Animating Humans on Handheld Devices for Interactive Gaming

A. Brosnan
Andrew.Brosnan@cs.tcd.ie

S. Dobbyn
Simon.Dobbyn@cs.tcd.ie

J. Hamill
John.Hamill@cs.tcd.ie

C. O'Sullivan
Carol.OSullivan@cs.tcd.ie

Trinity College Dublin
http://isg.cs.tcd.ie

Abstract

This paper examines the implementation and efficacy of impostor-based techniques on small-screen devices, and compares their effectiveness with that of the same techniques when used on desktop computers. Firstly, a real-time computer game has been implemented on a Personal Digital Assistant (PDA), featuring animated humans in a virtual city. Secondly—in order to examine the potential of impostors on future, more advanced small screen devices—psychophysical experiments were conducted. The results suggest that just as for PC-based systems, impostors are a useful technique for crowd and urban simulation on handheld devices.

Keywords: Impostors, Virtual Humans, PDA, Handheld devices, Games

1 Introduction

Small-screen, handheld devices such as PDAs, mobile phones, and portable gaming devices are becoming more powerful, with bigger and better displays. A system is described providing a navigable, 3D urban environment, populated by animated virtual humans. Image-based rendering involves replacing a high-rendering-cost geometric model with a simple 2-D image. A texture image is mapped onto a simple quad, and this "impostor" can be drawn to the screen much more quickly than the full 3-D model. The texture can be generated as a pre-processing step, storing visual detail in one image rather than in variously textured polygons, thereby minimizing the number of triangles to be rendered at run-time. Alternatively, the impostor-image can be rendered on-the-fly, but be re-used in successive frames until the scene changes such that the impostor image is no longer appropriate.

If impostor images are generated in advance, the decisions to be made at run-time are (a) when to use impostors, and (b) which impostor image to use. For (a), if the object in question is close to the camera, then the object must be rendered in high detail, and geometry is used. Developers will find it useful to know when impostors can be used without perceptible loss of visual fidelity. For (b)—which texture to use—the appropriate impostor image is chosen based on the relative positions and orientations of the object and the camera. Developers want to know how often the choice of impostor image must be updated, as the angle between camera and object changes. In practice viewers do not complain once the misalignment is beneath a certain threshold. Finally, those designing systems that use both impostors and geometric models in tandem will want to know whether there is a noticeable 'pop' on the transition from one to the other.

2 Background

2.1 Impostors

Tecchia et al. [1] describe a method of generating impostors of a human model from points on a hemisphere surrounding the model. From each of a discrete set of angles around the model, for each of a discrete set of elevations, and for each frame of walking animation, they render an impostor image. The appropriate image is chosen at run-time, and by this method they can render large crowds in real-time. The crowds populate a city, as in Dobbyn et al. [2], who also describe virtual humans in an urban simulation. The humans are represented by impostors when at a distance from the camera, but up close they switch to geometric models. This leverages the speed-up afforded by impostors when it is possible to do so without compromising the visual experience. We adopt a similar approach to Tecchia et al. and Dobbyn et al. but on a PDA.

2.2 Handheld Devices

Boier-Martin [3] describes a system of adaptive graphics, suitable for devices of varying types. Graphical data is stored in multiple levels of detail, right down to the level of Ascii art, and the appropriate representation is transmitted to the particular device accessing the data. Pham et al. [4] discuss the use of handheld devices for multimedia applications. Vainio et al. [5] and Rakkolainen et al. [6] combine a conventional map and 3D city model on a mobile (laptop) platform, and find 3D visualization useful in real-world navigation.

2.3 Perceptual Evaluation

In a related study [7], Hamill et al. conducted psychometric experiments to test the perceived quality of impostors. The main objective was to establish the circumstances under which IBR looks just as good as expensively-rendered geometry. Participants were shown both a ‘fake’ impostor and a ‘real’ geometric representation of a human or building, side by side. By analyzing the responses, they determined the distance beyond which a typical viewer cannot distinguish the full geometric model from the impostor.

