Perceptual Evaluation of Impostor Representations for Virtual Humans and Buildings

J. Hamill, R. McDonnell, S. Dobbyn, and C. O'Sullivan

Image Synthesis Group, Computer Science Department, Trinity College Dublin, Ireland

Abstract

In large-scale simulations involving complex scenes, such as cities inhabited by crowds, simplifications are almost always necessary to achieve interactive frame-rates. Level of Detail (LOD) techniques such as reducing geometric complexity, or substituting impostor images for geometry, are usually employed. Image-based or impostor techniques have been gaining in popularity in recent years, along with hybrid methods that combine impostors and geometry, but perceptual issues with respect to such representations have been largely neglected to date.

In this paper we evaluate the effectiveness of impostor representations for the real-time rendering and animation of static buildings and dynamic virtual humans. Using sets of psychophysical experiments, we establish some thresholds at which impostors are effective for static and dynamic objects, along with criteria for selecting transitions to geometry and update rates. We also compare the impact of two model representations (geometry and impostor), on the perception of human motion. We have found that impostors are an extremely effective substitute for detailed geometry in the target application area.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Virtual Reality, Perception

1. Introduction

Real-time applications such as games or urban simulations are often highly complex. User expectations grow year by year, with the desire for added detail and believability foremost in their minds. Computing and rendering hardware always manifest performance limitations, and so the use of simplification techniques to trade accuracy for performance is becoming more and more popular. The goal is to increase performance while maintaining visual fidelity.

Instead of simulating and rendering a small number of highly detailed models or characters (and potentially fading out models in the distance), a system can be designed to use many more low detail representations when they will be less noticeable to the user of a system, such as when they are far away from a viewer. The system can also switch between the representations as the object in question moves into the focus of the user's attention. While simplified geometric models provide a good tradeoff of performance against quality, the use of image based impostors can provide optimal performance, for minimal loss in visual detail. This can allow large animated crowds to be rendered, for example [DHOO05], or for cities and environments to continue to the horizon [WWS, SDB97]. In recent years the use of impostor techniques has become quite popular, but the evaluation of the effectiveness of this representation technique is an area that has been largely ignored thus far.

This paper details a series of psychophysical experiments in which the effectiveness of using impostor based representations is evaluated under various conditions. The experiments are designed to establish thresholds for the use of impostors in situations where they will not detract from the visual fidelity of the scene. The first group of experiments investigates when static impostors can be a valid substitute for large buildings and also evaluates impostor validity for replacing dynamic human models. The second group measures sensitivity to changes in human motion with geometric and impostor representations. The results of these experiments provide some concrete measures for the sensitivity of users to the use of impostor representations, and thus are of importance to developers of realistic real-time applications.

2. Background Work

Large scale environments such as cities are difficult to render in real-time applications. The high numbers of buildings combined with the large area covered sometimes leads to intractable hardware limitations. The use of image based impostors for representing urban environments has become quite popular in recent times as a way of alleviating some of these problems. Maciel et al. [MS95] used textured clusters for visualising large environments while Sillion et al. [SDB97] used image based impostors for the acceleration of the rendering of a model of the city of Paris. Decoret et al. [DSSD99] extended this with multiple layers to reduce parallax artifacts, and further to clouds of impostors useful for extreme simplification [DDSD03]. Wimmer and Wonka demonstrated point based impostors for urban simulations [WWS] while Jeschke and Wimmer showed layered environment-map impostors and textured depth meshes that are applicable to urban rendering and recently provided a comprehensive overview of such image-based rendering techniques in [JWP05].

In the work of Tecchia et al. [TLC02], a virtual human's geometric representation was replaced by a pre-generated impostor. This involved the offline rendering of a set of images of the human model from different viewpoints for multiple frames of animation. At run time, depending on the human's current frame of animation and position with respect to the viewer, the most appropriate impostor image is selected and displayed on a quadrilateral dynamically orientated towards the viewer. Using this approach, Heigeas et al. [HLTC03] simulated crowds containing up to 2000 people in an ancient Greek town. More recently, we [DHOO05] improved on this approach by implementing the shading of the impostor through graphics hardware and succeed in rendering large realistic crowds containing several thousand individuals in an urban city environment. We switch to a geometric representation according to a *Pixel to Texel ratio* when the impostor would become pixelated. A pixel to texel ratio refers to the number of screen pixels occupied by an image (e.g. a 256x256 pixel image, occupying exactly 256x256 pixels on screen has a one-to-one pixel to texel ratio). Aubel et al. [ABT00] and Schaufler et al. [Sch95] used dynamically generated approaches for impostor representation. In this case, no storage space is devoted to any impostor image that is not actively in use and results in less memory being consumed than pre-generated impostors, though the impostors do need to be updated at runtime.

Previous work on perception of human motion in the context of computer graphics has focused on the effect of animation quality on user perception. Particularly relevant to us, Hodgins et al. [HOT98] performed a series of perceptual experiments, the results of which indicate that a viewer's perception of motion characteristics is affected by the geometric model used for rendering. Furthermore Oesker et al. [OHJ00] demonstrated that there is psychological evidence that observers were affected by the level of detail of a character's animation in making judgments about the skill of virtual football players.

3. Experiment Introduction and Psychometric Techniques

An extensive user study looking at all aspects of impostors would be desirable, however this is a first step at establishing some useful thresholds for designers of impostor systems. Based on observations of our own impostor rendering system [DHOO05], we chose to focus on areas that we felt needed attention.

Firstly, we wished to know when users could notice the difference between an impostor and a geometric model to decide at what distance it would be appropriate to switch representations. In our initial system, a one-to-one ratio was used to control when an impostor changed to a geometric model. While this produced acceptable results, we felt that this was another issue that could be addressed with psychophysical experiments to evaluate the correctness of this choice. In an urban simulation, such as our system, large numbers of building impostors are required and the updating of these based on purely geometric criteria, such as Schaufler [Sch95], can be computationally expensive. We wished to evaluate user sensitivity to varying the update rate of such impostors, to determine whether this could improve performance. Finally, while impostors can replicate the appearance of geometric models at certain distances, we wished to evaluate whether or not they are also good at replicating the motion of the original model.

