Top-Down Visual Attention for Efficient Rendering of Task Related Scenes

Veronica Sundstedt, Alan Chalmers, Kirsten Cater and Kurt Debattista

Department of Computer Science
Computer Graphics Group
Woodland Road, BS8 1UB Bristol, UK
Email: {veronica,alan,cater,debattis}@cs.bris.ac.uk

Abstract

The perception of a virtual environment depends on the user and the task the user is currently performing in that environment. Models of the human visual system can thus be exploited to significantly reduce computational time when rendering high fidelity images, without compromising the perceived visual quality. This paper considers how an image can be selectively rendered when a user is performing a visual task in an environment. In particular, we investigate to what level viewers fail to notice degradations in image quality, between non-task related areas and task related areas, when quality parameters such as image resolution, edge anti-aliasing and reflection and shadows are altered.

Figure 1: Effects of a task on eye movements. Eye scans for observers examined with different task instructions: (left) Free viewing, (right) Count the number of loose stones.

1 Introduction

A major challenge in computer graphics is achieving high fidelity computer graphics imagery at real-time frame rates. Despite the availability of modern high performance graphics hardware, the increased complexity of the scenes being considered and the quality required means that they can simply not be computed in a reasonable time, let alone real-time. A key goal of perception-guided rendering techniques is to save significant computational time without compromising the resulting image quality as perceived by a human observer. Most computer graphics applications, in particular in virtual reality, computer games and the animation development process, contain some specific visual task and in the majority of cases objects relevant to the task can be identified in advance, for example looking for street signs in a driving simulation. To perform this task, the human visual system focuses its attention on these objects at the expense of other details in the scene. We can exploit such knowledge of the human visual system to save significant rendering time by selectively rendering, that is, simply not computing to a high quality those parts of a scene that the human will fail to notice [1].

In this paper we show the results of visual perception experiments to determine the level of quality needed in a task-related scene. The experiments aim to give empirical evidence about the possible bounds of directed attention in tasks that needs to be completed in a specific time. We investigate whether it is possible to render scene areas not related to the task with lower resolution and decreased edge anti-aliasing quality without the viewer noticing any perceptual difference. We also present a novel approach for creating and using *task-importance maps* (TIM) in selective rendering to control image detail.

2 Previous work

In order to determine the parts of an image that a human is likely to attend to, we need to understand the operation of the human visual system. In recent years, considerable efforts have been devoted to understanding the mechanisms driving visual attention [9, 16, 24, 27, 28]. When considering a complex environment, the retina of our eye converts the information about the scene from light waves into neural signals which the brain can process [4]. The center of the retina, the fovea, consists exclusively of densely packed color-sensitive cones and this part of the eye provides the highest spatial and chromatic resolutions than anywhere else. Outside this area, our peripheral vision has far less visual acuity. The visual angle covered by the fovea is only approximately 2°, about the size of eight letters on a typical page of text or the size of a thumbnail held at arm's length. To obtain detailed information from other parts of the scene, the eye has to be redirected so that the relevant parts of the environment fall on the fovea. The eye does not scan the scene in a raster-like fashion, but rather jumps so that these relevant objects fall sequentially on the fovea. These jumps are called *saccades*.

Visual attention is a coordinated action involving conscious and unconscious processes in the brain, which allow us to find and focus on relevant information quickly and efficiently. It is often compared to a spotlight, which enhances information within a selected region of the viewed image while suppressing information in the remaining regions [23, 35]. There are two general visual attention processes, labeled *bottom-up* and *top-down*, which determine where humans locate their visual attention [15]. In bottom-up processing the visual stimulus captures attention automatically without volitional control, for example a red apple in a green tree, a lamp in a dark room or the eyes and lips of a human face. Low level, bottom-up features which have been found to influence visual attention include contrast, size, shape, color, brightness, orientation, line ends and motion. Of these, motion has been found to be one of the strongest [16]. In contrast, the top-down process is under voluntary control, and is related to the observer's goal when studying a scene. Such goals may include searching for a street sign or counting the number of objects in a scene, as shown in figure 1. In 1967, Yarbus showed that there is a strong correlation between a viewer's eye movements and the visual task that he/she has been asked to perform [28].

