Third Irish Workshop on Computer Graphics (2002)
Eurographics Irish Chapter

Crowd and Group Simulation with Levels of Detail for Geometry, Motion and Behaviour.

Carol O'Sullivan[†], Justine Cassell[§], Hannes Vilhjálmsson[§], Simon Dobbyn[†], Christopher Peters, William Leeson, Thanh Giang and John Dingliana.[†]

[†]Trinity College Dublin, [§]MIT Media Lab.

Abstract

Work on levels of detail for human simulation has occurred mainly on a geometrical level, either by reducing the numbers of polygons representing a virtual human, or replacing them with a two-dimensional imposter. Approaches that reduce the complexity of motions generated have also been proposed. In this paper, we describe ongoing development of a framework for Adaptive Level Of Detail for Human Animation (ALOHA), which incorporates levels of detail for not only geometry and motion, but also includes a complexity gradient for natural behaviour, both conversational and social.

1. Introduction

Crowd simulations are becoming increasingly important in the entertainment industry. In movies, they can be used to simulate the presence of real humans. For example, in the movie Titanic, extensive use was made of virtual people, whose movements were generated from a library of pre-captured motions. Such technology can be used in situations where it is dangerous for real people to perform the actions, such as falling over 50 feet off a ship, or to reduce the complexity and expense of handling large numbers of human extras. In animated movies, crowd simulation really comes into its own, such as in the stampede scene in Disney's *The Lion King* and the colonies of ants in Dreamwork's *Antz*.

Recent research into crowd simulation has to a large extent been inspired by the flocking work of Craig Reynolds [17]. He revolutionised the animation of flocks of animals, in his case birds (or "Boids"), by adapting some ideas from particle systems, where each individual bird is a particle. More recently, Brogan and Hodgins [2] simulated groups of creatures travelling in close proximity, whose movements are controlled by dynamical laws. A key element of this type of animation is collision avoidance. The Computer Graphics group in Lausanne has also done significant research into the area of crowd simulation [13, 14]. In their ViCrowd system, they can create a virtual crowd, where the individuals have variable levels of autonomy i.e. scripted, rule-based, or guided interactively by the user. They have demonstrated the emergent behaviour of a crowd attending, for example, a political demonstration or a football match. Nevertheless, the realistic simulation of crowds of humans remains a challenge, due to the ability of humans to detect even slightly unnatural human behaviour. In particular, realistic gesture and interactions between the individuals in the crowd are important cues for realism. Individuals in the crowd should look as if they are conversing or communicating with each other in both verbal and non-verbal ways (see Figure 1).

Figure 1: *Natural conversation in a group.*

The crowd effects in the movie Antz were impressive. A extensive library of motion-tracked movements was used to assign random movement to the individuals in the crowd. Such a method looks good from a distance, but up closer it is obvious that the characters are behaving in a cyclical manner, and not interacting naturally with each other. In the movie, the main characters were animated manually by human animators, which would just not be possible in any kind of real-time, interactive application. Such random behaviour may be retained as a lowest level of detail, and increasingly realistic (and hence computationally expensive) techniques may be used for more important characters, e.g. those closer to the viewer, or characters that are more significant to the plot. Some good work has been done in Lausanne and University College London on levels of detail for crowd simulations [1, 19, 18], but this was done mainly on a geometrical level: i.e. by reducing the numbers of polygons representing a virtual human, or replacing them with a two-dimensional imposter. In the Image Synthesis Group in Trinity College Dublin, a framework for Adaptive Level Of detail for Human Animation (ALOHA) has been developed [9], incorporating levels of detail for geometry, motion, and behaviour. In a collaborative project with the Gesture and Narrative Language Group at MIT Media Lab, we are also adding levels of detail for gesture and conversational behaviour to our framework, see Figure 2.

2. Geometry

The requirement in interactive systems for real-time frame rates means that a limited number of polygons can be displayed by the graphics engine in each frame of a simulation. Therefore, meshes with a high polygon count often have to be simplified in order to achieve acceptable display rates. The number of polygons, i.e. the Level Of Detail (LOD), of the model needs to be reduced. This can be achieved in two ways: a representation of the object at several levels of detail with a fixed polygon count can be generated, although switching between such levels of detail can cause a perceivable "pop" in an animation; or special meshes, called multi-resolution or progressive meshes, can be built that can be refined at run-time, i.e. parts of an object may be refined or simplified based on the current view, thus allowing a smooth transition from lower levels to higher levels of detail.

