Ambient Fields: Representing Potential Sensory Information

Anthony Steed* University College London
Vijay Pawar† University College London
Sebastian Friston‡ University College London
Mandayam A. Srinivasan § Touch Lab, MIT University College London

Abstract

It is increasingly apparent that the traditional scene graph is not fulfilling the requirements of real-time interactive systems. The use of a single graph as a representation of the current state of the world means that display systems, that may operate at very different rates, or may need to predict ahead the state, need to be very tightly integrated with behaviour and semantics. In this position paper, we will propose a type of field called the "ambient field" which represents information proximate to the user's senses, which they could sample over short time periods. These fields might represent audio, video, haptic or other potentially sensed information. A display device can then sample these fields as necessary to construct the best representation possible at its own display rate. The ambient field draws on the concept of the ambient optical array from Gibson, light fields from computer graphics rendering and point-based physics simulations.

Keywords: virtual reality, real-time interactive systems, scene graphs, sensory fields

Index Terms: H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities;

1 Introduction

Modern real-time interactive systems for virtual reality tend to be dominated by the paradigm of a scene graph. In fact, the types of scene graph used by many game engines are often targeted primarily at the visual elements of the scene. There may be additional nodes for audi and other effects, and attributes on nodes, or individual nodes that represent behaviour or constraints. These scene graphs are typically based around surface representations (polygons, sub-division surfaces, etc.) because these representations are what current graphics cards are designed to support (e.g see [8]), but also because the process of creating and editing such representations is well understood. The scene graph has served us very effectively. Current consumer virtual reality systems are well serviced by game engines such as Unreal and Unity.

Modern virtual reality engines are not as simple as console games. The scene graph paradigm falls down in at least two areas: the representation of time, and representation of complex sensory effects.

In this position paper we provide a short critique of the current prevailing paradigm based on our experience in building some state of the art virtual reality systems and some requirements from systems that we wish to build in the next few years. We then present our position on what we believe is a starting point for a new set of data structures that can represent virtual environments for very short time periods suitable for re-sampling at different frame rates or at different spatial resolutions.

*e-mail:a.steed@ucl.ac.uk
†e-mail:v.pawar@ucl.ac.uk
‡e-mail:s.friston@ucl.ac.uk
§e-mail:srini@mit.edu

IEEE 9th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS) 2016
20 March, Greenville, SC, USA
978-1-5090-4275-3/16/$31.00 ©2016 IEEE

2 Critique of the Scene Graph

Although scene graphs model time-changing phenomena, they typically do not represent time within the scene graph. They may represent dynamics of rigid bodies, but the actual state of the scene graph will be based on a specific virtual clock time. This virtual clock times are often calculated at a fixed rate dependent on the frame rate of the rendering sub-system. Other processes such as physics simulation may work at the same but may integrate over smaller time slices for stability. Some processes such as routing algorithms for characters may run at a slower rate because the computation is very expensive.

The first concrete problem we have encountered is commonly experienced by those doing haptic simulations: haptic devices require output at 1kHz and this is incompatible with a game engine where the physics engine update rate is set by the requirements for visual rendering. Given lack of control over the internal update loop and flow control of the engine, implementations resort to including a second scene graph dedicated to haptic simulation and then synchronising the haptic and main engine scene graphics. Although this model can work, it has a significant overhead, and means that it can be difficult to add haptics simulations to existing demonstrations.

The second concrete problem we have encountered is in very low latency systems. The current paradigm for visual rendering is based around the scene graph where full images for display are calculated at fixed points in time. Ideally frames are available at maximum full frame display rate, say 90Hz, but even this is a challenge leading to many works that try to decouple display rate from frame rate, or using rendering at multiple frame rates (e.g. [9]). However the full frame needs to be communicated to the display over some video interface, and thus there is inherently a frame delay that is unavoidable. To get lower that this, pixels can to be computed on the fly (e.g. [3]), but each needs to be rendered at a separate time which is incompatible with a frame-based renderer.

The third concrete problem is that certain types of simulation fit badly with the scene-graph paradigm, or at least the use of surface representations. For example, De et al. present a critique and model for complex surgical training simulations [2]. They present a point-associated finite field method to unify various types of physical behaviour and interaction. Recently, NVidia has developed tools for unifying various types of physical simulation in point-based representation [7].

We can also foresee several problems that will be barriers in the short and mid-term. Firstly, in telepresence scenarios, the reconstruction of a scene-graph might be unnecessary if sensed stimuli (e.g. panoramic video) can be real-time deformed to compensate for latency of transmission. Secondly, haptic devices will get more degrees of freedom, and the combination of body-mounted, floor-mounted and encountered-style haptics will mean that it is difficult to represent the scene in such a way that a single physics engine can simulate each. Thirdly, as more senses are stimulated by displays, there is a need to consider how to represent, for example, radiant heat or smell, in a scene.

