Predicting and Evaluating Saliency for Simplified Polygonal Models

SARAH HOWLETT, JOHN HAMILL, and CAROL O'SULLIVAN
Image Synthesis Group, Trinity College Dublin

In this paper, we consider the problem of determining feature saliency for three-dimensional (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 the simplified models, 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 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. 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 focused upon during the naming, matching, and forced-choice preference tasks. Results demonstrated that the heads of natural objects received a significant amount of attention, especially during the naming task. We hope that our results will lead to new insights into the nature of saliency in 3D graphics.

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

Additional Key Words and Phrases: 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) 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 as well as geometry, into consideration. As the perceptual importance

Authors' address: email: Sarah.Howlett, John.Hamill, Carol.OSullivan@cs.tcd.ie.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 1515 Broadway, New York, NY 10036 USA, fax: +1 (212) 869-0481, or permissions@acm.org.
© 2005 ACM 1544-3558/05/0700-0001 $5.00

Fig. 1. The SMI EyeLink eye-tracking device.

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, 2000] on the first set of familiar objects and picture–picture matching times [Lawson et al. 2002] on the second set to determine if familiarity played a role; forced-choice preferences were used 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 toward the models simplified using the fixation data, especially at the lower LODs. In addition, in a final experiment, we used the eye-tracker to see which aspects of some objects received the most attention during the actual tasks of naming, matching, and forced-choice preferences. 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.

2. BACKGROUND

Recent work on this problem included 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. Thus, 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 unfamiliar 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.

Having used naming, matching, and preference tasks to test if the visual fidelity of models simplified, using the modified version of QSlim, was enhanced, 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. We wished to confirm our previous results and determine whether the different natures of these tasks influenced where a participant fixated.

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 (250 hz) manufactured by SensorMotoric Instruments was used to obtain 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. We then 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 microsaccades that are unintentional. 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 focused 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 focused 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 20×20 pixels represented a good zone of interest.

3.1 Participants and 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 four 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 nonsymmetric. 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; natural objects and 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; animals and car models; the second block contained all the fish and gear models. The models were randomized within each group.

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 black through to white with increasing total fixation length, increasing first-fixation length, and, finally, with increasing number of fixations (Figures 2–4).

As expected, perceptually important features like the heads, eyes, and 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 buttons of the blender. 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 mouth were fixated on for a significant amount of time. For the gears, there were only symmetric objects. Thus, it was not
[AQ1] clear whether 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 5). The method calculates a quadric **Q** for each vertex in the initial

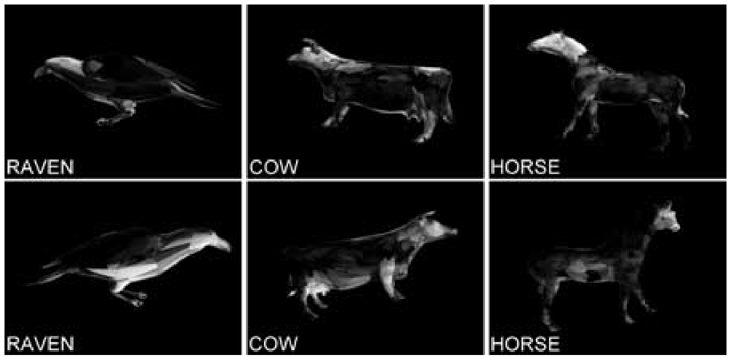

Fig. 2. Results from the saliency experiment (white representing the greatest number): the total length of fixations on the familiar natural objects.

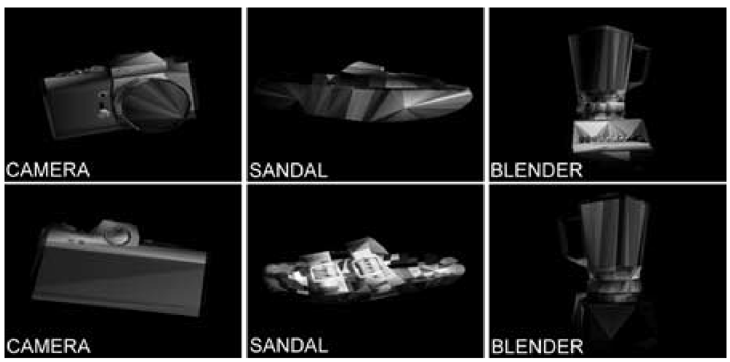

Fig. 3. Results from the saliency experiment (white representing the greatest number): the duration of the first fixations on the man-made artifacts.

