Adaptive Resolution Data Access For Visualization of Magnetohydrodynamic Simulation Data

Andrew Foulks
† University of New Hampshire
Department of Computer Science
July, 2008

Abstract

Interactive visualization of very large data sets remains a challenging problem to the visualization community. One promising solution involves using adaptive resolution representations of the data. In this model, important regions of data can be selected using either error analysis or by allowing the scientist to interactively zoom to important phenomena. Regions of interest are rendered in high resolution data, while low and/or medium resolution representations are used to render regions in which the data does not vary much. Our data comes from numerical simulation of the solar wind with the Earth's magnetosphere. We have developed software tools and support libraries to aid in the interactive visualization of large data sets by using our adaptive resolution data model. We show that with this software, we can achieve significant performance gains while maintaining good image quality and maintain interactive frame rates.

Categories and Subject Descriptors (according to ACM CCS): H.3.4 [Systems and Software]: Performance Evaluation

1. Introduction

One significant challenge remaining in the scientific visualization community is the size of the datasets that must be analyzed. Typically, today's simulations generate far more data than the scientist can effectively understand. Understanding the science behind the data of these complex datasets remains a prominent scientific goal. The large size of the data represents an obstacle to timely scientific research, even on powerful modern workstations. In this paper, we present our multi and adaptive resolution data model and demonstrate its use to effectively deal with interactive visual exploration of very large time series datasets.

† This work was supported in part by NASA grant AISR05-0071, and the National Science Foundation grant IIS-0082577

Interactive visualization of simulation data is a very effective technique for gaining insight into the phenomena behind the data. It is the huge size of these science datasets that limit the interactivity in the visualization environment. Often space science disciplines must deal with data that is much too large to be stored in main memory or to be represented on a display screen of only a few million pixels. One promising solution involves data representation at multiple resolution levels. Coarser resolutions are used for large-scale overview visualizations, and increasingly finer resolutions can be accessed as the scientist "zooms in" to regions that encompass a more limited spatial and/or temporal range.

The advantage of using large data sets is that they offer more precise details and allow the scientist to gain greater insights from the data. These large data sets, however, provide two major obstacles to effective interactive browsing...
by the scientist. First, the costs of transferring the data to the renderer can be prohibitive given the scope of typical scientific data sets today. Second, both human and display resolution limits restrict the amount of information that can be displayed and processed at any moment. Ideally, we would like to do the rendering on a common PC, but still reap all the advantages of using large, informative data sets.

Some of the very largest datasets come from space science research that requires the analysis of time varying scientific datasets. Fortunately, the temporal domain also provides a natural mechanism for limiting data access that can ease the data size issue to a certain extent. Our research focuses on interactive visualization of time series of spatial datasets such as the data produced by magnetohydrodynamics (MHD) simulations. The output produced is quite often many hundreds of gigabytes of three-dimensional time series data. The combination of both temporal and spatial data provides opportunities for the development of visualization applications, to the extent that the application can easily access a data representation based on a resolution in both the spatial and temporal domains that is appropriate for the current task.

We are currently developing adaptive resolution data management tools that help support interactive visualization of very large multiresolution and multisource scientific datasets. Our goal is to integrate our software into existing visualization applications tailored to the needs of specific research areas and tasks. In particular, we focus on using multi- and adaptive resolution data to support interactive visualizations.

2. Prior Work

Multiresolution techniques for visualization of time varying data are an active research area [GWGS02, FJS96]. Guthe et al. use multiresolution representations to achieve high performance volume rendering. Our research is similar in spirit, but we focus on distinguishing between spatial and temporal hierarchies and how they can be effectively integrated into a single software framework. Finkelstein et al. use some interesting multiresolution techniques, but they focus on video images, where we are more interested in scientific data.

Doleisch et al. [DGH03] use a feature based approach and other techniques from information visualization to keep the amount of data that is presented to the user manageable. Our goals are similar since we also wish to reduce the amount of data that is presented to the user. Our approach is different in that we have concentrated on multiresolution techniques (in both space and time), and have built an application that demonstrates how our framework supports these techniques effectively.

Silver and Wang [SW99] describe visualization algorithms to handle time series computational fluid dynamics data sets. In their approach, they use feature tracking to improve user interaction and enhance the visualization of time series data sets. They do not, however, incorporate the notion of multiresolution into the organization of their data model.

Recent work in adaptive resolution techniques comes from Ramachandran et al. [RLM+07], in which an intelligent data thinning algorithm is used to deal with large volumes of streaming satellite data. Their approach is similar to ours in that they retain data in regions of high spatial frequency while decimating regions with low variances. They focus on two dimensional atmospheric and climate satellite data, whereas our focus is primarily on three dimensional simulation data.