Constructing surfaces through subdivision has become popular within high end rendering packages over the last few years, and has been used by companies such as Pixar [6] to produce very effective results. These schemes solve many of the problems associated with other curved surface schemes such as NURBS and behave in a way similar to polygonal meshes. Character skins can be used almost directly with some subdivision schemes and others require modification in order to get a good representation of the original mesh. We have used subdivision surfaces as a means of improving the appearance of our virtual humans, along with some acceleration methods based on culling and preprocessing.

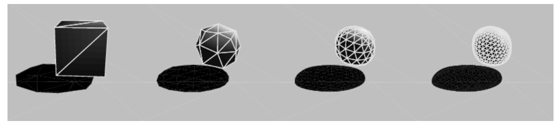

Figure 3: *Loop approximating subdivision surface for 4 iterations.*

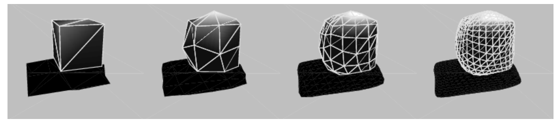

Figure 4: *Butterfly interpolating subdivision surface for 4 iterations.*

Subdivision schemes use a mask to define a set of vertices and corresponding weights, which are used to create new vertices or modify existing ones. Different masks are used for vertices on a boundary than the rest of the surface, because on a boundary edge some neighbouring vertices are absent. Other masks are also used to generate creases in a surface, making them capable of having sharp features. The masks are applied to each vertex in the mesh to produce a new mesh. After successive applications of the mask the mesh converges to a surface. The Loop[11] and Butterfly [8, 20] subdivision schemes are two techniques that can be used (see Figures 3 and 4). The masks can also be applied to the texture coordinates to generate the texture coordinates for each vertex. Further details may be found in Leeson 2002[10].

3. Motion

At the lowest level of detail in ALOHA, key-framed animations, exported from 3D Studio Max, are used to animate the Virtual Humans. When the characters are far from the viewer, these animations are chosen at random and changed at different intervals, in order to give an impression of varied activity to the crowd. When the viewer focuses on these characters, actions that are more meaningful are then chosen. We have implemented a real-time reaching and grasping system for autonomous virtual humans. The virtual human is endowed with a memory model based on the "stage theory" of memory from cognitive psychology, and has the ability to sense their environment using a virtual vision sensor. Virtual humans are not automatically aware of the exact characteristics of all objects in their environment or visual field, and must pay attention to them in order to be able to sample their attributes at a lower granularity. Given a command (e.g. "pick up the bottle"), the virtual human will become attentive towards the object, and will not only generate a goal-directed arm motion towards it, but will also remember the position and other perceived attributes of the object. Future

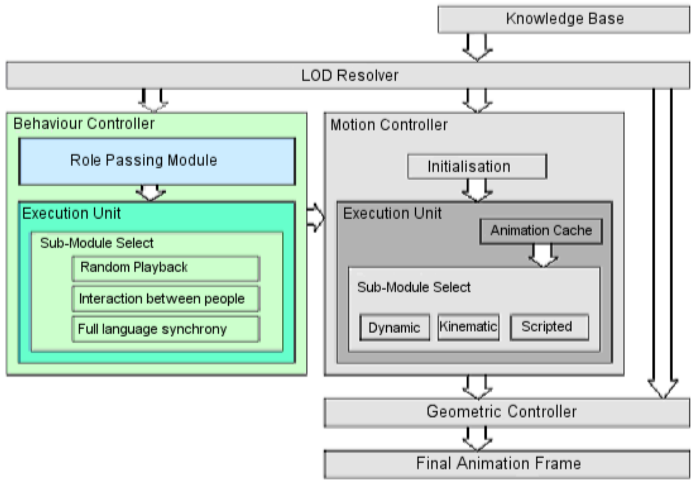

Figure 2: *Extended ALOHA framework with LODs for geometry, motion and behaviour.*

requests regarding the object can then be dealt with using the memorised state of the object; they will be able to make realistic reaching movements towards the remembered object location without having to look at it. The generation of these arm motions is based on results from neurophysiology. See Peters and O'Sullivan (2002)[16] for further details.