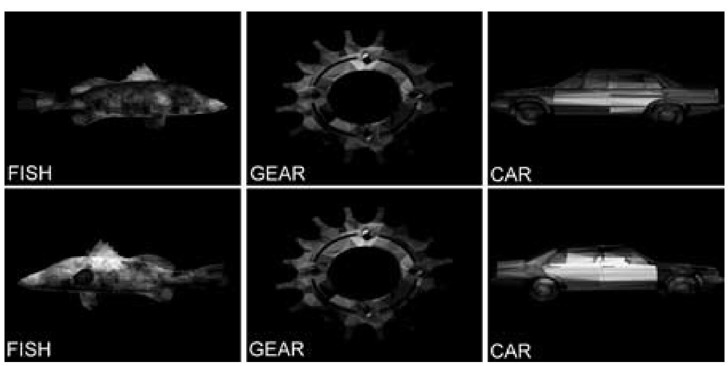

Fig. 4. Results from the saliency experiment (white representing the greatest number): the total number of fixations on the unfamiliar objects in the second set.

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.

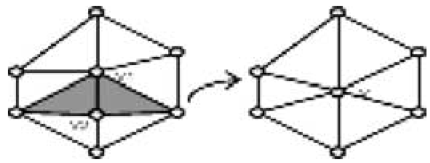

Fig. 5. Pair contraction. Selected vertices are contracted to a single point. Shaded triangles become degenerate and are removed.

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

With saliency data acquired from the eye-tracker, we created a modified quadric error metric that 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, and 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 involved someone seeing an object and then verbalizing the name that described that object, so the objects had to be of a familiar nature. Watson et al. [2000, 2001] carried out two sets of naming time experiments . 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

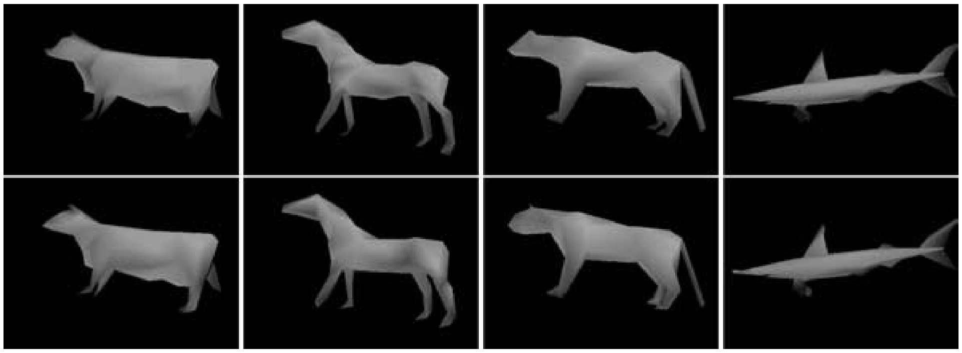

Fig. 6. Models simplified to 5% LOD using the original (*top*) and modified (*bottom*) simplification approach.

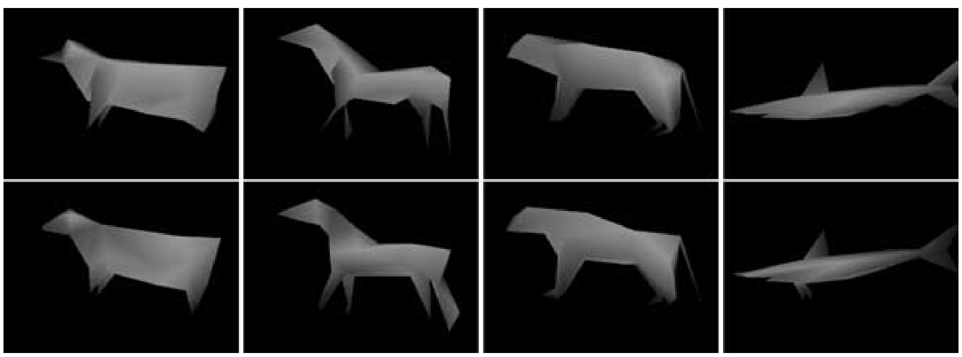

Fig. 7. Models simplified to 2% LOD using the original (*top*) and modified (*bottom*) simplification approach.

a much lower detail level than Watson (Figures 6 and 7). 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.

