EG UK Computer Graphics & Visual Computing (2016)
Cagatay Turkay and Tao R. Wan (Editors)

Collaborative Computational Projects – Visualisation Applications Survey

M.J. Turner,[1,2] R. Fowler[1] and T. Morris[2]

[1]Scientific Computing Department, STFC, Didcot, Oxfordshire, OX11 0QX
[2]The University of Manchester, School of Computer Science, Kilburn Building, Oxford Road, Manchester, M13 9PL

Abstract

This extended abstract presents initial outcomes from three visualisation user needs surveys, and includes an invitation for new communities to engage with follow-on surveys. Statistical and text cluster analysis have been used to assist specific computational groups; in order to select certain visualisation application packages for software development and to select which new algorithms to implement. This analysis is now also available for advising and creating recommendations to build a long term visualisation support service. The focus of these surveys and this work has been on looking at the use of software toolkits and application packages rather then surveying specific visualisation algorithm techniques.

Categories and Subject Descriptors (according to ACM CCS): I.3.4 [Computer Graphics]: Graphics Utilities—Application packages I.3.8 [Computer Graphics]: Applications—

1. Motivations

In 2015-2016 a Visualisation Tools survey was launched and analysis carried out aimed at certain High Performance Computing and collaborative computational user groups. This short extended abstract very briefly describes the initial results of this project with an invitation for communities to join the second phase during 2016-2017.

Through the EPSRC SLA project (Service Level Agreement within STFC's Scientific Computing Department http://www.scd.stfc.ac.uk) we were asked to coordinate some visualisation surveys across certain computational focused disciplines. This had the initial goal to inform specific current research software engineers of their current users' actual use, as well as indicate trends and development roadmaps for future areas of software development and collaborations. A longer term analysis is ongoing to consider the feasibility of a future funded visualisation service. Complete details are being stored within an explorable wiki site launched in 2015:

http://www.vizmatters.cs.manchester.ac.uk

Initial results have already informed software engineering work within certain groups; for example 3D and 4D tomographic (volume visualisation) quantification and evaluation algorithm developers are now focusing on a smaller and different subset of application for their plug-ins that impact the majority of their users; (the [Nag16, NF16, Wor13] development community now create plug-ins and have simple wrappers almost exclusively for ImageJ, ParaView and Avizo: a benefit to current and new developers is that there is experience for creating plug-ins that cross link between these applications).

2. Methodology

Three surveys have so far been issued: a global survey of all funded networks, as well as two specific surveys for Tomographic imaging and CFD respectfully. The surveys were broadly similar and presented in this extended abstract is part of the analysis for the main global survey.

2.1. Community response rates

Chosen as the main user base were the CCPs (Collaborative Computational Projects) who bring together leading UK expertise in key fields of computational research to tackle large-scale scientific software development, maintenance and distribution. Each project potentially represents many years of intellectual and financial investment as well as a diverse and expanding network of users [Jon16].

There have been over 20 CCPs funded in the past, ranging from macromolecular crystallography to condensed phase materials; from computational electronic structure of condensed matter to biomolecular simulation; and from material science tomographic imaging to medical PET-MR synergistic algorithms. Integrated visualisation and making a sensible choice of visualisation application is an important component within almost all the CCPs.

For this survey the questionnaire was distributed via each CCP's emailing list, and there were over 100 responses from 57 different