Another reason why human crowd simulations tend to lack realism is because of the flocking approach often adopted. Most crowd simulations implement collision avoidance, which often leads to crowd simulations where the crowd is too sparse, and therefore unconvincing. People in a crowd attending a concert or sporting event don't flock; They converse with each other, touch accidentally or unintentionally. Real people bump up against each other, or are squashed together, or even if it's not too tight, people who know each other will be walking along, chatting, holding hands, or occasionally touching. Also, people react differently if they bump into or touch a stranger. We need to incoporate proper collision detection and appropriate responses, based on behavioural rules, in order to achieve a more chaotic, less military look to the crowds. A real-time, adaptive approach is most suitable for this situation, and the techniques described in Dingliana and O'Sullivan (2000)[7] are being adapted to deal with this situation.

A hierarchy of rigid bounding volumes can be used for collision detection for virtual humans (see Figure 5). If the characters are modelled based on hierarchical transforms then these transforms can easily be used to update the positions of nodes in the collision detection hierarchy. Line swept spheres (LSS) are the volumes generated by sweeping a sphere across a line segment. A hierarchy of LSS nodes is an efficient way to model characters for collision detection as LSS's provide the best combination of ease of computation and tight-boundedness in the case of virtual humans. Although the characters' collision hierarchies are based on LSS's the rest of the environment can be modelled on more general volume representations using heterogeneous collision primitives such as spheres, planes and boxes. Once collisions with the humans have been detected, a realistic response is necessary. An optimal solution framework is currently being developed for the inclusion of physically correct responsive objects within the virtual space. This work is primarily concerned with the provision of a feasible real time level of detail hybrid impulse/constraint based solution (see Figure 6). To evaluate our collision handling techniques, psychophysical experiments similar to those described in O'Sullivan and Dingliana 2001[15] are being employed. We vary the ways in which collision processing can be speeded up and examine the effect of the resultant degradation upon the perception of the viewer.

4. Behaviour

An Embodied Conversational Agent (ECA) is an animated character that has the same properties as humans in face-to-face conversation, including the ability to produce and to respond to both verbal and non-verbal output. The Gesture and Narrative Languge Group at the MIT Media Laboratory has developed an ECA framework that deals with multimodal generation and interpretation of conversational content (e.g. what is being said) and conversational management (i.e. how turns are exchanged). Within this framework, the group has developed a nonverbal behavior generation toolkit called BEAT[5].

BEAT takes as input the text that is to be spoken by an

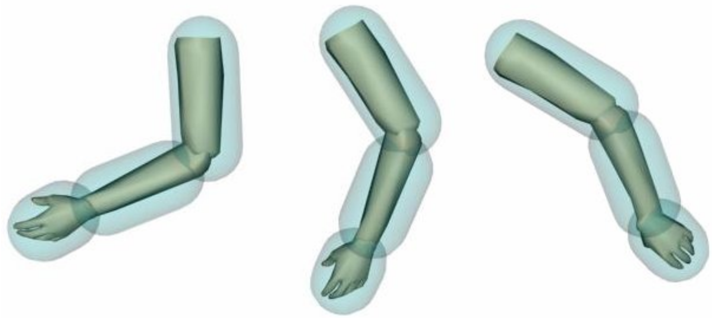

Figure 5: *Line Swept Spheres (LSS) are used to efficiently detect collisions with the virtual humans.*

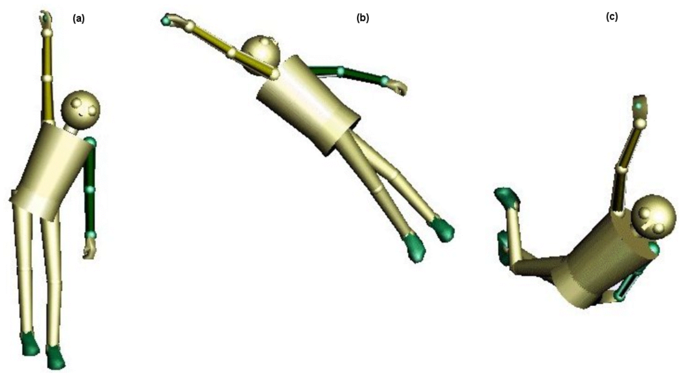

Figure 6: *Dynamic constraint-based solution applied to a hierarchical object - (a) figure is resting under gravity; (b) and (c) show external interaction with object - object is being dragged around by the right hand.*

animated human character, and produces as output appropriate nonverbal behaviors closely synchronized with either synthesized or recorded speech (see Figure 7). The input text could be from a prepared movie script, it could be automatically generated by a natural language planning module or it could be extracted from an ongoing conversation among humans users of a graphical chat system.

