Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Saliency Determination for Computer Graphics: An Experimental Approach

A Thesis

Submitted to the Office of Graduate Studies

of

Trinity College Dublin

in Candidacy for the Degree of

Doctor of Philosophy

December, 2005

by Sarah Howlett

Declaration

This thesis has not been submitted as an exercise for a degree at any other University. Except where otherwise stated, the work described herein has been carried out by the author alone. This thesis may be borrowed or copied upon request with the permission of the Librarian, Trinity College, University of Dublin. The copyright belongs jointly to the University of Dublin and Sarah Howlett.

Signature of Author

Acknowledgements

I am extremely grateful to my supervisor, Carol O'Sullivan, for giving me the opportunity to do my PhD. I really appreciate all her help and ideas over the past three years. I'd also like to thank all the "wonderful" people I have met in the Image Synthesis Group, Richard for his help with the hard stuff and everybody who took part in my experiments.

I'd like to acknowledge my house mates, because they made me, in order of preference; Anne S., Louise and Nicola. My friends; Anne H. (my walking buddy), Helen and Jenny (the crazies from Ballylanders) and not forgetting Caitriona. For friends who didn't get a mention, you should probably try a bit harder.

And finally, my mother for absolutely everything.

Abstract

In the computer graphics realm complex objects are abundant, but often need to be simplified in order to be displayed interactively. As the human visual system is far from flawless, advantage can be taken of its weaknesses by using perceptually adaptive graphics during the rendering of images or animations. In this thesis, we attempt to establish if visual fidelity can be improved by emphasising the detail of salient parts of models, found with an eye-tracking device, at the expense of unimportant areas. In an extension to this, we compared the effect of tasks on eye-movements in a real and virtual environment.

To begin with, we considered the problem of determining feature saliency for 3D objects and describe a series of experiments that examined if salient features existed and could be predicted in advance. To find these salient aspects an eye-tracking device was used to capture human gaze data. In general, results implied that the heads of natural objects were especially salient features. Following this, we investigated if the visual fidelity of simplified polygonal models could be improved by emphasising the detail of salient features identified in this way. In the evaluation of visual fidelity a set of naming time, matching time and forced-choice preference experiments were carried out. We found that perceptually weighted simplification led to a significant increase in visual fidelity for the natural objects at the lower levels of detail (LOD), however, in the case of the man-made artifacts the opposite was true.

As a further step, we carried out some confirmation experiments to examine if the prominent features found during the saliency experiment were actually the features focussed upon during the naming, matching and forced-choice preference tasks. Again results demonstrated that the heads of natural objects received a significant amount of attention, especially during the naming task. We therefore conclude that visually prominent features may be predicted in this way for natural

objects, but our results showed that saliency prediction for synthetic objects is more difficult, perhaps because it is more strongly affected by non-passive tasks that are more related to the objects.

Extending upon this, the next natural step would be to investigate what controls the salient features of man-made artifacts. Moreover, a large quantity of psychology research points to such prominent features being defined by the current task. Unfortunately, before these insights can be applied to computer graphics research, the differences between the effects of tasks in a real and virtual setup have to be recognised.

As a step towards finding salient features of man-made artifacts, the latter part of thesis concerns the framework we built, which was designed to allow the comparison of task performance in real and virtual environments. Realistic graphics, back projection, haptics and rapid prototyping were used to match the virtual scene to the real scene. Some placement tasks were carried out which were evaluated using eye-tracking. Preliminary findings established that attention differs between the real and virtual worlds. From analysis of the video overlay and the average fixation duration found, it is clear that eye-movements are more constrained in virtual circumstances than in the real world setup. In the virtual scenario attention is consumed far more by the object currently being manipulated.

In this thesis, we experimentally show that the visual fidelity of natural objects can be preserved by emphasising their salient features at the expense of unimportant areas. We hope that our results will be insightful to others performing mesh simplification. In addition, the framework developed should be a helpful tool for the examination of eye-movements during tasks. The preliminary experiments suggest that there is potential here and that further examination of tasks in a real and virtual situation is necessary. Perhaps ultimately, this could be extended to ascertain the salient features of man-made artifacts during tasks.

Related Publications

1. "A framework for comparing task performance in real and virtual scenes", S. Howlett, R. Lee, and C. O'Sullivan. In APGV '05: Proceedings of the symposium on Applied perception in graphics and visualisation, 2005.

2. "Predicting and evaluating saliency for simplified polygonal models", S. Howlett, J. Hamill, and C. O'Sullivan. In ACM Transactions on Applied Perception, 2005.

3. "An experimental approach to predicting saliency for simplified polygonal models", S. Howlett, J. Hamill, and C. O'Sullivan. In APGV '04: Proceedings of the symposium on Applied perception in graphics and visualisation, pages 57-64, 2004.

4. "Perceptually Adaptive Graphics", C. O'Sullivan, S. Howlett, Y. Morvan, R. McDonnell and K. O'Conor. Eurographics 2004, State of the Art reports, September 2004.

5. "Saliency determination for polygonal simplification", S. Howlett, J. Hamill, and C. O'Sullivan. Poster in European Conference on Eye Movements, 2003.

6. "Eye-movements and Interactive Graphics", C. O'Sullivan, J. Dingliana and S. Howlett. In The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research. Hyona, J. Radach, R. and Deubel, H. (Eds.) Elsevier Science, Oxford. pp. 555-571. April 2003.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The development of interactive graphics has brought with it a need for techniques to manage the cost of rendering. There is a plentiful supply of complex polygonal meshes currently available in computer graphics, therefore, highly detailed scenes can be created. Unfortunately, for interactive applications, lag is not an option as it degrades human performance [MW93], so the complexity of these scenes has to be controlled in order to save on the amount of computational power needed. Often, highly detailed scenes have to be simplified in order to be displayed in real-time. The major challenge is in maintaining visual fidelity under simplification. Simplifying models in these scenes based upon geometric properties alone may not be adequate if their distinguishing characteristics are rapidly lost, so, when a low polygon count is necessary other approaches need to be examined. One such area is perceptually adaptive graphics, where knowledge of the human visual system and its limitations are exploited during the rendering of images and animations.

Thus, the motivation for this work is based upon knowledge of visual perception. Visual perception seems to perform somewhere below conscious experience, it is almost effortless and operates most frequently without attention. Contrary

to common belief a fully detailed representation of the world around us is not kept in visual memory and only the currently attended objects are stored in any great detail. Although we have the impression of high-resolution over the entire visual field, vision is sharpest only in the fovea, therefore, the point of gaze is closely related to the course of attention and perception. The eyes are moved toward areas where high-acuity, central vision is required or toward objects of interest to the current task. When tasks are involved attention is largely consumed by this, with little or no visual attention focussing elsewhere. Therefore, observing eye-movements during different situations can provide insights into perception.

Thus, one promising solution to preserving visual fidelity under simplification, is to exploit knowledge of the human visual system and its weaknesses when displaying images and animations. If visual perception is considered, there is the opportunity of reducing the required computational resources. In this way, if only the aspects of a scene that receive visual attention or that are currently being focussed upon are maintained at a high level of detail (LOD), the computational power needed can be significantly reduced.

1.2 Overview

In a gaze-contingent system an eye-tracker is used to maintain the area currently under focus at a greater level of detail than the area in the periphery. However, most computer users do not have access to an eye-tracking device to use in this way. Therefore, our approach is to use the eye-tracker mainly to gain insights into the role of feature saliency in model simplification, which might be useful for others when simplifying. We use the device to automatically establish if salient features for 3D models exist and can be predicted in advance. Through experimental evaluation, we wished to determine if knowledge of these features could be exploited during the simplification process to enhance the visual quality of the simplified polygonal models. Moreover, we used eye-tracking to confirm

these results; to verify the evaluation studies and that the actual features found to be salient were indeed those focussed upon during the three tasks used in our evaluation.

Towards our ultimate goal of determining prominent features, inspired by previous task related research from the psychology domain, and our own experimental results, we designed an experimental framework to further examine these issues. Psychological research suggests that individuals have a tendency to process information from only one part of the environment with the exclusion of other parts and that this limited mental capacity is usually allocated to the task, at that given time. Research involving a range of tasks; including hand washing [HSMP03], food preparation [LH01a], driving [SHS01] and block copying [PHL01], all indicate that visual attention is nearly always consumed by the current task. Moreover, some work indicates that it might be the task related aspects of objects that receive attention [JWBF01]. This information would be very useful in determining salient features if it could be transferred directly to computer graphics research. However, it is likely that performance differences exist between tasks in a real and virtual environment, and it is therefore important to establish these before any further examination of salient features can be carried out.

As a starting point to addressing this issue, we describe the framework we built to examine task performance during a real and virtual setup. We used this framework to examine the nature of a participant's eye-movements while various 3D tasks were being carried out in a real world situation and in a matching virtual environment. The framework was designed using computer graphics, back projection, haptics and rapid prototyping. The idea here was to replicate as accurately as possible a purpose-built real world scene and the interaction with this world using haptics. Eye-tracking was used for the evaluation, to compare eye-movements during similar tasks in a real and virtual environment. Following this we discuss some preliminary experiments carried out and our results.

1.2.1 Methodology

The work in this thesis involves:

1. Determining if salient features for a set of 3D polygonal models exist and can be predicted in advance, using eye-tracking.

2. Evaluating, using some psychological metrics, if the visual fidelity of these models can be enhanced by taking saliency data into consideration during the simplification process.

3. Validating previously found salient features and our evaluation studies, using eye-tracking.

4. Finding how tasks control the salient aspects of objects, by building a framework to compare task performance in a real and virtual environment, and recording eye-movements during the evaluation.

Often it is necessary for computer graphics to operate under real-time constraints as well as maintaining realistic and dynamic scenes. Therefore it is useful to know what factors influence perception and can be allocated more resources at the expense of other aspects. The approach taken in this thesis, was to experimentally determine the saliency of 3D polygonal models using eye-tracking. In it we attempted to take advantage of some weaknesses of the human visual system. As perceptual importance is determined by the visual attention of the user, fixation data was gathered from participants, using an eye-tracker, while viewing a set of models at a high LOD. We predicted that, if we could ascertain the salient features of a set of objects in this way and maintain these aspects during simplification, we could improve the visual quality of a set of simplified models.

To this end, we used an SMI EyeLink eye-tracker to determine which features of two sets of models received the most attention. The different sets of models enabled us to examine the influence that factors such as object type and familiarity

had. Then, we investigated if the perceptual quality of these models could be enhanced by presenting these areas of high-acuity in greater detail, thus preserving the perceptually important regions. To do this, the recorded fixation data was applied as a weighting to the model simplification metric. In order to investigate if the models simplified this way actually did have higher perceptual quality, we used some experimental measures of visual perception.

The first psychological measurement gathered was *naming times* [WFM01, WFM00], on the set of familiar objects, this involved saying the name that described an object. In the case of *picture-picture matching times* [LBD02], pictures of a second set of objects were presented simultaneously and had to be matched. Finally, to determine if familiarity played a role *forced-choice preferences* experiments were carried out on both sets of models, this involved picking the stimulus with more of the experimenter-identified qualities. We wished to determine if there was a significant decrease in the naming or picture-picture matching times or a preference towards the models simplified using our approach, especially at the lower LOD's.

Additionally, in the final experiment we used the eye-tracker to see which aspects of objects received the most attention. To validate the previously found salient features and our evaluation studies, we recorded eye-movements during the actual tasks of naming, matching and forced-choice preferences. Moreover, it was possible that the nature of the task affected the results, so we examined the difference between where attention was focussed when viewing images and making image quality judgements through comparisons.

The goal during this phase of our research, was to use an eye-tracker to examine the role of feature saliency in model simplification and, as such, our results should provide insights which may be helpful for other approaches to perceptually guided simplification. This research demonstrates that, if the salient features of natural objects are preserved during simplification, the visual fidelity of these models can be improved (see Figure 1.1).

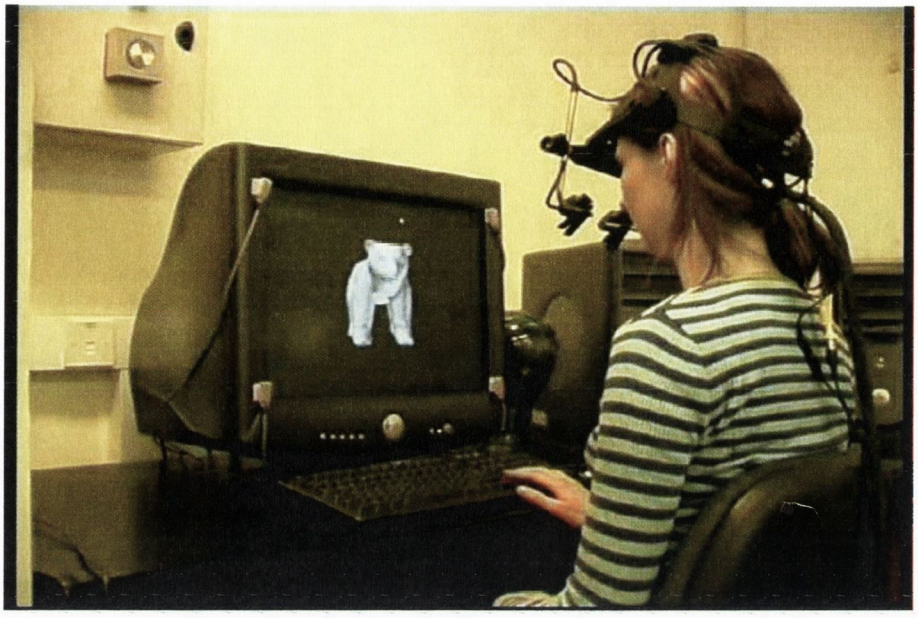

Figure 1.1: Predicting, evaluating and validating saliency for simplified polygonal models.

Psychological research suggests that individuals have a tendency to process information from only one part of the environment with the exclusion of other parts and that this limited mental capacity is usually allocated to the task, at that given time. Research involving a range of tasks; including hand washing [HSMP03], food preparation [LH01a], driving [SHS01] and block copying [PHL01], all indicate that visual attention is nearly always consumed by the current task. Moreover, some work indicates that it might be the task related aspects of objects that receive attention [JWBF01]. This information would be very useful in determining salient features if it could be transferred directly to computer graphics research.

However, contrary to the results for the natural objects, we found that the visually prominent aspects of man-made artifacts were not so easy to predict. A reason for this might be that man-made artifacts are generally related to different tasks and that prominent features may be defined in this way. It is well known that the task performed on an object plays an important role in determining

where attention is focussed. Many recent studies in psychology show that visual attention is controlled to the most part by the current task, this includes hand washing [HSMP03], food preparation [LH01a], driving [SHS01] and block copying [PHL01]. Furthermore, Johansson *et al.* [JWBF01] demonstrates that it is the task related aspects of objects that receive attention, which could be useful in determining the salient features of objects. Such insights would be useful if they could be applied to improve task performance in graphical systems. However, it is very likely that eye-movements patterns differ between a real world and virtual environment. Therefore, before it is possible to apply insights from psychological research in predicting the salient features of man-made artifacts, it is important that the differences in task performance between real and virtual environments are established.

Furthermore, extending from these insights and other background research on tasks from both the psychology and computer graphics literature, we implemented a framework for evaluating visual attention during tasks, in a real and virtual environment. It is true that our experiment is limited to one setup and that there are many factors that define a task, such as the environment, the nature of the task and the objects involved in the task. However, although we only examined tasks in one specific environment, during the implementation of our framework we tried to match these factors, such as the environment, nature of the task and the objects involve in the task using realistic graphics, back projection, haptics and rapid prototyping to compare task.

The aspects of task performance which we consider, include trial duration and eye-movement data during some object placement tasks. The eye-tracking device with scene camera was used during evaluation in the preliminary studies that we carried out. We wanted to create an environment where it would be possible to compare the effects on eye-movements during identical tasks in a real and virtual setup. In this framework we attempted to replicate as accurately as possible a real world scene which we created and the interaction with it using haptics. Some

Figure 1.2: Building a framework to examine task performance in real and virtual scenes.

possible future directions with this framework include; investigating how attention is captured for a variety of tasks, determining the differences in performance and strategies between real and virtual situations, finding the limitations of carrying out similar tasks in a virtual environment. Ultimately our aim is to to ascertain salient feature for man-made objects and to determine how the user experience could be improved, perhaps through previewing or by displaying salient object features [HHO04] or task related objects [CCW03] in greater detail.

In this part of the thesis we describe the implementation of our framework and our efforts to match the setup of the real and virtual environments as accurately as possible. Moreover, we describe some preliminary studies that have been carried out on the real world scene and its virtual counterpart (see Figure 1.2).

1.2.2 Contribution

This thesis improves the state of understanding about the role that saliency data can play in mesh simplification. We experimentally established that the heads of natural objects were particularly salient features, by obtaining perceptual information through eye-tracking. Expanding upon previous, user-guided simplification research, we weighted the simplification metric, not with user-defined aspects, judged by individual preferences and requirements, but with perceptual information gathered from a group of subjects, in order to produce perceptually guided simplified models.

We improved upon previous work by carrying out a thorough evaluation of our simplified models, and proved that the visual quality of natural objects is improved when saliency information is considered during the simplification process. We used three different experimental measures of visual fidelity, originating from the psychology domain, in our studies. Experimental metrics compared to automatic ones provide a more accurate measure of how similar surfaces actually look compared to automatic ones. Moreover, the results from our experimentation provide us with other less positive but useful information; that the visual fidelity of man-made artifacts cannot be preserved in this way, and that it merits further investigation. These insights into the important perceptual details are useful for others performing mesh simplification, *e.g.,* user-guided simplification. This work is described in Howlett *et al.* [HHO04].

The additional validation studies published in Howlett *et al.* [HHO05], provide a confirmation of the previous work. The eye-tracker was used as a validation tool to confirm that the heads of the natural objects were particularly salient and those focussed upon during the evaluation tasks of naming, picture-picture matching and forced-choice preferences.

In the final phase of this work, we take a step towards ascertaining how the salient features of man-made artifacts are determined. We do this by examining

the large amount of psychological research and more importantly through building a system, to compare task performance in a real and virtual environment. Considerable care was taken to recreate as accurately as possible a real world scene in a virtual environment and some interesting preliminary results were found when we used it for experimentation. Using eye-movement data as the indicator, we carried out some novel experiments, in which we found that there were performance differences between the real and virtual environments. We provide a useful framework for task comparisons and our results merit further investigation, providing a direction for future research in this area. This work has been published in Howlett *et al.* [HLO05]

1.2.3 Summary of Chapters

The remaining chapters are organised as follows:

Chapter 2 offers an introduction to eye-movements, visual attention and eye-tracking, in particular, gaze-contingent systems. Furthermore, it provides a detailed review of previous task related literature from the psychology and computer graphics domains.

Chapter 3 gives an in-depth description of the background information on LOD rendering techniques and model simplification as well as the use of perceptual models, saliency, the fixation metrics that we use, measures of visual fidelity and finally the limitations of virtual environments.

Chapter 4 describes the first set of experiments we carried out in which we used the eye-tracker to ascertain the salient features of 3D polygonal models and a discussion of our results.

Chapter 5 presents the modified version of QSlim, which was used to simplify models based on the saliency data found. This is followed by a summary of the three sets of experiments carried out to evaluate the visual fidelity

of the resulting models. These included *naming time*, *matching time* and *forced-choice preference* experiments.

Chapter 6 outlines the validation study, whereby we tracked the participants' eye-movements when they carried out the three tasks of naming, matching and forced-choice preferences, in order to confirm the saliency and evaluation results that we had previously found.

Chapter 7 describes the implementation details of our framework, which allows the comparison of similar tasks in a real and virtual setup, followed by some experiments carried out and a discussion of what was found.

Chapter 8 summaries the work discussed in the thesis and describes some possible future directions of this research.

Chapter 2

Attention, Eye-tracking and Tasks

2.1 Introduction

In our work we investigated what aspects of objects received the most *visual attention* by using an *eye-tracking device* to record *eye-movements*. We incorporated this into a simplification algorithm and performed an evaluation study. Additionally, for further examination of how attention is controlled by task, we built a framework that allows us to compare *tasks* in a real and virtual situation.

Therefore, in this chapter we provide a brief introduction to the human eye and how the human visual system works, eye-movements and visual attention. We then describe some of the current work in computer graphics that uses eye-tracking to take advantage of the weaknesses of the visual system, especially in a gaze-contingent way. Moreover, we outline some previous research on tasks, originating from the psychology domain and, more recently, from the field of computer graphics, in order to prepare the reader for the research we describe in the following chapters.

2.2 Human vision

2.2.1 Introduction

In our work, we attempted to take advantage of the weakness of the human visual system when rendering 3D models. We recorded eye-movements, to determine where visual attention was focussed for a number of reasons. Initially, we used an eye-tracking device to determine what aspects received the most visual attention while a participant was viewing a particular model. This eye-movement information provided us with the input weighting for the simplification algorithm. Following the evaluation of our simplified models, we used the eye-tracker in a validation manner, to confirm that the salient features found, were those actually focussed upon. Finally, as means of evaluation for the framework we built, we recorded eye-movements, which allowed us to compare task performance in a real and virtual environment. Therefore, we mention some general eye-movement terms and give an introduction to visual attention.

2.2.2 The human eye

The human eye (see Figure 2.1) is a specialised light sensitive organ, it receives visual images which are then carried to the brain. The eye has a spherical structure and is roughly 2.5 cm in diameter. Three layers of tissue make up the outer part of the eye, these include; the *sclera*: the outermost layer of protective coating, the *choroid*: a vascular layer which lines the back of the eye-ball and the *retina*: the innermost light-sensitive layer.

Light is allowed into the eye, through the *cornea*, a tough, layered membrane. The cornea is joined to the sclera. The cornea is separated from the *lens*; a flattened sphere made from layers of transparent fibers, by watery fluid called *aqueous humor*. The lens is surrounded by the *ciliary muscle*. The *iris* lies behind the cornea in front of the lens, muscles surrounding it control how much light is

permitted into the eye through the lens. The main body of the eye, behind the lens, is filled with a jellylike material called *vitreous humor*, therefore, the eyeball remains bulgy.

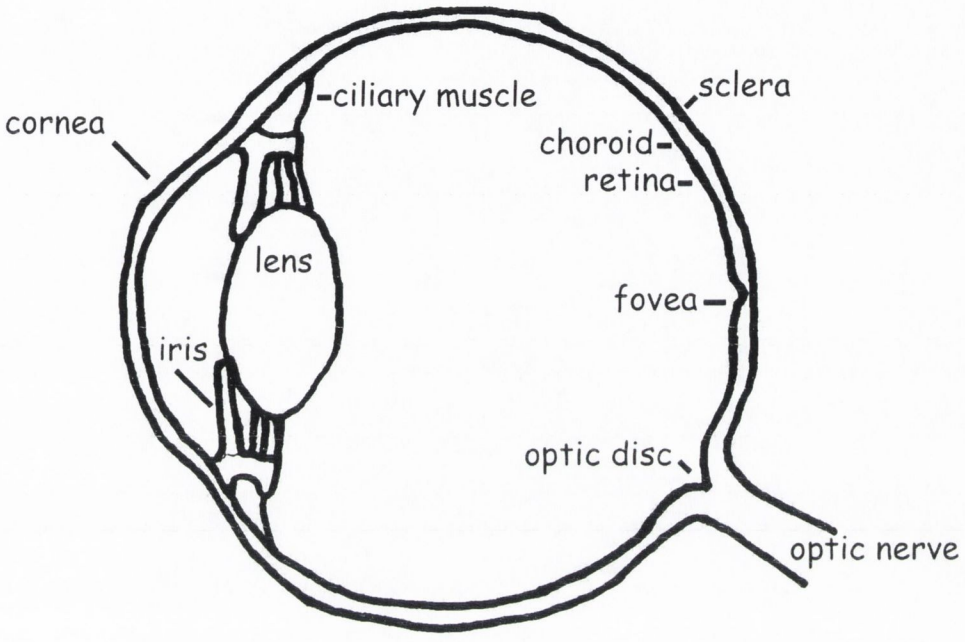

Figure 2.1: The human eye.

The *retina* lies and the back of the eye, and is made up largely of light-sensitive nerve cells, which are shaped like rods and cones. These photoreceptor cells capture light rays and convert them into electrical impulses, which travel along the *optic nerve* to the brain where they are turned into images. The *optic disk*, is where the optic nerve leaves the eye and has no visual receptors, therefore, it forms the blind spot. The *macula lutea*, lies directly behind the pupil and the *fovea* is located at the centre of the macula lutea. Cone-shaped rods make up the sensory layer at the centre of the fovea.

2.2.3 Working of the eye

The process which allows the eye to focus is known as *accommodation*. The ciliary muscle controls the focal length of the lens by flattening it. Distant objects can be seen without accommodation but for closer objects the lens is made increasingly round by the ciliary muscle. The lens brings objects to focus on the retina.

The visual field of the eye consists of a tiny central region of sharpness surrounded by an area of less visual acuity. Due to the neural structure of the retina, the eye sees with most clarity only in the region of the fovea. The reason for this is that this area is made up of cone-shaped cells. These cells are individually connected to other nerve fibers, so that stimuli to each individual cell are reproduced and fine details are distinguished. The cone-shaped cells give us the ability to appreciate colour.

In the surrounding area there are rod-shaped as well as cone-shaped cells, which increase in number with distance from the fovea. Rod shaped cells are joined in bunches, and therefore, respond to stimuli over a larger area. The rods cannot determine fine detail in the visual image but work well in poor lighting conditions. These rod-shaped cells have a high sensitivity to light and are responsible for peripheral and night vision. Dim objects can be seen at night on the peripheral part of the retina despite them being invisible to the most central region.

Despite this, we are not aware that our visual field contains a region of high acuity or central vision surrounded by an area of decreasing detail. This is because the eyes are always in motion, continually bringing different parts of the visual field into the foveal region of interest as the attention shifts from one object to another.

2.2.4 Eye-movements

Eye-movements carry the fovea and visual attention to parts of the scene to be fixated upon and processed at high resolution. A number of different types of eye-movements have been identified by researchers [BG85]. Here, we are particularly interested in *fixations*. Movements, called *saccades*, enable us to direct our eyes to different areas of our visual field to be focussed upon. Research has shown that when viewing an image or scene, the eye tends to fixate and saccade between a certain number of locations repeatedly [Yar67, NS79]. A sequence of these locations create what is referred to as a *scanpath*. Scanpaths are regarded as being idiosyncratic; that is, although people may share the same locations of interest in a scene, they will move around them in different sequences. Furthermore, the task at hand can affect the resultant eye-movements, as pointed out by Yarbus [Yar67].

2.2.5 Top-down and bottom-up attention

There are two distinct types of attentional processes; bottom-up and top-down attention. In the bottom-up case, visual attention is attracted automatically by the salient aspects of a scene. It is a rapid and involuntary process, the eye is drawn immediately towards visually prominent or important features, without a conscious decision being made. For example, one red dot among a set of green dots would attract the viewers attention. The eye responds well to a few parts of the scene and poorly to everything else [IK00].

In the other scenario, where attention is top-down, attention is directed voluntarily towards object of current interest. These are controlled by the particular goals of the viewer when a scene is being studied. There are many cases of task related research which show evidence of this [LH01a, SHS01, PHL01]. In addition, these mechanisms implement our longer-term cognitive strategies.

2.2.6 Visual attention

A great deal of research literature theorises that attention consists of a hierarchical, two-stage selection mechanism. The first stage is referred to as the pre-attentive stage and refers to the early processes that operate in parallel across the whole visual field. It is unlimited in capacity [IK00]. This stage precedes an attentive stage that is of limited capacity and can only process a limited number of items. When items move from the pre-attentive stage to the attentive stage, they are considered to be selected and regarded as having entered into the consciousness of the observer and been made available for higher level cognitive processing. The development of computational models of attention began with Feature Integration Theory [TG80], which viewed the perception of objects within the framework of the two-stage process.

More recently Rensink [Ren00] has proposed a theory of attention which involves three steps. The first step occurs before there is any focussed attention. He suggests that low-level objects which he names proto-objects are continually formed, rapidly and in parallel across the visual field. They can be complex, but are volatile objects and don't have any real memory. In the next step, he claims that focussed attention chooses a few of these proto-objects and stabilises them. This representation has a high degree of coherence over space and time. In the final step focussed attention is released, and objects lose coherence and return back to their constituent proto-objects. He says that there is hardly any consequence on an object of having been attended.

2.2.7 Change blindness and inattentional blindness

The role of attention in perception is a big topic within the psychophysical community. Studies on *change blindness* have shown that even somewhat large changes in a scene can go unnoticed when the view is otherwise interrupted. Numerous experiments have demonstrated that humans can miss large changes in their field of view when they occur simultaneously with brief visual disruptions, such as an eye saccade, flicker, shift of the picture, a film cut or a blink [Sim00, Ren02a, Ren02b].

Another perceptual phenomenon that has been recognised more recently is *inattentional blindness*. This occurs when a stimulus that is not attended is not perceived, even though the person is looking directly at it. This suggests that conscious perception is not possible without attention [MR98].

A major point from this work is that perception of the visual world is not as detailed as our subjective experience lets us believe. We do not have the capability to process everything in our visual field to the same degree, even though we are under the impression of having a fully detailed representation. We have a limited field of view, and the fovea our most acute visual area being only a small subset of this. Furthermore, it has been shown that focussed attention is needed for the conscious perception of change [ROC97]. Visual attention is a selective mechanism whereby some information receives enhanced cognitive processing.

2.2.8 Discussion

Taking on board the research on visual attention summarised above, we hypothesised that, when an object is being viewed, only certain aspects receive focussed visual attention and that this is important in determining what the user perceives. Therefore, we investigated whether, if these prominent feature were maintained at the expense of unimportant areas, the visual quality of simplified polygonal models was preserved. In our research, we also recorded eye-movements when we studied task performance. Eye-movements of subjects were compared while they

carried out similar tasks in real and virtual environments.

2.3 Eye-tracking

2.3.1 Introduction

In our research we used an eye-tracking device to record the necessary eye-movement data and found the aspects of models that received the most visual attention (see Figure 2.2). Eye-tracking provides information on where and how visual attention is focussed while viewing a scene. It is often used to analyse the perception of the participant, as it gives information that is often lost in a verbal report. Additionally, it shows where a participant fixates before they perform an action as well as the aspects that are focussed upon without any cognitive processing. Eye-tracking gives accurate information on what aspects receive focussed attention (*i.e.*, are visually important), for investigation, evaluation and testing purposes. Visual information is often a more accurate measurement than opinions and preferences, particularly when trying to predict what might happen in the future. Recently, eye-tracking has been used instead of traditional methods such as time, error measurements or subjective ratings in usability testing [PHG+04].

Figure 2.2: EyeLink II eye-tracker with scene camera.

Additionally, eye-tracking has been used in an attempt to compensate for the increasing demand on rendering power, this includes the development of gaze-contingent systems [Duc02] and peripherally degraded displays [Red98, WWHW97]. Gaze direction can be exploited by finding the area of screen space that corresponds to the foveal region of interest, which is tiny at approximately 2, and rendering only this in any detail. As single user can only focus on the portion of the display directly under the fovea, computational power can be saved by degrading the image quality in the periphery. Following, is an account of some of these techniques.

2.3.2 Gaze-contingent systems

Computer display resolution poses a constant challenge for the creators of rendering hardware, as more pixels consume more computational resources. Gaze-contingent systems manipulate the display so that the most informative details of the display are generated at the point of gaze and are more degraded in the periphery [Duc02]. The high resolution area moves with the user's attentional focus, so the area under scrutiny is always rendered at a higher resolution. Eye-tracking, with current sampling rates of up to 500hz, can be used to determine what is being attended and high quality can be preserved at the the foveal Region of Interest. In some of the earlier work from this area, Murphy and Duchowski [MD01] presented a gaze-contingent LOD rendering system. They describe an operational platform for real-time gaze-contingent rendering of multiresolution geometric objects. Their gaze-contingent viewing allowed near-interactive frame rates compared to frame rates that were too low to measure when the scene was drawn at full resolution. Parkhurst and Niebur [PN04] also evaluated a perceptually adaptive LOD technique, which renders the currently attended objects in greater detail on a standard desktop system.

2.3.3 Focus plus context screens

In addition to the regular gaze-contingent displays, Baudisch *et al.* [BDDG03] present several other attentive displays, that take advantage of the human visual system. They use eye-tracking and aim to match the subjective quality of a non-degraded display. They describe research into Focus Plus Context Screens an extension to normal gaze-contingent displays, which only degrade the resolution in the peripheral image regions (see Figure 2.3). Foveal regions of arbitrary shape or size can be created, with peripheral regions degraded by arbitrary means such as colour or contrast and not simply resolution. Additionally, the simultaneous display of multiple foveal regions is possible, which can be used for prediction. Usually, when peripheral content is rendered at low resolution, the display hardware is still the same resolution as any other part of the screen surface. However, in the case of a focus plus context screen, there is a difference in resolution between the focus and the context area. It contains a wall sized low-resolution display with an embedded high-resolution screen. The display content pans and can be brought into high resolution focus as required. This is interesting for large maps or chip design where certain areas need to be focussed upon.

