Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

An integrated visualization environment for the virtual observatory: Current status and future directions

E. Sciacca [a,*], U. Becciani [a], A. Costa [a], F. Vitello [a], P. Massimino [a], M. Bandieramonte [a,b], M. Krokos [c], S. Riggi [a], C. Pistagna [a], G. Taffoni [d]

[a] INAF - Astrophysical Observatory of Catania, Via Santa Sofia 78, 95123, Catania, Italy
[b] Dipartimento di Fisica e Astronomia, Università degli Studi di Catania, Italy
[c] School of Creative Technologies, University of Portsmouth, United Kingdom
[d] INAF - Astronomical Observatory of Trieste, Via G.B. Tiepolo 11, 34143, Trieste, Italy

ARTICLE INFO

Article history:
Received 30 September 2014
Received in revised form
10 January 2015
Accepted 20 January 2015
Available online xxxx

Keywords:
Scientific visualization
Data analysis
Science gateway
Workflows

ABSTRACT

Visual exploration and discovery applications are invaluable tools to provide prompt and intuitive insights into the intrinsic data characteristics of modern astronomy and astrophysics datasets. Due to the massively large and highly complex datasets, various technical challenges are involved to reach, e.g. interactivity, integration, navigation and collaboration. This paper describes a number of approaches to address these challenges, and focuses on the current status of VisIVO (Visualization Interface for the Virtual Observatory) concentrating on the provided tools ranging from a desktop application to a science gateway and a mobile application. We emphasize the latest developments made in the context of past and current international European funded projects and highlight planned future developments towards further integration within the framework of the Virtual Observatory.

1. Introduction

Modern astronomy and astrophysics produce massively large data volumes (in the order of petabytes) coming from observations or simulation codes executed on high performance supercomputers. Such data volumes pose significant challenges for storage, access and data analysis, leading to the development of a fourth data intensive science paradigm (Hey et al., 2009) (also known as eScience). A critical aspect in understanding, interpreting, and verifying the outcome of automated analysis and data mining processes is the visualization of the scientific results. Data visualization is a fundamental, enabling technology for knowledge discovery, and an important research field that covers a number of different topics such as: optical and radio imaging (Nita et al., 2011), simulation results (Labadens et al., 2012), multi-dimensional exploration of catalogues (Beaumont et al., 2013) and public outreach visuals (Kent, 2013).

Visual exploration of big datasets poses some critical challenges that must drive the development of a new generation of graphical software tools, specifically:

Interactivity The majority of existing astronomical analysis and processing solutions lack the ability to deal with datasets exceeding the local machine's memory capacity while visual exploration and discovery in complex, multi-dimensional datasets is more effective through real-time interaction although sizes may not fit the available memory (Jin et al., 2010);

Integration Most of the data analysis systems are implemented as a set of separate independent tasks that can interact and exchange information via stored files only. This will be a significant factor which delays or even prohibits day-to-day data analysis tasks over big data sizes. Visualization tools should be ideally fully integrated within the scientists' toolkit for seamless usage, abstracting from technical details freeing scientists to concentrate in doing science (e.g. see Goodman, 2009);

Navigation Some of the current data processing techniques depend on parameters tuning, which may not be easy to achieve with large data sizes due to processing power limitations. Adopted solutions should allow intuitive and sophisticated navigation among datasets by exploiting ubiquitous environments (e.g. see Greensky et al., 2008 and Keefe, 2010), such as tablets or motion controllers,

* Corresponding author.
E-mail address: eva.sciacca@oact.inaf.it (E. Sciacca).

http://dx.doi.org/10.1016/j.ascom.2015.01.006
2213-1337/

offering new Human–Computer Interaction paradigms to better tune the processing parameters;

Collaboration It will no longer be an easy job to develop a simple script or program to deal with such data. These tasks usually require a deep knowledge of technicalities and programming experience which are typical of computer scientists rather than of astronomers. Tools should be built into the processing pipelines in order to facilitate visualization, processing and analysis of big data in a collaborative manner (Mark et al., 2002; Balakrishnan et al., 2008; Isenberg et al., 2011).

Therefore new visualization tools and software should be designed to overcome the limits and the barriers of traditional software by exploiting the latest technological opportunities. Based on the aforementioned challenges a number of proposed objectives to design new generation graphical software tools by exploiting the latest technological advances (e.g. science gateways) are the following:

Interactivity — For complex visualizations the relevant computations should be performed close to the data to avoid time consuming streaming of large data volumes. This can be achieved via flexible distributed architectures striking a balance between local interactive exploration tools and remote services hiding data complexity.

