Swat Team Trip Report

Earth Space Research Group
Santa Barbara, CA
Friday, 11 September 1992

Mike Olson (mike_olson@postgres.Berkeley.EDU)

1. Introduction

On Friday, 11 September 1992, the Sequoia 2000 Swat Team visited the Earth Space Research Group (ESRG) in Santa Barbara. The meeting included swat team members (Jim Frew, Mike Olson, Peter Kochevar, Duane Waliser, and Allison Woodruff) and members of ESRG and the Earth science community (Terry Figel, Bill Ohirok, Marty Landsfeld, Jim Davidson, Catherine Gautier, Didier Jourdan, and Mike Byers).

At the meeting, the Earth scientists presented their work. They emphasized the areas in which they need support from the computer science community. The area requiring the most support is the visualization of scientific data.

2. Background

At the last Sequoia retreat, researchers proposed several pilot projects to investigate global change using Sequoia hardware and software. These projects include

- understanding air/sea interactions and their effect on global climate;
- an analysis of the seasonal climatic cycle in the tropics, comparing model data to observational and remote sensing data;
- an analysis and model of net primary productivity, a measure of the biomass that exists at the bottom of the food chain; and
- mapping wildfires in California from satellite imagery.

An additional project has since been proposed by Warren White at Scripps Institute of Oceanography. This project would investigate the effects of El Niño on global climate systems. The El Niño project is most likely a second-year effort.

Only the air/sea interaction and tropical cycle projects were discussed in any detail on the 11 September meeting. The others will be covered at later swat team meetings.

We would like to see preliminary results for the four original pilot projects by the January 1993 Sequoia 2000 retreat. In order to get results by January, we must develop data visualization tools that support the pilot projects. Visualization means that data from various sources (satellite imagery, model output, observational data from ship tracks, and processed versions of all three) must be viewable simultaneously. ESRG scientists unanimously agree that they do not have the tools or expertise to do this kind of visualization right now.

3. ESRG Hardware and Organization

At the time of the meeting, ESRG owned five DECstation 5000/240s, four X terminals, and one DECstation 5000/25. The local network is Ethernet. They are connected to the Center for Remote Sensing and Environmental Optics (CRSEO) on the UC Santa Barbara campus by a T1 line. ESRG is located a few miles off-campus, and is unlikely to have a better link to the rest of Sequoia anytime soon.

Catherine Gautier directs the ESRG research efforts. Frank Davis, Jeff Dozier, Dave Siegel, and Ray Smith are Sequoia investigators at the CRSEO lab on the UCSB campus.

4. The Pilot Projects

4.1. Air/Sea Interaction Project

The pilot project exploring air/sea interaction seeks to understand climate over open ocean. This project is of interest to several ESRG researchers. Three different research projects exploring different parts of air/sea interaction were presented.

4.1.1. Separating Snow and Ice from Cloud Cover using AVHRR

The Advanced Very High Resolution Radiometer (AVHRR) is a sensor on a weather satellite that collects images of the Earth in thermal and visual bands. Two satellites are normally flying at any time, both in polar orbits. A single AVHRR collects images of Earth surface temperature in a number of spectral bands. The satellite traces a curved swath over the surface of the globe, since it passes from pole to pole and the planet rotates beneath it. A single image spans the Earth's surface from horizon to horizon.

Viewed from space, clouds have a spectral and thermal signature that is very hard to distinguish from that of snow and ice on the ground. Catherine Gautier is trying to find a way to separate cloud cover from snow and ice in AVHRR images. In order to do this, she converts satellite images in different bands *a* and *b* to a *dimensionless parameter space*:

$$v = \frac{p_a - p_b}{p_a}$$

The values *v* in parameter space have a spatial relationship to the pixels *p* in image space.

Catherine would like to be able to look at parameter and image space simultaneously. She would like to use color to highlight particular values in parameter space, and see the corresponding pixels highlighted in image space. This operation also applies in the other direction — highlighting parts of an image should cause the corresponding parameter values to be highlighted.