In relation to Focus Plus Context interfaces, Lau *et al.* address the need to quantify the costs of such interfaces and their effect on visual perception. In their work they propose a new methodology, using the recently developed "shaker paradigm" [Ren04], which finds the threshold of perceptual invariance on scaling and rotations. They empirically investigate the effect of various Focus plus Context transformations on human perception and demonstrate that the "shaker paradigm" can be used to investigate the effects of nonlinear distortions in Focus+Context systems [LRM04].

Figure 2.3: A Focus Plus Context Screen. (Image from [BDDG03] courtesy of Andrew T. Duchowski.)

2.3.4 Attentive user interface techniques

Baudisch *et al.* [BDDG03] also describe attentive user interface techniques for directing a system's resources towards the scene components in real-time 3D graphics. Specifically, attentive 3D-rendering engines are discussed, which use a viewer's gaze position to vary the LOD at which an object is drawn (see Luebke *et al.* [LRC+02], for further details). Although similar to a gaze-contingent display, such approaches have one main difference; objects in an attentive 3D-rendering engine are simplified at the object geometry level instead of the image level.

2.3.5 Discussion

In all of the cases described here, the idea is to use the characteristics of human vision when designing computer displays, the most significant characteristic being the difference between foveal and peripheral vision. Using eye-tracking in variations of gaze-contingent displays increases display frame rates and responsiveness. Even though rendering and display hardware continuously improve, there will always be a need for alternative means to compensate for the demand for more power and resolution. However, we do not use the eye-tracker in a gaze-contingent way, but rather for determining the prominent features of a set of models in advance of rendering. Only a small group of people have access to an eye-tracking device, so the aim of our research is to gain valuable insights which might enable the design of better LOD strategies. As we found the use of visual saliency to be beneficial at lower LODs, this information could be used when rendering scenes that contain a large number of objects, like during crowd simulation. In certain highly detailed scenes with a large quantity of background or peripheral objects, when computational time is limited it is useful to know that the visual fidelity can be improved by enhancing the salient features of natural objects. Furthermore, although less positive, when designing systems it may be helpful to know that the salient features of man-made objects can not be found in this way. In our research, we also use the eye-tracker as a validation tool, to confirm previous results. In addition, it is used to record eye-movements in order to evaluate the difference between task performance in our real and virtual setup.

2.4 Task performance

2.4.1 Introduction

There is much research from the field of psychology suggesting that visual activity is largely controlled by task. Generally, the experiments carried out involve the

monitoring of eye-movements while a participant carries out a complex natural task in a real world situation. These include food preparation, block copying, sandwich making, natural manipulation and hand washing tasks.

2.4.2 Familiar tasks

Land and Hayhoe [LH01a] compared the results from two eye-tracking studies which investigated the relationship between eye and hand movements in food preparation tasks. They showed that tasks could be divided into a series of actions performed on objects. The next object in the sequence was often focussed upon before any manipulation action occurred. The eyes usually fixate the same object throughout the action upon it. However, they often fixated the next object in the task before the previous action was completed. They described the specific roles of the individual fixations:

1. Locating - establishing the locations of objects for future use.
2. Directing - establishing target direction prior to contact.
3. Guiding - supervising the relative movements of two or three objects.
4. Checking - establishing whether some particular condition is met, prior to the termination of an action.

They concluded that eye-movements during this kind of task were nearly always consumed by task related objects, so attention is primarily top-down and influenced very little by the bottom-up salience of objects. Generally, man-made artifacts are thought of in reference to a task, which is possibly why we did not find any positive results when we tried to ascertain their salient features.

In other work, Hayhoe *et al.* [HSMP03] recorded eye and hand movements during the task of sandwich making. Again results showed that almost all fixations focussed on the task. As change blindness suggests that the information

retained over fixations is limited, they investigate how much of the information used during a task was obtain during previous fixations. More specifically how much information is needed for guiding the movements of the hands and eyes. Their results are largely in agreement with earlier research that the information obtained is often transient and task-specific. They say that a lot of natural vision can be achieved with a "just-in-time" representations.

Johansson *et al.* [JWBF01] examined gaze-hand coordination in a natural manipulation task, where participants had to grasp and move a bar to a target, and found similar results. The bar had to be moved either directly or around an obstacle, and then returned to the support surface. Results showed that participants almost exclusively fixated certain landmarks critical for the control of the task. Compulsory landmarks included those at which contact events happened. Examples of these included, the grasp site on the bar, the target, and the support surface where the bar was returned after target contact. Optional landmarks included any obstacles in the direct movement path and the tip of the bar. They found that the moving bar was never fixated upon. Hand/bar movements following gaze and were linked concerning landmarks. They found that most of the fixations in their task were directing fixations, as described above [LH01a]. They conclude that gaze supports hand movement planning by marking points to which the grasped object are then directed. They demonstrated that participants nearly always directed gaze to objects involved in the task and, therefore, the salience of gaze targets arises from the functional sensorimotor requirements of the task.

In addition, Pelz *et al.* [PCBB01] recorded eye-movements during the familiar complex task of hand washing. Participants were instructed to walk to a bathroom, wash their hands, and return to the starting position. During this procedure the eye-movements were recorded. These experiments revealed a novel perceptual strategy, seconds before information was needed for a task, objects of future interaction were foveated. These look-ahead fixations were task dependant and used to achieve maximum efficiency during information gathering in a natural

task. They propose that look-ahead fixations are used to strategically distribute attention and visual resources to optimise information gathering during natural tasks.

2.4.3 Block-copying tasks

In their experiments, Pelz *et al.* [PHL01] studied the interactions of eye, head, and hand movements during a simple block-copying task. During the task there were some fixations dedicated to gathering information about the pattern. As well as these, other fixations were used to visually guide hand movements when picking up and placing down blocks. Participants used coordinated patterns of eye, head, and hand movements in a fixed temporal sequence. However, these patterns varied with respect to the immediate task context. Coordination was maintained by delaying the hand movements until the eye was available to guide it. Their results suggests that observers maintain coordination by setting up a temporary, task specific coordination between the eye and hand.

2.4.4 Driving tasks

Driving is another area of task related research which has received significant attention, one example of this being work by Shinoda *et al.* [SHS01]. They examined the ability of drivers to detect the presence of stop signs in a virtual world if these signs were present for only a limited period of time. They demonstrated that both the instructions and the local visual context largely controlled detection performance. This suggested that active search is necessary to notice a sign and that the frequency of this search is influenced by what knowledge was already known about the structure of the environment. They point out that, in the acquisition of visual data, top-down processing is an important contributing factor. The highly task-specific fixation patterns revealed in performance of natural tasks support this idea.

2.4.5 Additional task findings

In other interesting research, Hayhoe *et al.* [HBB98] showed that fixation durations revealed effects of the display changes that were not revealed in the perceptual report. In their experiment they made task relevant display changes during saccadic eye-movements. They changed the colour of the target object during a saccade. Despite, results showing that the length of fixations on the models pattern increased, depending on the point in the task that the change occurred, there was no verbal report of this. This indicates that the visual information that is retained across successive fixations depends on the task demands at that moment. This is consistent with previous suggestions that visual representations are limited and task dependent [LH01a].

In their research Ling *et al.* [LH04] investigated whether the characteristics, *i.e.,* the colour and size, of a 3D object interacted in an object similarity task. To investigate this, they used domelike objects, with varying shape and size as the stimuli. The tasks included two discrimination tasks and one object similarity task. Participants had to select the object which was bigger than, the same colour as, or most similar to a reference object. They found that objects with a more saturated colour appeared larger. Further, they demonstrated that during the object similarity task there was an interaction between the two attributes.

2.4.6 Discussion

The majority of this work supports the common idea that visual perception is virtually effortless, that it works below conscious experience and mostly without the help of attention. The eyes are moved both toward areas where high-acuity, central vision is required and toward objects of interest to the current task. Therefore, monitoring eye-movements during tasks can provide insights into the important aspects of objects, determined by their relevancy to the task at hand. In addition to recording and examining eye-movements in a real world situation, this testing

needs to be extended to the virtual world, if the knowledge is to be successfully used to aid rendering in computer graphics.

2.5 Tasks in graphics and perception

2.5.1 Introduction

Recent work in the field of computer graphics includes research that demonstrates experimentally that it is possible to render non-task related objects in less detail than task related ones. Other work investigates the relationship of delay and difficulty to user performance during a placement task. It also examines the effects of previewing, which occurs when certain aspects of a scene are shown in a prelude to a task. Moreover, we describe some work designed to investigate how the visual fidelity of real objects and self-avatars affects task performance in an immersive virtual environment.

2.5.2 Selective rendering during tasks

In the field of computer graphics, recent work by Cater *et al.* [CCW03] further supports the idea that visual attention is largely controlled by task. They showed experimentally that it is possible to render scene objects not related to the task at a lower resolution without the viewer noticing any reduction in quality. They carried out experiments involving a task on a still image. Participants were required to count the number of teapots in a computer generated office scene, which was rendered at three different levels of resolution; high (3072x3072), low (1024x1024) and selective level. At the selective level the scene was mostly rendered at a low level except for the visual angle of the fovea (2 degrees) centred on each teapot (see Figure 2.4). All scenes were exactly the same except for the position of the teapots. Results showed that, when carrying out a task, participants consistently failed to notice any difference between the high and the selective quality image.

Figure 2.4: A selective quality image, whereby it is mostly rendered at a low LOD except for the visual angle of the fovea (2 degrees) centred on each teapot. (Image from [CCW03] courtesy of Alan Chalmers.)

Twenty percent of observers even failed to notice the difference between the high and low quality images. Furthermore, when there was no task involved, the difference rarely went unnoticed. This demonstrates that people primarily attend to task related objects and the authors postulate that such objects can often be identified in advance, depending on the task. They show experimentally that it is possible to render scene objects not related to the task at lower resolution without the viewer noticing any reduction in quality. The main advantage to this approach is that attention is only dependent upon a specific task and not on the individual user. Therefore, no eye-tracker is needed as different people performing the same task should, the authors claim, be using similar visual processes. They show how

task semantics can be used to selectively render in high quality only the details of the scene that are attended to.

Sundstedt *et al.* [SCCD04] take this further by investigating to what level viewers fail to notice degradations in image quality, between task and non-task related areas. They consider how an image can be selectively rendered when a user is performing a visual task in an environment (see Figure 2.5). In particular, they 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, reflection and shadows are altered. Their results 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 effect on the viewers' perception of the overall quality of the rendered image. Additionally, this research shows that, when performing tasks, the low quality of non-task related areas even within the visual angle of the fovea, largely goes unnoticed.

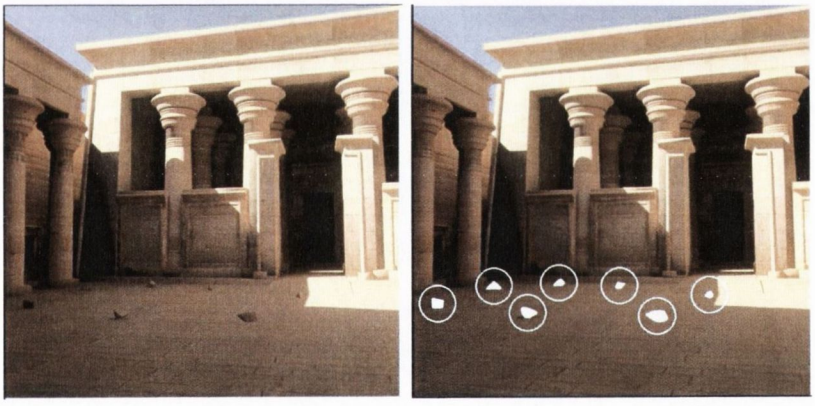

Figure 2.5: Kalabsha temple scene - high quality image on the left and on the right, white objects show the task quality areas and the surrounding white circles show the selective quality areas. (Image from [SCCD04] courtesy of Alan Chalmers.)

2.5.3 Performance gains during tasks

This task related research by Watson *et al.*, although not directly related to the area of visual fidelity, is an interesting approach which could perhaps be extended to incorporate it. Moreover, it is also interesting as tasks of this nature involve the user interacting with the scene, in comparison to a passive task such as counting. In their work, they measured placement errors and time in 3D object placement tasks, and demonstrate the effects of delay and difficultly on results. Moveover, they examined the effects of previewing, by indicating when an object was in the appropriate position. Participants wore a head mounted display and interacted with the environment using a plastic mouse gripped like a pistol. The environment consisted of two rectangular yellow pedestals. On top of the left pedestal there was a translucent box and on top of the right one there were two translucent squares, coplanar to two of the right pedestal's vertical sides (see Figure 2.6). In one experiment, participants had to place a sphere into the box on the right pedestal by releasing the mouse. For the second experiment, previewing was implemented by a colour change when the sphere was in the correct position over the box but had not yet been released by the mouse.

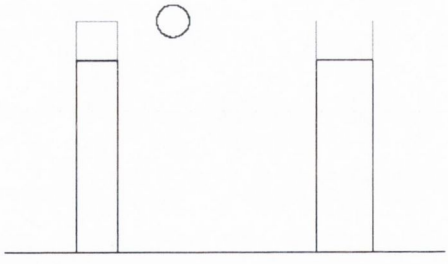

Figure 2.6: A schematic view of Watson's placement experiments - two pedestals with a translucent box on left and two translucent squares on the right. The spherical cursor is moving between them. (Image from [WWWR03] courtesy of Ben Watson.)

The study had 3 levels of delay implemented by adding a delay during each frame and 6 levels of difficulty implemented by varying the width of the right pedestal. For the second experiment, previewing was implemented by flashing the sphere when it was in position. Results for the first experiment showed that placement errors and time increased as delay and difficulty increased. Delay has a greater impact on performance when difficulty is high. Results for the second experiment shows that previewing reduces placement times and limits the effects of delay and difficulty. One possible direction with this work would be to examine eye-movements as well as measuring error and time, while participants carried out these tasks. It would be interesting to see how eye-movements are affected by delay, difficulty and previewing.

As described in Section 3.3.2, Parkhurst and Niebur [PN04] demonstrate that the behavioural costs associated with perceptually adaptive level of detail (LOD) techniques can be offset by the behavioural performance gains during visual search tasks on desktop systems. However, they make an important note that the nature of the task affects the amount of trade-off and behavioural research that is needed before perceptually adaptive rendering using LOD reduction techniques can be exploited fully.

2.5.4 Tasks in virtual environments

For other sorts of tasks, in particular those of a dangerous and expensive nature, virtual environments permit learning, training and practicing that would be otherwise impossible. However, Lok *et al.* [LNWB03] investigated some limitations, such as whether interactivity and effectiveness is decreased by having all objects virtual. They also examined if the visual fidelity of the virtual objects affect performance. They conducted an experiment to examine how the handling of real objects and self-avatars affects performance in a spatial cognitive task in an immersive virtual environment. Performance in a block arranging task was moni-

tored in a real-space setup, in several virtual environments and in a hybrid setup. They showed that, when the task was to manipulate real objects in a virtual environment, the task performance was brought closer to that of real space, than when manipulating virtual objects. Other research shows that providing generic self-avatars results in an increased sense of presence, compared to providing no self-avatar [SU93, SU94]. However, they hypothesise that, if the self-avatar was not accurate representation, the sense of presence would be reduced.

There have also been previous studies comparing performance in real and virtual setups. Thompson *et al.* [TWG+04] have shown that participants are significantly less accurate at judging distances in visually immersive environments than in the real world. Moreover, Mohler *et al.* [MTCR+04] carried out studies using treadmill-based virtual environments to simulate the perceptual-motor effects of actually walking around in the real world.

2.5.5 Discussion

We described some of the research into task performance and analysis in computer graphics. It demonstrates that task related aspects of images can be rendered in high detail, with little or no loss in the overall visual fidelity. This suggests that there is potential here to save on computational resources. We also provide some information about research on tasks in virtual environments. In our work we extend upon this by building a framework which can be used to compare tasks in a real and virtual scenario.

2.6 Concluding comments

In this chapter we gave a brief introduction to eye-movements and visual attention, as well as a description of some recent work using eye-tracking, especially in gaze-contingent display design. In our work, we do use eye-tracking, but not in a gaze-contingent way. Rather, we record eye-movements to ascertain the prominent

features of a set of 3D polygonal models. We also carried out some confirmation studies and preformed the evaluation using eye-tracking in our framework.

Moreover, we provided an account of some of the task related research from the computer graphics and the psychology domains. A lot of the psychological research involved tracking the eye-movements of participants while they carried out a natural task. Generally, results showed that attention was completely consumed by task related objects. Participants usually fixated on the current object with some look-ahead fixations. Furthermore, some computer graphics researchers have exploited attributes of the human visual system to limit what has to be rendered during tasks.

In our research, we wish to expand upon some of these approaches, by examining task performance in a truly interactive, multisensory environment. We try our best to maintain correspondence between a real and virtual situation. We recorded and compared the eye-movements of participants while they carried out similar tasks in a real world and virtual scenario. A description of the framework we built is detailed in Chapter 7.

Chapter 3

Simplification and Visual Fidelity

3.1 Introduction

For interactivity in computer graphics, the ideal is to have the most realistic dynamic scene possible while meeting real-time constraints. Often the computational cost of rendering complex models is too great when real-time graphics are required. To control processing time, simplification has to be preformed and a major challenge is in preserving the visual fidelity of simplified models. Purely geometric simplification can result in a rapid loss of important features. Therefore, when a simplified version of a polygonal model is required, it is sometimes necessary to look at other approaches. These include perceptually adaptive graphics, where the limitations of the the human visual system are exploited when displaying images and animations.

To this end, our work involves investigating whether the visual fidelity of a set of simplified models can be improved if *saliency* data, ascertained using eye-tracking, is taken into consideration during the *simplification* process. Some *experimental measures of visual fidelity* were used to evaluate this. In this chapter, we provide an overview of work relating to these topics. Initially we describe research on level of detail (LOD) rendering techniques, model simplification and

the reduction of model complexity based upon user selection. Following this, we provide a summary of some of the wide variety of work that takes models of visual perception into consideration during rendering. We give an account of work relating to saliency and some background information and reasoning to the fixation metrics that we chose to use in order to predict the salient features. We provide a summary of the typical ways to measure visual fidelity and discuss the three experimental measures we used for evaluation in our experiments. Finally, we give a brief introduction to virtual environments and their limitation.

3.2 Simplification and levels of detail

3.2.1 Introduction

With detailed polygonal meshes everywhere in computer graphics, numerous techniques have been developed to reduce scene and model complexity. Included are methods whereby displayed scenes have to be adjusted in real-time or in non-interactive situations, where the model geometry is simplified to contain fewer polygons.

For real-time applications such as computer games that operate on limited hardware or over a network, the size and complexity of the scene geometry has to be controlled. It is often the case that simplification has to be carried out in order to maintain interactivity. If there is too much detail the resulting lag caused in the system could have a huge effect on performance. There has been some research on lag: MacKenzie and Ware [MW93] discuss sources of lag and its effects on human performance. They show that it degrades human performance in motor-sensory tasks on interactive systems. They present a model which shows a strong multiplicative effect between lag and difficulty. These findings are of particular importance during the design of interactive 3D computing systems. If it is not possible to add more parallel processors or to get higher performance

hardware, methods of simplification are needed. Furthermore, a number of studies demonstrate that lag or a variance in the frame rate can effect the users' ability to carry out certain tasks [RPG99] and can lead to nausea and motion sickness if it occurs in head-tracked systems [BM98].

We describe research on LOD techniques, view-dependent simplifications of a polygonal model and interruptible rendering. Simplification based upon object geometry, specifically the quadric error metric, and more recent work on user-guided simplification systems are also discussed.

3.2.2 Level of detail (LOD) techniques and related work

The use of LOD techniques, described in detail by Luebke *et al.* [LWC+02], is one possible strategy to reduce scene complexity during rendering. LOD techniques try to find a balance between detailed virtual worlds and smooth animations, by adjusting the workload in real-time. Geometric objects are represented at a number of resolutions. There are numerous schemes for implementing LOD using selection criteria based upon an object's distance, size, velocity or eccentricity. These methods attempt to improve performance of applications by maintaining various representations of certain objects, each varying in complexity. Then the most suitable representation based upon some criteria is selected. Clark [Cla76] carried out the initial work in this area and since then there has been work on a number of selection criteria. These include selecting an appropriate LOD for a model based upon the distance from the viewpoint [CB97], the pixel size or area on the display device [Wer93], eccentricity (the degree to which it exists in the periphery) [WWH97] and velocity relative to the user [OYT96].

Some algorithms save hierarchical representations of objects and only render visible and close portions of the model in any great detail. Xia and Varshney [XV96] present an algorithm for performing view-dependent simplifications of a triangulated polygonal model in real-time. The simplifications are dependent on

viewing direction, lighting and visibility and are performed by taking advantage of image-space, object-space and frame-to-frame coherences. Taking this a step further, Klein and Schilling [KS99] describe a new approach for better estimation of normal deviations between the various LODs. This provides accurate lighting with a minimum number of triangles.

Another related idea is interruptible rendering, as described by Woolley *et al.* [WLWD03], which is a trade-off between fidelity and performance. It combines the spatial error caused by rendering and the temporal error caused by delay to create a single image-space error measure called dynamic visual error. Basically, a progressive rendering framework is used, which draws a coarse image to the back buffer. This is continuously refined while the temporal error is simultaneously checked. When the error due to the time delay becomes greater than the error due to the coarseness of the image, increasing the quality of the image any further is pointless, so the image is rendered (see Figure 3.1).

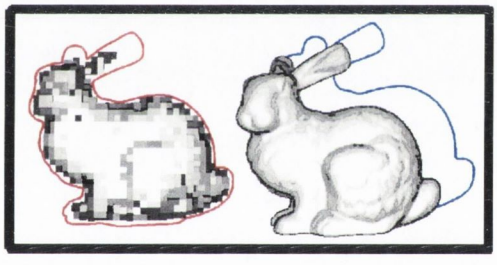

Figure 3.1: The ideal instantaneous image that reflects the latest input is shown in silhouette (coloured outlines). The left image is coarsely sampled, representing some spatial errors. The right image is finely sampled but as a result is quite late. The coarsely sampled bunny actually represents lower dynamic visual error.(Image from [WLWD03] courtesy of David Luebke.)

3.2.3 Geometric simplification

As objects can be captured at a high spatial resolution, which results in some models being hugely over-sampled, numerous methods have been developed to simplify geometries [CMS00]. In the past, geometric methods have been used to reduce the demand for time and memory during rendering [Rus01]. More specifically, the quadric error metric developed by Garland and Heckbert [GH97], which is used as the basis for the QSlim software, allows for a surface simplification algorithm which can rapidly produce high quality approximations of polygonal models. QSlim uses this to provide fast and accurate geometric simplification to automatically produce simplified models. It closes topological holes and joins unconnected regions. The algorithm uses iterative contractions of vertex pairs to simplify models and maintains surface error approximations using quadric matrices. By contracting arbitrary vertex pairs their algorithm joins unconnected regions of models.

Expanding upon this, Hoppe [Hop99] adapted the original approach which used the quadric error metric, to take appearance attributes into consideration when simplifying meshes. This new metric is based upon 3D geometric correspondence, needs less space, evaluates more rapidly and results in simplified meshes that are more accurate. Appearance attributes are not always continuous over the mesh, this results in surface creases and material boundaries. To handle this, they use a wedge-based mesh data structure which captures such discontinuities effectively.

Geometric simplification could be expanded for complex scenes, where there is a demand on resources, selectively processing graphics based upon geometric properties and other significant attributes. Moveover, there is the possibility to combine this with our saliency based simplification which could preserve important regions in the case of some models. A selective rendering system would be ideal because it would be useful to selectively ignore the saliency information for certain objects, such as man-made artifacts.

3.2.4 User defined simplification

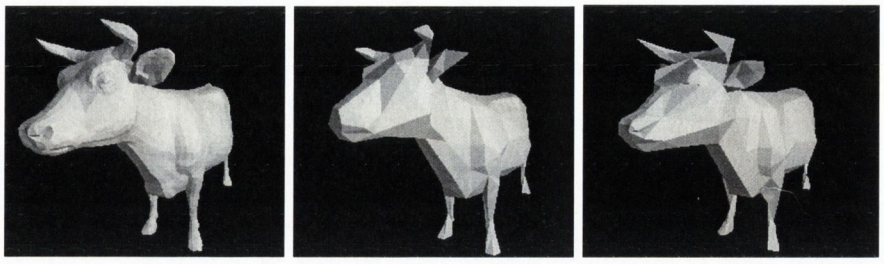

Figure 3.2: Reducing semantic blurring of the head. Original cow on left(10,000 faces), automatically simplified cow in middle (588 faces). Manually improved cow on right (588 faces). (Image from [LW01] courtesy of Ben Watson.)

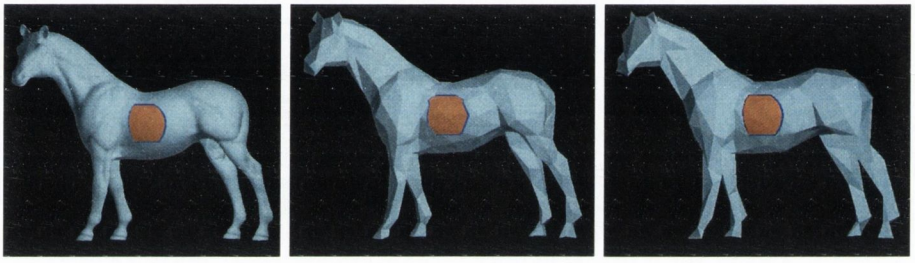

Figure 3.3: Reducing functional blurring. Here, the entire horse is covered with texture, but there is a strong colour discontinuity in the texture. The last two models have the same number of faces, the middle produced by qslim, the right with semisimp.(Image from [LW01] courtesy of Ben Watson.)

Expanding upon pure geometric simplification, Li and Watson [LW01] developed Semisimp, a tool for the semiautomatic simplification of highly detailed models. This tool allows users to improve the quality of extremely simplified models by manipulating the order in which primitive simplifications are applied, the vertex positions of simplified models, and the hierarchical partitioning of the model formed during simplification. The user is able to manually preserve semantically (see Figure 3.2) and functionally (see Figure 3.3) distinct model regions

that are blurred by automatic simplification algorithms, including facial details, regions bound to articulated skeletons, and details embedded in texture mapped images.

Following this, is work by Kho and Garland [KG03], which was also preceded by research from Cignoni *et al.* [CMRS98], in which they provide a user guided simplification system. This extension to QSlim gives the user the ability to select the importance of different areas on a model using a weighted quadric error metric. In addition, they provide a tool that allows the user to apply geometric constraints, which preserves features by guiding the placement of vertices on the approximation. Therefore prominent features of the user's choice are preserved which would be lost if fully automatic simplification was used. To demonstrate this, they apply weights around the eyes and constraints to the teeth of a dragon model in an attempt to preserve the scary aspects of the object. Furthermore, they measured error by the closest Euclidean distances from the original un-simplified model. These user-guided approximations show less error around perceptually important regions (*e.g.,* the eyes) with increased error on other less important aspects like the chin. However, the perceptually important aspects are defined by the requirements of the user and they provide no formal evaluation, to prove that the models simplified in this way have a higher level of visual quality in general.

Also expanding on Garland and Heckbert's quadric error metric is work from Pojar and Schmalstieg [PS03]. They present a tool for user-controlled creation of multiresolution meshes, allowing selective control of simplification. They attempt to find areas of high semantic or functional value by using weights as well as the quadric error metric during simplification. These importance weightings are supplied by the user for a particular mesh region thereby controlling the automatic simplification process. This is done through a Maya plug-in interactively, and the original Quadric Error Metric of Garland and Heckbert [GH97] is weighted by the user input during simplification. The resulting framework allows the user to improve the quality of a multiresolution mesh by taking semantic and functional

importance into account. Their interface works in real-time and is user-friendly.

3.2.5 Discussion

Although systems these days can generate hugely complex models, perhaps with even more detail than the user can actually perceive, maintaining high quality while displaying in real-time is still a concern. The ideal would be to simplify in order to reduce the computational load without losing any of the visual quality. Often simplification based upon geometry alone can create computer graphics renderings that appear very degraded to the human observer. Even manually selected user-defined simplification can result in rendering more polygons than are necessary, resulting in a waste of computational resources as many imperceptible details are produced.

Current user-guided simplification systems preserve prominent features during simplification at the expense of less important regions. These aspects are simply hand picked by the user, judged only by opinions and preferences, and maybe useful and necessary for many applications. However, as eye-tracking data represents more accurately what viewers perceive and process when viewing an object, it should provide a more accurate prediction of the viewers' expected behaviour *i.e.,* what aspects are likely to attract and maintain other viewers attention while examining that object. In our work we expand upon previous approaches by gathering perceptual information. We find the prominent features by examining where exactly the user attends, while viewing a model. Moreover, similar to previous user-defined simplification, we incorporate the data into the simplification procedure to maintain prominent features defined in this way at the expense of the less important regions. There is also the potential for our perceptually based simplification to be used in selective algorithms. The increased visual fidelity of natural objects at low levels of detail under our method could be exploited in specific circumstances. However, for other cases, geometric simplification or sim-

plification based upon different attributes may be more appropriate. Our negative results for the man-made artifacts when simplification is saliency based, suggests that this method would work best best in conjunction with other methods, which take other factors into consideration. In addition, previous studies on user-guided simplification do not provide an evaluation of the models produced in this way. In our work we expand upon this, by performing a thorough evaluation of the models produced using perceptually guided simplification, using some experimental measures of visual fidelity.

3.3 Research using perceptual metrics and models of visual perception

3.3.1 Introduction

There has been a significant amount of work on incorporating principles of perception in managing LOD for rendering meshes. Some of the LOD techniques described above have been adapted based upon perceptual criteria. Perceptual models can help improve simulations by optimising what is actually presented to the user, removing imperceptible details and saving the resources that could be used elsewhere. For example, spending time computing properties such as shape, shading and lighting *etc.* for objects located in our peripheral vision, which we are unlikely to perceive in real time task oriented operation, wastes resources. Other ways of exploiting the limitations of the visual system include studying methods preserving the visual fidelity of models under simplification. The background here provides some information on perceptually adaptive LOD rendering techniques, simplification driven by perceptual metrics and predicting fixations.

3.3.2 Perceptually adaptive level of detail rendering techniques

Parkhurst and Niebur [PN04] provide an evaluation of two perceptually adaptive LOD techniques on ordinary desktop systems. One is a velocity-dependent technique, which takes advantage of the fact that vision is less sensitive to the geometry or other properties of moving objects; and the other is a gaze-contingent technique (see Section 2.3.1 for more detail), which renders the objects being focussed upon in greater detail. These were implemented in the Unreal rendering engine on a standard desktop computer. In each experiment, the task was similar to the traditional visual search paradigm often used in psychological experiments [KJ94, WCG94, HH03], in that participants had to search for a target object in a display. Results suggest that reducing the detail can reduce target identification. However, the increase in frame rates helps with virtual interaction. Overall, the behavioural costs associated with perceptually adaptive LOD techniques can be offset by the behavioural performance gains on desktop systems. However, they also stress that the nature of the task is important in determining the exact cost-benefit trade-off.

Reddy [Red98] investigates how to optimally selected a specific LOD so that the user is not aware of any visual change, based upon data from the field of visual perception. They present a system for implementing LOD that is driven by models of visual perception and evaluate the systems performance. The results suggest that both velocity and eccentricity LOD should be implemented together, as their individual contributions, are negligible, and that distance optimises performance by far the most.

Although, a significant amount of research demonstrates that rendering aspects that are attended to in greater detail can improve visual fidelity, Watson *et al.* [WWH04] have more recently provide further insights into the importance of rendering in the periphery. They point out that much of the previous work on

LOD control is based upon perception at the threshold (see Figure 3.4), despite the fact that most LOD control happens above threshold. The threshold is exactly the point at which a stimulus first becomes perceivable. They highlight results from perception research that show how supra-threshold perception differs from perception at threshold. Their results show that LOD should often be increased in difficult situations, compensating for the challenging environment and maintaining a sufficient level of perceptibility. Threshold-based LOD control should only be used when supra-threshold contrast is low. When LOD control begins to affect task performance, detail should be preserved where sensitivity is lowest. Detail should be added to low contrast regions before high, and to eccentric regions before foveal (see Figure 3.4). Perhaps, to fully exploit the limitations of the human eye, current LOD systems which successfully use gaze data to enhance the currently attended objects and improve visual fidelity, could ultimately incorporate these insights on LOD in the periphery, to improve system performance.