4.2.1 *Participants and Method.* Participants consisted of 27 volunteers (21 male and 6 female) undergraduate and graduate students from the authors' department. 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. First, 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. Second, 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, with 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

Table I. Effects of Simplification Level on Naming Time Results

AVE BY	LOD(%)	ANOVA	p VALUE
objects	20 & 5	$F(1, 52) = 5.73$	0.02
objects	5 & 2	$F(1, 52) = 4.42$	0.04
participants	20 & 5	$F(1, 36) = 7.33$	0.01
participants	5 & 2	$F(1, 34) = 8.25$	<0.01

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. 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.* 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. Thus, 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 level of detail 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. Of all incorrectly named objects 58.1 and 25.6% 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] 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 an 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 (Tables I and II). 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 (Table III). This replicates previous psychological research, including by Watson et al. [2000].

Table II. Effects of Simplification Level on Number of Errors in Naming Time Experiment

AVE BY	LOD(%)	ANOVA	p VALUE
objects	20 & 5	$F(1, 52) = 11.35$	0.02
objects	5 & 2	$F(1, 52) = 48.24$	0.04
participants	20 & 5	$F(1, 36) = 04.95$	0.03
participants	5 & 2	$F(1, 36) = 17.63$	<0.01

Table III. Effects of Object Type and Simplification Type on the Naming-Time Results

VARIABLE	AVE BY	LOD(%)	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
simp type	objects	2	$F(1, 32) = 4.54$	0.04
simp type	participants	2	$F(1, 24) = 3.77$	0.06

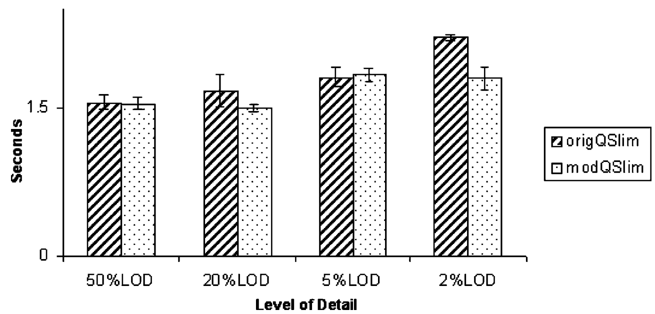

Fig. 8. Naming times for the natural objects.

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 or participants, there was a reduction in the naming time when modified QSlim was used (Table III).

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 8). 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

We then evaluated picture–picture matching time as a measure of visual quality and compared categories to examine the effects of familiarity, while bearing in mind that the number of polygons at each **[AQ2]** LOD was not uniform. 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, similar criteria were used for the cars and gears. This meant that at the lower LODs, objects **[AQ2]** 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 and Method.* A total of 28 participants were involved in this experiment, one-half for the original simplification method and one 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. 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 print-outs 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 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.)

Table IV. Effects of Simplification Level on Matching Time

AVE BY	LOD(%)	ANOVA
objects	Animals 14 and 5	$F(1, 26) = 06.79$, $p = 0.01$
participants	Animals 14 and 5	$F(1, 12) = 04.79$, $p = 0.04$
objects	Animals 5 and 2	$F(1, 26) = 07.35$, $p = 0.01$
participants	Animals 5 and 2	$F(1, 12) = 03.21$, $p = 0.09$
objects	Fish 5 and 2	$F(1, 26) = 04.20$, $p = 0.05$
participants	Fish 5 and 2	$F(1, 14) = 03.38$, $p = 0.09$
objects	Gears 14 and 5	$F(1, 26) = 13.49$, $p < 0.01$
participants	Gears 14 and 5	$F(1, 14) = 06.01$, $p = 0.02$
objects	Gears 5 and 2	$F(1, 26) = 06.80$, $p = 0.01$

Table V. Effects of Simplification Level on Number of Correctly Matched Objects

AVE BY	LOD(%)	ANOVA
objects	Animals 14 and 5	$F(1, 26) = 06.65$, $p = 0.01$
objects	Animals 5 and 2	$F(1, 26) = 13.70$, $p < 0.01$
objects	Fish 5 and 2	$F(1, 26) = 07.95$, $p < 0.01$
objects	Gears 14 and 5	$F(1, 26) = 21.92$, $p < 0.01$
participants	Gears 14 and 5	$F(1, 14) = 03.58$, $p = 0.08$
objects	Gears 5 and 2	$F(1, 26) = 25.86$, $p < 0.01$

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 and 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% level of detail, 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 (Table IV).

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 (Table V).

