Size Matters – Revealing Small Scale Structures in Large Datasets

T. Fogal, J. Krüger

Scientific Computing and Imaging Institute, University of Utah

ABSTRACT

The size of datasets generated in the medical imaging community is increasing faster than additional processing resources are made available. Even if we can surmount the hurdles in large data handling and processing, the amount of information encoded in these datasets is overwhelming. Therefore effective visualization techniques must allow a user to identify and focus on scientifically interesting subsets of the data. We describe a system for generating high-quality renderings of large volumetric data, and discuss the importance of scalable systems in the medical community.

Index Terms— volume, visualization, scalable, rendering

1. INTRODUCTION

There are two additional challenges that are presented once data grow beyond a certain size: first and foremost, we must be able to process and work with data using hardware resources that cannot advance as quickly as the size of our datasets. Secondly, the tools we create for processing these data must provide more efficient means to identify and focus on interesting structures within the data. The resolution obtained through modern data acquisition techniques is staggering, and therefore our tools must avoid presenting all of the information at once. It follows that those tools must allow users to efficiently focus their attention on interesting features of the dataset.

Aesthetics is an important and often overlooked secondary concern. While it is true that our primary motivation for using visualization tools is to create informative and useful visualizations, we argue that one should consider how the viewer will receive the generated media. Important considerations are ensuring the viewer is drawn to the feature of interest, is not distracted by extraneous information, and can immediately grasp the topic or issue that the visualization brings to light.

These concerns are frequently at odds with the challenges large data present. As the resolution of measured data improves, we begin to note many small scale structures that were not previously evident. In this paper, we describe a volume rendering system that addresses these concerns for large scalar data.

2. MOTIVATION

It is important to visualize data at their native resolution, despite the technical challenges in doing so. High resolution data can reveal features in a dataset which were not visible at lower resolutions. For example, the visible human male [1] was originally released in 1994, yet to date few people know about the tattoos which the male had on his chest and arms (see Figure 1). This was due to the difficulty processing and rendering the data at its full resolution. While a set of tattoos is arguably not a critical feature in a dataset, resolving structures on the same scale as a tattoo may be important for recognizing anomalies among the data, or useful in identifying components among a larger whole.

3. METHODS

3.1. Volume Rendering

We chose to implement our visualization tool based on the direct volume rendering method [2], to give the greatest flexibility to the user. Volume rendering allows users to map data values to colors and opacities of their choosing. This creates an empowering ability to create renderings which highlight areas of interest in exactly

Thanks to the NIH/NCRR Center for Integrated Biomedical Computing for funding.

O. Dössel and W.C. Schlegel (Eds.): WC 2009, IFMBE Proceedings 25/XIII, pp. 41–44, 2009.
www.springerlink.com

Fig. 1. Full-color visible human male dataset, with close-ups of the high-resolution tattoos visible on the chest and arms.

the manner that the user desires.

Unfortunately this flexibility comes with a price: volume rendering is one of the most computationally complex visualization techniques available [3]. We employ standard volume rendering techniques, such as empty space skipping and early termination [4], yet these techniques alone are not enough to achieve the real time rendering performance required for an exploratory visualization tool. To achieve this level of performance, we take advantage of the 3D texturing capabilities [5, 6, 7] and programmable shaders [8] of current graphics processing units.

3.2. Large Data

'Large Data' is a relative term. Such a term might refer to the physical extents of the domain under study, even if the dataset itself requires little disk space. It might refer strictly to the sampling rate used when acquiring that data. Or we might label data 'large' when it exceeds a particular threshold as stored on disk.

For our purposes, we use the following definition of large data:

- Data sets which are larger than the memory available on the computer used for visualization, or
- Data sets with a significant number of high frequency features.

These serve as our general guidelines, in that they imply all of the issues we encounter with large data. For example, large data might technically fit in the amount of memory available on a new workstation, yet conventional 32-bit computing architectures cannot handle data larger than two gigabytes. Further, the time to render datasets of this size would be prohibitively long for an interactive system. Yet data with high frequency components must be rendered at full resolution to effectively capture fine structures such as tattoos or bone texture. Clearly, these two issues represent competing design challenges for a visualization tool.

