THE AUGMENTED SCENE: INTEGRATING THE MAP AND THE ENVIRONMENT

James E. Mower
Department of Geography and Planning and
Geographic Information System and Remote Sensing Laboratory
University at Albany, State University of New York

ABSTRACT

This paper develops a theoretical and technological foundation for a new kind of interactive map, the augmented scene, that overlays acquired imagery of the environment on a perspective model of the surface in real time. The ground position and viewing direction of users are computed automatically, freeing them from the map orientation process. We define augmented scenes, propose some applications for them, discuss their construction and use, determine the effects of positional and directional error on system performance, explore their data structure, and review a prototype system under development.

INTRODUCTION

In the field, map interpretation depends upon the ability of the user to orient a sense-acquired view of the environment with its mapped representation. By establishing correlations between visible objects in the scene and their abstractions on the map, the user registers the two views and builds a cognitive projection between them, reconciling the viewing parameters of the map with the perspective view of the scene. Unfortunately, many potential map users have no training in map orientation and interpretation. Even the attempts of skilled users are sometimes frustrated by suboptimal viewing conditions.

This paper will develop a theoretical and technological framework for a new type of map, *the augmented scene*, that projects cartographic symbols onto the user's view of the environment in real time. Orientation of the user to the environment is handled automatically by a global positioning system (GPS) receiver, a digital compass, and a digital Abney level. Employing a video camera as an imaging device, the user can scan the environment and choose to overlay selected features of any view with cartographic symbols. The video image provides a graphical index to a GIS database of the scene, stored on a portable computer. The user selects geographic objects for augmentation by clicking on them with a mouse or another graphic pick device.

To see how augmented scenes might be implemented and used, we will

- define augmented scenes,

- suggest some application domains,
- outline the steps required to construct and use them,
- determine the effects of positional and directional errors,
- explore the underlying GIS data structure, and
- view the overall construction of a prototype system.

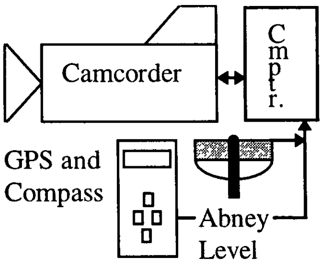
DEFINITION OF AUGMENTED SCENE

An augmented scene is an interactive, symbolized, perspective view of the user's environment seen from his or her current field position that serves as a graphical index to an underlying spatial database. The view may be acquired from captured imagery or simulated with a perspective rendering of the associated 3D surface model. The primary goal of the augmented scene is to provide the user with a direct experience of the map and the environment as a single entity and thereby simplify navigation and map-based queries.

The overlay of acquired imagery on surface models has long been a standard component of computer-assisted map revision systems. Horn (1978) describes a method for registering satellite imagery to surface models for planimetric viewing projections. Drasic and Milgram (1991) discuss techniques for overlaying a user-controlled pointer on captured stereoscopic video imagery to locate objects in the 3D world for robotic navigation within lab settings. This paper extends the use of image acquisition and overlay to perspective views of the general environment.

System overview

The author is currently developing a prototype augmented scene system as illustrated in Figure 1a. Briefly, an augmented scene is created by taking the user's field position (provided by a GPS receiver), horizontal look direction (from a digital compass), and vertical look direction (from a digital Abney level) and combining it with the current focal length and field of view of the imaging device to create a perspective model of the underlying digital elevation model (DEM) registered to the user's actual view. Upon the user's request, the imaging device (in this case a standard video camcorder) captures a single image (frame) of the environment. By default, the computer transmits the frame back to the ocular (or viewfinder) of the imaging device and overlays it with an interactive pointer icon. The user manipulates the pointer in the manner of a mouse or joystick to select a pixel or group of pixels in the image for symbolization (Figure 1b). The computer uses the position of the pointer to look up features in the spatial database located at the world coordinates represented by the selected display coordinates. If poor environmental conditions prevent the acquisition of visual data, the system will use the known positional, directional, and optical parameters to display a perspective map of symbols representing user-selected themes within the current viewing region. A detailed discussion of these procedures follows in the section "Constructing an Augmented Scene."