Catherine needs to be able to explore different parameter spaces and spectral bands easily. This means that the manipulation and visualization tools she uses must be interactive and relatively easy to use.

4.1.2. Monte Carlo Modelling of Photon Absorption

Bill Ohirok has a probabilistic model of atmospheric and Earth-surface photon absorption. He constructs a cube containing the volume of interest, and subdivides it into smaller cubes. Each of the smaller cubes has its own smaller coordinate system, and the coordinate systems of adjacent cubes may be different, so moving between them can cause a change in spatial resolution. Each subcube contains a mixture of aerosols (airborne particles), atmospheric gases, and water.

Bill shoots photons into the big cube. According to a probabilistic model, a photon bounces off a cube at a new angle or is absorbed by the cube. Some photons escape to space, some are absorbed by cubes in the model, and some are absorbed by the surface of the Earth. An absorbed photon contributes energy to the cube that absorbs it, affecting climate.

After a modelling run, Bill would like to visualize photon absorption and the trajectories of individual photons in three spatial dimensions and time. He would like to rotate the cube, fly through it, and incorporate results from separate runs of his model (for

example, different wavelengths of light) in a single visualization.

At present, a single model run is for 100,000 photons. Eventually, Bill would like to do much larger (10^6 or bigger) modelling runs. He also plans to make the model more complex by making the Earth's surface non-uniform by introducing vegetation and soil cover.

The biggest problem Bill faces with respect to visualization are the size of his model output and the bandwidth required for interactive visualization of his results.

4.1.3. Analysis of Latent Heat Flux

Didier Jourdan is working on a project that requires high-speed database access from his climate modelling software. He would like to visualize both his source data and the results he produces simultaneously.

Didier would like to analyze *latent heat flux* over the ocean. *Heat flux* is the transfer of heat from one object to another; here, it refers to the transfer of heat between the Earth's surface and the atmosphere. *Latent* heat flux is heat transferred by the evaporation or condensation of water.

If you know the surface temperature of the sea, the wind speed, and the amount of water vapor in the air, you can compute the amount of evaporation or condensation that will occur. Didier's model uses sea surface temperature, wind speed, and water vapor content data to produce a map of latent heat flux over the ocean.

Didier would like to be able to select a point in his result and see all the input values that contributed to it. For example, he may want to answer the question, ''Why does this output value look bogus?'' This means that the modelling tool or database must understand the relationship between the source data and the model output. Also, Didier could ask for any of the ancestor data at any point, so the database should optimize its storage of ancestor data cleverly to make these lookups fast. For Didier's purposes, a minute is too long a time to wait for answers to questions of this sort.

4.2. Tropical Seasonal Cycle Project

Duane Waliser presented work in progress at UCLA on understanding the seasonal cycle in the tropics. This project uses data from different sources to predict or record seasonal climatic variations, and tries to understand the differences between the models and observations.

One data source is output from Roberto Mechoso's global climate model (GCM). GCM output is in three dimensions; the model records data at various elevations above given latitude and longitude positions. These data are regular (evenly spaced across the Earth's surface), span the entire globe, and have coarse spatial but fine temporal resolution.

Another data source is satellite imagery. Images have fine spatial resolution in two dimensions, and arbitrary temporal resolutions. In addition, pixels in satellite images may not be at regular intervals over the field of observation.

A third source of data is climatology observations from ship traffic. These data are irregular; observations in shipping lanes are more frequent, and hence more reliable, than observations at less well-travelled locations. In order to compensate for the irregularity of observations, monthly averages are computed and used as input to the tropical seasonal cycle project.

The main problem faced by researchers at UCLA with respect to visualization is the heterogeneity of their data. They have different spatial and temporal resolutions, and different data types, and they want to visualize them all simultaneously and compare individual

values from each. In addition, they would like to do temporal animation on each data source.