Two groups of experiments were carried out. The first group (consisting of three experiments) aimed to establish thresholds for the use of impostor representations for static building models and for animated virtual human models. In the second group there was one experiment that aimed to evaluate the effect of model representation on the perception of human motion. In previous work [ODGK03] a staircase procedure was effective at homing in on user thresholds for limited numbers of participants. Due to the number of experimental trials required, we were limited in the number of participants that could take part in each experiment. While it is difficult to select a representative group of participants, we felt that approximately ten was a reasonable number and range of participants, and assessed their results to ensure consistency. The participants were drawn from those both familiar and unfamiliar with computer graphics.

3.1. Psychophysics

Psychophysics is a mathematical approach relating the internal psychic and the external physical world on the basis of experimental data [Tre95]. The basic procedure for performing psychophysical experiments is to present participants with stimuli and record their responses. Typically par-

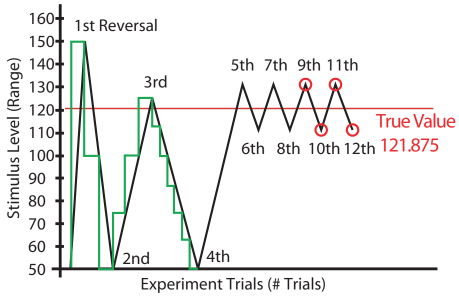

Figure 1: *Example of a staircase procedure in operation. Reversal Values: 150, 50, 125, 50, 131.25, 112.5, 131.25, 112.5, 131.25, 112.5, 131.25, 112.5*

ticipants are asked to compare stimuli to a reference stimulus and report if they thought that they were the "same" or "different", or to choose between two stimuli based on some criteria. The results are recorded and typically plotted as the percentage of responses that are correct for a particular stimulus, such as distance.

Experimental data can be gathered efficiently using a staircase procedure [Cor62, Lev71]. This is a method whereby a stimulus is alternately increased and decreased until it passes a participant's detection level. An adaptive staircase reduces the step-size used over the course of the experiment to home in on the true value for the staircase. Typically, dual staircases starting from opposite ends of the experimental range are used and a result is accepted if both staircases converge to approximately the same answer. In our experiments, we make use of several randomly interleaved staircases during each experiment run to prevent the participant guessing the next correct response. Other examples where staircases have been used to evaluate graphical systems are Mania et al. [MAEH04] and O'Sullivan et al. [ODGK03].

A typical staircase procedure uses equal step-sizes for increasing and decreasing the stimulus and is called a One-Up, One-Down (1U-1D) Staircase. In the case where the participants' responses will be random beyond a certain level (such as when choosing between items that are very far away), a Weighted Up-Down staircase may be used. This method uses different step-sizes for upward and downward steps so the staircase will approach the detection point even on random responses when the participant is guessing. The staircase will still react appropriately to correct responses when the participant is certain about their choice. A typical Weighted Up-Down ratio would be 3:1.

An example of typical staircase operation is shown in Figure 1. The user is initially shown a stimulus, e.g. an impostor and a geometric model side by side, at a minimum range (here 50) and is asked to choose which they prefer. The staircase then repeatedly moves the stimulus farther away (initially in steps of 100 units) until the user changes their response (the 1st Reversal - in this case at 150 units away). The range is then decreased at a smaller step-size (now 50 units) until the user changes their selection response again (2nd Reversal - at 50 units away). The staircase switches direction again and again, halving the step-size each time. After a set number of reversals (here four) the step-size is fixed to avoid too many repetitions of virtually the same stimulus which would occur if the step-size became too small. In this case the user sees the stimulus moved towards and away in steps of 6.25 units. Finally, the staircase is stopped after eight further reversals and the average of the final four reversals is taken to be the result of the staircase. This is compared with the final result of a second staircase that runs the same experiment from the opposite direction (far to near initially). The staircases are said to have converged if these results are close in value, and the experimental data is considered good.

If the two staircases for a particular participant have converged then their responses may be combined together. During each experiment run the number of correct responses at each point (such as successfully choosing the impostor) can be recorded for a participant along with the number of times a particular stimulus was seen, as the user sees the same stimulus several times. This is then used to calculate the *percentage correct* for the observed stimuli and, when plotted, produces a graph like Figure 2 from which thresholds may be estimated. In such a curve a value of 100% means the user always chose the *correct* or *acceptable* response while a value of 0% similarly indicates a constant *incorrect* choice or *rejection* by the user. Should the percentage correct value remain at 50% then the user's choices are a random spread of correct and incorrect (e.g. in the previous example, being unable to distinguish the impostor). Multiple users' curves may be combined to produce final averaged results.

$$P(x) = 1 - \left(\gamma \times \left(\frac{1}{1+(\frac{x}{\alpha})^{-\beta}}\right)\right) \qquad (1)$$

A Psychometric Function is a statistical curve fitted to the cumulative responses of an experiment. For evenly distributed stimulus data (where the percentage of correct results varies between 0% and 100%), a simple ogive inverse normal distribution may be fitted and results estimated from the graph. For experimental data where the probability of detection of a stimulus falls below a chance detection level (e.g. when discriminating between two items, that eventually look the same and selection is random), a logistical distribution such as that described in Treutwein [Tre95] may be fitted instead. We use a slightly modified (inverted) version of their logistical psychometric function as shown in Equation 1, where α is the stimulus at the halfway point, β is the steepness of the curve, and γ is the probability of being correct by chance (50% when there are two random choices).

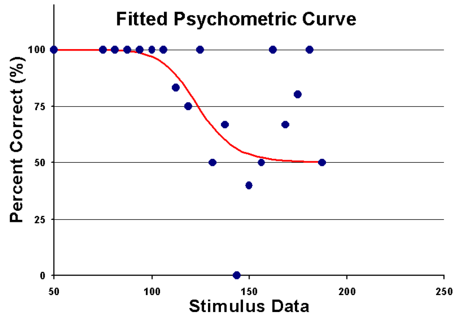

Figure 2: *Example of a fitted Psychometric Function*

Figure 2 shows an example of such a fitted curve for a set of real world user responses.