DOI: 10.2312/cgvc.20161292

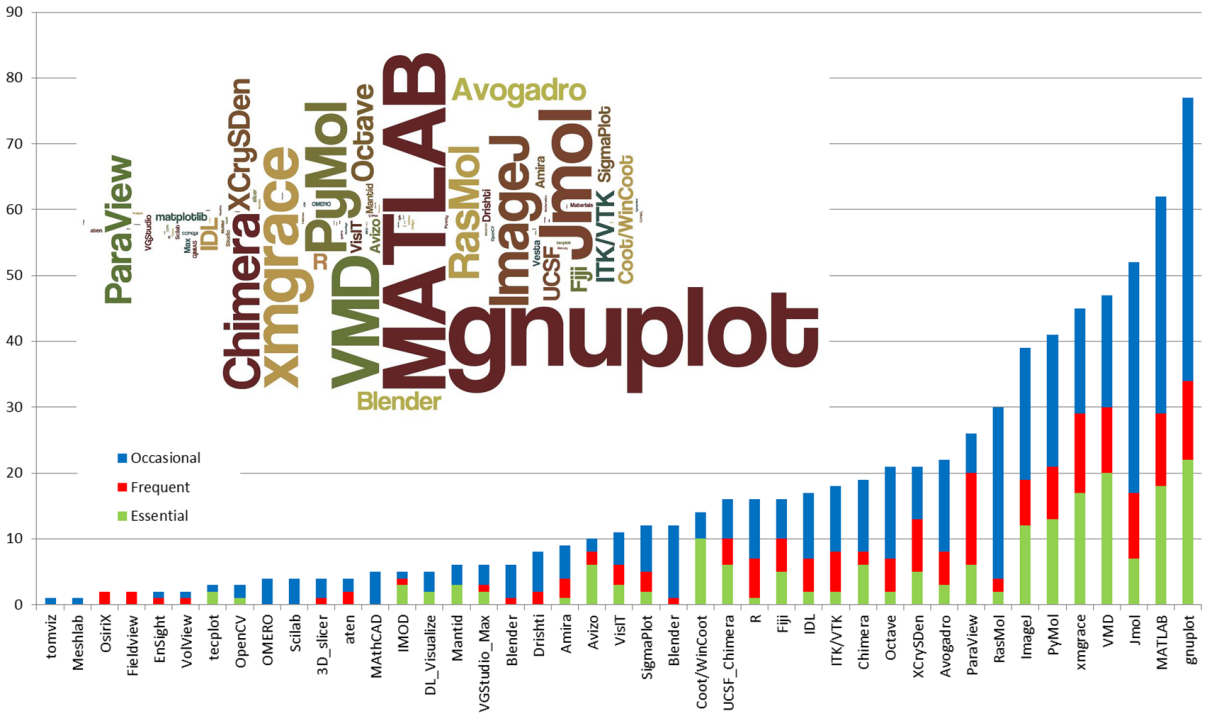

Figure 1: *When asked what software do you use for visualisation of data? we looked at the usage for 'essential', 'frequent' and 'occasional' categories that all had a long tail distribution. This figure indicates via the colour how different the graph would look when just considering the 'essential' tools. This can also indicate how each user community has its favoured tool that can be quite specific.*

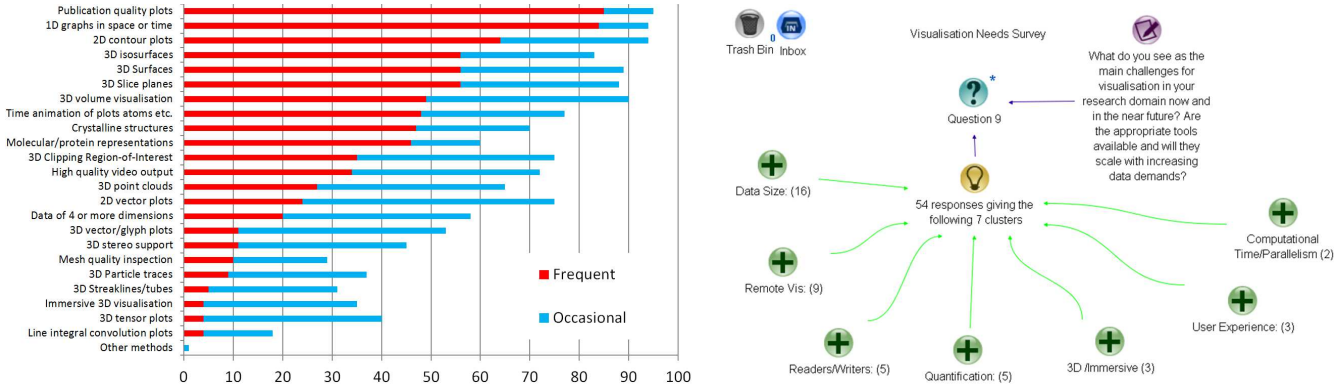

Figure 2: *Asked what visualisation techniques are important to your work responses were one of: 'occasionally', or 'frequently' used. The histogram represents the number of responses ordered by the summation of the 'frequent' replies. Also shown is a screenshot of the CompendiumNQ free-form text clustering results.*

distinct institutions/groups. This represented respondents across the world, with a bias towards Europe and the UK. The raw numbers show a long tail distribution that is not a normal power law relationship. As expected the newer and therefore more current organisations returned a larger response; with CCP5 (Computer Simulation of Condensed Phases) and CCP9 (Study of the Electronic Structure of Condensed Matter) being popular source code repositories and ISIS (Neutron Spallation Source) and CLF (Central Lasers Facility) being popular user communities.

2.2. Questions and analysis

The global survey had the following ten questions with an extensive list of selectable options provided for further statistical analysis, as well as open ended sections after each question for adding comments. For the open ended sections, we utilised the CompendiumNG mindmapping tool [SSS*01] to cluster comments on keywords and themes employing a semi-guided analysis to define specific clusters of issues.

