SIMULATION DRIVEN VIRTUAL OBJECTS IN REAL SCENES

Erol Gelenbe and Khaled Hussain
School of Electrical Engineering and Computer Science
407-823-0345, {erol, khaled}@cs.ucf.edu
Billy Foss, Institute for Simulation and Training
407-658-5540, bfoss@ist.ucf.edu
Niels Lobo, School of Electrical Engineering and Computer Science
407-823-2873, niels@cs.ucf.edu
University of Central Florida
Orlando, FL 32816

Hubert Bahr, U.S. Army STRICOM, Orlando, FL
407-384-3874, hubert_bahr@stricom.army.mil

Abstract

STRICOM, together with the School of Electrical Engineering and Computer Science (SEECS) and the Institute for Simulation & Training (IST) at the University of Central Florida are developing a system to allow virtual objects to be placed in live images in real time. The proposed approach is simulation driven in that it will use a geometric database of the site of the live scene to drive a simulator which will be used to predict the location of the synthetic object in the real scene at each instant of time. The research we are conducting involves object identification in the real world scene using registered overlays, registration of the real world view with the synthetic view of the virtual terrain data, placement of the virtual object with the synthetic terrain and then the natural view using simulation, and finally realistic integration of the synthetic object into the live scene. New techniques are being developed to determine the occlusion of virtual objects based on their relation to terrain features in the live scene. This paper describes the whole process used in the project, discusses the basic algorithms and presents novel techniques used for recognition and placement of the objects.

Biographical Sketch:

Erol Gelenbe, Ph.D. (Polytechnic University), DSc (University of Paris). He is a Professor of Computer Science, Director of the School of EECS and Associate Dean of Engineering at the University of Central Florida in Orlando. Previously he was the Nello L. Teer Jr. Professor and Chair of Electrical and Computer Engineering at Duke University. His honors include Fellow of the IEEE (1986), Chevalier de l'Ordre du Merite (France, 1992), Dott. Ing. ``Honoris Causa'' of the University of Rome (Italy, 1996), Grand Prix France Telecom (French Academy of Sciences, 1996), Science Award of the Parlar Foundation (Turkey, 1995).

Khaled Hussain received his Baccalaureate and Masters degrees in Electrical Engineering from the University of Assiout (Egypt). He is currently a doctoral candidate in Computer Science at UCF. His interests are in neural networks, learning algorithms and image processing.

Billy Foss is an Associate Research Computer Scientist at the Institute for Simulation and Training. He is currently researching voice recognition in an HLA environment and virtual object overlay into lives scenes. He also is investigating applications of cognitive modeling in distributed simulation. Mr. Foss received a B.S. and a M.S. both in computer science from the University of Central Florida. He is currently pursuing a Ph.D. His research interests are in distributed computing, software architecture and simulation.

Niels da Vitoria Lobo is an Associate Professor in the School of Electrical Engineering and Computer Science at the University of Central Florida. He received the B. Sc. (Honors) degree in Mathematics and Computer Science from Dalhousie University, Canada, and the Ph.D. in Computer Science from the University of Toronto.

Simulation Driven Virtual Objects In Real Scenes

Erol Gelenbe and Khaled Hussain
School of Electrical Engineering and Computer Science
407-823-0345, {erol, khaled}@cs.ucf.edu
Billy Foss, Institute for Simulation and Training
407-658-5540, bfoss@ist.ucf.edu
Niels Lobo, School of Electrical Engineering and Computer Science
407-823-2873, niels@cs.ucf.edu
University of Central Florida
Orlando, FL 32816

Hubert Bahr, U.S. Army STRICOM
407-384-3874, hubert_bahr@stricom.army.mil

BACKGROUND

This research will investigate several issues critical for the success of the Inter-Vehicle Embedded Simulation Technology (INVEST) Science and Technology Objective (STO). These issues include terrain registration and image processing. This approach to solving these problems involves techniques for real time capture and processing of natural imagery for display (synthetic and natural modes), storage of a local terrain database and culture, and multi-scale image analysis. This approach will preclude having to construct a terrain database with sub-meter resolution, and avoid the expense of excessive instrumentation. The product of this research will be a Target Overlay System (TOS) that will support INVEST technology demonstrations in May 2001. Using the TOS, a demonstration vehicle, using display devices representing the M1A2 SEP tank, will view virtual targets moving realistically over live terrain. This will be achieved while the demonstration vehicle is moving over the live terrain. This research effort will involve UCF faculty from the School of Electrical Engineering and Computer Science and the Institute for Simulation and Training. The UCF team will also work closely with Carmel Applied Technology, Inc. (CATI) due to their existing role as the INVEST image generation lead.

