Pure appl. geophys. 163 (2006) 2455–2465
0033–4553/06/122455–11
DOI 10.1007/s00024-006-0132-0

© Birkhäuser Verlag, Basel, 2006

Pure and Applied Geophysics

Visualization and Analysis of Multi-terabyte Geophysical Datasets in an Interactive Setting with Remote Webcam Capabilities

BENJAMIN J. KADLEC,[1] DAVID A. YUEN,[1] and GORDON ERLEBACHER[2]

Abstract—Visualizing and analyzing datasets in the geosciences is becoming increasingly complicated as their volumes are growing explosively. This poses a challenging problem for researchers who must sift through terabytes of data to discover useful relationships inside the information. There is a great need for geophysicists to interactively explore their data sets. Conventional visualization systems lack adequate bandwidth and rendering capabilities necessary for the largest data sets. CAVE and Powerwall display devices are necessary for researchers to explore their data sets in an immersive setting. We describe a utilitarian system targeted specifically at the cost-effective interactive exploration of data sets tens of terabytes in size and harness this system for visualization and analysis of geophysical simulations. Webcams can be used as a steering device to track a local region of interest, which is useful for remote visualization of large data sets. This system will be employed as a web-service under the auspices of Narada-Brokering, while using webcam technologies to enable remote visualization for collaborating researchers. Webcams can be incorporated in a point-to-point network for rapid exchange of information and quickly announcing natural disasters, such as tsunamis, landslides and earthquakes.

Key words: Remote visualization, interactive visualization, webcam, grid computing.

1. Introduction

Multi-terabyte data sets are becoming commonplace in the geosciences due to the proliferation of global observatories and large-scale numerical simulations. Projects such as EarthScope [http://www.earthscope.org] currently provide a wealth of observations on physical processes controlling earthquakes, volcanic eruptions, and tsunamis. Satellite missions, a part of EarthScope, will contribute 200GB/day of data to ongoing projects. The present progress and future needs in computational geosciences have been summarized in a recent report (COHEN, 2005). As

[1] Minnesota Supercomputing Institute and Virtual Laboratory for Earth and Planetary Materials, University of Minnesota, 413 Walter Library, 117 Pleasant St. SE, Minneapolis, MN 55455-0219, U.S.A. E-mail: kadlec@msi.umn.edu
[2] School of Computational Sciences and Information Technology, Florida State University, Tallahassee, FL 32306-4120, U.S.A. E-mail: davey@krissy.geo.umn.edu; erlebach@csit.fsu.edu

described in this report, computational geoscientists continue to run simulations resulting in increasingly large data sets. For example, numerical simulations associated with earthquakes and the predictions of 3-D tsunami wave propagation (WARD, 2004) can result in hundreds of gigabytes of data. On a larger scale, simulations of geophysical phenomena such as 3-D mantle convection are already producing terabyte-size data sets. This multitude of data is easily realized when considering a simulation on a 1000^3 grid with 5 output variables and 1000 time steps results in 20 Tbytes of data ($1000^3 \times 5 \times 1000 \times 4$ bytes $= 2 \times 10^{13}$ bytes). Computing multiple runs of the same simulation will then result in hundreds of terabytes of data. The results of simulations, in particular those relating to natural disasters, must be available in a timely fashion by using fast and efficient techniques for storing and visualizing these data sets. The International Solid Earth Research Virtual Observatory (iSERVO) (http://www.iservo.edu.au, AKTAS *et al.*, 2005) is a globally scalable grid-infrastructure that can provide remote access to data sets and visualization applications through a web-portal interface. We describe a cost-effective data exploration system using remote webcam technologies that can be integrated into the iSERVO network as a web-based component. The remote webcam service will be based on the framework of our Web-based Data Interrogation System (WEB-IS) (GARBOW *et al.*, 2003; WANG *et al.*, 2004). This system will allow visualizations and the analysis to be quickly transmitted to collaborating researchers, decision makers, and the broader community. The aims of this paper are as follows:

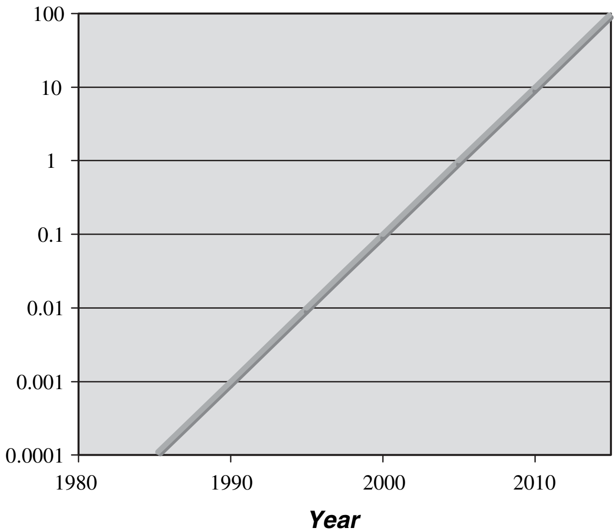

Figure 1
Exponential growth of data sets for 35 years starting from 1980.

- Point out the cost-effectiveness of webcam technology for remote visualization.
- Describe how the use of webcam will fit into the greater framework of WEB-IS and iSERVO.

2. Visualizing Multi-terabyte Datasets

The limiting process in obtaining scientific understanding from computer simulation is not the time required to perform the calculation, but instead the considerably longer time required to analyze and visualize the resulting data. Therefore the main objective of our work is to decrease the time needed to understand scientific data, which typically sets the rate of scientific progress, in the hopes of increasing scientific productivity. This urgency is particularly important when considering the promulgation of information related to impending natural disasters, such as tsunamic wave propagation in realistic ocean basins. Simulation and visualization are tightly coupled and visualizations must closely model the underlying computation (2-D or 3-D). We draw examples from techniques used in 2-D and 3-D simulations of slab dynamics and mantle convection, but other areas in computational geosciences, such as floods, volcanic eruptions and tsunami wave generation, also will produce voluminous data sets.

We have studied the two-dimensional multiscale dynamics of hydrous cold plumes at subduction zones (RUDOLPH *et al.*, 2004) using techniques for extremely high-resolution 2-D visualization, involving on the order of a billion unknowns. Commercial software, such as Matlab, Origin, and Tecplot, was found to be extremely hard-pressed for visualizing this type of large-scale data sets, involving more than 10^7 unknowns, and therefore new tools have been developed for addressing both local and remote visualization solutions. A web-based, zoomable image service (WEB-IS) has been developed to allow a user to explore high-resolution 2-D subduction zones through time, across many thermo-physical properties, and through different spatial scales (RUDOLPH *et al.*, 2004). We use a 13 million-pixel Powerwall display device for comprehending the relationships between multiple physical and chemical properties. The Powerwall enables parallel visualization of multiple fields in an immersive environment (Fig. 2). Recently available large (30-inch) LCD displays with resolutions of over four million pixels have opened the doors for extremely high-resolution powerwalls. The Electronic Visualization Laboratory [http://www.evl.uic.edu] at the University of Illinois at Chicago is developing a large-scale display with a wall of 30-inch LCD screens and a capacity of 4 Megapixels on each panel, essentially creating a powerwall with around one hundred million pixels! These new display walls will be enthusiastically received as simulations are already surpassing resolutions resulting from computations of one billion tracers.

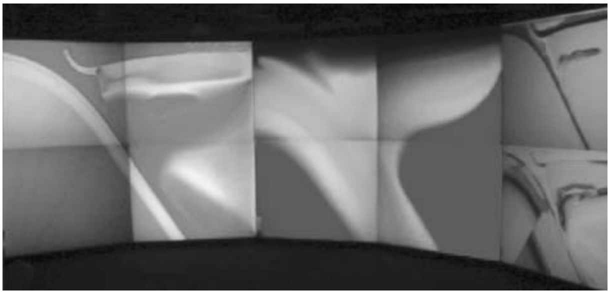

Figure 2
Multi-physics visualization of slab dynamics: Here we show four fields simultaneously on the Powerwall display device. The larger images from left to right are represented by five million pixels and show the viscosity and temperature fields, respectively. For further details see RUDOLPH *et al.* (2004).