Integration — Tools should be coupled with advanced high performance computing (HPC) resources to deal with requests to archives through scientifically meaningful lightweight versions of the datasets obtained by analysis/processing operators since full data sizes may not fit the available memory to allow real-time interaction.

Navigation — Local exploration tools should enable interactive visualization optimized for ubiquitous computing environments, intuitively controlling the resulting visualization.

Collaboration — Tools should be combinable within e.g. science gateway technologies (Kacsuk, 2014) to allow collaborative activity between users and provide customization and scalability of data analysis/processing workflows, hiding underlying technicalities.

A number of different analysis and visualization tools and software have been developed for astronomers and astrophysicists during the last decades (refer to Hassan and Fluke, 2011 for a comprehensive review on existing tools). Many of them are designed to run on desktop computers, which have a finite memory size, and they are not suitable for exploring large scientific datasets. A solution to this problem is the use of distributed visualization, where a networked computing cluster shares the visualization tasks. For example, ParaView[1] or VisIt[2] can be used either as stand-alone local tool or they can run on a cluster in a client–server configuration. Remote visualization can imply losing real-time interactivity due to data transfer over the network.

Another important aspect is the collaborative visualization that enables multiple users to share a visualization experience. To successfully achieve this aim, it is necessary to provide high-speed network connections and effective communication protocols, e.g. in AstroSim (Nakasone et al., 2009).

When dealing with a large dataset, additional benefits may be achieved using workflow driven applications, e.g. in VisTrails (Freire et al., 2014). A workflow manages data and metadata of visualization products and it simplifies the complex problem of creating visualizations that becomes subsequent tasks of a workflow engine.

To reduce the output data size some authors propose an approach based on in situ visualization. For example Kageyama and Yamada (2014) developed a method for the interactive analysis of in situ visualization images produced by a batch simulation job, however this approach lacks integrated data analysis functionalities for effective data exploration.

A high-performance, GPU-based framework is presented in Hassan et al. (2013) for the analysis and visualization of terabytes sized 3D images. The framework was one of the first distributed environments but it lacks any collaborative facility and, most importantly, it relies only on powerful HPC infrastructures i.e. GPU clusters.

Finally, the astronomy and astrophysics community often relies on customized codes and scripts that are created to overcome limitations in other software to handle specific astronomy data formats or other domain specific issues related e.g. to the kind of visualization they want to obtain, see for example Labadens et al. (2012) and Szalay et al. (2008).

VisIVO software is an innovative tool that accounts for all the criteria highlighted above. VisIVO offers a unique integrated ecosystem for visualization including: services for collaborative portals, mobile applications for visualization and data exploration. It includes a number of key components such as workflow applications, analysis and mining functionalities, which can be re-thought under the big-data paradigm (see Section 5), focusing on: (1) an effective merger of the filtering/visualization pipeline so that the user can see in real time the effect of operations or interactions with the visualization results and (2) optimizing data movement and memory usage.

This paper presents the latest developments of VisIVO. Section 2 describes the basic technical details and features of VisIVO software tools and Section 3 discusses the latest developments made in the context of past and current international European funded projects. Section 4 is dedicated to present some scientific cases in which the support of data analysis and visualization is of primary importance. Section 5 presents the planned developments of VisIVO in the context of the Virtual Observatory framework to provide an integrated visual analytics tool for big data analysis and visualization.

2. The VisIVO ecosystem

VisIVO is an open source collection of graphics applications, which blend high performance multidimensional visualization, data exploration and visual analytics techniques. VisIVO tools have been developed to allow multidimensional data exploration of very large-scale datasets in order to visualize and discover a priori unknown data characteristics e.g. searching for outliers, characteristic regions or special properties. The underlying design based on commonly used toolkits (e.g. the Visualization ToolKit[3]) makes VisIVO a visualization framework useful also in contexts outside astronomy and astrophysics.

The main VisIVO tools are the following:

- VisIVO Desktop (Comparato et al., 2007), a stand-alone application for interactive visualizations running on PCs;
- VisIVO Server[4] (Becciani et al., 2009), a grid-enabled high performance data exploration and visualization command line toolkit;
- VisIVO Library (Becciani et al., 2012b), an application programming interface that has been developed to port VisIVO Server on gLite middleware[5] but can be installed on common Distributed Computing Infrastructures (DCIs);

1 www.paraview.org.
2 https://wci.llnl.gov/codes/visit.
3 http://www.vtk.org.
4 http://sourceforge.net/projects/visivoserver.
5 http://glite.cern.ch.

Please cite this article in press as: Sciacca, E., et al., An integrated visualization environment for the virtual observatory: Current status and future directions. Astronomy and Computing (2015), http://dx.doi.org/10.1016/j.ascom.2015.01.006