Visualization for eResearch: Past, Present and Future

Stuart M. Charters
Applied Computing Group, Lincoln University, New Zealand

Abstract

Visualization has been a part of computing for a long time, however with the growth in data produced by researchers and the computing resources available visualization capability has not developed in a way that provides researchers with the ability to include visualization as part of their standard analysis of these very large datasets. To address this issue and to understand how visualization systems must adapt to meet the new needs the eResearch brings we examine the past, present and begin to look into the future at visualization systems and architectures, to aid in understanding how visualization may be used or wish to be used a scenario involving a variety of Earth Science researchers working at a variety of locations to collect data and conduct analysis is presented. Reviewing traditional visualization systems, in particular, Modular Visualization Environments and Visualization Toolkits, to understand the heritage of visualization systems and the challenges that researchers have identified face. We look at current visualization systems that begin to take advantage of grid computing technologies, including those that modify traditional systems, those that a new architectures and those that have been developed in a bespoke manner for particular eResearch projects. Whilst these current visualization systems address some of the challenges of visualization for eResearch several challenges still exist and we examine ways in which these systems need to develop into the future to meet these challenges relating to use of multiple datasets, display devices, variation in bandwidth availability, the need for interaction and the role that predictive rendering can play in this, the need for new and revised algorithms, a focus on the end to end performance of visualization pipelines and the ability to integrate in to a researchers workflow rather than be an additional activity.

1 Introduction

Visualization has been a part of computing since its earliest days [Brodlie et al., 2004], increasingly it has also become an important part of research where it is used for analysis and communication. eResearch has seen a greater push towards interdisciplinary and multidisciplinary research and also an increased use of grid computing technologies to enable that research.

We can divide visualization into two broad types or categories, Scientific Visualization and Information Visualization. Scientific Visualization is concerned with the conversion of numeric data, usually from scientific experiments and simulations to a graphical representation. Information Visualization is concerned with the conversion of other forms of data, structured and unstructured text, images and video, to an appropriate graphical representation. The representations generated by the visualization process can be abstract in nature or can be based on a physical representation or a real world metaphor.

Butler et al. [Butler et al., 1993] describe three categories of user oriented visualization task:

- Descriptive Visualization, when the phenomena represented in the data is known, but when the user needs to present a clear visual verification of the phenomena.
- Analytical Visualization, or directed search, is the process we follow when we know what we are looking for in the data; visualization helps determine whether it is there.
- Exploratory visualization, or undirected search, is necessary when we do not know what we are looking for; visualization may help us understand the nature of the data by demonstrating patterns in that data.

Visualization has many advantages that it brings to a research investigation, five of these are [Ware, 2000]:

- Ability to comprehend huge amounts of data
- The perception of emergent properties
- Shows errors in the data easily
- Understanding large scale and small scale features
- Facilitation of hypothesis formation

These advantages show visualization has an important role to play in eResearch, however the development of visualization technologies has not kept pace with the changing needs of visualization users. This paper looks at a brief history of visualization before examining the current research efforts and visualization technologies available. We then look forward and discuss some of the developments that need to occur to allow visualization technologies to be integrated with the research process in a transparent manner.

2 Past Visualisation Systems

Despite visualisation being part of computing from its early history, it was the development of the visualization pipeline that has resulted in most modern visualization systems and toolkits.

2.1 Haber McNabb Pipeline

The Haber McNabb Pipeline [Haber and McNabb, 1990] consists of four parts, these are shown in Figure 1 and are Data, Filter, Map, and, Render.

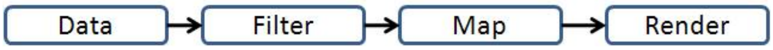

Figure 1: Haber-McNabb Visualization Pipeline

Each stage in the pipeline is concerned with performing some operation to its inputs to produce its output. The data stage forms the start of the pipeline and provides the data for visualization to the filter stage. At the filter stage the data may be manipulated to obtain the format, structure and types that the map stage requires. The map stage generates a representation of the input data which is the input to the render stage, responsible for producing the final output. In some visualization systems the pipeline model in not always explicitly exposed to users but the model exists in the implementation of the system.