Knowledge of the human visual system is being increasingly used in computer graphics to improve the perceptual quality of the rendered images, for example [5, 6, 17, 19, 20, 22, 25, 33]. In addition, other research is investigating how details in our peripheral vision can be significantly reduced without any reduction in the viewer's perception of the quality of the resultant images. Watson et al. exploited peripheral vision in a peripherally degraded display, rendering those areas outside the viewer's foveal visual angle at a lower quality [30]. Work by McConkie and Loschky showed that photographic images filtered with a window radius of 4.1° produced results statistically indistinguishable from that of a full high-resolution display [10, 11, 12]. However, they also showed that the viewer will notice the difference in image quality if the area of fixation was not displayed to the highest quality within 5 milliseconds of the fixation.

Figure 2: Saliency maps and VDP images: (top left) Saliency map of the Kalabsha temple, (top right) Saliency map of the red balls scene, (bottom left) VDP image between HQ and LQ of the Kalabsha temple, (bottom right) VDP image between HQ and LQ of the red balls scene.

Koch and Ullman [9] introduced the idea of a *saliency map* (SM) to accomplish preattentive selection, figure 2. This is an explicit two-dimensional map that encodes the saliency of objects in the visual environment. Competition among neurons in this map gives rise to a single winning location that corresponds to the most salient object, which constitutes the next target for the eyes. If this location is subsequently inhibited, the system automatically shifts to the next most salient location, endowing the search process with internal dynamics. Itti and Koch [7, 8] developed a saliency-based computer model, and used it to predict the saliency of selected objects in a scene. Yee [29] took this work further and used a visual attention model to improve the efficiency of indirect lighting computations in the *Radiance* system [26] for dynamic environments. A spatiotemporal error tolerance map, which Yee calls the Aleph map, is constructed from spatiotemporal contrast sensitivity and a low-level saliency map for each frame in an animation. The Aleph map is then used as a guide to indicate where more rendering effort should be spent in computing the lighting solution, significantly improving the computational efficiency during animation. Subsequent work by Marmitt and Duchowski [21] showed, however, that such bottom-up visual attention models do not always predict attention regions in a reliable manner. Their work also showed that any attentional model used in perception-guided rendering should be augmented with a memory of previously seen image regions or perhaps previously inspected objects. Most perception-driven rendering algorithms have so far exploited only peripheral vision and the bottom-up visual attention process. Recent work in perception-based guidance of global illumination solutions has, however, considered volition-controlled and task-dependent attention, to produce so-called *task maps*, which is especially important for interactive applications [1, 14]. Randall et al. has also proposed the idea of using a task map to guide virtual humans in a complex environment by making some perceptual objects more salient than others [32].

Cater et al. showed that conspicuous objects in a scene that would normally attract the viewer's attention in an animation are ignored if they are not relevant to the task at hand [2]. This failure of the human to see unattended items in a scene, is known as *inattentional blindness* [18, 34]. In their experiments, viewers were presented with two animations. One was a full, high-quality rendering, while in the other, only the pixels in visual angle of the fovea (2°) centered around the location of a task object within the environment were rendered at high quality. The high quality was then blended to a much lower quality for the rest of the image. This work showed that when observers were performing a task within an animation, their visual attention was fixed exclusively on the area of the task, and they consistently failed to notice the significant difference in rendering quality between the two animations.