Further experiments were conducted to detect the distance inside which the viewer notices a popping artifact, when an impostor-based human or building, moving towards the camera, switches to a geometric model. In another experiment, an impostor-based building was displayed, rotating continuously in one place. The update frequency of the impostor image was varied. This established a threshold for the use of ‘stale’ impostors (*i.e.,* those designed for a previous camera-object viewing angle) without viewer dissatisfaction.

Hamill et al. found, as expected, that impostors can be used in some circumstances without perceived degradation; thresholds were established for such situations. These results are useful to designers of PC-based graphical systems who wish to use impostors to increase frame-rate. In this paper, we conduct similar tests, but the device on which the impostor system is displayed and on which the experiments are conducted is a PDA with a 4”/10cm screen. The Hamill et al. study involved desktop PCs with 19-21”/48-53cm screens.

3 Interactive Game

Interactive games which take place in fully realized 3D worlds are now commonplace on PCs. This section describes a similar system on a mobile platform. A night-time simulation of a city center is populated with virtual drunkards. The player moves through this city zapping said drunks, whereupon they turn into besuited businessmen and go respectably on their way.

3.1 Human Animation on Handheld

Impostors are used to display crowd scenes at interactive frame-rates on a device with limited power. The system displays four human characters—drunk, businessman, male player and female player. Each uses a walking animation, with the drunk enacting an exaggerated staggering motion.

The impostor images are created in a preprocessing step. In previous work [1] and [2], Tecchia et al. and Dobbyn et al. describe a system to generate impostor images from an animated geometric model, for use in a PC-based Virtual City. Images are taken from all sides of the model, and for all frames of the walking animation. The handheld system simplifies this process by restricting the user to ground level. Thus, textures are needed only from one angle of elevation of the camera relative to the

Figure 1: Handheld game with drunks swarming towards the player

model. Therefore, for each model, impostor images are required from 16 discrete positions around the model, for each of 10 frames of animation. However as the human models are symmetrical left and right, the number of images can be halved—the images are mirrored when required—and each model requires 4MB of textures (the device has 64MB RAM).

A 3DS Max plug-in renders the impostor images required to display a full walking animation of a given model. Selection of the correct impostor image to use at run-time is achieved by the consideration of two factors. Firstly, the relative positions and orientations of the camera and the human. As in [1] the quad on which the impostor image is displayed is oriented towards the camera. The angle between the camera's look-vector and the human's direction-of-movement vector determines the choice of impostor image (*e.g.* front-on view of the human, side-on view of the human), along with one other factor, namely the current frame of the walking animation. A further simplification over the PC-based system is achieved by not dynamically lighting the impostors.

3.2 Urban Environment

The human models move through a virtual city. The player is free to move around the large cityscape, using the directional pad on the handheld device to navigate. Most of the buildings in this world are described by geometric models, with much of the architectural detail encoded by the textures rather than complex geometry. Textures are limited by the graphics API (PocketGL, described later) to a maximum of 128x128 pixels, and to the *.tga* image file format (uncompressed). Complex buildings are represented by impostors. As with the human models, textures are pre-rendered from a variety of angles around the full geometric model, and the appropriate one is chosen at run-time and displayed on a simple quad. Transparency in the impostor images is provided by PocketGL.

3.3 Handheld Device Issues

The virtual world is implemented with a coordinate space, light source and camera, as in a desktop OpenGL application. PocketGL 1.2 is used. It is a very limited subset of OpenGL for mobile devices without fast floating point arithmetic capability. Positions and orientations are described using integer values, *i.e.* the *glTranslate()* and *glRotate()* functions take integer parameters. It is desirable to achieve subtlety of movement, lest the humans jump from point to point rather than walking smoothly. Therefore the whole world is scaled up in the coordinate space. Apparently smooth movement of the humans can now be achieved in integer increments. PocketGL takes rotation angles in integers, but uses a system in which 360° equals 2048 custom rotation units, providing minimum rotation increments of 0.175°. This is found to be sufficiently detailed, both for smooth turning of the camera and for impostor image selection.