3 Ambient Fields

Our proposal is to consider creating an intermediary set of data structures that represent the current fields of information that the human senses could sample. That is, instead of representing the scene at all times and over large volumes, we focus on that information that the user could experience. This would include the light rays near the eyes, the audio waves near the ears, the forces encounterable near the skin of the user, the composition of molecules in the air that could reach the nose and mouth. We call these the *ambient fields* as they are inspired by the concept of the ambient optic array described by Gibson ([4, 5]).

Of course such a representation, if complete and time-varying would substitute for the scene graph. However, we target the generation of very short-lived fields that can be re-sampled at extremely fast rates by display hardware at very low computational cost. The main concept would be that a scene-graph or composition of scene-graph and other representations would generate ambient fields as fast as possible, but the core logic (scripts, constraint resolution) would not represented in that ambient field, except for, potentially, very simple forward change.

For computer graphics, similar *l*ight field representations have been investigated [6]). Light fields represent the light in an environment as a 4D field. An ambient light field would only be concerned with light that could be modulated by some display device to enter the eyes of the user. Thus some representation such as a variant of surface light fields might be appropriate [10].

For other senses, other field representations might be more appropriate, such as frequency representations (e.g. audio) or simple voxel-like structures (e.g. molecules in air for smell reproduction). These fields would need to be calculated at an appropriate resolution and extent for the sense of the user, so they only need to be sampled at positions where the senses can move. Thus given we can model the movement of the user, we might model different volumes: the head can move at certain rates and thus the volume through which the pupil of the eye can potentially move can be calculated. For haptics, we would use the space that the limb in question could reach in a short time period and might have varying densities of representation depending on the spatial sensitivity of the body (e.g. high resolution near the fingertips, low in the mid-back area).

The field might be time-varying with some notion of simple interpolation between spatial samples, or might contain simple forward dynamics (e.g. a Langrangian point or a Eulerian grid representation). However the constraint on constant or bounded, low re-sampling would need to be adhered to.

4 NEXT STEPS

This position paper calls for a discussion about how to define intermediate representations that are highly renderable. Effectively these would replace, or be at a lower level than standards such as OpenGL, OpenAL or Chai3D [1].

There are two main potential benefits for this approach. The first is that we can decouple application-level simulation from the rendering and display issues. A side effect would be that it would allow hardware designers or labs to focus on reproduction of simple representations, rather than deal with a full-stack of software and driver issues. A second is that it gives a more direct way to reason about the mis-match between sensory fidelity and display. Currently we compare systems on crude metrics such as field of view, frame rate and frequency of response; whereas the human senses have complex non-separable spatial and temporal responses. Thus an ambient field representation might allow us to reason about the difference between the potential for sensing and the display capabilities.

ACKNOWLEDGEMENTS

REFERENCES

[1] Chai3D. `http://www.chai3d.org/`. Accessed: 2016-01-30.

[2] S. De, Y.-J. Lim, M. Manivannan, and M. A. Srinivasan. Physically realistic virtual surgery using the point-associated finite field (PAFF) approach. *Presence: Teleoper. Virtual Environ.*, 15(3):294–308, June 2006.

[3] S. Friston, A. Steed, S. Tilbury, and G. Gaydadjiev. Construction and evaluation of an ultra low latency frameless renderer for vr. *Visualization and Computer Graphics, IEEE Transactions on*, PP(99):1–1, 2016.

[4] J. J. Gibson. Ecological optics. *Vision research*, 1(3):253–262, 1961.

[5] J. J. Gibson. *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press, 2014.

[6] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42. ACM, 1996.

[7] M. Macklin, M. Müller, N. Chentanez, and T.-Y. Kim. Unified particle physics for real-time applications. *ACM Trans. Graph.*, 33(4):153:1–153:12, July 2014.

[8] H. Sowizral. Scene graphs in the new millennium. *Computer Graphics and Applications, IEEE*, 20(1):56–57, 2000.

[9] J. P. Springer, C. Lux, D. Reiners, and B. Froehlich. Advanced multi-frame rate rendering techniques. In *Virtual Reality Conference, 2008. VR'08. IEEE*, pages 177–184. IEEE, 2008.

[10] D. N. Wood, D. I. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. H. Salesin, and W. Stuetzle. Surface light fields for 3d photography. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 287–296, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.