An Experimental Approach to Predicting Saliency for Simplified Polygonal Models

Sarah Howlett, John Hamill, Carol O'Sullivan*
Image Synthesis Group, Trinity College Dublin

Abstract

In this paper, we consider the problem of determining feature saliency for 3D objects and describe a series of experiments that examined if salient features exist and can be predicted in advance. We attempt to determine salient features by using an eye-tracking device to capture human gaze data and then investigate if the visual fidelity of simplified polygonal models can be improved by emphasizing the detail of salient features identified in this way. To try to evaluate the visual fidelity of models simplified using both metrics, a set of naming time, matching time and forced-choice preference experiments were carried out. We found that our perceptually weighted metric led to a significant increase in visual fidelity for the lower levels of detail (LOD) of the natural objects, but that for the man-made artifacts the opposite was true. We therefore conclude that visually prominent features may be predicted in this way for natural objects, but our results show that saliency prediction for synthetic objects is more difficult, perhaps because it is more strongly affected by task. We hope that our results will lead to new insights into the nature of saliency in 3D graphics.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism; I.3.5 [Computer Graphics]: Computational Geometry and Object Modelling

Keywords: visual perception, model simplification, salient features.

1 Introduction

For interactivity in computer graphics, the ideal is to have the most realistic dynamic scene possible while meeting real-time constraints. As more computational power is not always available, highly detailed models must be simplified in order to be displayed interactively and the major challenge is to maintain the visual fidelity of the models under simplification. Simplifying models 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 promising solution is to use perceptually adaptive graphics where knowledge of the human visual system and its weaknesses are exploited when displaying images and animations. To this end, we used an SMI EyeLink eye-tracker (Figure 1 of the color plate) to determine which features of two sets of models received the most attention and then investigated if the perceptual quality could be enhanced by presenting these aspects in greater detail. We wished to determine if higher visual quality is maintained when simplification takes fixation data into consideration as well as geometry. As the perceptual importance of an object is determined by the user, fixation data was gathered from participants while viewing a set of models at a high LOD. Then, using this data while minimizing the number of polygons, we hoped to create a model with a higher perceptual quality. To do this we weighted the model simplification metric with fixation data, thus preserving the perceptually important regions. In order to determine the visual quality of these simplified models, we gathered some psychological measurements: naming times [Watson et al. 2001; Watson et al. 2000] on the first set of familiar objects, picture-picture matching times [Lawson et al. 2002] on the second set to determine if familiarity played a role and forced-choice preferences on both sets of models. 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, especially at the lower LOD's. The goal of our research is to use an eye-tracker to examine the role of feature saliency in model simplification and, as such, our results should provide insights which will be helpful for other approaches to perceptually guided simplification.

*e-mail: Sarah.Howlett, John.Hamill, Carol.OSullivan@cs.tcd.ie

2 Background

Recent work on this problem includes reducing model complexity based on geometry, perceptual models [Luebke and Hallen 2001] or input taken directly from the user [Kho and Garland 2003]. There has also been major work on gaze contingent systems [Duchowski 2002] and peripherally degraded displays [Reddy 1998; Watson et al. 1997]. There has been much previous research into saliency [Itti et al. 1998; Yee et al. 2001]. A lot of the initial work on simplification used geometric methods [Rushmeier 2001], especially the quadric error metric developed by Garland and Heckbert [1997], which is used as the basis for the QSlim software. Expanding on Garland and Heckbert's quadric error metric is work from Pojar and Schmalstieg [2003]. They present a tool for user-controlled creation of multiresolution meshes. Recent work by Kho and Garland [2003], which was preceded by work from Cignoni et al. [1998] and Li and Watson [2001], also uses weights that can be specified by the user. It gave the user the ability to select the importance of different areas on a model, thus preserving the prominent features of their choice which would be lost if fully automatic simplification was used.

In our research we expanded upon some of the previous approaches by using an eye-tracking device, not in a gaze contingent way, but to ascertain the prominent features of a model. So when examining saliency, unlike much previous work, we focus on the salient features of particular objects and not on saliency in a scene. We used three metrics to determine attention. The first was the total duration of all fixations on a region while a scene is being viewed [Henderson and Hollingworth 1998]. Henderson also suggests that a better fixation measure is the duration of the first fixation on an

object [Henderson 1992], our second metric. The third metric was the number of fixations on each triangle in the mesh. According to Fitts et al. [1950], 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.

