Research Masters Internship (Stage de Recherche MASTER)

Development of an Audio-Visual Saliency Map

INRIA Sophia-Antipolis, France

REVES (http://www-sop.inria.fr/reves)
Supervisors : George Drettakis George.Drettakis@inria.fr / Nicolas Tsingos

General Presentation of the Research Domain

The focus of the REVES research group is on image and sound synthesis for virtual environments. Our research is on the development of new algorithms to treat complex scenes in real time, both for image rendering (for example the capture and rendering of trees using an image-based technique [1]) or for sound (for example using perceptual masking and clustering to render complex sound scenes [2]). We are coordinating the new EU project IST/FET CROSSMOD, which starts on December 1st 2005, on the perceptual interaction between the audio and visual channel and the effects of this interaction on rendering and user attention for both sound and images.

Objectives

In recent research computational models for visual [3] and audio [4] attention have been developed, which permit the prediction or identification of the "salient" parts of an image or a soundscape. In computer graphics, such *saliency maps* can be used to guide rendering of a scene, by allocating more resources to the more visually important parts of a scene. Understanding the mechanisms leading to increased saliency can also be used to guide user attention to a specific part of the scene. In the past, visual saliency maps have been used in this goal; audio saliency is an extremely recent development, and thus is still unexplored.

In this M.Sc. we have to main objectives:

1. Accelerate the computation of saliency maps, both for images and for audio. In particular, we believe that a large part of the processing for these saliency maps can be performed on the GPU. In addition, we can use approximations, both for the rendered images (in the spirit of "Snapshot" [5]) and in the computations performed, thus greatly accelerating the overall computation. The validity of the approximations made will be validated using user-experimental procedures, confirming that the predictive capacity of the approximate saliency maps is maintained. Ideally, the computation of the approximate saliency map will be fast enough for use in real-time.

2. Given the visual and audio saliency maps, we will develop a joint, or cross-modal, saliency map, which incorporates both senses. As a first step, implementations of the two maps will be simply used as a weighted combination, and simple user-experiments will be designed to verify their pertinence. Depending on the consistency and coherence of the results, a more involved combination will be developed, and potential unpredictable interactions between the two sensory channels will be modelled, at least for the specific cases investigated.

Methodology

The successful candidate will use C++, OpenGL or other graphics API's to develop the software required for this research. An implementation of [3] is publicly available; the audio saliency map [4] will be implemented by the student, based on our in-house audio API. The experimental verifications will be implemented in the Ogre3D [6] engine, which is supported in our group. Knowledge of psychophysical experimental setups and statistical analysis methods will be a plus for the analysis of the user experiments.

The design of the experiments will be performed in collaboration with the neuroscientist partners of the CROSSMOD project, at the UMR 7593 CNRS at the Hopital Salpetriere in Paris.

References

[1] Reche, A., Martin, I., & Drettakis, G. (2004, August). *Volumetric reconstruction and interactive rendering of trees from photographs*. ACM Transactions on Graphics (SIGGRAPH Conf. Proceedings), 23(3).

[2] Tsingos, N., Gallo, E., & Drettakis, G. (2004, August). *Perceptual audio rendering of complex virtual environments*. ACM Transactions on Graphics (SIGGRAPH Conf. Proceedings), 23(3).

[3] Itti, L., and Koch, C. (1998). A model of Saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Learning 22, 1254–1259.

[4] Christoph Kayser, Christopher I. Petkov, Michael Lippert, and Nikos K. Logothetis, Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map Current Biology, Vol. 15, 1943–1947, November 8, 2005, DOI 10.1016/j.cub.2005.09.040

[5] Peter Longhurst, Kurt Debattista, Alan Chalmers: Snapshot: A Rapid Technique for Driving Global Illumination Rendering. WSCG (Short Papers) 2005: 81-84

[6] http://www.ogre3d.org