3 Psychophysical study of task areas

In this paper we extend the work by Cater et al. and investigate whether people would indeed even fail to notice parts of a scene within the visual angle of fovea rendered in low quality in the presence of a high-level task focus, i.e. if it is the visual angle of the fovea that is important or only the objects related to the task. The importance of this result is that the portion of an image covered by the foveal angle is related to the distance the viewer is from the screen and this area may contain many more objects than those related to the task. Our hypothesis was that viewers would not notice visible degradations in an image that did not affect the clarity of the task related objects. If true, our hypothesis would show that only the quality of the task objects were important when exploiting inattentional blindness in selective rendering, and not the quality of all of the scene covered by the foveal angle. For this experiment only still images were used, however, previous work has proven that this method also works for animations [2].

The study involved two rendered environments, figure 5. The task chosen was to count the number of loose stones in the temple of Kalabsha scene and the number of red balls in the geometrical object scene. To test our hypothesis we decided to conduct two experiments, one of which considered altering the sampling resolution of the images and the other with different edge anti-aliasing qualities. A pre-study was run with 10 participants to discover how long subjects took to perform the task. This was found to be on average 2.75 seconds to count the 7 stones in the Kalabsha temple image and 4 seconds to count the 10 red balls in the geometrical scene. Two pilot studies were then conducted to determine an appropriate image resolution and edge anti-aliasing quality.

3.1 Pilot studies

In the first pilot study 10 participants were shown 24 pairs of images at random, and asked if they could distinguish a change in quality for the scene with stones. Figure 2, (bottom left), shows the perceptual difference between high and low image resolution quality for the Kalabsha temple scene computed using Daly's Visual Difference Predictor (VDP) which is an approach to predicting perceived differences between a pair of images [13, 31]. Each image was displayed for 2.75 seconds, the value found in the pre-study. One image was always the high quality image rendered at a 3072×3072 sampling resolution, whilst the other image was one selected from images rendered at sampling resolutions of 256×256, 512×512, 768×768, 1024×1024, .., 2816×2816. In half of the pairs of images, there was no change in resolution; i.e. they saw two 3072×3072 resolution images. The pilot study showed that 100% of the participants could detect a difference in resolution at 512×512 and lower.

The second pilot study was run to determine at what level the viewer could always notice an edge anti-aliasing quality difference in the red balls scene. In this case the scene was rendered in Alias' Maya with resolution 1024×1024. Figure 2, (bottom right), shows the perceptual difference between highest and low edge anti-aliasing quality for the red balls scene computed using Daly's VDP. 10 participants were shown 12 pairs of images at random, and asked if they could distinguish a change in quality for the scene with the red balls. Each image was displayed for 4 seconds which was the value found in the pre-study. One image was always the high quality image rendered with edge anti-aliasing set to highest quality, whilst the other image was one selected from images rendered with edge anti-aliasing low, medium or high quality. In half of the pairs of images, there was no change in edge anti-aliasing; i.e. they saw two images with edge anti-aliasing at the highest quality. The pilot study showed that 100% of the participants could detect a difference only in the images with low quality edge anti-aliasing, figure 6, when displayed for 4 seconds.

We also decided to investigate a third case, when the red balls scene was rendered completely without shadows and reflections, to see whether the observers could see this difference while counting the red balls.

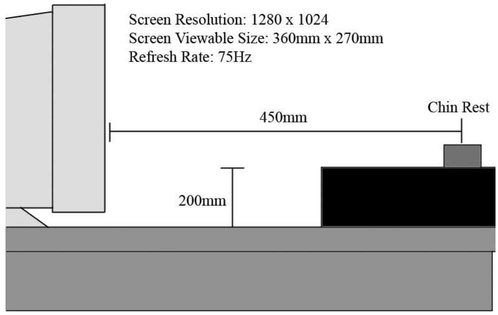

Figure 3: Experimental setup.