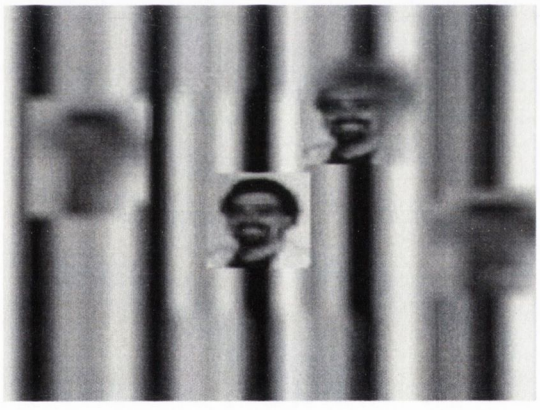

Figure 3.4: A view presented in the second experiment. Here the periphery uses the 20 x 15 LOD, while the lowest contrast background is used. The central area is (always) displayed at the highest HMD resolution. Four distractors are shown. (Image from [WWH04] courtesy of Ben Watson.)

3.3.3 Simplification driven by perceptual metrics

Human vision sensitivity varies with spatial frequency. Some algorithms try to take advantage of this by limiting the spatial frequencies that can be impacted by a change caused by simplification [Red96].

Luebke *et al.* [LHNB00] present a unique polygonal simplification method whereby local simplification operations are driven directly by perceptual metrics, rather than the geometric metrics. The effect each operation has on the resultant image is judged by the contrast the operation causes in the image and the spatial frequency of this change. If the effect of the operation is perceptible, as judged by equations derived from psychophysical studies, the operation is not carried out. They also incorporated gaze-directed rendering into their system, which allows the image to be simplified more in the periphery than at the centre of vision.

In more recent work by Luebke and Hallen [LH01b], they demonstrate a novel approach to reducing model complexity that is driven by perceptual criteria. They use a psychophysical model of visual perception to create a framework that improves interactive rendering and is used for multiresolution rendering techniques. The circumstances under which simplification will be perceptible are determined, and those that are deemed perceptible are not carried out. Their framework is applied to view-dependent polygonal simplification and factors such as imperceptible simplification, silhouette preservation and gaze-directed rendering are taken into account. Their results demonstrate that imperceptible simplification was achieved with a limited reduction in polygon count when this method was used. In their evaluation it was found that the probability of seeing a difference was no better than chance. They claim that models could potentially be reduced even more *i.e.*, up to three times further, without a degradation in perception due to the conservative estimate of the spatial frequency at present (see Figure 3.5).

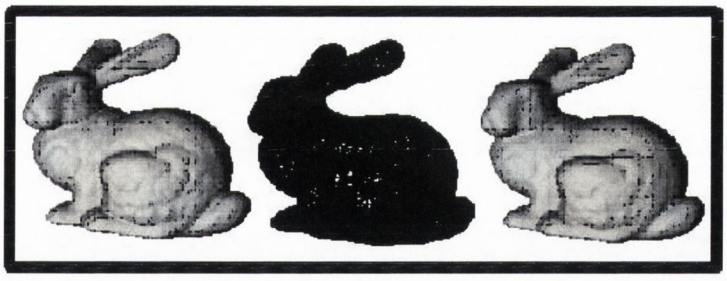

Figure 3.5: Original Stanford Bunny (69,451 faces) and a simplification by Luebke and Hallens perceptually driven system (29,866 faces) (Image from [LH01b] courtesy of David Luebke.)

Closely related is work from Williams *et al.* [WLC+03], who describe a best-effort simplification of polygonal meshes based on rules of visual perception. Best-effort rendering is a form of time-critical computing where processing must occur within a certain time budget. This work applies to a wider range of models and accounts for textures and dynamic lighting. They use parameterised texture deviation to measure distortion more accurately, leading to better simplifications for a certain number of polygons. The simplification of lit models is improved by accounting for both specular and diffuse effects, under both Gouraud-shaded vertex lighting and per-pixel normal-map lighting. Here the focus is not so much on imperceptible simplification, but on the approach of perceptually-guided best-effort rendering to a budget. The most obvious advantage of this approach is on vertex-lit models, because the distortion and tessellation artifacts in specular highlights are highly perceptible. Normal maps are used to maintain smooth highlights even at low resolutions. The system has the ability to simplify low-contrast regions and to preserve high-contrast areas such as silhouettes.

Reddy developed a view-dependent LOD system, which uses a model of visual perception to remove non-perceptible regions when rendering terrain [Red01]. In other to predict the user's limited vision, he uses a model of visual acuity based upon the extent to which a feature is in the periphery and velocity. He shows that

his system saves on computational power and optimises performance.

3.3.4 Predicting fixation

In some cases, a model of visual attention has been used to predict fixations instead of tracking the user's gaze. Privitera and Stark [PS00] have investigated and developed a methodology to automatically identify a subset of algorithmically detected regions of interest using different Image Processing Algorithms, and appropriate clustering procedures. The goal of their work is to replace human visual attention and eye-movement patterns with an engineering approach. Although only a small portion of those available, they demonstrated that the set of algorithms they used accurately predicted eye fixations within their experimental constraints.

Similar to this, Marmitt and Duchowski [MD02] have developed and evaluated a new method for the comparison of human and artificial scanpaths recorded in virtual reality (see Figure 3.6). They use a string editing methodology for the evaluation of human-human or human-artificial scanpaths. They compare the sequence of regions of interest identified using Itti *et al's.* attentional model [IKN98] with those recorded from a human observer. The experiment examined three different scenarios; a simple cube, a panorama, and a more complex graphical environment, which participants were allowed to free-view. They showed that, for all three situations, the similarities between the human and the artificial scanpaths are less than expected. Although this attentional model works reasonably well for still images, it does not accurately predict human fixations in a virtual reality environment. They found that the attentional model assigns attention to a wider area of the image, whereas observers pay more attention to the central region of the display.

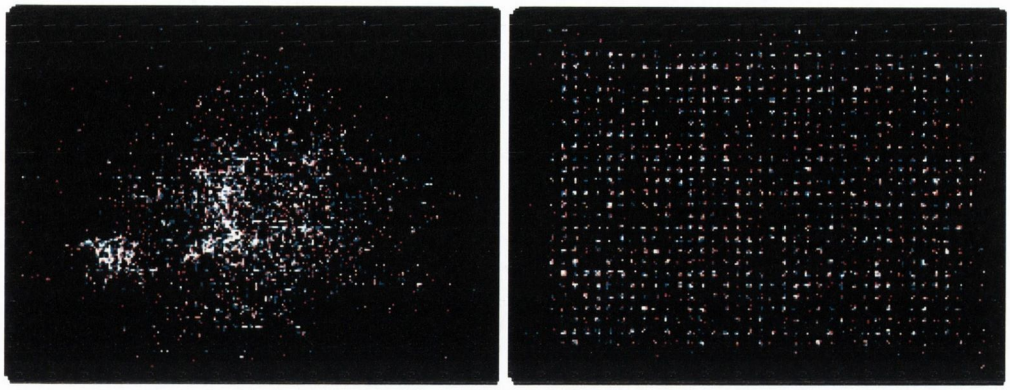

Figure 3.6: Human (left) and artificial (right) scanpaths. (Image from [MD02] courtesy of Andrew T. Duchowski.)

3.3.5 Discussion

Evidently, from the work described above, the use of knowledge of visual perception appears to offer the possibility of maintaining high fidelity scenes in a reasonable time. As our visual system is far from flawless, there is a great opportunity to exploit this. By taking into consideration the fact that the final image will be viewed by the human visual system, rendering time can by saved by not computing those parts of the scene which will not be focussed upon. In our work we wish to determine what aspects of a set of polygonal models receive the most visual attention and take advantage of this data during the simplification procedure.

3.4 Saliency

3.4.1 Introduction

Past research on saliency included efforts to locate prominent object features in images [WCT98], and image based techniques to find areas of high salience [IKN98, YPG01]. In our work we are interested in saliency not in image-based approach

but at the object geometry level [CT99, LVJ05]. However, unlike other work, we used an eye-tracker to find the salient features of a set of 3D objects. As previous research demonstrates that focussed attention is important in determining what the user perceives and that fixations are a good measure of attention, we are interested in finding fixation data for each object. We give a description of prior work on saliency, both image and object based, and some background on eye-movements and the three fixation metrics we decided to employ.

3.4.2 Previous research on saliency

Some of the previous work on saliency includes that of Walker *et al.* [WCT98], who describe object features as being salient, when it is unlikely that they are misclassified with any other feature. In their work, they locate salient object features by using a probabilistic measure of saliency, calculated from several objects to select objects features. In their approach they take advantage of a set of training images, assuming a correspondence between them is known. They train statistical models for each feature, using vectors taken from a number of training examples. The probability of misclassifying a feature is found using these feature models, with salient features having a low probability. They demonstrate that salient features can be relocated better than features chosen by hand or using other methods.

Also, described by Itti *et al.* [IKN98], is a perceptually-based technique which results in a significant reduction in the computational load. Their visual attention system was guided by insights from the behaviour and the neuronal architecture of the early primate visual system. The model used is related to the Feature Integration Theory [TG80], explaining human visual search strategies. One topographical saliency map is created by combining multi-scale image features. A dynamical neural network then selects attended locations in order of decreasing saliency. Additional uses of this method include image base rendering, realistic

image synthesis and geometry LOD selection.

More recently, Yee *et al.* [YPG01] present a method, which is based upon the model of Itti *et al.*, that exploits the limitations of the human visual system, to accelerate global illumination calculation in pre-rendered animations. Spatiotemporal sensitivity determines the amount of error that is tolerable. A spatiotemporal error tolerance map is created to accelerate rendering, using psychophysical data based on velocity dependent contrast sensitivity. This map is then extended with a model of visual attention so that areas where attention is focused is rendered more accurately. When applied to animation sequences, results indicate an order of magnitude of improvement in computational speed.

In a lot of work on saliency involving scenes, salient features cannot be located in a physical sense. Cutzu and Tarr [CT99] find salient object features by physically locating these features on object geometries. They judge the perceptual object features as the salient areas on an objects surface. In their work they present an algorithm which uses either goodness-of view scores measured at several viewpoints, or perceptual similarities among several object views, to find the relative perceptual saliences of the features of a 3D object. The salient regions found, using the assumption that salience varies slowly on a surface, empirically shows the object structures important in human 3D object perception.

Very recent work by Lee et al. [LVJ05] incorporates a model of low-level human visual attention. Instead of using mathematical measures such as mesh curvature in their calculation, they use a saliency measure. They present the idea of mesh saliency, as a measurement of the regional importance, for 3D meshes, as well as a method to compute it. With low-level human visual system cues in mind, they use centre-surround filters with Gaussian-weighted curvatures in their calculations. They demonstrate that incorporating mesh saliency can improve the visual fidelity of several graphics tasks including mesh simplification and viewpoint selection. Their preliminary results indicate that mesh saliency may capture features of 3D models that are important to the human visual system.

3.4.3 Fixation metrics we used

In order to measure the saliency of particular aspects of objects, we recorded the eye-movement of participants while viewing certain objects. We used the fixation data to determine saliency, as fixations indicate the user's spatial focus of attention over time. Loftus and Mackworth [LM78] suggest that the eyes fixate on areas that are surprising, salient or important through experience, therefore we used three fixation metrics to measure the saliency of objects. Laarni *et al.* [LRS03] described some temporal measures of fixations including fixation duration and the number of fixations.

One interesting metric was the total duration of all fixations on a region while a scene is being viewed [HH98]. The total amount of fixation time on an area of interest is generally interpreted as the amount of interest a viewer has in that particular visual element, and it is also interpreted as the amount of time spent processing the information [Lat88]. Baker and Loeb [BL73] found correlations between ratings of the importance of sections of geometric forms and durations of fixations on those sections.

Another metric was the number of fixations on each triangle in the mesh. According to Fitts *et al.* [FJM50], the number of fixations on a particular display element should reflect the importance of that element, so more important display elements will be fixated more frequently. Also Goldberg and Kotval [GK99] show that in visual search, once the participant has found what they are interested in, the number of fixations indicates the amount of interest in a visual area. Furthermore, Mackworth and Morandi [MM67] made comparisons between visual fixations on, and verbal estimates of, the relative importance of regions within photographs. They found that the regions that were rated highly for informativeness produced the highest fixation frequency.

Henderson suggests that the duration of the first fixation on an object [Hen92] is also a good fixation metric, our final metric.

3.4.4 Discussion

As our aim is to find salient object features determined by the human observer and as fixations indicate the user's spatial focus of attention, we used an eye-tracking device to gather some relevant fixation data. Bearing in mind former eye-movement research, we recorded fixations and decided that the prominent features of the objects should be chosen based on some combination of these values. Values were recorded for the total length of all fixations, the duration of the first fixation and the total number of fixations.

3.5 Measures of visual fidelity

3.5.1 Introduction

In order to ascertain how good a simplification is, its visual fidelity has to be measured. To find the best simplification the differences between representations of the same object need to be measured against the original. However, some methods measure perceptually related parameters, but not the perceived quality of a representation. Geometric measures are not perfect measures of how much surfaces will look alike. It is also possible to tell if a simplification is adequate by using experimental measures.

3.5.2 Automatic fidelity evaluation

There have been automatic ways developed to measure visual fidelity. Tools have been developed to measure the geometric distance between surfaces in a variety of ways. Cignoni *et al.* [CRS98] present a system called Metro. This tool allows the comparison of a pair of surfaces. Their approach is based upon surface sampling and calculating point to point surface differences. It is designed as a very general system and works regardless of how meshes were created. Metro supplies visual

results, in addition to numerical results *e.g.,* mesh areas and volumes, maximum and mean error.

Geometric measures such as these are not perfect measures of how much surfaces will look alike, as they are not perceptual metrics. What is needed is some automatic way of measuring how close they are perceptually. Insights from perception need to be applied to these measurements. Browse *et al.* [BRA01], show an example of some work on shape perception in computer graphics, but it has not been used for the purpose of measuring visual fidelity.

Another approach has been to use image-based metrics to predict the perceived degradation of simplified 3D models. However, Rogowitz and Rushmeier [RR01] carried out experiments in which they compared the perceived quality of animated 3D objects and their corresponding 2D still image projections. Their results demonstrate that 2D judgements do not provide a good predictor of 3D image quality, and identify the need to develop better quality metrics.

3.5.3 Experimental fidelity evaluation

In an alternative approach to determining whether one simplification is actually better than another, Watson *et al.* [WFM01] looked at techniques that experimentally and automatically measured and predicted the visual fidelity of simplified models. In their work a set of 36 3D polygonal models were simplified using two different simplification methods (QSlim and Vclust) to two LODs (50% and 80% of the original detail). The stimuli were divided into two different object categories; natural objects and man-made artifacts. Three experimental measures were used to measure the fidelity of these images; naming time (time taken to verbalise the name of an object), ratings and forced-choice preferences. All measures were affected by simplification level and type of simplification. Naming times were longer with increasing simplification and it took longer to name objects simplified using Vclust. When ratings were measured, participants were sensitive to simplifica-

tion level and also rated objects simplified by QSlim as closer to the ideal. The preference results showed that there was a greater preference for Qslim-simplified stimuli, which was greater for the animal models and greater for the objects at 80% detail (see Figure 3.7).

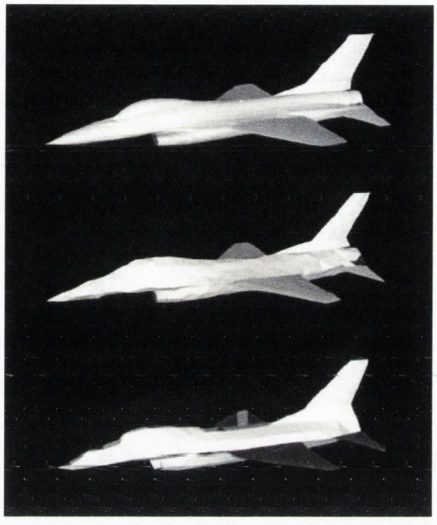

Figure 3.7: One set of stimuli from Watson's experiment: Original (top), QSlim at 80% (middle), Vclust 80% (bottom) (Image from [WFM01] courtesy of Ben Watson.)

The effect of object type was particularly interesting. Firstly, it took longer to name the natural objects, which was consistent with earlier results. Furthermore, the ratings results showed that the animal models were more like the standard when simplified using QSlim, but that the artifacts were more like the standard when Vclust had been used during simplification. Regarding preferences, the preference for QSlim-simplified stimuli was greater for the animal models than for the artifact models and for the 80% simplified models than the 50% objects.

3.5.4 Experimental measures of visual fidelity we used

In our research it was necessary for us to determine the visual fidelity of a set of models. We used three different experimental metrics, from the field of psychology, during our evaluation. The first metric we used was naming time. This involved someone seeing an object and then verbalising the name that described that object, so the objects had to be of a familiar nature. Watson *et al.* [WFM00] carried out experiments to confirm that naming times are affected by model simplification. They present evidence that naming times are sensitive to simplification and model quality. Naming times have also been used in psychological experiments as a measurement. One example is work by Humphreys *et al.* [HRQ88] which concerns the processes involved in picture naming. Forced-choice preference was the second experimental measurement used [WFM00]. In this case, participants had to judge which of two simplified models was more similar to the original version of that model. As our second set of stimuli included some non-familiar objects, it was necessary to use a third metric, a picture-picture matching method [LBD02, Bar76] to determine the visual quality of these models because no verbalisation is required. Kelly *et al.* [KGS98] also used a form of visual-visual identity matching, namely picture-picture, in their research.

3.5.5 Discussion

Following the integration of saliency data into the simplification process, we had to evaluate the visual fidelity of these simplified models. Either automatic or experimental techniques could have been used to measure the visual fidelity of our simplified models. However, it has been shown that automatic measure often do not provide a good predictor of what is perceived by the human visual system, they are not exact measures of how alike surfaces will look [RR01]. In order to measure object recognition and similarity we used experimental measures during our study for increased accuracy in our results.

In our first experiment we wanted to measure how quickly participants could identify a set of familiar objects. The experimental measure of visual fidelity we used here was naming time, as it is the best indicator of ease of recognition [WFM00]. In order to measure the visual fidelity of a set of unfamiliar objects, picture-picture matching was used to compare objects to an original, as this required no verbalisation or prior knowledge of these objects [LBD02]. Forced-choice preference was the final measurement from the field of psychology that we used because of the increased sensitivity of this metric and because we wanted to use both sets of models during this experiment.

3.6 Virtual environments

3.6.1 Introduction

In the final piece of research, the measurements we used for evaluation were values obtained through eye-tracking. Psychological research, which suggests that the areas in a real scene that receive attention are dependent upon task, could provide useful insights if it could be ascertained that these principles also held true in a virtual world as well. We designed a virtual framework which matched as accurately as possible a real world scene, to evaluate the performance difference between real and virtual task performance by using eye-movement data. Moreover, any general insights found with respect to the limitations of virtual environments may be useful to others working with virtual environments. We give a brief overview of why it is necessary to establish the distinctions between the virtual world and the real world that is being emulated.

3.6.2 Benefits and limitations of virtual environments

Virtual environments provide significant support to many fields including the visualisation of scientific data, art, architecture, industry and learning. For industrial

purposes virtual simulations even make it feasible to test expensive or dangerous environments. They permit the evaluation and validation of many different designs more rapidly and cheaply than experimenting in the real world. Frequently, it is safer to carry out experiments in a virtual environment. However, despite these advantages, care must be taken when making assumptions about the correspondence between the real and the virtual world. It is important to realise the variations between the two, especially if training and practising are high risk or costly tasks. Neglecting to account for distinctions could have a very expensive or even disastrous effect when a user transfers from a virtual to a real environment.

In addition, if participants spend a lot of time and cognitive load learning to interact with a virtual environment, the overall effectiveness of the virtual world may be reduced or even pointless. Lok *et al.* [LNWB03] pointed out that having interactions with real objects within a virtual environment has its benefits. As current virtual setups are far from being ideal, and before intuitions from the real world can be used in a virtual environment, or assumptions can be made, tests have to be carried out to determine the extent of the correspondence between the real and virtual setup. If not, there is potential for something disastrous to happen.

3.6.3 Discussion

In our case, although there is no possibility of something disastrous happening, it is nevertheless necessary for us to carry out comparisons. As the role of the framework is to enable us to understand the differences between performance in a real and virtual setup and eventually find a way of using the insights found by researchers from the field of psychology to understand how the salient features of man-made artifacts can be predicted. Therefore, it is important to establish the differences between a real and virtual situation before further progress can be made.

3.7 Concluding comments

As there is a constant demand for high frame-rates in interactive environments, simplification plays an important role. In the past it has been based upon several criteria including geometry, perceptual models and more recently user-guided simplification.

However, in our work we expand upon this by performing simplification based upon perceptual information, obtained through eye-tracking. In a similar fashion to user-guided simplification, the simplification algorithm is weighted, however, in this case with where precisely a participant focusses when viewing a particular model. In addition, it is equally important that the perceptual quality of these simplified models are evaluated. Therefore, we use experimental measurements of visual fidelity in order to provide an accurate measurement of the perceived quality of our simplified models. In addition to our evaluation, we carry out a validation study, using eye-tracking, as a further confirmation of our results. In this work, we hope to find general insights into the role of saliency during simplification.

The remaining part of our research was driven by the results we found, and much research from the psychology field, which suggests that visual attention is often task dependant [HBB98], as well as other research demonstrating that gaze is directed towards areas that are important to the control of the task [JWBF01]. Therefore, if aspects of objects that are directly related to the task could be predicted, this may be useful in finding the salient features of objects. However, before this can be investigated, it has to be determined if attention is consumed similarly by tasks in the real and virtual world, or else these insights would not be relevant. Hence, a framework that allows the comparison of eye-movements in real and virtual environments is necessary.

To this end, the remainder of this thesis documents the saliency determination experiments carried out, followed by a description of the evaluation and validation studies. In addition to this we provide a description of the framework we built,

to study task performance.

Chapter 4

Perceptually Guided Simplification

4.1 Introduction

Inspired by and based upon knowledge of vision and visual perception, the first idea in this thesis was to simplify models, not only based upon geometry, but geometry weighted by a perceptual value. Fixations were used as a measure of visual attention or saliency. Features thus defined are viewpoint independent, since each triangle in the mesh will have a weight. As the perceptual importance is determined by the user, fixation data was gathered from participants while viewing a set of models at a high levels of detail (LOD). Three measurements regarding fixations were recorded, which included the total length of all fixations, the duration of the first fixation and the total number of all fixations on each face in the mesh. Then, using a combination of this fixation information and model geometry during the process of minimising the number of polygons, we hoped to create a model with a higher perceptual quality, thus preserving the perceptually prominent features at the expense of unimportant regions.

The initial step was to use the eye-tracker to record the salient features of the

models automatically. At any instant the eye is either fixating on something or making a saccade (an eye-movement). A saccade can be detected by measuring the difference between the current eye position and the average of the last six eye positions. We had to obtain information about what exactly the salient features of the models were. The eye-tracking device was used to get information on where a participant was fixating when viewing a particular model, for example a viewer may spend significantly more time examining the head of an animal than its body. It would therefore make sense that these features should receive more detail at the expense of less important aspects.

To measure a fixation, the size of the visual angle was computed, and if this was greater than some threshold a saccade was recorded. We kept track of the faces in the polygonal model that were focussed upon since the last saccade until a new one was detected, then we updated these with the fixation data. The threshold value for saccade generation had to be large enough to deal with a phenomenon referred to as the "Midas Touch" problem by Jacob [Jac93]. Even when fixating, the eye makes tiny jittery movements called micro-saccades that are not intentional. Therefore we have to keep the threshold high enough so that this jittery movement does not cause a saccade to be generated while a real saccade is detected correctly.

We obtained information regarding fixations, the total number of fixations, the total length of each fixation and the duration of the first fixation on each face. A false colouring method was used to determine which faces were being focussed upon. Faces were drawn (without lighting) to a back buffer with a unique colour associated with them. When the point under the EyeLink gaze was found, the colour under the corresponding region in the back buffer was read back. As colours were unique, the face or faces being focussed upon could be determined. Furthermore, by expanding the region under scrutiny, the neighbouring faces to the fixation point could be determined easily. From observation (using triangle highlighting while viewing the models), we determined that a square region of

20x20 pixels represented a good zone of interest.

4.2 Apparatus (Eye-tracking device)

In this saliency determination experiment it was necessary to get information on where a participant was fixating when viewing a particular model. An SMI EyeLink eye-tracker developed by SensorMotoric Instruments was used (see Figure 4.1). It was connected to a PC, with a video sampling rate of 250hz. It was a remote eye-tracking device that used infrared light and provided online analysis of eye-movement data into saccades, fixations and blinks. It consisted of a head-mounted band, two miniature high-speed cameras that took images from both eyes, and a third camera that tracked four infrared markers that were mounted on the monitor in order to provide head motion compensation and true gaze position tracking. The subject's PC and monitor output were linked to an operator PC. Here, an image processing system synchronously analysed the images from all three cameras in real-time and provided the position of the pupil from both eyes and the marker.

Figure 4.1: The initial SMI EyeLink eye-tracking device.

Recently this has been upgraded to the new EyeLink II eye-tracker which has a video sampling rate of 500hz and a built in scene camera (see Figure 4.2). This device has a high resolution with a noise limit of less than 0.01 degrees. Also, its data rate of 500 samples per second, means it can be used for saccade analysis and smooth pursuit studies. The eye-tracker can also be used in a gaze-contingent way, as gaze position data is available within delays as low as 3 milliseconds. Moreover, eye events such as saccades and fixations are available within 25 ms to the display computer. The EyeLink scene camera works with an external video overlay box, which generates the overlay graphics with video inputs from scene camera and a portion of the EyeLink VGA display, indicating the current gaze position. In addition to allowing the recording of eye-movements at a fixed viewing distance (*e.g.*, computer monitor, TV screen, *etc.*), the EyeLink scene camera option allows tracking of participants' eye-movements at different viewing depths with high accuracy in the same recording. The EyeLink II tracker interface consists of a set of setup and monitoring screens (see Figure 4.3), which are navigated from the host PC or from the Display PC using the Scene Camera DV Application.

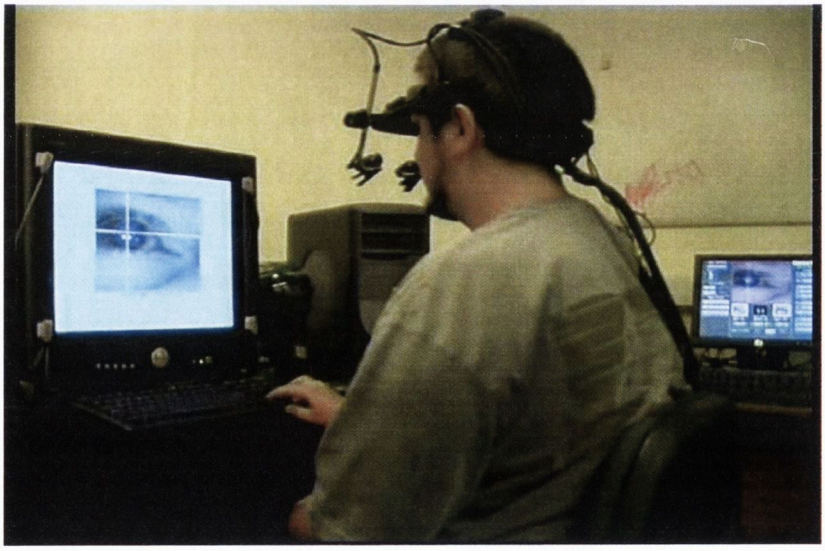

Figure 4.2: The new EyeLink II eye-tracker with scene camera and setup screen.

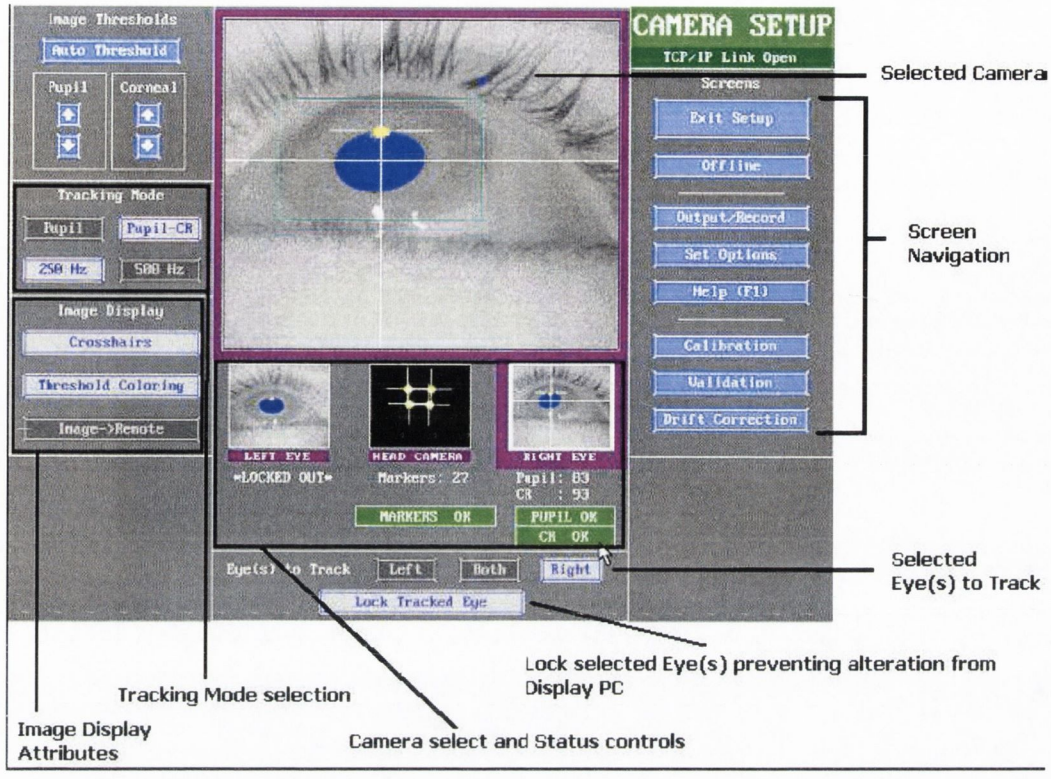

Figure 4.3: The EyeLink II setup screen.

4.3 Participants and apparatus

There were 20 participants involved in the saliency determination experiment; 8 males and 12 females, ranging in age from 19 to 27, from various backgrounds. All had either normal or corrected to normal vision and were naïve to the purpose of the experiment.

There were two different sets of models for viewing. The first set contained 37 familiar objects; 19 natural objects (see Figure 4.4) and 18 man-made artifacts (see Figure 4.5), which were in the public domain, and the same stimuli as those used in Watson *et al's.* [WFM01] experiment with one additional model. Using QSlim [GH97], all 37 of these objects were simplified to have an equal number of faces. The second set contained 30 models which were divided

into 4 categories; 4-legged animals, fish, cars and gears (see Figure 4.6) (models in the public domain - http://www.toucan.co.jp (fish), http://www.3dcafe.com/, http://3dmodelworld.com/). These models could be classified in several ways; natural and man-made, familiar and unfamiliar and symmetric and non-symmetric.

Using QSlim, all the animal objects were simplified to have 3700 faces, the fish, cars, and gears to 5200, 7868 and 1658 faces respectively so that the number of faces per model was uniform only within each category. The number of faces were selected to provided an accurate representation of these objects and were regarded as the standard model at highest LOD (*i.e.*, with the most polygons).

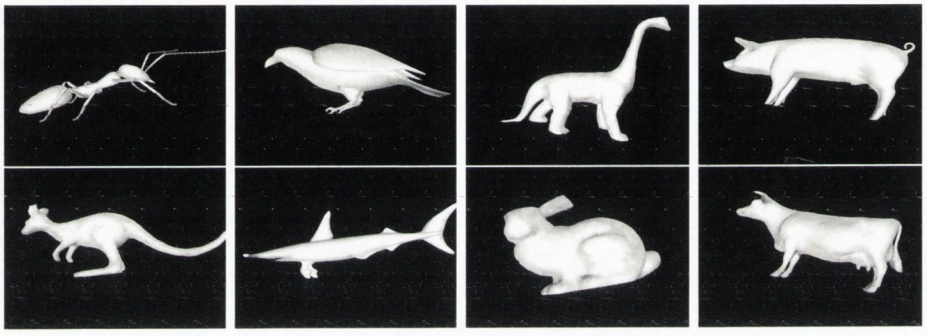

Figure 4.4: A subset of the natural objects used.