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 matched in 3.14 s, followed by the fish (4.51 s), then the gears (4.74 s); the cars were the slowest (6.40 s).

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 and a significant reduction at 5 and 2%. When averaged by participant at 5%, there was also a marginally significant reduction (Table VI).

Table VI. Effects of Simplification Type on Results for Matching Time

AVE BY	LOD(%)	ANOVA
objects	Animals 14	$F(1, 26) = 03.68$, $p = 0.07$
objects	Animals 5	$F(1, 26) = 06.06$, $p = 0.02$
participants	Animals 5	$F(1, 12) = 03.70$, $p = 0.08$
objects	Animals 2	$F(1, 26) = 13.14$, $p < 0.01$

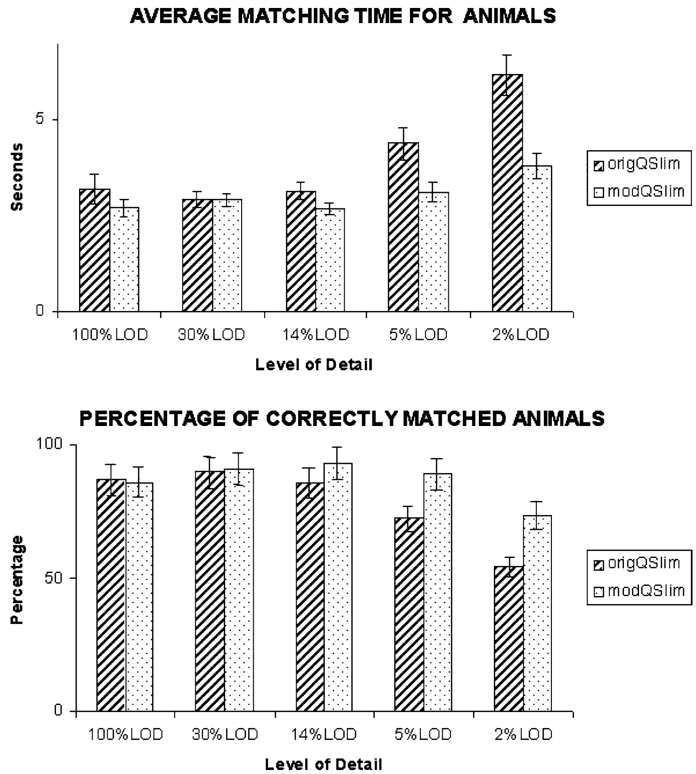

Fig. 9. Comparing the percentage of correctly matched and the average matching times for the animal models.

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 9). There was a significant increase in the number of correctly matched fish, when averaged by objects at 30% (p value < 0.05). There was also 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 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

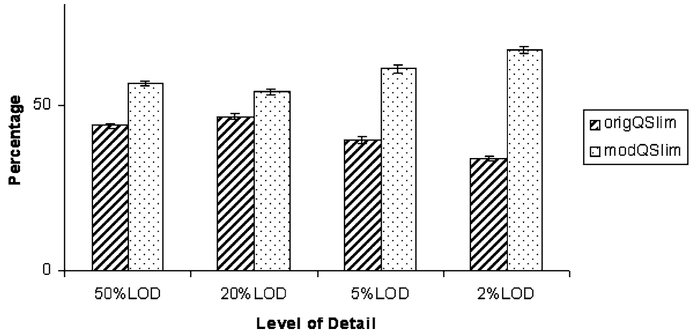

Fig. 10. Percentage preferences for the natural objects.

[AQ4]

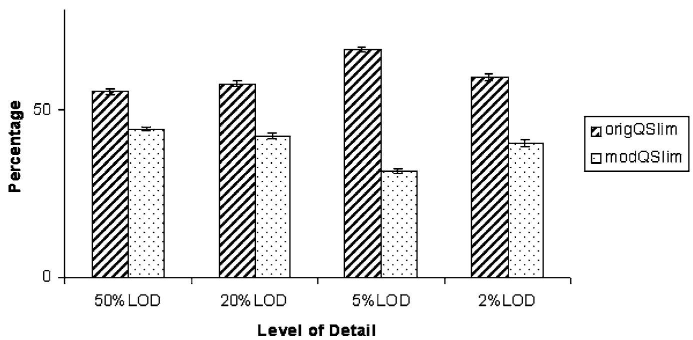

Fig. 11. Percentage preferences for the man-made artifacts.