The game has also been implemented on a PC. As described by Rossi et al. [8], desktop and handheld machines communicate with each other via a server, using WiFi. Multiple players on these different platforms can compete against one another in real-time.

3.4 Results

Two basic criteria determine the merit of the handheld system—picture quality, and frame-rate. The picture quality can be seen in Figure 1. The buildings are recognizable as the real buildings on which they are modeled. The humans display a reasonably smooth walking motion. At close quarters the impostor humans do not look as good as geometric models would. However, they are more than acceptable in the context of the game, and look good at middle- and long-range. The frame-rate is maintained at around 20 fps with buildings and 10 humans on screen, and with player movement and game-play mechanics also in operation.

4 Experiments

Psychophysical experiments are conducted to gauge the viewers' abilities to distinguish impostor from geometry; to notice popping artifacts; and to detect delay in updating of impostor images, on the smaller screen of the PDA. Rather than evaluating perception on the current system, in real-time, the experiments simulate as-yet unavailable hardware, and use more advanced techniques (including dynamic lighting of the virtual human impostors). The experimental stimuli were videos, displayed on the handheld device. The videos were generated on a desktop PC, using a system which employs methods described in [1] and [2] for desktop crowd and urban simulation. The experiments were conducted to examine issues that will arise for developers taking advantage of newly available capabilities.

All experiments were conducted as follows. The display of the HP iPaq hx4700 palm-top computer was set to landscape mode, giving a screen resolution of 640x480 pixels. The experiment videos were displayed in full screen using Windows Media Player 9 for Pocket PC. Each 4–6 second video was rendered at 640x480 resolution and 25 frames per second. The viewer looked at a series of videos and gave a binary response—left or right, pop or no pop, smooth or jerky—according to the experiment. The verbal responses were noted by a research assistant.

The videos of each individual test scenario automatically played one after another, in random order, with a short blank screen in between each one. As these were 'two alternative forced choice' (2AFC) experiments, the viewer was forced to give one or the other response, even when unsure. The objective of the 'left or right' experiments was to discover the point at which a viewer is just guessing, and thus gives the 'correct' response 50% of the time. All participants have normal or corrected-to-normal vision, and come from a variety of backgrounds. The age-range in each case is 20 to 53 years.

4.1 Human Impostor vs. Geometry

We sought to establish thresholds for the discrimination of impostor-based and geometric models of animated, virtual humans. Nine participants (six male, three female) were shown videos of two virtual humans, one a geometric model, the other an impostor. Both humans perform a walking animation, on the spot, and turn slowly through 360°. The subject was asked to respond either "left" or "right" depending on which human "looks better". Having first seen examples where the humans were very close (easy to tell apart), and far away, the participant proceeded to view a series of these videos, with the humans between 4 and 19.5 virtual meters' distance from the camera. The increments between distances was 0.5m, and the videos were shown in random order.

The results suggest that viewers can easily tell the difference at close range, but not at long range. We had thought that the small screen size, and consequent proximity of the geometric and impostor models to each other, might allow the viewer to look at both representations at once, making it easier to contrast the images. However the results are in line with Hamill et al's results on a PC, see Figure 2. The results are compared and contrasted using the pixel-to-texel ratio rather than absolute distance, as this ratio takes into account the differing screen resolutions on PC (800x600) and PDA (640x480). The mean Point of Subjective Equality (PSE)—*i.e.* the point at which the pop is unnoticeable—across all participants is almost identical to that on PC. The mean Just Noticeable Difference (JND)—*i.e.* the change in distance required to go from noticeable to unnoticeable—is also similar.