Q1 Home institutions
Q2 Which CCP(s) are you involved with?
Q3 What software do you use for visualisation of data?
Q4 What visualisation techniques are important to your work?
Q5 Comments on the respondents' most used visualisation tool.
Q6 Same as question 5, for any other tool used.
Q7 Visualisation requirements. How important do you see the provision of the following?
Q8 Requirements for high performance/advanced visualisation facilities. Do you have any need for access to:
Q9 What do you see as the main challenges for visualisation in your domain now and in the near future?
Q10 Any other comments?

2.3. Summary results

Briefly listed here are some key outcome results that have been acted upon:

1 When asked what software do you use for visualisation of data, Figure 1 shows the distribution and length of the tail (only top 30 applications shown). The survey followed guidance from previous social science surveys on geographic visualisation that forced responders to consider tools as either; 'essential', 'frequent' or 'occasional' [DMT08]. The top three 'frequent' tools (gnuplot, MATLAB, Jmol) account for 26% of the responses. Similarly the top four 'essential' tools (gnuplot, MATLAB, VMD, xmgrace) account for 42% of the responses. The tail is longer than shown in Figure 1 with a further 41 other tools, with less than 3 responses, not displayed (these are being incorporated in the wiki). A question we need to consider is why there is a long tail and if this could be because there are extra useful features provided by the less popular applications that are useful to unique groups.

2 For their favoured visualisation tools there were a series of comments requested (that are extractable on the website). For these tools (top five; VMD, MATLAB, Avizo, Gimias, PyMol), using text keyword clustering a few observations can be drawn from the responses;

- Users will prefer software that is written specifically for their domain of interest.
- Large datasets must be handled efficiently.
- Scripting or other ability to extend the tool is required.
- Publication quality output of images or graphs is a valued bonus

3 Users second most favoured packages are often general purpose visualisation tools (top six; gnuplot, Materials Studio, MATLAB, paraview, PyMOL, VMD) and these had the following observations;

- Users seem to prefer software that is general purpose.
- Large datasets must be handled efficiently.
- Good quality documentation/tutorials is required.
- Ability to read multiple formats is useful.

4 We provided a list of what we believed to be the important visualisation techniques used by members of the CCPs. The question asked about how often these were used (Figure 2) and looking at the frequently used techniques: the most common technique is to produce publication quality output and the facility to produce line graphs is equally important, then there are a set of specific visualisation methods that are also required.

5 The future issues question was totally open ended, and the following issues are in order of importance;

- The ability to handle large amounts of data
- The ability to operate in a distributed environment.
- Package needs to contain many useful Readers/Writers
- Quantification tools are expected
- 3D Immersion available
- Enhanced User Experience
- Computational/Time Parallelism

In the general comments fields of the survey, again using the text cluster analysis (see Figure 2), open source was not always the most important issue, but the easy creation of plug-ins, new readers and writers, as well as analysis tools have all been requested.

3. Success, and Future Work

In further analysis via data mining we can now extract specific information for certain groups which has been requested; but more raw data is required. The process has been shown to be useful and a series of further surveys and follow-up questions are now planned and we invite further communities to join. The wiki site [MFT14] is being expanded upon with other related tools and exploratory content.

With an era of research software engineering principles of continual beta development and constant agile design it is important to understand how and when to act upon change requirements; and a take home message should include within the use of continual software development process the need for continual user surveying.

References

[DMT08] DODGE M., MCDERBY M., TURNER M. (Eds.): *Geographic Visualization: Concepts, Tools and Applications*. Wiley/Blackwell, Chichester, 4 2008. 3

[Jon16] JONES D.: Collaborative Computational Projects portal. STFC / Scientific Computing Department, Jun 2016. http://www.ccp.ac.uk/. 1

[MFT14] MORRIS T., FOWLER R., TURNER M.: Visualisation Matters Survey Results. Wiki on Computer Science, University of Manchester, Dec 2014. http://www.vizmatters.cs.manchester.ac.uk. 3

[Nag16] NAGELLA S.: Collaborative Computational Project in Tomographic Imaging. STFC / Scientific Computing Department, Jun 2016. http://www.ccpi.ac.uk/. 1

[NF16] NAGELLA S., FOWLER R.: Image Quantification Algorithms SourceForge. EPSRC collaborative software development environment, Jun 2016. http://ccpforge.cse.rl.ac.uk/gf/project/iqa/. 1

[SSS*01] SELVIN A. M., SHUM S. B., SIERHUIS M., CONKLIN J., ZIMMERMANN B., PALUS C. J., DRATH W., HORTH D., DOMINGUE J., MOTTA E., LI G.: Compendium: Making meetings into knowledge events. In *Knowledge Technologies* (Mar. 2001). http://www.compendiumng.org/about/. 2

[Wor13] WORTH D.: *Developing an image analysis plugin for VolView*. Tech. Rep. 7, STFC Rutherford Appleton Lab.), 6 2013. RAL-TR-2013-007. Persistent URL http://purl.org/net/epubs/work/65673. 1