Two common values that can be estimated from the curve are the Point of Subjective Equality (PSE) and the Just Noticeable Difference (JND). The PSE is the stimulus that the observer indicates to be the same as the standard. This is typically estimated as the 50% detection level on the best fit psychometric function (or the 75% level where the curve covers a range of 100% to 50%). The JND is the smallest difference between two stimuli that enables consistent discrimination between them. The JND is typically the amount of additional stimulus needed to increase a participant's detection rate from 50% to 75% on the fitted psychometric function. The steeper the function, the smaller the JND.

4. 1st Group: Evaluation of Static and Dynamic Impostors

4.1. Model Representations Used in Experiments

For our experiments, we made use of two different types of impostor representation (pre-generated and dynamic), along with a geometric representation. For the building impostors, we use a dynamically generated representation [Sch95]. Typically, building impostors remain unchanged for long periods of time, thus the impostor can be generated at runtime, and then be reused over many frames of rendering. This method uses little memory, at the expense of occasional processing requirements to update the impostor. The virtual human's impostor representation uses a pre-generated approach as detailed in Dobbyn et al [DHOO05] and Tecchia et al [TLC02]. This approach consumes memory, but has very little runtime processing requirements. The virtual human's geometric representation is a skinned skeleton mesh, animated with keyframe animation.

4.2. Experimental Apparatus and Setup

The equipment used was a high end commodity PC with an NVidia GeForce graphics accelerator card. The OpenGL test application was displayed to the participants on a 19-inch Monitor at a resolution of 800x600 pixels with a screen refresh-rate of 85 Hertz. User input for the experiments was provided by a USB gamepad featuring two trigger buttons allowing the participant to make their selection.

The experiment environment consisted of a black grid for the ground plane with a white background. The 3D world was configured for a standard 45° field of view of the environment. Experimental participants were positioned approximately 28"-30" from the screen at zero elevation and so the full display subtended a visual angle of approximately 26°. All models were displayed in grey-scale, as in these experiments we wished to determine people's ability to discriminate detail. Colour would complicate this issue by introducing further confounding factors, so we will leave this aspect for future examination.

Four building models (*Museum*, *Library*, *Old Building* and *Bank*) were used in the building set. These models were chosen to represent varying types, with ascending complexity as shown in Figure 3.

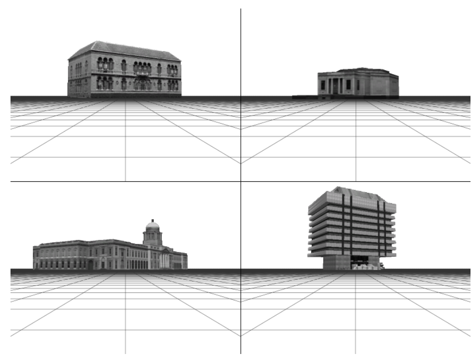

Figure 3: *The four buildings used in the static impostor experiments: Top Left - Museum; Top Right - Library; Bottom Left - Old Building; Bottom Right - Bank*

The same virtual human model was used throughout the human set. Both representations were animated with the same one second walk-cycle consisting of one frame of animation every 100 milliseconds. Models were lit by a directional light source pointing towards them, positioned directly behind the camera.

4.3. Set 1: Discrimination of Impostors

We aimed to establish the point at which an impostor becomes noticeable to a user. We tested this by finding the PSE for the discrimination of a geometric model displayed beside an impostor representation. These experiments were carried out separately for both the building models and the dynamic virtual human model. The goal in establishing such a threshold is to provide a guide for developers in their decision as to when they should switch level of detail representations.

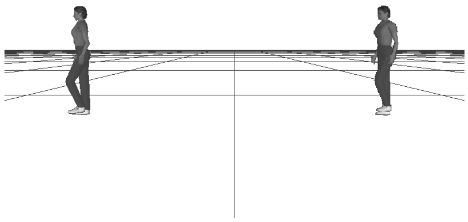

Figure 4: *Discrimination Between Virtual Human Geometry & Impostor Experiment*

The participants were shown the geometric model and impostor side by side at different ranges from the viewer (as shown in Figure 4). This involved employing a two alternative forced choice (2AFC) paradigm whereby the subject was asked to choose which representation "Looked Better". We considered the geometric model to be the "correct" response. Two staircases were used for each model, one ascending and one descending. Each staircase ran for twelve reversals, halving the step-size after each of the first four reversals, and the staircases for different models were randomly interleaved. A 3:1 Weighted Up-Down procedure was employed, so each time the participant guessed "incorrectly" (i.e. picked the impostor model), the stimulus was moved closer by three times the step-size, otherwise a "correct" response caused the stimulus to be moved away by just the step-size. This mechanism gave us an estimate of the 75% detection threshold, which represents the PSE for these scenarios (i.e. the point where the impostor just becomes noticeable).

The 75% threshold is a standard psychometric data point when the percentage of correct responses would vary between 50% and 100%. It was hypothesised that beyond the point of one-to-one pixel equivalence of the impostor image and the projected geometric model, the participants would be unable to correctly discriminate between the representations. The one-to-one pixel to texel ratio equivalent distance for each of the models is listed in Table 1 along with the two-to-one distance for comparison.

A psychometric curve as shown in Equation 1 was fitted to the experimental data using $\gamma = 0.5$. This produces a curve ranging from 100% to 50%. The 75% level on the graph is a good measure of the Point of Subjective Equality (PSE) for the stimulus. To find the Just Noticeable Difference (JND) for the threshold, we take the 87.5% point and subtract the 75% point. The PSE is a generally accepted psychometric measure, though once the logistical function has been computed, other data points (for example, when people cannot discriminate 90% of the time) can be simply extrapolated.