Having used the eye-tracker to gather this data on saliency, the original version of QSlim was modified to use this information. To find out if our method actually works, it was necessary to measure the visual fidelity of the new models. There are several common ways of measuring visual fidelity, namely automatic and experimental measures. Experimental measures include forced-choice preferences, ratings and naming times, all described in detail by Watson et al. [2001]. The experimental measures we used were naming times and forced-choice preferences. Watson et al. [2000] 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. As our second set of stimuli included some non-familiar objects, we chose to use a picture-picture matching method [Lawson et al. 2002] to determine the visual quality of these models because no verbalization is required.

3 Finding the Salient Features

The initial step was to attempt to determine the salient features of the models automatically. An SMI EyeLink high-speed eye-tracking system (250hz) manufactured by SensorMotoric Instruments was used to get information on where a participant was fixating when viewing a particular model. At any instant the eye is either fixating on something or making a saccade (an eye-movement), so we detected saccades by measuring the difference between the current eye position and the average of the last six eye positions. If the size of the visual angle was greater than some threshold then a saccade was recorded. We kept track of the faces in the polygonal model that were focused 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 [1993]. 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 coloring method was used to determine which faces were being focussed upon. Faces were drawn (without lighting) to a back buffer with a unique color associated with them. When the point under the EyeLink gaze was found, the color under the corresponding region in the back buffer was read back. As colors were unique, the face or faces being focussed upon could be determined. Furthermore, by expanding the region under scrutiny, the neighboring 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.

3.1 Participants & Method

There were 20 participants involved in this 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 and 18 man-made artifacts, which were in the public domain, and the same stimuli as those used in Watson et al's [2001] experiment with one additional model. Using QSlim [Garland and Heckbert 1997] 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; animals, cars, fish and gears (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).

For both sets of models, each participant viewed each model twice for approximately 30 seconds, from two different initial orientations. The two initial positions were front and back facing but participants were free to change the orientation using the arrow keys, as Watson [2003] in new work investigates how image rotation reduces simplification effectiveness. For the first set there were 74 trials per participant, which were organized 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 randomized.

During the experiments participants had to wear the eye-tracking device in order to record the necessary data. Before each experiment, calibration and drift correction had to be carried out to ensure the information was reliable. Also prior to each model being displayed, drift correction was performed again. Participants were told to examine each of the models carefully for the time they were displayed, bearing in mind that they would need to recognize them at a later stage. Models were displayed on a 21-inch monitor with diffuse, grey shading.

3.2 Results

While some trials had to be omitted due to calibration error, this was only 1.6% of all results. The information on fixations was summed over participants giving us the final data for each object. The results over all participants are best seen visually with a color map, which shows the important fixation data we use. The color map ranges from red through yellow, green, cyan, blue, magenta and finally to white with increasing total fixation length, increasing first fixation length and finally with increasing number of fixations (Figures 4, 5, 6 of the color plate).

As expected, perceptually important features like the eyes and the mouth, in the case of the natural objects, were viewed considerably more than the less salient features. For the man-made artifacts, prominent features include the straps of the sandal and the keys of the piano. For the second set, the cars' prominent features included the door handles and side mirrors. For the fish, attention appeared to be primarily focused on the upper fins and, like the animals, the eyes and the mouth were fixated on for a significant amount of time.

For the gears, the only symmetric objects, it was not clear that there were any prominent features, suggesting that this method may not be suitable for symmetric objects. Next we incorporated this data into a simplification method and evaluated the visual fidelity of each of these models.

4 Evaluation

4.1 Quadric Error Metric and Modifications

The method proposed by Garland and Heckbert [1997] utilizes iterative vertex pair contraction guided by a Quadric Error Metric (Figure 1). 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 [1997] 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.

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 neighboring pairs have their costs updated.
5. Steps 3 and 4 are continued until the model reaches the desired level of simplification.

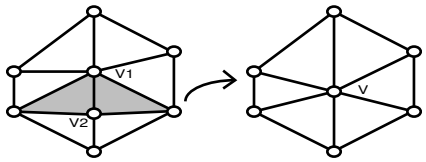

Figure 1: **Pair Contraction** - Selected Vertices are contracted to a single point. Shaded Triangles become degenerate and are removed.