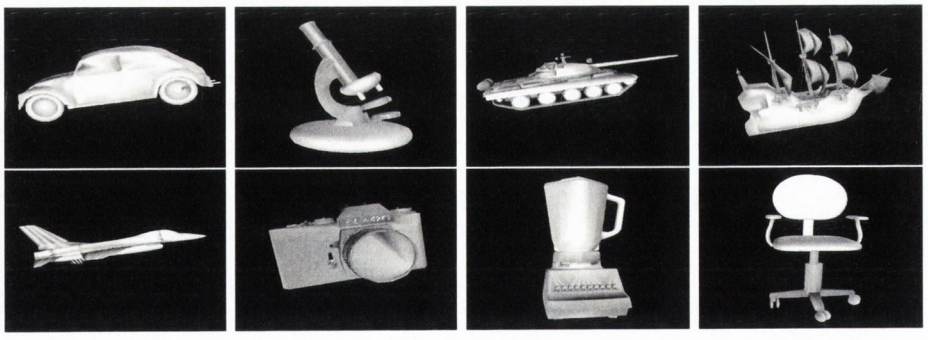

Figure 4.5: A subset of the man-made artifacts used.

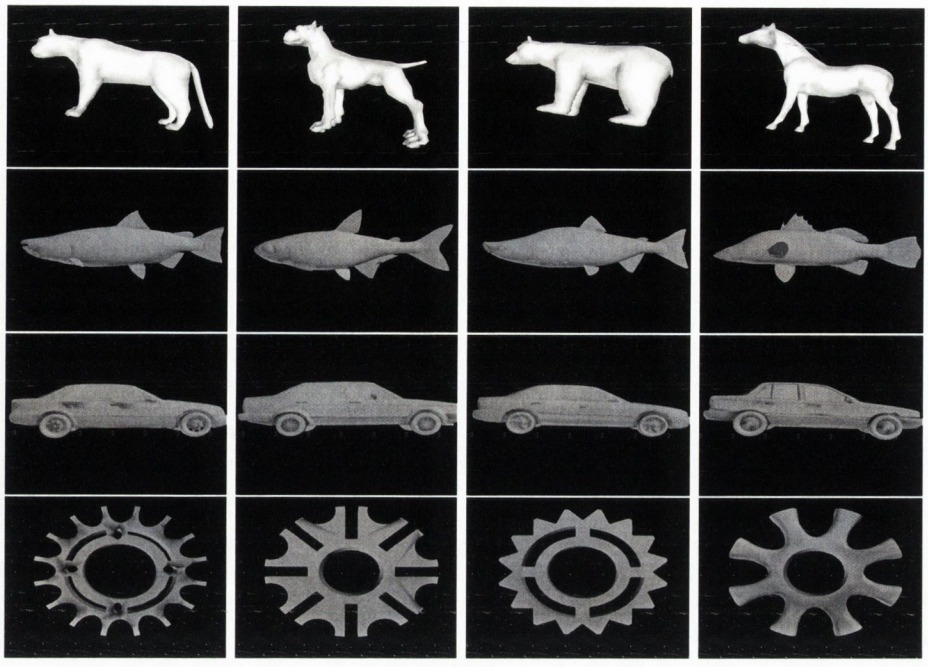

Figure 4.6: A subset of the animal, fish, car and gear models used.

4.4 Method

For both sets of models, each participant viewed each model twice for approximately 30 seconds, from two different initial orientations (see Figure 4.7). The two initial positions were front-facing and back-facing but participants were free to change the orientation using the arrow keys, as Watson [Wat03] in new work investigates how image rotation reduces simplification effectiveness. For the first set there were 74 trials per participant, which were organised into four blocks for viewing. Each block was made up of two groups; a group of natural objects and a group of man-made artifacts. For the second set there were 30 models and therefore 60 trials. This time models were only divided into two blocks, each containing two groups; the first one the animals and the car models and the second block containing all the fish and gear models. Within each group the models were randomised.

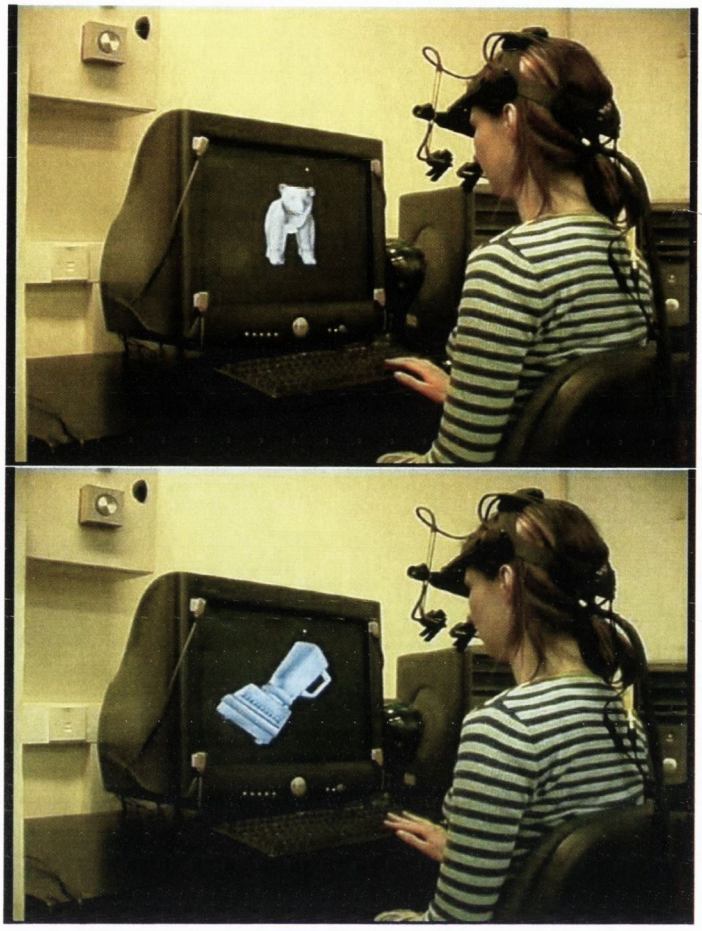

Figure 4.7: An example of a participant performing the saliency determination experiment.

Each session began with the seating of the participant approximately 60cm in front of the PC. Participants were informed of their task and asked to minimise their head movements. The experimental session started as soon as the participant indicated their readiness verbally. The eye-tracker headband was then placed on the participant and calibrated using the EyeLink calibration procedure. Before each experiment, calibration, validation and drift correction had to be carried out to ensure the information was reliable. The calibration phase of the eye-tracking experiment consists of finding a suitable threshold for detection of the

participant's pupil so that accurate eye-tracking may take place. Calibration was carried out for both eyes. The calibration and validation procedure involved the participant following a black dot around a grey screen and focussing on it. The experimenter did not proceed until a good calibration and validation was acquired to ensure a high quality of eye-tracking. Also prior to each model being displayed, drift correction was performed again. Almost all eye-tracking systems have a tendency to drift over time, meaning that the error from the participant's actual gaze position increases. Drift can be caused by a number of factors, such as head movement. The drift correction procedure displays a grey screen with a black-dot at the centre and requires the participant to look at this position; they are not permitted to continue the experiment until they are looking straight at the dot. Participants were told to examine each of the models carefully for the time they were displayed, bearing in mind that they would need to recognise them at a later stage. Models were displayed on a 21-inch monitor with diffuse, grey shading. There was no other ambient lighting present in the room.

4.5 Results

We are aware that the scope of this experiment is limited to the shape of an object and that other factors such as colour, viewpoints, textures and context, which we do not consider, also play a role in determining visual fidelity. However, the goal of our research was not to examine all aspects but to try and improve upon purely geometric simplification, which only considers shape, by making use of eye-movement data during the simplification of 3D polygonal models. Of course, any insights we find could be combined with other factors when rendering scenes, in order to obtain a scene a with the best possible visual quality. While some trials had to be omitted due to calibration error, found by examining the videos with the overlayed data, this was only 1.6% of all results. The information on fixations was summed over all participants, giving us the final data for each object. The

results over all participants are best seen visually with a colour map, which shows the important fixation data we use. The colour map ranges from black through to white with increasing total fixation length, increasing first fixation length and finally with increasing number of fixations.

In the case of the familiar objects, results showed that, for the natural objects, there were some general prominent aspects. As expected, perceptually important features like the heads, eyes and the mouth, were viewed considerably more than the less salient features (see Figure 4.8).

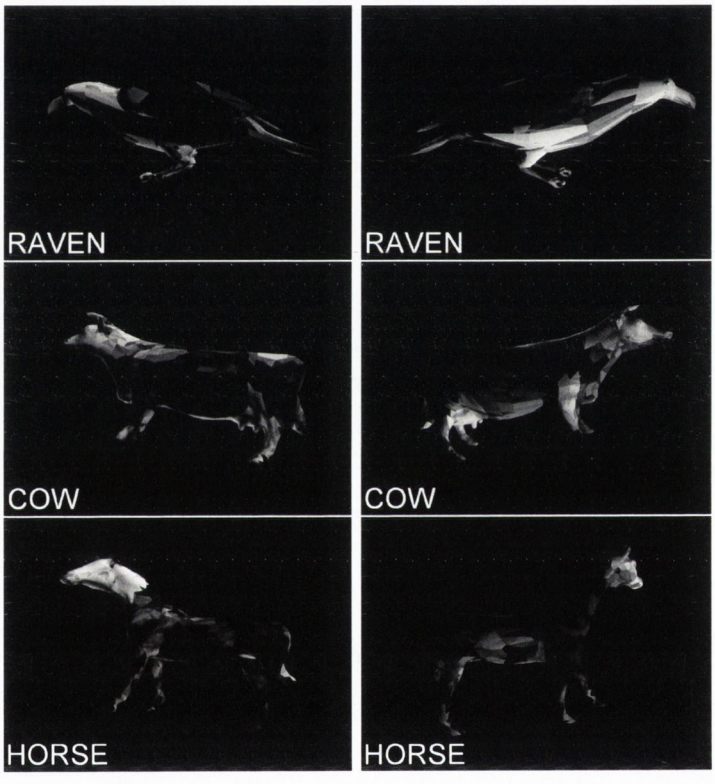

Figure 4.8: Results from the saliency experiment (white representing the greatest number): the total length of fixations on the familiar natural objects.

For the man-made artifacts they also appeared to have important aspects. However, not surprisingly due to the nature of the this set of objects, prominent features varied from object to object. These included the straps of the sandal, the keys of the piano and the buttons of the blender (see Figure 4.9).

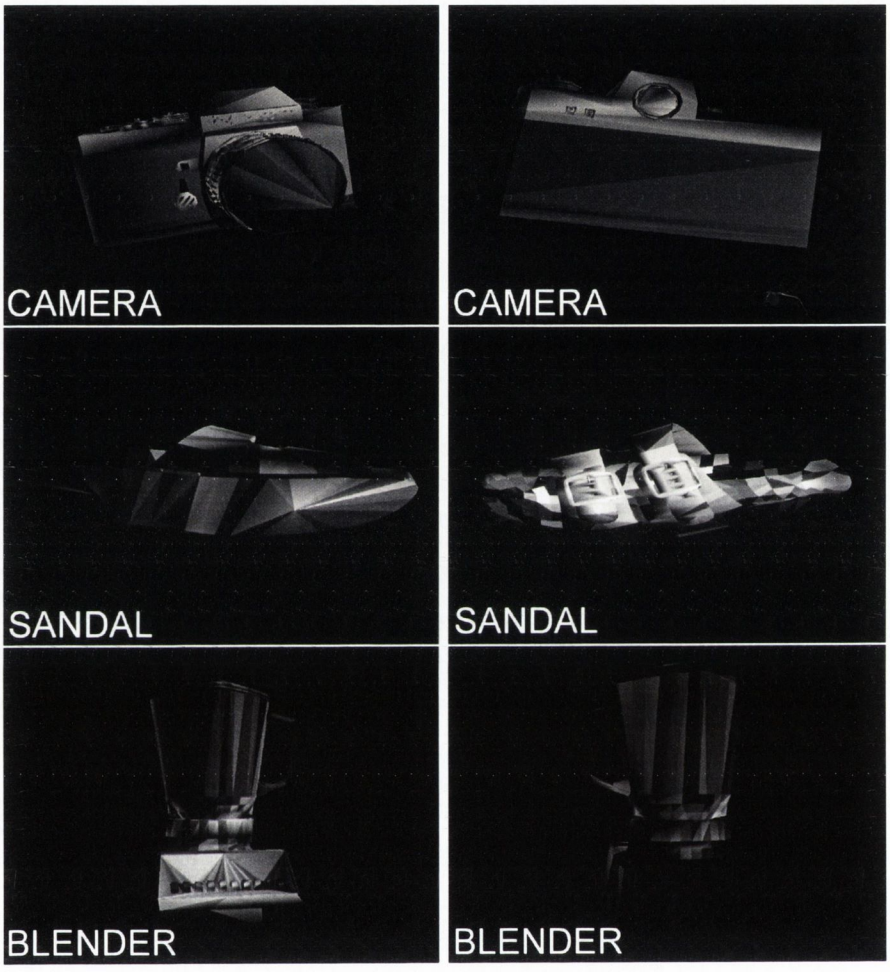

Figure 4.9: Results from the saliency experiment (white representing the greatest number): the duration of the first fixations on the man-made artifacts.

The second set of objects contained four types of models; these include a selection of animal models as well as unfamiliar fish, car and gear models (see Figure 4.10) .

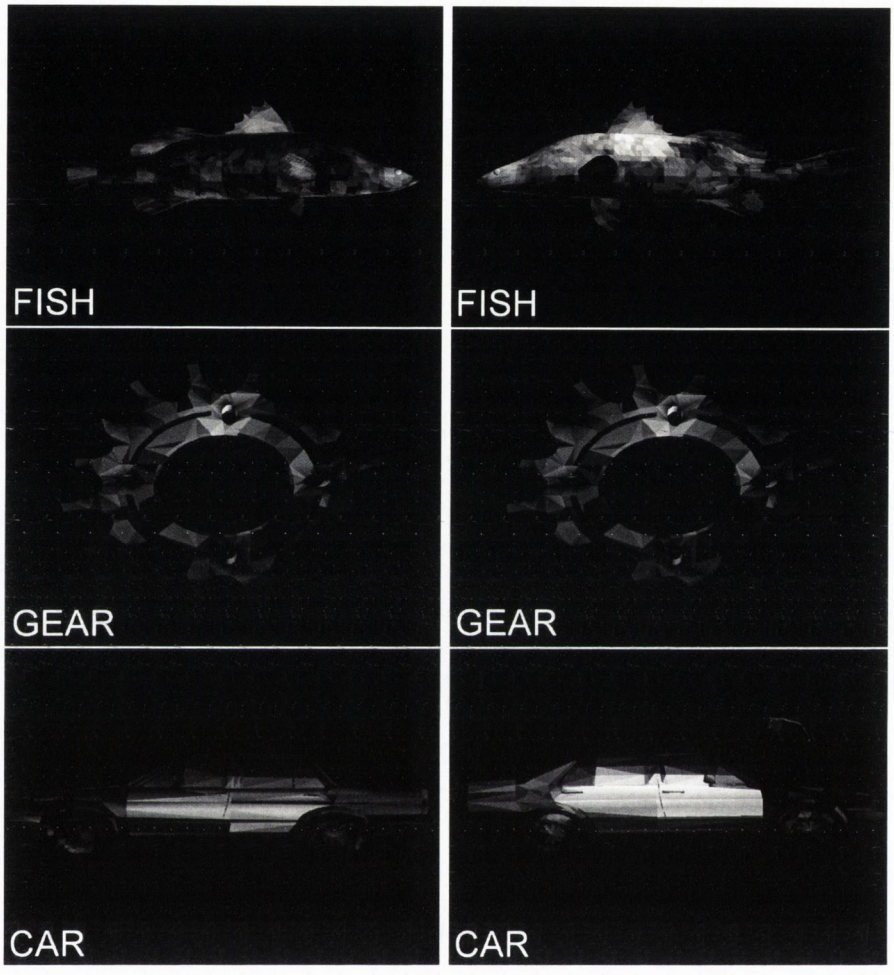

Figure 4.10: Results from the saliency experiment (white representing the greatest number): the total number of fixations on the unfamiliar objects in the second set.

The cars' prominent features included the door handles, side mirrors and the front and back registration plates (see Figure 4.11).

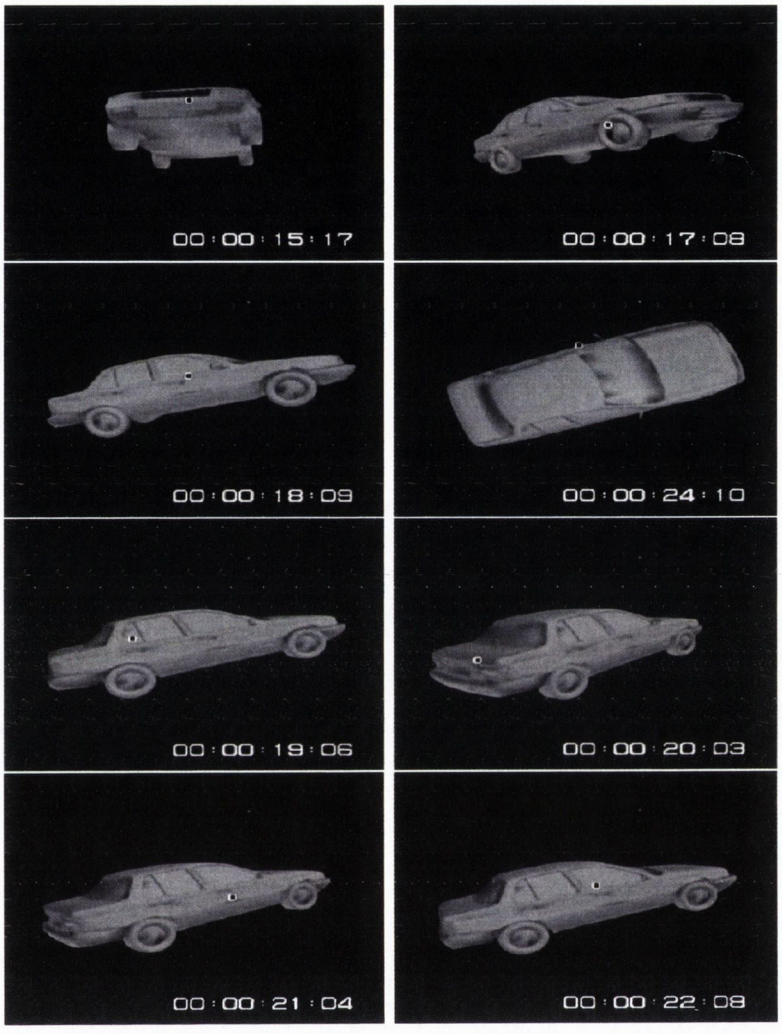

Figure 4.11: Images of the Video Curvid Overlay of one participant on a car object.

For the fish, attention appeared to be primarily focussed on the upper fins, the other fins also appeared to get some attention and, like the animals, the eyes and the mouth were fixated on for a significant amount of time (see Figure 4.12).

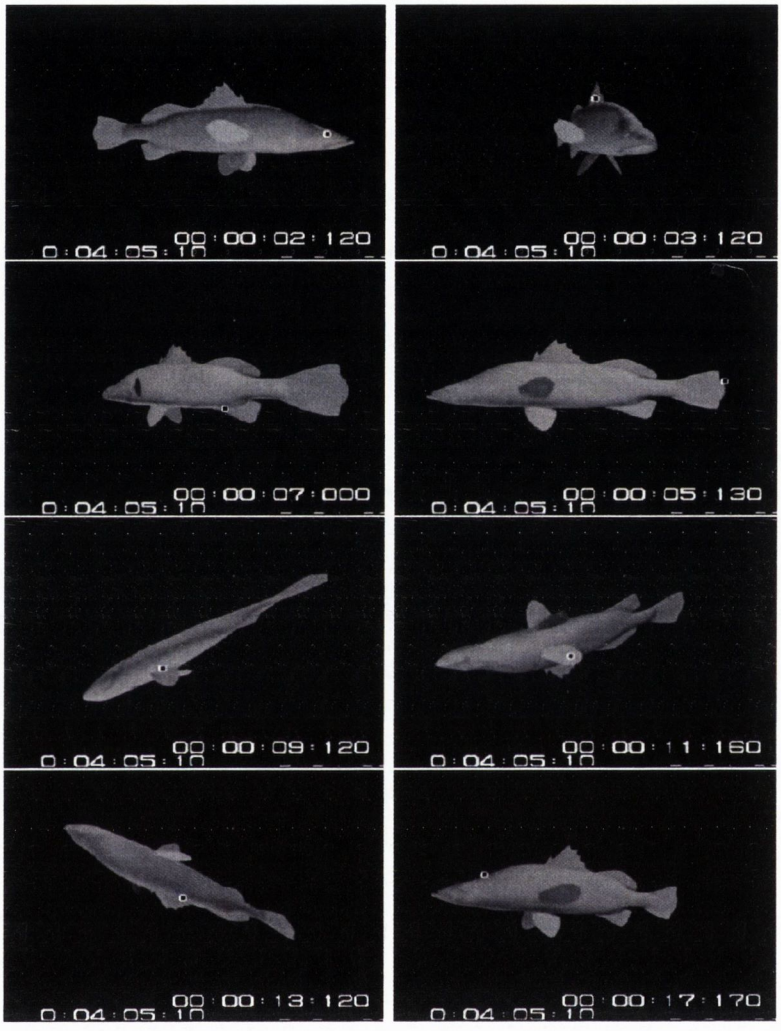

Figure 4.12: Images of the Video Curvid Overlay of one participant on a fish object.

For the gears, the only symmetric objects, it was not clear that there were any prominent features. Attention was more widespread over different parts of the model, suggesting that this method may not be suitable for symmetric objects (see Figure 4.13).

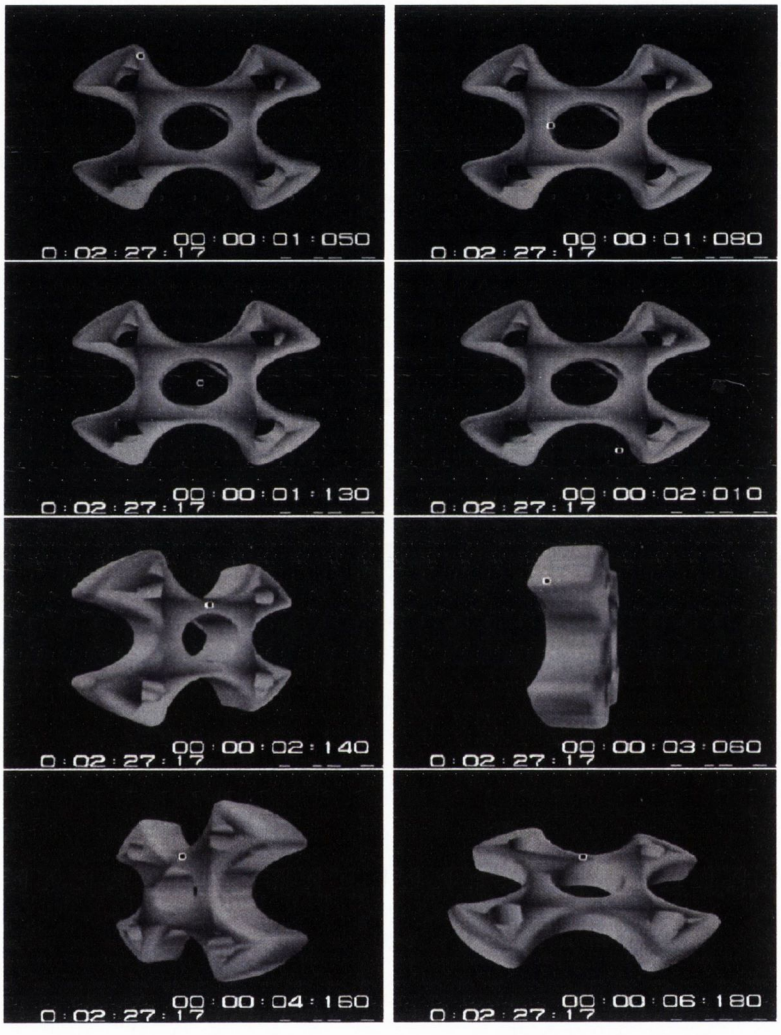

Figure 4.13: Images of the Video Curvid Overlay of one participant on an animal object.

4.6 Modified quadric error metric

Having obtained the fixation data, the next step was to incorporate the data from the saliency experiment into the quadric error metric in order to make it possible to perform perceptually guided simplification. In order to do this, the QSlim software developed by Garland and Heckbert [GH97] was altered by John Hamill [HHO04] to incorporate the information regarding fixations that we found.

4.6.1 Quadric error metric and modifications

The original method proposed by Garland and Heckbert [GH97] utilises iterative vertex pair contraction guided by a Quadric Error Metric (see Figure 4.14). The method calculates a quadric **Q** for each vertex in the initial model, which is the sum of squared distances to planes of that vertex and the planes of faces meeting at the vertex. See Garland and Heckbert [GH97] for a full description of Quadrics and their properties. Valid pairs of vertices for contraction are chosen from those vertices linked by an edge, or those whose separation is below a user-defined threshold.

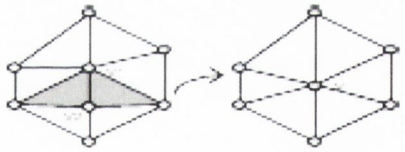

Figure 4.14: **Pair Contraction** - Selected Vertices are contracted to a single point. Shaded Triangles become degenerate and are removed.

The main algorithm then follows this sequence:

1. All valid pairs (v_1, v_2) suitable for contraction are selected.

2. An optimal contraction point $\overline{v}$ for each pair is computed. Its quadric $\overline{Q} = Q_1 + Q_2$ is the cost of contraction of the pair.

3. All pairs are inserted into a heap and sorted by contraction cost $\overline{Q}$.

4. Pairs are removed and contracted by cost, and neighbouring pairs have their costs updated.

5. Steps 3 and 4 are continued until the model reaches the desired level of simplification.

4.6.2 Modification to the quadric error metric

With saliency data acquired from the eye-tracker, a modified quadric error metric which incorporated this data was created. The method chosen was to weight the quadrics of vertices in the initial model based on a combination of the eye data captured by the eye-tracker. As in the original algorithm the quadric for each vertex is calculated from a combination of the quadrics for all faces that contain that vertex. Therefore, as the captured fixation data was based on the faces of the evaluated model, the weighting was applied to quadric of the face before the summation took place.

For each face in the initial model the following equation was applied:

$$Q_w = Q_f + \omega(F_f)$$

Where Q_w is the *Weighted* quadric produced, Q_v is the initial quadric for that face and $\omega(F_f)$ is the *Weight* associated with the face. Following this, the weighted quadrics for all faces that contain a vertex were summed in order to find the weighted quadric for each vertex.

The weight $\omega(F_f)$ is derived from a combination of data consisting of the total number of fixations on a face, the total duration of all such fixations and the duration of the first fixation on a face. Despite the fact that all three of these fixation values correlated very well for all object types (see Figure 4.15), we still decided to use a combination of all three metrics in our weighting. This decision

was based upon a quick survey we carried out. A group of 10 people were shown examples of models simplified using a weighting from each individual metric and a combination of all three.

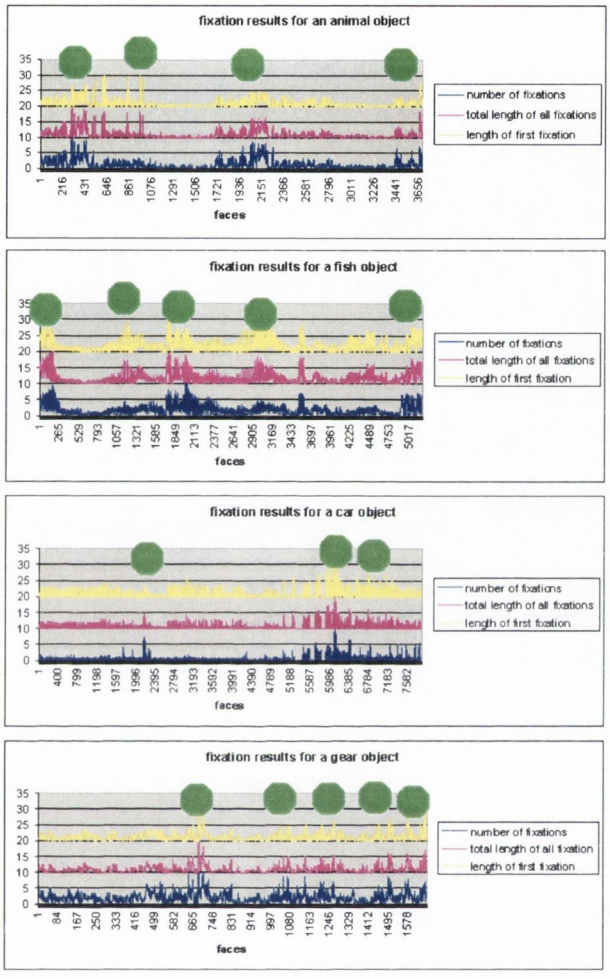

Figure 4.15: Fixation results for a typical animal, fish, car and gear model respectively. These show the correlation between the three metrics we used. The green dots indicates the areas that received the most attention.

The models simplified using all three metrics were preferred by the majority of people. However, it should be noted that other combinations of these metrics or a more sophisticated approach to integrating the results of saliency guided simplification into QSlim, similar to that of Kho and Garland [KG03], might also be effective, but further testing would be needed to investigate this.

For the three metrics, each value was normalised by the maximum value obtained for that metric and combined as follows:

$$\omega(F_f) = \overline{TotalFix} + \overline{DurationAllFix} + \overline{DurationFirstFix}$$

This weighted metric was applied to the QSlim 1.0 implementation of Garland and Heckbert's quadric based simplification. Data files generated from EyeLink data were associated with models and loaded into the QSlim program to weight the simplification process. Following this we evaluated the quality of the models simplified using both simplification types; the modified version of QSlim which produced perceptually guided simplified models (modified) and the original version of the QSlim 1.0 software (original).

4.7 Concluding comments

Having found the aspects of these models that received the most visual attention, by measuring how much participants fixated on each triangle in the mesh, the information was then incorporated into a simplification algorithm. Two sets of models were then simplified to various LODs using both this new form of simplification and the original version of the QSlim software. Following this, an evaluation was carried out, using some experimental measures of visual fidelity. The visual fidelity of models created using both forms of simplification were compared.

Chapter 5

Evaluation

5.1 Introduction

For measuring the visual fidelity of our simplified models, we had the choice of automatic or experimental techniques as described in section 3.5. As our aim was to find the objects with the highest visual quality as determined by the human observer, experimental measures were more appropriate. The first experimental measure of quality used was naming time on a set of familiar objects. It is the best indicator of ease of recognition and it has been shown that automatic measures are not successful at predicting naming time [WFM00]. In addition, Watson *et al.* [WFM01], had carried out a similar set of experiments in which they used this metric. The second indicator used was picture-picture matching time [LBD02] to determine if familiarity played a role. We chose this because our seconds set of models contained unfamiliar objects, therefore, naming times could not be used. Forced-choice preference, also used by Watson *et al.* [WFM01], was the final experimental measure used. It was possible to obtain relative judgements using both sets of models. We used two way forced-choice preferences, which are more sensitive because they force a decision to be made. Even when unsure participants have to choose which of two stimuli they prefer. Alternative metrics that we could

have used include ratings, 3 way forced-choice preferences or forced-choice preferences with confidence ratings. We wished to determine if there was a significant decrease in the naming or picture-picture matching times or a preference towards the models simplified using the fixation data.

5.2 Finding the naming times

5.2.1 Introduction

As described in Section 3.5.3, experimental measures include forced-choice preferences, ratings and naming times, all described in detail by Watson *et al.* [WFM01]. For our first experiment, the measure we used was naming time. Watson *et al.* [WFM00] carried out experiments to confirm that naming times are affected by model simplification. They present evidence that naming times are sensitive to simplification and model quality. In this experiment, naming times were used as a measurement of visual quality. Naming involves someone seeing an object and then verbalising the name that described that object, so the objects had to be of a familiar nature. Watson *et al.* carried out two sets of naming time experiments [WFM00, WFM01]. Using the same stimuli as Watson *et al.* [WFM01] plus one additional model, we carried out a similar experiment to examine if naming time is an accurate measure of model quality and how results are affected by object type. Furthermore, in our experiments we also used stimuli created by reducing these models to a much lower detail level than Watson (see Figures 5.1 and 5.2).