Model	1-1 Ratio Range	2-1 Ratio Range
Museum	132.4 units/meters	264.8 units/meters
Library	100.5 units/meters	201.0 units/meters
Old Building	315.2 units/meters	630.4 units/meters
Bank	131.5 units/meters	263.0 units/meters
Virtual Human	11.0 units/meters	22.0 units/meters

Table 1: *One-to-one pixel to texel ratio distances for the four test buildings (256x256 impostor) and the virtual human with a 45° Field of View at 800x600 Resolution*

4.3.1. Building Impostor Discrimination

For the building models, the range of the experiment stimulus was set between 50 and 500 units in the virtual world using preliminary observations, with an initial step-size of 250 units. World units were equivalent to real world meters in terms of the scale of the models. The final step-size (after halving four times) was 15.625 units. The *Old Building* model was much larger than the other buildings and was tested on a range of 250 to 700 units, with the same initial step-size of 250 units.

4.3.2. Building Discrimination Results

There were 10 participants for the static building experiment (5M-5F, aged 22-26), 8 of whose experimental data converged properly. Experimental participants were drawn from staff and students of the authors' institution both familiar and unfamiliar with graphics. All participants had normal or corrected to normal vision.

The PSE values, as shown in Figure 5a, demonstrate that, for all buildings, the observed PSE values exceeded the expected one-to-one ratio threshold, usually by around 20%, but with better discrimination for the detailed *Bank* building. This additional sensitivity is believed to be due to the geometry aliasing effects on this model. The JND values (Figure 5b) represent a change in distance of approximately 20 units in the virtual world. Single factor Analysis of Variance (ANOVA) was performed to compare the mean PSE values of the four buildings, showing these differences were statistically significant ($F_{3,28} = 21.83, p \approx 0$). Thus, for the most part impostors may successfully be used at distances where their pixel to texel ratio is greater than 1.25, but care must be taken with models of great complexity to ensure they are correctly represented. Using a geometric representation for a greater distance is advised in such cases.

4.3.3. Human Impostor Discrimination

For the virtual human models, the range of the experiment stimulus was set between 5 and 31 units, with an initial step-size of 2.5 units. The final size (after halving) was 0.15625 units. Both representations were rotated 5.625° every 100

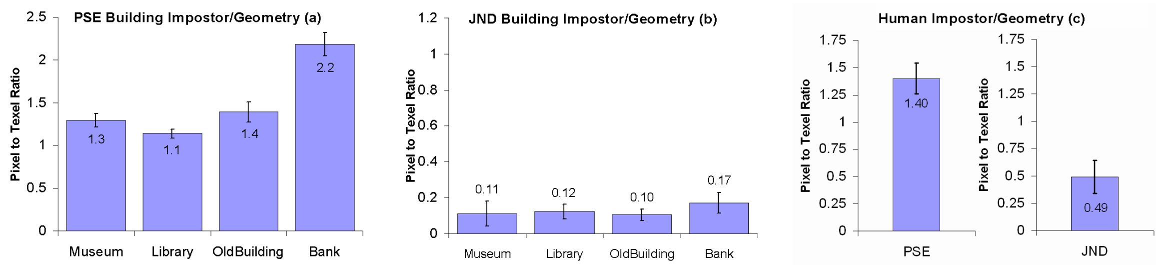

Figure 5: *Results of the Discrimination Experiments (showing the Points of Subjective Equality (PSE) and the Just Noticeable Differences (JND) in terms of Pixel to Texel Ratio) for Buildings (a,b) and Humans (c)*

milliseconds in a random direction so that the subject was not comparing them based on a single viewpoint with respect to the camera to eliminate directional bias. The virtual humans were separated by a fixed number of screen pixels to keep the distance between the representations constant.

4.3.4. Human Discrimination Results

For the virtual human experiments, there were eleven experimental participants (9M-2F, ages 22-39), nine of whose experimental data converged properly. All participants had normal or corrected to normal vision. The mean PSE calculated (shown in Figure 5c), was greater than the hypothesised value of one-to-one by approximately 40%. However, the mean JND is quite large indicating that the participants were not sensitive to small changes to the pixel to texel ratio at which the impostor was being displayed. Due to impostor artefacts, the participants were able to perceive the difference between the representations when displayed side-by-side. These artefacts may be caused either by aliasing, loss of depth information, or using a fixed number of pre-generated viewpoint images. Since the participants were capable of differentiating between representations through direct comparison, both representations should not be displayed at the same distance in our system. Since only one virtual human model was used for this set there were no means to compare hence an ANOVA was not performed.

4.4. Set 2: Sensitivity to Level of Detail Changes

Next, we aimed to establish the point at which a user might notice a transition from an impostor to a geometric representation. In this case, the percentage of correct responses varies from 100% to 0% so the PSE was estimated by the 50% detection threshold for both the static building models and the dynamic virtual human model separately. While having thresholds for the believability of impostor representations is useful, so called ‘popping’ artifacts often manifest during the transition from impostor to geometry. Thus, obtaining threshold information about the transitions provides developers with a guide as to when such transitions will be acceptable when deploying multiple levels of detail.

For each trial, a single model was displayed, starting at a specific distance from the viewer, then moving at a constant speed towards the camera, and finally stopping at a specific distance. At some point during the interval the model switched from an initial impostor representation to a geometric model. This switching range was varied according to two staircases (one ascending, one descending) for each model. The participants were asked if they noticed a “Definite Change” in the model, giving yes/no responses using the gamepad to signal their 2AFC response. Each staircase ran for 12 reversals where the step-size was halved for the first four reversals. The staircases were 1U-1D, allowing for a 50% detection threshold to be determined. The same logistical curve as in the first experiment was fitted to each user’s responses (with the γ value set to 1.0 for a 0% to 100% range).

4.4.1. Building Transition Detection

The same four buildings were used as in the first experiment. Isolated buildings were displayed facing the user, starting at a range of 600 units, and then moved at a speed of 200 units/sec toward the screen. The stopping point was a range of 50 units from the screen. After the first four reversals, the final step-size was 12.5 units.