2.2 Modular Visualization Environments and Problem Solving Environments

Modular Visualization Environments and Problem Solving Environments are both types of visualization application designed with a graphical user interface that allows visualization pipelines to be built using drag and drop interaction. Typically these types of system run on a single computer or a cluster at a single location. Examples of this type of environment are IRIS Explorer, SciRun and Paraview. These systems are typically designed to be used by a single user at one location and are limited by the type of resource they are executing on as to the size of dataset they can manipulate.

2.3 Toolkits

Visualization toolkits, such as VTK, provide a more flexible environment for constructing bespoke visualizations than modular visualization environments. Similar to modular visualization environments however visualization toolkits are typically designed to be developed into applications for a single user on a single computer or cluster. They have the advantage that they can be use to develop visualization applications that perform well for a particular problem but have the disadvantage that they require a greater level of programming ability than modular visualization environments to achieve a result.

2.4 Visualization Challenges

Hibbard presented 10 visualization problems in 1999 these are summarised below [Hibbard, 1999]:

1 Make the spatial and temporal resolution of visual displays indistinguishable from reality.

2 Integrate virtual reality with physical reality.

3 Integrate visualization with networking, voice, artificial vision, computation and data storage.

4 Optimize physical resources used to perform visual interactions.

5 Find effective ways to visualize numerical information of high dimension.

6 Find effective ways to visualize non-numerical information.

7 Find effective visual idioms for direct manipulation user interactions with visualizations.

8 Find effective visual idioms for for collaborative interactions among multiple users.

9 Define effective abstractions for the visualization and user interaction process.

10 Present abstractions to users in ways that reconcile expressiveness and ease-of-use.

Of these challenges it is challenge three, challenge four, and challenge eight that have most impact on visualization architectures, how they are designed and how they operate. We will revisit these challenges after an examination of current visualization systems to see how far we have progressed.

3 Current Visualization Systems

In this section we look at some current visualization systems making use of grid technologies, a more comprehensive review can be found in [Brodlie et al., 2004].

Shalf and Bethel 2003 present a scenario of heterogeneous resources on the grid being used for visualization and argue that a "fundamentals paradigm shift is needed to create fully global visualization applications". They discuss a scenario for a scientist undertaking earthquake research. A similar type of scenario that is expanded to encompass multiple visualization outputs, data inputs and collaboration models is shown in Figure 2.

The scenario in Figure 2 is based on the work of an Earth Scientist. The research team is working at a variety of locations, gathering data offshore, at a coastal location, from satellite imaging, from simulations and from previous survey work. This data needs to be visualized in a variety of different ways, including on a PDA by the coastal imaging team to ensure they have collected all the required data. Various collaborators around the world are viewing the results and performing new calculations at different times and communicating their results back to the group. One group of researchers are gathered together in at a single location and are using an immersive collaborative visualization environment to explore aspects of the data. Following the analysis stage the researchers make use of the the visualizations they have produced to communicate to a variety of different audiences including conferences and external agencies.

Three types of visualization architecture have been developed to take advantage of the developments in grid computing systems these are:

- Modification to existing platforms
- New architectures
- Bespoke Developments

Each of these architectures has benefits and deficiencies which will be discussed in more detail below.