[AQ4]

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, since 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-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 and Method.* 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 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—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 ten different random orderings of the models, giving rise to 40 unique web pages, which were assigned to participants in sequence. On each page, one-half of the original versions of the models were on the left and one-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, upon 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. The participant then 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 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 modified over the original models. 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
[AQ3] significant preference for the modified simple models (Table VII).

[AQ5]

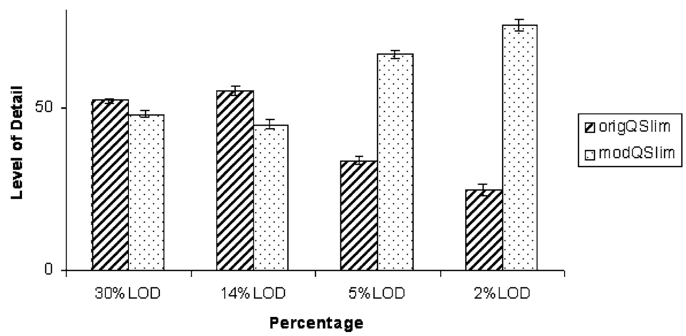

Fig. 12. Percentage preferences for the fish objects.

Table VII. Significant Effects of Simplification Type on the Preferences (all $P < 0.05$).

PREFERENCE	AVE BY	LOD(%)	ANOVA
modified simp	participants	natural 50	$F(1, 36) = 07.7$
modified simp	participants	natural 5	$F(1, 36) = 05.8$
modified simp	participants	natural 2	$F(1, 36) = 20.9$
original simp	participants	man-m 5	$F(1, 34) = 07.4$
original simp	participants	man-m 2	$F(1, 34) = 24.2$
modified simp	participants	fish 5	$F(1, 14) = 09.9$
modified simp	participants	fish 2	$F(1, 14) = 08.9$

Forced-choice preferences responded more strongly than the other predictors used in the experiments. It demonstrates that saliency-guided simplification can work for 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 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 objects. 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. This suggests 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.

5. VALIDATION OF RESULTS

5.1 Face "Pop-out"

Based on our results indicating that the heads of natural objects were particularly salient, we decided to investigate this further. We also wished to examine further 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.

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 similar to other basic elements, like color, e.g., when a green dot pops out of a field containing red dots. “Pop-out” can be proved through visual search tasks by showing that processing is done in parallel and preattentively, when an increase in the number of distractors results in a minimal increase in reaction time. Hochstein et al. [2004] and Hershler and Hochstein [2003] found that human faces popped out from a background of varied photograph distractors. However, they demonstrated that this effect does not generalize to animal faces.

Contrary to this, there is much evidence to suggest that faces are not accessed preattentively in parallel. In his experiments, Nothdurft [1993] 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 [1994] showed that none of the search conditions involving distractors containing face features resulted in “pop-out” and only nonface 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, as there is a certain biological significance of human faces and our familiarity with them. Suzuki and Cavanagh [1995] showed that facial organization preempted combined low-level features (curvature features). They showed that rapid search processes operated on high-level representations, such as faces, even when low-level features were more efficient. Brown et al. [1997] 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 benefitted from training with the stimuli. Recent clinical research may suggest that the specialized cells thought to be only for human face processing may stretch to other classes of stimuli. The term “prosopagnosia,” refers to a selective impairment in face recognition, but not object recognition. Work with patients suffering from this by Levine and Calvanio [1989] demonstrated impaired object recognition, in some cases, across three classes of stimuli: objects, animals, and faces and not just with human faces as previously thought. This could mean that there may be a possibility that animals or maybe even animal faces could be treated in a similar fashion to human faces.

5.2 Validation Experiments

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 focused on during the three types of experiments. However, for the naming and matching tasks, we made the decision to only use models at full 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 in the 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 it. 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 spread more all 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. 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.

5.3 Participants and Method

Ten participants were involved in this experiment. There were 8 males and 2 females, ages 22–27, with 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 randomized.

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. Comparing models on a sheet of paper was not required, otherwise, it would have been very difficult to accurately analyze 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 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 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.

