Augmented Virtual Environments (AVE) for Visualization of Dynamic Imagery

Ulrich Neumann, Suya You, Jinhui Hu, Bolan Jiang, and JongWeon Lee

Integrated Media Systems Center
University of Southern California
Los Angeles, CA 90089-0781

Abstract

An Augmented Virtual Environment (AVE) provides a means of fusing dynamic imagery on a 3D model substrate. The AVE approach allows users to visualize and comprehend multiple streams of imagery (video and still images) in a four-dimensional context. The addition of projected live or recorded imagery to an otherwise static virtual environment creates an AVE. Our methods are described in the context of a prototype system for visualizing activities on the USC campus.

Introduction

Realistic three-dimensional environment models are used for virtual environment (VE) applications in engineering, mission planning, training simulations, and entertainment. In many cases, the value of the VE is increased if both the geometric information and the appearance of the virtual environments are accurate and realistic analogues of an analog in the real world. However, VE models capture only a snapshot of the real world; a VE model of a campus does not reflect the movements of cars and people moving around the buildings and streets. Textures can produce realistic building facades and street surfaces, but these are snapshots and therefore lack any representation of dynamic events and activities occuring in the scene. A static VE model is augmented with live or recorded images (video or repeated snapshots) to create an augmented virtual environment (AVE) that provides a visualization of both the geometry and dynamic imagary in a comon 3D context. By projecting real-time video onto the 3D surface of the scene model, the AVE not only adds the video information to the visualization, but adds it in the 3D context of the model. The use of a comon 3D context allows multiple image streams to be visualized simultaineously without the usual cognitive load and difficulties associated with users fusing multiple images taken from differing viewpoints.

As with any virtual environment, an AVE allows users to freely move their viewpoints from a “god’s-eye” view that visualizes an entire region of an environment to a specific area of interest, such as a building entrance. From any viewpoint, users observe multiple video streams from moving aerial or ground-level cameras projected onto the model, painting real-time views of the actual events and activities occuring in the real world.

Figure 1 - An AVE system screen snapshot showing three video projections within a campus area. The top-left window shows a novel viewpoint and the three windows show the rendered views from the three sensor viewpoints.

Figure 1 shows a simple AVE visualization of a campus building complex with three video streams projected onto the model. In an aerial view (top left), three moving cameras are depicted by their red wireframe viewing frustums to show their current positions and orientations in the world. Note that the visualization viewpoint is campletely arbitrary, and the aerial view aids users in fusing and comprehending the multiple camera images and their relationships to the scene.

We envision such aerial views as providing a user with navigation and browsing capabilities for all the imagery available from cameras in the environment. While an individual camera image is easy to understand, people have difficulty in comprehending the relationships of multiple images to a scene and the replationships of the images to each other.

As the number of cameras and other networked sensing systems viewing the world proliferates, there is a need for a framework for processing the aggregate information that is captured by these sensors for presentation and comprehension by people. The Augmented Virtual Environment is pursued as one approach to this problem. Informal demonstrations of our prototype system show that the AVE approach has potential for fusing a great deal of information in a fashion that is easily browsed and interpreted by users.

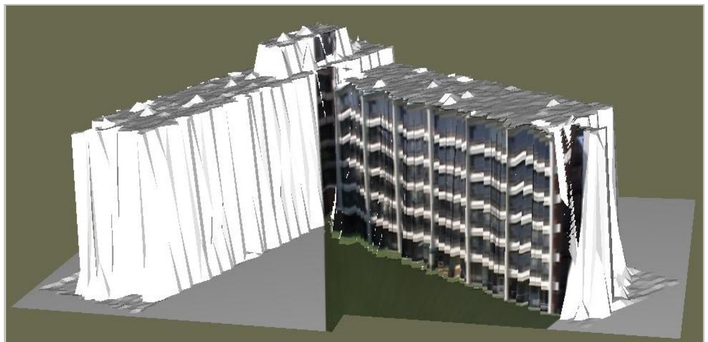

Figure 2 – Texture projection on reconstructed building model from LiDAR data shows artifacts