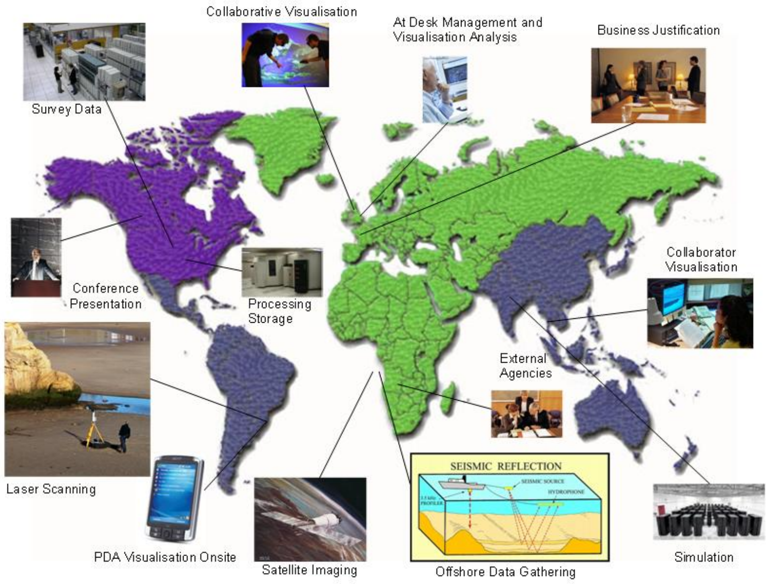

Figure 2: eResearch Visualization Scenario

3.1 Modified Platforms

Several existing visualization platforms have been modified to take advantage of grid computing, the most notable of these is the gViz project [Aslanidi et al., 2005] which has modified the IRIS Explorer platform to allow modules in the system to be run on remote computers. The advantage of modifying existing architectures over developing new architectures is the lower cost of entry, both in terms of development effort and the effort for researchers to adapt to using the system. The disadvantage of modifying existing platforms in this way is that it is harder to take advantage of all grid computing developments. In particular as the applications are centered on one user establishing, controlling and driving the visualization it is harder to implement some collaboration models and the execution of the visualization pipeline is tied to the resource executing the modified application.

3.2 New Architectures

A variety of new architectures have been developed to take advantage of grid computing technologies, a range of applications are discussed below.

The RAVE project [Grimstead et al., 2005] concentrates its efforts on the final render stage of the visualization pipeline in allowing the visualization to make use of a remote server to produce the visualization and a local client to display it.

A service based architecture as shown in Figure 3 has been developed by [Charters, 2006] which decomposes the visualization pipeline into constituent parts each of which can be encapsulated as a service and run on a different resource.

This concept has been taken and revised to use notifications by [Wang et al.,

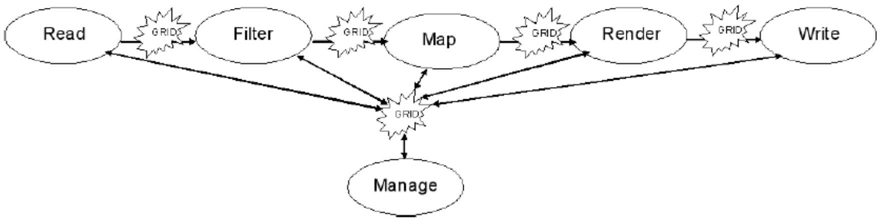

Figure 3: Service Oriented Visualization Architecture Pipeline

2006]. The advantages of this type of architecture is that the resource used for different stages of the pipeline can be tailored to the operation being performed to improve throughput and performance. The disadvantages relate to the effort of discovering and composing remote services into a pipeline to undertake the visualization and also the potential for networking delays to adversely impact the visualization.

3.3 Bespoke Development

Many grid computing projects requiring visualization capability have developed bespoke visualizations integrated into their application. An example of such a project is the RealityGrid project [Kalawsky et al., 2005], this project was funded through the UK e-Science Pilot Programme initially and produced an excellent demonstrator of what eResearch can achieve. The implementation of the project however can be considered a 'hero effort' with a large number of resources devoted to achieving the project, whilst this is useful in demonstrating what can be achieved through eResearch the level of investment required is outside that available to many researchers. For the majority of projects, the focus is upon a specific research outcome, rather than building the tools required to help achieve that research outcome. Bespoke development therefore is not suitable for the majority of projects.

3.4 Collaboration

eResearch projects have a greater potential make use of a variety of collaboration models in dealing with collaborators at the same institution, within the same country and internationally. Applegate 1991 presents a useful model of collaboration defining collaboration possibilities as (same place or different place) and (same time or different time).

