

Data Retrieval Through Virtual Experimentation

Andrew Johnson and Farshad Fotouhi

ABSTRACT:

In this paper we describe a new interface to scientific databases, the SANDBOX: Scientists Accessing Necessary Data Based On eXperimentation. The SANDBOX is a virtual reality tool allowing an investigator to visualize the contents of a scientific database while retrieving data. As the data in these databases was typically collected through experimentation, an investigator can use the SANDBOX to retrieve data from the database by placing virtual instruments into a virtual reenactment of the original experiment. These instruments give visual and auditory feedback, allowing the user to browse through the data, setting up and running experiments until they have collected the data they need. We have implemented a prototype of the SANDBOX on NASA's FIFE scientific database using the CAVE virtual reality theatre.

KEYWORDS:

Information retrieval, browsing, virtual reality, scientific databases, visualization, sonification

1 Introduction

Scientific databases store large amounts of information typically collected through experimentation. This information is then made available to investigators from a wide range of disciplines, mostly unfamiliar with databases and their associated query languages. These investigators need to search for specific pieces of data quickly; they need to browse through related information to see if it is of value; they need to relate information from various parts of the database.

Since the information was originally collected through experimentation, we can use virtual reality to recreate the original experiments. This allows an investigator to collect data from the scientific database in much the same way that the original data was collected. The investigator places virtual instruments into a virtual environment and collects data from the scientific database without ever typing in a query.

Each investigator uses familiar measuring instruments and collects data in a familiar way. As the environment is virtual the investigator can repeat experiments multiple times using different combinations of instruments to see how different values relate. The instruments allow the investigator to visualize the contents of the database, and browse through those data values, before retrieving information from the database. We call this interface the SANDBOX: Scientists Accessing Necessary Data Based On eXperimentation.

The SANDBOX allows the user to become immersed in the database. The user is surrounded by the data values themselves visualized by the instruments. The SANDBOX gives the investigator an environment to work in, rather than just a screen to look at.

Section 2 of this paper describes the general framework of the SANDBOX. Section 3 presents the details of a specific application of the SANDBOX to a scientific database. Sections 4 discuss our initial evaluation of this interface and give some user reaction to it. Finally, section 5 gives our conclusions and future directions.

2 SANDBOX

The SANDBOX [4, 5, 6] gives the user a virtual laboratory that can be configured for different experiments on different scientific databases by loading in different sets of instruments, and environments. The exact instruments, and the way space and time are modeled, will depend on the environment being recreated.

In the SANDBOX, the user interacts with the scientific database by running experiments in the virtual

environment. When the user places virtual instruments into the virtual environment, they react as the real instruments would, giving the user feedback. The investigator can use this feedback to add additional instruments to the experiment, move the instruments to other locations, or remove unnecessary instruments. Once the user has placed the appropriate instruments into the environment and set the appropriate time interval, the information is retrieved from the database and stored in an external file for further use. The SANDBOX rewards the investigator's familiarity with the environment, not with artificial database concepts.

Scientific databases often store information in multiple ways: files, relational databases, object-oriented databases. The database can contain numeric data, textual data, and graphical data. The database may also include meta-data (notes, drawings, diagrams, maps, photographs, sounds, etc.) This data can be integrated into the virtual reality interface, giving the user additional information of any kind. As the user recreates the experiments the appropriate information is retrieved from the appropriate source, allowing the user to simultaneously see and relate multiple types of information from multiple sources.

2.1 Instruments

As the user recreates the experiments with the virtual instruments, the SANDBOX retrieves the appropriate information from the appropriate source. Some instruments are linked to numeric and textual experimental data; some to graphical, audio, and other experimental data; some to meta-data.

Some generic instruments include:

- fillers - fill/empty according to values
- expanders - expand/contract according to values
- pointers - point in a specific direction
- brighteners - brighten/darken according to values
- enumerators - more/fewer according to values
- colourers - change colour according to values
- spinners - spin fast/slow according to values
- densers - density varies according to values
- texters - display alphanumeric characters
- listeners - produce audio
- viewers - display graphical images
- markers - mark a location the user wants to remember

Each scientific database requires its own set of instruments. The more familiar an instrument is to the user, the more obvious is its function. This may mean that different users of the same scientific database will use different instruments to access the same data. For example, Figures [1](#) and [2](#) show census data from the Great Lakes states using expanders and enumerators. With the expanders, the height of the column rising out of each state reflects that state's population. With the enumerators, each person-counter represents 1 million residents.