There are numerous solutions to these problems. The first and most popular approach is to downsample the data to a size which is manageable given the current

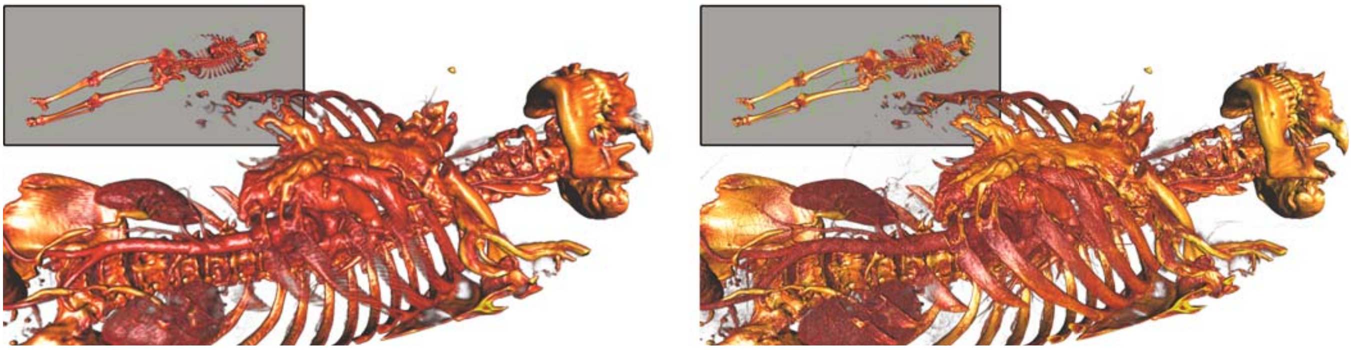

Fig. 2. Renderings at low (left) and high (right) resolutions reflect the importance of viewing data at native resolution. Note the additional texture available in the high-resolution rendering. The full dataset is shown in the background, with green outlines depicting bricks.

computing resources. This is undesirable for many reasons, the most important of which is the loss of important, high frequency features in the source dataset (see Figures 2 and 3) .

Another solution is to crop or extract a subset of the data. A dataset would be split into multiple parts, for example by extracting only the head and shoulders. While this can significantly impact the size of the data to be visualized, we argue that potentially important context information may be lost in this process. Furthermore, used exclusively this strategy is unable to cope with the largest of datasets: a third of the visible human female is around twelve gigabytes, too large for even the most capable current workstations.

3.3. Scalable Tools

The solution we have implemented is a bricked, level of detail volume renderer. Bricking the dataset involves dividing it into chunks of a predefined size, and committing to working with only one such chunk at a time. The advantage of this scheme is that it scales independently of the dataset size. Our system can render datasets up to tens of gigabytes as easily as it renders datasets comprised of only tens of megabytes.

Using level of detail techniques allows real time performance, even for datasets on the size of the visible human male. By presenting the user with a reduced resolution rendering of the dataset during interaction, immediate feedback is given when rendering parameters change. We emphasize that this feature is critical for identifying regions and features of interest. While visualization systems should encourage an exploratory style of interaction, without real time feedback for the available parameters users will inevitably resort to a guess and check style of interaction.

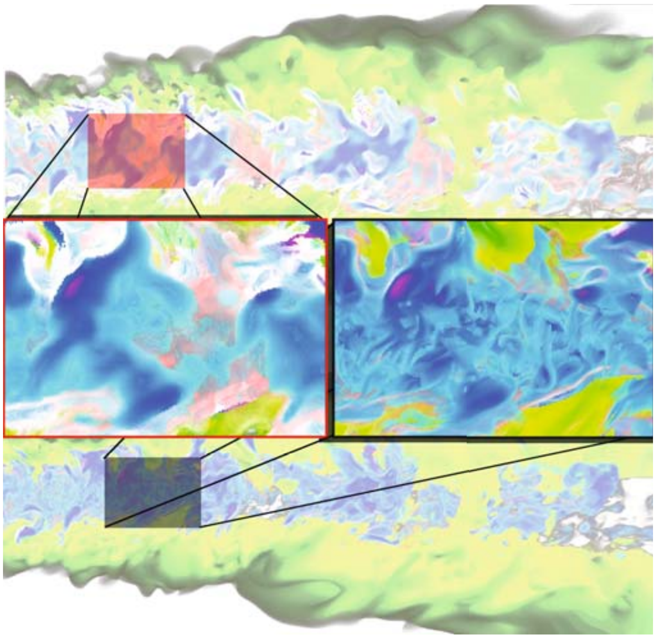

Fig. 3. Low (top) and high (bottom) resolution results from a combustion simulation. Some of the larger structures implied in the low resolution version (left) are not present in the data at its native resolution (right).