Other researchers have addressed the notion of combining temporal and spatial multiresolution data representations for specific applications such as dynamic polygon meshes [SP01] as well as isosurface and direct volume rendering generation for time-varying fields [She98, SH99, SCM99]. Some have tried representing vector fields hierarchically [HWH99], storing error information during decomposition. Their data would fit well into our framework, because our approach is neither application-specific nor does it depend on a particular technique for creating the multiresolution hierarchy.

3. Multiresolution Representation

We have developed a multiresolution scientific data model [RBS01] that incorporates spatial and temporal semantics with localized error and we have implemented a prototype database system (the Granite Scientific Database System) based on that model [RBS02, RTBS05a, RTBS05b, Tan04, Ye04, Jia02]. The model’s semantics are common to many scientific applications and therefore are valuable to a variety of disciplines. The model supports integrated views of multi-file datasets that may be distributed over a network. Both point-based and cell-based data organizations are supported.

The model includes a standard multi-level data hierarchy for large scientific datasets. Original data is stored permanently at a repository site or sites. A preprocessing operation generates a multiresolution representation with increasingly coarse representations and the associated localized error information at each level. Users can download arbitrary subsets from any representation level, trading off accuracy and dataset size based on the needs of particular tasks.

One particularly effective way to generate multiresolution data is to apply wavelet transformations to such data [Dau92, Mal89]. A wavelet transformation is applied to the entire data set uniformly and results in a new summary data set that is \(1/2^d\) the size of the original where \(d\) is the dimensionality of the data. The wavelet transformation also produces a detail component that can be used to estimate...
the error that results from using the summary data to represent the original [WB95]. Decomposition techniques using wavelets lead to a natural multiresolution hierarchy in which the original (highest resolution) data is the root of the hierarchy and each successive child is the result of applying a wavelet transformation (or other uniform decomposition operation) to its child.

Our model supports both the spatial and temporal domains. These resolutions are integrated into a single comprehensive data representation. Our prototype implementation is in our STARgen and STARview tools [FBM05]. STARview supports multiple spatial wavelet decompositions of each step of a time series dataset, which can itself be a member of a multiresolution time series decomposition. Figure 1a shows a temporal data reduction by applying a one-dimensional time series data reduction for each spatial position. Figure 1b shows the full space-time multiresolution hierarchy. Given a time step we can generate $n$ spatial resolutions for that time step. We complete the hierarchy by generating a coarser temporal resolution for each spatial resolution created.

4. Adaptive Resolution Representation

Adaptive resolution representations of the data vary the resolution of different subsets of the data depending on the local and global behavior of the data [CDM94, LB01]. A coarse representation may be used for a region of the dataset if the amount of change in that region falls below a user specified error tolerance $\delta$. Typically, the scientist with knowledge of the dataset will specify $\delta$ based on the level of coarseness that is acceptable. The effect of this representation is that regions of high interest are shown in a finer resolution, while less important regions can be stored at coarse resolution.

Using low resolution data to represent regions of uninteresting data can vastly reduce the size of the data that needs to be read in and rendered. The dynamic nature of our adaptive resolution model allows the scientist to choose any region of interest and display it at an arbitrarily high resolution (up to the original data size). This representation can therefore help to minimize I/O and rendering pipeline computation costs. Figure 2 shows a general adaptive resolution hierarchy.

5. Data Visualization

5.1. MHD Simulation Data

Space science researchers have built and simulated numerical models of the solar wind and its interaction with the earth’s magnetosphere for only the past 20 years or so [LTKD78, LBFP81]. The data we are currently working with comes from simulations run at the Space Science Research center at the University of New Hampshire. Their simulation of solar wind activity records several physical attributes, including bulk plasma velocity, current density, magnetic field, and pressure. The data is a 3D time series in which points are sampled on a stretched cartesian grid [Rae95]. The particular dataset used for the examples in this paper contains 87 recorded time steps spanning a numerical simulation of 5220 time steps. The total size of this dataset is 15GB.

Figure 3 reveals the bow shock of the solar wind magnetosphere simulation data using a mapped pseudocolor
slice rendering. The image is shown at the full resolution (392x112x112). The image represents a single time step from Geospace General Circulation Model (GGCM) density simulation data [RBAA98, Rae03].