4.4.2. Building Transition Detection Results

Nine participants took part in the static building impostor ‘popping’ experiment (6M-3F, aged 22-26). For the two simpler building classes (the *Museum* and *Library*), few participants (3-4) showed good convergence. This is believed to indicate that for those simpler models, no definite detection of the transition occurred, with various screen aliasing and rendering artifacts dominating instead. The other models showed good convergence for all participants.

The JND values for the building ‘popping’ detection experiment were established by subtracting the PSE 50% value from the 75% point on the fitted psychometric curve. The results, as shown in Figure 6(a,b), indicate pop detection around the one-to-one pixel to texel ratio point but from a

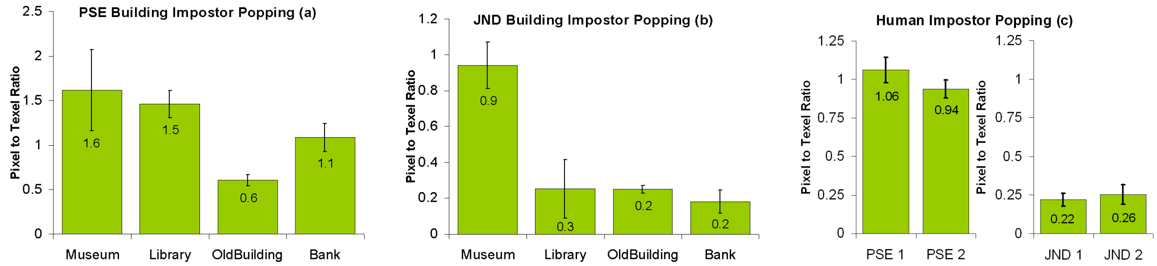

Figure 6: *Results of the Popping Detection Experiments (showing the Points of Subjective Equality (PSE) and the Just Noticeable Differences (JND) in terms of Pixel to Texel Ratio) for Buildings (a,b) and Humans (c) facing viewer(1) and spinning(2)*

ratio of 0.6 to 1.6. With only three and five successfully converged staircases for the *Museum* and *Library* buildings respectively, we can be much less certain of the results for these buildings, but overall it does seem to indicate that successful detection of an impostor to geometry transition is more difficult for these simpler buildings. Single factor ANOVA was performed to compare the mean PSE per building ($F_{3,20} = 6.80, p < 0.002$). Based on the converging results, to be certain a user will not detect a transition (with building type impostors), it is suggested that the switch should occur where the pixel to texel ratio exceeds 1.5. This should ensure that a majority of people will not detect the transition, though this may be further improved by utilising blending.

4.4.3. Human Transition Detection

The same virtual human model was also used as in the first experiment. Two experiments were carried out with the model either facing the user or spinning on the spot at a rate of 5.625° every 100 milliseconds in a randomised direction. For both experiments, the model started at a range of 36 units, and then moved at a speed of 6 units/sec toward the screen. The stopping point was a range of 1 unit from the screen. After the first four reversals, the final step-size was 0.3125 units. The virtual human switched from its impostor to its geometric representation at a switching distance ranging from 6 to 31 units.

The results of pilot experiments were used for setting the speed of the camera. It was found that when the virtual human approached the camera too quickly, the resulting rate of change in the texture detail of the geometric representation caused the participants to perceive a switch where there was none. While the effect of popping artifacts may be reduced by blending such as in Ebbesmeyer [Ebb98], we aimed to establish baseline thresholds were this would not be necessary. For urban simulations (which generally are constrained to the ground plane), transitions typically occur in the distance where the change in depth information is small due to perspective, and for virtual humans the overall change of depth information is similarly small. A further investigation of the effect of blending on transition detection is desirable.

4.4.4. Human Transition Detection Results

For these experiments, participants had normal or corrected to normal vision, and were both familiar and unfamiliar with graphics. For the first case, where the virtual human faced the viewer, there were seventeen experimental participants (13M-4F, ages 12-39), 10 of whose experimental data converged properly. The mean PSE calculated (shown as PSE1 in Figure 6c), was approximately the predicted one-to-one value with a small mean JND (shown as JND1), indicating that the participants were quite sensitive to subtle changes in the pixel to texel ratio at which the popping occurred. (This is lower than in Set 1, probably because the two representations are never compared side by side). For the second case, where the virtual human spun, there were 10 experimental participants (8M-2F, ages 12-39), nine of whose experimental data converged properly. The mean PSE calculated (shown as PSE2), was less than for PSE1, suggesting the spinning was a distracting factor. However, the differences were not significant for the PSE ($F_{1,17} = 1.46, p > 0.3$) or the JND values ($F_{1,17} = 0.22, p > 0.7$). The large number of diverging results in the first case, however, suggests that the participants noticed other artefacts, which were masked in the second case when the virtual human was spinning.

4.5. Set 3: Building Impostor Update-Rate Sensitivity

The third experiment for building impostors aimed to establish a 50% threshold for the acceptability of variations in rotational update rate for the impostor representation. This is important if one considers the typical use of a building impostor to represent a structure in the distance. As the user moves through a virtual world, their angle with respect to the impostor will change, necessitating an update of the impostor representation. Finding the threshold of acceptability for updates should allow impostors to be refreshed only as frequently as necessary. Schaufler [Sch95] first addressed the topic of dynamic impostors and geometric metrics for their updating. Perceptual metrics establishing lower update rates for such impostors would increase performance for urban simulations allowing resources to be devoted elsewhere.

Each building impostor was displayed on the screen at a fixed distance and no geometric representations were used. The models appeared on screen rotating in a random direction to avoid any directional bias and at a constant speed of 90° per second.