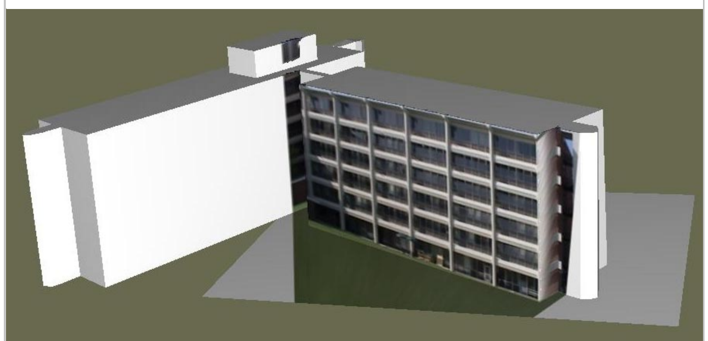

Figure 3 – Refined model with projected texture

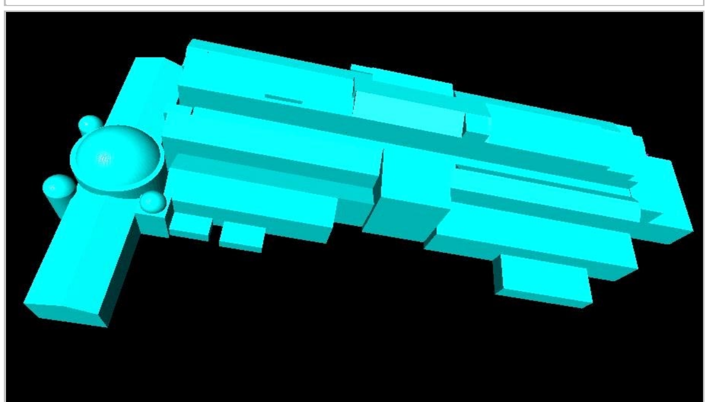

Figure 4 – Refined model of LA Natural History Museum

AVE System

Since AVE visualization is a reflection of a real world scene, we require accurate models of real scenes. Real world scene models are often obtained from images and range data [1]. Many methods also add static texture imagery to produce photorealistic visualizations [2]. Our approach is to use an active airborne laser sensor to acquire building footprints and roof data. The raw data samples are a point cloud with ~0.5-1.0 meter ground-spacing and centimeter height accuracy registered to a world coordinate system. Multiple passes of the aircraft are merged to ensure good coverage. This data is pre-processed to create a consistent 3D geometric mesh model of the environment.

Figure 2 shows a building mesh produced by processing LiDAR data. The texture projection emphasizes the artifacts that can be corrected by using knowledge of building construction (Figure 3). Our semi-automated method fits linear and high-order (superquadric) surfaces to the mesh data to produce constrained models that replace the LiDAR data in the enclosed regions. Figure 3 shows the visual quality benefits obtained with texture projection on these models. Our software allows users to create building models in a few minutes by selecting a few points in the LiDAR data. Figure 4 shows the LA Natural History Museum across the street from the USC campus. Note the inclusion of multiple slanted roof segments and a dome in this model. Figure 5 shows a full model of our campus and University Park area including the Coliseum, LA Arena, museums, and gardens.