Conclusions

Analyzing and rendering data at their native resolution is paramount for gaining insight into fine structures

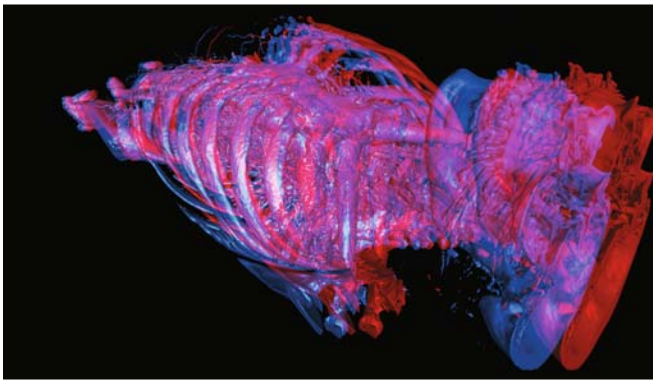

Fig. 4. Stereo rendering of a CT scan. Using anaglyph glasses increases the information presented by offering an illusion of depth.

that frequently go unnoticed. Despite the importance, many if not most tools cannot cope with the size of data currently being generated. To ensure that data size does not become a limiting factor in future research, new tools must be developed, or existing tools retrofitted, to explicitly cope with large data.

As we move forward with larger datasets, tools which can only load and render the data will become obsolete. Since the number of scientifically interesting features of a dataset scales with the dataset size, it follows that the limits of human analysis and cognition will soon become the limiting factor [9]. Therefore scientific visualization tools of the future must embrace a human-assisted computational analysis pattern, to help focus the rare resource of human analysis on the components of a dataset which are most interesting.

Acknowledgements

We thank Jacqueline Chen for the combustion dataset pictured in Fig. 3, and the visible human project for the dataset pictured in Fig. 1. This work was made possible in part by the NIH/NCRR Center for Integrative Biomedical Computing, P41-RR12553-07.

References

[1] V.M. Spitzer, M.J. Ackerman, A.L. Scherzinger, and D.G. Whitlock, "The visible human male: a technical report," *J. Am. Med. Inform.*, pp. 118–130, 1996.

[2] Robert A. Drebin, Loren Carpenter, and Pat Hanrahan, "Volume Rendering," in *SIGGRAPH '88: Proceedings of the 15th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1988, pp. 65–74, ACM.

[3] Nelson Max, "Optical Models for Direct Volume Rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, no. 2, pp. 99–108, 1995.

[4] Marc Levoy, "Efficient Ray Tracing of Volume Data," *ACM Trans. Graph.*, vol. 9, no. 3, pp. 245–261, 1990.

[5] Brian Cabral, Nancy Cam, and Jim Foran, "Accelerated volume rendering and tomographic reconstruction using texture mapping hardware," in *VVS '94: Proceedings of the 1994 symposium on Volume visualization*, New York, NY, USA, 1994, pp. 91–98, ACM.

[6] Timothy J. Cullip and Ulrich Neumann, "Accelerating volume reconstruction with 3D texture hardware," Tech. Rep. TR93-027, University of North Carolina at Chapel Hill, 1994.

[7] R. Westermann and T. Ertl, "Efficiently using graphics hardware in volume rendering applications," in *ACM SIGGRAPH 1998*, 1998.

[8] Jens Krüger and Rüdiger Westermann, "Acceleration Techniques for GPU-based Volume Rendering," in *Proceedings IEEE Visualization 2003*, 2003.

[9] Hank Childs and Mark Miller, "Beyond meat grinders: An analysis framework addressing the scale and complexity of large data sets," in *SpringSim High Performance Computing Symposium (HPC 2006)*, 2006, pp. 181–186.