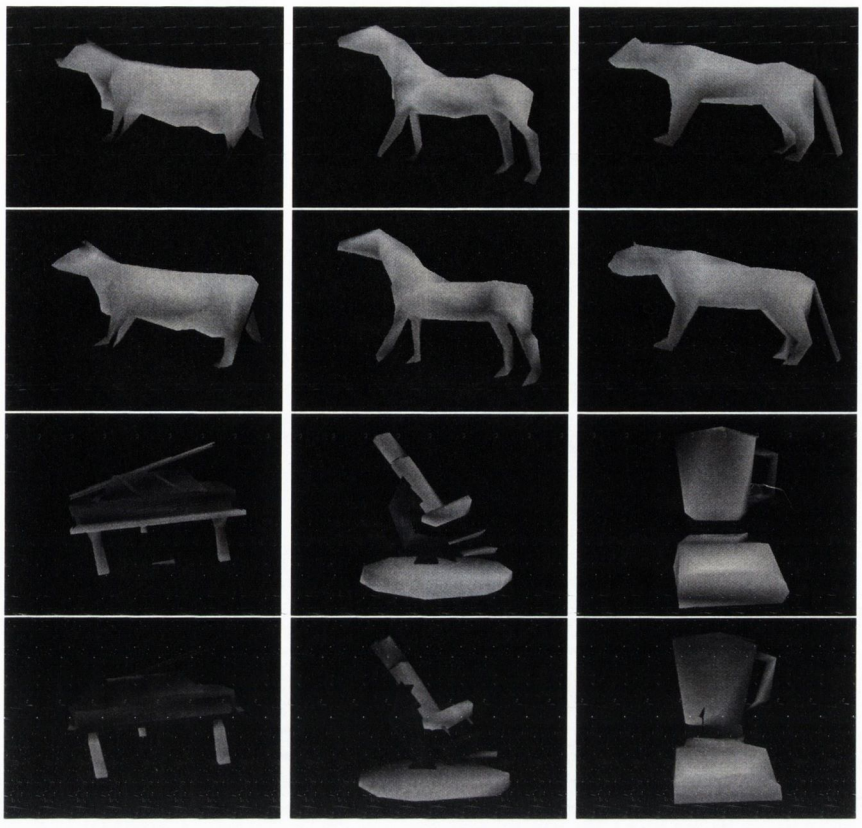

Figure 5.1: Natural objects simplified to 5% LOD using the original (1st row) and modified (2nd row) simplification approach. Man-made artifacts simplified to 5% LOD using the original (3rd row) and modified (4th row) simplification approach.

Bearing in mind that we used more versions of each model, we tried to follow their experimental method and evaluation procedure as accurately as possible. Finally, we investigated if the visual fidelity of the models was improved by using captured saliency data.

5.2.2 Participants and apparatus

Participants consisted of 27 volunteers, undergraduate and graduate students from the Computer Science department; 21 male and 6 female. All were naïve partici-

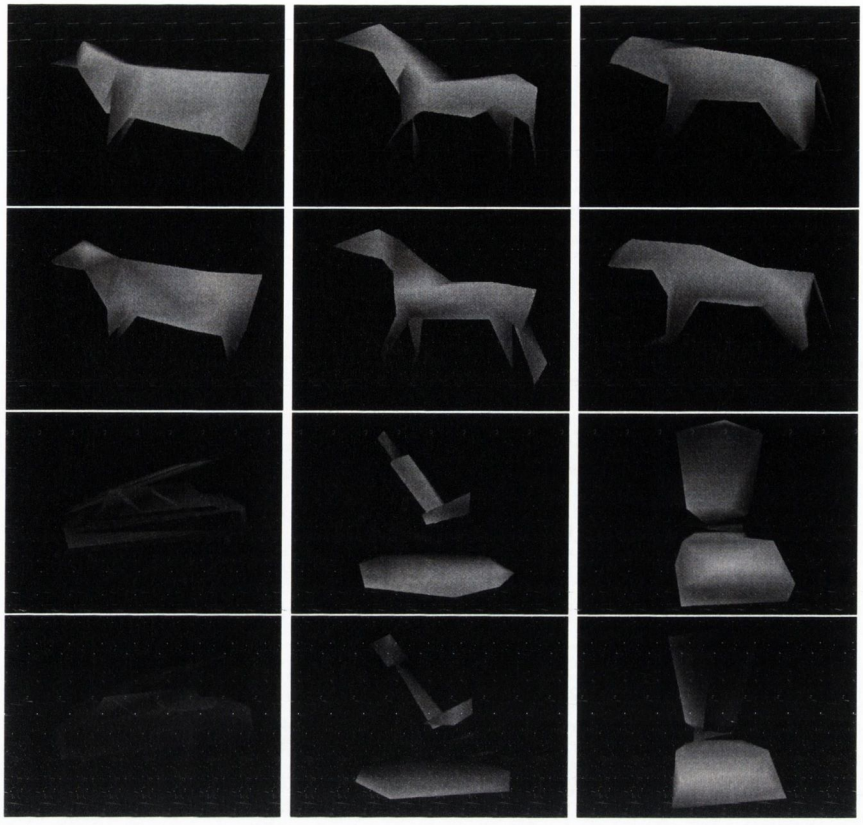

Figure 5.2: Natural object simplified to 2% LOD using the original (1st row) and modified (2nd row) simplification approach. Man-made artifacts simplified to 2% LOD using the original (1st row) and modified (2nd row) simplification approach.

pants with either normal or corrected to normal vision.

Stimuli consisted of the 37 familiar 3D polygonal models used in the previous experiment. Using 3D Studio Max, all models were rotated in order to achieve a canonical or optimal view. As described before, all 37 models were simplified using QSlim to have a standard 3700 polygons. Firstly, a set of models was made by simplifying the standard to various levels: to have 50% (*i.e.,* 1850 polygons), 20%, 5% and 2%, using the original version of QSlim. Secondly, a similar set of models was created, but this time using the software that took fixation data as well as geometry into consideration during the simplification process. There were

nine examples of each model giving a total of 333 stimuli.

5.2.3 Method

Prior to each experiment there was a test run. Stimuli for the test run were different from the experimental stimuli and these were present at different levels of detail (LOD). Each participant saw a total of eight models during the test run so that they clearly understood the procedure. Each of the 27 participants viewed a total of 37 models in which there was only one representation of each model. Therefore it took 9 participants to view all 333 stimuli once. Each participant saw at least four objects from each of the nine possible scenarios of simplification (including the standard models, and the two simplification types over the four simplification levels) and no more than five from any one scenario. The models within each experiment were then randomised and were static *i.e.,* participants were not permitted to rotate the models.

Participants viewed the diffuse-shaded models on a 21-inch monitor and a Labtec AM-22 microphone was used to obtain the naming times (see Figure 5.3). Participants were told that they would see a set of familiar objects made up of natural objects and man-made artifacts at various levels of detail. They held the microphone themselves and were told to name the models as quickly and as accurately as possible. They were also informed that some of the stimuli would appear very simplified. There were 37 trials in each experiment. A trial involved the experimenter pressing a key and a fixation cross appearing for a short time, the model appearing on the screen, the participant verbalising the name of the model, which triggered the microphone so the naming time could be recorded. Following this, the object disappeared and the experimenter, by pressing the appropriate button, recorded the accuracy of the response and caused the next trial to begin.

Figure 5.3: An example of a participant performing the naming time experiment.

5.2.4 Results

Participants each viewed a total of 37 models, but were only allowed to view one of the nine possible versions of a specific model. This was for familiarity control, as viewing a stimulus once reduces its subsequent naming time. Therefore, in order for all 27 participants to see some models simplified under each of the nine possible scenarios and for all 9 versions of the 37 models (*i.e.,* 333 models) to be viewed the same number of times, we designed the experiment so that no

two participants viewed the exact same set of models. All versions of the 37 models were named by three different random participants, essentially making the model condition a between-subject condition. We therefore applied between subject ANOVAs (ANalysis Of VAriance across groups) to all of the results. So, for example, during the evaluation of the effect of simplification type on results, although the same participants did not see the exact same models, the participants whose results were compared all saw some selection of models of the same type and at the same LOD and each of the models compared was viewed by three different participants.

We recorded the naming time and the number of incorrectly named objects. We examined how results were affected by simplification level, object type and simplification type. The number of incorrectly named objects made up 11.7% of all results. Spoiled trials, which occurred when the participant failed to trigger the microphone or triggered the microphone accidentally, made up 4.9% of all results. 58.1% and 25.6% of all incorrectly named objects were those at 2% and 5% respectively. Incorrectly named objects and spoiled trials were excluded from the naming time results. The near misses, which were acceptable as correct, occurred when similar names within a semantic category were used *e.g.,* when a hound was called a dog.

Unlike Watson *et al.* [WFM01], we found that only results at low LODs were significantly affected by level of simplification. Between 20% and 5% and between 5% and 2% there was an effect of simplification level on results and there was a significant increase in the naming times and the number of incorrectly named objects at low LODs when averaged by participants or objects (see Table 5.1 and 5.2). When comparing by object type there was an interaction effect, at 100% detail on naming time. Results averaged by either participants or objects, showed that it took significantly longer to name natural objects than man-made artifacts (see Table 5.3). This replicates previous psychological research, including that by Watson *et al.* [WFM00]. We found only one significant effect of simplification

AVERAGED BY	LEVEL OF DETAIL	ANOVA	P-VALUE
objects	20% and 5%	$F(1,52) = 5.73$	0.02
objects	5% and 2%	$F(1,52) = 4.42$	0.04
participants	20% and 5%	$F(1,36) = 7.33$	0.01
participants	5% and 2%	$F(1,34) = 8.25$	< 0.01

Table 5.1: The effects of simplification level on the naming time results.

AVERAGED BY	LEVEL OF DETAIL	ANOVA	P-VALUE
objects	20% and 5%	$F(1,52) = 11.35$	0.02
objects	5% and 2%	$F(1,52) = 48.24$	0.04
participants	20% and 5%	$F(1,36) = 04.95$	0.03
participants	5% and 2%	$F(1,36) = 17.63$	< 0.01

Table 5.2: The effects of simplification level on the number of errors in the naming time experiment.

type. There was an interaction effect between LOD and simplification type on naming time for the natural objects at a very low LOD. At 2% LOD, when averaged by objects or participants, there was a reduction in the naming time when modified QSlim was used (see Table 5.3).

VARIABLE	AVERAGED BY	LEVEL OF DETAIL	ANOVA	P-VALUE
object type	objects	100%	F(1,46) = 5.29	0.03
object type	participants	100%	F(1,34) = 6.42	0.02
simplification type	objects	2%	F(1,32) = 4.54	0.04
simplification type	participants	2%	F(1,24) = 3.77	0.06

Table 5.3: The effects of object type and simplification type on the naming time results.

We found that overall results were only affected by simplification level at low LODs, suggesting that naming time may not be a good indicator of fidelity in these circumstances. Further results show that, for natural objects at very low detail, saliency information retained can improve visual fidelity (see Figure 5.4).

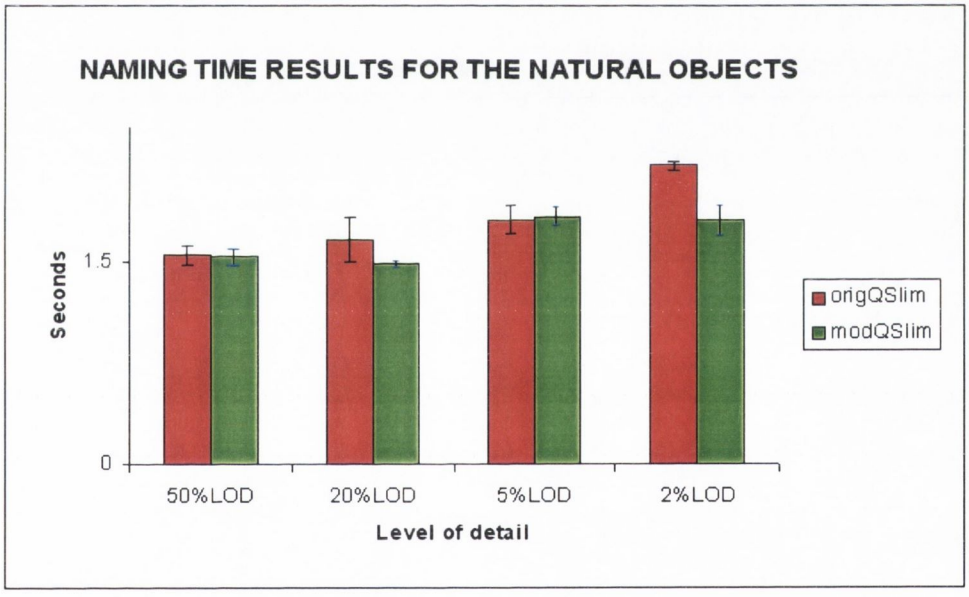

Figure 5.4: Naming times for the natural objects.

5.2.5 Discussion

Although in our particular case, naming time did not indicate great differences in visual fidelity, these experiments still show promising results for natural objects at low LODs, which merit further study. In this situation, it was possible to improve the visual fidelity of these familiar natural objects when saliency data, ascertained through eye-tracking, was considered during the simplification procedure. Following these results for familiar natural objects, we were interested to see if the familiarity of these objects had any effect on the results. Naming time could not be used here, as objects have to be of a familiar nature in order to be named. This leads us to the next set of evaluation experiments, in which we used picture-picture matching times as a measurement of visual fidelity, to examine different categorical effects.

5.3 Acquiring the picture-picture matching times

5.3.1 Introduction

We evaluated picture-picture matching time as a measure of visual quality and compared categories, while bearing in mind that the number of polygons at each LOD was not uniform, and examined the effects of familiarity. Finally, and most importantly, we compared the matching results to determine if there was any improvement when the saliency data was used during simplification. The idea was to have the objects in each category as similar as possible. All the animals were four legged creatures, while the fish were all roughly the same shape with mostly the fins being the distinguishing characteristics and similarly for the cars and gears. This meant that, at the lower LOD's, objects within a category were hard to distinguish from each other.

Picture-picture matching involves matching two pictures presented simultaneously with no verbalisation. We used picture-picture matching rather than naming

times here because most of these models were not familiar. Participants could not be expected to know or even remember the names of these objects as that would require an expert in the given field. Lawson *et al.* [LBD02] used this measurement in experiments on matching similarly and dissimilarly shaped morphs from different as well as identical views. Picture-picture matching is commonly used in research on participants with mental retardation [DMdMT03, GSM97]. In our experiment, the participant had to choose which of the two images of the simplified models was most similar to the image of that model at full LOD. The sample stimuli appeared on the screen and the comparison pictures on a sheet of paper. This process does lead to high response times and the difference between the luminance on the display and on the sheet might also have effected this. However, the length of time is not relevant to our study as it is the relative difference in performance across our two conditions that we are interested in.

5.3.2 Participants and apparatus

A total of 28 participants were involved in this experiment, half for the original simplification method and half for the modified version, ages ranging between 19 and 27 from various backgrounds. There were 18 males and 10 females with either normal or corrected to normal vision. Some of these participants had taken part in the experiment to find the salient features of these models, using the eye-tracking device. Those who had not taken part first viewed the models using an identical procedure for the same amount of time (only without using the eye-tracker), in order to counteract learning effects and for familiarity control.

We used the set of 30 models on which the saliency data had been acquired. The four categories of models, as described in the previous section, were prepared under the headings of animals, cars, fish and gears. The animal objects were a subset of the natural object set used in the naming time experiment. The animals and the fish categories had five detail levels 100%, 30%, 14%, 5% and 2%. Within

each category the number of faces an object had at each level was uniform but not across categories. This was because the idea was to have models that were accurate representations of the objects. For example, fewer polygons would be needed to make a good animal model than a more complex model such as a car. Therefore all animal objects at 100% had 3700 faces and at 30% had 1110 faces and so forth. At 100% or highest LOD the fish models had 5200 faces. Initially the car models were rendered at the same percentage LODs with 7868 faces being the highest level. However, after some test runs were carried out, it was obvious that even with high detail it took quite a long time to recognise the individual cars and at the lowest detail they were no longer recognisable as cars. So it was decided that the four levels the cars should be rendered at were 100%, 75%, 50% and 25%. In the final category, the objects called gears were also shown at four LODs, 100% (1658 polygons), 30%, 14% and 5%. Again, these models were displayed using diffuse shading on a 21-inch monitor.

There were two versions of this experiment, one for each type of simplification, with identical procedures. With each of the 30 models rendered at the different levels, each participant viewed a total of 135 stimuli. These were divided into four different blocks, one for each category. Within each category the models were randomised *i.e,* all LODs were mixed up within their own category only. All models were static.

5.3.3 Method

Participants were seated in front of the computer and given print-outs containing screen shots of the models as they appeared only at the highest LOD (see Figure 5.5). Beside each model was a name and a number. Taking one category at a time, participants were told to complete the task. This involved viewing the models on the screen one at a time, comparing them to those on the sheet and finally pressing the number on the keyboard assigned to that particular model on

the sheet. Participants were told to press the correct button as accurately and as quickly as possible. As soon as the button was pressed, a new model appeared, and this process was repeated until each model had been displayed once at each LOD in a random order. After each category was displayed on the screen, there was a small pause when the paper copies were replaced with those displaying the new category. (Although the different luminance's between the screen images and the sheets might have had an effect, So perhaps in the future, if a similar experiment was being carried out, it would be more practical to use a second screen instead of the paper copies.)

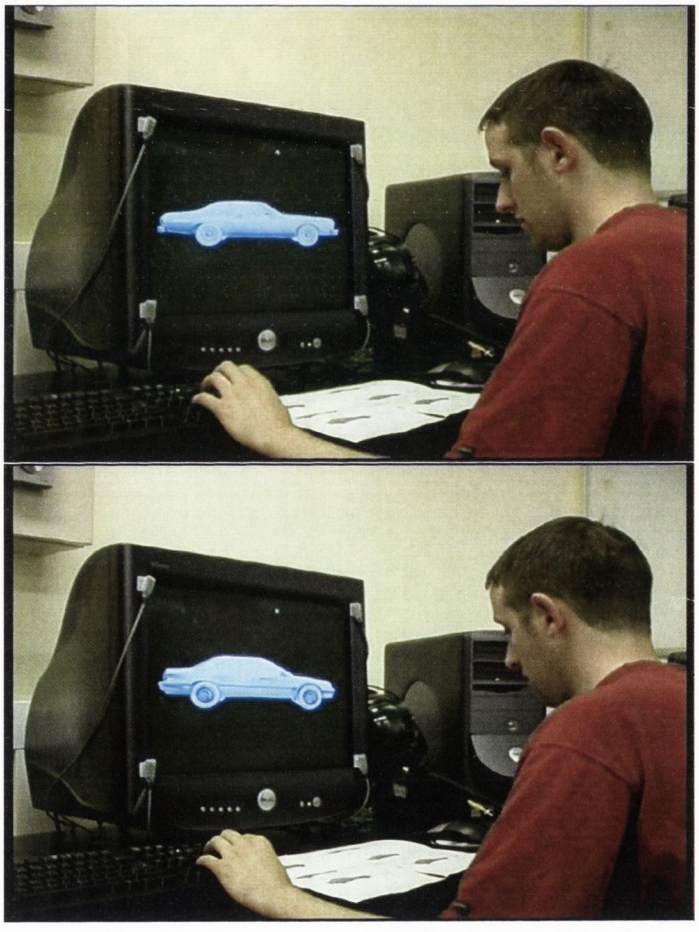

Figure 5.5: An example of a participant performing the matching time experiment.

5.3.4 Results

We recorded the average matching times and the number of correctly matched objects. We used a split-plot ANOVA design (*i.e.*, between subject ANOVAs for the simplification type factor and within subject ANOVAs for the simplification level and object type factors). No statistically significant differences in results were obtained for the car models. The results averaged over simplification type for the animal, fish and gear models were affected by simplification level at the lower LODs.

For the animal models between 14% and 5% LOD, there was a significant increase in the matching times when averaged by objects and participants. For these models between 5% and 2% LOD, when averaged by objects there was a significant difference and a marginally significant one when averaged by participants. For the fish objects between 5% and 2% LOD, when averaged by objects and participants there were marginally significant results. For the gear objects between 14% and 5%, when averaged by objects and participants there was a significant result. Between 5% and 2% when averaged by objects there was a significant result (see Table 5.4).

Regarding the number of correctly matched objects; for the animal models averaged by objects there was a significant decrease between 14% and 5% LOD and between 5% and 2%. For the fish objects, averaged by object there was a significant result between 5% and 2% LOD. For the gear objects between 14% and 5% there were significant results when averaged by objects and marginally significant results when averaged by participants. Again when averaged by objects, between 5% and 2% there was a significant decrease (see Table 5.5).

Next, bearing in mind that the number of polygons was not uniform across categories or LODs, we compared all four categories averaged over the first four LODs. There was a significant difference in the matching times for all categories except the fish and gears (P-value < 0.05). The animal objects were the fastest

AVERAGED BY	LEVEL OF DETAIL	ANOVA	P-VALUE
objects	Animals 14% and 5%	$F(1,26) = 06.79$	0.01
participants	Animals 14% and 5%	$F(1,12) = 04.79$	0.04
objects	Animals 5% and 2%	$F(1,26) = 07.35$	0.01
participants	Animals 5% and 2%	$F(1,12) = 03.21$	0.09
objects	Fish 5% and 2%	$F(1,26) = 04.20$	0.05
participants	Fish 5% and 2%	$F(1,14) = 03.38$	0.09
objects	Gears 14% and 5%	$F(1,26) = 13.49$	< 0.01
participants	Gears 14% and 5%	$F(1,14) = 06.01$	0.02
objects	Gears 5% and 2%	$F(1,26) = 06.80$	0.01

Table 5.4: The effects of simplification level on the matching time.

AVERAGED BY	OBJECTS	LEVEL OF DETAIL	ANOVA	P-VALUE
objects	Animals	14% and 5%	$F(1,26) = 06.65$	0.01
objects	Animals	5% and 2%	$F(1,26) = 13.70$	< 0.01
objects	Fish	5% and 2%	$F(1,26) = 07.95$	< 0.01
objects	Gears	14% and 5%	$F(1,26) = 21.92$	< 0.01
participants	Gears	14% and 5%	$F(1,14) = 03.58$	0.08
objects	Gears	5% and 2%	$F(1,26) = 25.86$	< 0.01

Table 5.5: The effects of simplification level on the number of correctly matched objects.

AVERAGED BY	OBJECTS	LEVEL OF DETAIL	ANOVA	P-VALUE
objects	Animals	14%	F(1,26) = 03.68	0.07
objects	Animals	5%	F(1,26) = 06.06	0.02
participants	Animals	5%	F(1,12) = 03.70	0.08
objects	Animals	2%	F(1,26) = 13.14	< 0.01

Table 5.6: The effects of simplification type on the results for matching time.

to be matched in 3.14 sec, then the fish (4.51 sec), then the gears (4.74 sec) and the cars were the slowest (6.40 sec).

Regarding simplification type, there was a marginally significant reduction in the matching time for the animal models when averaged by objects at 14% (see Figure 5.6) when modified QSlim was used and a significant reduction at 5% and 2% . When averaged by participant at 5% there was also a marginally significant reduction (see Table 5.6).

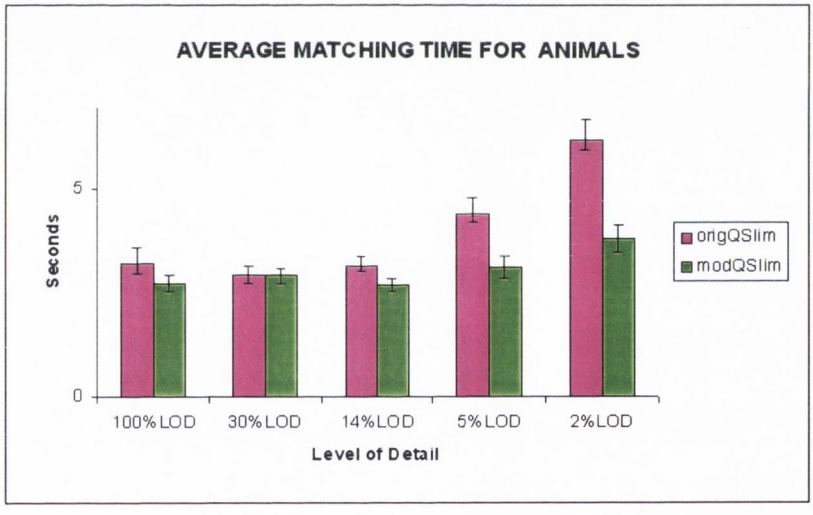

Figure 5.6: Comparing the average matching times for the animal models.

Results for the number of correctly matched animal objects (see Figure 5.7) at 14% averaged by objects show a marginally significant increase in the number of

AVERAGED BY	OBJECTS	LEVEL OF DETAIL	ANOVA	P-VALUE
objects	Animals	14%	F(1,26) = 03.96	0.06
objects	Animals	5%	F(1,26) = 11.17	< 0.01
participants	Animals	5%	F(1,12) = 03.40	0.09
objects	Animals	2%	F(1,26) = 07.61	0.01
objects	Fish	30%	F(1,26) = 06.76	0.02

Table 5.7: The effects of simplification type on the results for the number of correctly matched objects.

correctly matched objects when modified QSlim was used. For the animal models averaged by objects at 5% and 2% there was a significant increase. Again at 5%, when averaged by participants, there was marginally significant increase. There was a significant increase in the number of correctly matched fish when averaged by objects at 30% (see Table 5.7).

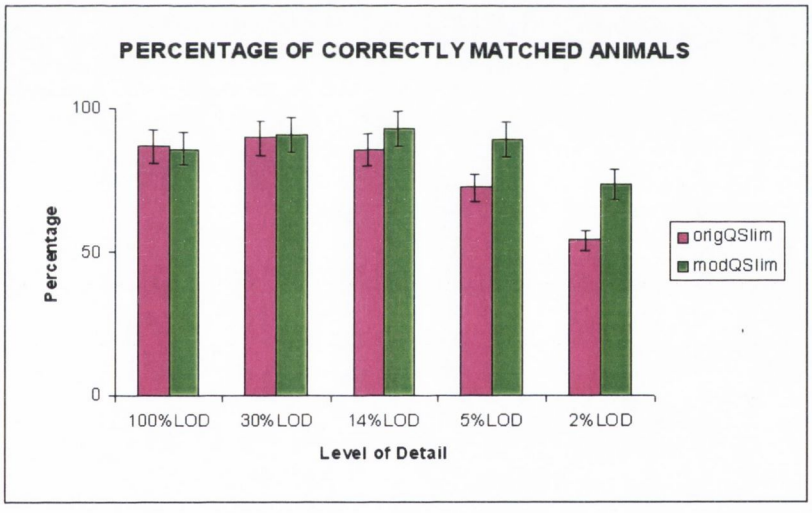

Figure 5.7: Comparing the percentage of correctly matched animal models.

5.3.5 Discussion

Matching time results show that, like naming time, there is an effect of simplification level only at the lower LODs. The lowest matching times were achieved for the animal models, possibly because these were the only familiar category of objects used in the picture-picture matching experiment. Moreover, they could be classified as a basic level category, as opposed to a subordinate one [Ros76]. For example, basic level objects (*e.g.,* chair, car, dog, kangaroo) are objects at the most inclusive level at which there are attributes common to most members of the category. At this basic level of abstraction, cue validity is maximised [TH84]. As described by Rosch [Ros76], a category with a high cue validity is more differentiated from other categories than one with low cue validity. Categories one level more abstract will be superordinate categories (*e.g.,* furniture, vehicle, animals) whose members share only a few attributes among each other. Subordinate categories have lower total cue validity than basic categories, because they share most attributes. According to Tversky [Tve77], they tend to be combined because the weight of the added common features tends to exceed the weight of the distinctive features. Categories below the basic level will have predictable attributes and functions and contain many attributes that overlap with other categories (for example, a Mercedes shares most of its attributes with other kinds of car).

At the lower LODs for the animal objects there was an interaction effect, i.e., variation among the differences between means for different levels of one factor over different levels of the other factor, as there were significantly less errors and significantly lower matching times for the animal models when the modified version of QSlim was used for simplification. These results further suggest that perceptually guided simplification can enhance the visual quality of natural objects from basic level categories at low details.

The results for the natural category of fish indicate that category level and familiarity play a role, as at 30% there is one significant result, perhaps because

below this level objects are too similar and cannot be distinguished. However, further tests would be needed to investigate this further.

There were no significant results for the car models. A reason might be that, even at the lowest LOD, these models were rendered at 25% of the original detail (this was however necessary due to the nature of the models). The car models were by far the slowest to be named even though they had the greatest amount of detail; this may be due to the category resemblance or the low cue validity. Perhaps the car models we used could be described as subordinate level objects because they share more attributes in common than the other categories and hence the low cue value. Perhaps, in the future it would be interesting to see if there differences in results between males and females for such a category of objects.

The lack of significant results for the gear objects, as suspected from the previous saliency determination experiment, could be resulting from the symmetry of the objects.

5.4 Forced-choice preferences experiments

5.4.1 Introduction

Finally, we carried out an experiment in which both sets of models, familiar and unfamiliar, could be included. The experimental technique used was forced-choice preference. Preferences obtain relative judgments; participants have to choose the stimulus with more of the experimenter-identified qualities, in this case similarity to the actual model. We used a web-based interface for this experiment (see Figure 5.8). All models under the two types of simplification were compared at the same simplification level.

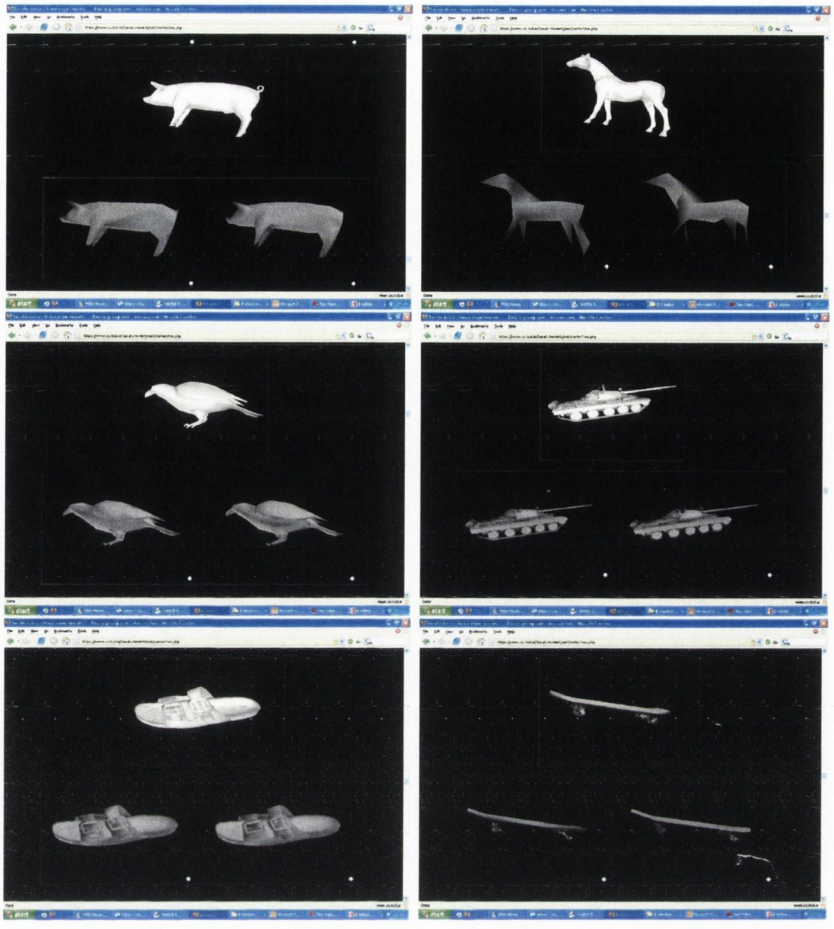

Figure 5.8: Screen shots of trials from the web-based forced-choice preference experiments.

5.4.2 Participants and apparatus

Sixty eight people participated in each part of this experiment. Sixty males and 8 females in the first part and 51 males and 17 females in the second part. There were both graduate students and staff from the Computer Science department. All had either normal or corrected to normal vision.

There were two separate web-based experiments. Stimuli for the first one included two types of images, screen shots of the stimuli from the naming time

experiment (those of natural objects and man-made artifacts, see Section 5.2.2). Images were created from the standard and the simplified models. Images of the models created using the original QSlim and the modified version of the software were compared to the standard at the four simplification levels; 2%, 5%, 20% and 50%. There were 37 different models and four different levels giving 148 unique comparisons.