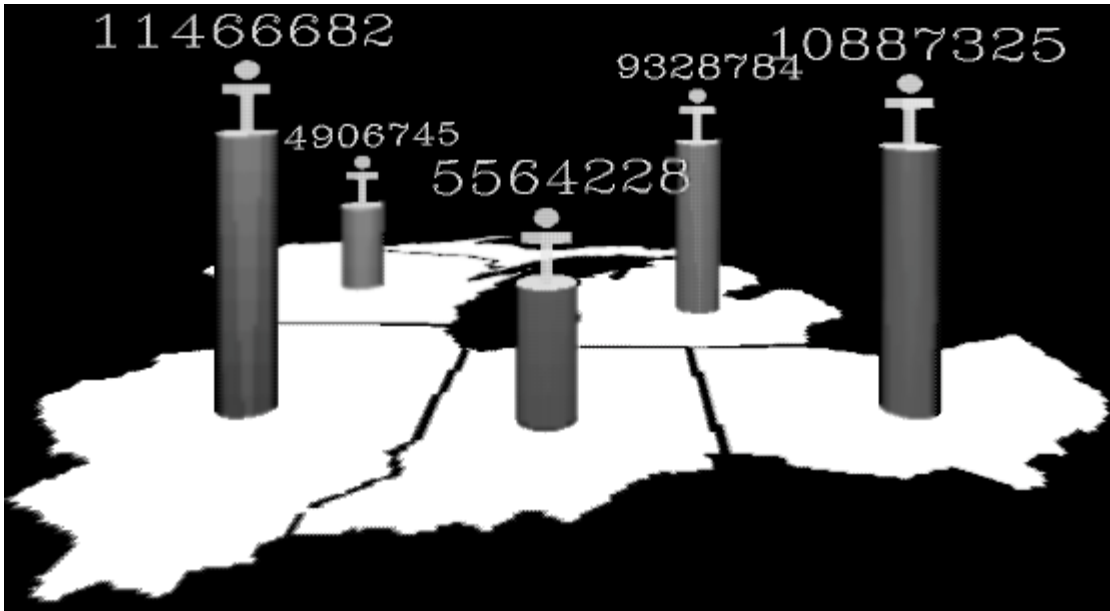


Figure 1: Census Data With Expanders



Figure 2: Census Data With Enumerators

As the instruments are virtual, they can be virtual representations of actual instruments. They can also visualize and retrieve information from a combination of actual instruments, or display statistical information derived from actual instruments.

The actual data values stored in the database were collected from specific locations or sites during the actual experiment. In the SANDBOX, a virtual instrument can be placed at any site where the information it measures was collected during the actual experiment. Each virtual instrument placed at a site, maintains its own data. This includes the data currently being visualized by this instrument (e.g. a thermometer instrument stores temperatures for several days at a single site) and the current visualization settings on this instrument (e.g. the minimum and maximum temperatures to be displayed.)

Each instrument is linked to the scientific database. This linkage is a combination of an instrument class, an access function, and a set of filter functions.

The instrument classes convert data into visual and auditory form, allowing the user to see and/or hear the values. Each instrument is a member of some instrument class (e.g. each individual thermometer is a member of the thermometer instrument class which is a subclass of the generic fillers class.) The class maintains static information common to all instruments of that class. Each instrument class has a G-function to represent the instrument and its current value(s) in terms of graphics primitives such as cubes, cylinders, spheres, and pyramids. Each instrument class has an A-function to represent the instrument and its current value(s) in terms of audio primitives such as external sound files, or functions which synthesize sounds.

The access functions retrieve data from the database. An access function takes a site identifier, and a time range and retrieves information in the form {time, data values} from a specified part of the scientific database. This access function could retrieve data from a tabular file, a file containing a single large data value (e.g. a satellite photograph), a table in a relational database, or an object in an object-oriented database. Time in the scientific database may have many formats and each experiment may have a unique rate of collection. Each of these times is converted to a common time index to decrease access time. Time in the SANDBOX is taken as absolute time from the beginning of the actual experiments with the granularity of time dependent on the individual experiments.

The filter functions perform transformations on the data. The SANDBOX uses two types of filter functions: access filters and display filters. Access filters filter the data values as they are retrieved from the scientific database before they are placed into the local memory. They are used to ensure all the data stored in the local memory has the same format (e.g. all temperatures are in degrees Celsius), and to perform time consuming operations on large data sets (e.g. cropping, rotating, and enhancing a satellite photograph.) Display filters filter individual data values before they are displayed by a virtual instrument. They give the users control over the settings on the virtual instruments (e.g. setting the minimum temperature a thermometer can display.)

Each scientific database requires its own set of instrument classes, access functions and filters. These can be stored in a library and used again in future interfaces for databases in similar fields of study.

2.2 System Overview

The operation of the SANDBOX is centered around visualizing the appropriate data for the current frame to be displayed. For each frame the sequence is shown below:

```
repeat
    User input (movement and/or actions)
    Data retrieval for the current virtual time
    Visualization and virtual environment update
until user exits
```

An overview of the system is shown in Figure 3. The SANDBOX is composed of three main components: the Virtual Reality (VR) Interface, the Preprocessor, and Local Memory.

The VR Interface is responsible for the maintaining the virtual environment, representing the virtual instruments visually and audibly, as well as monitoring the user's actions within the virtual environment. The primary component of the VR Interface is the Environment Manager. The Environment Manager manages the virtual environment, keeping track of the user's position and actions. Based on these actions the Environment Manager sends requests for data to the Data Manager in the Preprocessor. It also uses Display Data from each of the instruments in Local Memory to display images and sounds in the AudioVisual Environment.

The Preprocessor is responsible for interfacing with the various components of the scientific database to quickly retrieve data according the needs of the VR Interface and store it in Local Memory. The primary component of the Preprocessor is the Data Manager. The Data Manager receives requests from the Environment Manager and sends requests to the appropriate database access module to retrieve that data. This data is then moved to the appropriate place in the Local Memory.

The Local Memory maintains all the information necessary to support the VR Interface. This includes information on the user, the various tracking devices, and instruments that the user has placed in the virtual environment.

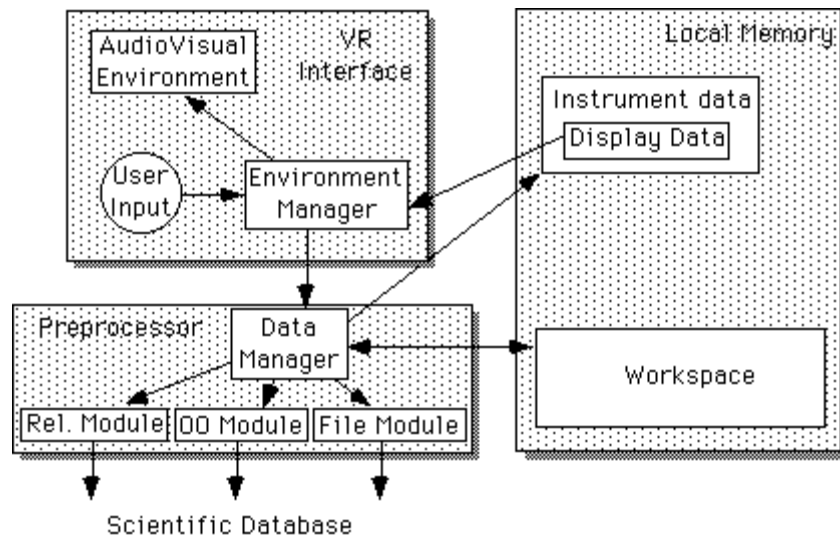


Figure 3: Overview of the SANDBOX

2.3 Data Storage

Current database access methods are not fast enough to support the needs of virtual reality. For smooth movement of a 3D image at least 15 frames per second must be generated for each eye. Generating a frame involves accessing all of the relevant information, converting it to graphical form, and then drawing the scene.