5.4 Results

5.4.1 *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

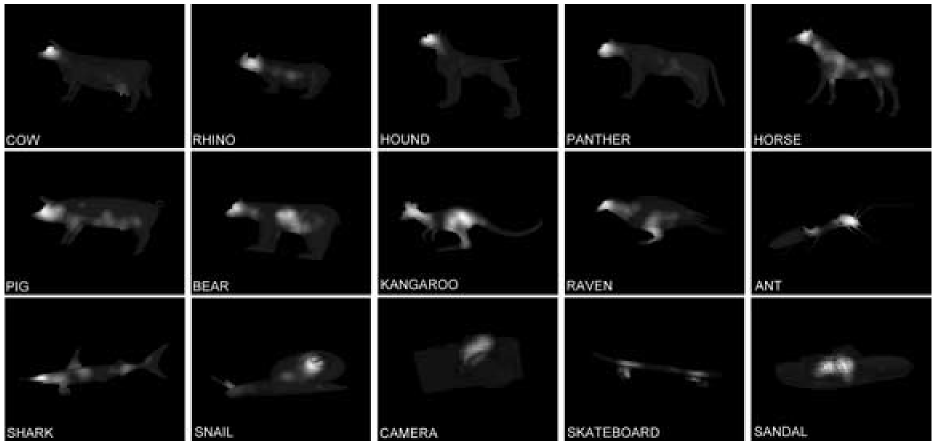

Fig. 13. Fixation maps of all fixations for some objects in the naming–time experiments.

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.

First we examined all fixations on each of the natural objects (Figure 13). We cannot claim that the heads of the animals "popped-out" preattentively, but they were definitely salient features that 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. 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 focused on the most.

We also examined whether attention was immediately drawn to the head of the natural objects. 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 focused upon the neck region. Perhaps the participants were saccading toward the head, the most prominent feature, but underestimated the distance and fixated half way there (Figure 14). However, when the average of the first and second fixations were used, it was indeed the case, as before, that faces were immediately fixated upon.

Following this, we examined the fixation maps for the man-made artifacts (see the bottom row of Figure 13). In most cases, attention was not drawn to any specific features. Generally the participants' attention was centered on the area around the fixation dot that was present prior to the experiment. For the camera, it seems that attention was drawn toward the writing, for the truck and the fighter jet toward 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, and the straps of the sandals. 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 earlier. Objects disappeared from view as soon as they were named, nor were participants allowed to

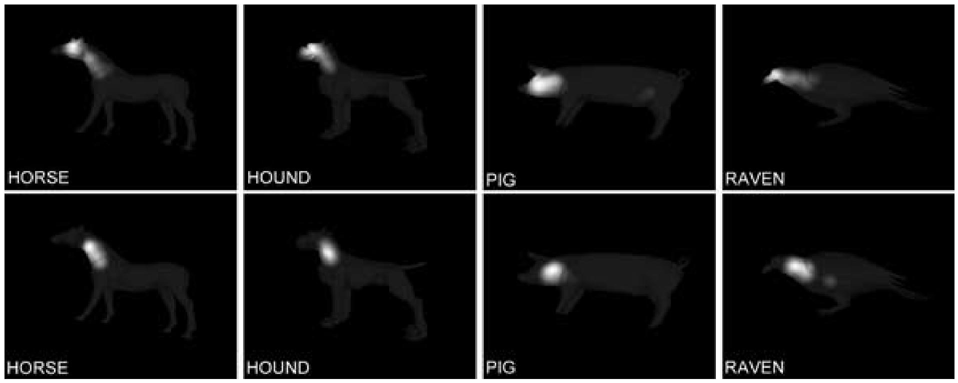

Fig. 14. Fixation maps of the combined first and second (*top*) and first (*bottom*) for some natural objects in the naming–time experiments.

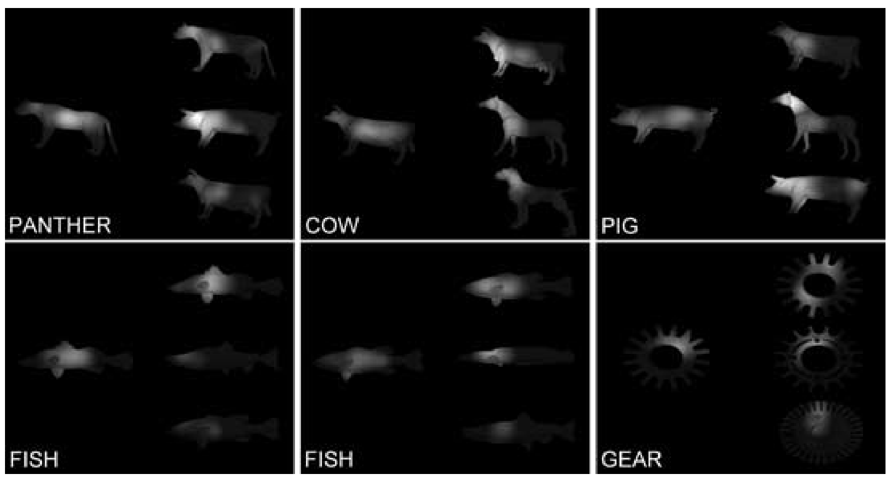

Fig. 15. Fixation maps of all fixations for the matching–time experiments for three of the animal objects, two of the fish objects, and one of the gear objects, respectively.

