation to visual stimuli. One of these approaches is known as Subtle Gaze Direction.

Subtle Gaze Direction is a term that was initially presented in [BMSG09]. The paper presents a technique of using eye-tracking to ascertain the effectiveness of an image modulation procedure for directing what a viewer looks at in an image. The technique

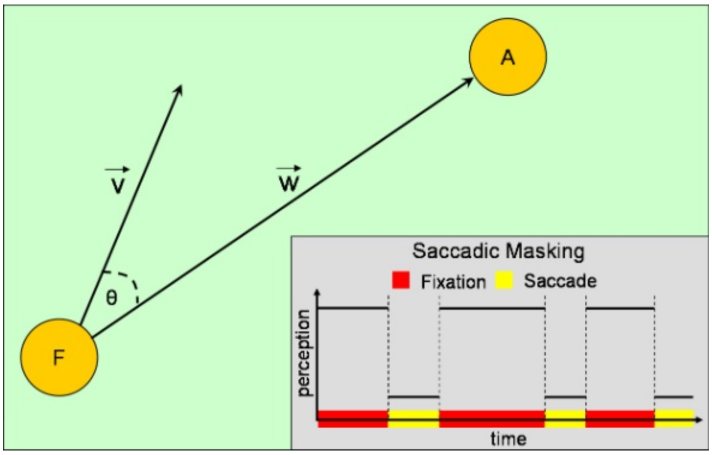

Figure 2.3: A hypothetical image with current fixation region F and predetermined region of interest A. Saccadic masking illustrated in table.[BMSG09]

works off the knowledge that a human being's peripheral vision has a very poor acuity(ie. capacity to see fine detail) compared to their foveal vision(ie. the portion of our range of vision offered by our eyeballs that offers 100% visual acuity - the centre of one's line of sight, in other words). An additional adjustment made in the process of the image modulation is that the modulation of the image is reduced to the point of termination via the process of saccadic masking, where a saccade describes the fast movement of the eye. The result is the viewer not perceiving any change in the image. Naturally, this additional functionality utilizes the eye tracker to detect the saccadic velocity of the viewer's gaze.

This paper points to various previous papers that deal with the pattern of eye movements, as well as the manipulation of such movements. One notable source here would be [PN+03], wherein the research presented strongly suggests that our gaze is naturally drawn to areas of an image that are of high contrast. [OM] also shows that peripheral vision tends to react faster to visual stimuli than foveal vision.

The subtle gaze technique operates by using the eye tracker to detect the current saccade of the viewer's vision and using vector calculations to discern the orientation and distance of the current saccade relative to the point of interest, which is predetermined(ie. it is the point of the image that is intended to be viewed by the viewer via image modulation). Image modulation occurs in the form of luminance modulation or warm-cool modulation on pixels that encompass an area around the point of interest.

Upon initializing modulation, saccadic velocity is monitored using the eye tracker, ie. the saccades of the eye from one refresh of the eye tracker to the next is used to detect the saccadic velocity. The angle between the current saccade and the point of interest is also calculated at each updated with a simple dot product calculation, using the two vectors that form the current saccade and the vector from the current saccade to the point of interest. This is visualized in figure 2.3. The resultant angle can be used to detect whether the current saccade is drawing close to the point of interest(a value of less than or equal to 10 degrees strongly suggests such a scenario).

The actual modulations applied to the image are described as simple alternating interpolations of the region of interest(ie. the pixels around the point of interest). In the case of luminance modulation, these interpolations alternate the pixel with black and white, while in the case of warm-cool modulation, they are interpolated with a warm color and a cool color. The technique described in the paper has these interpolations occur at a frequency of 10 Hz.

Warm-cool modulation uses the following equation for the interpolation of a pixel p = (x,y), where (x,y) is the pixel coordinate:

$$col(p) = ((w * i) + col(p) * (1 - i)) * f(p) + col(p) * (1 - f(p))$$

where col(p) is the color of p, w is a warm color, f(p) is a Gaussian falloff function(described in [You87]), and i is some scalar value in the range [0,1], which controls the intensity of the modulations. In the experimental step with participants, i was adjusted in steps of 0.005 to locate a value where modulations were just noticeable. The cool modulations for the cool interpolation of the cycle, and the black and white modulations of the luminance modulation cycle are analogous to this equation. The region of interest is set with a radius of 32 pixels around a point of interest.

The results found the image modulation approach to be successful in manipulating the attention of the viewer every time it was applied to the images. That said, it was not always successful in drawing the viewer's gaze all the way to the modulated region every time, generally falling to within one to two perceptual spans. This was believed to be a result of the modulation terminating upon saccades getting closer to the region of interest. On top of this, viewers were questioned on whether they found anything unusual in the images, a term which was not defined by the researchers. The term was attached to an overall perception of the image quality, and the mean image quality was found to be lower on average for the modulated images than for the unmodulated images, likely due to the potentially distracting nature of the image modulation.

In the papers conclusion, the viability of extending the technique to video is mentioned. The idea is presented that the presence of motion would serve to further suppress the subtle modulations, which would result in the modulated videos not appearing to suffer from the perceived sense of image quality degradation that occurred in the user experiments for the image modulation. Potential issues are pointed out, such as the challenge of maintaining frame-to-frame coherency of modulations and adjusting the positions of the modulations to follow moving regions of interest to ensure consistency.

[MBG09] builds upon the base technique of Subtle Gaze Direction by presenting a new set of experiments that incorporate the presence of distractions into the user experiments.

The paper acts as a re-tread of sorts, re-describing the concept and technique of subtle gaze direction, and once again implementing a set of user experiments that initially don't incorporate distractions, before bringing them into the fold, so to speak.

The paper re-affirms the motivation behind subtle gaze direction in the introduction. It justifies the emphasis of suggesting areas of a screen to observe over flat-out presenting areas to the user by maintaining the visual experience of the user as a high priority. The paper once again utilizes the low visual acuity of peripheral vision as a basis for the application of the image modulation in certain parts of the image, using a eye tracker to detect saccades from point to point of the viewer's gaze. Of particular note is the use of the bubble objects as the targets of the search task presented.

At first, three distinct user experiments were performed; firstly, without modulation, secondly, with subtle modulation, and thirdly, with excessive modulation. Excessive modulation was achieved by extending the modulation radius to 0.125 degrees of visual angle, as opposed to 0.04 degrees for the subtle modulation.

The results of the experiments indicated a notable improvement in search task performance for the modulated experiments over the non-modulated experiment. In the process of seeking targets within images, users noted that, although they could identify the occurrence of modulations in the obvious modulation experiments(while not identifying the modulations in the subtle modulations experiment), they did report being able to locate targets in the image they would not have able to otherwise.

An ANOVA for the differing effect of condition concluded with a notable variation between the three groups of modulation. Essentially, the results indicate that group 1 (no modulation) varied considerably to groups 2 and 3, while group 2 and 3 possessed little variation compared to one another. The findings gave credence to the hypothesis presented by the paper that the mere presence of modulation is the most notable effector - in terms of assisting with search task performance - as opposed to the level of modulation applied.

Following these experiments, distractions are introduced to the search tasks. The motivation behind their introduction is the issue of false positives that can appear in the application of image processing techniques that are regularly used today. The authors of this paper note that the high level of usefulness offered by these visual algorithms despite the presence of false positives, and this becomes the motivation behind testing the effectiveness of their gaze direction technique in the presence of distractions.

The distractions are defined as extra image modulations that do not occur at the points of interest within the image. The experiment was the same as that of the subtle modulation experiment(the second experiment) with the only change being the addition of distractions. The locations for the distractions were randomly applied, and three distractions were added to each image.

False positives were accounted for in the experimental results. However, it was found that the negligible variation in performance between the subtle modulation experiment and the obvious modulation experiment was again present when those two experiments were compared to the subtle modulation with distractions experiment. In certain cases (ie. in the cases of certain images), the distractions even appeared to benefit in the search tasks, namely in the images that depicted interior and office scenes. This leads the authors to consider the undertaking of a follow-up experiment exploring the effect of scene context and whether certain characteristics of a scene can favour the presence of distractions in the modulation.

One particularly interesting detail the authors take from their results is the fact that the presence of distractions resulted in a higher degree of accuracy in the search tasks. They theorize that the distractions caused viewers to distribute their gaze more evenly over the image. The issue of in-attentional blindness [MR98] is brought up, where a person's ability to detect salient features in an image is greatly diminished if their gaze is focused elsewhere on the image. The authors believe that their technique causes viewers to move their gaze over larger portions of the image, instead of focusing on a single portion of the image for long amounts of time.

2.3 Other attention guidance

The paper [NI05] presents a computational model for the task-specific guidance of visual attention in real-world scenes. Naturally, such a paper is of considerable relevance to the topic of gaze direction. The paper builds upon the foundations presented by [IKN+98] as regards constructing saliency maps that emulate the workings of the human visual system. This paper focuses on taking the concept further by presenting a system that adopts various practises that are considered natural characteristics of the human eye and brain as regards vision.

The paper describes the process of modelling how task influences attention. The processes starts out by first determining what to look for. An ontology, which is a knowledge base containing entities and relationships to yield task-related entities and corresponding relationships. Next, the relevance of the of the task-related entries are determined and the most task-relevant entry in the scene is looked for.

The next step is of particular interest. The model presented in the paper is intended to detect a given target quickly and reliably in a given scene. This is achieved via a process of modulating the bottom-up salience of the target. In the context of images, or parts of images, modulating is generally inherently a top-down approach, and that remains true in this context as well. The model biases the low-level visual system with the known features of the target so as to render it more salient, after which the most salient scene location is then chosen as the focus of attention. The idea is that the aforementioned biasing that occurs for the target of choice should result in it being the considered the most salient portion of the input image.

The steps following this one focus on the process of recognizing the entity that is at the focus of attention following biasing. However, the biasing is, as previously mentioned, of notable worth, so that will be discussed here in more detail. Section 5 of the paper describes the top-down biasing method for detecting desired objects. The process begins by describing a representation of the image, or more specifically, objects in the image, in diverse, complex, backgrounds. The description is constructed from the initial low-level feature maps tuned to color, intensity and orientation, just like those mentioned in [IKN$^+$98].

The system begins by learning the object representation. In other words, it achieves this by calling upon what is described as its Symbolic Long Term Memory, or symbolic LTM. The LTM is essentially a memory bank of the learned visual representation of the target. A knowledge base, in other words. The combination of the LTM and the visual Working Memory (or WM), which is a descriptor for short-term memory in the context(ie. only stores information on a maximum of three-four objects at a time) create a representation of the object.

The learned representation of the object that is stored in the LTM is then used to bias the salience map. Let us denote a feature by f. f is considered to be relevant and reliable if its main feature value is high and its feature variance is low. A feature's contribution to the salience map is weighted as $\mathrm{R}(f)$, where

$\mathrm{R}(f) = \textit{relevance of feature } f = \frac{\mu f}{1+\sigma(f)}$.

This is the point where the various classes of features in several processing channels(color, intensity, etc.) are used to create a channel hierarchy H, where H90) is the set of all features at different spatial scales, H(1) is the set of subchannels formed by combining features of different spatial scales, H(2) is the set of all channels formed by combining subchannels of the same modality; ... H(n) is the salience map(where n is the height of H). Each parent channel then promotes itself proportionally to the maximum feature weight of its children channels. The result of this is that channels that have little weighting on the salience of a particular feature (eg. a long straight line in a region where color varies only slightly), allowing the existence of actual features to take precedent in the weighting.

At each level of the hierarchy, the child maps of each parent map are summed into a unique map, all the way to the root, where a salience map is formed. The nature of this salience map causes it to prune away features where there is too many competing features from different maps while promoting features that experience little competition of equal magnitude. The paper refers the reader to

[IK99] presents four different strategies to constructing a bottom-up salience map: 1) Simple normalized summation, 2) Linear combination with learned weights, 3) global nonlinear normalization followed by summation, and 4) local nonlinear competition between salient locations followed by summation.