To prevent repeated exposure to the same model, each participant saw only one version of each model *i.e.,* a total of 37. Therefore we needed four different versions of the experiment to cover all the comparisons, each set having one quarter of its images from each of the four LODs. These four sets contained 10 different random orderings of the models, giving rise to 40 unique web pages, which were assigned to participants in sequence. On each page, half of the models simplified using the original version QSlim were on the left and half on the right, in random order. The left and right positions of the models simplified using the original version of QSlim were distributed evenly throughout the different pages.

5.4.3 Method

Participants, on going to the web page, carried out the version of the experiment that they were assigned. Each participant had to make 37 choices. Participants were asked to choose which of the two images of the simplified models was more similar to the image of that model at 100% detail, which was displayed on top in the centre (see Figure 5.9). The two simplified versions (original and modified) were displayed below, side by side. Participants entered their responses by checking the left or right box. Then the participant scrolled down to the next set. The web address of the experiment was sent via e-mail. Each person to visit the page was assigned one of the 40 versions of the experiment. They were asked to give some additional information including name, age, gender and vision quality for validity and statistical purposes. Their identity was validated and only gen-

uine entries were accepted. Participants therefore viewed the images on a range of display sizes and resolutions. We examined results from the first 68 genuine entries.

Figure 5.9: An example of a participant performing the forced-choice preference experiment.

In the second experiment, there were three types of images, those of fish, cars and gears. As before, the images were screen shots of the unfamiliar models used in the picture-picture matching time experiment and simplified as before using the original and the modified versions of QSlim. The fish models were compared at four levels, the car and the gear models at three. Again, it was in the form of an online experiment with the same design as before but on a smaller scale as

there were only 8 fish, 7 car and 6 gear models used. Each participant made their choices as in the previous forced-choice experiment and we examined results from the first 68 genuine entries.

5.4.4 Results

We applied single-factor within-subject ANOVAs on the results. For the first experiment, less than 0.7% of all results had to be excluded where participants failed to choose either of the images. Results were averaged by participants (see Figures 5.10, 5.11 and 5.12). We found an interaction effect of simplification type on the preference results. For the natural objects at 50%, 5% and 2% , there is a strong preference for the models simplified using the modified version over the original version of the QSlim software.

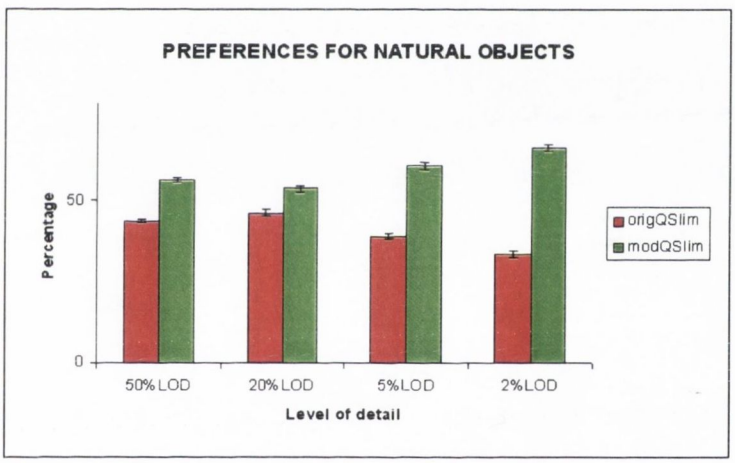

Figure 5.10: Percentage preferences for the natural objects.

However, results for the man-made artifacts show that marginally significantly more people chose the models simplified using the original version of QSlim at 20% and significantly more chose them at the 5% and 2% LODs. In the second web-based experiment, less than 0.9% of results had to be excluded. The only significant result was an interaction effect that showed that, in the case of the

fish objects at the lower levels, there was a significant preference for the models simplified using modified QSlim (see Table 5.8).

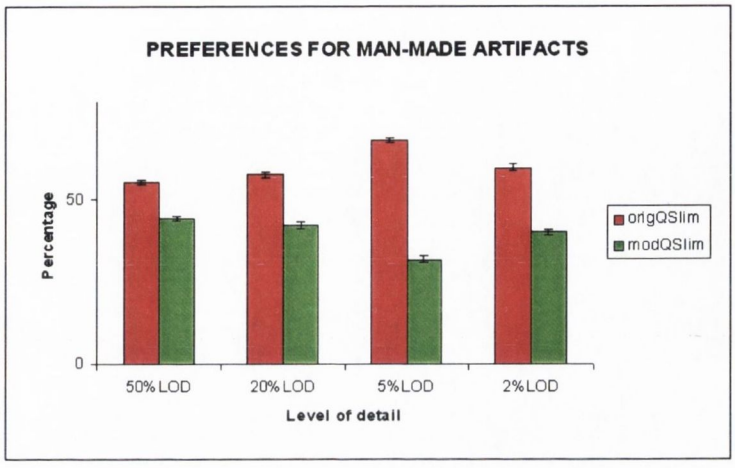

Figure 5.11: Percentage preferences for the man-made artifacts.

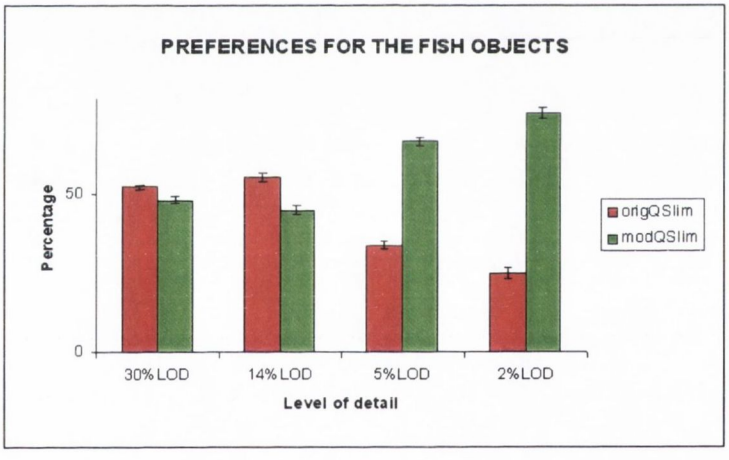

Figure 5.12: Percentage preferences for the fish objects.

PREFERENCE	AVERAGED BY	OBJECT	LEVEL OF DETAIL	ANOVA
modified simplification	participants	natural	50%	F(1,36) = 07.7
modified simplification	participants	natural	5%	F(1,36) = 05.8
modified simplification	participants	natural	2%	F(1,36) = 20.9
original simplification	participants	man-made	5%	F(1,34) = 07.4
original simplification	participants	man-made	2%	F(1,34) = 24.2
modified simplification	participants	fish	5%	F(1,14) = 09.9
modified simplification	participants	fish	2%	F(1,14) = 08.9

Table 5.8: The significant effects of simplification type on the preferences (All P-values < 0.05).

5.4.5 Discussion

Forced-choice preferences responded more strongly than the other predictors used in the experiments. Unlike naming, preferences do not measure ease of recognition, but measure the visual fidelity by judging visual similarity to the original. This metric is a particularly useful predictor in our study because it forces the participant to make a relative judgement, returning results that may be lost by using other metrics. In this case, forced-choice preferences provide us with the data we need; it represents the conscious decision of the participant as to which simplified model is more visually similar to that of the original one. It demonstrates that saliency guided simplification can improve the appearance of unfamiliar natural objects as well as familiar ones, which was not apparent from the matching time results. It also produces some preferences at higher LODs for the modified natural objects and the original man-made artifacts.

Importantly results show that, while saliency based simplification does work for natural objects, it actually reduces the visual quality of familiar man-made artifacts, a reason for this may be that in a lot of cases, man-made artifacts are related to a task and that prominent features may be defined by this. A wide

variety of psychological research [HSMP03, LH01a, SHS01, PHL01] indicates that when tasks are involved attention is generally top-down, and not affected by the bottom-up salient properties. For example, when a participant's eye-movements were tracked while making a snack [Hay00], results showed that almost all of the fixations focussed on the task, rarely focussing elsewhere; suggesting that visual activity is largely controlled by the task, so various tasks would mean various different sets of prominent features. Cater *et al.* [CCW03] also recently showed how task semantics can be used for selective rendering of scenes. Results also confirm our initial hypothesis that this method would not work so well on the symmetric gear objects.

5.5 Concluding comments

Our evaluation results suggest that simplifying to a low LOD using perceptual information to guide the process can enhance the visual quality of natural objects at a low LOD. However, results demonstrate that for the man-made artifacts this was not the case. One possible reason for this might be that we adapted QSlim to incorporate the perceptual information. Perhaps, if another tool (*e.g.,* Vclust) had been used, this might have resulted in more positive results for the man-made artifacts. Another reason that suggests this could be the case, are the findings of Watson *et al.* [WFM01]. They point out that natural objects were more like the standard when simplified using QSlim. However, in the case of the man-made artifacts, objects were judged more similar to the standard when Vclust was used to produce the simplified models. This deserves further investigation and it would be interesting to see if different results could be found if Vclust, as opposed to QSlim, was incorporated with the fixation data. Moreover, a selective simplification algorithm could be developed in which, depending upon the object type and attributes, different methods of simplification could be applied in order to produce the best solution.

We decided as a further validation step to use the eye-tracker to gather the fixation data for participants while actually performing a version of these tasks, as it would be interesting to see if the prominent features determined through just viewing the objects are the same as those found during the tasks of naming, picture-picture matching and forced-choice preferences. We wished to confirm our previous results and determine whether the different natures of these tasks influenced where a participant fixated. Furthermore, we hoped that it would provide some better insights as to how salient features for man-made artifacts can be predicted. Perhaps, when there is some sort of task involved, like matching, prominent features will be studied and compared.

Chapter 6

Validation

6.1 Introduction

Results from both the initial saliency determination experiment, described in Chapter 3 and out evaluation studies, described in Chapter 4, indicated that the heads of natural objects were particularly important features. We decided, as a final confirmation step, that it would be interesting to investigate if the actual features focussed upon during the three tasks fit in with these previous findings. We also kept in mind the lack of positive results for the man-made artifacts and the suggestion that perhaps prominent features for these objects would be heavily dependent on task.

6.2 Background on face 'pop-out'

As discussed in Section 3.3, it is theorised by a great deal of researchers that attention is a two-staged selection mechanism. Processing in the pre-attentive stage of vision is thought to have unlimited capacity. Furthermore, it performs in parallel across the whole visual field; that is, it operates on a number of different locations at the same time in a scene. One interesting phenomenon here is the contrast effect, where stimuli defined by a single feature dimension appear to

'pop-out' of the display, regardless of their position. For example, one red dot among a set of green dots would attract the viewer's attention. The parallel levels of processing respond well to a few parts of the scene and poorly to everything else [IK00]. The responses at this level of processing are highly dependent on the context. Specifically, attention is attracted if stimuli are well contrasted with their neighbours instead of being considered in isolation. Finally, pre-attentive processing seems to behave in a bottom-up manner, independently of strategic control. The distribution of attention is therefore stimulus driven, as the scene draws attention to locations or objects of possible interest for further attentive processing.

'Pop-out' of this kind has been evident during visual search tasks [WCG94]. It can be demonstrated by showing that processing is done in parallel and pre-attentively, when an increase in the number of distractors results in a minimal increase in reaction time during a task. Research on the perception of human faces suggests that high-level representations such as faces are processed in a different manner to other objects, and that they activate unique cells in the brain.

One suggestion is that human faces have a 'pop-out' effect in a similar manner to other basic elements like colour. Brain-injured patients, brain imaging studies, and behavioural results all appear to indicate a face-specific recognition system. One type of brain injury, termed prosopagnosia, refers to a selective impairment in face recognition only, not in object recognition. It results in an inability to recognise individual faces [Far92]. Tarr [Tar00] presents further evidence of a specialised recognition processes for recognising faces. He shows that a prosopagnosic subject performed better at matching inverted faces than upright faces, the opposite of normal subjects. Farah *et al.* [FWDT94] provide additional evidence of this, where, in the case of a prosopagnosia patient, impairment was again greatest with upright faces. More recently, Hochstein *et al.* [HBH+04] and Hershler and Hochstein [HH03] found that human faces popped out from a background of varied photograph distractors. Furthermore, they demonstrated that this effect does

not generalise to animal faces.

In direct conflict with this, there is also evidence to suggest that faces are not accessed pre-attentively in parallel and do not 'pop-out' during visual search [BHF97, SC95]. In his experiments, Nothdurft [Not93] found that test reaction time increased steeply with sample size, therefore suggesting serial search and no evidence to support face 'pop-out'. In their tests to find out whether special face cells exist, Kuehn and Jolicoeur [KJ94] showed that none of the search conditions involving distractors containing face features resulted in 'pop-out' and only non face distractors allowed a face target to 'pop-out'. However, they conclude that they cannot rule out the possibility that humans may have cells that respond selectively to faces.

Despite much research that suggests that faces do not 'pop-out', there is much evidence that faces are treated differently to other objects by the human visual system, as there is a certain biological significance of human faces and our familiarity with them.

In their study, Suzuki and Cavanagh [SC95] used stimuli consisting of three features (up and down arcs) organised to form schematic faces or patterns of no significance. They showed that faces cannot be ignored when presented as search non-targets, and that facial organisation has precedence over combined low-level features, even during a feature search where low-level features would be more efficient. In the case of the 'change-blindness' paradigm, changing faces capture attention more than other types of changing objects [RRL01]. Brown *et al.* [BHF97] showed that the ability of participants to search for a face in peripheral vision could be learned by training, but only for upright faces. Even though they did not find evidence that faces 'pop-out' immediately, their work suggested that faces have a special status in tasks that require learning, *i.e.*, participants benefited from training with the stimuli.

Although not addressing the issue of face 'pop-out', Lavie *et al.* [LRR03] conclude that faces do play a special role in attention. In their work it is demonstrated

that irrelevant faces are especially distracting, interfering even under conditions of high attentional load, conditions that eliminate other kinds of interference. They present evidence that face processing may be automatic and mandatory, but their results are still inconclusive.

Utilising other object categories might also result in perceptual performance similar to those found for faces [GT97, GSGA00]. Further, this could be acquired for other stimuli after some practice. Some research suggests that the specialised cells thought to be only for human face processing may be activated by other classes of stimuli. Other research with prosopagnosia patients demonstrated impaired object recognition for objects and animals as well human faces [LC89]. This suggests that there might be a possibility that animals or maybe even animal faces could be treated in a similar fashion to human faces.

6.3 Validation experiments

6.3.1 Introduction

We carried out some further experiments to examine the eye-movements of participants while performing the tasks of naming, matching and making forced-choices. The idea was to examine if the prominent features found in the saliency experiment, which we took into consideration during simplification, were also the features that the participants focussed upon during the three types of experiments. However, for the naming and matching tasks, we made the decision to only use models at full levels of detail (LOD). Although using simplified models might have been quite interesting, we chose this approach because only models at full detail were used during the original saliency experiment. We were interested to see if the prominent features, such as the heads of the natural objects, the fins of the fish or the sides and the door handles of the cars still received a lot of attention. Furthermore, we were interested to see if the aspects of the objects that received

the most attention differed from task to task, as it is well known that task strongly affects eye-movements.

However, in the case of the forced-choice preference task, as in the original experiment, participants had to view two simplified models and choose which of these models they thought to be more similar to the original model. Otherwise, the task would have been meaningless, as both choices would have been identical.

The fundamental difference between the experimental tasks was that naming was more like a memory task, as opposed to the matching and preference experiments, which were comparison tasks. For naming, participants were required to look at an object and recall from memory what the name of that object was. Naming times measured ease of recognition. As there was no other requirement other than to look at the objects, we could examine if these objects had any particularly salient features that captured the participants' immediate or total attention, *e.g.*, the heads/faces of the natural objects, as it was likely that they would look for a specific salient feature which would identify that object. For the matching time and the forced-choice preference experiments, participants simply had to compare objects presented to them, with which they did not have to be familiar, so it was less likely that attention would be drawn to any specific features but be more spread out over the objects. Unlike naming, these two predictors measured visual difference rather than visual recognition. The difference between the matching and the preference task was that, for the matching task, the participant had to choose an identical model and there was an obvious choice, but for the preference task, participants had to choose which of two simplified models they thought to be more similar to the original model.

6.3.2 Participants and apparatus

Ten participants were involved in this experiment. There were 8 males and 2 females with ages ranging from 22 to 27 from various backgrounds. All participants

had either normal or corrected to normal vision. The 10 participants took part in all three experiments, naming time, picture-picture matching time and forced-choice preference, in random order. These experiments were carried out as described before with some adjustments. We did not record responses, as these were determined in the previous experiments. Here we were only interested in where the participants fixated while doing the task. Participants were therefore required to wear the eye-tracker while carrying out the experiments. There were less trials in these experiments, and within each experiment trials were randomised.

6.3.3 Method

There were 37 trials in the naming time experiment and we used images of the 37 models, consisting of 19 natural objects and 18 man-made artifacts at full LOD. In the validation experiment participants did not have to name any simplified models.

For the picture-picture matching experiment there were 20 trials, including five each for the animals, fish, cars and gears. All models were displayed on the screen, the example model on the left and the comparison models on the right, so comparing with models on a sheet of paper was not required. Otherwise, it would have been very difficult to accurately analyse fixation data. Furthermore, our aim was to compare results to the original saliency experiment which did not involve consulting any sheets of paper. Certainly the eye-movements will be quite different under these circumstances, but the areas of the objects fixated upon should not be, as the task still was to find the visual difference between objects and to match the ones that were the same. Again, during this experiment all models were at full detail for the same reason as stated above, so the task was to match the model on the left with the one on the right that was exactly the same. Another difference from the earlier experiment was that, instead of pressing a key, the choice was made by saying top, middle or bottom. This further reduced head

movements and other actions that could effect eye-movements as participants did not have to fixate on the keyboard but only on the screen.

For the forced-choice preference experiment there were 28 trials: 8 natural objects, 8 man-made artifacts, 4 fish, 4 cars and 4 gears. These were similar to the matching time experiments. Participants had to view the screen with the example model on the top, and two comparison models on the bottom of the screen and had to choose between the two by verbally specifying either left or right. The models were simplified to either 5% or 2% of the original detail using the modified and the original version of QSlim, similar to the original experiments. Before each experiment there was a calibration procedure and participants had to focus on a dot prior to each trial for drift correction.

6.3.4 Results

Results for the naming time experiments

The EyeLink data viewer was used to generate fixation maps which allowed a "landscape" view to be created for a group of trials with the same background image in order to identify the informative parts of the display. This tool allows the user to set the standard deviation of the Gaussian distribution for each fixation point when creating a map, set the contrast between the fixation hotspots and the background and set the number of standard deviations extended for each fixation point when creating the map. Using the system recommended default values, we generated fixation maps for each image combining the results from all the participants.

Firstly we examined all fixations on each of the natural objects (see Figure 6.1). We cannot claim that the heads of the animals 'popped-out' pre-attentively but they were definitely salient features which attracted almost all of the participants' attention, especially for the four-legged creatures, as was the case in the original saliency experiment. For models such as the ant, spider and raven this was also

the case. For the shark, fish and dolphin, the head/face area received a significant amount of the viewers' attention.

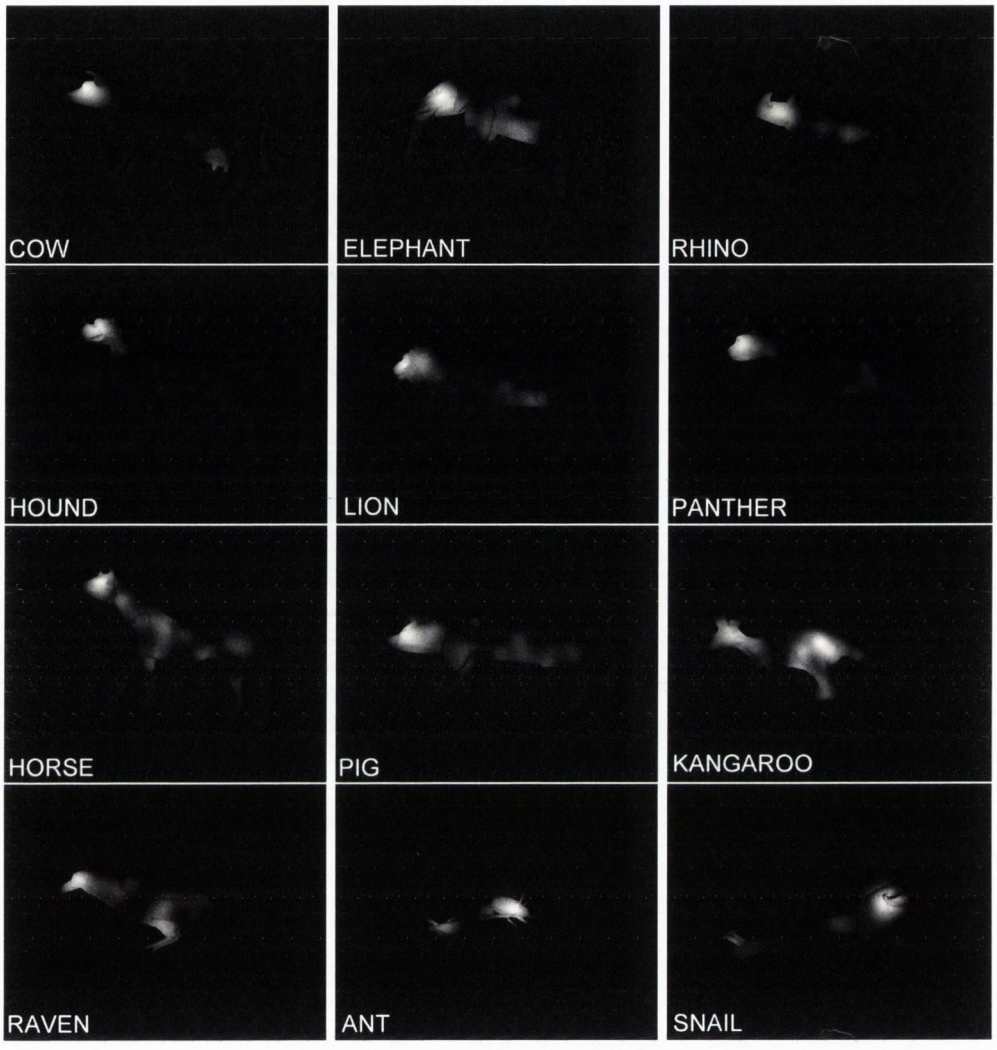

Figure 6.1: Fixation maps of all fixations for some natural objects in the naming time experiments.

For the kangaroo, even though the head was an important aspect and received quite a lot of attention, the most prominent feature appeared to be the stomach area. Perhaps this was where participants would expect to see the pouch, a defining characteristic of a kangaroo. Similarly, the shell of the snail seemed to be focussed on the most.

We also examined whether attention was drawn to the head of the natural objects immediately. As in our earlier saliency experiment, where this was the case, we examined first fixations averaged over all participants. When examining only the first fixations, attention seemed to be usually focussed upon the neck region (see Figure 6.2).

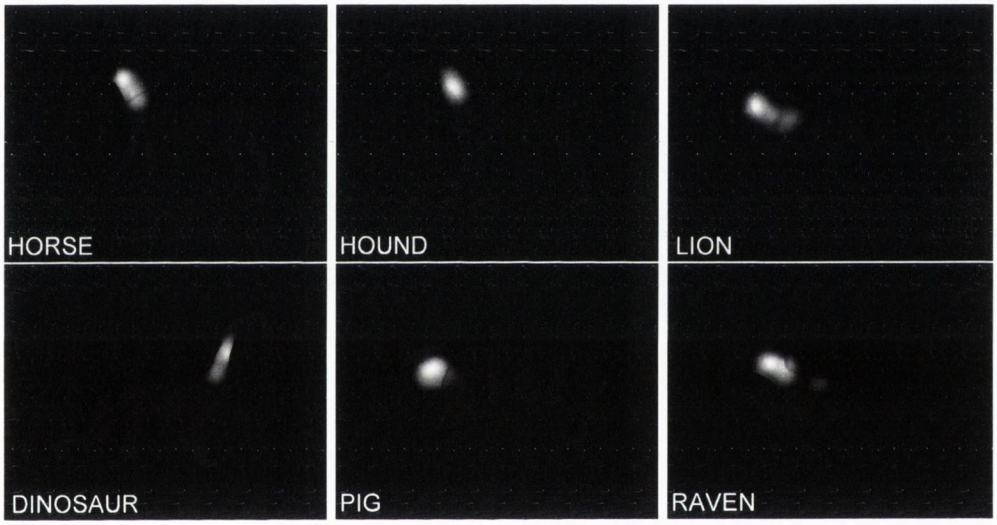

Figure 6.2: Fixation maps of all first fixations for some natural objects in the naming time experiments.

Perhaps the participants were saccading towards the head, the most prominent feature, but under-estimated the distance and fixated half way there. However, when the average of the first and second fixations were used, it was indeed the case, as before, that faces were fixated upon immediately (see Figure 6.3).

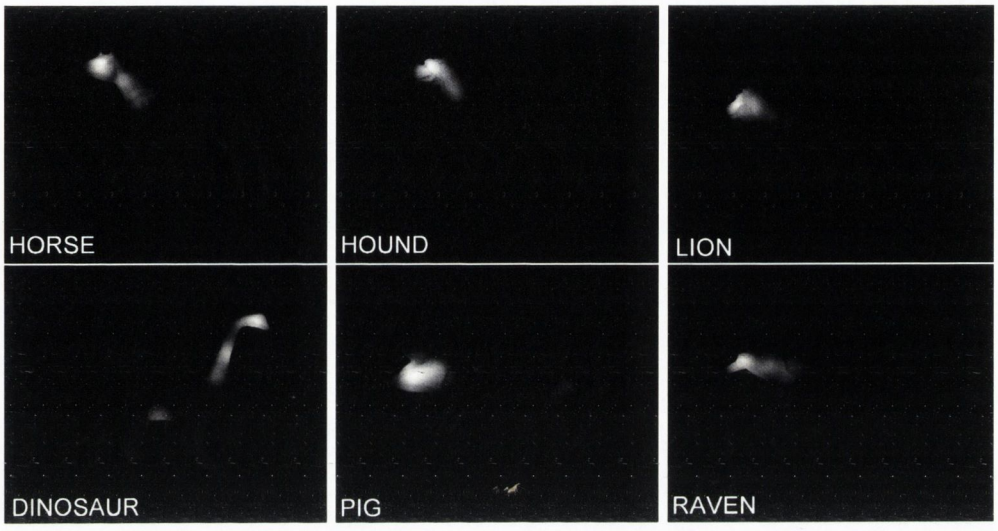

Figure 6.3: Fixation maps of all first and second fixations for some natural objects in the naming time experiments.

Following this we examined the fixation maps for the man-made artifacts (see Figure 6.4). In most cases attention was not drawn to any specific features. Generally the participants' attention was centred on the area around the fixation dot that was present prior to the experiment. For the camera it seems that attention was drawn towards the writing, for the truck and the fighter jet towards the front, for the skateboard to the wheels and in these few cases attention was drawn to these aspects immediately. During the original saliency experiment there were a lot more salient features found, for example, the keys of the piano, the straps of the sandals *etc.* A reason for this might be that, in this new task, participants were not forced to spend a certain amount of time examining the objects, as before. Objects disappeared from view as soon as they were named, nor

were participants allowed to rotate them. Perhaps when a participant was forced to examine an object for a specific amount of time, with the task of memorising it, more features were focussed upon whereas, in the new task, the objects were recognised straight away without any in-depth examination of any features. This would make sense as it seemed that in many cases participants only focussed on the centre of the screen where the fixation dot appeared.

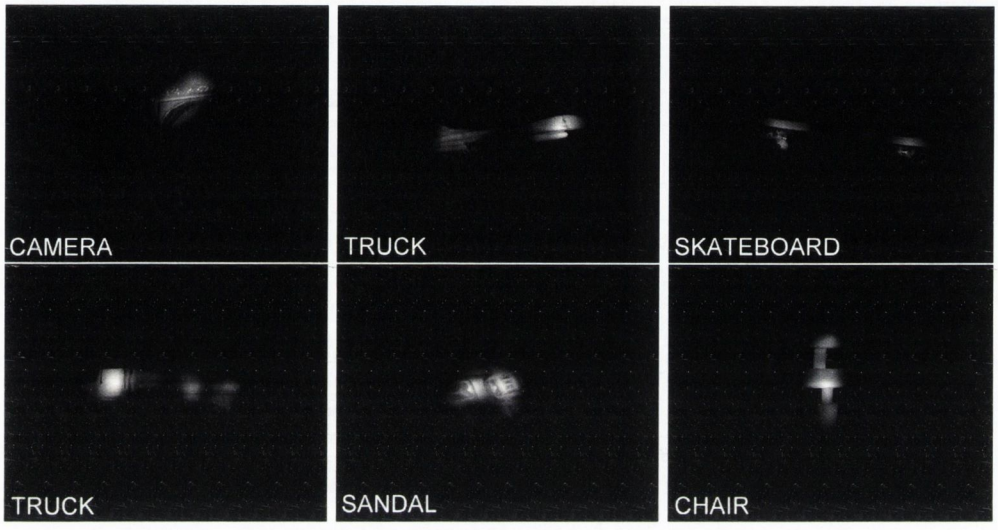

Figure 6.4: Fixation maps of all fixations for some man-made objects in the naming time experiments.

Results for the picture-picture matching experiments

These were carried out on the animal, car, fish and gear objects. We examined the animal objects first to see if, similar to the saliency and naming time experiments, the heads of these objects received the most attention (see Figure 6.5). During the task of matching, the example models on the left only received a very small amount of attention, mainly around the centre of the object; the heads did not appear to be prominent features at all. However, for the comparison models on the right of the screen, the heads were the focus of attention, as before.

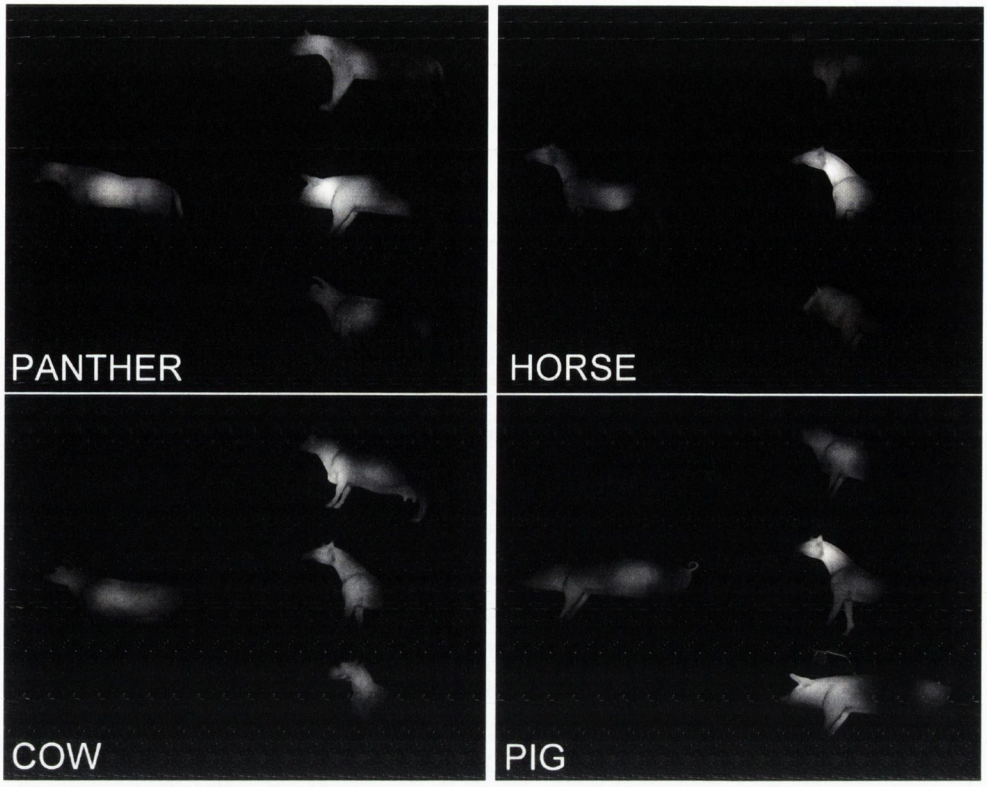

Figure 6.5: Fixation maps of all fixations for the matching time experiments for some animal objects.

The region of interest seemed to be much larger than in the naming task, so the upper body also got a significant amount of attention. This could be due to the nature of the task, as participants might have overshot while looking back and forward. From examining a few of the individual results, we detected a pattern of saccading from where the eyes landed on an object to the neck and then to the face region before moving onto the next object. However, this effect would need further investigation to confirm. Another effect of the task nature was that the objects in the middle seemed to get most of the attention, regardless of whether they were the correct object or not. Initially attention was drawn to either the body of the example model or the head of the middle comparison model. After

five fixations, the example objects seemed to be examined only a little more and the upper body of all three comparison models had received attention, but mostly the middle one and the correct answer.