Typically in scientific visualization all of the relevant data is brought into RAM before the visualization begins. That is clearly not possible here given the huge amounts of data involved. Since we are using visualization while the data is being retrieved, not just afterward, the entire database needs to be accessible.

Data storage in the SANDBOX is therefore hierarchically organized into four levels:

- Display data - The data to be converted into visual and auditory form for this frame. Each instrument placed at a site maintains its own display data.
- Instrument data - The data currently being indexed by a single instrument - including all of the data for this instrument at this site for the currently selected time interval. For each frame, a subset of this data becomes the Display data. The Instrument data also includes the display filters and the settings on the display filters.
- Workspace - A cache for the Database data. It holds data that is not currently Instrument data, but might become Instrument data in the near future.
- Database data - The scientific database itself. The database may be local or remote, and is probably a heterogeneous environment.

As the user chooses instruments, times, and other experimental parameters, the VR interface passes this information along to the preprocessor. In the SANDBOX `where' is determined by the choice of site, `when' is determined by the choice of time, and `what' is determined by the choice of instrument. The preprocessor determines which parts of the database are likely to be accessed in the near future. Relational tables can be partitioned vertically and horizontally, objects can be isolated, and files can be marked. These blocks of information can then be moved into local memory before they are needed by the visualization system.

The SANDBOX requires rapid access to the data values in memory based on time. Given a time index, the data structure must be able to quickly find the nearest stored time and its associated data values. In order to achieve fast access to the instrument data values we are using a modified version of the T-tree data structure proposed by Lehman and Carey [7]. Each instrument placed at a site keeps its own T-tree of instrument data ordered by time.

3 Implementation

We have implemented a prototype of the SANDBOX using the CAVE [2, 3] virtual reality theatre at the Electronic Visualization Lab at the University of Illinois at Chicago as shown in Figure 4. The CAVE is a projection based virtual reality system. The user enters a 10 foot by 10 foot by 10 foot room where images are projected onto the three of the walls and the floor. When the user dons a pair of lightweight StereoView LCD shutter glasses, the projected images fill the room and surround the user. The user can move around the room reasonably unencumbered, and can walk around or through the virtual objects in the CAVE. Since they can see their own bodies, users have a true sense of being inside the virtual environment. The user carries a physical three button wand to interact with the virtual objects in the CAVE.