The paper acknowledges the process of top-down voluntary control towards locations of cognitive interest, even though these may not be particularly salient, a form of image influence that is quite similar in description to Bailey and McNamaras image modulation techniques for their Subtle Gaze Direction system[BMSG09].

That said, the paper largely makes abstraction of the top-down component for controlling the focus of attention. The bottom-up, scene-driven component of visual attenuation is the primary focus of this paper. In other words, the paper is less concerned with how to actually manipulate one's attention(or gaze) to a particular point of a given image, and more concerned with understanding - in biologically plausible computational terms - what influences the initial raw manipulation of attention in the first place. The paper justifies study by pointing out that bottom-up cues would form the primary influence of the influence of a viewer's attention within the first few hundred milliseconds. The paper refers the reader various articles and papers.

One of those papers is [Wol94], which describes a search model intended to emulate the behaviour of a typical human when faced with a rudimentary visual search task.

Of note are some observations made on the biological processes behind performing search tasks such as these. Firstly, the human brain does not contain the capacity to perform all visual functions at all locations in the visual field at a time. Hence, input received from the eyes tends to be discarded in large quantities, save for the information received from the fovea, which receives much more detailed sampling. Secondly, information received from the eyes tends to be processed selectively. In other words, certain visual recognition functions may only be performed at particular regions of the visual field. The fact that it is nigh-impossible to read two separate parts of a body of text at the same time is a good example of this extreme prioritization of certain parts of the visual field over others.

The paper describes a manner of deploying limited capacity visual processes only on certain parts of the visual field that are deemed worthy of attention in the context of whatever the current search task may be. A system is described that attempts to model the way the human brain interprets data when it comes to visual search tasks, as well as presenting a computer simulation that produces data to closely resemble that of the data produced by humans.

The paper gives a rather comprehensive overview of guided search modelling at the time of its release(1994).

The stated basic goal of the Guided Search Model is to explain and predict the results of visual search experiments. The idea is that the model ought to be able to explain our ability to find a desired stimulus in a normal, continuous scene.

The term feature maps appears in this paper, also. Their definition remains essentially the same as that mentioned for other papers, though which feature maps are chosen for this paper can differ slightly. A single, multi-dimensional map is considered to be capable of representing all of these feature maps as one. The guided search model presented in this paper employs one map per feature type, though the paper acknowledges evidence to suggest the possibility of differences between features(eg. there may be a need for multiple color maps but only one orientation map, similar to how Itti-Koch approach feature maps).

In the simulation of the guided search model presented in this paper, hereby referred to as GS2 (as in the paper), only feature maps describing color and orientation are used.

Each map is an 8x8 array, with numerical values designating the orientation or color at each location. For orientation, values are given in degrees clockwise from vertical. The values for color are an arbitrary mapping of spectral colors from red, yellow, green, and blue onto numerical values.

The information from these maps is assumed to be sufficient for the task of specifying colors, orientations and depths. Guided Search operates by attempting to discern parts of the image considered worthy of further attention, a basic visual process which is performed in parallel. This is modelled as differential activation of locations in a feature map. A greater activation at a location in a map denotes a likelihood that more attention will be directed to that position in the map.

There are two types of activation that are to be considered; bottom-up and top-down. Bottom-up activation measures how unusual an item is in its present context. The similarities between the concept of unusual the definition of salience should be noted. High-level knowledge related to the overall search task or the object being searched for is not considered in the case of bottom-up activation. The feature maps mentioned previously are heavily utilized here. The process of analyzing them for what would constitute unusual is described as non-trivial in the paper, and there are multiple complex steps involved in performing bottom-up activation.

Top-down activation is the process of guiding attention to a specific item in the visual field utilizing more descriptive properties of the item being sought than primitive properties such as color and orientation. The feature maps are still utilized, but they are generally used as steps towards the final goal of the search. For example, if a green crayon was being searched for in an image, a top-down activation approach would begin its searching by establishing the sufficiently positive results of the green feature map as an initial stepping stone. From here, the orientation map would be considered for the process of discern whether objects represent would would constitute a green crayon.

The model also refers to an Activation map, for which the goal is to guide attention. In terms of the guided search model, the activation map operates as expected, summing the activations in all the feature maps. The processes under attentional control are influenced by the activation map before coming to a conclusion that an particular item in question is what is being searched for. The paper makes reference to the concept of pop-out, where an item being searched for appears to leap out at the viewer. Such an effect is believed to be the result of the target(the item being searched for) having a particularly high activation compared to the distractors(in other words, items in the image that clash significantly with the target).

The modulation of bottom-up and top-down activation is also touched upon in the paper. References are made to papers that discuss various methods of modulation to capture attention.

2.4 NPR attention guidance

Though this project is intended to result in the implementation of a gaze direction system for use in photo-realistic environments, there are various works that deal with the process of manipulating user attention in Non-Photorealistic Rendering (NPR). [DS02] presents a method of NPR that focuses on the use of bold edges and large regions of constant color along with the use of a hierarchical structure to connect meaningful parts of an image to adjust user attention. It, like many of the attention guidance papers mentioned here, is based on bottom-up human perception models. Similarly, [SD02] presents the use of initial, rapidly collected eye-tracking behaviour and human perception modelling to adjust the painterly effect rendered to an image. [WOG06] builds upon this work by applying similar abstraction techniques to real-time video.

[CDF+06] presents an interactive attention directing system intended to work using stylized rendering to emphasize certain aspects of a scene, while de-emphasizing others. The rendering method focuses on adjusting line density in the scene to perform an effect similar to that of the focus pull implemented with real cameras, with the intention of drawing one's attention to certain parts of the image.

2.5 Other papers

One paper relevant to this project is [SSWR08], which presented a search task designed with the intention of analyzing user behaviour when faced with a game-oriented task. Of note was the finding that no difference presented itself between eye fixations during active gameplay and passive observation.

A curious paper related to the topic of gaze direction is [ZGSS12]. As the title implies, the paper approaches the concept of representing space with the intention of constructing a navigation model that could be utilized to aid people with Alzheimer's Disease (AD) dealing with the problem of disorientation and getting lost behaviour. The model is based on the process of training the person's tendency to pay attention to salient and autobiographically important landmarks, as well as the routes between them.

The paper makes use of experiments revolving around subjects navigating virtual environments. The motivation here is to better understand the relationship between the real world and a mental representation of the real world, which is described in the paper as the cognitive map. The virtual maps are tested to check whether they obey the laws of Euclidean geometry. Certain virtual maps were designed to intentionally break these laws, and as such were defined as impossible virtual environments. These violations of Euclidean metrics or planar topology were then checked to discern whether or not they caused problems in navigation performance.

The system that is presented in the paper utilizes salience-based models with regards to pathfinding. The virtual environment it constructs mainly consists of building exteriors of various shapes and sizes. A mapping operation is presented that can translate the different weighting or representations of the objects in the environment. This is used in conjunction with a distance measure for each facade (building face in the virtual environment) to estimate a salience measure for each building facade within a particular context. This and various other experimental methods employed in this paper could be of benefit as regards navigation of 3D virtual environments.

Chapter 3

Design and Implementation

3.1 Attention Guidance in OpenGL

3.1.1 Expanding SGD to 3D

The following section describes how the attention guidance method is implemented. This process entails [BMSG09] essentially being re-fitted for use in 3D scenes. As that paper describes, SGD defines a particular pixel within an image as a pixel of importance. It is from this single location that the part of the image to be modulated is decided. All other pixels in the image are subjected to the following equation:

$col(p) = ((w * i) + col(p) * (1 - i)) * f(p) + col(p) * (1 - f(p))$

$col'(p)$ is the new color of the pixel. w is the color to be applied to the image(and is an interpolated value between two colors). i is the intensity of the image modulation, and dictates the strength of the image modulation. $col(p)$ is the original pixel color.

Finally, $f(p)$ is a Gaussian falloff function. The presence of this function is to adjust the intensity applied to a particular pixel by its relation to the pixel of importance. Generally the relation is in terms of distance(ie. the further away the pixels is, the greater the degree to which the Gaussian function diminishes the the intensity). This results in a blending effect with pixels in a region of importance, which is the image space around the pixel of importance that experiences modulation. Pixels closer to the pixel of importance are effected more heavily than pixels further away but still fall within the region of importance.

Implementation of this SGD method in a 3D virtual scene is done in such a way that the person constructing the scene can specify objects of importance within the scene. The image-based nature of the implemented method is still retained in the 3D version of SGD. Hence, the solution here is to provide the OpenGL application with a screen coordinate to define as the pixel of importance.

The application operates as follows; all world objects(ie. meshes) within the scene are provided an 'importance' attribute, simply a value signifying whether an object is intended to be considered of importance to the user navigating the environment.

Upon rendering the scene, the importance attribute is passed along with the rest of the relevant scene data to the shaders. This is where an extra render target texture is used, called the Importance Texture. As for the rendering pass it is associated with, it renders, for each pixel, the value of the importance attribute attached to the world object the pixel is rendering. The result is a binary image, highlighting important objects in the scene. According to [BMSG09], adjusting the positions of the modulations to follow moving regions of interest to ensure consistency is raised as a concern for the SGD method's use in video. The use of this importance texture ensures this consistency.

The next step is to find the location of the important object(s) in relation to screen coordinates. This requires a searching of the importance buffer texture before passing the screen coordinates of the centre of the important object out to the final shaders that apply SGD. The reason for this is that the 3DSGD is still ultimately applied to an image, hence the necessity to supply the shaders implementing it with screen coordinates.

The importance texture is fetched for searching. The implementation and performance of the fetch process in OpenGL is described in the next section.

The search for the pixel of importance is performed once the importance texture is available to read. Performance gains could be made here by ending the search upon discovery of the important pixels, but such an optimization would result in unstable performance that depends on where the important object is in image-space.

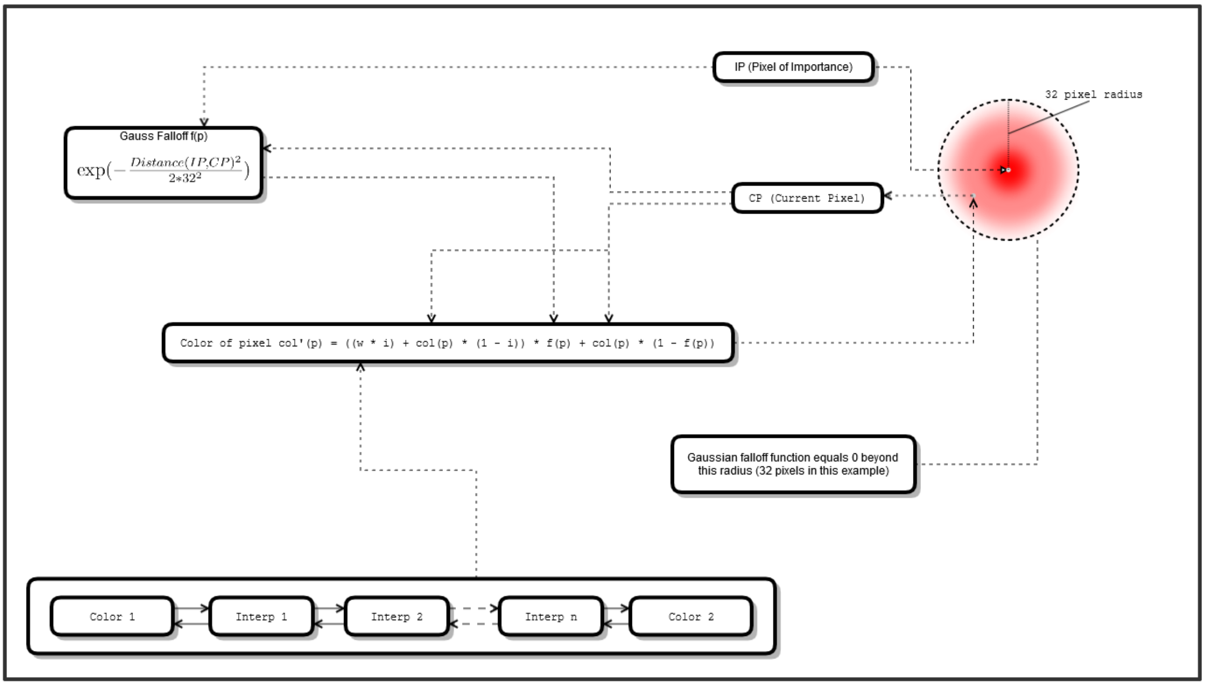

Figure 3.1: Application of SGD to a particular pixel in an example scene

The texture coordinates of the centre of the object of importance are then passed as a uniform variable to the final shaders that apply SGD on the rendered image. The final render target takes the rendered image as it is intended to appear without gaze direction and applies SGD to the image. Figure 3.1 demonstrates how a particular pixel's color is adjusted for an example scene containing a visible important object. The current value of w is also passed for the SGD implementation. Its value is acquired per frame by traversing back and forth across an array of interpolations between the two colors of the SGD method, at rate of 10 Hz, similar to the SGD method described in [BMSG09](hence, in Figure 3.1, the n would be 8).