We now consider visualization techniques for 3-D mantle convection simulations (Fig. 3) carried out on a large grid space (ERLEBACHER *et al.*, 2002). We have used a large database of 3-D simulations computed at the Earth Simulator Center (KAMEYAMA *et al.*, 2004). For modest data sets, the fast and interactive 3-D environment of Amira [www.amiraviz.com] works excellent when visualizing mantle convection data sets on workstations. Unfortunately, large time-evolving data sets are too big when using workstations with Amira and therefore more sophisticated software and hardware are required for creating large-scale time-evolving 3-D visualizations. Until now the solution has been to work with a down-sampled or smaller subset of the original data set, but this approach does not allow researchers to see their simulation at the highest resolution possible and can result in missed information. We will show that an interactive visualization system coupled with a large-scale display device can solve this problem.

We see promising results in using a CAVE (CRUZ-NEIRA, 1993) display device for 3-D simulations. Its immersive design allows for a natural exploration of 3-D data

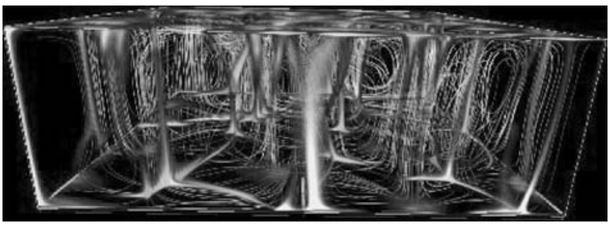

Figure 3
Visualization of 3-D mantle convection at a Rayleigh number of 10^8, using 400^3 evenly spaced points (ERLEBACHER *et al.*, 2002). Larger data sets from runs taken from the Earth Simulator Center (KAMEYAMA *et al.*, 2004) would further aggravate this data-extraction and visualization problem, as there are close to one billion unknowns to be sorted through at each time step. The cold descending flows and the hot rising plumes are depicted by the blue and yellow colors, respectively; streamlines show the directions of the instantaneous flow motion.

sets, such as those from mantle convection or geodynamo simulations. A CAVE is effectively being used for looking interactively at spherical shell convection at the Earth Simulator Center under the direction of Dr. Akira Kageyama [http://www.es.jamstec.go.jp/esc/research/Solid/members/kage/index.en.html]. The visualization centers supporting large displays such as the CAVE or Powerwall have systems tailored to storing and analyzing large data sets. We use the interactive visualization system at the Laboratory for Computational Science and Engineering (LCSE) [http://www.lcse.umn.edu] at the University of Minnesota for exploring our data sets.

The LCSE system uses a Hierarchical Volume Rendering (HVR) software for storing large geophysical data sets at multiple time steps and in full resolution. HVR (Fig. 4) is designed for simulations that require visualizing volumetric data sets, such as those from mantle convection simulations. The software works by saving a simulation's variables at every grid cell as compressed graphics files that contain all the information needed to make a snapshot of any variable over the problem domain. The resulting HVR-files (Figs. 5 and 6) are used for loading data sets into the interactive visualization system. Using a single byte per value, this hierarchical format is able to produce volume rendered images very quickly (PORTER *et al.*, 2004) and enables interactive rotation and zoom of large data sets. We have adapted our data sets to fit into this hierarchical data framework by using a Block of Bytes (BOB) (CHIN-PURCELL, 1995) intermediate file format as a way to create a hierarchical layout of bytes. The HVR software runs on a cluster of 10 rendering machines at the LCSE, each of which is connected to one of the Powerwall's ten display panels. Each node is equipped with 16 Gigabytes of RAM, a high-performance graphics card, and

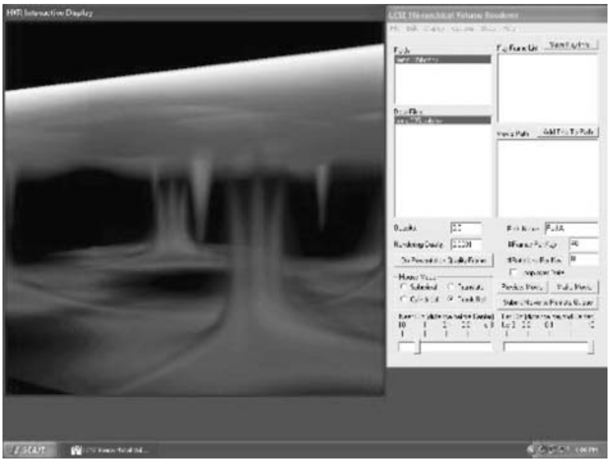

Figure 4
Using the HVR software for visualizing mantle convection data in the LCSE system. The situation depicted is layered convection with a Rayleigh number of 10 billion.