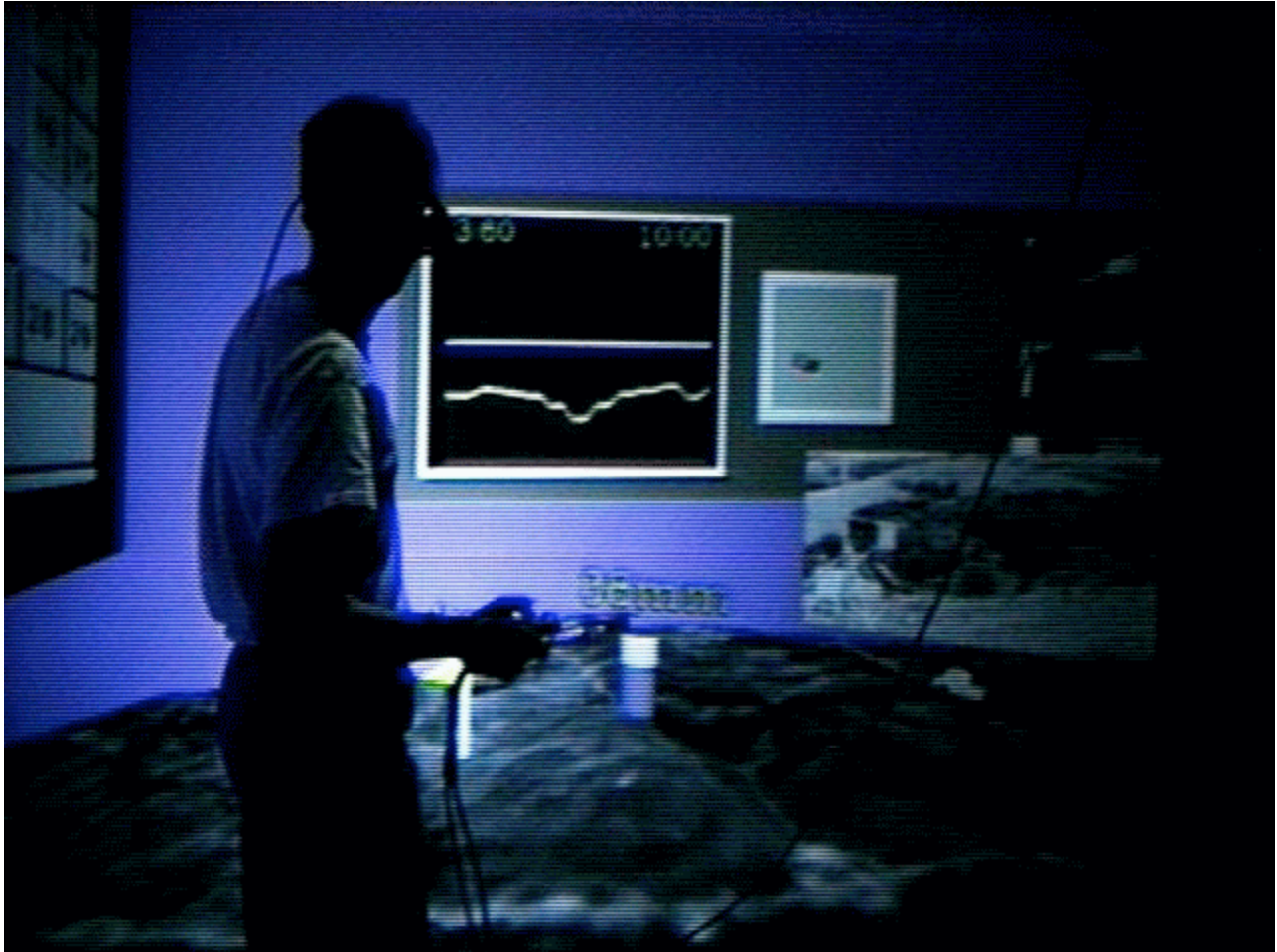


Figure 4: Photograph of SANDBOX in the CAVE

We implemented this prototype on a subset of the FIFE scientific database. Developing techniques to determine surface climatology from satellite observations is the objective of the ISLSCP (International Satellite Land Surface Climatology Project.) FIFE (First ISLSCP Field Experiment) was undertaken at a 20km by 20km square site near Manhattan, Kansas in 1987 and 1989. Its purpose was to gather enough data to allow the creation and testing of models to develop these techniques [13].

120 gigabytes of data was collected (300 megabytes textual data, the rest image data including satellite photos, and site photographs taken on the ground.) The textual data fills over 100 tables in a relational database. Each experiment is given its own table with the attributes containing the numeric data collected during that experiment. These tables typically have 10-30 columns each, and over 100,000 rows.

In this reenactment of the FIFE field experiments, the user is surrounded by an elevated 3D plane showing the 20km by 20km square experiment site, a pallet of instruments to choose from on the right wall, and a calendar to choose dates from on the left wall. The 3D plane initially shows an enhanced satellite view of the experiment site. See Figure 5.

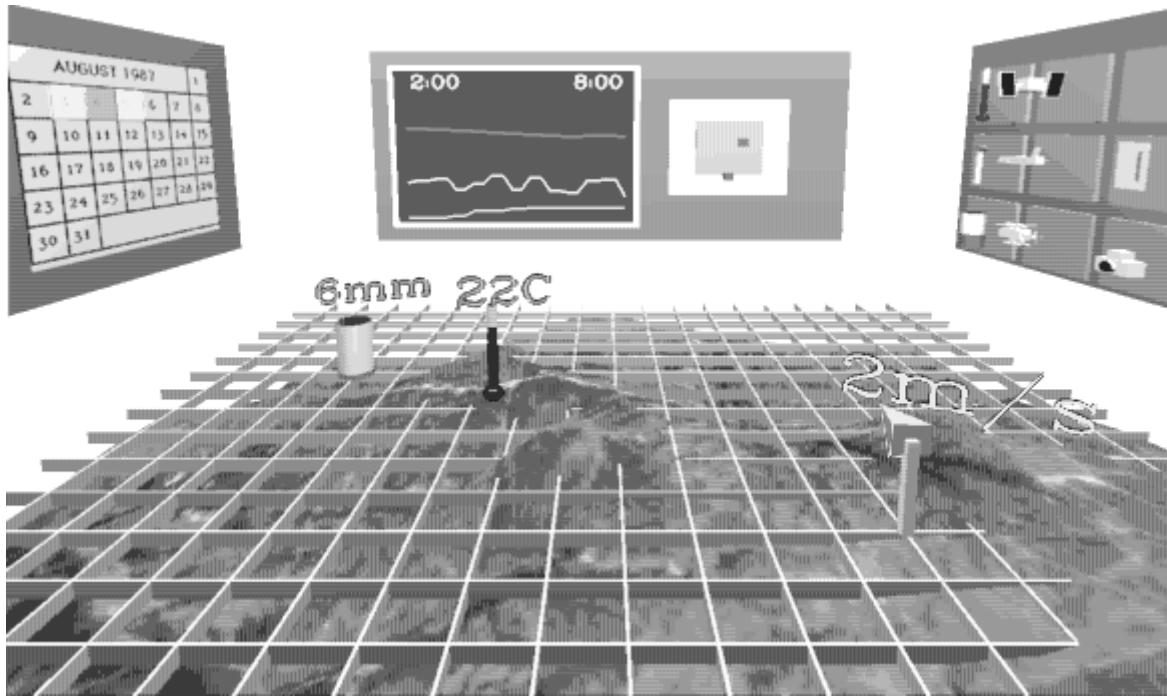


Figure 5: Overview of the Virtual Environment

3.1 Instruments

Figure 6 shows a close-up view of the instrument pallet. Some of the instruments (thermometer, wind-sock, and water beaker in the left hand column) are linked to columns in the relational database. Some (LANDSAT satellite, airplane, helicopter in the center column) are linked to graphics files. Some (notepad, camera in the right hand column) are linked to meta-data. The instruments on the pallet are 3D and animated (e.g. the beaker fills with water) to improve their recognizability. The instruments have obvious affordances [9].

In a large scientific database there would be many possible instruments. A second, much larger instrument pallet would be needed to hold all of the possible instruments. The user would be able to scroll through this pallet to the appropriate instruments, and then carry instruments over to the smaller pallet. This pallet could also contain additional meta-information about the instruments themselves, helping the user choose the appropriate instrument.

The user can literally pick an instrument off the instrument pallet, carry it over to a site, and place it there. The user directly manipulates the objects in the virtual environment [10]. There is a continuous visual representation of the instruments on the pallet, placed at sites, and in the user's hand. The user physically places instruments instead of typing in queries and can immediately see the result of an action. This allows a novice user to quickly begin using the system to run experiments.

While the amount of data in the FIFE database is very large, the number of sites for each experiment is typically small (usually around 10 to 15 sites), giving sparse coverage of the total experiment area. Because of this, conventional visualization techniques can not be employed (i.e. there is not enough information to draw meaningful colour mapped surfaces over the 3D plane.)