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

5.4.2 *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 (Figure 15). During the task of matching, the example models on the left only received a very small amount of attention, mainly around the center 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. The region of interest seemed to be much larger than in the naming task, so the upper body also received 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, although generally the middle objects did get some attention (Figure 15). Perhaps this was due to the similarity of all the fish object (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 focused on the example models or the upper half of the middle comparison ones. The fish objects results do not compare as 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; attention was 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 divided all over the gear objects.

5.4.3 *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 (Figure 16). 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 toward 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 mainly focused on the center of the objects. 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.

5.5 Discussion

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 generalize to animal faces [Hershler and Hochstein 2003] and that the degree of animal face pop-out is extremely variable [Hochstein et al. 2004], our results have demonstrated that the heads of these natural objects are particularly salient. We are not claiming that they are processed preattentively or in parallel. However, when the task involves recognizing 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, also the matching and forced-choice tasks. This is further backed up by our previous

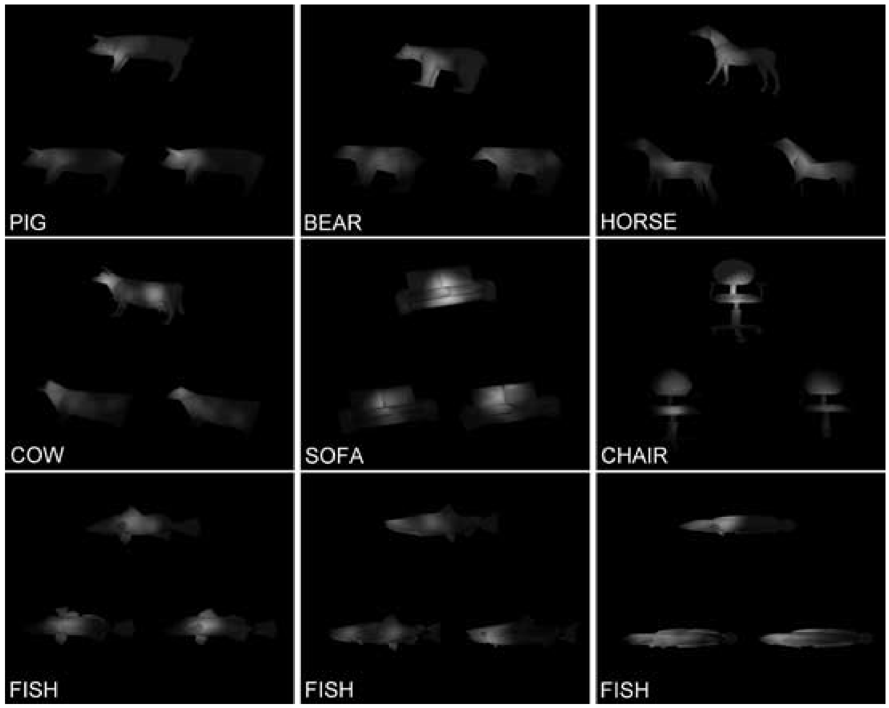

Fig. 16. Fixation maps of all fixations for the forced-choice experiments for four of the animal objects, two of the man-made artifacts, and three of the fish objects, respectively.

evaluation experiments which shows positive naming, matching, and forced-choice preference results for the natural objects at low levels of detail 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-overestimation 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 focused on the center of these objects. For the man-made artifacts and car objects results do not correspond exactly to those of the original saliency experiment, however, this might explain why no positive results were found for these object during the naming, picture–picture matching, and forced-choice preference experiments.

6. CONCLUSIONS AND FUTURE WORK

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 showed that the modified form of simplification produced better naming-time results on familiar natural objects at a low LOD. Matching times also suggest that low-level familiar natural objects can have their visual quality enhanced by using saliency data.

Forced-choice preference results show that our experiments had more of and effect on visual difference than visual recognition. Preferences responded most strongly in our experiments and results show that saliency-based simplification can work for unfamiliar 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.