3.2 Task-Importance Maps

In order to selectively render task related scenes we use a *task map* (TM) that specify what parts of a scene that are perceptually important for performing the task. Previous definitions of task maps have either been manual [1] or done by predicting non-diffuse objects as salient [14]. In our system, modellers are able to select task specific parts of the scene during the modelling process. Prior to rendering, the system allocates a perceptual weighting to all parts of the scene based, not only on this selection by the user, but also on the saliency of the scene and the similarity of other objects in the scene to the task object in terms of color and shape. Furthermore, indirect task areas, including reflection and shadows are identified using a simple environment map. The output from this pre-rendering system is a series of maps at key-frame points on the motion path, including a *saliency map* (SM) and task map. These maps are then combined into the resultant *task-importance map* (TIM) using a weighted average with possibilities to control the weighting between both maps. Each pixel is now selectively rendered according to this combined map. Bright areas in the map will result in preferential rendering of the pixels by shooting a variable number of rays depending on the pixel value. For our implementation we extend *Radiance* to accept TIM as input, together with the maximum possible number of rays to be shot per pixel. The reference lighting solution of a complex environment, such as the Kalabsha temple, took 14.5 minutes to compute, whereas the perceptually-based solution for the SM, TIM and TM took 4.5, 3 and 2 minutes respectively (the same

Acronym	Description
HQ	High Quality: Entire image rendered at a sampling resolution of 3072×3072 in *Radiance* with a high quality setting.
LQ	Low Quality: Entire image rendered at a sampling resolution of 512 (found in the threshold experiment) in *Radiance* with a high quality setting.
SQ	Selective Quality: A sampling resolution of 512 (found in the threshold experiment) all over the image apart from in the visual angle of the fovea (2°) centered around the stone, which are rendered to a sampling rate of 3072×3072.
TQ	Task Quality: A sampling resolution of 512 (found in the threshold experiment) all over the image apart from the objects related to the task, which are rendered to a sampling rate of 3072×3072.
HQ	High Quality: Entire image rendered in Alias' Maya with edge anti-aliasing at highest quality.
LQ	Low Quality: Entire image rendered in Alias' Maya with edge anti-aliasing quality set to low.
SQ	Selective Quality: An edge anti-aliasing quality of low quality (found in the threshold experiment) all over the image apart from in the visual angle of the fovea (2°), which have edge anti-aliasing quality set to highest.
TQ	Task Quality: An edge anti-aliasing quality of low quality (found in the threshold experiment) all over the image apart from the objects related to the task, which have edge anti-aliasing set to highest.
RQ	Reflection Quality: Entire image rendered in Alias' Maya without shadows and reflections.

Table 1: Image pairs shown in the experiment: (top) Kalabsha temple scene (bottom) Red balls scene (1) HQ/HQ, (2) HQ/LQ, (3) LQ/HQ, (4) HQ/SQ, (5) SQ/HQ, (6) HQ/TQ, (7) TQ/HQ, (8) HQ/RQ, (9) RQ/HQ.

weighting was used for the SM and the TM to compute the TIM rendering). The same pre-computed irradiance cache was used for each rendering and the maximum rays per pixel value was set to nine.

3.3 Main experiment

For the main experiment only task maps were used and the bottom-up approach not taken into account. In the actual study, a total of 160 participants were considered. Each subject saw two images, each displayed for the time found in the pre-study. Table 1 describes the conditions tested in the experiment. To minimize experimental bias, 8 trials were run in the morning and 8 in the afternoon for each condition, altering the sex of the participants. Subjects had a variety of experience with computer graphics, and all exhibited normal to corrected to normal vision in testing. Before beginning the experiment, the subjects read a sheet of instructions on the procedure of the particular task they were to perform. After each participant had read the instructions, they were asked to clarify that they understood the task. They then placed their head on a chin rest that was located 45cm away from a 17-inch monitor, figure 3. The chin rest was located so that their eye level was approximately level with the center of the screen. The first image was then displayed and the participant stated out loud how many stones/red balls they saw. Following this, the second image was displayed, during which the task was repeated. There were the same number of stones/red balls in both experiment images, but the objects did change location slightly. On completion of the experiment, each participant was asked to fill out a questionnaire which asked for some personal details including age, sex, and level of computer graphics knowledge. The participants were then asked questions about the quality of the two images they had seen.