With saliency data acquired from the eye-tracker, we created a modified quadric error metric which incorporated this data. 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 captured data was based on the faces of the evaluated model, not its vertices, weighting must be applied equally to each vertex of a face. For each vertex in the initial model the following equation was applied:

$$Q_w = Q_v + \omega(F_v)$$

Where Q_w is the *Weighted* quadric produced, Q_v is the initial quadric at the vertex and $\omega(F_v)$ is the *Weight* associated with the face that vertex v is a member of.

The weight $\omega(F_v)$ 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. To choose what combination of metrics to use, a quick survey was carried out. A group of 10 people were shown examples of models simplified using each individual metric and a combination of all three. 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 [2003] might also be effective but further testing would be needed to investigate this. For the three metrics, each value was normalized by the maximum value obtained for that metric and combined as follows:

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

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).

4.2 Finding the Naming Times

In these experiments, naming time was used as a measurement of visual quality. This involves someone seeing an object and then verbalizing the name that describes that object, so the objects must be of a familiar nature. Using the same stimuli as Watson et al. [2001] 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 (Figures 2 and 3 of the color plate). Finally, we investigated if the visual fidelity of the models was improved by using captured saliency data.

4.2.1 Participants & Method

Participants consisted of 27 volunteers, undergraduate and graduate students from the authors' department; 21 male and 6 female. All were naïve participants 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.

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 LODs. 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 randomized 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. 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 verbalizing 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.

4.2.2 Results

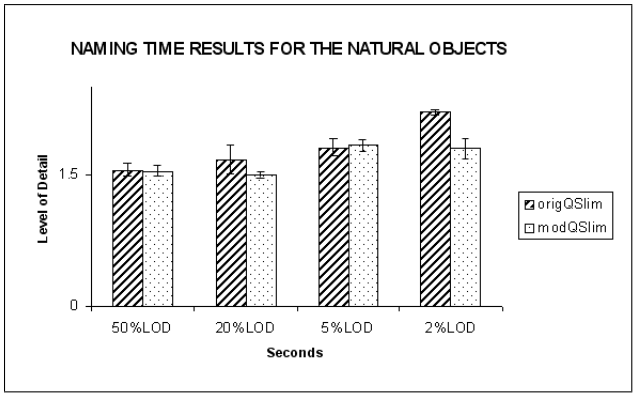

Figure 2: Naming times for the natural objects.

We recorded the naming time and the number of incorrectly named objects and applied within-subject ANOVAs (ANalysis Of VAriance across groups) to all of the results. 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. [2001] we found that only results at low LODs were significantly affected by level of simplification i.e., between 20% and 5% and between 5% and 2% there was a main effect of simplification level on results, 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 (all P-values $<$ 0.05). When comparing 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 (both P-values $<$ 0.05). This replicates previous psychological research, including Watson et al. [2000]. We found only one significant effect of simplification type. There was an interaction effect on the naming time for the natural objects at a very low LOD. At 2% LOD when averaged by objects (P-value $<$ 0.05) or participants (P-value $<$ 0.1) there was a reduction in the naming time when modified QSlim was used.

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 (Figure 2). Following the interesting results for familiar natural objects at a low LOD, we carried out further experiments to examine different categorical effects, this time using picture-picture matching time in the evaluation.

4.3 Acquiring the picture-picture matching times

Next, 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 verbalization. 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. [2002] 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 [Davis et al. 2003; Geren et al. 1997]. 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, but 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.

4.3.1 Participants & Method

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 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 less

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 recognize the individual cars and at the lowest detail they were no longer recognizable 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 randomized i.e, all LODs were mixed up within their own category only. All models were static.

Participants were seated in front of the computer and given printouts containing screen shots of the models as they appeared only at the highest LOD. 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 and 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 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. (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.)

4.3.2 Results

We recorded the average matching times and the number of correctly matched objects. We used 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 significant 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 (P-value < 0.05) and participants (P-value < 0.05). For these models between 5% and 2% level of detail, when averaged by objects there was a significant difference (P-value < 0.05) and a marginally significant one when averaged by participants (P-value < 0.1). For the fish objects between 5% and 2% LOD, when averaged by objects and participants there were marginally significant results (both P-value < 0.1). For the gear objects between 14% and 5%, when averaged by objects and participants there was a significant result (all P-values < 0.05). Between 5% and 2% when averaged by objects there was a significant result (P-value < 0.01).