3.1.2 Performance Analysis and Improvement

The 3DSGD procedure is based on image-processing techniques, and as such is itself image-based in nature. This promotes the use of the procedure as a generalized method for user guidance that can be implemented in modern 3D engines. However, the image-based nature of the procedure does present some hurdles when it comes to performance. By the very nature of the image-processing method, it must sample the rendered image before discerning the location of the object of importance in terms of the screen's pixel coordinates.

In the case of this procedure, the buffer texture being sampled is the importance buffer, which, as you will recall, is a binary image. The importance buffer is the same size, in terms of resolution, as the buffer texture of the rendered image. Modern screen displays generally tend to possess native resolutions of 1920x1080 and upwards. In the case of 1920x1080, that is a total of 2073600 pixels to be sampled. The importance buffer must also be fetched before performing the search for the 'important pixels'. All these operations take a sever toll on the performance of the application in OpenGL.

The pixel data consisting of bools cannot be improved upon. The size of the pixel data, however, is another story. The SGD method presented by Bailey et. al. is based on the process of implementing color modulations in regions of the screen where the user's foveal vision is not focused(ie. their periphery). The color modulation described there is deactivated the moment the eye is detected to be moving in the direction of the object of interest. As such, the exact position of the centre of the region of image modulation does not appear to be deemed of absolute importance, since the color modulation described is merely intended to focus a viewer's vision towards a specific region of the image, as opposed to focusing their vision to an exact spot on the screen.

This serves as a useful observation with regards to improving the performance of the OpenGL application. Namely, the resolution of the importance buffer is not required to be identical to the resolution of the final output image, and can in fact be several times smaller. Discovery of the pixel coordinates of important objects in the lower resolution importance buffer can then be extrapolated out to gain an approximate of the impor-

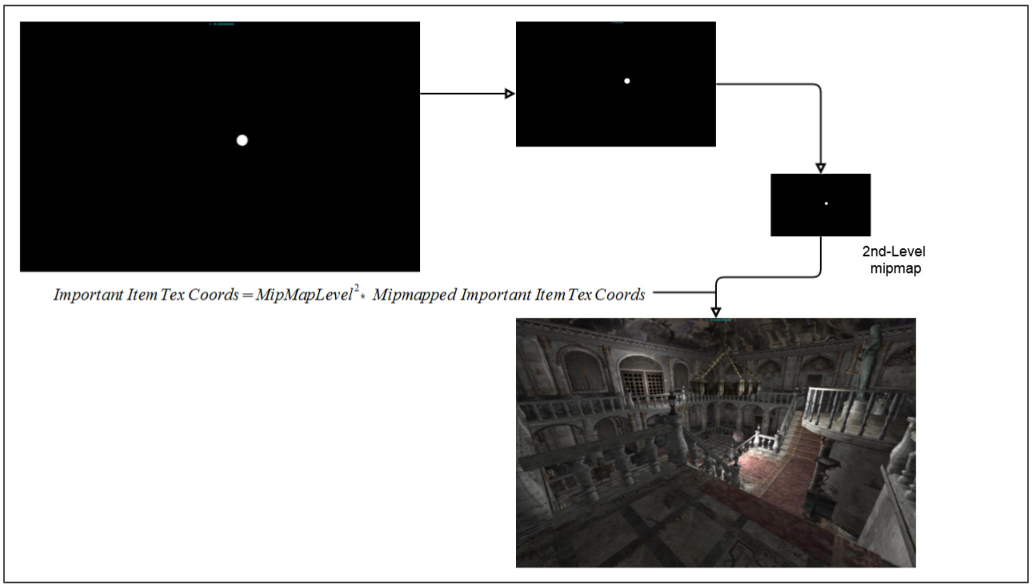

Figure 3.2: Demonstration of using important item coordinates from mipmapped buffer texture

tant pixel coordinates of the rendered image to apply the image modulations.

The process of reducing the importance buffer can be achieved in OpenGL via mipmapping, a common feature amongst modern 3D engines. Before fetching the importance buffer with `glGetTexImage()`, mipmaps are generated using the `glGenerateMipmap()` function, with `GL_TEXTURE_2D` as the target provided to the function following the binding of the importance buffer texture. This function automatically constructs mipmaps of various levels(See Figure 3.2).

The image level of the mipmap to use is provided as one of the parameters to the `glGenerateMipmap()` function. An increase of 1 in texel image level halves both the horizontal and vertical dimensions of the texture resolution of the previous level.

Analysis of the performance of the OpenGL application at varying image levels found level 2 removed all bottlenecks associated with the fetching and searching of the importance image. For a screen resolution of 1920x1080, that becomes a resolution of 1920/4 x 1080/4 = 480 x 270 = 129600 pixels, 16% the number of pixels originally. Running the application with this change results in a performance boost of 20-30 fps on a machine with an Intel i7 procesor and an Nvidia GTX 460.

It must be noted that the search for the important object texture coordinates are not the sole reason for the performance improvement, and the generation of the mipmaps results in a quicker fetch operation for the (smaller) importance texture as well.

3.2 Experiment Procedure

3.2.1 Setup

The following section outlines how the user experiment was conducted to gauge the effectiveness of the 3DSGD method.

Evaluating the effectiveness of the method required the testing of two distinct results; how effective the method is at actually guiding user attention in 3D scenes, and how subtle the image processing appears to participants during the experiment.

In terms of evaluating the effectiveness of the tool for drawing attention, a search task was proposed. The paper that introduces Subtle Gaze Direction [BMSG09] does not present users with a search task requiring them to look at particular features in the images. It simply analyses their eye saccades with the image modulation turned on and off.

[MBG09], which followed this initial paper, presented the use of small 'bubbles' as the objects of interest in a search task experiment. The use of these kinds of objects as objects of note seemed suited for use in a 3D virtual scene. By their nature, spherical objects that refracted and reflected light would be somewhat challenging to identify from their surroundings. They would not instantly stand out in an environment the way, for example, a red 'coin' would, which was an initial idea for the search task objects.

Following the decision to use 'bubble' objects as the objects of importance in the search task, a decision was made to build the experiment around two main types of search task; one where the camera was stationary in the scene(the static trials), and one where the camera traversed the scene on its own accord, following a set path and panning around the scene as it went(the dynamic trials).

Recall that the initial motivation behind this dissertation was to construct a system for guiding user attention in an interactive scene, much like a videogame. The difficulty involved with constructing such an experiment became apparent upon creating initial tests involving a scene consisting of a house with a garden and shed(Figure 3.3), acquired from [hou].

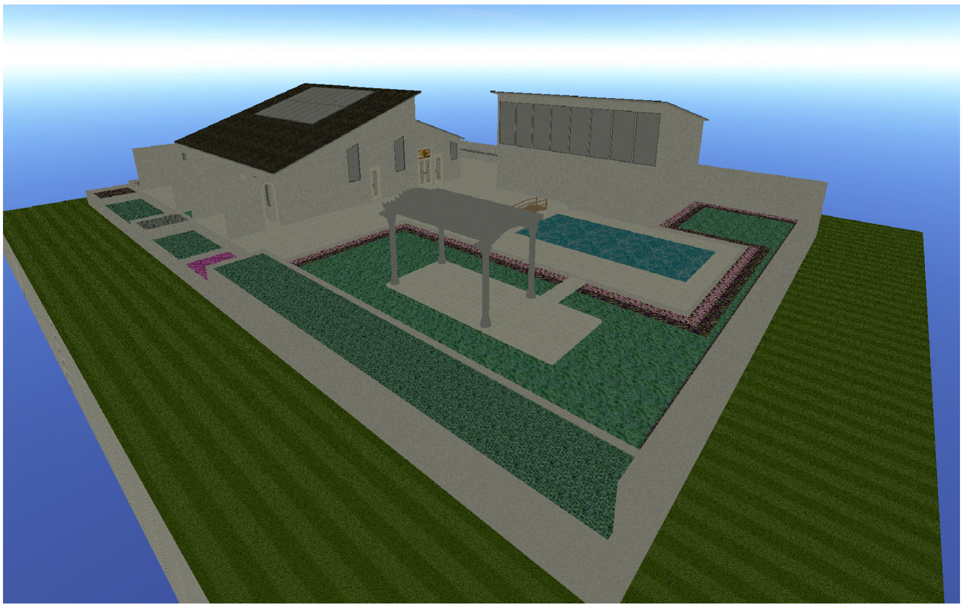

Figure 3.3: Screenshot of initial environment to test experiments in

This presented some issues. Namely, presenting participants in a user experiment with such an environment, where users - no matter where they were located in the scene - would always be occluded from many other parts of the environment. For example, were a participant to find themselves in the house interior (which is fully modelled) in the scene, they would not be able to see a large majority of the rest of the scene, like the garden, the shed interior, other rooms in the house, etc. . The result was a concern of the existence of too many variables due to the participant having too much freedom in their choices of were to go and what to look at.

Certainly, the end-goal behind the construction of this attention guidance method is its use in 3D videogames where players navigate vast, complex environments. However, successful evaluation of the effectiveness of this attention guidance method requires more constrained experiment parameters. A relatively large environment where the participant can move through doors and traverse the scene in a variety of different paths expands the amount of variables (as regards where the participant can go and what they themselves may find interesting in the scene) in the experiment drastically. This provided a strong motivator to scale down the size of the environment the participant would find themselves in. This resulted in using a single room model (Figure 3.4) from Resident Evil: Umbrella Chronicles[Cap07]. Lighting was setup for this scene to give it a more photorealistic appearance appearance, and a collision and navigation system was implemented to allow the camera to traverse the environment using the mouse and keyboard as controls, much like a videogame. A demonstration of this movement can be seen on the accompanying disc.

Figure 3.4: Screenshot of environment [hou] to test experiments in

Another potential cause of too many variables would be the inclusion of user interactivity into the user experiments. Naturally, this may come as a surprise for an attention guidance system intended to be implemented in videogames. However, first-person controls can consist of many differing forms of control(using a gamepad, using a mouse and keyboard, using only arrow keys, etc.). As a result, there is always an issue of participants potentially possessing considerably different levels of familiarity with these control methods. Constructing trials familiarizing participants with a chosen control scheme to the degree that all participants are roughly equally comfortable with the control scheme would not be achieved in any reasonable time frame.