Figure 6: The Instrument Pallet

Once the instrument is placed at a site, it begins to operate. The mercury level in the thermometer rises and falls with the temperature. The water level in the beaker rises and falls with the rainfall. The orientation of the wind-sock changes with the direction, and speed of the wind. The user can see how different measurements relate (e.g. the mercury level in a thermometer begins dropping as a nearby water beaker begins filling with water.)

The user can also hear the instruments. A beaker makes a 'drip' sound when its water level rises. The faster the water level is rising, the louder the 'drip.' The average strength of the wind measured by all the wind-socks is used to generate a 'whoosh' sound - the strength of the wind determines the pitch of the 'whoosh.' The average temperature measured by all the thermometers is used to generate a 'cicada' sound - the higher the temperature, the louder the 'cicada.' An instrument draws attention to itself when there is a change in the value it is monitoring.

If the user needs to see a record of how the measured values change over time, the values can be displayed in a graph on the graph wall. This allows the user to see trends that may not be apparent from the individual data values. If the user requires quantitative numeric, as well as qualitative graphical values, they can be displayed above each instrument. These quantitative values rotate horizontally and vertically to follow the user, so they are always readable wherever the user is standing in the CAVE.

The user views textual meta-data (e.g. site information, notes) with the notepad, and graphical meta-data (e.g. photographs taken at a site) with the camera. A user wishing to see information about a site can place the notepad at a site. A page with the text then appears above it. A user wishing to see a photograph taken at a site can place the camera at a site. The picture of that site then appears above it. The meta-data is integrated with the numeric information and the graphic information. In the actual scientific database the user would have to integrate this information manually.

Figure 7 shows four instruments placed in the virtual environment: a beaker measuring rainfall, a wind-sock measuring wind speed and direction, a thermometer measuring temperature, and a camera displaying a photograph taken of a site. The beaker, wind-sock, and thermometer have their current values displayed graphically, audibly, and numerically. Their values over the last six hours are shown in the graph. The user is simultaneously seeing numeric data from the database visualized by the instruments, graphical data showing the features of the 3D plane, and meta-data showing an actual site photograph.

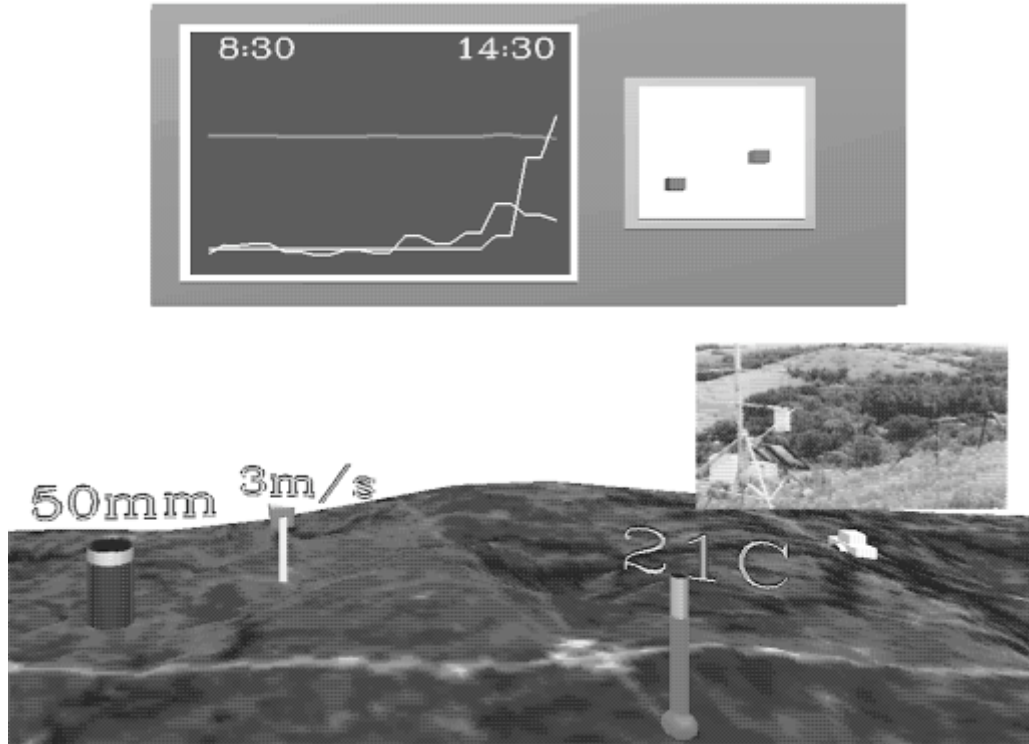


Figure 7: The Instruments Giving Feedback

The user can change the settings on a virtual instruments using the menu shown in Figure 8. The user can set the minimum and maximum values displayed by the instrument. The user can turn on or off whether the instrument displays quantitative values overhead, whether its values are shown on the graph, and whether the instrument makes sound. Each instrument maintains its own individual settings.

The interface employs an object-oriented paradigm. Existing virtual reality systems [11] allow for the creation of sliders, menus, and button panels. These are simply transferences of 2D systems into the 3D virtual environment where operations on objects are performed using a separate detached interface control panel. In our approach each object in the virtual environment maintains its own state, and has its own interface control panel. This reduces user errors during interaction by giving users access only to those functions which are appropriate to the object [9].