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% (P-value < 0.05 and P-value < 0.01). For the fish objects, averaged by object there was a significant result between 5% and 2% LOD (P-value < 0.01). For the gear objects between 14% and 5% there were significant results when averaged by objects (P-value < 0.01) and marginally significant results when averaged by participants (P-value < 0.1). Again when averaged by objects, between 5% and 2% there was a significant decrease (P-value < 0.01).

Next, bearing in mind 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 to be named 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% when modified QSlim was used (P-value < 0.1) and a significant reduction at 5% and 2% (P-value < 0.05 and P-value < 0.01 respectively). When averaged by participant at 5% there was also a marginally significant reduction (P-value $<$ 0.1). Results for the number of correctly matched animal objects at 14% averaged by objects show a marginally significant increase in the number of correctly matched objects when modified QSlim was used (P-value < 0.1). For the animal models averaged by objects at 5% and 2% there was a significant increase (all P-values < 0.05). Again at 5% when averaged by participants there is marginally significant increase (P-value < 0.1) (Figure 3). There was a significant increase in the number of correctly matched fish when averaged by objects at 30% (P-value < 0.05). Also there was a significant increase in the number of correctly matched gears when averaged by objects and participants at 30% (both P-values < 0.05).

Matching time results show that, like naming time, there is a main effect of simplification level only at the lower LODs. However, 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 probabilistic concept known as cue validity. As describe by Rosch [1976], a category with a high cue validity is more differentiated from other categories than one with low cue validity. Perhaps the cars could be described as a subordinate category because they share more attributes in common than the other categories and hence the low cue value. 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, or because they could be classified as a basic level category, with a higher cue validity as opposed to a subordinate one [Rosch 1976]. At the lower LODs there was an interaction effect, 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.

4.4 Forced-choice preferences experiments

Finally, we carried out an experiment in which both sets of models could be included. The experimental technique used was forced-

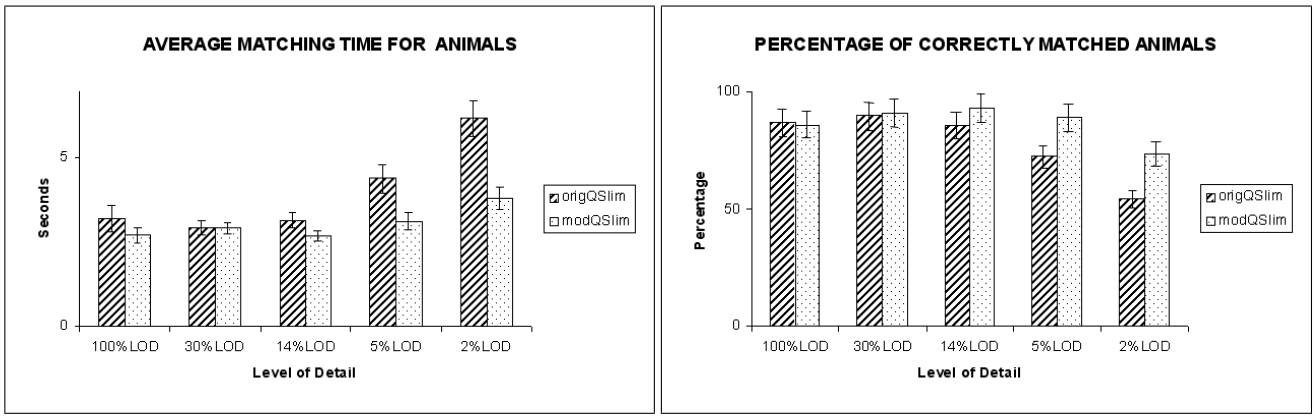

Figure 3: Comparing the percentage of correctly matched and the average matching times for the animal models.

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. All models under the two types of simplification were compared at the same simplification level.

4.4.1 Participants & Method

Sixty eight people participated in each part of this experiment. Sixty male and 8 female in the first part and 51 male and 17 female in the second part. There were both graduate students and staff from the authors' 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, those of natural objects and man-made artifacts. The images used were screen shots of the 333 stimuli from the naming time experiment. Images were created from the standard and the simplified models, resulting in nine examples of each model. 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 original versions of the models were on the left and half on the right, in random order. The left and right position of the original (modified) model was distributed evenly throughout the different pages.

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 center. 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 genuine 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.

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.

4.4.2 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 and can be seen in Figure 6. 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 modified over the original models (all P-value < 0.05). However, results for the man-made artifacts show that marginally significantly more people (P-value < 0.1) chose the models simplified using the original version of QSlim at 20% and significantly more chose them at the 5% and 2% LODs (all P-values < 0.05). 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 (all P-values < 0.05).