As a result of the above concerns, a decision was made to scale down the size of the environment used for the user experiments, as well as to remove interactivity from the experiment to avoid the potential fluctuation of results that could come from introducing so many variables with these experiment features. Hence, the user experiments consist of various trial-runs, where the participant observes a 3D scene. For certain trial-runs of the scene, the camera remains stationery, looking in a static direction for the duration of the trial(a static trial). Other trials have the camera moving throughout the environment(a dynamic trial), adjusting where it is looking at the same time. This gives the effect of emulating a play-through of the scene as if it were being controlled by another user.

In the OpenGL application the camera is provided with an array of position way-points, which it moves through in sequence. It is also provided with an array of view target way-points, which it rotates towards in sequence. The camera's view target is defined as a normalized 3D vector. The view target array consists of normalized 3D points intended to be relative to the camera. The desired axis upon which to rotate the camera is achieved by normalizing the cross product of the camera's view target and its desired view target. The speed at which the camera rotates per frame is proportional to the magnitude of the angle between these two vectors, resulting in a smooth panning motion as the camera re-orients from target to target.

Placement of the glass spheres in the scene was initially intended to be automatic for each run-through of the trial. However, this raised several concerns. The first was that there was no way to place the spheres in positions where they would be guaranteed to not be occluded outside of performing expensive collision tests with the camera and the geometry. A more difficult issue to deal with was how the automatic placement of spheres would result in scenarios where certain trials would inevitably present an easier search task than others due to the automatic placements of the spheres. This presented a strong motivator behind the decision to manually place the spheres for each trial.

Manually dictating where spheres appear at all times ensures all participants have a fair chance to perform the search task. This did, however, raise a new issue. The goal here was to evaluate the effectiveness of the attention guidance method with a search task experiment involving glass spheres. The issue here is that the re-arrangement of the placement of the glass spheres results in another degree of 'unfairness' when it comes to comparing various differing degrees of intensity/color choice in the image region modulation process. Comparing multiple different parameters for this image modulation must ideally be carried out on the same trial-run with the same path, the same camera view targets, and the same sphere placements.

The solution proposed here was to construct a set amount of trial-runs with unique sphere counts and placements, and to mix up the order in which the participant completes these trials, so as to prevent them from detecting a pattern in the number of spheres in the scene. However, simply mixing up the order in which participants complete trials with different parameters was deemed insufficient to prevent participants from realizing they were running through repeated scenes. Hence a solution was implemented to apply a mirror-flip effect to the entire final rendered image on certain trials. In doing so, participants would still be completing the same trials, and in doing so, ensuring the 'fairness' the experiment, but they would also be unaware they were doing so.

Following the experiment, participants were informed of the number of separate sphere placement trials and the use of mirror-flipping to perform the same trials multiple trials. No participants reported detecting the re-use of sphere placements or the use of mirror-flipping to hide said re-use.

3.2.2 Experiment Process and Parameters

The experiment consisted of a total of ten trial-runs, in the same order for all participants, arranged as seen in Figure 3.5:

As mentioned previously, 'static' and 'dynamic' refer to a stationary camera that looks

TRIAL #	# OF SPHERES	INTENSITY VALUE	MODULATION TYPE	MIRROR FLIP
1(static)	14	0.07	WARM-COOL	NO
2(static)	14	0	N/A	YES
3(dynamic)	25	0.07	WARM-COOL	NO
4(dynamic)	20	0.07	WHITE-BLACK	NO
5(dynamic)	25	0	N/A	YES
6(dynamic)	25	0.07	WHITE-BLACK	NO
7(dynamic)	20	0.07	WARM-COOL	YES
8(dynamic)	25	0.1	WARM-COOL	YES
9(dynamic)	20	0	N/A	YES
10(dynamic)	20	0.1	WARM-COOL	NO

Figure 3.5: Trial Setup

in one direction throughout the trial, and a camera that traverses the environment, looking at various parts of the scene, respectively.

As can be seen in Figure 3.5, just two sphere placements are used for the dynamic scene. Mirror-flipping of the image is used to avoid participants detecting this re-use(as described in the previous section). In [BMSG09], an intensity level of 0.1 was deemed just noticeable enough for a viewer's periphery. Thus two trials consisting of this intensity were included. However, since that paper utilized an eye tracker in its application of SGD, it could focus on ensuring that the modulated region would never be present in the foveal vision of participants. To this end, another intensity value was chosen for evaluating 3DSGD when the user's current target of vision cannot be discerned, ie. a lower intensity value. 0.07 was chosen as this value. This was because small pilot experiments showed 0.07 to be the ideal pay-off between retaining a notable effect on the image while being so subdued as to be potentially unnoticeable to the unaware viewer.

Participants were provided with an information sheet describing what they would be doing and what would be required of them in the experiment. The documents relating to the experiment can be seen in the appendix of this document. Following this, they were provided a questionnaire form. The info sheet and questionnaire can be seen

As can be seen from these documents, participants were required to rate the degree of distraction they felt they observed during the experiment, despite not being explicitly told what the nature of the expected distractions would be. This is a rather difficult aspect to accurately evaluate, and so we must rely on participants being unfamiliar with the nature of the 3DSGD to effectively ascertain how overt they find the effect. Users were left to formulate their own notion of what would be considered distracting in the image. They were, however, shown two small trial scenes to familiarize themselves with the environment and the glass spheres, and were informed that none of what they saw then would be considered distracting in the sense of what was being looked for in the experiment. The idea behind this procedure was to avoid participants becoming distracted by any aspects of the scene such as aliasing artifacts or close-ups of lower resolution textures.

Following each trial, the participants wrote their answers into the questionnaire. The results of the experiment are shown in the next chapter.

3.3 Eye Tracker Implementation

Although a significant motivation behind this project was the implementation of attention guidance techniques without the reliance on eye trackers, a particular piece of hardware presented an interesting opportunity to use the previously discussed work in new ways.

The Eyetribe eye-tracker [Eye14] is a modern piece of hardware that approaches eye-tracking with an eye towards ease of use, portability and commercial viability. This makes it an ideal candidate for a project such as this, where the user is intended to experience the subtle attention guidance in a familiar and comfortable environment(at home, basically).

Inclusion of the eye-tracker into the project came somewhat late. However, the relatively easy-to-use API for C++ meant it could be incorporated into the project quite quickly. [BMSG09] describes the analysis of rapid eye saccades to detect eye movement in the direction of the object of interest. It relies on the relatively low visual acuity of human peripheral vision. This characteristic is also taken advantage of in this part of the project. The Eyetribe API can be used to ascertain the target of the eye's gaze per frame. A quick calculation of the distance between the reported target and the pixel of

Figure 3.6: Eyetribe eye-tracker[Eye14]

importance can be used to adjust the intensity of the image modulation as necessary. The result is a more subtle form of the gaze direction presented in this project, though at the cost of some degree of intrusiveness with regards to the physical requirements of the hardware.

A video displaying the use of the eye tracker, where the mouse cursor represents the target of the eyes, can be found on the disc provided with this document. The eye tracker was not included in the experiment described in this project since the necessary collaboration required to prepare each participant was deemed to be too time consuming to be implemented satisfactory.

Chapter 4

Results and Evaluation

The following chapter details the results gained from the experiment, as well as an evaluation of said results. 9 participants took part, 7 male and 2 female. The results are discussed in section 4.3. The static trials and the dynamic trials of the experiment are evaluated separately.

4.1 Search Task Performance

4.1.1 Search Task Performance in Static Trials

Figure 4.1 displays the average results of the the first two trials, where the camera remained static throughout the trial.

As can be seen, there is a slight discrepancy when it comes to static scenes. The static scenes consist of 14 glass spheres in total, and it is unclear whether participants performed better with or without the presence of image modulation. The most likely reason for this issue was an insufficient amount of training users with initial trial-runs.

As mentioned previously, participants ran through two initial trial-runs that contained spheres with no attention guidance cues, for the purposes of familiarizing themselves

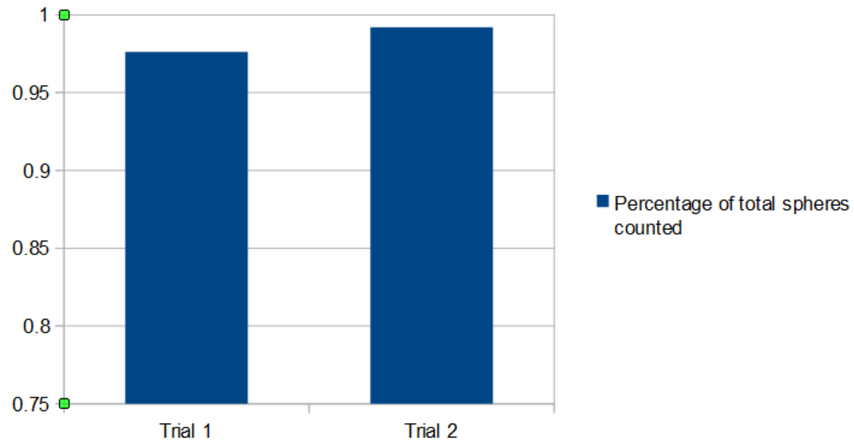

Figure 4.1: Average performance per static trial

with the scene and their objective, while allowing them to remain unaware of the attention guidance cues. In retrospect, ideally, more of these training trials would have been performed prior to beginning the trial run. Also, the speed at which spheres fade in and out of the scene may have caught participants unaware.

4.1.2 Search Task Performance in Dynamic Trials

With the previous section in mind, the search task performance of users for the remaining 8 scenes is notably more promising. There is a definite improvement in search task performance in trials that contained attention guidance cues compared to scenes that do not. Figure 4.2 shows the average search task performance per trial, while Figure 4.3 shows the average search task performance per image modulation technique.

Trials 5 and 9 were the trials in which no attention guidance cues were present. Trial 5 consisted of the scene with the 25 sphere placement, while trial 9 consisted of the scene with the 20 sphere placement. Initial observations reveal that certain participants performed notably better for Trial 4 than Trial 5, where Trial 4 was the 20 sphere placement with a relatively moderate(0.07) level of intensity and White-Black color modulation. However, Trial 9 - at which point any potential cause for concern due to necessary training(as mentioned in the previous section) would be nullified -

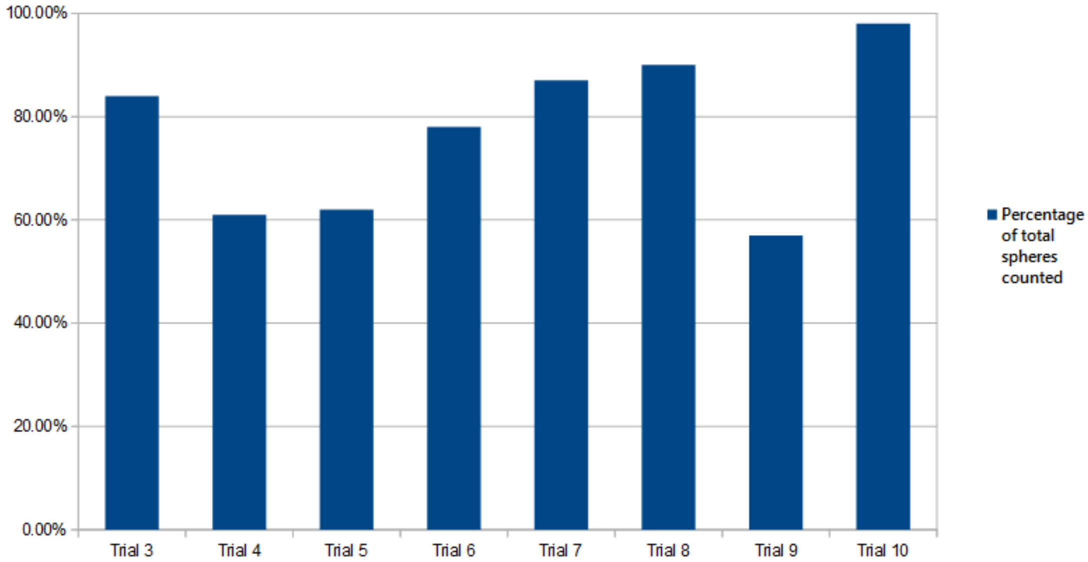

Figure 4.2: Average performance per dynamic trial

demonstrated a significant drop in search task performance in the absence of attention guidance cues.