As the building represented by the impostor rotated, the impostor (which always faced the user) had to be updated. The frequency of these updates was controlled using an ascending and descending staircase as before. The range of update frequency was between 0° and 20° and the experiment varied the update amount between these two limits according to the staircase. Again the staircases followed a 1U-1D pattern, providing for a 50% threshold, and were randomly interleaved to prevent user training. The initial step-size was 10°, reducing to 0.625° by halving at the first four reversals. As before, eight further reversals were then counted. The participants were asked to specify whether they perceived the motion as "Smooth" or "Jerky", and gave their responses using the gamepad.

4.5.1. Building Impostor Update-Rate Sensitivity Results

With update frequency running the range from 0° to 20°, a 50% threshold can be established by fitting the logistical function to each user's responses (with $\gamma = 1.0$). The JND was again found by subtracting the 50% point from the 75% point. There were 12 participants in this experiment (6M-6F, aged 22-39), and all their results showed good convergence. As in the previous experiments, the PSE and JND values listed in Figure 7 are the mean of the converged participants' responses after curve fitting.

The resulting PSE values show that most people found an update rate of around 8° to be smooth enough for use, and this seems to be irrespective of the building used. Single factor ANOVA was performed on the PSE ($F_{3,43} = 0.219, p > 0.88$) and JND ($F_{3,43} = 0.158, p > 0.92$) values to determine if the type of building model played a role in the results and it was found that it did not. This indicates that these results are independent of the building model used and thus these thresholds should apply to any building of similar type. While it is expected that these results apply to objects such as buildings (which are mostly cubical or slightly oblong in shape), their applicability to objects which are greatly distorted in one direction (such as a train) cannot be predicted. Future psychometric experiments are suggested to shed light on this matter.

5. 2nd Group: Evaluation of the Effect of Model Representation on the Perception of Human Motion

A final set of experiments was conducted to evaluate how accurately the impostor replicates the motion of the virtual human that it represents. In the target application area, impostors and virtual humans are present together in scenes, so it is important that the motion of the impostors is a good representation of the motion of the virtual humans. As previously mentioned, Hodgins et al. [HOT98] showed that a difference in model complexity affected user perception, with the participants in their study being more sensitive to changes in motion when viewing a complex model than a simple model. Their results suggest that sensitivity to motion changes could be a good metric for evaluating the visual fidelity of an animated human model. Our experiment was similar to theirs, except that we used impostors rather than stick figures as the simple model. Also, a different psychophysical technique was used.

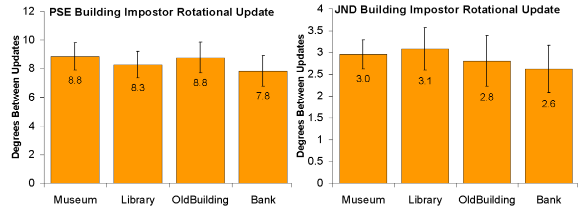

Figure 7: *Results of the Rotational Update Rate Experiments (showing the PSE and JND in terms of degrees between updates)*

The experiment was a between-groups experiment, where one group viewed the polygonal model and the other viewed the impostor. We favored a between-groups experiment because we felt that, if the participant viewed both the impostor and the polygonal model, they may have formed a strategy as to what sort of change in motion was being evaluated. For example, the artefacts on the impostor images may cause the participant to focus on the artefacts instead of the overall motion if they had already seen the geometric model which shows less artefacts. We separately tested the ability of people to notice subtle changes in motion on the impostor to their ability on the polygonal model. We then compared the performances of the two sets of participants to see how close the results of the impostor representation were to those of the polygonal model.

5.1. Experimental Apparatus and Setup

Sixteen participants (13M 3F, aged 17-25) took part in the experiment. All participants were naïve as to the purpose of the experiment and had normal or corrected to normal vision. Participants viewed the motion sequences on a 21 inch flat screen C.R.T. monitor. A greyscale checkerboard floor plane was used so that the movement of the model could be seen clearly. Lighting and rendering conditions were constant throughout the experiment. We evaluated three different types of motion variation: torso rotation, dynamic arm motion and dynamic leg motion.

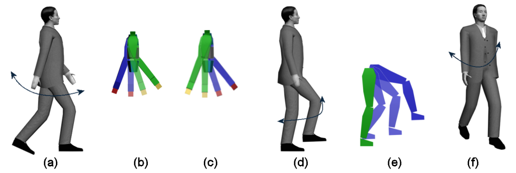

Figure 8: *(a) Rotation of arm at shoulder about horizontal axis relative to torso. (b) Skeleton showing both arms (L: blue, R: green) at step 0 in one keyframe of kAl, overlayed by step 10. (c) Skeleton showing arm at step 0 in one keyframe of kAr overlayed by step 10. (d) Translation of leg along horizontal and vertical axes. (e) Skeleton showing the alteration of the left leg at step 0, step 5 and step 10 (f) Rotation of torso along horizontal axis relative to pelvis.*

Assessing the arm motion variation involved comparing a reference motion to a set of motions which altered the distance of the arm from the body at certain keyframes. The reference motion was a single cycle of a keyframed walk repeated a number of times so that 3 seconds of motion were recorded. Keyframes *kAl* were the keyframes in the original walk motion sequence where the left arm was furthest away from the body in the positive direction (Figure 8b). The upper left arm joint in *kAl* was altered to create the modified biped motion sequences, and the right arm was altered by the same amount in the reverse direction. Ten discrete motion sequences were constructed, representing the ten different steps in the staircase analysis. Each step was created by iteratively rotating the upper arm joint at the shoulder along the horizontal axis at *kAl* by a fixed number of degrees. The poses of the skeleton at *kAl* were copied and the inverse of these poses were pasted onto the corresponding poses at keyframe *kAr* (the keyframe where the right arm is furthest away from the body in the positive direction - Figure 8c).

The keyframe motion sequences were then exported into an OpenGL rendering system and applied to our polygonal model which was a deformable mesh as before, but a different human figure was used. The original keyframe sequence was altered to ensure that it was cyclic, so that it and all the motions created by altering it could be repeated a number of times to create 10 3-second movies. The ten impostor sequences of the polygonal model were then rendered and recorded as movies (a lengthy process taking some hours).