Forced-choice preference seems like the better predictor as it demonstrates that saliency guided simplification can work for unfamiliar natural objects as well as familiar ones, which was not apparent from the matching times results. It also produces some preferences at higher LODs for the modified natural objects and the original man-made objects. Importantly results show that, while saliency based simplification does work for natural objects, it actually reduces the visual quality of familiar man-made artifacts and does not produce any significant results for the car and the gear ob-

jects. A reason for this may be that man-made artifacts are generally related to a task and that prominent features may be defined by this and not the specific object. As described by Hayhoe [2000], when a participant's eye-movements were tracked while making a snack, results showed that almost all of the fixations focused on the task, rarely focusing 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. [2003] 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.

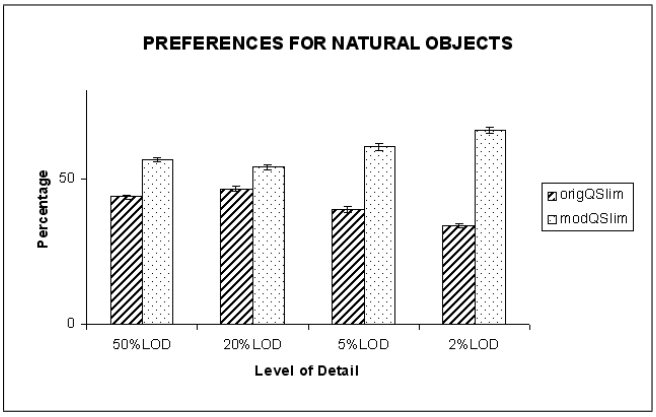

Figure 4: Percentage preferences for the natural objects.

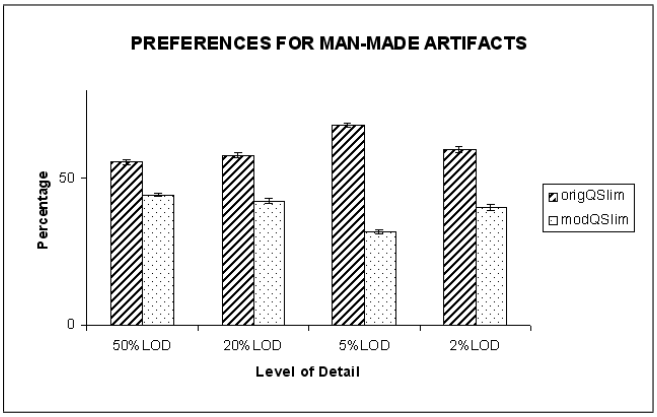

Figure 5: Percentage preferences for the man-made artifacts.

5 Conclusions and Future Work

In this paper we described our research in which we examined whether visual fidelity would be improved by emphasizing the detail of automatically-detected salient features of models at the expense of unimportant areas. The saliency data ascertained using the eye-tracking device showed that there were prominent features in the case of some objects. We examined naming times, picture-picture matching times and forced-choice preferences for models simplified using the original version of QSlim and the modified version of this software, to see if our saliency guided simplification method works on certain categories of models. The first set of evaluation results show that the modified form of simplification produces better naming time results on familiar natural objects at a

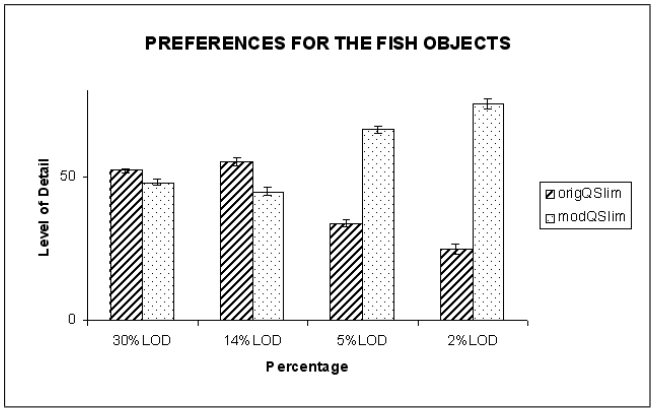

Figure 6: Percentage preferences for the fish objects.

low LOD. Matching times also suggest that low level familiar natural objects can have their visual quality enhanced by using saliency data. Results suggest that forced-choice preferences are the best indicators of visual fidelity and these results show saliency based simplification 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 recognizing these objects is made easier.