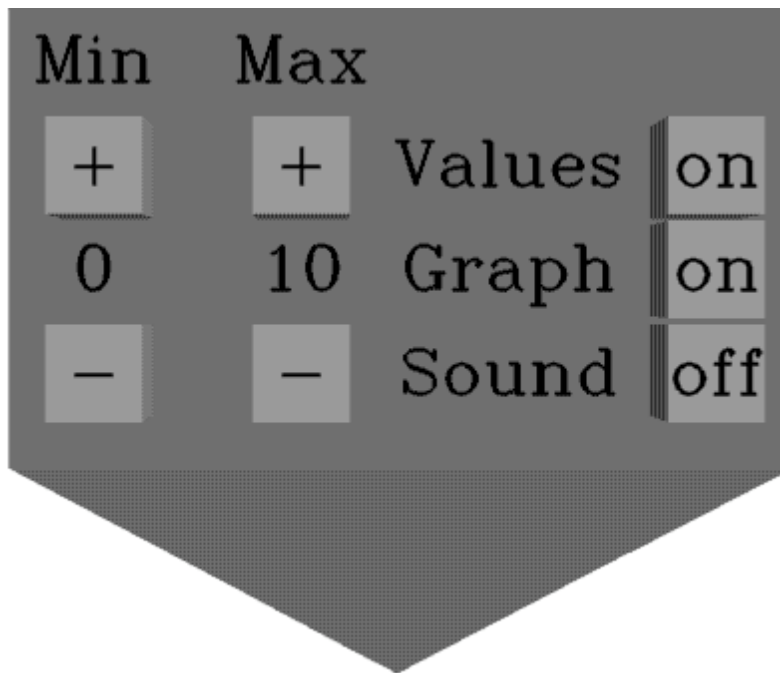


Figure 8: Menu for an Individual Instrument

3.2 Environment

The user can alter how the 3D plane is displayed. The user can enlarge or shrink the 3D plane depending on whether the user wants an overview of the entire experiment, or a close-up view of a certain area. The user can place the 3D plane on the floor to look down at it from high above, or at waist height to get a better sense of the topography, or above the user's head to look at the terrain from below.

When the user grabs the satellite instrument and places it in the sky above the 3D plane, the user can choose which LANDSAT band to view the landscape in. LANDSAT photographs from the database are mapped onto the 3D plane. In the actual scientific database the user must refer to a site using its site ID number. In the SANDBOX the user can see where the sites are located. The user can immediately see if a site is on top of a hill, or in a valley, near a river, or near a road. The graphical information is integrated with the numeric information. In the actual scientific database the user would have to integrate this information manually.

The user selects days from the calendar by clicking on them with the wand. The current virtual day cycles through the selected days. A virtual sun orbits the 3D plane rising out of the east each morning at 6am, and setting in the west at 6pm. As the sun dips below the western horizon, the moon rises in the east and continues to orbit the 3D plane opposite the sun. While the Earth's moon does not exhibit this behavior, it gives the user better control over their virtual environment. By grabbing the sun or moon with the wand, the user can adjust the behavior of time, speeding it up, slowing it down, or stopping it.

The sky changes color as the virtual day progresses, changing from black at midnight, to purple at dawn, to bright blue at noon, to purple again at dusk, and back to black at midnight. This gives the user additional temporal feedback in a form natural to the environment.

3.3 Viewing External Information

All of the information that an investigator needs to run their experiments may not be in the scientific database. Since the user is not isolated from the real world while interacting with the virtual world inside the CAVE, the user can bring items into the virtual environment. The user may want to bring in a list of experiments to run, or a map of sites to examine.

Unfortunately the CAVE must be kept dark, making it difficult to see anything brought inside. To alleviate this

problem, the user can turn on a virtual spotlight in the SANDBOX. This spotlight illuminates the area surrounding the user and follows the user's movements around the CAVE.

3D Desktop Version

This method of extracting data through recreation of experiments, and the integration of multiple information types does not rely on virtual reality. While virtual reality provides more of a direct manipulation immersive environment, these same techniques could be applied to a more common, and less costly, 2D or 3D desktop environment.

The SANDBOX code could also be adapted to a Head Mounted Display or a BOOM. Both of these hardware platforms isolate the user from the real world making it impossible to see external information while interacting with the virtual environment. A "fish tank" desktop system [1] would probably be the best alternative for those wanting a currently affordable hardware platform.

A typical desktop version includes the following hardware: an SGI Extreme and StereoGraphics-ready monitor, a pair of StereoGraphics' CrystalEyes glasses with infra-red emitter, a head tracker, and a Logitech 6D mouse with ultrasonic emitter. The monitor and glasses allow the user to see the SANDBOX in 3D. The head tracker allows the scene to shift appropriately as the user's head moves increasing the illusion of 3D. The 6D mouse replaces the wand in the CAVE allowing the user to interact with objects in the 3D environment. In effect, this hardware set up brings one of the CAVE walls to your desktop as shown in Figure 9.



Figure 9: Desktop Version of the SANDBOX

The CAVE and the desktop version each have their advantages and disadvantages.

The CAVE is immersive; the desktop version is not. The CAVE user can turn and look around the environment; the desktop user must look at the screen. There is a more natural mapping between the user's actions and their effects in the CAVE than in the desktop version.

Both the CAVE and the desktop version can generate sounds that appear to emanate from a particular point in space. The CAVE however has a more natural mapping between the visual space and the audio space. If a sound is emanating from an instrument on your right, you can turn to the right and see the object generating that sound. With the desktop system all of the visual cues are in front of you on the screen. Sounds can still come from behind you, but you must turn the virtual environment to see the object generating it.

The desktop version is overwhelmingly cheaper than the CAVE. A desktop version of the system capable of running the prototype would cost about twenty times less than the CAVE. There is also a range of possible desktop hardware allowing the user to pay for only the performance needed. The desktop version is more mobile, less prone to hardware failure, and built from more common hardware components.

The CAVE and desktop users have similar encumbrances. The glasses are identical, and the wand is very similar to the 6D mouse. The desktop user does have the advantage of sitting down while using the interface which could be important during long experiments.

The CAVE user can bring items into the CAVE but may have trouble seeing them. The desktop user has easy access to external information through books, notes, the telephone, and other computer programs. If most of the information that the investigator needs to work with is in the database then the CAVE gives a much richer environment to interact with that information. If the investigator needs access to a lot of external supplementary material, then the desktop version gives the user more restricted access to the virtual environment without limiting access to the supplementary material.

4 Evaluation

In this section we will use several examples to illustrate how the SANDBOX can be a more effective data retrieval tool than traditional access methods in certain situations. It is difficult to do direct comparisons because traditional access methods are designed to retrieving data for specific queries, where the SANDBOX is designed to provide an environment where the user can browse through related data.