A similar test was conducted to examine the performance of the participants in distinguishing larger and smaller leg motions for both representations. A further set of 20 motion sequence movies were created in a similar manner to the arm motions, except that the leg was altered by iterative translations along the longitudinal and vertical axes (Figure 8d,e).

Finally, the ability of the participants to distinguish alterations to the torso was tested. A final 20 movies were created by making kinematic alterations to the reference walk motion. In this instance, the alterations were made by iteratively rotating the lower spine of the skeleton by a fixed number of degrees around the longitudinal axis (Figure 8f).

5.2. Actual Experiment

The experiment consisted of 3 ascending staircases and 3 descending staircases randomly interleaved, i.e. an ascending and descending staircase for each of the motion variation types. Participants viewed pairs of motions, and were asked to specify whether they thought that the motions were the "same" or "different". Eight of the participants viewed the polygonal model, while the other 8 viewed the impostor. The first motion sequence was viewed for 3 seconds, after which the participant pressed a "view next" button on-screen, using the mouse. The next motion sequence was then viewed for 3 seconds and the participant had to choose whether they thought that the motions were the "same" or "different" and press the corresponding button on-screen.

The ascending staircases began with a comparison of the reference motion R to itself, and the descending staircases began with a comparison of R with the most exaggerated motion sequence (i.e. step 10 of the staircase). For the ascending staircases, a simple Up-Down staircase was employed so that for every correct response (i.e. when the stimuli were the same and the user chose "same", or when they were different and the user chose "different"), 2 steps were added to the current step, and for every incorrect response, 1 step was subtracted from the current step. We adapted the step-size after the first reversal so that only 1 step was added or subtracted. For the descending staircases, the same procedure was employed, but with the steps decreasing in the opposite direction. Once the refinement to the step-size was made, the procedure was continued until 8 reversals were recorded. Staircases were randomly interleaved, and participants were randomly shown either R or the motion sequence at the current step-size. This gave a 50% detection threshold which is used as the PSE.

5.3. Finding a Uniform Scale for the Steps

In order to compare the performance of the polygonal model to the impostor across the arm, leg and torso motion variations, a more uniform scale than 0-10 was needed for the steps. As the alteration to the motion was fixed at each step, a single scaling factor for each of the motion variation types was needed. The scaling factor we used was based on the amount of alteration that was made to each of the joints between 2 steps of a staircase. Lee et al's [LCR*] distance metric defines the difference between frames of animation by computing the changes in orientation of the joints. We used this distance metric, together with our own empirical joint weighting scheme as a scaling factor for the arm, leg and torso staircase step-sizes.

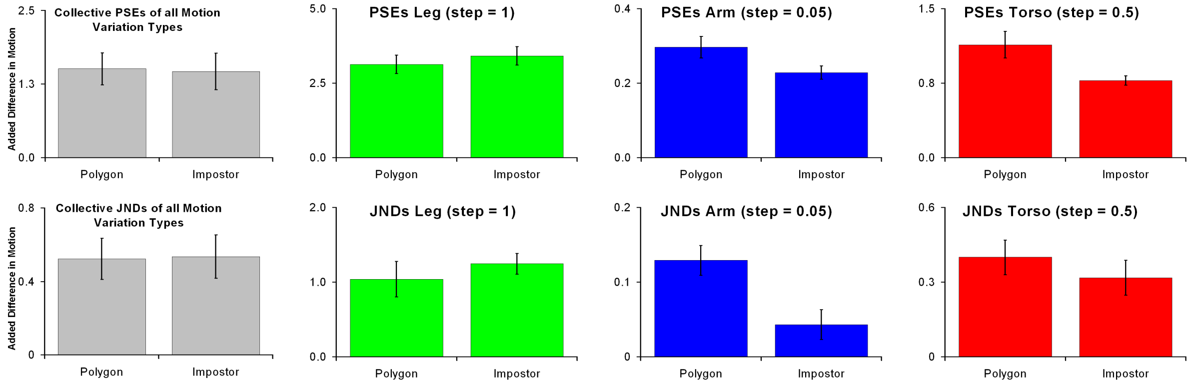

Figure 9: *Results of the Perception of Human Motion experiments (showing the Points of Subjective Equality (PSE) and the Just Noticeable Differences (JND)). The vertical axis shows differences in motion as estimated in section 5.3*

5.4. Perception of Human Motion Results

For each participant, the number of correct responses was recorded, along with the number of times that they viewed a pair of motions at each of the stimulus levels. The percentage of correct responses was then calculated and plotted against the stimulus level values. A psychometric curve was then fitted to the resulting dataset and a JND for each participant was found by calculating the difference between the 50% and the 75% detection levels on the psychometric curve. Average JND and PSE values are shown in the graphs (Figure 9).

We first compared the mean PSEs of the 8 participants who viewed the impostor with those of the 8 participants who viewed the geometric model for the leg motion variation, and the ANOVA showed no statistical significance as a result of model type ($F_{1,14} = 0.45, p > 0.5$). Furthermore, no significance was recorded for the mean JND values ($F_{1,14} = 0.46, p > 0.6$). This implied that, for the type of leg motion tested, there was no difference in the ability of participants to perceive motion variation in the impostor to that of the polygonal model.

For the arm motion variation, the ANOVA revealed a statistical significance between the JNDs ($F_{1,14} = 9.77, p < 0.01$). The average JND for the polygonal model was higher than that of the impostor, implying that for small arm motion variations, the participants were more sensitive to these changes when viewing the impostor than viewing the polygonal model. The ANOVAs for the Torso motion variation showed statistically significant differences between the mean PSE values ($F_{1,14} = 6.52, p < 0.025$). This indicates that people started noticing changes in the motion of the impostor before the polygonal model. The leg motion variation was an example of large motion alteration, having the greatest calculated change in motion, and consequently the greatest step-size of the experiment ($s = 1$). The torso motion alteration was a more subtle type of motion variation, ($s \approx 0.5$). Finally, the arm motion variation test looked at the most subtle level of motion variation, ($s \approx 0.05$). We conclude that in the case of the leg motions, the big difference in step-size meant that participants arrived at the same detection threshold for both the impostor and the polygonal model. When more subtle changes in motion were examined, like in the case of the arm motions, they came closer to the threshold at which there is a difference between the performance with the impostor to that with the polygonal model.