3.4 Results

Figure 4 shows the overall results of the experiment. Most participants, except for one only watching the red balls scene, did not report any difference in quality between the HQ/HQ images (they were the same). The results also show that 75% and 82% failed to notice the HQ/SQ difference whilst performing the task in the Kalabsha temple scene and the red balls scene respectively. These percentages dropped to 63% and increased to 88% in the HQ/TQ condition for the Kalabsha temple scene and the red balls scene respectively. Surprisingly, nearly 69% of the viewers, in the HQ/LQ case, in both scenes were so engaged in the task that they failed to notice any difference between these very different quality images. In addition, most observers, nearly 94%, failed to notice a quality difference between the HQ/RQ condition while counting the red balls. This can be compared to 25% while only watching the images.

3.5 Statistical analysis

We analysed the results statistically to determine any significance. The appropriate method of analysis is a "paired samples" t-test, and since each subject had a different random selection of images, an unrelated t-test was chosen [3]. We compared the image pairings, see table 1, to the HQ/HQ data while performing the task to ascertain whether the viewers perceived the difference in quality. For a two-tailed test with the *df* = 30 (*df* is related to the number of subjects), t must be ≥ 2.043 for significance with p < 0.05 (less than 5% chance of random occurrence). For the Kalabsha scene and the image resolution difference experiment, for HQ/TQ t=3, HQ/SQ t=2.36 and HQ/LQ t=2.61 and thus the viewers could distinguish a difference for all selectively rendered and low quality images. However, for the red balls scene and the edge anti-aliasing experiment, for HQ/TQ t=1.15, HQ/SQ t=1.86, HQ/RQ t=1 and HQ/LQ t=2.61. Thus the viewers failed to notice the difference for the TQ, SQ and RQ cases, but did notice the difference when the image was completely rendered in low quality. We also analyzed the results statistically between viewers that were simply watching the images with the viewers performing a task and naturally the results were statistically significant for all cases.

The results for the Kalabsha temple images failed to show that the difference in image resolution quality was not noticeable to the viewers. Further work needs to be done to understand why this was the case, and in particular whether this is a result of the sharp contrast between the high quality stones and the low quality background or our ability to identify objects by using high spatial frequency background information. The viewers, however, did fail to notice the edge anti-aliasing quality difference for the red balls scene when the images were selectively rendered and rendered without reflections and shadows while the complete low quality image was noticeable.

3.6 Verification with an eye-tracker

To confirm that the observes actually did focus on the task objects when performing the counting task, the experiment was repeated for both scenes using the Eyelink Eyetracking System developed by SR Research Ltd. and manufactured by SensoMotoric Instruments. This system has got an average gaze position error of $< 0.5°$. Figure 1 shows the scan paths of two participants. One viewer is simply watching the Kalabsha temple scene, whilst the other is performing the counting loose stones task for 2.75 seconds. The X's in the images are the viewer's fixation points and the lines are the saccades.

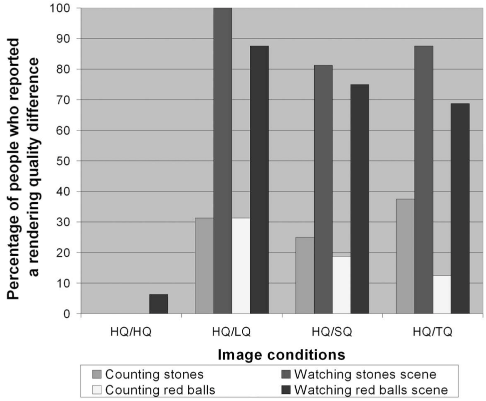

Figure 4: Experimental results for the two tasks: counting objects vs. simply viewing the images.