Many visualization systems such as COVISA, described in more detail below, support Same Time, Same or Different Place collaboration. That is collaboration where the participants are connected to the visualization system at the same time, even if they are viewing the visualization remotely. eResearch increases the possibility that collaboration will occur Different Time, Different Place particularly with international collaborations across multiple time zones. This changes the requirements for visualization systems supporting collaboration, they now need to be accessible at any time by any collaborator rather than being under the control of a lead collaborator.

COVISA [Wood et al., 1997] is a collaborative enhancement to the IRIS Explorer environment which allows a user to export output from modules in the pipeline to other users to take as inputs. This type of collaboration enhancement is suitable for researchers who want to remain in control of what is exported and exposed to other collaborators and retain control over when access to the visualization is allowed. The exporting of outputs also requires each collaborator to have suitable resources for undertaking the computation for all stages of the pipeline downstream of the output exported. In the simplest scenario this is just the computation required to perform the final render stage, in the most complex it is the computation required for perform all stages downstream of the data read.

4 Challenges Revisited

The visualization challenges are listed in section 2.4, of these challenges numbers 3, 4 and 8 were identified as those that have most impact on visualization architectures. Challenge three about integration has begun to be addressed through the use of service oriented architectures which allow multiple different services to be integrated to perform a visualization. However this is not yet developed to include such services as voice and artificial vision. Challenge four relates to the optimization of resources, this is a challenge that is beginning to be addressed through grid computing which will allow greater utilization of available resources. Challenge eight relates to collaborative interactions, this an area which has seen much research both for visualization and generally, however few visualization systems incorporate effective collaboration support to date.

5 Future Developments

Having reviewed some of the current visualization systems, this section looks to the future and the ways in which visualization and visualization systems need to adapt to meet the needs of researchers and the developments required to achieve this.

5.1 Multiple Datasets

Increasingly researchers need to aggregate multiple datasets for analysis. This is illustrated in Figure 2 with Seismic Data, Laser Scanning, Simulation results and historical survey data being aggregated into a visualization. The use of multiple datasets brings with it issues of data formats, data alignment and registration. The datasets that researchers wish to access have also increased in size, these datasets are bigger than those that can be accessed by a desktop computer, sometimes in the order of Petabytes. This increase in size coupled with aggregation of multiple datasets means that strategies for accessing parts of the dataset and for distributed processing of the dataset are required.

5.2 Small and Large Format Displays

The range of displays and output devices available to researchers have grown as have the desires and expectations of researchers. The devices available range in size from Small Form Factor devices such as PDAs and Smartphones through to Large Format Displays such as Multimega pixel display walls. The capabilities of these devices has also grown from standard 2D views to autostereoscopic multiview 3D displays, immersive environments and volumetric displays. These devices each present unique possibilities for visualization and researchers have a desire to make use of novel display device where these provide insight into their data. The challenge for visualization systems is that many of these devices require special configuration to allow images to display correctly, for current visualization systems that configuration is often not possible. The challenge for new visualization systems is the ability to support a wide range of devices and multiple users each using a different device that requires a different configuration.

5.3 Bandwidth Flexibility

In a similar manner to the wide variety of display devices available, the range of network connection types that users have available and could make use of it also particularly diverse. At the limited bandwidth end of the spectrum are mobile data services running over 2G mobile networks to the high capacity, low latency, jitter free links offered by switched light path optical networks as offered by networks such as UKLight in the UK, StarLight in the US, and EN4R in Australia. The variety of different network capabilities mean that new visualization architectures will have to adapt to the network conditions they are presented with to provide the best level of service for the current available resources.