Traditional access methods favor those who are familiar with the structure and content of the database. They are appropriate for those who know what data they want, where it is in the database, and how to retrieve it. They are appropriate for those who need to retrieve data based on complicated combinations of information. The SANDBOX favors those who are familiar with the original experiments. It is appropriate for those who are not sure what data they are interested in and need to browse through the data to find appropriate data to retrieve. It is appropriate for those who need to integrate multiple types of data. This means there are certain situations where the SANDBOX will be most appropriate and certain situations where a query language will be most appropriate.

Data retrieval is only one aspect of the scientific analysis process [8, 12]. The SANDBOX integrates data retrieval with some visualization capability and represents them using familiar metaphors, reducing cognitive load and enhancing the scientific analysis process.

Investigators using a traditional interface must first take their understanding of the experiments and convert it into an understanding of the database structure (or schema.) As the schema is designed to facilitate efficient storage of information it may be very unrelated to the way users naturally think of the data. Then the investigators need to learn the query language to retrieve the data they require. Then, once they have the required data, they need to visualize that data in a way that is meaningful to them.

Investigators using the SANDBOX have several advantages. The database schema is hidden from the user, while still giving the user access to the data. They deal with familiar concepts and instruments in a more natural environment. They can perform visualization while retrieving data.

The FIFE database uses SQL as its query language. For these tests we used an Oracle database on a Macintosh Quadra 700. This was compared against the SANDBOX environment described previously. Users taking part

in the SQL tests were familiar with databases in general and with SQL in particular, though they were not everyday users. This avoided the training time that investigators less familiar with databases would incur.

Users of both interfaces were given an introduction to the FIFE experiment in general, and an overview of the interface. They were told which tables were important in the SQL interface and what their columns contained, and how instruments were placed and dates selected in the SANDBOX

The SQL queries were performed twice. The first time, the users were given only the information discussed above. The second time took place several days later, to see how fast the users could repeat the query.

4.1 Example 1: Visualizing Numeric Data

Let us assume an investigator is interested in knowing what directions the wind was blowing on August 28th, 1987. With the existing interface the user must go through the following sequence of actions:

1. Finds AMS_DATA_87 table contains wind data
2. Finds appropriate columns in AMS_DATA_87 table
3. Queries database for list of all sites in AMS_DATA_87 table on August 28th, 1987 returning a list of 10 sites
4. Chooses site (e.g. site 3)
5. Queries database to extract wind information from site 3 on August 28th, 1987 from AMS_DATA_87 table

With the SANDBOX the user goes through the following sequence of actions:

1. Clicks on August 28th from the calendar
2. Grabs wind-sock instrument. All of the sites where wind measurements were taken are displayed
3. Places wind-sock at one of the sites

Looking at the resulting file, the user of the query language will find information in the form of x, y, and z components of the wind speed (e.g. 1.23, 4.27, 0.85.) The user must now compute the actual wind direction from these values.

Placing a wind-sock at site 3, the user of the SANDBOX would quickly see the direction that the wind was blowing and how that direction changes throughout the day.

User testing showed users familiar with SQL took an average of 5 minutes to get their results the first time, and 2 minutes the second time. Novice SANDBOX users took only 26 seconds on average.

4.2 Example 2: Integrating Numeric Data and the Environment

Let us assume an investigator is interested in measuring rainfall at the highest point on August 4th, 1987. With the existing interface the user must go through the following sequence of actions:

1. Finds AMS_DATA_87 table contains rainfall data
2. Finds appropriate columns in AMS_DATA_87 table
3. Queries database for list of all sites in AMS_DATA_87 table on August 4th, 1987
4. Finds FIFE_SITE_REF table relates site numbers to elevations
5. Finds appropriate columns in FIFE_SITE_REF table
6. Queries database to find elevations of all the sites in the AMS_DATA_87 table
7. Chooses site 11
8. Queries database to extract rainfall information from site 11 on August 16th, 1987 from AMS_DATA_87 table

With the SANDBOX the user goes through the following sequence of actions:

1. Clicks on August 4th from the calendar
2. Grabs beaker instrument. All of the sites where rainfall measurements were taken are displayed

3. Looks at the 3D plane to see where the highest site is
4. Places beaker at the highest site

The SANDBOX lets the user to deal with familiar instruments such as beakers and familiar concepts such as elevation as opposed to artificial concepts like tables and site IDs.

User testing showed users familiar with SQL took an average of 16 minutes to get their results the first time, and 7 minutes the second time. Novice SANDBOX users took only 18 seconds on average.

4.3 Example 3: Integrating Numeric Data and Graphical Data

Let us assume an investigator is interested in comparing ground temperatures to the infra-red LANDSAT photograph on August 16th, 1987 to see how they correlate. With the existing interface the user must go through the following sequence of actions:

1. Finds AMS_DATA_87 table contains temperature data
2. Finds appropriate columns in AMS_DATA_87 table
3. Queries database for list of all sites in AMS_DATA_87 table on August 16th, 1987 returning a list of 10 sites
4. Finds FIFE_SITE_REF table relates site numbers to site locations
5. Finds appropriate columns in FIFE_SITE_REF table
6. Queries database to find site locations of all the sites in AMS_DATA_87 table
7. Finds IR LANDSAT photo for August 1987
8. Finds site locations on IR LANDSAT photo
9. Chooses sites 3, 7, 25, and 29
10. Queries database to extract temperature information from chosen sites on August 16th, 1987 from AMS_DATA_87 table

With the SANDBOX the user goes through the following sequence of actions:

1. Clicks on August 16th from the calendar
2. Grabs satellite instrument and places it in the sky
3. Clicks on satellite to activate its menu and selects the IR band. The 3D plane now displays infra-red landscape
4. Grabs thermometer instrument. All of the sites where temperature measurements were taken are displayed
5. Places thermometer at appropriate sites

Looking at the resulting file, the user of the query language will find that only site 3 gave measurements throughout the day. The user must now go through the same sequence of actions again to re-query the database for information from other sites.