5.2. VisIt Toolkit

We use the VisIt visualization toolkit to render the images of the scientific data. VisIt is a general purpose visualization environment developed at Lawrence Livermore Laboratories. Its aim is to give researchers scientific visualization tools for visualizing scientific datasets. VisIt is built largely upon the Visualization Toolkit (VTK) [SML96] libraries, extending the interface and providing a comprehensive environment to the scientist. VisIt supports many different types of data and has a modular architecture that allows users to build data plugins to access other types of data. We have implemented a database plugin for VisIt that can read our multiresolution hierarchy and provide VisIt renderers with multiple resolutions of data.

We have extended the functionality of the toolkit and have integrated multiresolution and adaptive resolution representations. We utilize their API so that any existing rendering plugin can use our multi and adaptive resolution data without modification. Our software consists of several modules, including a Database plugin, responsible for reading our multi and adaptive resolution format; and user with a graphical user interface widget to control the multiresolution level of refinement and area of interest specification.

5.3. STAR Data Model

We term our data “STAR” data (Space Time multi/ Adaptive Resolution). In this data model, spatial and temporal multiresolution hierarchies are integrated in a global tree structure in which the base (finest) resolution sits at the root of the tree. Each component of a data hierarchy consists of uniform data at various resolutions. Adaptive resolution data hierarchies are metadata files built on the multiresolution data. A data reduction tool and a software library were developed to form the foundation necessary to allow users to generate and access by arbitrary combinations of spatial and temporal decomposed data. Figure 4 shows a VisIt visualization for the low resolution data (98x28x28) generated by our tools.

5.4. Adaptive Resolution Implementation

In this implementation, we use a uniform spatial subdivision octree data structure to build an arbitrary, dynamic adaptive resolution data representation from a standard multiresolution hierarchy. Using uniform subdivision results in a trade-off. The advantage is that an AR representation can be built dynamically from existing MR data stored on disk. The drawback is that regions of interest, specified either by an error tolerance $\delta$, or by the scientist performing a zoom operation, are limited to pre-chunked octants of data. In some cases, the interesting data is located on the boundary of two octants, requiring the scientist to render both in high resolution. In a general AR hierarchy model, only the most interesting data would be stored in high resolution. However, an additional preprocessing step would need to be performed for each given $\delta$ to create such an AR representation.

The first image 5 shows a uniform rendering of density data in which each domain of the octree is rendered at full resolution. This can be contrasted with the adaptive resolution data rendering, in figure 6 showing the density data using adaptive resolution representation. The upper right octant of data (as shown) is rendered in high resolution, but the other octants are rendered in low resolution. (The slice operation only reveals 4 of the 8 octants). This image is intended to highlight the difference between the high and low resolution regions of the data. The octant subdivisions achieve this contrast, although in reality this subdivision scheme may...
Figure 5: Uniform Chunked (octree) Resolution Rendering

seem a bit arbitrary. In this case, for example, the scientist might be more interested in viewing the bow shock in high resolution, but the rest of the data at a coarse level.

Figure 6: Adaptive Resolution Rendering

In the adaptive resolution images shown here, there are visible gaps between regions of different resolutions. This is an artifact of our use of wavelet transformations to generate the lower resolution data representations. In this context, the application of a wavelet transformation to a rectilinear grid of data points has the characteristic that data points of the output grid are not located at the same positions in the spatial domain as the original data points. A naive combination of data points from different resolutions results in gaps between regions at different resolutions that need to be filled with additional cells. Fang & Weber [FWC’04] describe an approach that inserts non-rectilinear cells in a way that avoids T-vertices. We have developed an approach that maintains rectilinearity at the expense of some additional cells. If the adaptive resolution data generation algorithm is tuned to favor relatively large blocks of uniform resolution data, the added computational cost of filling the gaps is negligible. In these examples, we have left the gaps unfilled in order to show clearly the boundaries between the different resolutions.

Figure 7: Adaptive Resolution Mesh

6. Results

6.1. Tests

Two experiments were run to compare the performance of rendering the original data produced by the simulation with the same data in our adaptive resolution representation. The results from this test are shown in Table 1.

The first trial measured the I/O and rendering time of 840 timesteps from magnetohydrodynamic simulation density data, each time step 784x224x224 sample points. The simulation data are sampled on a uniform rectilinear grid. Six trials were run and their times averaged. For the two data sets, both uniform data and adaptive resolution data were measured. The uniform data consists of 8 octants of 392x112x112 sample points rendered by performing a slice operation on a pseudocolor map of the density values. Adaptive resolution data was built from one high resolution octant at 392x112x112 sampled on the original rectilinear grid. The other octants were rendered at low resolution 49x14x14 blocks, sampled on low resolution rectilinear grids that were created using the same wavelet transformations used to decimate the data. Therefore, for high resolution blocks the region of interest selected by the scientist is half the size of the original data in each dimension. This region is displayed at full resolution (392x112x112). Other regions, deemed “uninteresting” are displayed at a coarse resolution (49x14x14).