4.2 Building Impostor vs. Geometry

A similar experiment is conducted using buildings instead of humans. In this instance the buildings are static. There were eleven participants, eight male and three female. Four different building models were used, representing a range of polygonal complexity. Videos of all four buildings were interleaved in one experiment set. As expected, impostor and geometry become harder to distinguish as distance increases. The handheld results are compared and contrasted with those for the desktop in Figure 3, again using pixel-to-texel ratio PSE. Note that pixel-to-texel ratio is inversely proportional to distance. The comparison suggests that for buildings, the impostor is indistinguishable from the geometric model at a closer distance on PDA than on PC. The building for which the ratios are

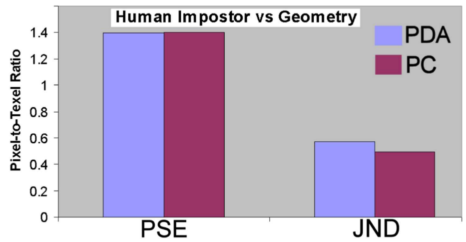

Figure 2: PDA vs. PC pixel-to-texel ratio

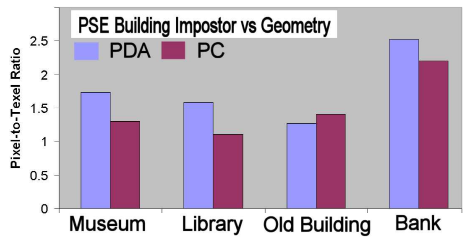

Figure 3: PDA vs. PC pixel-to-texel ratio

similar on desktop and handheld is a more complex model, suggesting that impostors are comparatively more useful for less complex models.

4.3 Building Rotational Update

The next experiment features an impostor of a single building rotating at a constant speed of 90° per second. The building rotates either clockwise or anti-clockwise around the y-axis. As the building rotates, the impostor image should be updated to compensate for the fact that the angle between it and the camera is changing. However the impostor is only updated when the discrepancy goes beyond a certain threshold value. This value was varied for each video. There were ten participants, seven male and three female. The participant was asked to respond either "smooth" or "jerky". He or she was first shown an example where the building was being continuously updated (smooth), and another where the impostor was only updated after a discrepancy of 20° had developed (jerky). The objective was to establish the threshold at which viewers accept this imperfect updating. The results for each building show a steady drop in acceptance as update angle increases. In this case the experiments differed from the desktop experiments in that the buildings were positioned further from the camera. At close range the videos files were too large, and thus did not display smoothly on the PDA. The greater distance allowed more effective video compression. For this reason, comparisons are left to future work. The PSE angles for the four different building models used are shown in Figure 4.

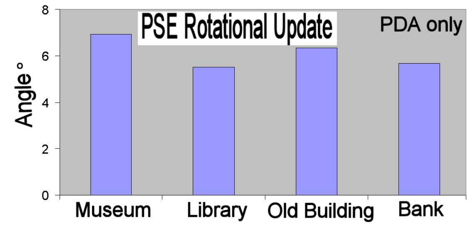

Figure 4: PDA Rotational Update Angle

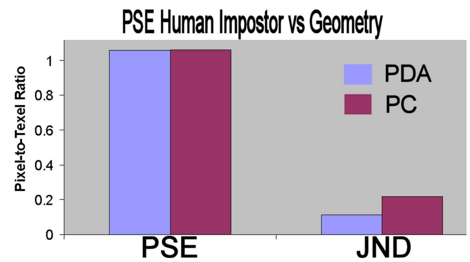

Figure 5: PDA vs PC pixel-to-texel ratio

4.4 Human Impostor Popping

The final experiment examines the effect of switching between impostor and geometry as the model moves closer to the camera. A single human model moves at a constant speed towards the camera, and switches at some point on its path. The distance at which it switches was varied. Nine participants, six male and three female, were asked whether they noticed an instantaneous "pop" or "switch" between a coarser and finer representation of the human. Two (male) participants could not see the pop even when it took place close to the camera, and their results have been excluded. As expected the switch was more noticeable when it happened up close. Once again, the mean Point of Subjective Equality across all participants is almost identical to that on PC. However the mean Just Noticeable Difference across all participants is smaller. The small JND value suggests that a clear threshold has been established for the point at which the popping becomes unnoticeable. See Figure 5 for PDA vs. PC results. In both cases the movement of the human towards the camera is a distracting factor.