Placing thermometers at sites 7, 25, and 29, the user of the SANDBOX would immediately see that those instruments were not giving any feedback. The user could then move those thermometers to other nearby sites where the instruments give feedback. The SANDBOX allows the user to easily browse through the data itself, not just the structure that has been imposed on the data.

4.4 User Comments

Users of the SQL interface were frequently frustrated by that interface. They needed scrap paper to remember table names and attribute names, and even then mistyped names and commands were common. They were constantly scrolling back through their previous queries to see what they had done already, and trying to get an overview of the data they were dealing with. Mistakes often meant screenfulls of useless data scrolling by or terse error messages. The interface was described as 'annoying.'

On the other hand, the SANDBOX seemed very natural to the users and they were able to retrieve the necessary data much quicker, and easier in these examples.

4.5 Reaction

Our initial testing in the CAVE suggests that users find this paradigm to be very natural. Picking up and placing instruments appears to be very easy and intuitive. We've observed users bending down to place their instruments onto the 3D plane, even when they can just stand over the site and push the button on the wand. The environment and the instruments are realistic enough that the users are treating them as real objects.

Unfortunately this 'realism' can be physically tiring. Users must walk over to the instrument pallet to choose instruments, must walk over to the calendar to choose dates. Allowing the user to choose instruments or dates from anywhere in the CAVE should further reduce this fatigue. Some users have found the calendar and instrument pallet to be uncomfortably high, others found the plane to be uncomfortably low. This suggests that the virtual environment should be more flexible.

It is also currently very repetitive to place an instrument at multiple sites. Future versions should allow a user to select multiple sites simultaneously. While less 'realistic', this would be more efficient.

A couple users have objected to the moon orbiting the 3D plane opposite the sun, as the Earth's moon does not do this. An alternative may be to use the same direct manipulation technique on the hands of a virtual analog clock. The rotating hands of the clock would then serve the same function as the rotating sun and moon.

In the actual experiment scientists could place clusters of instruments simultaneously. The user should have the ability to combining existing instruments into instrument packages. These packages would then be available on the instrument pallet. Each of the individual instruments would function independently but they could be placed and removed as a single package.

5 Conclusions and Future Directions

Using virtual reality the SANDBOX allows the user to bypass the current structuring of the data and access the contents of a scientific database by recreating the original experiments. The user deals with familiar concepts in a more natural environment and interacts with the data directly. The user can browse through the data, and integrate different types of data.

We are currently working on enhancing the SANDBOX in the following ways: 1. Increasing the visualization capabilities, 2. Allowing the user to make annotations to the database and access them in a convenient manner, 3. Increasing the capabilities of the instruments to allow for more aggregate functions, and 4. Applying the concept of instrument-based retrieval to other types of databases.

Acknowledgments

We would like to thank Jason Leigh, Thomas DeFanti and the other members of the Electronic Visualization Lab at the University of Illinois at Chicago for their assistance with the CAVE.

References

- 1 K. Arthur and K. Booth. Evaluating 3D task performance for fish tank virtual worlds. *ACM Transactions on Information Systems*, 11(3):239-265, July 1993.
- 2 C. Cruz-Neira, D. Sandin, and T. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Computer Graphics (SIGGRAPH '93 Proceedings)*, volume 27, pages 135-142, Aug. 1993.
- 3 C. Cruz-Neira, D. Sandin, T. DeFanti, R. Kenyon, and J. Hart. The CAVE automatic virtual environment. *Communications of the ACM*, 35(2):64-72, jun 1992.

- 4 A. Johnson and F. Fotouhi. The SANDBOX: A virtual reality interface to scientific databases. In *Proceedings of the Seventh International Working Conference on Scientific and Statistical Database Management*, pages 12-21, Charlottesville, Virginia, Sept. 1994.
- 5 A. Johnson and F. Fotouhi. The SANDBOX: Scientists Accessing Necessary Data Based On eXperimentation. Demonstration in the VROOM of SIGGRAPH '94, July 1994.
- 6 A. Johnson, F. Fotouhi, J. Leigh, and T. DeFanti. SANDBOX: an interface to scientific data based on experimentation. In *Proceedings of the Fifth Eurographics Workshop on Visualisation in Scientific Computing*, Rostock, Germany, May 1994.
- 7 T. Lehman and M. Carey. Query processing in main memory database management systems. In *Proceedings of ACM SIGMOD '86 Conference*, pages 239-250, Washington, D.C., May 1986.
- 8 G. Marks. How do analysts work? user interface issues. *Data Engineering*, 7(1):10-17, 1984.
- 9 D. Norman. *The Design of Everyday Things*. Doubleday, 1988.
- 10 B. Schneiderman. Direct manipulation: a step beyond programming languages. *IEEE Computer*, 16(8):57-69, Aug. 1983.
- 11 C. Shaw, J. Liang, M. Green, and Y. Sun. The decoupled simulation model for virtual reality systems. In *Proceedings of CHI '92*, pages 321-328. ACM, May 1992.
- 12 R. Springmeyer, M. Blattner, and N. Max. A characterization of the scientific data analysis process. In *IEEE Visualization 92*, pages 235-242, 1992.
- 13 D. Strebel, J. Newcomer, and J. Ormsby. Data management in the FIFE information system. In *Proceedings of International Geoscience and Remote Sensing Symposium*, pages 42-45, Vancouver, Canada, July 1989.

Andrew Johnson received his Ph.D. from the Department of Computer Science at Wayne State University in Detroit, Michigan. His current research interests include virtual reality, hypermedia systems, scientific databases, user interfaces, and artificial intelligence.

Farshad Fotouhi is an Assistant Professor in the Department of Computer Science at Wayne State University in Detroit, Michigan. He received his Ph.D. in Computer Science from Michigan State University. His current research interests include database management systems, object-oriented databases, temporal and multi-media databases, query optimization, and parallel processing.