5. Implications for Visualization

The pilot projects described in the last section all require good data visualization tools to succeed. Researchers must be able to visualize

- satellite images in multiple projections (for example, polar projections or equatorial projections spanning the prime meridian);
- geophysical products from processed satellite data (for example, Catherine's dimensionless parameter space from AVHRR images);
- observational data (shipboard temperature measurements); and
- model outputs (cloud and radiative heat transfer models).

These data sources may be very large, have varying temporal and spatial resolutions, and change resolution at internal boundaries. The scientists at ESRG would like to learn a single visualization tool that permits them to examine data from all of these sources, and to construct visualization images interactively.

5.1. Problems with Existing Packages

Terry Figel presented a summary of the problems with visualizing data from different satellite sources. Different satellites have different resolutions, both spatial and temporal. In addition, panoramic optical effects and surface curvature of the Earth distort images, particularly at the edges.

Existing packages for visualizing data have a number of problems, and no single package handles all data sources well. Among the problems with packages like IDL and TeraScan are

- projection singularities (strange behavior at the poles or prime meridian);
- excessive resource consumption — space or compute time — for examining large datasets;
- artificial limits of magnitude, such as being built for small-scale, rather than global, data visualization; and
- inflexible defaults, making it hard to build a visualization interactively.

5.2. Spatial Transformation on Images

Images of interest to Sequoia researchers are in different *geodetic projections*. Pixels in an image have a well-known spatial relationship to locations on the surface of the Earth. In some cases, pixels may be at known latitude and longitude. Pixel locations in any geodetic projection may always be converted to and from latitude and longitude.

When transforming an image from one geodetic projection to another, it may happen that either the source or the destination image contains holes. If an image is transformed into one with twice as fine a spatial resolution, the spaces between the original pixels must be filled in with sensible values. This problem is known as *resampling*.

Doing spatial transformations on images and resampling are well-understood. Sequoia researchers use images from many different sources, however, and no existing package handles transformations on all these sources. ESRG researchers need a generic spatial transform package that they can use to convert any image to and from some standard projection.

6. Action Items

Over the course of the day, the group agreed to pursue the following action items.

6.1. Specify a Generic Spatial Transform Package

The most important problem confronting ESRG scientists is the lack of a good tool for doing spatial transformation. By the January 1993 Sequoia retreat, we will write a document that describes

- existing software;
- its shortcomings; and
- whether we will write a generic spatial transform package ourselves.

In order to make this happen, everyone interested in spatial transformation will send their requirements to Jim Frew. Frew will generate a merged specification and search the network and literature to find out what exists now. Once the specification is ready, Mike Olson will find someone to write the software if we decide that we want to do that.

6.2. AVS Training

One powerful visualization tool with which we need more experience is AVS. This is an extensible system; it allows users to write modules that manipulate their data and to link these modules directly into AVS. Peter Kochevar suspects that software to handle many of the Sequoia visualization problems is already available for AVS.

Peter has volunteered to learn the system and to present a workshop on it before the January 1993 Sequoia retreat. Attendees will bring their own datasets with them, and Peter will show them how to use and extend AVS to manipulate their data.

6.3. Toolbox Interface for Visualization

ESRG scientists would like an interactive, Mac-like toolbox for constructing visualizations. For example, they would like to select from a palette of colors and icons for representing individual data values in a visualization.

Peter Kochevar will talk to the Earth scientists and will produce a specification for the toolbox interface. If possible, we will use an existing visualization package, like AVS, as the layer underneath this toolbox.

6.4. Bigfoot Dataset Newsletter

Jim Frew will produce a newsletter at arbitrary intervals describing the datasets that have been loaded on Bigfoot.

6.5. Data Loading

Duane Waliser and Terry Figel will load scientific project data onto the Bigfoot server in Berkeley. Researchers at ESRG will experiment with this arrangement to see how much more (or less) convenient it is than storing data on 9-track tapes locally. In addition, Peter Kochevar will use Duane's datasets to experiment with AVS.

7. Future Swat Team Meetings

The concensus was that the first swat team meeting was a success, and that the meetings should be continued. Visits to San Diego, UCLA, CRSEO in Santa Barbara, and Davis are planned for the near future.