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 is 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 insight which could be used when rendering scenes which contain a very large number of objects, like during crowd simulation. Results show, 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 [Kho and Garland 2003; Pojar and Schmalstieg 2003], we have experimentally established that using this data as weights in the simplification process can help to preserve the visual fidelity of low quality models for longer.

References

CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: Exploiting visual tasks for selective rendering. In *Proceedings of the 2003 EUROGRAPHICS Symposium on Rendering*, EUROGRAPHICS, P. Christensen and D. Cohen-Or, Eds., 270–280.

CIGNONI, P., MONTANI, C., ROCCHINI, C., AND SCOPIGNO, R. 1998. Zeta: A resolution modeling system. In *GMIP: Graphical Models and Image Processing 60*, vol. 5, 305–329.

DAVIS, M. H., MOSS, H. E., DE MORNAY, P., AND TYLER, L. K. 2003. Spot the difference: Investigations of conceptual structure for living things an artificacts using speeded word-picture matching. In *Department of Experimental Psychology, Universityof Cambridge*.

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

FITTS, P. M., JONES, R., AND MILTON, J. L. 1950. Eye movements of aircraft pilots during instrument-landing approaches. In *Aeronautical Engineering Review*, vol. 9(2), 24–29.

GARLAND, M., AND HECKBERT, P. 1997. Surface simplification using quadric error metrics. In *Computer Graphics Proceedings, Annual Conference Series*, 209–216.

GEREN, M. A., STROMER, R., AND MACKAY, H. A. 1997. Picture naming, matching to sample, and head injury: a stimulus control analysis. In *Journal of applied behaviour analysis*, vol. 30, 339–342.

HAYHOE, M. 2000. Vision using routines : A functional account of vision. In *In Visual Cognition 2000*, vol. 7(1/2/3), 43–64.

HENDERSON, J. M., AND HOLLINGWORTH, A. 1998. Eye movements during scene viewing: An overview. In *G. Underwood (Ed.), Eye Guidance in Reading and Scene Perception*, 269–294.

HENDERSON, J. M. 1992. Object identification in context: The visual processing of natural scenes. In *Canadian Journal of Psychology*, vol. 46(3), 319–341.

ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-based visual attention for rapid scene analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20(11), 12541259.

JACOB, R. J. K. 1993. Eye-movement-based human-computer interaction techniques: Toward non-command interfaces. In *Advances in Human-Computer Interaction*, vol. 42, 151–190.

KHO, Y., AND GARLAND, M. 2003. User-guided simplification. In *Proceedings of ACM Symposium on Interactive 3D Graphics*.

LAWSON, R., BÜLTHOFF, H., AND DUMBELL, S. 2002. Interactions between view changes and shape changes in picture-picture matching. In *Technical Report No. 095*.

LI, G., AND WATSON, B. 2001. Semiautomatic simplification. In *In ACM Symposium on Interactive 3D Graphics 2001*, 43–48.

LUEBKE, D., AND HALLEN, B. 2001. Perceptually-driven simplification for interactive rendering. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*.

POJAR, E., AND SCHMALSTIEG, D. 2003. User-controlled creation of multiresolution meshes. In *Proceedings of ACM Symposium on Interactive 3D Graphics*.

REDDY, M. 1998. Specification and evaluation of level of detail selection criteria. In *Virtual Reality: Research, Development and Application*, vol. 3(2), 132–143.

ROSCH, E. 1976. Natural categories. In *Cognitive Psychology*, vol. 4, 328–350.

RUSHMEIER, H. 2001. Metrics and geometric simplification. In *Siggraph Course Notes*.

WATSON, B., WALKER, N., HODGES, L. F., AND WORDEN, A. 1997. Managing level of detail through peripheral degradation: Effects on search performance with a head-mounted display. In *ACM Trans. on Computer-Human Interaction*, vol. 4, 323–346.

WATSON, B. A., FRIEDMAN, A., AND MCGAFFEY, A. 2000. Using naming time to evaluate quality predictors for model simplification. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 113–120.

WATSON, B., FRIEDMAN, A., AND MCGAFFEY, A. 2001. Measuring and predicting visual fidelity. In *Computer Graphics Proceedings, Annual Conference Series*, 213–220.

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

YEE, H., PATTANAIK, S., AND GREENBERG, D. P. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. ACM Press, vol. 20, 39–65.