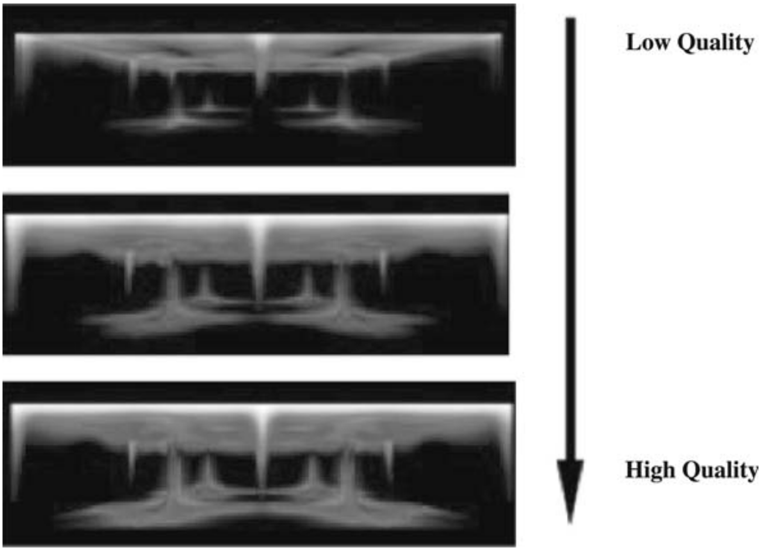

Figure 5
Mantle Convection Data Viewed in a Hierarchical Rendered Format. Rayleigh number of 10^{10} and 500^3 evenly spaced points have been employed (ERLEBACHER *et al.*, 2002).

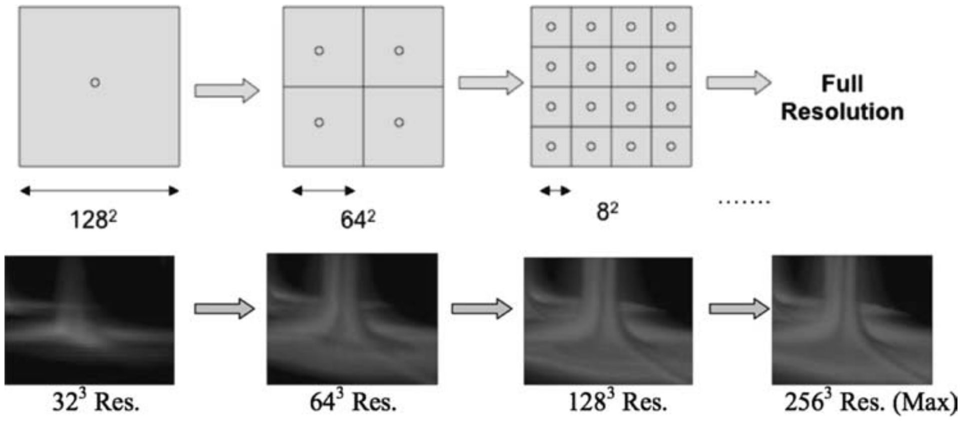

Figure 6
Storing data sets in a hierarchical data format for fast volumetric rendering (top, left to right) (PORTER *et al.*, 2004). The base of a 3-D mantle plume (ERLEBACHER *et al.*, 2002) displayed at increasing resolution (bottom, left to right).

connected to a hundred-terabyte RAID system for staging the largest data sets. This cluster of visualization machines works in-sync to allow interactive visualization of time-evolving data on the Powerwall using HVR.