4 Conclusions

In order to complete a visual task in an environment, a user's eyes focus on the specific task-related parts of the scene. The results of our experiments confirm that, at least for edge anti-aliasing, inattentional blindness can in fact be exploited to significantly reduce the rendered quality of a large portion of a scene without having any affect on the viewer's perception of the overall (high) quality of the rendered image. The results go further than previous work by showing that, when performing the task, a high percentage of the observers even fail to notice low quality of non-task related areas within the visual angle of the fovea. This applies to a number of quality parameters including edge anti-aliasing and reflection and shadows. It is interesting that people notice rendering quality but fail to notice grosser changes, such as the elimination of shadows and reflections. These results advances the state of understanding in task specific rendering and how the user's task can influence the best choice of rendering parameters in realistic rendering. A further understanding is still needed of the complex interaction between the bottom-up and top-down visual attention processes of the human visual system in order to better combine the existing saliency models with our task map approach, into the new task-importance map. Even though many factors which influence visual attention have been identified, little quantitative data exists regarding the exact weighting of the different factors and their relationship. It is especially important to investigate the weighting of these different factors within the area of computer generated images since most previous models have been developed for use with photographs. Such a task-importance map can then be used to accurately guide our selective rendering, ensuring perceptually high fidelity images in reasonable, perhaps even real-time frame rates. We are also planning to investigate the case when task objects are partially occluded and how shadows and reflections of the task objects will affect the visual perception of a scene in more detail. Future work will also include progressive rendering techniques whereby pixels are rendered in order of their contribution to the final image, thus producing the best perceptual images in the given time frame.

5 Acknowledgements

We would like to thank Dimitris Agrafiotis and the Electronic Engineering Group at the University of Bristol for the loan of the Eye-tracking device and his support on using the equipment. We are indebted to Pete Longhurst for the VDP images, Greg Ward, Kate Devlin and Patrick Ledda for your advice and everyone that participated in the experiments, we would not have got any results without you. This work was supported by 3C Research.

References

[1] K. Cater, A. Chalmers and G. Ward, "Detail to Attention: Exploiting Visual Tasks for Selective Rendering", *Proceedings of the Eurographics Symposium on Rendering*, pp. 270-280, 2003.

[2] K. Cater, A. Chalmers and P. Ledda, "Selective quality rendering by exploiting human inattentional blindness: looking but not seeing", *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 17-24, ACM Press, 2002.

[3] H. Coolican, "*Research Methods and Statistics in Psychology*", Hodder & Stoughton Educational, U.K., 1999.

[4] JE. Dowling, "*The retina: An approachable part of the brain*", Cambridge: Belknap, 1987.

[5] J.A. Ferwerda, S.N. Pattanaik, P. Shirley and D.P. Greenberg, "A model of visual adaptation for realistic image synthesis", *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 249-258, ACM Press, 1996.

[6] D.P. Greenberg, "A framework for realistic image synthesis", *Commun. ACM*, vol. 42, no. 8, pp. 44-53, ACM Press, 1999.

[7] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention", *Vision Research*, vol. 40, pp. 1489-1506, 2000.

[8] L. Itti, C. Koch and E. Niebur, "A model of Saliency-Based Visual Attention for Rapid Scene Analysis", *Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1254-1259, 1998.

[9] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry", *Human Neurobiology*, vol. 4, pp. 219-227, 1985.

[10] L.C. Loschky, G.W. McConkie, J. Yang and M.E. Miller, "Perceptual Effects of a Gaze-Contingent Multi-Resolution Display Based on a Model of Visual Sensitivity", *Advanced Displays and Interactive Displays Fifth Annual Symposium*, pp. 53-58, 2001.

[11] L.C. Loschky and G.W. McConkie, "Contingent Displays: Maximizing Display Bandwith Efficiency", *ARL Federated Laboratory Advanced Displays and Interactive Displays Consortium, Advanced Displays and Interactive Displays Third Annual Symposium*, pp. 79-83, 1999.

[12] G.W. McConkie and L.C. Loschky, "Human Performance with a Gaze-Linked Multi-Resolutional Display", *ARL Federated Laboratory Advanced Displays and Interactive Displays Consortium, Advanced Displays and Interactive Displays First Annual Symposium*, pp. 25-34, 1997.