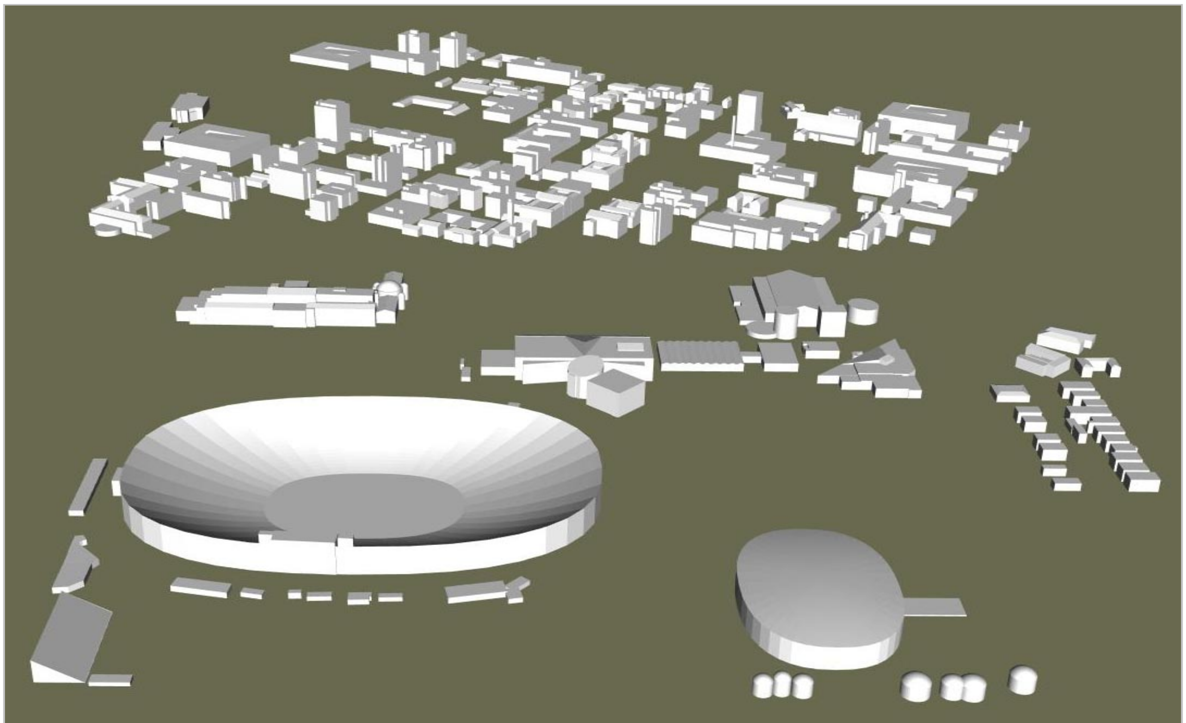

Figure 5 – Complete refined models of USC campus and surrounding University Park area.

Texture mapping heightens the realism of models; however, texture maps are often static and their mapping to the geometry is fixed and must be established prior to rendering. Texture mapping is not appropriate for AVE visualization since the cameras capturing texture imagery are often moving in a scene, and therefore, the mapping from image to geometry is dynamic. Texture projection computes this mapping during scene rendering [3] from the internal camera parameters and the camera's position and orientation (or pose) for each frame.

While texture projection is a powerful approach to integrating dynamic imagery with 3D models, there are potential pitfalls. Simple texture projection results in textures on all surfaces within the frustum of projection. So, visibility information is needed to modulate the projection process. Visibility processing has to be fast in order to support real-time visualization. Fortunately, depth-map shadows offers an approach to fast visibility detection that is supported by many high performance graphics cards, such as NVIDA's Geforce series GPU that supports 24-bit shadow maps. A P4 2GHz PC achieves real time rendering (26 Hz) of 1280x1024 visualizations with four concurrent texture streams projected onto our campus model. Figure 6 shows a projection of high-resolution imagery onto refined campus building models.

Camera tracking is needed to provide pose for texture projection. We combine a high resolution Firewire (IEEE 1394) stereo camera head (MEGA-D from Videre Design), a differential GPS receiver (Z-Sensor base/mobile from Ashtech), 3DOF inertial sensor (IS300 from Intersense), and a laptop computer to produce a tracking system coupled to a video camera. The 6-DOF (degree of freedom) pose of the camera is tracked by the sensors in real-time as the system moves in an open outdoor environment. The tracking data is encoded and stored with the camera video stream in real-time on the computer hard-drive. The complete camera, tracking, and recording system is housed in a backpack for mobility (Figure 7).