Trials 3, 4, 5, and 7 consisted of moderate(0.07) intensity levels. Trials 3 and 7 applied warm-cool color modulation(ie. red and blue), while Trials 4 and 6 applied white-black modulation. The intention here was to evaluate and compare the performance of the two distinct color choices. From Figure 4.3 it appears that warm-cool modulation has a distinct advantage over white-black modulation, at least in terms of search task performance. It must be noted, however, that this could certainly be due to the nature and overall color tone of the environment. The mostly dull grey color tones could contribute to the warm-cool modulation being the most effective in drawing user attention in this particular scene.

Trials 8 and 10 consisted of relatively heavy (0.1) intensity with warm-cool modulation and certainly resulted in the most effective performance in the search task. This is notable in that there is still a distinct improvement in performance when the intensity level is increased from 0.1 to 0.07. However, pushing the intensity any higher naturally results in an attention guidance effect that could not feasibly be considered subtle. As a result, using higher intensity values is considered unnecessary and 0.1 is considered

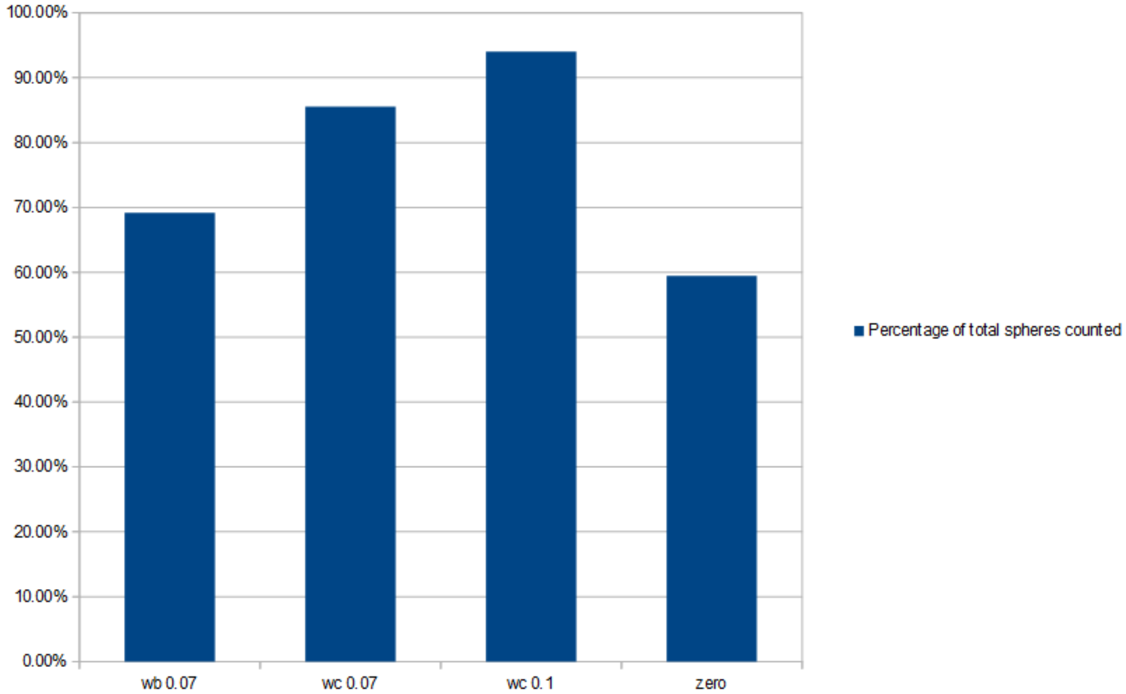

Figure 4.3: Average performance in dynamic scenes per image modulation technique

the ceiling of feasibly subtle attention guidance.

Certainly, the degree to which the intensity level would need to be increased is likely quite low, considering the success rate of participants in the two trials with 0.1 intensity. Trial 8 reports an average sphere count rate of 90.2%. Trial 10 reports an average sphere count rate of 96.67%, with 7 out of 9 participants successfully counting all spheres in the scene.

4.2 Level of Distraction

As described in the previous section, the experiment produces promising results for the actual effectiveness of the 3DSGD method for attention guidance in search tasks in dynamic scenes. However,when it came to the evaluation of user distraction during the trials, the results were not quite as clear-cut.

4.2.1 Static Scene Distraction

Figure 4.4 shows the average Level of Distraction reported per static trial (recall that level of disturbance was measured on a scale from 1 to 5, where 1 is completely non-distracting, and 5 is extremely distracting).

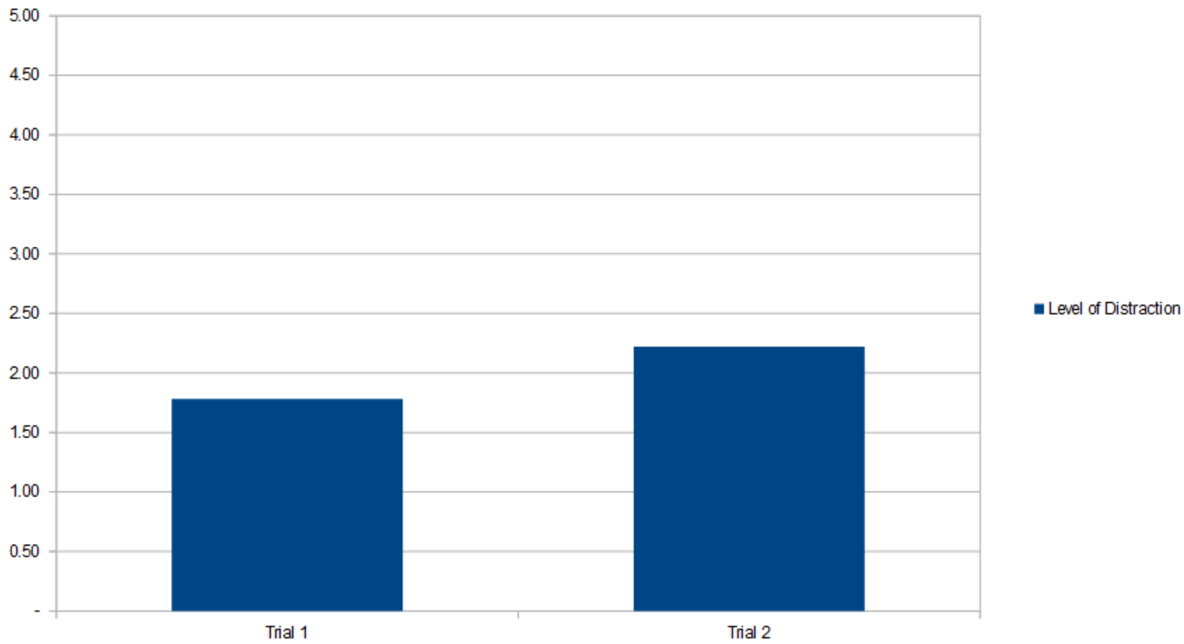

Figure 4.4: Average Level of Distraction reported per static trial

Though the results from the search task performance are quite demonstrative of the effectiveness of the 3DSGD as an attention guidance method, the results for the level of distraction are rather ambiguous as to what participants found distracting. An immediate observation to consider is that participants, on average, found Trial 1 to contain more distractions than Trial 2. This, of course, raises some questions. Trial 1 consisted of warm-cool color modulation with moderate intensity level, while Trial 2 consisted of no attention guidance at all.

As with the search task performance of the first two trials, the issue of insufficient training is potentially re-occurring here. The first trial contains attention guidance cues, and the second trial does not. The motivation behind this was to avoid users becoming too aware of the nature of the 3DSGD too quickly in the experiment by providing them a blatant 'off'/'on' pair of trials to begin with. By beginning the experiment with a moderate intensity trial, then following up with a non-affected trial, the intention was to ensure participants would remain unaware of the exact nature of the 3DSGD, at least until the final few trials, where the intensity level was increased to 0.1.

4.2.2 Warm-Cool/White-Black Distraction compared

When it comes to the average Level of Distraction for the remaining trials(the dynamic trials), there is still an issue of uncertainty as regards what participants found distracting. Let's first consider how the warm-cool and white-black modulations compared to one another. Figure 4.5 shows the average level of distraction reported for each sphere placement per color modulation type for the trials with moderate intensity.

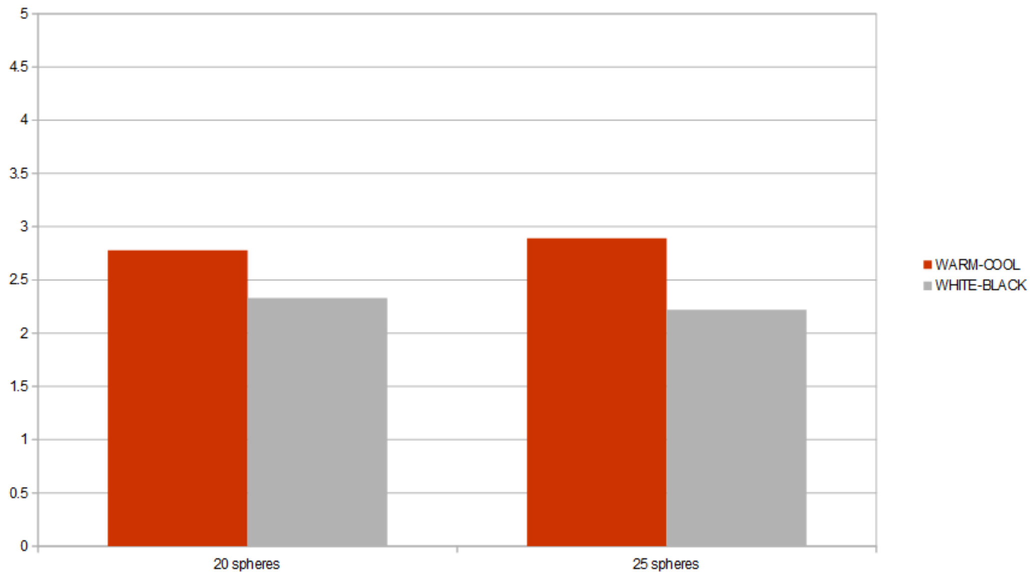

Figure 4.5: Average Level of Distraction reported per color modulation type in moderate intensity trials

First of all, the average of the two warm-cool trials of moderate intensity, Trial 3 and Trial 7, are 2.78 and 2.89 respectively. The Level of Distraction can range from 1 to 5, and 0.07 is what is considered the moderate intensity of this experiment. These results could be considered ideal in the sense that they represent what participants considered to be the approximate moderate Level of Distraction. In fact, these results appear to compare only slightly less favourably to the white-black trials, which report a slightly lower level of distraction on average.Comparing the average of the warm-cool 0.07 search task performance to the average of white-black 0.07 search task performance yields an improvement of 123.99% from white-black to warm-cool search task performance. Comparing the average of the warm-cool 0.07 level of distraction reports to the average of the white-black 0.07 level of distraction reports yield an increase of 124.6% from white-black to warm-cool.

These increases indicate white-black as a potentially more favourable color modulation, since the increase in search task performance with warm-cool corresponds to a smaller increase in the reported Level of Distraction. It must be noted, however, that at standard deviation of 0.115 and 0.145 for white-black and warm-cool respectively, there is always the potential for these observations of performance-per-distraction-level are entirely suitable for such a small pool of test cases.

4.2.3 All or nothing

Now let's consider the afore-mentioned moderate intensity trials in relation to the heavy(0.1) intensity and 0 intensity trials. As can be discerned from Figure 4.2, the search task performance of the various modulation techniques generally fall in line with expectations, with 0 intensity trials resulting in a considerable drop in search task performance, compared to the other three trials that contain attention guidance methods. The average level of distraction reported per dynamic trial can be seen in Figure 4.6.