[13] K. Myszkowski, "The Visible Differences

Predictor: Applications to global illumination problems", *Eurographics Workshop on Rendering*, pp. 223-236, 1998.

[14] J. Haber, K. Myszkowski, H. Yamauchi and H.-P. Seidel, "Perceptually guided corrective splatting", *Computer Graphics Forum*, vol. 20, pp. 142-152, 2001.

[15] W. James, "*A saliency-based search mechanism for overt and covert shifts of visual attention*", Principles of Psychology, 1890.

[16] E. Niebur and C. Koch, "Computational Architectures for Attention", *The Attentive Brain*, pp. 164-186, MIT Press, Cambridge, MA., 1998.

[17] A. McNamara, A. Chalmers, T. Troscianko and I. Gilchrist, "Comparing Real and Synthetic Scenes using Human Judgements of Lightness", *Proceedings of the Eurographics Workshop in Brno*, Eurographics, 2000.

[18] A. Mack and I. Rock, "*Inattentional Blindness*", MIT Press, 1998.

[19] K. Myszkowski, T. Tawara, H. Akamine and H-P. Seidel, "Perception-Guided Global Illumination Solution for Animation Rendering", *SIGGRAPH*, pp. 221-230, 2001.

[20] K. Myszkowski, R. Przemyslaw and T. Tawara, "Perceptually-informed Accelerated Rendering of High Quality Walktrough Sequences", *Eurographics Workshop on Rendering*, pp. 13-26, 1999.

[21] G. Marmitt and A.T. Duchowski, "Modeling Visual Attention in VR: Measuring the Accuracy of Predicted Scanpaths", *Eurographics 2002, Short Presentations*, pp. 217-226, 2002.

[22] M. Ramasubramanian, S.N. Pattanaik and D.P. Greenberg, "A perceptually based physical error metric for realistic image synthesis", *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., pp. 73-82, 1999.

[23] R. Sekuler and R. Blake, "*Perception*", McGraw-Hill, 1994.

[24] J.K. Tsotsos, S.M. Culhane, Winky Yan Kei Wai, Y. Lai, N. Davis and F. Nuflo, "Modeling visual attention via selective tuning", *Artificial Intelligence*, Elsevier Science B.V., pp. 507-545, 1995.

[25] S.N. Pattanaik, J.A. Ferwerda, M.D. Fairchild and D.P. Greenberg, "A multiscale model of adaptation and spatial vision for realistic image display", *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ACM Press, pp. 287-298, 1998.

[26] G. Ward, "*The RADIANCE Lighting Simulation and Rendering System*", SIGGRAPH, vol. 40, pp. 459-472, 1994.

[27] S. Yantis, "Converging operations in the study of selective visual attention", *Attentional Capture in Vision*, Washington, DC: American Psychological Association, pp. 45-76, 1996.

[28] A.L. Yarbus, "Eye movements during perception of complex objects", *Eye Movements and Vision*, pp. 171-196, 1967.

[29] H. Yee, S. Pattanaik and D.P. Greenberg, "Spatiotemporal sensitivity and Visual Attention for efficient rendering of dynamic Environments", *ACM Transactions on Computer Graphics*, vol. 20, pp. 39-65, 2001.

[30] B. Watson, B. Friedman and A. McGaffey, "An Evaluation of Level of Detail Degradation in Head-Mounted Display Peripheries", *Presence*, vol. 6, pp. 630-637, 1997.

[31] S. Daly, "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity", *Digital Images and Human Vision*, A.B. Watson, MIT Press, Cambridge, MA, pp. 179-206, 1993.

[32] Randall W. Hill, Jr., Y. Kim and J. Gratch, "Anticipating where to look: predicting the movements of mobile agents in complex terrain", *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, ACM Press, pp. 821-827, 2002.