3. *Remote Visualization Using Webcam*

As methods for large-scale interactive visualization are quickly being developed, there is increasing motivation for remotely visualizing data at interactive rates, using protocols such as virtual network computing [http://www.realvnc.com]. This need is especially dire in the geosciences, where global collaboration has quickly become commonplace with organizations arising such as ACES [http://www.aces-workshop-2004.ac.cn] and the iSERVO [http://www.iservo.edu.au, ATKAS *et al.*, 2005] group in earthquake physics. We focus now on enabling collaborative visualization by sharing with remote users. Data storage, graphic rendering, and high-resolution display hardware commonplace at regional visualization centers could provide a great benefit to distant researchers with limited access to advanced hardware and software. Webcam technology can be employed as a cost-effective solution for remote collaborative visualization by sharing these resources with remote users, who are looking at different portions of a Powerwall display (Fig. 7).

We have devised a solution for remotely viewing data sets located at distant computing centers and displayed on a large display device, such as the Powerwall, by using webcam technology. Using a method called remote scanning, visualizations are captured from a display device using a webcam equipped with a controllable TrackerCam base [http://www.trackercam.com] and a stream of normal-sized images is broadcasted at a rate of around 10 frames per second to remote users through a java applet (Fig. 8). A demonstration of this visual broadcasting technology was presented during the 2004 ACES Meeting in Beijing, China [http://www.aces-workshop-2004.ac.cn, KADLEC *et al.*, 2004]. Webcams achieve significant data

Figure 7
Schematic representation of collaborative visualization using remote webcam capabilities. Webcam scans geophysical data from the Powerwall and broadcasts it individually to remote users (laptops) through the Naradabrokering middleware. The viewers can look at different portions of the grand image.

compression by using photons to transmit visual information rather than bulky raw data files. The data compression inherent to webcams allows fast transmission of large visualizations, although some information of fine structures in the data set may be lost. Proprietary webcams, such as the Apple iSight®, are cost-effective solutions preferred over more expensive video conferencing hardware, since they are cheap enough to make widely available to the general public. The Apple iSight® can capture video with resolutions up to 1024×768 pixels. This is the maximum number of pixels that can be captured by a webcam when scanning a visualization at its highest resolution, which amounts to an area of 3 ft by 2 ft on the Powerwall. Larger areas of a visualization can be captured when the webcam reverse-zooms and scans a down-sampled version of the larger visualization area. This approach is effective for fast viewing of simulation results when high-resolution is not as necessary as understanding the greater structure of what is being visualized. Computational down-sampling is not necessary with webcam visualizations, and therefore a significant amount of time and resources are saved using this cost-effective approach. The remote java client (Fig. 8) captures a stream of images directly from the webcam and currently only supports a resolution of 640×480 pixels. The effective resolution on the client side is highly dependent on the bandwidth available to remote users, since greater video frame rates and resolutions correspond to a greater load on the network. Latency also needs to be considered when discussing network issues since as bandwidth increases, network latency decreases to under milli-seconds and allows for still greater webcam resolutions to be transmitted.

Recent events following the Sumatran Earthquake on December 26, 2004 have greatly increased the need for a system to quickly broadcast information on pending and ongoing natural disasters. The consequences that follow are well known when these systems are not in place. Therefore, the importance of using webcams for

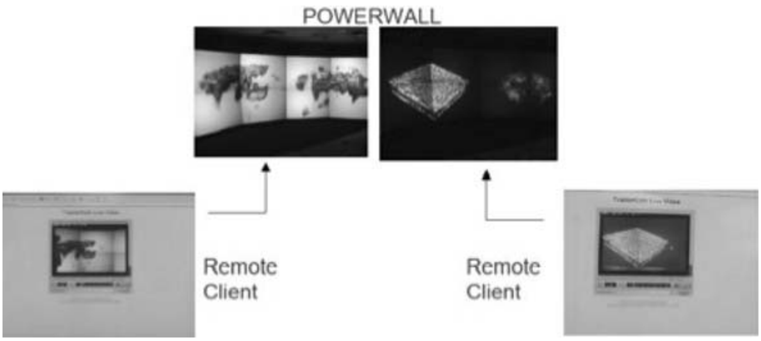

Figure 8
Remote scanning of the Powerwall using webcam. The upper middle images show visualizations of 3-D seismic tomography (top left) and 3-D mantle convection (top right) being scanned by two webcams while displayed on the Powerwall. The lower images are taken from a remote client viewing the webcam visualizations through an interactive java applet at 640×480 pixels.