The School of Electrical Engineering and Computer Science (SEECS) will investigate the overlay of synthetic moving targets into natural scenes. The purpose is to implement synthetic target overlay into real scenes in real-time so as to develop embedded training systems, which combine real and synthetic vision, as well as real-time simulation and training with synthetic fire, using operational terrain databases.

To simplify the discussion we introduce the following terminology. A "viewer" will be a vision-based device, which renders on a screen the natural scene being viewed either in the visible frequency range, or in the infrared range, or both. The viewer is assumed to be equipped with a targeting device designed to allow the human observer to point a weapon at a particular location in the scene. This research task will not address issues related to the algorithms used for actually pointing the weapon itself at the target. We will only deal with the pointing device in the viewer, and assume that the transformation of the pointing and aiming mechanism into a firing procedure is handled automatically by other means. However, this task will provide the pointing angle to the visual system.

This research will address the following important components:

1) Initial identification of relevant areas of the terrain from the scene observed through the viewer. Typically, the terrain database will initially call data covering an area, which is larger than the terrain being viewed.

2) Registration of the terrain being viewed with the data from the terrain database. This work will call upon specific pattern recognition algorithms, followed by more precise local searches.

3) Techniques to realistically mask or occlude virtual entities as seen from the viewer with terrain features, cultural objects, buildings and other real world objects in a live scene. The research will address techniques to wholly or partially occlude virtual entities resulting from a virtual entity moving behind a real world object, an object moving in

front of a virtual entity or a change in viewing angle causing the virtual entity to be masked.

4) The terrain database will be included in a simulation of one or more potential targets, which are maneuvering in the terrain database. At some point in time, one or more potential targets being simulated in the synthetic terrain may enter the area, which has been registered with the natural terrain being viewed through the viewer. At that point, the potential targets will be rendered in the field of vision of the viewer, in realistic positions and with realistic motion within the field of vision.

We intend to experiment with algorithms for deterministic and stochastic alignment of the real image with a synthetic database. Deterministic alignment would be with terrain and discrete objects such as roads, streams, lakes, and buildings. Stochastic processes would be applied to 'softer' features such as trees, forests, rocks, etc. Deterministic or stochastic categorization would be based on a sampling of SEDRIS feature classification codes. Our work will begin with the ownship in a static position and, therefore, be oriented to angular excursion of the turret. Future work may add complexity to include ownship movement, other live vehicles, and multiple virtual vehicles.

We expect that phase one research will address points 1) and 2). During this phase of the project, we will use static target locations which are derived from the terrain database, and demonstrate a system where (a) the vehicle from which training is being conducted is not moving, (b) the viewer or viewing device can be moved vertically and horizontally while the terrain and the target remain static. Technical papers and computer software that implements the algorithms researched will be the products of this research. We will work with other researchers integrate the software into a prototype system used to demonstration this capability.

The second phase will cover points 3) and 4). We will integrate this work with the INVEST version of ModSAF so as to demonstrate the use of moving synthetic targets. These targets will be generated by SAF and registered, using the terrain database, with the real terrain being viewed in the system. The targets will then be viewed within the natural scene being observed.

ARCHITECTURE OVERVIEW

During the planning of this project, we developed several architectures to demonstrate the algorithms during the development stages. Some of these architectures were used to quickly illustrate the concepts, while others are intended to provide real-time interaction between components.

The first challenge was to determine a representation of the live and virtual environments that could be compared to each other. We evaluated the current representations of each.

Live Representation

The easiest representation of the live environment is through a video camera. This model provides a two-dimensional representation of the scene given to the soldier. It also provides the closest model to the final representation seen by the soldier. The main problem with this representation is that it provides little information about the objects in the scene that could occlude virtual objects

Another approach is to survey the live environment with real-time instrumentation. This method provides more information about objects and their distances in the live environment, but requires more equipment and costs. SRI took this approach in previous work.

Virtual Representation

The primary representation of the SAF terrain environment is through topological terrain formats. The Modular Semi-Automated Forces (ModSAF) simulation uses the Compact Terrain Database (CTDB) format. (Smith, 1995) This format provides terrain height and feature representations. It also provides detailed information about objects in the virtual environment. The virtual targets to be placed in the live scene are of primary interest.