The same issue that crops up in the two static trials appears to be evident in the dynamic trials as well. Trials 5 and 9 did not consist of any attention guidance, while Trials 8 and 10 consist of color modulation with heavy (0.1) intensity.

As can be seen from Figure 4.6, participants appeared to find the 0 intensity trials to have a total average Level of Distraction of 2.615. Meanwhile, trials with heavy intensity were reported to have an equal average Level of Distraction of 2.56. Obviously,

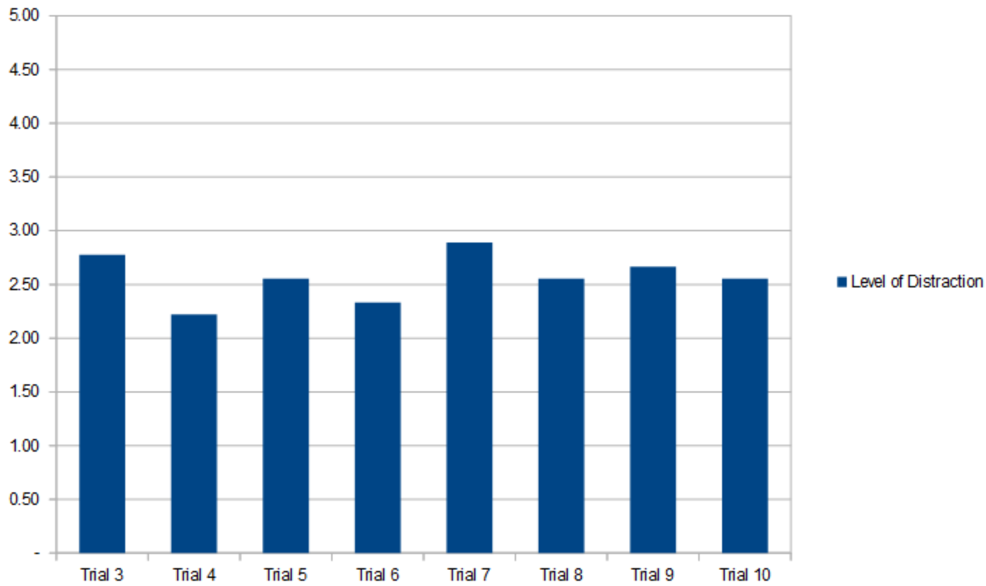

Figure 4.6: Average Level of Distraction reported per dynamic trial

these results do not appear clear in their indication of how distracting the respective trials should ideally be. For example, trials 5 and 9, considering their complete lack of any attention guidance image adjustment whatsoever, would ideally receive Level of Distraction reports of 1, and no more than 2 at the very most. Meanwhile, trials 8 and 10 would ideally receive Level of Distraction reports of 3 or more, indicating a notable-to-extreme level of distraction. As it so happens, participants appear to consider trials that contain no attention guidance to consist of more distracting visual elements than trials that contain the highest intensity of 3DSGD out of all trials.

It is difficult to analyze why participants reported the levels of distraction as they did. It must be noted that the average Level of Distraction reported for each trial was no higher than 0.11 away from that of the other trial with different sphere placements and identical intensity level/color modulation type. This does indicate uniformity in participants' interpretation of distraction. However, the high reported distraction for the 0 intensity trials and the low reported distraction for the heavy intensity trials still suggests issues in how participants perceived their goal.

Ultimately, one of the main reasons that may be suggested for the inconclusive results here is that participants were not sufficiently informed on the nature of the attention guidance that they would be subjected to. All the documentation participants were provided during the experiment with can be seen in the appendix. Participants were only made aware that "image processing techniques" would be in effect during the course of the experiment. Upon completing the two initial training trial runs, participants were informed that nothing seen within those two trials contained what was intended to be considered distracting. This was to avoid distractions inherent to the 3D scene being reported(aliasing artifacts, potentially low texture resolutions at close range, etc.). They were then informed that an interpretation of what would be considered distracting in the scene would be left up to them to formulate.

It seems possible, then, that this attempt to leave participants to define distracting visual elements in the scene for themselves may have been the cause of the inconclusive distraction results.

4.3 Image Saliency

The previous sections detail the results and analysis of the user experiments. We now focus on an evaluation of the actual method of attention guidance itself. Whereas the previous section considers the perceived aspects of the attention guidance method(ie. in terms of interpreting results obtained from the user experiments), this section examines the attention guidance method with regards to more empirical performance analysis (where the performance of the method here would refer to its overall effectiveness as a subtle attention guidance process).

4.3.1 Saliency Toolbox

Recall from chapter 2 that the salience of an image is the state or quality by which an object or feature of the image stands out relative to its neighbours. In the case of the 3DSGD method, the salience of the image modulation technique would be the image feature of concern.

The Saliency Toolbox [WK06], was developed as part of [Wal06], and is an effective tool that can be used to evaluate the saliency of a 2D image. The underlying architecture behind the construction of the saliency model is based on [IKN+98]. The Saliency Toolbox breaks down a 2D image and constructs a saliency model in a similar method to that seen in Figure 4.7. This tool can be used to analyze screencaps of the OpenGL application to detect salient features. The model is based upon early primate vision, and so can give us some concrete results on what would be considered conspicuous regions of an image as dictated by a system emulating human vision.

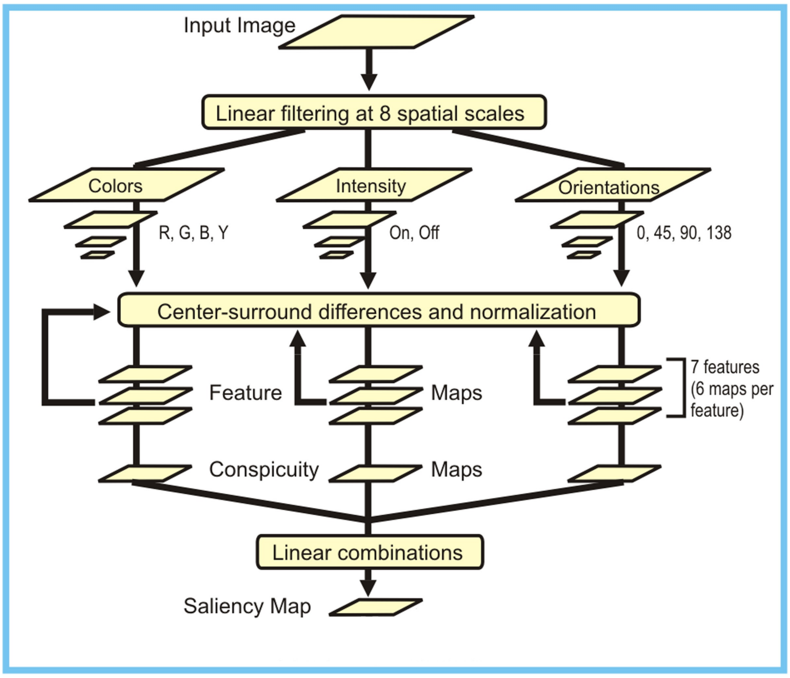

Figure 4.7: Saliency model architecture [sal14]

The Saliency Toolbox constructs a saliency map by first constructing three conspicuity maps for three separate image feature channels; color, intensity, and orientation. These three maps detect the strongest regions in which a feature in the image stands out from its neighbours, in terms of the features that channel responds to. For example, a part of an image where a light source creates a bright light in comparison to its surroundings would result in a strong response in that region of the image for the intensity conspicuity map.

Saliency Toolbox with heavy intensity

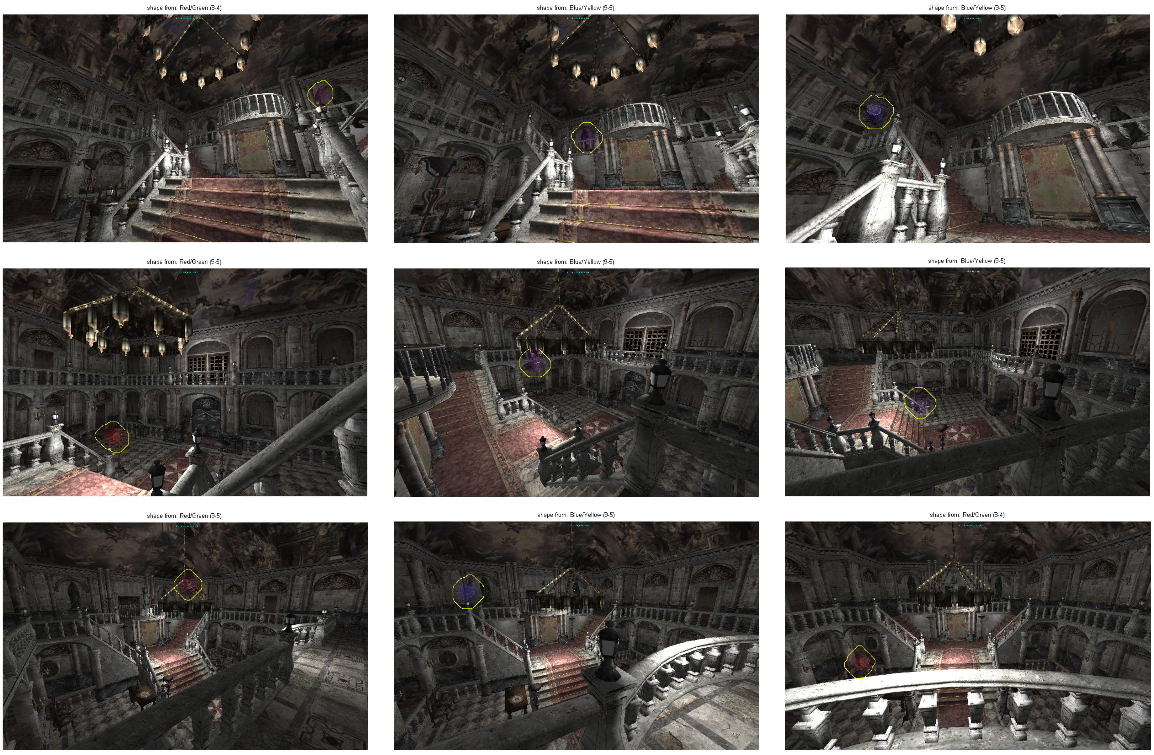

Figure 4.8: Saliency Toolbox Saliency map results for random screencaps of 3D scene w/ 0.2 intensity. The yellow outline denotes the most conspicuous region reported by the toolbox. The text above each image refers to the feature map with greatest impact on the overall saliency map. For example, the first image has its strongest features detected in the Red-Green feature map[IKN+98].

Images of the scene were captured at random intervals throughout the scene and given to the Saliency Toolbox to discern the most salient image feature. Figure 4.8 displays the results of the toolbox applied to 9 randomly chosen images from a large sample of screencaps of the application running, at a warm-cool intensity level of 0.2. The results of this analysis are rather promising; the region of interest(the modulated region) is considered the 'winner' in terms of saliency, for all 9 images, by the Saliency Toolbox.

Saliency Toolbox weighting adjustment

These results are quite promising, but 0.2 intensity is still too high for a feasible level of subtlety. Figure 4.9 displays the results of providing 9 screencaps from the 3D scene with 0.1 intensity.