For the fish objects it seemed that the tendency was to examine the correct answer more than the middle object, though generally the middle objects did get some attention (see Figure 6.6).

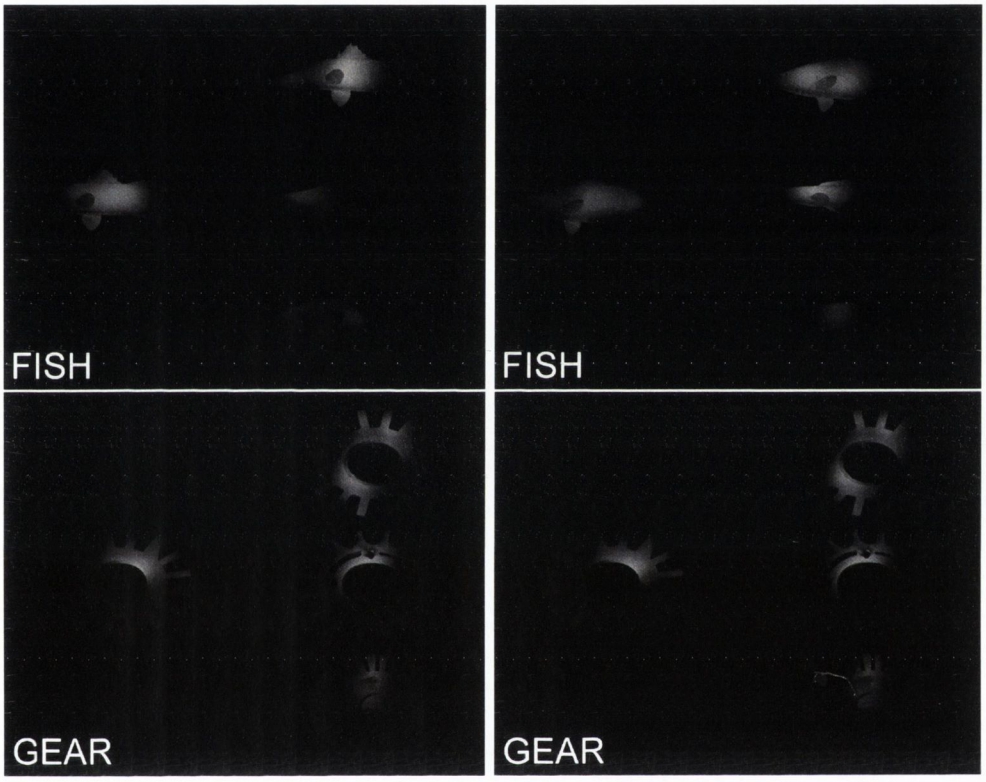

Figure 6.6: Fixation maps of all fixations for the matching time experiments for some fish and gear objects.

Perhaps this was due to the similarity of all the fish objects, (these were all exemplars of fish objects so were more similar to each other than the natural objects). It is therefore possible that it required more attention to confirm that the correct object had been chosen for an unfamiliar object than a familiar one. Attention was more widespread than for the animal objects but the front half still received more attention. This could also be dependent on the familiarity of the objects, as perhaps more aspects of these objects had to be examined to confirm that the correct object had been chosen. As with the animal objects, initially attention was focussed on the example models or the upper half of the middle comparison ones. For the fish objects, results do not compare so well to the original saliency experiment. The fins of the fish got attention in a few cases but were not particularly prominent features.

Despite the fact that in the original saliency experiments the car models did appear to have some prominent features, there were none found here, with attention being distributed equally all over the full model. Perhaps this was due to the task nature or that participants were not allowed to rotate the objects. Similar to the original saliency experiment, attention was spread all over the gear objects.

Results for the forced-choice preference experiments

In the forced-choice experiments, we compared two different simplifications of the same model. Again, like matching, attention was more widespread than for the naming task. In these tests the heads of the natural objects received attention but to a lesser extent (see Figure 6.7). However, in most cases participants examined the upper body more and the head of at least one of the objects. There does not appear to be any bias towards the left or the right object or any additional time spent studying the original or the modified versions of the models. However, from an examination of individual results, the tendency is to examine the left object before the object on the right, which fits in with the left right reading bias.

For the man-made objects, similar to the naming time tasks, attention was

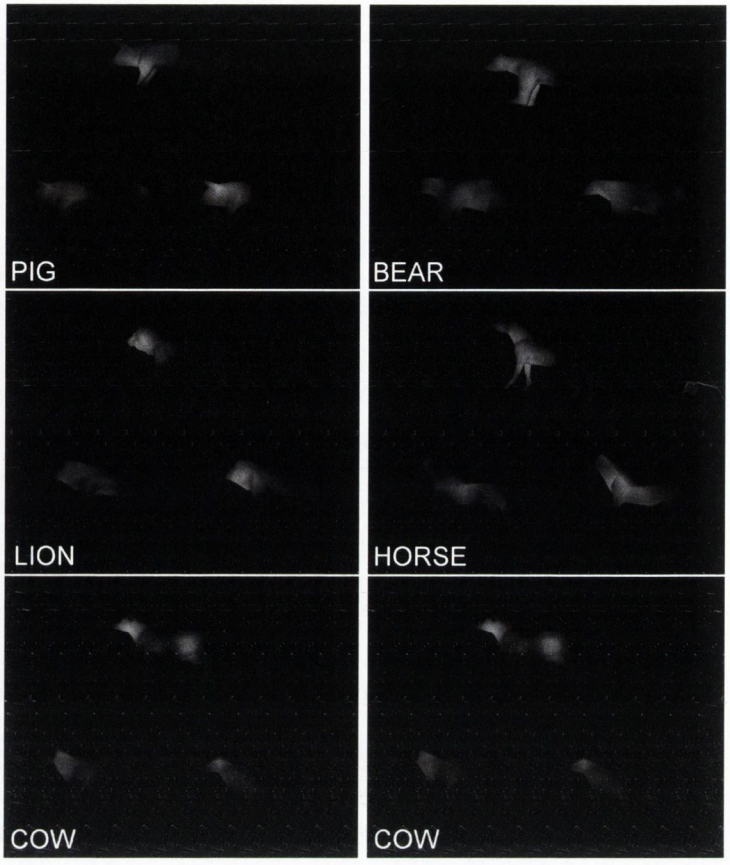

Figure 6.7: Fixation maps of all fixations for the forced-choice experiments for some animal objects.

mainly focussed on the centre of the objects (see Figure 6.8). For the fish objects, to some extent the front half of the comparison objects received more attention. Again, for the cars and gears, there does not appear to be any prominent features.

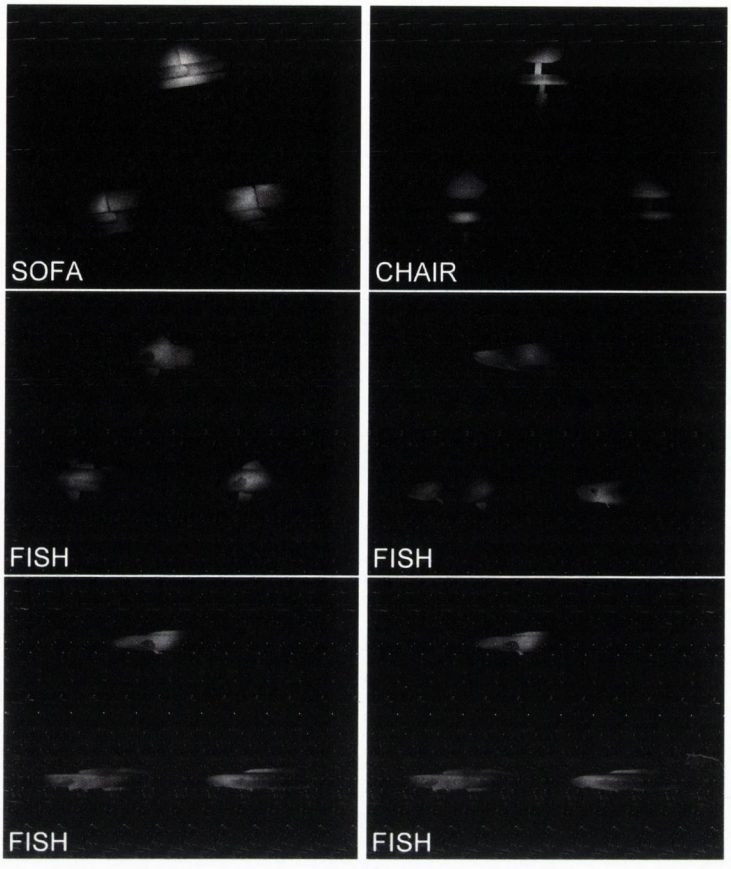

Figure 6.8: Fixation maps of all fixations for the forced-choice experiments for some man-made (1st row) and fish objects (2nd and 3rd row).

6.4 Concluding comments

These results confirm our hypothesis that the heads of natural objects are very important features. Although it has been shown before that the 'pop-out' effect for highly salient categories such as human faces does not generalise to animal faces [HH03] and that the degree of animal face 'pop-out' is extremely variable [HBH+04] our results have demonstrated that the heads of these natural objects are particularly salient. We are not claiming that they are processed pre-attentively or in parallel. However, when the task involves recognising one

of these natural objects, as in the naming time experiments, attention is drawn immediately to the head and is retained there, particularly for the four-legged animals. Although not all of the results reported are in agreement with the original saliency experiment, results for naming the natural objects in particular are, and to a lesser extent the matching and forced-choice tasks. This is further backed up by our previous evaluation experiments which shows positive naming, matching and forced-choice preference results for the natural objects at low LODs when they are simplified using the modified version of QSlim. However, the type of task affects the extent of this, with the area of focus extended in the case of the comparison task. The under/over-estimation of the distance may be the reason that the upper body of the objects received a lot of the attention as opposed to just the head. Another consequence of the layout of the matching task was that the middle object received extra attention.

With the exception of a few cases, in the naming time experiment participants did not pay special attention to any features of the man-made artifacts, car or gear objects. Regardless of the task, participants generally focussed on the centre of these objects. For the man-made artifacts and car objects, results do not correspond exactly to those of the original saliency experiment. This might explain why no positive results were found for these objects during the naming, picture-picture matching and forced choice preference experiments. Unfortunately, these results don't provide any further insights into what determines the prominent features of these man-made objects, except that they are not affected by tasks of this nature. It does not seem that attention is drawn to any specific aspects. Perhaps this is a result of the passive nature of these tasks, as there is no user interaction with objects involved here. Despite this, there is a large amount of research from the psychology field, which clearly shows that attention is controlled by a task [LH01a]. It is very likely that more promising results could be obtained with a placement or a more natural task. Perhaps, salient features defined by a task are only found when there is some form of manipulation involved in the

task and not when the objects are just viewed, as was the case in these studies. However, we have to confirm that this holds true in a virtual setup before salient features of such objects can be ascertained. To this end, we built a framework so that we could compare task performance in a real and virtual environment.

Chapter 7

Comparing Task Performance in Real and Virtual Scenes

7.1 Implementation

7.1.1 Introduction

A lot of the psychological research on tasks, demonstrates that participants only fixate aspects related to the current task. Therefore, as attention is directed voluntarily towards object of current interest control is mostly top-down, and hardly influenced by the fundamental salience of the objects. This may also explain why we had difficultly finding any salient features for the man-made artifacts in our previous study. Perhaps, a study of the landmarks critical for task control as described by Johansson *et al.* [JWBF01] would be more insightful. As described in the literature, we want to examine the eye-movements of participants while they carry out various 3D tasks in a real world situation, but we want to extend this idea by comparing the eye-movement results to those found for a similar task carried out in a matching virtual environment. In our framework, we try our best to maintain as much correspondence as possible between our real and virtual environment.

Driven by previous research and as a step towards the goal of finding ways to automatically detect salient features of man-made artifacts, this chapter is dedicated to the framework we implemented and related experimentation. We wish to examine task performance in a truly interactive, multisensory environment. We give an account of how we used realistic graphics, back projection, haptics and rapid prototyping to replicate as accurately as possible a real world scene which we have created and the interaction with this world using haptics. For our investigation regarding tasks, we recorded eye-movements in a real and virtual situation and compared them. Following this, we describe some experiments carried out using eye-tracking in the evaluation and discuss some of the results that we found.

7.1.2 Real environment

The real scene consisted of a five-sided box and a selection of physical models. The box is painted matte white, of dimensions 90x90x90 cm, as described by Morvan and McNamara [MM03]. The box is divided equally into three regions using two horizontal shelves and is placed on a table 73 cm above the ground (see Figure 7.1). The environment is lit by a single 150 W bulb placed above a square 7x11 cm opening in the top of the box. The two shelves are partitioned at the centre so that each box region has a different illumination level. Directing the light in this way makes it easier to realistically model the lighting conditions on the computer (harder shadows, less indirect light, *etc.*). An adjustable chair is positioned in front of the box.

The selection of physical models in the scene were generated from freely available data. We had five model types; the Stanford bunny, Utah teapot, cow, dragon and a block object. These models were fabricated using a Dimension 3D printer, as follows; 3D models are constructed from the bottom up, one layer at a time, with acrylnitrile butadene styrene (ABS) plastic. Catalyst software is used to import STL files. Then the device slices and orients the parts and makes the support

Figure 7.1: Real environment.

structures needed. The printer then follows the exact path that the software plots for it. The plastic is heated to a semi-liquid state and deposited as very fine layers and, finally, all support structures are removed by hand.

These models were painted in matte grey to increase contrast with the background environment. A range of luminances were provided by painting them in 5 different shades of grey.

Interaction with the scene was carried out using a 54 cm long plastic hand-held rod. Plastic hooks were attached to the top of these models to allow them to be picked up and repositioned easily using the plastic rod. This form of interaction was chosen as it closely resembles the sensation experienced when lifting up virtual objects with the haptic input device that we used in the virtual environment.

7.2 Virtual environment

The experience of interacting with the real environment was created on computer by back projecting an OpenGL application onto a projection screen. In an attempt to match accommodative and vergence distances between the real and virtual environments we place this screen the same distance from the participant as the physical box in the real world experiments, and adjust the size of the projected box so that it matches the dimensions of the front face of the physical box.

The application was back projected onto a Filmscreen 150 canvas using a high quality DLP projector. However, we experienced a major problem using the eye-tracker's scene camera with a single-chip DLP projector. These projectors beam white light through a spinning colour wheel which filters it into red, green and blue components sequentially in time. Strobing effects result due to the insufficient sampling rate of the scene camera. A solution to this problem would be to use a three-chip DLP projector instead which should prevent colour cycling since red, green and blue colour components are displayed simultaneously.

Models of the box and rod were reproduced in Autodesk 3ds Max and the rest of the objects were appropriately rescaled, and appended with hook models, to match their 3D-printed, real world counterparts.

The scene was viewed from a fixed camera pose, chosen to match the viewpoint of the real world setup. Since the box and the surrounding environment remained static for the duration of the experiment, we decided to use a background image in place of an OpenGL rendering of the box for increased realism (see Figure 7.2). The sequence of rendering steps is as follows. First we render the box geometry. Next we copy a full-screen image of the box into the frame-buffer, with depth buffer writes disabled. Finally we render the remaining object models. Note that, even though we overwrite the OpenGL rendering of the box geometry, this step is necessary in order to update the depth buffer so that the other objects are correctly sorted into the background plate.

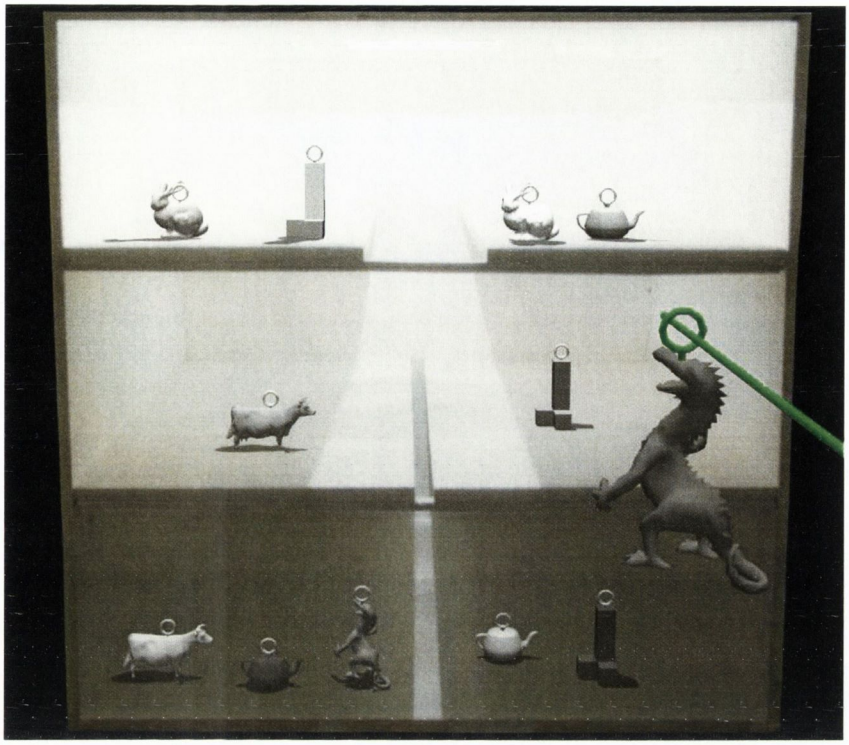

Figure 7.2: Virtual environment.

To capture the background image, we photographed the real box with a digital camera which was matched to the virtual camera pose by adjusting its position, orientation and focal length. It was difficult to obtain a single photograph without over or under exposing certain areas so, in order to gain full control over the local exposure, a series of photographs were taken, from which a high dynamic range photograph of the box was constructed in HDRShop. This was then converted into a tone-mapped low dynamic range image to account for the detail visible to the human eye under the extreme range of brightnesses present in the real environment. Finally, the image was warped to correct the perspective and barrel distortion.

The dynamic scene elements were lit using a single, downward-pointing OpenGL directional light, appropriately attenuated to account for the differing light levels

in each box region. We found that this simple local illumination model provided sufficient realism. More advanced lighting models, such as those based on pre-computed radiance transfer [SKS02], provide greater physical accuracy and may be more appropriate for other scenes. Using the stencil buffer and clipping planes, semi-transparent hard shadows were rendered by plane projecting model geometry onto the shelves in the direction of the light source, and fading their intensity based on the height of the object above the shelf. Again, we found that this simple shadow model suited our needs even though more complex alternatives such as shadow mapping or shadow volumes [HLHS03] would have improved realism by potentially supporting effects such as shadow casting between objects, self-shadowing and soft shadows.

A Phantom Premium 6DOF device was used to interact with our scene. Using this haptic device, it was possible to move a rod model around the scene (see Figure 7.3). For manipulation, objects were "picked up" by pressing and holding the Phantom switch when the rod was within range of the object's hook. Once selected, participants felt inertia, weight and collision forces on the Phantom. Due to the difficulty of computing stable haptic forces between groups of arbitrary meshes in contact, we chose to simplify the collision detection and force feedback problems by treating all objects as axis-aligned boxes, including the geometry of the five-sided box.

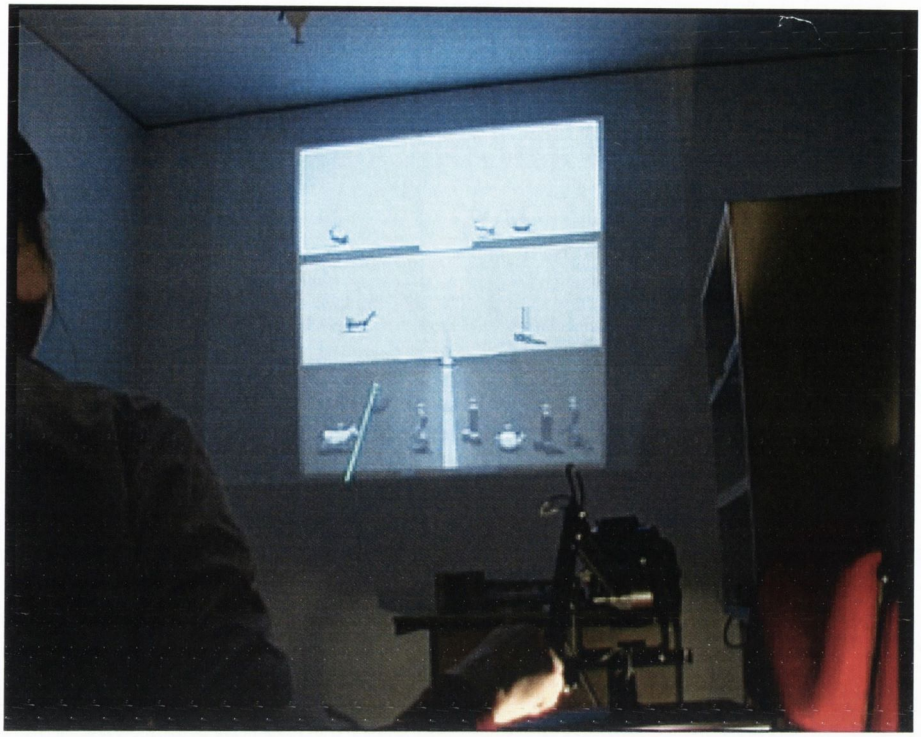

Figure 7.3: Projected virtual environment (front projected onto a white wall to produce a higher quality photograph) with the Phantom haptic device.

Haptic forces were computed by treating the selected object as a dynamic rigid body to which weight, damping and penalty-based collision forces are applied. Under the assumption of constant density, we set the mass of each object proportional to its computed volume. Collision detection was performed by analytically computing intersection times between static boxes and swept boxes. Even though the objects were always drawn in a collision-free position, internally, the boxes were allowed to intersect each other in order to provide penetration depth information from which stable haptic forces could be computed. The Phantom feedback force that resulted from movement of the dynamic body was calculated by coupling the body's centre of mass to the Phantom position with a simulated spring and setting the feedback force proportional to the spring tension force.

This model accurately simulates the sensation of lifting and dragging objects with the rod in the real world.

7.3 Preliminary experiments

7.3.1 Participants and stimuli

Eight people participated in this experiment, 7 males and 1 female, all from a Computer Science background. All participants had either normal or corrected to normal vision. Four of these performed tasks in the real environment and 4 in the virtual setup.

Apparatus for this experiment consisted of the framework described in the previous sections. During the experiments, the EyeLink II eye-tracker with scene camera was used to obtain the necessary eye-movement data (see Section 4.2).

Stimuli for this experiment consisted of 5 different models; cow, bunny, teapot, dragon and a block object. For each, there was the 3D printed version and the corresponding computer generated version. We used at least 3 copies of each of these models but no more than 5 copies of any specific template. There were five different shades of grey used for the painting. We will refer to the shades as dark, dark-medium, medium, medium-light and light shaded models (increasing in luminance from dark to light). All five models were presented in at least one dark, medium and light shaded form. All painted models had an appropriately shaded virtual counterpart.

7.3.2 Method

Prior to the real experiment, the participants were seated in front of the box, and the height of the chair was adjusted to control eye level and field of view. Participants were instructed to interact with the scene by manipulating a selection of physical models. They were given a demonstration, followed by a practice

session to move the objects around using the plastic rod. Similarly, in the virtual case, participants were seated in front of the display, the chair was adjusted, they were given a demonstration and a trial run to interact with the scene using the Phantom device. This lasted as long as it took for the participants to feel comfortable with the device. Model configurations were designed in advance and saved using an interactive editor built into our application.

Following either of these scenarios, participants were seated in front of the subject PC. Next, eye-tracker and scene camera setup was performed, this included scene camera alignment, display area detection, calibration and depth correction using the scene camera DV application on the subject PC. Following this, the participants were moved back to the previously adjusted seat in front of the real/virtual setup.

Participants carried out two placement tasks in the real/virtual environment. In one task, participants had to organise the models on the shelves according to luminance. Fifteen of the models were used in this experiment. The 5 dark, medium and light shaded models were placed randomly in the box. Participants were given the instruction to place the dark shaded objects on the top shelf, the medium shaded objects on the middle shelf and the light shaded objects on the bottom shelf.

For the second task we used 16 of these models, 8 natural objects and 8 man-made artifacts. Prior to the experiment, objects were placed randomly on the top and bottom shelves of the box. In this task, participants were told to arrange the objects depending on their type. They were told to place all natural objects on the left side of the middle shelf and man-made artifacts on the right side.

7.3.3 Preliminary results

The EyeLink II system generates output files in the EDF (EyeLink Data File) format. This output file contains information on event types such as fixations,

saccades, blinks *etc.* The EyeLink Data Viewer is the tool we used during our analysis that allows the display, filtering, and report output of the EyeLink II EDF data files. This software has three viewing modes; a spatial overlay view, a temporal graph view, and an animation view. The user can specify which event types to display, including fixations, saccades, blinks, messages, and buttons. In our analysis we used the information regarding fixations and saccades. We found the total number of fixation and saccades and the average fixation duration and saccade amplitude from the moment the participants began the task until they signalled completion. Moreover, we studied the video footage from the the scene camera, which included an overlay of each participant's gaze position.

The tasks took longer to perform in the virtual environment, which is not surprising. The number of saccades and fixations could not be compared, as the trials took different lengths of time. However we found interesting results for the average fixation duration. In both tasks, the average, minimum and maximum fixation duration was longer in the virtual setup (see Figure 7.4). Additionally, for the luminance task only the the average, minimum and maximum saccade amplitude was greater in the virtual environment(see Figure 7.5).

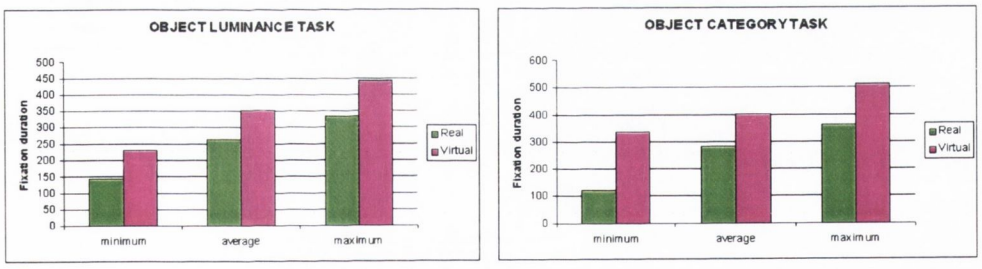

Figure 7.4: Comparing fixation duration in the real and virtual world.

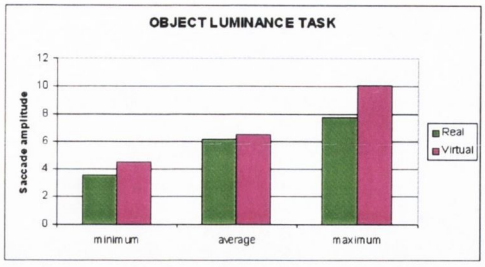

Figure 7.5: Comparing saccade amplitude in the real and virtual world.

In the real and virtual environments, the average saccade amplitude, which is the distance covered by a saccade, was greater for the luminance task than the object task for all participants (see Figure 7.6). However, the fixation duration was only greater for the luminance task than the object task in the virtual environment (see Figure 7.7).

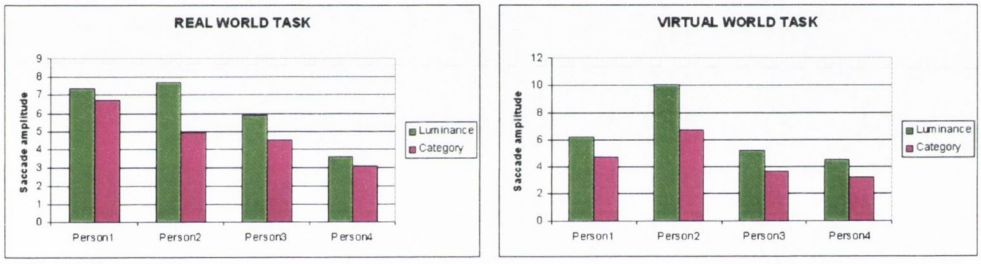

Figure 7.6: The effects of task type on saccade amplitude.

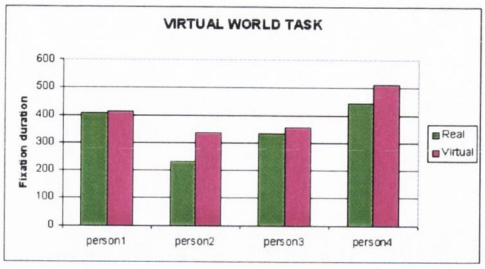

Figure 7.7: The effects of task type on fixation duration.

TASK	SETUP	TASK DURATION	FIX DURATION	SACCADE AMPLITUDE
Luminance	real	77secs	264msec	6.1deg
Luminance	virtual	136secs	353msec	6.5deg
Object type	real	53secs	284msec	4.8deg
Object type	virtual	79secs	403msec	4.5deg

Table 7.1: Average results over all participants for the task duration, saccade amplitude and fixation durations during the trials.

This demonstrates that the nature of the task has a similar influence on saccade amplitude in the real and virtual setup (see Table 7.1).

Some patterns were identified in the eye-movement data of the majority of participants during the examination of the scene camera data. In the real world setup, we also found evidence of look ahead fixations as in [PCBB01]. This occurs when objects of future interactions were foveated before they were needed, *i.e.*, when the next object to be picked up was fixated during the placement of the current object. It appears that, prior to placement, participants look ahead for free space to place the current object down. Surprisingly, it appears that the majority of participants fixate on the objects on either side of the free space and not on the free space itself. Furthermore, if a possible collision object can be picked up next, participants generally go for it with little or no regard for other pick-up options. They do not examine if there is free space for an object until after

it is picked up. There is relatively little fixation on the rod, as the objects seem to get the majority of attention. When placing an object, participants perform advance planning, by fixating the next object they will pick up. There are some differences between participants, as some are more efficient and some people tend to look around. Most people seemed to adopt a strategy of picking up objects sorted by distance from the previously placed one.

In the virtual situation fixations were far more concentrated on the objects involved in the current manipulation. Unlike the real world scenario, participants did not look-ahead and plan for future actions when the object was fixed on the rod, but strictly followed the object being moved with their eyes. Hence, the tendency was not to plan for the next step with fixations, but to deal with one object at a time. Participants almost exclusively attended to objects that were part of the task in hand with no consideration for future events, in the virtual setup. This was supported by the fact that the average fixation duration was greater in the virtual world.

7.3.4 Concluding comments

It is important to note that results reported here are only from preliminary experiments, to provide some initial insights. As the study uses a between-subject design and the number of participants involved was too small, we did not perform any statistical significance tests on the results. In the future, the study should be repeated on a bigger group of participants and a proper statistical evaluation preformed.

The most striking information found in these exploratory results was that, in the virtual scene, fixations were more concentrated on the object currently being manipulated. Moreover, the average fixation duration was longer in the virtual setup. Subjects generally looked around when the object was on the rod in the real scenario, but eye-movements strictly followed the object that was being moved in

the virtual world. In the virtual setup, eye-movements were a lot more limited, as the pattern was to deal with one object at a time with little or no advanced planning.

These results indicate that manipulation in a virtual environment of this kind requires focussed attention, limiting the participants ability to plan ahead as is the natural thing to do in a real world situation. Also, the figures found for saccade amplitude would suggest that the nature of the tasks have similar effects in the real and virtual environment. However, an investigation using a larger number of participants, on how performance varies after some practice runs would be interesting, as it is likely that novice users require more focused attention to carry out a specific task than an experienced one.

Another possible direction of this research would be to find ways to counteract the limitations of the virtual environment, in comparison to its real world counterpart. Extensive practice in the virtual setup may have an effect. An additional idea would be to, perhaps, compensate in some way for the consumption of attention by the current manipulation by using well known techniques such as previewing [OF04]. Whereby, the participant would be assisted by viewing certain properties of the scene prior to the task.

Conversely, there is much room for exploitation here. It has been shown that tasks consume attention [PHL01], and that non-task related areas of an image can be rendered in less detail [CCW03] with no loss in fidelity. This suggests that the visual fidelity of many aspect of a virtual environment could be rendered at a lower quality, reducing the computational power needed, as attention is so totally consumed in this situation by the current task, to an even greater extent than in the real world.

Chapter 8

Conclusions and Future Work

This chapter provides an overall discussion of the work described in this thesis. Moreover, some limitations of this work are pointed out as well as some avenues for future research.

8.1 Summary

In this thesis we described our research in which we examined whether visual fidelity could be improved by emphasising the detail of automatically-detected salient features of models at the expense of unimportant areas. There were positive results for natural objects at a low level of detail. Following this we were still interested in finding out how the salient features of man-made artifacts were determined. Psychological research points to them being task related. Therefore, to study the effects of task further, we built a framework which allowed the comparison of task performance in a real and virtual situation.