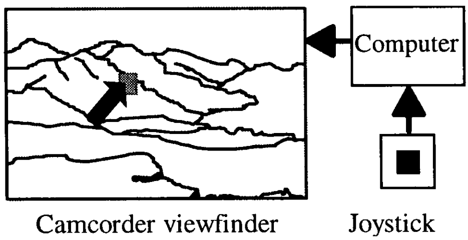


Figure 1a. Augmented scene system components.

Figure 1b. View through system viewfinder. Joystick moves selection pointer.

APPLICATIONS OF AUGMENTED SCENES

Having developed a definition and structural description of augmented scenes, we can consider some of their potential applications. A few of the most interesting applications include:

<u>Location of utility infrastructure.</u> Utility workers could establish an initial framework for construction or repair of underground conduits by creating an augmented scene from a ground position. The scene would overlay symbolic representations of utility features on an acquired image.

<u>Forest fire management</u>. Teams could query the system for evacuation routes under heavy smoke conditions. Users would see a fully-symbolized perspective view of their environment when smoke conditions preclude the acquisition of useful visual information. They would require the ability to "look behind" objects to navigate beyond their current viewshed.

<u>Recreational uses.</u> Hikers could query the system for the names of features in their environment and their distance to them. Their primary interest will likely be the determination of names and distances to visible features in their environment. They may also ask, "What's behind that hill over there?"

<u>Urban applications.</u> Given a sufficiently rich urban GIS, users could query a building within an image for ownership, structural information, and escape routes. This would require the augmented scene to shift the virtual viewing position anywhere along the user's line of sight.

CONSTRUCTING AN AUGMENTED SCENE

The construction of an augmented scene consists of 1) determining the user's position, horizontal look direction, and vertical look direction; 2) determining lens parameters of the imaging device; 3) building a perspective model of the landscape within the scene; 4) capturing a representative frame from the video camera; 5) registering the captured frame with the perspective model; and 6) transmitting the computer display of the frame (including the mouse pointer and menu structure) to the imaging device ocular.

1) Establishing user position and view angles. The user's position is acquired through a GPS receiver that transmits its data to the system computer. GPS is the most suitable technology for this project because it is highly portable and, under real-time differential operation, accurate to within several meters of true ground position.

Urban environments pose well-known problems for GPS signal reception. The accuracy of a positional calculation depends upon the clear, straight-line reception of timing signals from 4 or more satellites. Buildings can block signals entirely or reflect them, artificially increasing the travel time of the signal to the receiver (Leick 1995). For now, we must allow the manual input of user world coordinates where a GPS solution is not available.

The user's horizontal look direction is determined by a digital magnetic compass which sends its output to the system computer. Like a standard magnetic compass, a digital compass measures the angle to a point relative to the user's magnetic meridian. The magnetic bearing is typically corrected to compensate for magnetic declination. The accuracy of digital compasses is comparable to other magnetic compasses and subject to the same types of error generated by local magnetic disturbances or improper handling.

The vertical look direction can be measured with an Abney level built to provide its data in digital format. At the time of this writing, the author was unable to find a commercially-available, digital Abney level. Since digital theodolites are commonly used to measure both horizontal and vertical angles, it is reasonable to assume that the much simpler Abney level could be produced in a digital format as well. For this project, we will assume that digital input is available but the prototype will require the user to input the vertical angle manually.

2) Determining lens parameters. The prototype system uses a video camcorder as the imaging device. Most standard consumer-grade video camcorders allow the user to change lens focal length and field of view in a single zoom operation. Unfortunately, none but the largest and most expensive video cameras provide digital information about these parameters to an external source. If we

know the number of video frames that occur between a fixed-speed zoom from the shortest to the longest focal length, we can make a very rough estimate of the current focal length as a function of time from one of the two extremes. The project prototype will not attempt to provide this type of zoom capability but will instead limit imaging to the known shortest and longest focal lengths for the project camcorder.