4.5 Discussion

The length of the experiments is short enough to allow the participant to maintain concentration throughout.There is a one second gap (black screen) between each video to allow the viewer to give their response, if they have not already done so. Viewers did not object to the length of time given for, and between, each video. The decision was made to allow experiment subjects to hold the PDA as they wished, rather than at a prescribed distance from their eyes. In this way, each person used the device as they would normally. The videos themselves look clean and sharp, and run smoothly on the machine.

5 Conclusions and Future work

We conclude from our work that realistic crowd and urban simulations are possible on handheld devices, and that impostors are suitably employed to this end. The two principal elements of this paper—a fully implemented crowd and urban simulation on a mobile device, and experiments relating to more complex impostor-based and combined impostor/geometry rendering—will converge as more advanced machines become available. The usefulness of the more complex techniques has herein been examined. The advent of dedicated graphics/rendering hardware on handhelds will enhance the feasibility of 3D crowd simulations and provide scope for future research. A better graphics API will allow more detailed textures at reduced filesizes, and greater variation in the models. Their walking animations should also be varied by offsetting some models' animation cycles.

The results of the experiments are individually compared and contrasted with those obtained from earlier research where similar experiments were conducted on desktop PCs. Those experiments used an adaptive staircase methodology. The experiments described in this paper used pre-determined sets of pre-rendered videos. It would be useful to react to the responses given and alter the stimuli shown to the participant.

The comparisons of the desktop and handheld results suggest that impostors are equally (or, in the case of more complex models, more) useful on high-quality handheld devices. Future psychophysical work will involve evaluating impostors within a more complete environment. Instead of just one human or building being the focus of attention, subjects may be asked about whole scenes, or elements within whole scenes.

Acknowledgements

This research was funded by Enterprise Ireland and the Irish Higher Education Authority.

References

[1] Tecchia F., Loscos C., and Chrysanthou Y. Image-based crowd rendering. *IEEE Comput. Graph. Appl.*, 22(2):36–43, 2002.

[2] Dobbyn S., Hamill J., O'Conor K., and O'Sullivan C. Geopostors: a real-time geometry / impostor crowd rendering system. In *SI3D '05: Proc. of the 2005 symposium on Interactive 3D graphics and games*, pages 95–102.

[3] Boier-Martin I. M. Adaptive graphics. *IEEE Comput. Graph. Appl.*, 23(1):6–10, 2003.

[4] Pham B. and Wong O. Handheld devices for applications using dynamic multimedia data. In *GRAPHITE '04: Proc. of the 2nd intl. conf. on Computer graphics and interactive techniques in Australasia and SE Asia*, pages 123–130.

[5] Vainio T. and Kotala O. Developing 3d info systems for mobile users: some usability issues. In *NordiCHI '02: Proc. of the 2nd Nordic conf. on HCI*, pages 231–234.

[6] Rakkolainen I. and Vainio T. A 3d city info for mobile users. *Computers & Graphics*, 25(4):619–625, 2001.

[7] Hamill J., McDonnell R., Dobbyn S., and O'Sullivan C. Perceptual evaluation of impostor representations for virtual humans and buildings. *Computer Graphics Forum (Eurographics 2005), To Appear*.

[8] Rossi C., Cummins A., and O'Sullivan C. Distributed mobile multi-user urban simulation. In *GRAPH '03: Proc. of the SIGGRAPH 2003 conference on Sketches & applications*, pages 1–1.