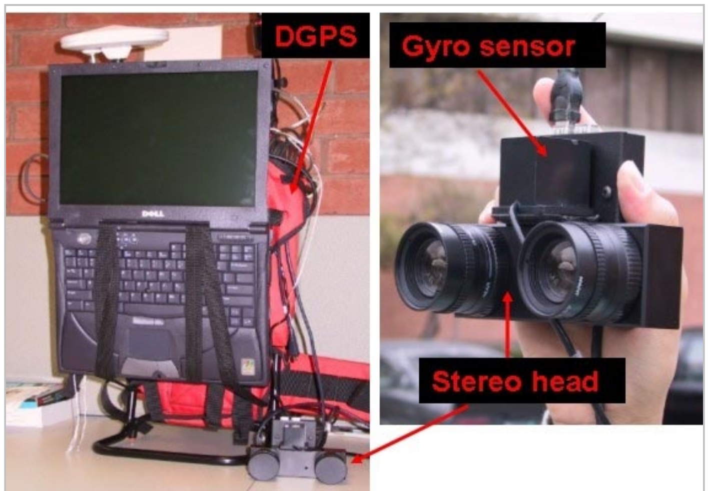

Figure 7 - Portable data acquisition and tracking system

While the sensor platform provides good initial pose estimates, a vision-based pose refinement improves the stability and accuracy of texture projection from moving cameras. Line features (or edges) are detected in the camera images and corresponded to 3D model edges based on proximity and correlations in a projected-model image created by using the sensor pose estimate. The pose estimate is refined in a Kalman filter to minimize the corresponding projected edge distances [4]. This pose refinement is needed for our backpack sensor package, however, better sensors, or constrained or calibrated camera motions may not need such refinement.

An integrated prototype system demonstrates the concept of AVE visualization. The display consists of an 8x10 foot screen, back-projected by a Christie sequential-frame stereo video-projector. A ceiling tracker is used to couple the rendering viewpoint to user's head position, providing a user with a highly immersive visualization environment (Figure 8).

Open issues and ongoing research

In closing, some comments are appropriate on the limitations of AVE systems. Clearly, objects that are not part of the model are only properly displayed from the camera's viewpoint. For example, lamp poles, cars, and trees are projected onto the building and ground models and they look warped and distorted from viewpoints other than the camera's. Subjectively, however, people seem to adapt to those distortions, perhaps because they are local and the camera viewing frustums are visible. Clearly, human perception and performance testing is required to better understand the benefits and shortcomings of AVE visualizations.

Another issue is the lack of texture data from projections seen a few moments ago. Even casual users comment that they would prefer that the projection data persist after camera pose changes. Static textures could be prepared to cover the entire scene (as shown in Figure 6) and projections would only modify scene regions in view of cameras, however the dynamic events would still be lost. Other approaches seem feasible, are we are pursuing them. AVE visualization provides a rich area for future research.

Lastly, performance is an issue. A major issue is video bandwidth. AVE visualization is data-intensive. If scaled to modest levels of 20 concurrent video streams, the video data exceeds the movement and rendering capabilities of any existing PC and graphics system.

Figure 8 - AVE system offers users a dynamic visualization environment

Acknowledgments

This work was supported by a Multidisciplinary University Research Initiative (MRUI) on "Next Generation 4-D Distributed Modeling and Visualization", and by a 2002 U.S. Office of Naval Research (ONR) DURIP award. We thank the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, for their support and facilities. We thank Todd Stennett at Airborn1 Inc. for providing us with the USC LiDAR data. Our thanks also go to NVIDIA, Intel, and Microsoft for equipment support.

References

1. C. Fruh and A. Zakhor, "3D Model Generation for Cites Using Aerial Photographs and Ground Level Laser Scans", *IEEE Conference on Computer Vision and Pattern Recognition*", 2001.

2. P. E. Debevec, C.J.Taylor and J. Malik, "Modeling and Rendering Architecture from Photograph: A Hybrid Geometry-and Image-based Approach", *SIGGRAPH' 96*, pp. 11-20, 1996.

3. Segal, M., Korobkin, C., Van Widenfelt, R., Foran, J. and Haeberli, P., "Fast shadows and lighting effects using texture mapping", *SIGGRAPH '92*, pp. 249–252, 1992.

4. B. Jiang, U. Neumann, "Extendible Tracking by Line Auto-Calibration," *International Symposium on Augmented Reality,* pp. 97-103, New York, October 2001.