3) Building a perspective model. Once the user's position, horizontal look direction, and vertical look direction are established, we must create a viewing model for the DEM that matches these parameters and incorporates information about the current focal length and field of view of the video camera lens.

Elevations in a DEM are measured with respect to the surface of the ellipsoid specified by a given datum. Since the surface of the ellipsoid is curved, the height of distant objects will appear lower than close objects of the same height, beyond corrections for perspective scale reduction. Equation 1 is a standard correction applied to elevations read from a staff with a surveyor's level in a geodetic survey (Bannister, Raymond, and Baker 1992). It is used here to lower the reported height of a sample in the DEM as a function of its distance from the viewpoint. In Equation 1, z is the original elevation reported in the DEM in meters, d is distance between the viewpoint and the object in kilometers, and $z_{correct}$ is the adjusted elevation for the sample in meters. Additional error in the perceived height of an object due to atmospheric refraction are not taken into account in Equation 1.

$$Z_{correct} = Z - (.078d^2) \quad (1)$$

A perspective projection is created by first transforming 3D world coordinates (elevation values registered to a planar coordinate system such as UTM) to a 3D rectangular eye coordinate system in which the Z axis is collinear with the user's line of sight. Visual perspective is simulated by scaling eye coordinates with respect to their distance from the view point.

3D rendering libraries provide these transformations and allow the user to specify the focal length and field of view of the "camera" representing the user's point of view. By providing the lens parameters of the actual video camera to the rendering functions, the augmented scene can scale the perspective view of the DEM to the video image in the camcorder viewfinder.

4) Capturing a representative frame. For most applications, an augmented scene system needs to capture an individual frame representing the user's viewshed and overlay an interactive mouse pointer and cartographic symbols. The prototype uses a video capture chip on the system computer to read a video sig-

nal from the camcorder and load it into the frame buffer. An augmented scene only requires the capture of individual frames (video stills).

5) Registering the frame with the perspective model. Since the lens parameters of the video camcorder match those of the DEM perspective model, the captured image and the model should be registered if the user's position and view angles are correct. To ensure that overlain symbols will fit the surface of the captured image exactly, the image can be applied as a texture wrap over the DEM perspective model. Rendering libraries use texture wrapping functions to take a 2D image (the captured frame) and drape it over the surface of a 3D object (the perspective model).

6) Transmitting the computer display of the frame to the imaging device ocular. Once the image has been captured into the frame buffer of the system computer, the augmented scene software must transmit the image back to the video camera viewfinder along with its menu structure and the mouse pointer. The prototype accomplishes this by simply redirecting its video monitor output to the camcorder.

VIEWING AND QUERYING AN AUGMENTED SCENE

After the construction of an augmented scene, the user is presented with an image of the landscape which he or she can use as an index to geographic information. By moving a graphics pointer onto a part of the image and clicking, the user can ask questions like, "What am I looking at?" or "How far is that fire tower from here?" To support these queries, the system must 1) allow the user to select an object theme and an operation to act upon it; 2) reverse project the 2D image coordinates of a selected pixel to establish a vector in the world coordinate system parallel to the eye Z axis; 3) allow the user to select one of the set of surfaces intersecting the eye Z vector; 4) symbolize a feature at the selected location if it is within the user's viewshed, otherwise render the view along the vector as seen from the selected surface; and 5) pan and zoom on the current viewshed.

1) Selecting a theme and an operation to act upon it. The interactive operation of an augmented scene begins when the user selects a theme for querying and displaying. Thematic operations are selected by clicking on items in a menu system. A hiker may elect to activate a **Campsite** theme and select the **Display** operator that will symbolize all campsites in the viewshed. If he or she then selects the **Distance** operator, the straight-line distance from the current location to subsequently selected campsites (or the ground distance along a path) will appear in a text window.