In our first experiment, the saliency data ascertained using the eye-tracking device indicated that there were prominent features in the case of certain model types. We examined naming time, picture-picture matching time and forced-choice preference values for models simplified using the original version of QSlim and the modified version of this software, to test if our saliency guided simpli-

fication actually improves the visual fidelity of simplified models. Our first set of evaluation results showed that the modified form of simplification produced better naming time results on familiar natural objects at a low level of detail (LOD). Matching times also suggest that low level familiar natural objects can have their visual quality enhanced by using saliency data. Preferences responded most strongly, showing that our methods had more of an effect on visual difference detection than visual recognition. This metric is a more efficient predictor for this type of study because it forces the participant to make a relative judgement. It reports, which objects the participant consciously chose to be a better representation of the original, thus, determining which form of simplification produces the object with the higher visual quality. It provides insights, not into how easily an object can be recognised but into its visual similarity to the original highly detailed model. Results show that our saliency based simplification approach can work for non-familiar natural objects as well as familiar ones, but not for man-made artifacts. There are promising results for natural objects at low LODs and it seems that, if their prominent features are preserved, the task of recognising these objects is made easier.

In the final experiment involving these objects, we confirmed previous results found. The prominent features found during the saliency experiment were actually those focussed upon during the tasks of naming, matching and forced-choice preferences, but only in those cases where the visual fidelity of the model was improved. Fixation data from the naming task demonstrated this in particular; when the task was simply to identify the natural objects, attention was immediately drawn to the heads of these objects and almost all fixations were focussed here, with little or no attention given elsewhere in many cases. Although in the saliency experiment it appeared that there were prominent features for the cars and some of the man-made artifacts, these results showed that participants tended to look only at the centre of the object, which was consistent with our evaluation results: *i.e.*, that saliency information on man-made artifacts retained during sim-

plification does not improve the naming time, picture-picture matching time or forced-choice preference results for man-made artifacts, car or gear models. Perhaps we would be more likely to find a positive result for the man-made artifacts if the task was specifically related to an object's function rather than the more general tasks of recognising and comparing. For example, it is possible that the spout of a teapot would received particular attention when pouring a cup of tea.

We are aware that it is not feasible to perform eye-tracking on every known object and that other factors such as viewpoints and textures play a role in visual fidelity too. Furthermore, the goal of our research was not to convince others to use an eye-tracker - rather it serves to provide further insights into the role of saliency in model simplification. Although the use of visual saliency does not appear to be beneficial at all LODs, it provides useful insights which could be used when rendering scenes that contain a very large number of objects, like during crowd simulation. Results show that this may also be relevant for user-guided simplification, as similar difficulties would arise when attempting to select salient features for such models by hand. Given that we know the salient features of models, either by eye-tracking or user selection like in recent work [KG03, PS03], we have experimentally established that using this data as weights in the simplification process can help to preserve the visual fidelity of low quality natural models for longer.

Following these findings, we realised that it would to be more complicated to find the salient features for the the man-made artifacts. In fact, it is more likely that the salient features of such objects vary depending upon the current task. It has been shown that visual attention is largely controlled by task during many studies from the field of visual perception. We want to examine what controls the salient features of man-made artifacts. However, before it is possible to transfer these insights about tasks from the psychology domain to the computer graphics field, further investigation is needed. In an ideal virtual system, a participant would believe they were performing the real world task. Current virtual reality

systems are not yet developed to such a degree of reality, as shown by several studies [TWG+04], so it is almost certain that task performance in a virtual environment will differ from the real world. With this in mind, we developed a framework in which we tried to maintain as much correspondence as possible between the two setups, in order to establish the differences in eye-movements during real world and virtual tasks.

We described the framework in detail followed by an account of some experiments that we carried out on it. Preliminary findings that the average fixation duration was greater when tasks were performed in the virtual world established that there are indeed some significant differences between the two situation. From further analysis of the video overlay of participants during these tasks, it is clear that eye-movements are far more constrained under the virtual circumstances. Perhaps, this could be improved with more practice. Considerable more focus is needed for the task in hand in the virtual situation.

The study described is very preliminary; a broader investigation, involving a variation of tasks within this framework is necessary. Despite the fact that airtight conclusions cannot be drawn from these exploratory tests, fixation duration and eye-movement patterns do indicate there is room for further investigation in this area and that there is potential for some interesting results. We hope that this will encourage further tests in the area and that our framework will be useful for finding some of the effects on eye-movements that arise for tasks in a virtual environment.

8.2 Limitations

1. It is not plausible to perform eye-tracking on every known object, therefore general insights have to be found in order for this to be useful, *e.g.*, the heads of the natural objects being prominent features.

2. There are other factors such as contexts, viewpoints and textures which play a role in visual fidelity which we haven't taken into consideration.

3. Regarding our framework, there are still significant differences between the real and virtual worlds. Extensions that could be made to the current framework include a back projection with stereoscopic display, sound and head tracking.

4. In order to compare tasks in a manner in which the real and virtual worlds were similar, objects had to be manipulated using a rod, which is not a very natural of task.

8.3 Future work

Similar experiments could be carried out under different scenarios, *e.g.,* the table scene shown in Figure 8.1, without requiring major modifications to our framework. However, using the rod as a means of interaction, as described above, would be inappropriate in this case.

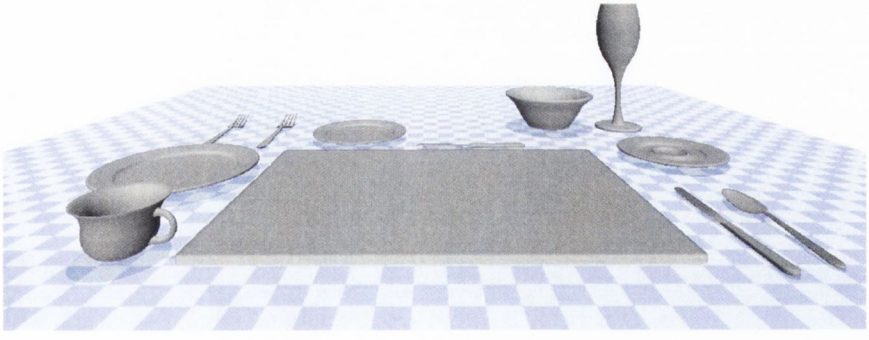

Figure 8.1: The table setup.

It is difficult to find a means of interacting with real world objects which can be faithfully reproduced in the virtual world. Other options include the use of data

gloves and motion capture, which may be more appropriate for scenes that require more control over object manipulation. Hand [Han97] examines some interaction techniques which have been developed for object manipulation, navigation and application control in 3D virtual environments.

The most important discrepancy between our current real and virtual environments may be the lack of stereo vision in the virtual environment. We would like to address this by augmenting our framework with a back projection stereoscopic display which will allow us to measure the effect of stereopsis on task performance in virtual environments. Alternatively, the real world could be made more like the virtual situation by simply covering up one of the participants eyes. We have recently installed a Vicon optical motion capture system and will use this to examine motion parallax effects by tracking the head movement of participants, thereby allowing a dynamic viewpoint in our software.

The influence of sound on task performance is another area of interest. It was found that sound can affect visual perception in certain circumstances [Red01]. Participants incorrectly counted a small number of visual stimuli when they were accompanied by a short beep, which illustrates a complex interaction between the senses. Clearly, there are many sensory factors that can interact and affect how we perceive our world [Wel].

From an evaluation perspective, more experiments need to be carried out on the framework. So far, we have only examined eye-movements for a small number of participants during two placement tasks. It would be interesting to see the difference in fixation patterns between various types of tasks. Future experiments on the framework could involve more participants carrying out a wider variety of tasks, *e.g.,* passive, manipulation, counting and memory, in addition to placement tasks. It would be interesting to see if there were any prominent features found for passive tasks, as we didn't find any for the man-made artifacts during the naming, matching and forced-choice tasks. Finally, it would be useful to further investigate how eye-movements varied during different sorts of manipulation tasks.

The interesting observation about average fixation length merits further study. Perhaps in the future this could be used as a measure to compare how similar the real and virtual setups are. As it seems like more attention to the current object is needed to perform the same task in the virtual environment, taking this a step further it would be interesting to review ways to facilitate the user, to make manipulation easier and closer to the real world experience.

Possibilities include looking for the salient object features during specific tasks and determining whether enhancing these aspects would increase the visual fidelity of simplified models if they were used during a tasks [HHO04]. Perhaps, like Cater *et al.* [CCW03], enhancing the task related object at the expense of less relevant objects could make it easier for the participant. Also, it would be interesting to investigate how the salient features of the natural objects would be determined during a task, *i.e.,* would it be a combination of top-down and bottom-up visual attention, a mixture of the salient features such as the head and the task related aspect of these objects?

Other phenomena such as priming could be looked at, which is a change in the ability to process an object produced by prior exposure to that object [SB01]. In general, prior exposure to an object results in faster identification of that object the next time it is encountered. Maljkovic and Nakayama [MN94] showed that search was facilitated when the colour of the target was repeated on the next trial. Thus, a stimulus should be found more quickly the second time that display is viewed; the prime doesn't have to be identical to the target in order to assist priming (*e.g.,* semantic priming). Practices such as priming through previewing various aspects of objects also has the potential to be interesting. Olds and Fockler [OF04] examined the effects of previewing or viewing one aspect of a search display prior to a task, either colour or orientation, in order to determine what subset of display information was most useful. Few positive results were found until both previews were presented in succession. They showed that conjunction previewing was most effective, more specifically when first the colour and then the orientation

was previewed.

In our current system previewing is implemented, by changing the colour of a models hook when the rod is in a position to select or pick it up. This design decision was taken to make the task easier for the user. In addition, other aspects of the scene could be exposed in a prelude to a task. Especially, where the users are novices, colour changes could be used to indicate what objects should be picked up next or prior indicators of where objects should be placed could also facilitate the user. On the other hand, priming has a parallel with the users experience. Our results are taken only from as small number of novice users. It needs to be explored further whether or not eye-movement patterns would be affected by increases in familiarity with our framework.

Watson *et al.* [WWWR03] showed that it is crucial to control delay when placement is difficult. Furthermore, they demonstrated that previewing decreases placement times directly and with previewing, more delay can be tolerated even when the task is difficult. Overall performance is improved with previewing, by reducing the effects of difficulty and limiting the impact of delay, so perhaps it is possible that other aspects of performance may also be affected, such as length of fixation when carrying out some simple tasks. Maybe this mechanism could be used to moderate the effects that a virtual world has on performance as measured by fixation length compared to the real world. Another example of the benefits of previewing is in the case of vehicle control tasks, where previewing the necessary input can compensate for delay [Wic86].

Finally, another interesting idea might be guiding the users attention during a task. Halper *et al.* [HMDS05] show that non-photorealistic rendering can play a subtle, yet effective, role in guiding judgement and subsequent interactions. An examination of additional cues that could be used to augment the virtual environment may also result in improved task performance.

Bibliography

[Bar76] D. J. Bartram. Levels of coding in picture-picture comparison tasks. In *Memory & Cognition*, volume 4(5), pages 593–602, 1976.

[BDDG03] P. Baudisch, D. DeCarlo, A. T. Duchowski, and W. S. Geisler. Focusing on the essential: considering attention in display design. *Commun. ACM*, 46(3), 2003.

[BG85] V. Bruce and P. Green. *Visual Perception: physiology, psychology, and ecology.* Lawrence Erlbaum Associates, London, 1985.

[BHF97] V. Brown, D. Huey, and J. M Findlay. Face detection in peripheral vision: do faces pop out? In *Perception*, volume 26, pages 1555–1570, 1997.

[BL73] M. A. Baker and M. Loeb. Implications of measurement of eye fixations for a psychophysics of form perception. In *Perception & Psychophysics*, volume 13, pages 185–192, 1973.

[BM98] M. Bolin and G. Meyer. A perceptually based adaptive sampling algorithm. In *Computer Graphics (Proc. Siggraph 98)*, volume 32, pages 299–309, 1998.

[BRA01] R. Browse, J. Rodger, and R. Adderly. Perception of object shape in computer graphics displays. In *Journal of Electronic Imaging*, volume 10(1), pages 181–187, 2001.

[CB97] R. Carey and G. Bell. *The annotated VRML 2.0 reference manual.* Addison-Wesley Longman Ltd., Essex, UK, UK, 1997.

[CCW03] K. Cater, A. Chalmers, and G. Ward. Detail to attention: Exploiting visual tasks for selective rendering. In *Proceedings of the 2003 EUROGRAPHICS Symposium on Rendering*, pages 270–280. EUROGRAPHICS, 2003.

[Cla76] J. H. Clark. Hierarchical geometric models for visible surface algorithms. *Commun. ACM*, 19(10):547–554, 1976.

[CMRS98] P. Cignoni, C. Montani, C. Rocchini, and R. Scopigno. Zeta: A resolution modeling system. In *GMIP: Graphical Models and Image Processing 60*, volume 5, pages 305–329, 1998.

[CMS00] P. Cignoni, C. Montani, and R. Scopigno. A comparison of mesh simplification algorithms. In *Computers and Graphics*, 2000.

[CRS98] P. Cignoni, C. Rocchini, and R. Scopigno. Metro: Measuring error in simplified surfaces. In *Computer Graphics Forum*, volume 17(2), pages 167–174, 1998.

[CT99] F. Cutzu and M. J. Tarr. Inferring perceptual saliency fields from viewpoint-dependent recognition data. In *Neural Computation*, pages 1331–1348, 1999.

[DMdMT03] M. H. Davis, H. E. Moss, P. de Mornay, and L. K. Tyler. Spot the difference: Investigations of conceptual structure for living things an artificacts using speeded word-picture matching. In *Department of Experimental Psychology, University of Cambridge*, 2003.

[Duc02] A. T. Duchowski. A breadth-first survey of eye tracking applications. In *Behavior Research Methods, Instruments, and Computers*, pages 455–470, 2002.

[Far92] M. J. Farah. Is an object an object an object? Cognitive and neuropsychological investigations of domain-specificity in visual object recognition. In *Current Directions in Psychological Science*, volume 1, pages 164–169, 1992.

[FJM50] P. M. Fitts, R.E. Jones, and J. L. Milton. Eye movements of aircraft pilots during instrument-landing approaches. In *Aeronautical Engineering Review*, volume 9(2), pages 24–29, 1950.

[FWDT94] M. J. Farah, K. D. Wilson, M. Drain, and J. R. Tanaka. The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. In *Vision Research*, volume 35(14), pages 2089–2093, 1994.

[GH97] M. Garland and P. Heckbert. Surface simplification using quadric error metrics. In *Computer Graphics Proceedings, Annual Conference Series*, pages 209–216, 1997.

[GK99] J. H. Goldberg and X. P. Kotval. Computer interface evaluation using eye movements: methods and constructs. In *International Journal of Industrial Ergonomics*, volume 24, pages 631–645, 1999.

[GSGA00] I. Gauthier, P. Skudlarski, J. C. Gore, and A. W. Anderson. Expertise for cars and birds recruits brain areas involved in face recognition. In *Nature Neuroscience*, volume 3, pages 191–197, 2000.

[GSM97] M. A. Geren, R. Stromer, and H. A. Mackay. Picture naming, matching to sample, and head injury: a stimulus control analysis. In *Journal of applied behaviour analysis*, volume 30, pages 339–342, 1997.

[GT97] I. Gauthier and M. J. Tarr. Becoming a "greeble" expert: Exploring mechanisms for face recognition. In *Vision Research*, volume 37, pages 1673–1682, 1997.

[Han97] C. Hand. A survey of 3-d interaction techniques. In *Computer Graphics Forum*, volume 16(5), pages 269–281, 1997.

[Hay00] M. Hayhoe. Vision using routines : A functional account of vision. In *Visual Cognition 2000*, volume 7(1/2/3), pages 43–64, 2000.

[HBB98] M. M. Hayhoe, D. G. Bensinger, and D. H. Ballard. Task constraints in visual working memory. In *Vision Research*, volume 38(1), pages 125–137, 1998.

[HBH+04] S. Hochstein, A. Barlasov, O. Hershler, A. Nitzan, and S. Shneor. Rapid vision is holistic [abstract]. In *Journal of Vision*, volume 4, 2004.

[Hen92] J. M. Henderson. Object identification in context: The visual processing of natural scenes. In *Canadian Journal of Psychology*, volume 46(3), pages 319–341, 1992.

[HH98] J. M. Henderson and A. Hollingworth. Eye movements during scene viewing: An overview. In *G. Underwood (Ed.), Eye Guidance in Reading and Scene Perception*, pages 269–294, 1998.

[HH03] O. Hershler and S. Hochstein. High-level effects in rapid 'pop-out' visual search. In *ECVP 2003 abstract*, 2003.

[HHO04] S. Howlett, J. Hamill, and C. O'Sullivan. An experimental approach to predicting saliency for simplified polygonal models. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 57–64. ACM Press, 2004.

[HHO05] S. Howlett, J. Hamill, and C. O'Sullivan. Predicting and evaluating saliency for simplified polygonal models. In *ACM Transactions on Applied Perception*, 2005.

[HLHS03] J. Hasenfratz, M. Lapierre, N. Holzschuch, and F. Sillion. A survey of real-time soft shadows algorithms. In *Eurographics*. Eurographics, 2003.

[HLO05] S. Howlett, R. Lee, and C. O'Sullivan. A framework for comparing task performance in real and virtual scenes. In *APGV '05: Proceedings of the 2st Symposium on Applied perception in graphics and visualization*. ACM Press, 2005.

[HMDS05] N. Halper, M. Mellin, D. Duke, and T. Strothotte. Implicational rendering: Drawing on latent human knowledge. In *ACM Transactions on Applied Perception*, 2005.

[Hop99] H. Hoppe. New quadric metric for simplifiying meshes with appearance attributes. In *VIS '99: Proceedings of the conference on Visualization '99*, pages 59–66, 1999.

[HRQ88] G. W. Humphreys, M. J. Riddoch, and P. T. Quinlan. Cascade processes in picture identification. In *Cognitive Neuropsychology*, volume 5(1), pages 67–103, 1988.

[HSMP03] M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. Pelz. Visual memory and motor planning in a natural task. In *Journal of Vision*, volume 3, pages 49–63, 2003.

[IK00] L. Itti and C. Koch. Feature combination strategies for saliency-based visual attention systems. In *Journal of electronic imaging*, volume 10(1), pages 161–169, 2000.

[IKN98] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 20(11), page 12541259, 1998.

[Jac93] R. J. K. Jacob. Eye-movement-based human-computer interaction techniques: Toward non-command interfaces. In *Advances in Human-Computer Interaction*, volume 42, pages 151–190, 1993.

[JWBF01] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan. Eye-hand coordination in object manipulation. In *The Journal of Neuroscience*, volume 21(17), pages 6917–6932, 2001.

[KG03] Y. Kho and M. Garland. User-guided simplification. In *Proceedings of ACM Symposium on Interactive 3D Graphics*, 2003.

[KGS98] S. Kelly, G. Green, and M. Sidman. Visual identity matching and auditory-visual matching: A procedural note. In *Journal of Applied Behavior Analysis*, volume 21, pages 237–243, 1998.

[KJ94] S. M. Kuehn and P. Jolicoeur. Impact of quality of the image orientation and similarity of the stimuli on visual search for faces. In *Perception*, volume 23, pages 95–122, 1994.

[KS99] R. Klein and A. Schilling. Efficient rendering of multiresolution meshes with guaranteed image quality. In *Visual Computer*, volume 15, pages 443–452, 1999.

[Lat88] C. R. Latimer. Eye-movement data: Cumulative fixation time and cluster analysis. In *Behavior Research Methods, Instruments, and Computers*, volume 20(5), pages 437–470, 1988.

[LBD02] R. Lawson, H. Bülthoff, and S. Dumbell. Interactions between view changes and shape changes in picture-picture matching. In *Technical Report No. 095*, 2002.

[LC89] D. N. Levine and R. Calvanio. Prosopagnosia: A defect in visual configural processing. In *Brain and Cognition*, volume 10, pages 149–170, 1989.

[LH01a] M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activites? In *Vision Research*, volume 41, pages 3559–3565, 2001.

[LH01b] D. Luebke and B. Hallen. Perceptually-driven simplification for interactive rendering. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, 2001.

[LH04] Y. Ling and A. Hurlbert. Color and size interactions in a real 3d object similarity task. In *The Journal of Vision*, volume 4, pages 721–734, 2004.

[LHNB00] D. Luebke, B. Hallen, D. Newfield, and B.Watson. Perceptually driven simplification using gaze-directed rendering. In *University of Virginia Technical Report CS-2000-04*, 2000.

[LM78] G. R. Loftus and N. H. Mackworth. Cognitive of determinants of fixation location during picture viewing. In *Journal of Experimental Psychology: Human Perception and Performance*, volume 4, pages 565–572, 1978.

[LNWB03] B. Lok, S. Naik, M. Whitton, and F. P. Brooks. Effects of handling real objects and avatar fidelity on cognitive task performance in virtual environments. In *Proceedings of the IEEE Virtual Reality*, 2003.

[LRC+02] D. Luebke, M. Reddy, J. Cohen, A. Varshney, B. Watson, and R. Huebner. *Level of Detail for 3D Graphics*. Morgan-Kaufmann, 2002.

[LRM04] K. Lau, R. A. Rensink, and T. Munzner. Perceptual invariance of nonlinear focus+context transformations. In *APGV '04: Proceed-*

ings of the 1st Symposium on Applied perception in graphics and visualization, pages 65–72, 2004.

[LRR03] N. Lavie, T. Ro, and C. Russell. The role of perceptual load in processing distractor faces. In *Psychological Science*, volume 14(5), 2003.

[LRS03] J. Laarni, N. Ravaja, and T. Saari. Using eye tracking and psychophysiological methods to study spatial presence. In *Proceedings of the 6th International Workshop on Presence*, 2003.

[LVJ05] C. H. Lee, A. Varshney, and D. W. Jacobs. Mesh saliency. In *Proceedings of the 2005 ACM SIGGRAPH Symposium on Interactive 3D Graphics*, 2005.

[LW01] G. Li and B. Watson. Semiautomatic simplification. In *ACM Symposium on Interactive 3D Graphics 2001*, pages 43–48, 2001.

[LWC+02] D. Luebke, B. Watson, J. D. Cohen, M. Reddy, and A. Varshney. *Level of Detail for 3D Graphics*. Elsevier Science Inc., New York, NY, USA, 2002.

[MD01] H. Murphy and A. T. Duchowski. Gaze-contingent level of detail rendering. In *Eurographics 2001 (Short Presentations)*, 2001.

[MD02] G. Marmitt and A. T. Duchowski. Modeling Visual Attention in VR: Measuring the Accuracy of Predicted Scanpaths. In *EuroGraphics 2002 Proceeding*, 2002.

[MM67] N. H. Mackworth and A. J. Morandi. The gaze selects informative details within pictures. In *Perception & Psychophysics*, volume 7, pages 173–178, 1967.

[MM03] Y. Morvan and A. McNamara. Assessing the visual perception impact of indirect lighting under different levels of illumination. a psychophysical experiment. In *Proceedings Eurographics Ireland*, 2003.

[MN94] V. Maljkovic and K. Nakayama. Priming of pop-out: I. Role of features. In *Memory & Cognition*, pages 657–672, 1994.

[MR98] A. Mack and I. Rock. *Inattentional blindness.* Cambridge, MA: MIT Press, 1998.

[MTCR+04] B. J. Mohler, W. B. Thompson, S. Creem-Regehr, H. L. Pick, W. Warren, J. J. Rieser, and P. Willemsen. Visual motion influences locomotion in a treadmill virtual environment. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 19–22, 2004.

[MW93] S. MacKenzie and C. Ware. Lag as a determinant of human performance in interactive systems. In *Proceedings of the ACM Conference on Human Factors in Computing Systems - INTERCHI '93*, pages 488–493, 1993.

[Not93] H. Nothdurft. Faces and facial expressions do not pop out. In *Perception*, volume 22, pages 1287–1298, 1993.

[NS79] D. Noton and L. W. Stark. Scanpaths in eye movements during pattern perception. In *Science*, pages 308–311, 1979.

[OF04] E. S. Olds and K. A. Fockler. Does previewing one stimulus feature help conjunction search? In *Perception*, volume 33, pages 195–216, 2004.

[OYT96] T. Ohshima, H. Yamamoto, and H. Tamura. Gaze-directed adaptive rendering for interacting with virtual space. In *VRAIS '96:*

Proceedings of the 1996 Virtual Reality Annual International Symposium (VRAIS 96), page 103, Washington, DC, USA, 1996. IEEE Computer Society.

[PCBB01] J. B. Pelz, R. Canosa, J. Babcock, and J. Barber. Visual perception in familiar, complex tasks. In *Proceedings of the ICIP*, 2001.

[PHG+04] B. Pan, H. A. Hembrooke, G. K. Gay, L. A. Granka, M. K. Feusner, and J. K. Newman. The determinants of web page viewing behavior: an eye-tracking study. In *ETRA'2004: Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications*, pages 147–154. ACM Press, 2004.

[PHL01] J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. In *Exp Brain Res*, volume 139, pages 266–277, 2001.

[PN04] D. Parkhurst and E. Niebur. A feasibility test for perceptually adaptive level of detail rendering on desktop systems. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 49–56, 2004.

[PS00] C. M. Privitera and L. W. Stark. Algorithms for defining visual regions-of-interest: Comparisons with eye fixations. In *IEEE Transaction on patteren analysis and machine intelligence*, volume 22, 2000.

[PS03] E. Pojar and D. Schmalstieg. User-controlled creation of multiresolution meshes. In *Proceedings of ACM Symposium on Interactive 3D Graphics*, 2003.

[Red96] M. Reddy. Scrooge: Perceptually-driven polygon reduction. In *Computer Graphics Forum*, volume 15(4), pages 191–203, 1996.

[Red98] M. Reddy. Specification and evaluation of level of detail selection criteria. In *Virtual Reality: Research, Development and Application*, volume 3(2), pages 132–143, 1998.

[Red01] M. Reddy. Perceptually optimized 3d graphics. In *IEEE Computer Graphics and Applications*, 2001.

[Ren00] R. A. Rensink. The dynamic representation of scenes. In *Visual Cognition*, pages 17–42, 2000.

[Ren02a] R. A. Rensink. Change detection. In *Annual Review of Psychology*, pages 245–277, 2002.

[Ren02b] R. A. Rensink. Visual attention. In *Encyclopedia of Cognitive Science*, 2002.

[Ren04] R. A. Rensink. The invariance of visual search to geometric transformation. In *Journal of Vision*, 2004.

[ROC97] R. A. Rensink, J. K. O'Regan, and J. J. Clark. To see or not to see: The need for attention to perceive changes in scenes. In *Psychological Science*, pages 368–373, 1997.

[Ros76] E. Rosch. Natural categories. In *Cognitive Psychology*, volume 4, pages 328–350, 1976.

[RPG99] M. Ramasubramanian, S. N. Pattanaik, and D. P. Greenberg. Perceptually based physical error metric for realistic image synthesis. In *Computer Graphics (Proc. Siggraph 99)*, volume 33, pages 73–82, 1999.

[RR01] B. E. Rogowitz and H. E. Rushmeier. Are image quality metrics adequate to evaluate the quality of geometric objects? In *Proceedings*

of SPIE Vol 4299 Human Vision and Electronic Imaging VI, pages 340–349, 2001.

[RRL01] T. Ro, C. Russell, and N. Lavie. Changing faces: A detection advantage in the flicker paradigm. In *Psychological Science*, volume 12, pages 94–99, 2001.

[Rus01] H. Rushmeier. Metrics and geometric simplification. In *Siggraph Course Notes*, 2001.

[SB01] D. L. Schacter and R. D. Badgaiyan. Neuroimaging of priming: New perspectives on implicit and explicit memory. In *Current Directions in Psychological Science*, pages 1–4, 2001.

[SC95] S. Suzuki and P. Cavanagh. Facial organization blocks access to low-level features: An object inferiority effect. In *Human Perception and Performance*, volume 21, pages 901–913, 1995.

[SCCD04] V. Sundstedt, A. Chalmers, K. Cater, and K. Debattista. Top-down visual attention for efficient rendering of task related scenes. In *Vision, Modelling and Visualization*, 2004.

[SHS01] H. Shinoda, M. M. Hayhoe, and A. Shrivastava. What controls attention in natural environments? In *Vision Research*, volume 41(25-26), pages 3535–3545, 2001.

[Sim00] D. J. Simons. Current approaches to change blindness. In *Vision Cognition*, pages 1–16, 2000.

[SKS02] P. Sloan, J. Kautz, and J. Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 527–536, 2002.

[SU93] M. Slater and M. Usoh. The influence of a virtual body on presence in immersive virtual environments. In *Virtual Reality International, Proceedings of the Third Annual Conference on Virtual Reality*, pages 34–42, 1993.

[SU94] M. Slater and M. Usoh. Body centred interaction in immersive virtual environments. In *Artificial Life and Virtual Reality*, pages 125–148. John Wiley and Sons, 1994.

[Tar00] M. J. Tarr. Visual pattern recognition. In *Encyclopedia of Psychology*, 2000.

[TG80] A. Treisman and G. Gelade. A feature integration theory of attention. In *Cognitive Psychology*, volume 12, pages 97–136, 1980.

[TH84] B. Tversky and K. Hemenway. Objects, parts, and categories. In *Journal of Experimental Psychology: General*, volume 113(2), pages 49–56, 1984.

[Tve77] S. Tversky. Features of similarity. In *Psychological Review*, pages 327–352, 1977.

[TWG+04] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when judging distances in visually immersive environments? In *Presence: Teleoperators and Virtual Environments*, pages 560–571, 2004.

[Wat03] B. Watson. Frontiers in perceptually-based image synthesis: Modeling, rendering, display, validation. In *Siggraph Course, San Diego*, 2003.

[WCG94] Q. Wang, P. Cavanagh, and M. Green. Familiarity and pop-out in visual search. In *Perception & Psychophysics*, volume 56, pages 495–500, 1994.

[WCT98] K. N. Walker, T. F. Cootes, and C. J. Taylor. Locating salient object features. In *British Machine Vision Conference*, 1998.

[Wel]

[Wer93] J. Wernecke. *The Inventor Mentor: Programming Object-Oriented 3d Graphics with Open Inventor, Release 2.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1993.

[WFM00] B. A. Watson, A. Friedman, and A. McGaffey. Using naming time to evaluate quality predictors for model simplification. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120, 2000.

[WFM01] B. Watson, A. Friedman, and A. McGaffey. Measuring and predicting visual fidelity. In *Computer Graphics Proceedings, Annual Conference Series*, pages 213–220, 2001.

[Wic86] C. D. Wickens. The effects of control dynamics on performance. In *Handbook of Human Perception and Performance*, volume 2, pages 1–39, 1986.

[WLC+03] N. Williams, D. Luebke, J. Cohen, M. Kelley, and B. Schubert. Perceptually guided simplification of lit, textured meshes. In *Proceedings of the 2003 ACM SIGGRAPH Symposium on Interactive 3D Graphics*, 2003.

[WLWD03] C. Woolley, D. Luebke, B. Watson, and A. Dayal. Interruptible rendering. In *Proceedings of the 2003 ACM SIGGRAPH Symposium on Interactive 3D Graphics*, 2003.

[WWH97] B. Watson, N. Walker, and L. F. Hodges. Managing level of detail through head-tracked peripheral degradation: a model and resulting design principles. In *VRST '97: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 59–63, New York, NY, USA, 1997. ACM Press.

[WWH04] B. Watson, N. Walker, and L. F. Hodges. Supra-threshold control of peripheral lod. In *Proceedings of ACM Siggraph 2004*, 2004.

[WWHW97] B. Watson, N. Walker, L. F. Hodges, and A. Worden. Managing level of detail through peripheral degradation: Effects on search performance with a head-mounted display. In *ACM Trans. on Computer-Human Interaction*, volume 4, pages 323–346, 1997.

[WWWR03] B. A. Watson, N. Walker, P. Woytiuk, and W. Ribarsky. Maintaining usability during 3d placement despite delay. In *VR '03: Proceedings of the IEEE Virtual Reality 2003*, page 133, 2003.

[XV96] J.C. Xia and A. Varshney. Dynamic view-dependent simplification for polygonal models. In *Proceedings of IEEE Visualization*, pages 327–334, 1996.

[Yar67] A. L. Yarbus. *Eye Movements and Vision.* Plenum Press, New York, 1967.

[YPG01] H. Yee, S. Pattanaik, and D. P. Greenberg. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. In *ACM Trans. Graph.*, volume 20, pages 39–65. ACM Press, 2001.