Another approach to representing the virtual environment is by using visual databases. These visual databases are used in stealth and immersive environment applications. Often, they are more detailed than the topological databases because they were designed to be viewed by the human eye. This approach also gives a two-dimensional representation of the virtual environment. This representation is much closer to the available live and expected final representations, but we still have the problem with finding objects and determining occlusion.

ALGORITIHM FOR OBJECT DETECTION

This section presents a method to detect objects in a live camera scene by using image processing algorithms in conjunction with a correlated virtual view. Image processing techniques are used to match certain boundaries in the live and virtual two-dimensional views. Once we determine a reasonable match, virtual targets are placed in the live scene. The process of determining occlusion will be aided by depth information in the virtual image. Many stealth packages use OpenGL (SGI, 1997) or some other three-dimensional rendering API. We plan to use the Carmel Applied Technologies Inc. (CATI) X-IG Stealth.(CATI, 2000) It uses OpenGL as its API to the graphics card, which renders the three-dimensional virtual scene in two-dimensions. The rendering is often done by hardware, which is much faster than software rendering.

As part of the rendering process in OpenGL the depth (Z-buffer) of each object in the scene is determined. This depth is then used to calculate which objects are viewable from the current viewpoint.

Hardware Acceleration

One unique aspect of this project is the use hardware acceleration to render scenes that will be processed by software algorithms. The terrain representation of the virtual world is represented as a set of polygons in three-dimensional space. In order to get a two-dimensional image, these polygons need to be processed though several complex transformations. Newer video cards provide hardware specifically tuned to perform these calculations to improve on-screen rendering of complex 3D environments, like those found in virtual reality games. Several APIs exist so that game developers can write programs that take advantage of the hardware acceleration. Two of the most common are DirectX and OpenGL. We are focused on the OpenGL API.

Normally, video cards render a scene for display on the screen. Our application is unique in that we do not need to or want to render the scene on a visual display. We have identified some functions within the OpenGL API that allow us to read the rendered image into memory before it is displayed on the screen. This allows the image processing algorithms to read the virtual image from memory buffers and the live image from the camera buffers and then perform the matching algorithms on these images.

The glReadPixels() function will read the image from the video card after rendering without displaying the image to the screen.

ALGORITHM FOR IMAGE PROCESSING

During the early stages of this project, we performed analysis on various cameras, video capture boards, stealth packages, and computer systems. This analysis provided the foundation for current development, but there was some delay between the final decisions to order certain components and their arrival and configuration. During this time, we identified several methods to generate live and virtual images for use in developing the image processing algorithms.

Correlated Terrain Development

The first priority was to develop a method to get reasonably correlated live and virtual terrain. Traditionally virtual terrain information has been developed to loosely model the visual perceptions found on the actual terrain. For example, a tree in the virtual environment is typically represented by a generic instance of a tree from the database repository and may have little in common with the live tree. Locations of objects in the virtual world need to properly match the locations found in the real world.

To solve this correlation problem, several solutions were presented. The first solution was to model a local area of terrain so that the virtual terrain could be tweaked to accurately model the real world. The main obstacle with this solution was the generation of both an OpenFlight database for the X-IG Stealth software and a CTDB database for ModSAF to use when simulating the vehicles. Several OpenFlight databases existed for areas of the University of Central Florida (UCF) campus. However, there were no corresponding CTDB databases. Several tools are available to convert from OpenFlight to CTDB, but they require significant manual tweaking of the input data. Another obstacle was the ability to isolate aspects of the live terrain. Variable weather would limit the times we could use the live image. The coming and going of cars, bikes, and pedestrians would also change the environment. Even larger structures change with the continual new constructions at UCF.

The second solution was to develop a tabletop model of the real world that could be used for the live camera scene. The tabletop model could be modeled after an area that already has correlated

OpenFlight and CTDB databases. This approach allows the tabletop to be adjusted to correlate with the virtual terrain. We can also adjust the OpenFlight and CTDB databases to fix any miscorrelation that is discovered. The tabletop also protected the development from stray pedestrians and adverse weather.

For initial development, we chose to develop a tabletop model.