The nonverbal behaviors are assigned on the basis of actual linguistic and contextual analysis of the text, relying on rules derived from extensive research into human conversational behavior. Such a principled procedural approach guarantees consistency across modalities that is hard to achieve through stochastic methods and practically impossible to achieve through a purely random behavior assignment.

The modular nature of BEAT makes it easy to add new rules to generate behaviors as well as rules to filter out conflicting behaviors or behaviors that meet certain characteristics. For the purpose of animating groups of people having a conversation, BEAT has been extended to include rules for turn-taking, speaker explicit addressing, feedback elicitation and corresponding listener feedback.

With regard to level of detail (LOD), BEAT lends itself well to controlled generation of behaviors based on visual and functional properties. At the generation level, it is easy to choose which behavior generators are active when processing a given conversation. Filters can also be used to remove generated behaviors based on LOD criteria. At the animation level, since each behavior generator annotates its generated behavior with a visual salience parameter, the LOD framework can selectively drop behaviors as the animated character moves further away or out of the focus of attention.

We are also working on integrating an intelligent agent based role-passing technique into the ALOHA framework. This technique allows intelligent agents to take on different roles depending on the situation in which they are found. Role-passing operates by layering a role on top of a very basic agent. The basic agent is capable of simple behaviours such as moving through a virtual world, using objects and interacting with other characters. The agent may, or may not, have a collection of motivations that drives its behaviour and a number of attributes describing personality traits. Level of Detail Artificial Intelligence (LODAI) techniques can be used to avoid updating complex behaviours for less important characters, while still maintaining some basic functionality as they may become more salient later. See McNamee et al. (2002)[12] for further details.

5. Conclusions and Future Work

This paper has presented a framework for achieving Level Of Detail human animation. This allows refinement of the animation both at a macro level, whereby full components of the system may be activated or deactivated based on importance heuristics, and within the individual components themselves. At the geometrical level, subdivision techniques can be used to achieve smooth rendering LOD changes, while other objects can be pre-emptively simplified. At the motion level, the movements themselves can be simulated at adaptive levels of detail. For behaviour, LODAI can be employed to reduce the computational costs of updating the behaviour of characters that are less important. Furthermore, the knowledge embedded in the system can be used to allow the gesture generation engine to make informed decisions about which features are most important to retain based on the salience of an agent or group of agents, thus allowing graceful degradation of the gesture repertoire.

Although significant gains can be achieved by using level of detail techniques for all tasks involved in the simulation of a complex graphical environment, inbuilt redundancies could be introduced if the system fails to share the prioritisation information generated by one process with the others. We want to make the LOD resolver more generic, so that truly polymorphic LOD control can be achieved. This will involve developing a seamless interface to the knowledge base that can be used to schedule the processing of each of the constituent components of the system.

Acknowledgements

This work has been supported by Enterprise Ireland and the Higher Education Authority.