remote collaboration is relevant now more than ever. Since webcams can be cheaply implemented in remote regions of the world, they remain an effective tool for transmitting visual information to sparsely located researchers and decision makers. By deploying webcams as a web-service in the iSERVO framework, they can be used for interactive remote visualization and broadcasting visual information on disaster events. Integrating additional controls with a web-based remote visualization service has the potential to facilitate additional remote interaction features with this service. The remote webcam service can be modeled after the Web-Based Data Interrogation System (WEB-IS) (GARBOW *et al.*, 2003 and WANG *et al.*, 2004) in the hopes to apply remote webcam scanning to iSERVO using the web-services middleware Narada Brokering [see http://www.naradabokering.org].

This approach follows popular methods currently being used for distance audio/video communication between researchers (KADLEC *et al.*, 2004). Recent advances in the Apple ® Macintosh operating system 10.4 has made it possible for the iChat [http://www.apple.com/ichat] webcam software to be used for multi-user conferencing with both Macintosh and PC users. Additional updates by other webcam softwares (MSN Messenger ®, AOL Instant Messenger ®) over the last year, have primed webcam for greater utility in collaborative movements by allowing for easy webcam communication in environments consisting of diverse computing systems (Windows ® PC, Mac®, Linux, etc.). Online virtual laboratories, such as the iSERVO and the Virtual Laboratory for Earth and Planetary Materials [http://www.vlab.msi.umn.edu], are now in a position to embrace webcam in order to fulfill goals aimed at educational outreach and bridging gaps between disciplines in the geosciences. By allowing different webcam systems to communicate with each other, less commonly used operating systems, such as Macintosh and Linux, are now able to take part in collaboration and outreach where they had previously been neglected.

4. *Conclusion*

Without any doubt visualization will assume an ever more prominent role in computational endeavors in the foreseeable future. As discussed already in the Cohen report (COHEN, 2005), many young researchers in the geosciences are already enthusiastic to use advanced visualization, but they lack the funding for expensive visualization equipment. They require the use of expensive interactive visualization systems for their simulation results, but unfortunately these resources are only available to a few privileged and well endowed Earth science departments and institutes. Therefore, regional visualization centers would satisfy this need by providing via remote visualization and webcam technologies the means for accessing advanced visualization tools and storing large data sets for many researchers at less well-supported universities. The Cohen report also highlights the need for conveying rapidly visual information for interpreting simulations and transmitting essential results. The jump to greater bandwidth and Internet-III will allow these aspirations to be satisfied.

In light of recent tragedy from the Sumatran Earthquake, it was shown that the means were not available to disseminate visual information warning of an impending tsunami disaster. In preparing for future geophysical events, simulations like those of tsunami wave propagation will need to be computed and visualized quickly so that results can be broadcasted to decision makers located in remote regions. To this end, we stress the necessity of the webcam as a fast and cost-effective implementation for making visual information available to a broader worldwide community. Through integration with visualization centers and the web-services devoted iSERVO framework, the webcam becomes an invaluable tool for researchers and globally wide communities.

Acknowledgements

We would like to thank Geoffrey Fox, Maxwell Rudolph and Charley Kameyama for stimulating discussions. This research has been supported by the Virtual Laboratory for Earth and Planetary Materials, Math-Geo, and ITR grants from the National Science Foundation.

References

Aktas, M., Aydin, G., Donnellan, A., Fox, G., Granat, R., Grant, L., Lyzenga, G., McLeod, D., Pallickara, S., and Parker, J., Marlon Pierce iSERVO: Implementing the International Solid Earth Research Virtual Observatory by Integrating Computational Grid and Geographical Information Web Serv Pure Appl. Geophys. *163*, 2281–2296.

Bollig, E.F., Kadlec, B.J., Erlebacher, G., Yuen, D.A., and Palchuk, Y.M. (2004), *Interactive collaborative visualization in the geosciences*, Eos Trans. AGU *85*(46), Fall Meet. Suppl., Abstract NG43A-0443.

Cohen, R.E. (2005), *High performance computing requirements for the computational solid earth sciences*, Workshop on *Computational Geoinformatics*, http://www.geo-prose.com/computational_SES.html, May 2–4, 2004.