2) Reverse projecting the 2D image coordinates. To make this type of selection work, a mouse click in the image must generate the 3D world coordinates associated with the selected pixel. Most 3D graphics rendering libraries provide functions to reverse project image coordinates to world coordinate space. However, a single pixel in a 2D scene maps to a 3D vector passing through the horizontal and vertical coordinates of the pixel and orthogonal to the plane of the viewing screen (Figure 2). This is similar to a view of a tree that we might see out of the window of a house. Branches that appear to cross one another are, of course, at different distances from the window but project, at the area of apparent overlap, to the same region of 2D coordinate pairs on the glass. If we want to identify one of the overlapping branches, we would simply indicate "the closer one" or "the further one."

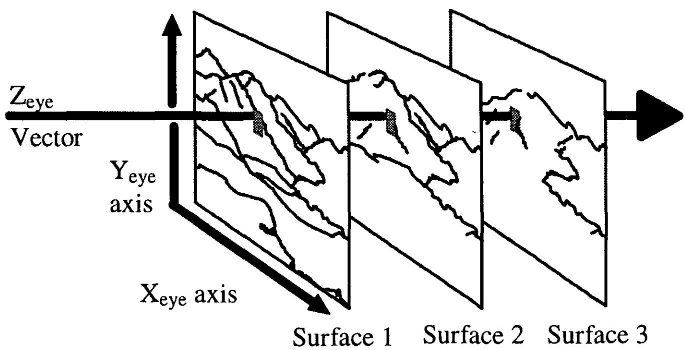


Figure 2. A selected pixel in 2D screen space represents an infinite number of points along the Z vector in eye space. The Z vector may intersect a number of surfaces in the perspective model.

3) Selecting a surface. An augmented scene must provide criteria for selecting which of the infinite points along the eye Z vector is to be referenced by a given pixel. The user may only want to see information on objects that are visible from the current viewing position. Alternatively, it may be important to see information on obscured objects. 3D rendering libraries handle these issues by allowing the user to select one of the set of objects "covered" by a given pixel. The hiker looking for a trail on a hill slope doesn't need (and doesn't care) to see a representation of the hill substructure. Yet the same hiker might ask, "What's on the other side of the hill?" or perhaps more precisely, "What is the view in this direction from the other side of the hill?" We encounter the same types of queries in urban environments. Will the user want to know the name and owner of a building or an elevator shaft inside of it?

The prototype uses the Microsoft DirectX graphics libraries for image modeling and rendering support. DirectX provides a library function that creates a list of surfaces, sorted by depth from the viewpoint, that project to a selected pixel. By default, the prototype models (and renders, if the system cannot capture a video image) only the surfaces in the user's current viewshed. If the user right-clicks on a pixel, the user's point of view will shift to a location on top of the next surface in the list at the selected pixel coordinates. The viewing parameters that were used to create the initial surface will be used again to render the new viewshed from the updated viewpoint.

4) Symbolizing features of selected themes. Symbols in an augmented scene refers to any graphic marks other than the background video frame or rendered surface. They highlight visible features captured in the current video frame or locate features on a rendered surface (for which no frame exists).

Selected objects are symbolized by opaque, 2D polygons defined as closed lists of vertices and faces. Predefined symbols will typically represent objects in the current theme though users will have the option of changing the default symbol sets. Rendering library functions overlay symbols on the video frame or rendered surface. Whenever a user deletes a symbol, the windowing system redraws any underlying graphics.

5) Panning and zooming on a surface. The user may elect to change the scale of the image (zoom) or move the image viewpoint (pan) to any location. The DEM will be rendered over any portions of the new view not covered by the current video frame.

DETERMINING THE EFFECTS OF POSITIONAL AND DIRECTIONAL ERRORS

Positional accuracy in an augmented scene is most critical for areas in the viewshed nearest the user. If the user wants to see an image of the pipes extending from immediately below his or her feet to the edge of the viewshed, positional error will be most visible in objects projected onto the scene nearest the viewpoint. As objects recede from the viewer, perspective scale reduction will continue to decrease the distances between their visible representation in the captured image and their projected locations in the perspective model.