References

1. A. Aubel, R. Boulic and D. Thalmann. "Real-time Display of Virtual Humans: Level of Details and Impostors", *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on 3D Video Technology* **10**(2), pp. 207–217 (2000). 2
2. D.C. Brogan and J.K. Hodgins. "Group Behaviors for Systems with Significant Dynamics", *Autonomous Robots* **4**, pp 137–153 (1997). 1
3. J. Cassell. "More than Just Another Pretty Face: Embodied Conversational Interface Agents", *Communications of the ACM* **43**(4), pp 70–78 (2000).
4. J. Cassell and H. Vilhjálmsson, "Autonomy vs. Direct Control: Communicative Behaviors in Avatars". *Autonomous Agents and Multi-Agent Systems* **2**(1), pp 45–64 (1999).
5. J. Cassell, H. Vilhjálmsson and T. Bickmore. "BEAT: Behavior Expression Animation Toolkit." *Proceedings SIGGRAPH 2001*, pp 477–486 (2001). 3
6. T. DeRose, M. Kass and T. Truong. "Subdivision Surfaces in Character Animation", *Proceedings SIGGRAPH 1998*, pp 85–94 (1998). 2
7. J. Dingliana and C. O'Sullivan. "Graceful Degradation of Collision Handling in Physically Based Animation", *Computer Graphics Forum. (Proc. Eurographics 2000)* **19**(3), pp 239–247 (2000). 3
8. N. Dyn, J.A. Gregory and D. Levin. "A Butterfly Subdivision Scheme for Surface Interpolation with Tension Control", *ACM Transactions on Graphics* **9** (2) pp 160–190 (1990). 2
9. T. Giang, R. Mooney, C. Peters and C. O'Sullivan. "ALOHA: Adaptive Level Of Detail for Human Animation: Towards a new framework", *Eurographics 2000 short paper proceedings*, pp 71–77 (2000). 2
10. W. Leeson. "Subdivision Surfaces for Character Animation", *Games Programming Gems III*, (To Appear 2002). 2
11. C. Loop. "Smooth subdivision surfaces based on triangles", *Master's thesis, University of Utah, Department of Mathematics* (1987). 2
12. B. MacNamee, S. Dobbyn, P. Cunningham and C. O'Sullivan. "Men Behaving Appropriately: Integrating the Role Passing Technique into the ALOHA System", *Proceedings of the AISB'02 symposium: Animating Expressive Characters for Social Interactions (short paper)* (To Appear 2002). 5
13. S.R. Musse, F. Garat and D. Thalmann. "Guiding and Interacting with Virtual Crowds in Real-time", *Proceeding of the Eurographics Workshop on Computer Animation and Simulation (CAS '99)* pp 23–34 (1999). 1
14. S.R. Musse and D.Thalmann. "A Behavioral Model for Real Time Simulationof Virtual Human Crowds", *IEEE Transactions on Visualization and Computer Graphics* **7**(2) pp 152–164 (2001). 1
15. C. O'Sullivan and J. Dingliana. "Collisions and Perception", *ACM Transactions on Graphics* **20** (3) pp 151–168 (2001). 3
16. C. Peters and C. O'Sullivan. "Vision-Based Reaching for Autonomous Virtual Humans", *Proceedings of the AISB'02 symposium: Animating Expressive Characters for Social Interactions (short paper)* (To Appear 2002). 3

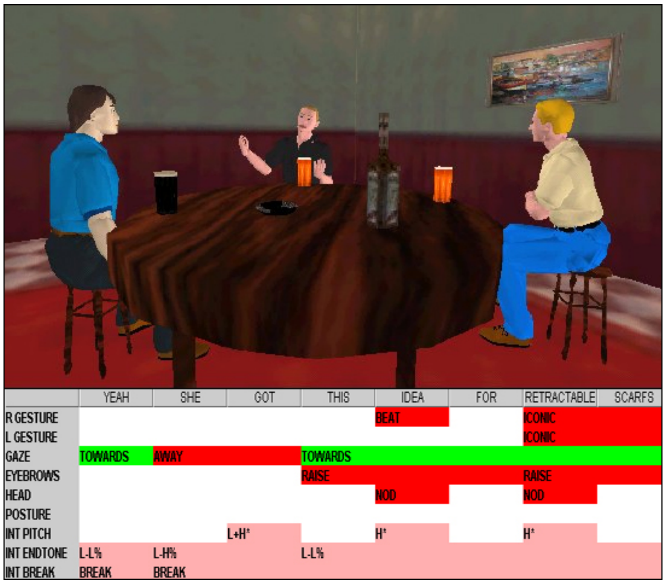

Figure 7: *The BEAT gesture toolkit showing speech synchronized output for the speaker.*

17. C. Reynolds. "Flocks, Herds, and Schools: A Distributed Behavioral Model", *Proceedings SIGGRAPH 1987* pp 25–34 (1987). 1

18. F. Tecchia, C. Loscos and Y. Chrysanthou. "Real Time Rendering of Populated Urban Environments", *ACM Siggraph'01 technical sketch* (2001). 2

19. F. Tecchia and Y. Chrysanthou, "Real-Time Rendering of Densely Populated Urban Environments", *Rendering Techniques'00 (Proceedings 10thEurographics Workshop on Rendering.* pp 45–56 (2000). 2

20. D. Zorin, P. Schröder and W. Sweldens. *Interpolating Subdivision for Meshes with Arbitrary Topology*, *Proceedings SIGGRAPH 1996* pp 189–192 (1996). 2