Cave (2005), *Cave Automatic Virtual Environment*, Electronic Visualization Laboratory at University of Illinois, Chicago, http://www.evl.uic.edu/pape/CAVE/.

Chin-Purcell, K. (1995), *BOB-Brick Of Bytes*, University of Minnesota Army High Performance Computing Research Center.

Cruz-Neira, C., Sandin, D., and DeFanti, T. (1993), *Virtual Reality: The design and implementation of the CAVE*, Proceedings of SIGGRAPH 93 Computer Graphics Conference 08/01/1993 - 08/01/1993, ACM SIGGRAPH, pp. 135–142.

Erlebacher, G., Yuen, D.A., and Dubuffet, F.W. (2002), *Visualization and analysis of high Rayleigh number 3D convection in the Earth's mantle*, Proceedings of IEEE Visualization Conference, Oct.–Nov. pp. 493–496, IEEE Publishers.

Erlebacher, G. and Yuen, D.A. (2004), *A wavelet toolkit for visualization and analysis of large data sets. In Earthquake Research*, Pure Appl. Geophys. *161*, 2215–2229.

GARBOW, Z.A., YUEN, D.A., ERLEBACHER, G., BOLLIG, E.F., and KADLEC, B.J. (2003), *Remote visualization and cluster analysis of 3-D geophysical Data over the internet using off-screen rendering*, Visual Geosciences.

KADLEC, B.J., YUEN, D.A., and SEVRE, E.O.D. (2004), *Webcam for remote visualization and scanning of Giga- to Tera-byte size data sets on large-scale display devices*, APEC Cooperation for Earthquake Simulation (ACES), July 9–14, 2004.

KADLEC, B.J., SEVRE, O.D., YUEN, D.A., LILLI, X.L., BOLLIG, E.F., WANG, Y., ERLEBACHER, G., and RUDOLPH, M. (2004), *Web cam's potential for collaborative activities in the earth sciences*, Visual Geosciences.

KADLEC, B.J., SEVRE, O.D., YUEN, D.A., LILLI, X.L., BOLLIG, E.F., WANG, Y., ERLEBACHER, G., and RUDOLPH, M. (2003), *Web cam as an easy-to-use way for distant collaborative activities*, University of Minnesota Supercomputing Institute Research Report UMSI 2003/194.

KAMEYAMA, M., YUEN, D.A., and KAGEYAMA, A. (2004), *High-resolution 3-D numerical studies on the interplay between variable thermal conductivity and post-Perovskite phase transition*, Eos Trans. AGU *85*(47), Fall Meet. Suppl., Abstract MR23A-0194.

PORTER, D.H., WOODWARD, P.R., and IYER, A. (2004), *Initial experiences with grid-based volume visualization of fluid flow simulations on PC clusters*, University of Minnesota Supercomputing Institute Research Report UMSI 2004/114.

RUDOLPH, M.L., GERYA, T.V., YUEN, D.A., and DEROSIER, S. (2003), *Visualization of complex multiscale phenomena at subduction zones*, Eos. Trans. AGU *84*(46), Fall Meet. Suppl., Abstract NG11A-0167.

RUDOLPH, M.L., GERYA, T.V., YUEN, D.A., and DEROSIER, S. (2004), *Visualization of multiscale dynamics of hydrous cold plumes at subduction zones*, Visual Geosciences.

RUDOLPH, M.L., GERYA, T.V., and YUEN, D.A. (2004), *Vizualization challenges of a subduction simulation using one billion markers*, Eos Trans. AGU *85*(47), Fall Meet. Suppl., Abstract NG43A-0449.

WANG, Y., BOLLIG, E.F., KADLEC, B.J., GARBOW, Z.A., ERLEBACHER, G., YUEN, D.A., RUDOLPH, M., YANG, X.L., and SEVRE, E.O.D. (2004), *WEB-IS (Integrated System): An overall view*, University of Minnesota Supercomputing Institute Research Report UMSI 2004/66.

(Received December 30, 2004, revised August 26, 2005, accepted August 30, 2005)
Published Online First: December 20, 2006

To access this journal online:
http://www.birkhauser.ch