Finally, we performed a two-factor ANOVA with repeated measures on the full data set using all the recorded JND and PSE values. Model type and motion variation type were the two factors, with repetition (8 participants) on the motion variation type. Results showed that, when comparing the JND values of the 8 participants that viewed the impostor to those that viewed the polygonal model, for all motion variation types there was no statistical significance due to model type ($F_{2,42} = 0.08, p > 0.9$). Similarly, no significance was recorded when the PSE values were compared ($F_{2,42} = 0.02, p > 0.9$). Our results indicate that, using sensitivity to motion changes as a metric for evaluating human model representations, impostors are perceptually equivalent to high detail geometry for perception of human motion.

6. Conclusions and Future Work

The main objective of this paper was to evaluate the effectiveness of impostors for use in large-scale simulations involving complex scenes, such as cities inhabited by crowds. Perceptual issues with respect to such representations have been largely neglected to date, and we have redressed this situation in the current paper. To this end, we have carried out a set of perceptual experiments to find appropriate detection thresholds and to evaluate the factors affecting the perception of static and dynamic models using impostor and geometric representations. For static models, such as buildings, the pixel to texel ratio at which to switch was approximately 50% greater than the hypothesised one-to-one ratio. For animated virtual humans, the pixel to texel ratio at which to switch was equal to the hypothesised value of one-to-one previously used in our system. Finally, we evaluated the impostor's ability to replicate the motion of the associated geometric model and found they are equivalent in terms of perception of motion. As such, our results provide a baseline from which to evaluate all subsequent factors. In the application areas considered (urban simulation), buildings and humans are seldom displayed in isolation and colour is used extensively. Future work will investigate the effect of such factors on perception. We hope that this work will stimulate others to investigate these and other perceptual issues in computer graphics.

This work has been supported by the HEA of Ireland.

References

[ABT00] AUBEL A., BOULIC R., THALMANN D.: Real-time display of virtual humans: Levels of detail and impostors. *IEEE Transactions on Circuits and Systems for Video Technology 10*, 2 (2000), 207–217.

[Cor62] CORNSWEET T.: The staircase method in psychophysics. *American Journal of Psychology 75* (1962), 485–491.

[DDSD03] DÉCORET X., DURAND F., SILLION F. X., DORSEY J.: Billboard clouds for extreme model simplification. *ACM Trans. Graph. 22*, 3 (2003), 689–696.

[DHOO05] DOBBYN S., HAMILL J., O'CONOR K., O'SULLIVAN C.: Geopostors: A real-time geometry / impostor crowd rendering system. *In Proc. ACM Siggraph Symposium on Interactive 3D Graphics and Games* (April 2005), 95–102.

[DSSD99] DECORET X., SCHAUFLER G., SILLION F., DORSEY J.: Multi-layered impostors for accelerated rendering. *Computer Graphics Forum 18*, 3 (September 1999), 61–73.

[Ebb98] EBBESMEYER P.: Textured virtual walls - achieving interactive frame rates during walk-throughs of complex indoor environments. *VRAIS '98: Proceedings of the Virtual Reality Annual International Symposium* (1998), 220–228.

[HLTC03] HEIGEAS L., LUCIANI A., THOLLOT J., CASTAGNÉ N.: A physically based particle model of emergent crowd behaviors. *Graphicon* (2003).

[HOT98] HODGINS J. K., O'BRIEN J. F., TUMBLIN J.: Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics 4*, 4 (1998), 307–316.

[JWP05] JESCHKE S., WIMMER M., PURGATHOFER W.: Image-based representations for accelerated rendering of complex scenes. *In Eurographics 2005 STAR Report 1* (2005).

[LCR*] LEE J., CHAI J., REITSMA P. S. A., HODGINS J. K., POLLARD N. S.: Interactive control of avatars animated with human motion data. *ACM Trans. Graph 21*, 3.

[Lev71] LEVITT H.: Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America 49* (1971), 467–477.

[MAEH04] MANIA K., ADELSTEIN B., ELLIS S., HILL M.: Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. *In Proc. ACM Siggraph Symposium on Applied Perception in Graphics and Visualization* (2004), 39–47.

[MS95] MACIEL P., SHIRLEY P.: Visual navigation of large environments using textured cluster. *In Proc. ACM Siggraph Symposium on Interactive 3D Graphics and Games* (April 1995), 95–102.

[ODGK03] O'SULLIVAN C., DINGLIANA J., GIANG T., KAISER M. K.: Evaluating the visual fidelity of physically based animations. *ACM Trans. Graph. 22*, 3 (2003), 527–536.

[OHJ00] OESKER M., HECHT H., JUNG B.: Psychological evidence for unconscious processing of detail in real-time animation of multiple characters. *Journal of Visualization and Computer Animation 11* (2000), 105 – 112.

[Sch95] SCHAUFLER G.: Dynamically generated impostors. *GI Workshop on "Modeling - Virtual Worlds - Distributed Graphics"* (1995), 129–136.

[SDB97] SILLION F., DRETTAKIS G., BODELET B.: Efficient impostor manipulation for real time visualization of urban scenery. *Computer Graphics Forum 16*, 3 (September 1997), 207–218.

[TLC02] TECCHIA F., LOSCOS C., CHRYSANTHOU Y.: Visualizing crowds in real-time. *Computer Graphics Forum 21*, 4 (December 2002).

[Tre95] TREUTWEIN B.: Adaptive psychophysical procedures. *Vision Research 35*, 17 (1995), 2503–2522.

[WWS] WIMMER M., WONKA P., SILLION F.: Point-based impostors for real-time visualization. *Proceedings of the Eurographics Workshop on Rendering 2001*, 163–176.