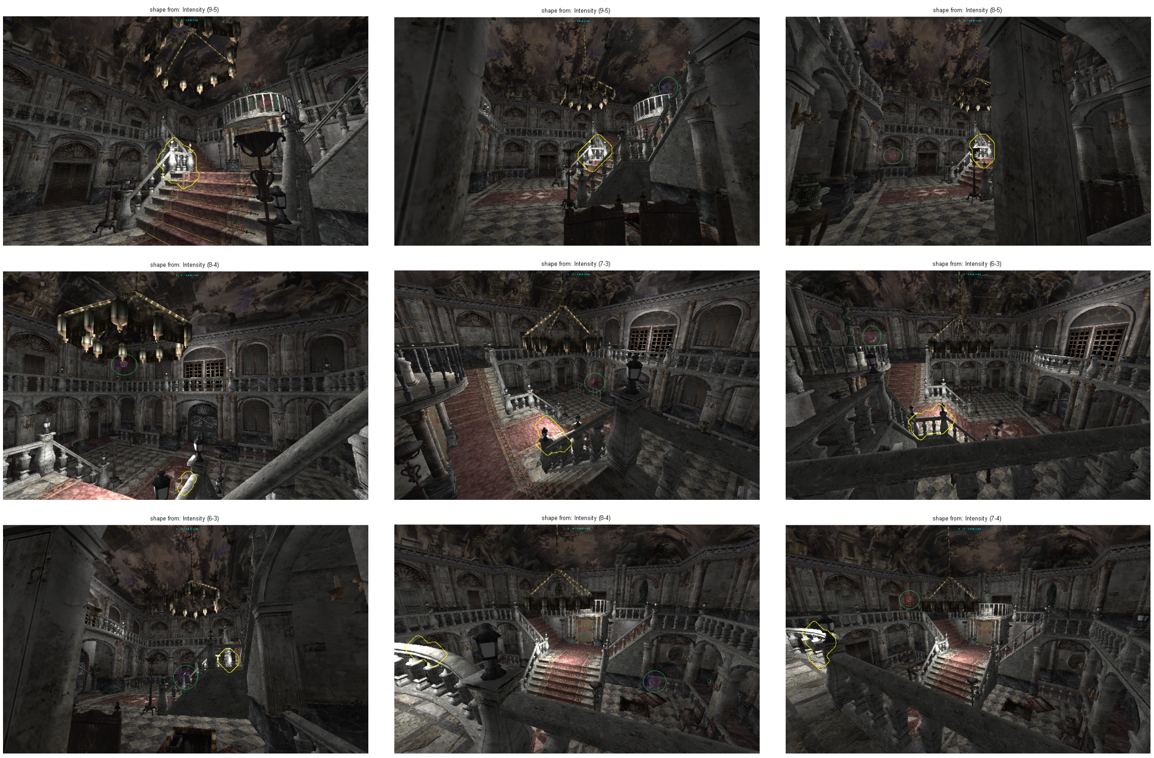

Figure 4.9: Saliency Toolbox Saliency map results for random screencaps of 3D scene w/ 0.1 intensity. The green circle denotes the location of the region of importance.

The results are considerably less promising compared to Figure 4.8, with 0% of the sample images having the region of importance be deemed the most salient feature at such a relatively low intensity. This ties into the findings of [PN^{+}03], seeing as the scene does consist of some areas of notably strong lighting. However, the Saliency Toolbox also allows a user to adjust the weightings of these three conspicuity maps in such a manner that it will consider the responses in a certain conspicuity map to be of greater relative importance when constructing the salience map. As we can see from the text above each image, the most salient regions detected in the images are a result of the most salient features being considered to lie in the intensity maps. By

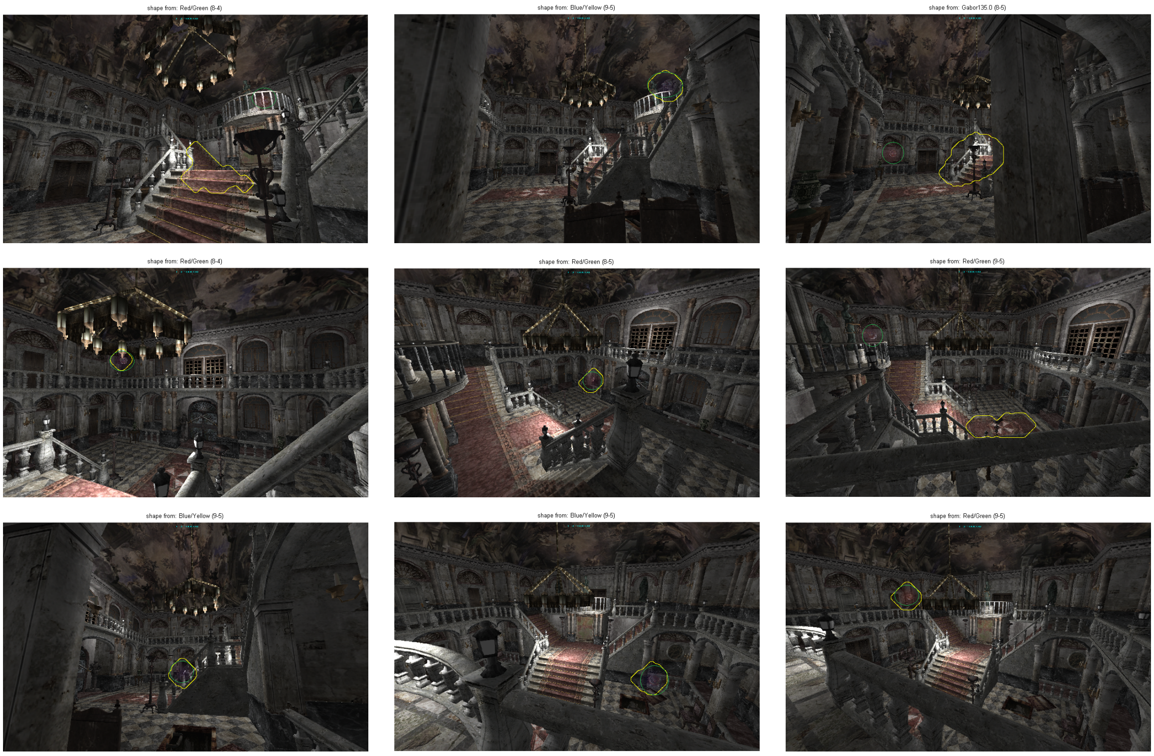

Figure 4.10: Saliency Toolbox Saliency map results for random screencaps of 3D scene w/ 0.1 intensity, with toolbox parameters adjusted to weight intensity maps less. Region of importance is notably more successful in its detection as the winner.

adjusting the Saliency Toolbox weighting so that intensity has a weight of 0 (while color and orientation maintain the default weighting of 1), a more promising set of results is acquired(Figure 4.10). The Saliency Toolbox successfully considers the region of importance to be the most salient feature in 6 out of the 9 sample images.

Meanwhile, one of the sample images(image 3) is considered to have its most salient feature produced by the orientation conspicuity map (the orientation feature maps are constructed using Gabor filters).

These results, due to the adjusted weighting of the Saliency Toolbox parameters, suggest the potential for inhibiting various features of the scene that result in heavy responses from the intensity and orientation channels. This is left as future work, however.

Chapter 5

Conclusion/Future Work

This dissertation presents a method of user attention guidance for use in 3D virtual environments.

The original motivation behind the project was to construct something that would be feasible for use in a videogame. Evaluating the effectiveness of such a system inherently comes with certain restrictions that complicate the evaluation process. When it came to user experiments, allowing interactivity on the part of the participants introduces various variables that would hamper the analysis of experiment results. As explained in chapter 3, the experiment needed to be conducted under the knowledge that participants would possess varying levels of familiarity with conventional first-person controls. Allowing users to navigate the environment with a gamepad, or even simply the arrow keys, would require an infeasible amount of training to ensure participants were equally familiar with such controls. Indeed, accurately ascertaining the familiarity participants may possess with any controls would itself necessitate extensive pre-experiment training. This was the underlying motivation behind the decision to implement the experiment as a series of non-interactive scenes where the camera navigates along a set path.

As for the actual experiment results themselves, the 3DSGD method presents promising results with regards to improvement of search task performance in dynamic 3D scenes where the camera is moving. When it came to these scenes, implementing the method with various different parameters always resulted in an improvement in search task performance compared to that of trials with no attention guidance implemented. Merely increasing the intensity parameter of the color modulation method resulted in an improvement in search task performance for the warm-cool color choice of the modulation method. However, the distinct improvement of the modulated trials over the non modulated trials lend credence to the hypotheses proposed by [MBG09] that the mere presence of the image modulation is more important than the intensity with which it is applied. Meanwhile, experiment results appear to suggest white-black color modulation to provide a possible improvement of ratio of search task performance to level of distraction, when compared to warm-cool. However, an analysis of how distraction levels scale with search task performance is necessary to validate this possibility. This could, of course, be related to the choice of environment, and evaluation of this attention guidance method within differing 3D environments is a definite step to take in future work.

A particular note to make as regards the experiment would be the ambiguity with which participants were provided when it came to formulating their own definitions for visual distraction. The intention here was to gauge a user's performance where they were almost completely unaware of what was happening to the image to guide them. Ultimately, the language used to gauge their level of distraction was likely too ambiguous to acquire concrete results on the reported level of distraction users reported. Future work would benefit from being more forward in its presentation of what is happening 'behind the curtain', so to speak. Training participants in familiarizing themselves with the nature of the color modulations would likely result in improved results for the distraction evaluation, at the cost of potentially more genuine results gained by user unfamiliarity with the attention guidance method.

The Eyetribe eye-tracker was a late addition to the project, but one that proved to be very suited to implementing with the OpenGL application. This form of eye-tracker, which is geared more towards ease of use, comfort and commercial viability - as opposed to absolute accuracy like other eye-trackers that require more complex setups and physical constraints from the user - seems ideally suited for further research in incorporating eye-tracking into a system such as that presented here. Initially, successful implementation of attention guidance methods without the need for eye-trackers was one of the main motivations behind eschewing the use of eye-trackers in the project. However, eye-tracking hardware such as this presents a compelling approach to non-intrusive eye-tracking. Prioritizing immersion into a 3D environment without distracting player guidance elements is a major aspect of the initial motivation, and this formed the reasoning behind the inclusion of this hardware into the project.

Ultimately, there are a multitude of ways the work presented here can be built upon. For example, the findings of [MBG09] argue that the presence of modulation distractions can result in improved search task performance owing to the increased pace at which eyes are motivated to traverse the screen per frame, and this technique could be incorporated into the current system. As mentioned above, evaluation of various environment types is a particularly important aspect to consider. Re-introducing user interactivity, allowing potential participants to navigate a 3D scene for themselves, is another. Indeed, the color modulation method [BMSG09] this work is based on appears to be quite suited to implementation in a manner in which the periphery of a screen is the only portion ever affected. Such an approach would seem ideally suited to an interactive environment, where getting a user to look where you want them to may be half the battle.

Appendix

The appendix consists of various documents relating to the experiment involved with this project. They are as follows:

- Research project proposal for the Ethics Committee

- Information sheet for participants undertaking the experiment

- Copy of questionnaire provided to participants to record their results.

GAZE DIRECTION IN 3D VIRTUAL SCENES

TRINITY COLLEGE DUBLIN

RESEARCH PROJECT PROPOSAL

TITLE OF PROJECT:

Gaze Direction in 3D Virtual Scenes

PURPOSE OF PROJECT:

The goal of this project is to investigate techniques of shading and illumination in 3D virtual scenes in order to subtly direct the viewer's gaze to particular objects within the scene. In particular, the project extends upon various image processing techniques that have been proven to effectively achieve the above intended effects in the context of 2D imagery with an eye tracker [Bailey09]. The goal of this project is to expand upon that work by implementing that work in the context of 3D scenes, without the use of an eye tracker. At the same time the goal is for the image modifications to be relatively subtle so as to not affect the flow or immersive-ness of a game or virtual experience.

Participants will be presented with an information sheet describing the specifics of the experiment they are partaking in. They will be placed in front of a computer monitor and instructed to follow the on-screen instructions, which will namely consist of locating objects in a 3D scene. The scene they will be shown is typical of 3D environments in current commodity 3D games. They will be informed in the information sheet that image processing techniques may or may not be in effect during their experiment, though they will not be informed on the exact nature of the image processing techniques.