Manual Image Generation

The second priority was to get some sample images that could be used in initial evaluation of the image processing algorithms. We decided to take screenshots of the virtual terrain by using an OpenFlight visualization tool, perfly, which comes with SGI Performer. We positioned the viewpoint manually to match that of the tabletop viewpoint (see section Development of Correlated Terrain). This manual matching should be more automated once the correlated visual terrain and tabletop model are synchronized. The camera position and angle will eventually be fixed so that we can determine the correct angle and viewpoint on the virtual terrain.

We have been able to manually collect several images from perfly and compare them to carefully positioned live camera images of the tabletop model. A screenshot from the perfly view is shown in Image 1. An image form the live camera view is shown in Image 2. The virtual image was scaled to match the live image and is shown in Image 3. Finally, we used the two images to calculate a difference image as seen in Image 4. This difference image shows some areas to refine the image matching.

Image 1: Virtual Scene

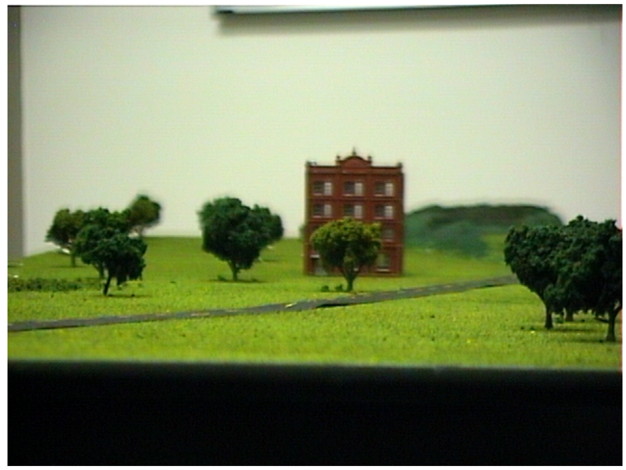

Image 2: Tabletop Camera Scene

Image 3: Scaled Virtual Scene

Image 4: Error/Difference Image

ALGORITHMS FOR OBJECT OCCLUSION

A primary concern with placing virtual objects within a real live scene is the proper placement of the virtual objects in front of or behind live objects. The object occlusion problem requires detailed knowledge of the objects and their placement in the live scene. While the two-dimensional images of the both the live and the virtual scenes provide little information about the objects in the scene, we anticipate using the depth (Z-buffer) from the virtual image to detect objects at different distances. These depths will be used to determine the boundaries between objects in the virtual world. We will use the color image of the virtual scene to match the object boundaries to object boundaries in the lives scene. The depth information from the virtual scene will be matched to the depth information in the live scene, which will give relative depths of objects in a live scene without additional instrumentation to calculate depths of the live world surrounding the observer.

Once this depth information about the virtual world is determined, the insertion of a virtual object or target will be based on the distance at which the target should be placed. Figure 1 illustrates this insertion between partitions.

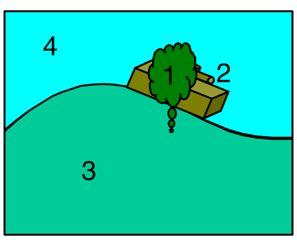

Figure 1: Depth Layering

In this image, the tree was identified as the closest object and given a distance of one. The background sky is farthest away and labeled at distance four. The target virtual tank is placed at distance two. In this image, the tree occludes the tank, while the tank slightly occludes the farther away parts of the ground.

STATUS

We recently developed a prototype implementation, which places virtual tanks into an image of a live scene of the tabletop database. The image of the virtual tank is generated from the CATI X-IG application using a virtual database correlated to the tabletop model. The tank images are then placed into the live static image. We continue progress toward integrating live camera images correlated with camera position and viewpoint information.

CONCLUSIONS

The current work has produced several different areas of interest. Primarily, we have introduced a novel technique to register virtual terrain to live terrain by using correlated two-dimensional images of each. Along the development path, we have identified several methods to work around current limitations of correlated live and virtual terrain. We also discussed methods for using hardware acceleration to speed up the complex analysis of three-dimensional environments.

REFERENCES

CATI (2000, June 26), X-IG Image Generator. <http://catinet.com/products/xig/Default.htm> (2000, Aug. 29).

Smith, J. (1995). Recent Developments in ModSAF Terrain Representation. Proceedings of the Fifth Conference on Computer Generated Forces and Behavioral Representation. Orlando, Florida, May 9-11, 1995.

SGI (1997). OpenGL Man Pages (online). http://www.opengl.org/Documentation/Specs.html (2000, July).