5.4 Prediction & Algorithms

One of the advantages that visualization can give during the analysis process is that of interactivity. The ability to manipulate the dataset to view it from different angles or to add and remove data from the display or to change parameters and understand the effect that the change has on the dataset. To achieve this when dealing with distributed systems and in visualizing large datasets where rending of the final image and transmission of the image across the network can be time consuming requires an approach other than on-demand rendering and transmission. One solution to this problem is to develop prediction and pre-fetching mechanisms that allow images to be rendered remotely and cached at the local client before the user has requested that they be displayed. [Chen et al., 2008] have looked at this problem and achieved a 233% effective speedup in image display through using a pre-fetching mechanism. They noted that pre-fetching was most effective when a user was conducting an orderly navigation in one dimension through the visualization and least effective, but still an improvement, when random navigation was undertaken. Mechanisms of this type will be required in future visualization systems to improve the performance and interactivity as perceived by the user.

Much work has been undertaken on algorithms of tightly coupled computing resources such as clusters and shared-memory machines, to take full advantage of grid computing developments, algorithms that scale well across loosely coupled heterogeneous architectures are required. These type of algorithms will be required across all stages of the visualization pipeline from data through to render.

5.5 Workflow

The main challenge to visualization for eResearch is the adoption of visualization by researchers. One of the major barriers to adoption is the fact that current visualization systems often do not integrate well into the workflow of a researcher. It requires specialist software to be installed and learnt, data conversion to be undertaken and pipelines to be built. If the researcher then wants to make use of specialist equipment or display the visualization in a collaborative environment the visualization may have to be recreated using another set of tools. For visualization to become a regular part of a researchers analysis it must integrate well into their workflow and require minimal extra effort to begin using. Extra effort may be worth expending once the value of the visualization has been demonstrated, however in the first instance the barriers to use must not be too great.

5.6 End to End Performance

Evaluation of many systems, including those used as components in visualization systems, is based on performance metrics, such as frames per second, data throughput, dataset size. From a user perspective raw performance of individual components of a visualization is not important, it is the overall performance of the visualization system that is important. Therefore for future visualization systems it will be the end to end performance of the visualization pipeline that needs to be measured. The performance can be measured in time to produce first view of the visualization, time to update the visualization and time to respond to other updates to the visualization pipeline.

6 Requirements for Grid Middleware

Visualization Architectures place requirements of the development of grid computing middleware that are different from systems that have been developed to date. These requirements are derived from the topics listed above and highlight those areas that impact grid middleware systems on top of which visualization architectures will be built.

- Interactivity
- Unknown computational load
- Collaboration
- Reliability and Quality of Service

Current grid middleware solutions are designed with batch computing in mind. Visualization users require a level of interaction from their visualization systems that batch computing is unable to provide, this leads to a requirement for middleware to support interactive execution of tasks. This may be achieved through scheduling systems that can give visualization tasks a high priority and run them in preference to other tasks.

Visualization tasks may also have a computational load that is harder to predict over time due to the interactivity requirement. This will make resource allocation and scheduling more complex, the addition of collaboration requirements exacerbates this issue as multiple users may be required to be supported at any one time.

Visualization tasks also require a level of reliability that is greater than other grid computing tasks, if a component of the visualization system fails, it needs to be replaced immediately rather than simply rescheduled for execution later as can occur with other batch computing jobs. Visualization tasks may also place quality of service requirements on grid systems, for example, a need to provide a certain frame rate or to process a dataset within a certain timeframe.

7 Design of Future Systems

Future visualization systems capable of supporting the eResearcher are likely to be based around a service oriented architecture for ease of discovery, and composition of pipelines, however behind these services will be implementations of algorithms and software that run on a variety of resources, using a variety of execution models to deliver the service.

The services would be semantically rich and handle all negotiations for security, quality of service and other requirements. The services would in many ways provide the "public face" for the components of the visualization pipeline with the main computation being performed behind the scenes.

The user interface to the online visualization pipelines could be delivered as a traditional software application on a users desktop or as a Rich Internet Application that can be accessed from any web browser to design, deploy, interact with and share a visualization pipeline. The display component of the pipeline can be a fat client using local resources to aid with rendering or a thin client which simply accepts image data from a remote service. This approach allows for flexibility in dealing with a variety of display devices, from PDAs to fully immersive environments.