A second trial was run using data sampled on a 392x112x112 grid, half the size (in each dimension) of the data used for the first trial. Again 840 timesteps of MHD density data were timed in six trials and their times were averaged.

The large data (784x224x224) takes on average 75 milliseconds to read the data from disk. From there it takes .35
For which we can render out-of-core images of both the uniform as well as the adaptive resolution data.

7. Conclusions and Future Work

We have developed a data model and tools that help the scientist to achieve interactive visualization of very large scientific data sets. To demonstrate the effectiveness of our model, we have integrated our preliminary implementation with the existing visualization toolkit, VisIt. Data represented by an adaptive resolution representation yields demonstrable performance gains. Coarse resolution areas are selected by the scientist to represent regions of uninteresting data. Interesting regions of data are shown at native resolutions to minimize the artifacts of the visualization at the coarse resolution.

Table 1: Test Results

<table>
<thead>
<tr>
<th>Data Size 784x224x224</th>
<th>Read Time</th>
<th>Render Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Data</td>
<td>0.074718</td>
<td>0.352907</td>
</tr>
<tr>
<td>Adaptive Resolution Data</td>
<td>0.020218</td>
<td>0.163336</td>
</tr>
<tr>
<td>Performance</td>
<td>3.69X</td>
<td>2.16X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Size 392x112x112</th>
<th>Read Time</th>
<th>Render Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Data</td>
<td>0.004707</td>
<td>0.115619</td>
</tr>
<tr>
<td>Adaptive Resolution Data</td>
<td>0.001664</td>
<td>0.102088</td>
</tr>
<tr>
<td>Performance</td>
<td>2.82X</td>
<td>1.13X</td>
</tr>
</tbody>
</table>

In spite of the large reduction in the size of the data, the adaptive resolution image (Figure 6) holds up well against the uniform data rendering (Figure 5). Though the images look good, often a question arises concerning the quality of the image. We are currently developing uncertainty visualization algorithms to address this issue, and we hope to integrate these algorithms with our visualization environment in the near future.

6.2. Trials

The smaller data (392x112x112) also saw performance gains. The adaptive resolution data is read in almost 3 times as fast and rendered 13% faster. The performance is to be expected considering that the amount of data processed is significantly less. A single timestep of uniform data contains 4,917,248 sample points of data. By contrast, the adaptive resolution data only contains 681,884 samples, a reduction by a factor of 7.2.

In spite of the large reduction in the size of the data, the adaptive resolution image (Figure 6) holds up well against the uniform data rendering (Figure 5). Though the images look good, often a question arises concerning the quality of the image. We are currently developing uncertainty visualization algorithms to address this issue, and we hope to integrate these algorithms with our visualization environment in the near future.

In fact we expect that as the size of the data further increases to get even better results. Once each timestep is too large to fit into memory, out-of-core rendering techniques must be used to render the high resolution images. We feel the AR representation could conceivably reduce the size of the data to be rendered to keep the rendering in-core, and thus keep interactive rendering rates possible for such large data sets. We are actively researching suitable environments for which we can render out-of-core images of both the uniform as well as the adaptive resolution data.

7. Conclusions and Future Work

We have developed a data model and tools that help the scientist to achieve interactive visualization of very large scientific data sets. To demonstrate the effectiveness of our model, we have integrated our preliminary implementation with the existing visualization toolkit, VisIt. Data represented by an adaptive resolution representation yields demonstrable performance gains. Coarse resolution areas are selected by the scientist to represent regions of uninteresting data. Interesting regions of data are shown at native resolutions to minimize the artifacts of the visualization at the coarse resolution.

Though we feel that our model and approach shows promise, there remain several unresolved issues with our current implementation. First, we would like to integrate our data model into a visualization environment that supports out-of-core rendering. We feel that there is a lot of promising research in this area. Second, we are currently integrating algorithms that effectively deal with the gaps that form naturally in AR data into the visualization environment. Finally, we foresee uncertainty visualization as a key part of AR visualization techniques. Good uncertainty visualizations should highlight the quality of the data being rendered to the scientist. These techniques show promise in dealing with unreliable visualization of the coarse data.

Additional future work calls for extending the data model to a full AR representation not based on uniform octree subdivisions but instead a k-D tree representation that better captures local regions of interesting data.

References