Our final experiment confirmed the evaluation experiments carried out previously. 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 focused here, with little or no attention given elsewhere, in many cases. Although in the saliency experiment it appears that there are prominent features for the cars and some of the man-made artifacts, the results found here that participants tend to look only at the center of the object is consistent with our evaluation results: that saliency information on man-made artifacts retained during simplification 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 comparison task, if, for example, the spout of a teapot 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 also play a role in visual fidelity too. Furthermore, the goal of our research is not to convince others to use an eye-tracke. 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 that could be used when rendering scenes which 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 [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.

ACKNOWLEDGMENTS

This work was supported by the Higher Education Authority (HEA).

REFERENCES

BROWN, V., HUEY, D., AND FINDLAY, J. M. 1997. Face detection in peripheral vision: do faces pop out? *Perception 26*, 1555–1570.

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*, P. Christensen and D. Cohen-Or, Eds. EUROGRAPHICS, 270–280. **[AQ6]**

CIGNONI, P., MONTANI, C., ROCCHINI, C., AND SCOPIGNO, R. 1998. Zeta: A resolution modeling system. *GMIP: Graphical Models and Image Processing 60*, 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 artifiacts using speeded word-picture matching. Department of Experimental Psychology, University of Cambridge, Cambridge.

DUCHOWSKI, A. T. 2002. A breadth-first survey of eye tracking applications. *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. *Aeronautical Engineering Review 9*, 2, 24–29.

GARLAND, M. AND HECKBERT, P. 1997. Surface simplification using quadric error metrics. *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. *Journal of Applied Behaviour Analysis 30*, 339–342.

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

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

HENDERSON, J. M. AND HOLLINGWORTH, A. 1998. Eye movements during scene viewing: An overview. In *Eye Guidance in Reading and Scene Perception* G. Underwood, (Ed.), 269–294. **[AQ7]**

HERSHLER, O. AND HOCHSTEIN, S. 2003. High-level effects in rapid 'pop-out' visual search. In *ECVP 2003 Abstract*.

HOCHSTEIN, S., BARLASOV, A., HERSHLER, O., NITZAN, A., AND SHNEOR, S. 2004. Rapid vision is holistic [abstract]. In *Journal of Vision 4*. **[AQ8]**

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 20*, 11, 1254–1259. **[AQ9]**

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

KHO, Y. AND GARLAND, M. 2003. User-guided simplification. In *Proceedings of ACM Symposium on Interactive 3D Graphics*. **[AQ10]**

KUEHN, S. M. AND JOLICOEUR, P. 1994. Impact of quality of the image, orientation, and similarity of the stimuli on visual search for faces. *Perception 23*, 95–122.

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

LEVINE, D. N. AND CALVANIO, R. 1989. Prosopagnosia: A defect in visual configural processing. *Brain and Cognition 10*, 149–170.

LI, G. AND WATSON, B. 2001. Semiautomatic simplification. 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*. **[AQ11]**

NOTHDURFT, H.-C. 1993. Faces and facial expressions do not pop out. *Perception 22*, 1287–1298.

POJAR, E. AND SCHMALSTIEG, D. 2003. User-controlled creation of multiresolution meshes. In *Proceedings of ACM Symposium on Interactive 3D Graphics*. **[AQ11]**

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

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

RUSHMEIER, H. 2001. Metrics and geometric simplification. In *Siggraph Course Notes*. **[AQ12]**

SUZUKI, S. AND CAVANAGH, P. 1995. Facial organization blocks access to low-level features: An object inferiority effect. *Human Perception and Performance 21*, 901–913.

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

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 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.

YEE, H., PATTANAIK, S., AND GREENBERG, D. P. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. In *ACM Trans. Graph.* Vol. 20. ACM Press, New York, 39–65.

Received December 2004; revised March 2005; accepted March 2005

Author Queries

Title: Predicting and Evaluating Saliency for Simplified Polygonal Models

Authors: Sarah Howlett, John Hamill, and Carol O'Sullivan

AQ1: Is this your meaning?
AQ2: OK changes.
AQ3: OK?
AQ4: OK to delete infor. found in legend.
AQ5: OK to delete?
AQ6: Location of Publisher?
AQ7: Pls. supply Publ. & location.
AQ8: Pages?
AQ9: OK pages: 1254–1259.
AQ10: Pages?
AQ11: Pages?
AQ12: More infor? Publisher? Page? San Diego, CA?