8 Conclusion

Visualization systems have developed considerably over the years and recent developments to take advantage of grid computing has made some progress. However to meet the needs of researchers and achieve the paradigm shift needed "to create fully global visualization applications" [Shalf and Bethel, 2003] more research needs to be done to address the issues outlined and to achieve the goal of effective visualization systems for eResearch.

References

L. M. Applegate. Technology support for cooperative work: A framework for studying introduction and assimilation in organizations. *Journal Organizational Computing*, 1(1):11–39, 1991.

O V Aslanidi, K W Brodlie, R H Clayton, J W Handley, A V Holden, and J Wood. Remote visualization and computational steering of cardiac virtual tissues using gviz. In *Proceedings of UK e-Science All Hands Meeting 2005*, 2005. ISBN 1-904425-53-4. http://www.allhands.org.uk/2005/proceedings/papers/433.pdf.

K. W. Brodlie, D. A. Duce, J. R. Gallop, J. P. R. B. Walton, and J. D. Wood. Distributed and collaborative visualization. *Computer Graphics Forum*, 23 (2):223–251, 2004. doi: 10.1111/j.1467-8659.2004.00754.x.

David M. Butler, James C. Almond, R. Daniel Bergeron, Ken W. Brodlie, and Robert B. Haber. Visualization reference models. In *VIS '93: Proceedings of the 4th conference on Visualization '93*, pages 337–342, 1993. ISBN 0-8186-3940-7.

Stuart M. Charters. *Virtualising Visualisation: A distributed service based approach to visualisation on the Grid*. PhD thesis, University of Durham, 2006.

Jerry Chen, Ilmi Yoon, and Wes Bethel. Interactive, internet delivery of visualization via structured prerendered multiresolution imagery. *IEEE Transactions on Visualization and Computer Graphics*, 14(2):302–312, 2008. ISSN 1077-2626. doi: http://dx.doi.org/10.1109/TVCG.2007.70428.

Ian Grimstead, Nick Avis, Roger Philp, and David Walker. Resource-aware visualization using web services. In *Proceedings of the UK e-Science All Hands Meeting 2005*, 2005. ISBN 1-904425-53-4.

R.B. Haber and D.A. McNabb. Visualization idioms: A conceptual model for scientific visualization systems. In B. Shriver, Nielson G.M, and L.J. Rosenblum, editors, *Visualization in Scientific Computing*, pages 74–93. IEEE, 1990.

Bill Hibbard. Top ten visualization problems. *SIGGRAPH Comput. Graph.*, 33(2):21–22, 1999. ISSN 0097-8930. doi: http://doi.acm.org/10.1145/326460.326485.

R. S. Kalawsky, J. O'Brien, and P. V. Coveney. Improving scientists interaction with complex computational visualization environments based on a distributed grid infrastructure. *Phil. Trans. R. Soc. A.*, 363(1833):1867–1884, 2005. doi: 10.1098/rsta.2005.1616.

John Shalf and E.Wes Bethel. The grid and future visualization system architectures. *IEEE Computer Graphics and Applications*, 23(2):6–9, 2003. ISSN 0272-1716. doi: http://doi.ieeecomputersociety.org/10.1109/MCG.2003.1185573.

Haoxiang Wang, Ken Brodlie, James Handley, and Jason Wood. Service-oriented approach to collaborative visualization. In *Proceedings of UK e-Science All Hands Meeting 2006*, 2006. ISBN 0-9553988-0-0. http://www.allhands.org.uk/2006/proceedings/papers/651.pdf.

Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2000. ISBN 1558605118.

Jason Wood, Helen Wright, and Ken Brodlie. Collaborative visualization. In R. Yagel and H.Hagen, editors, *Proceedings of IEEE Visualization 1997 Conference*, pages 253–260. ACM Press, 1997. ISBN 1-58113-011-2.