The preliminary data gathered in this pilot experiment should allow us to determine how different techniques and different intensities of visual cues perform against each other and the baseline (no visual cues), with regard to drawing attention and relatively subtlety of the effects.

DESCRIPTION OF METHODS AND MEASUREMENTS:

All measurements taken from the experiments will be in the form of numbers and simple answers to questions that the subject will fill out in paper questionnaires.

The participant will be instructed to report the number of occurrences in the 3D scene of a target object (a glass sphere).

Participants will also be asked to answer on a 5 point Likert scale the degree to which they found certain elements in the scene (the image processing methods) noticeable (where 1 represents not noticeable, and 5 represents very noticeable).

PARTICIPANTS

Participants will be recruited via noticeboards and emails throughout the college and through personal acquaintances. A note acknowledging a potential conflict of interest will be included in the Information Sheet for Participants. If sending an email to a mailing list, the list owners consent will first be obtained. Beyond being over 18 years of age to provide consent to participate, there is no restriction placed on participants as regards age, gender, etc.

10 participants are expected for this study. This is the number of participants used in the original paper by Bailey et al [Bailey09], which we extend upon in this research. We expect that this number of participants should be enough for this pilot experiment which should provide us data for further development.

DEBRIEFING ARRANGMENTS:

Participants will be advised in the information sheet to raise any questions and concerns they have about the experiment at any point during their participation.

ETHICAL CONSIDERATIONS:

Participants are highly unlikely to experience any physical or mental discomfort or ill-effects however, as the experiment is conducted on a computer screen participants with epilepsy or a family history of epilepsy are advised in the information sheet to proceed at their own risk.

RELEVANT LEGISLATION TO THE PROJECT:

This study complies with the Data Protection Act.

REFERENCES:

[Bailey09] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. 2009. Subtle gaze direction. ACM Trans. Graph. 28, 4, Article 100 (September 2009), 14 pages. DOI=10.1145/1559755.1559757 http://doi.acm.org/10.1145/1559755.1559757

GAZE DIRECTION IN 3D VIRTUAL SCENES

TRINITY COLLEGE DUBLIN

INFORMATION SHEET FOR PARTICIPANTS

Name of Course:

MSc in Computer Science (Interactive Entertainment Technology)

Project this Experiment Belongs to:

Subtle Gaze Direction in 3D Virtual Scenes

- This experiment is designed to ascertain the effectiveness of an image-processing method for the purpose of guiding viewer attention and navigation in 3D virtual scenes, typical of what is found in modern computer games.
- As a participant, you are entitled to withdraw yourself from the experiment at any time without penalty should you see fit. You are also entitled to omit individual responses to questions without penalty.
- Participants are being recruited from amongst work colleagues and fellow students at TCD. We acknowledge that this may present a conflict of interest in this regard.
- The experiment in total is expected to take no longer than <u>20 minutes</u> including time for setting-up, etc.
- The experiment will consist of you looking at a graphical 3D virtual scene on a computer monitor. You will initially be met with a splash screen explaining what your goal in each trial-run is. The goal will be to locate and count all glass spheres you see appear in the scene. The spheres will appear sequentially at random points throughout the 3D environment. A maximum of one sphere will be on screen at any time, and <u>spheres will appear for no longer than a second at a time</u>. The spheres may be located close to or far away from the "camera", i.e. the point-of-view in the 3D scene of the camera, but will always be visible in some capacity.
- In addition some of the 3D scenes will feature some simple image processing techniques (dynamic changes in colour and brightness in parts of the scene) which may serve as visual distractors to draw your attention.
- **IMPORTANT NOTE:** Though the nature of the rendering and image processing is not likely to result in any risk to individuals, the experiment will be conducted on a computer screen. As a result, participants with a history of epilepsy are advised to proceed at their own risk.
- In the unlikely event that you experience any discomfort (e.g. motion sickness) whilst viewing the animated scenes, you should notify the researcher and the experiment will be stopped immediately.
- You will first take a series of short initial trial-runs to familiarize yourself with the 3D scene and your goal. You will not provide results for these.

- Following each trial-run of the experiment itself, you will be instructed to answer one or more questions related to what you saw in the 3D virtual scene. You will be provided an answer sheet to write your answers down on. You will have been instructed to take these questions into account before the experiment begins and an intro message will remind you of the goal in each scene. You proceed to the next trial-run by following on-screen prompts.

- All participant answers will be recorded anonymously and results will only be analysed, publicised, or presented as collective aggregations of all participant results.

- Should you have any questions or concerns following completion of the experiments, do not hesitate to ask for a debriefing/further information after participation.

GAZE DIRECTION IN 3D VIRTUAL SCENES

TRINITY COLLEGE DUBLIN

TRIAL-RUN QUESTIONNARE

Below are questions for each run-through of the experiment. Please enter your answers to the questions. Please be honest with your answers.

NOTE: Each question is optional. Feel free to omit a response to any question; however the researcher would be grateful if all questions are responded to.

For each trial-run, after completing the run, please fill in the number of glass spheres you counted in the scene and then tick the box of the number below that best describes how noticeable you found any visual modifications:

1 – Non-existent

2 – Very slight

3 - Moderate

4 - Strong

5 – Very strong

Trial-Run Number	Number of Glass Spheres counted	Level of Distraction				
		1	2	3	4	5
1	____________________	❒	❒	❒	❒	❒
2	____________________	❒	❒	❒	❒	❒
3	____________________	❒	❒	❒	❒	❒
4	____________________	❒	❒	❒	❒	❒
5	____________________	❒	❒	❒	❒	❒
6	____________________	❒	❒	❒	❒	❒
7	____________________	❒	❒	❒	❒	❒
8	____________________	❒	❒	❒	❒	❒
9	____________________	❒	❒	❒	❒	❒
10	____________________	❒	❒	❒	❒	❒

Bibliography

[Bet12] *Dishonored.* Bethesda Softworks [Optical disc, digital distribution], 2012

[BMSG09] Bailey, Reynold ; McNamara, Ann ; Sudarsanam, Nisha ; Grimm, Cindy: Subtle gaze direction. In: *ACM Transactions on Graphics (TOG)* 28 (2009), Nr. 4, S. 100

[Cap07] *Resident Evil: Umbrella Chronicles.* Capcom [Nintendo Wii, Sony Playstation 3], 2007

[CDF+06] Cole, Forrester ; DeCarlo, Douglas ; Finkelstein, Adam ; Kin, Kenrick ; Morley, Keith ; Santella, Anthony: Directing gaze in 3D models with stylized focus. In: *Proceedings of the 17th Eurographics conference on Rendering Techniques* Eurographics Association, 2006, S. 377–387

[DS02] DeCarlo, Doug ; Santella, Anthony: Stylization and abstraction of photographs. In: *ACM Transactions on Graphics (TOG)* Bd. 21 ACM, 2002, S. 769–776

[e1m] *Classic Doom, Doom Maps.* Copyright ©2006. `http://www.classicdoom.com/doommaps.htm`

[EA07] *Crysis.* Electronic Arts [Optical disc, digital distribution], 2007

[Eye14] Eyetribe: *The Eye Tribe Developer Site. N. p.* `http://dev.theeyetribe.com/general`, August 2014

[GKG98] Gottlieb, Jacqueline P. ; Kusunoki, Makoto ; Goldberg, Michael E.:

The representation of visual salience in monkey parietal cortex. In: *Nature* 391 (1998), Nr. 6666, S. 481–484

[hou] *The Free 3D Models, Homepage.* Copyright tf3dm.com ©2014. `http://tf3dm.com`

[Id93] *Doom.* Id Software [Floppy disks, CD-ROM, ROM cartridge, digital distribution], 1993

[IK99] Itti, Laurent ; Koch, Christof: A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems. In: *Journal of Electronic Imaging* 10 (1999), S. 161–169

[IKN+98] Itti, Laurent ; Koch, Christof ; Niebur, Ernst u. a.: A model of saliency-based visual attention for rapid scene analysis. In: *IEEE Transactions on pattern analysis and machine intelligence* 20 (1998), Nr. 11, S. 1254–1259

[MBG09] McNamara, Ann ; Bailey, Reynold ; Grimm, Cindy: Search task performance using subtle gaze direction with the presence of distractions. In: *ACM Transactions on Applied Perception (TAP)* 6 (2009), Nr. 3, S. 17

[MR98] Mack, Arien ; Rock, Irvin: *Inattentional blindness.* The MIT Press, 1998

[ND13] *The Last of Us.* Sony Computer Entertainment [Optical disc, digital distribution], 2013

[NI05] Navalpakkam, Vidhya ; Itti, Laurent: Modeling the influence of task on attention. In: *Vision research* 45 (2005), Nr. 2, S. 205–231

[OM] Ogden, T. E. ; Miller, R. F.: Studies of the optic nerve of the rhesus monkey: Nervefiber spectrum and physiological properties.

[PN+03] Parkhurst, Derrick J. ; Niebur, Ernst u. a.: Scene content selected by active vision. In: *Spatial vision* 16 (2003), Nr. 2, S. 125–154

[RP92] Robinson, David L. ; Petersen, Steven E.: The pulvinar and visual salience. In: *Trends in neurosciences* 15 (1992), Nr. 4, S. 127–132

[sal] *Model of bottom-up attention, The theory.* Copyright ©2000 by the

University of Southern California, iLab and Prof. Laurent Itti. http://ilab.usc.edu/bu/theory/index.html

[sal14] *Monetary Bay Aquarium Research Institute, Flow diagram of a typical model for the control of bottom-up attention.* Copyright ©1996-2014 MBARI. http://dev.theeyetribe.com/general. Version: Februar 2014

[SD02] Santella, Anthony ; DeCarlo, Doug: Abstracted painterly renderings using eye-tracking data. In: *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering* ACM, 2002, S. 75–ff

[SqE11] *Deus Ex: Human Revolution.* Square Enix [Optical disc, digital distribution], 2011

[SSWR08] Sundstedt, Veronica ; Stavrakis, Efstathios ; Wimmer, Michael ; Reinhard, Erik: A psychophysical study of fixation behavior in a computer game. In: *Proceedings of the 5th symposium on Applied perception in graphics and visualization* ACM, 2008, S. 43–50

[Ubi13] *Assassin's Creed IV: Black Flag.* Ubisoft [Optical disc,digital distribution], 2013

[Val04] *Half-Life 2.* Valve Software [Optical disc, digital distribution], 2004

[VaP04] *Portal.* Valve Software [Optical disc, digital distribution], 2004

[Wal06] Walther, Dirk: *Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics*, California Institute of Technology, Diss., 2006

[WK06] Walther, Dirk ; Koch, Christof: Modeling attention to salient proto-objects. In: *Neural networks* 19 (2006), Nr. 9, S. 1395–1407

[WOG06] Winnemöller, Holger ; Olsen, Sven C. ; Gooch, Bruce: Real-time video abstraction. In: *ACM Transactions On Graphics (TOG)* 25 (2006), Nr. 3, S. 1221–1226

[Wol94] WOLFE, Jeremy M.: Guided search 2.0 a revised model of visual search. In: *Psychonomic bulletin & review* 1 (1994), Nr. 2, S. 202–238

[You87] YOUNG, Richard A.: The Gaussian derivative model for spatial vision: I. Retinal mechanisms. In: *Spatial vision* 2 (1987), Nr. 4, S. 273–293

[ZGSS12] ZETZSCHE, Christoph ; GERKENSMEYER, Torben ; SCHMID, Falko ; SCHILL, Kerstin: Sensorimotor Representation of Space: Application in Autonomous Systems and in a Wayfinding Assistant for Alzheimer's Disease. In: *Computer and Information Science (ICIS), 2012 IEEE/ACIS 11th International Conference on* IEEE, 2012, S. 219–224