[33] R. Dumont, F. Pellacini and J.A. Ferwerda, "Perceptually-driven decision theory for interactive realistic rendering", *ACM Trans. Graph.*, ACM Press, vol. 22, no. 2, pp. 151-181, 2003.

[34] D.J. Simons and C.F. Chabris, "Gorillas in our midst: Sustained inattentional blindness for dynamic events", *Perception*, vol. 28, pp. 1059-1074, 1999.

[35] R.A. Rensink, "Visual Attention", *Encyclopedia of Cognitive Science*, London: Nature Publishing Group, 2002.

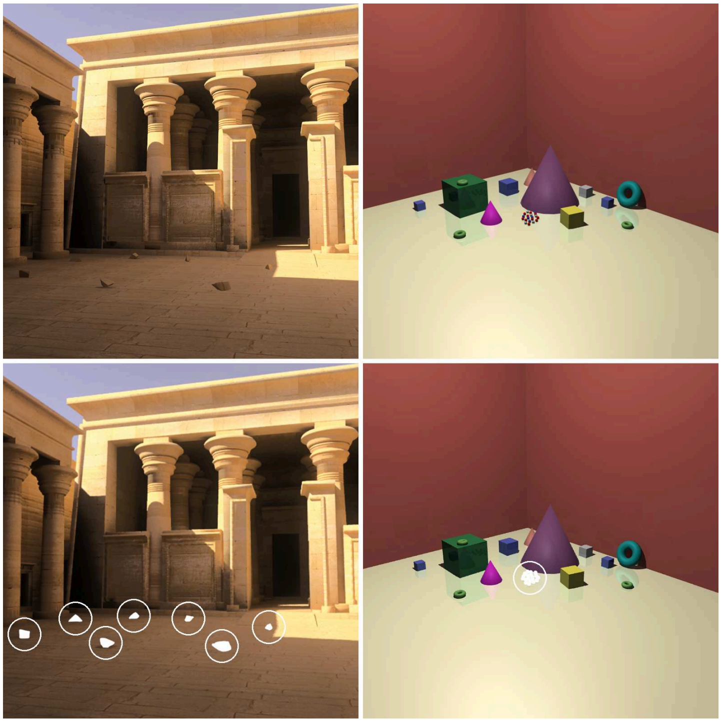

Figure 5: Experimental images: (top left) Kalabsha temple scene in HQ, (top right) Red balls scene in HQ, (bottom left) White objects show the TQ areas in the Kalabsha temple scene whilst the surrounding white circles show the SQ areas, (bottom right) White objects show the TQ areas in the red balls scene whilst the surrounding white circle shows the SQ area.

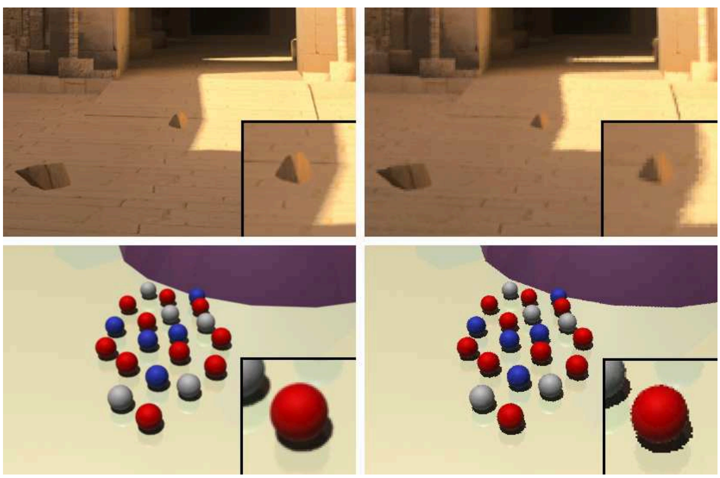

Figure 6: Quality setting: (top left) Resolution 3072×3072. (top right) Resolution 512×512. (bottom left) Edge anti-aliasing quality Highest. (bottom right) Edge anti-aliasing quality Low.