In contrast to positional error, error in the look direction angles becomes more critical as distance increases away from the user. At a distance of 1 kilometer, a 1° error in either look direction will offset a feature approximately 17.5 meters in 3D world space from its true position. At a distance of 10 kilometers, the error will be approximately 175 meters.

DESIGN OF THE DATA STRUCTURE

The data for an augmented scene will consist of a DEM and a set of thematic coverages. To calculate the amount of data that will be required, we need to establish a sampling density and a maximal viewshed over which the system must operate.

Using Equation 1, we find that an object rising 500 meters above a viewpoint has a corrected elevation of 0 meters at a distance of approximately 80 km from the viewpoint. Does this mean that we would require DEM and thematic data over a radius of 80 km away from this point? Most users will probably show little interest in features that are close to disappearing below their horizon, yet some (like a driver on a highway) may want to know the name of the mountain range that just appeared above it. Assuming a 30 meter grid sampling interval, we would need approximately 22,000,000 DEM grid cells alone to cover a circle of this radius. A 15 meter interval will require 4 times this amount. Visible objects rising higher than 500 meters above the viewpoint will extend the viewshed boundary and the required DEM coverage even further. Although data sets of this size can easily be stored on current CD-ROM media, constructing a perspective model for a large viewshed may require prohibitive amounts of computing time.

To solve this problem, an augmented scene must either limit its working range or build perspective models by resampling data at increasing intervals with distance from the viewpoint. In a perspective view, the ground area covered by a unit area of the viewing screen increases with the distance of the ground from the screen. Pixels representing areas at large distances from the screen can cover many cells in the database. Therefore, information about distant features can be resampled at a greater interval than the default sampling interval of the DEM. Resampling can be handled as a continuous function of distance from the viewpoint or by establishing a single threshold distance between two sampling interval regions.

Most rendering libraries use the concept of clipping planes (parallel to the viewing screen) to identify visible regions within a viewing pyramid extending from the user's eye. Only the region in the viewing pyramid between the front clipping plane (near the viewer's eye) and the back clipping plane (far from the viewer's eye) is rendered. Instead of using the back plane for clipping, the system uses its location to separate regions of higher and lower resolution sampling. The default location is based on the user's preference with regard to system processing speed and memory capacity. The user can move the default location of the back clipping plane to handle specific imaging requirements.

BUILDING THE PROTOTYPE SYSTEM

The prototype for this project uses an IBM ThinkPad 755CD as the system computer. The 755CD uses an on-board video capture chip to receive input from a Sony CCD-TRV70 camcorder. The camcorder also receives video output from the 755CD to see the captured image superimposed with menu and pointer data. A Silva GPS Compass provides both positional and horizontal look direction data. Data from the GPS Compass is transmitted to the 755CD over through a standard serial communication port. The system software is composed of a main program, Urhere, written by the author in C++ for the Microsoft Windows 95 operating system.

CONCLUSION

The intent of this paper has been to define augmented scenes, suggest some applications for them, and to propose a model for their construction and use. The most important topics for future research includes:

- reducing positional and directional error;
- reducing the size of data sets;
- incorporating non-perspective (plan) views;
- determining a suitable data distribution format; and
- refining the user interface through subject testing.

It is expected that technological developments in portable computing and position finding devices will greatly increase the feasibility of augmented scenes in the near future.

REFERENCES

Bannister, A., S. Raymond, and R. Baker (1992). *Surveying, 6th ed.*. Longman Scientific and Technical, Essex, England, pp. 76-79.

Drasic, D. and P. Milgram (1991). Positioning accuracy of virtual stereographic pointer in a real stereoscopic video world. In: *SPIE Stereoscopic Displays and Applications II*. Vol. 1457, pp. 302-312.

Horn, B. and B. Bachman (1978). Using synthetic images to register real images with surface models. In: *Communications of the ACM*. Vol. 21 No. 11, pp. 915-924.

Leick A. (1995). *GPS Satellite Surveying, 2nd ed.* John Wiley & Sons, New York, pp. 311-313.