Interactive visualization of multiscale biomedical data: an integrated approach

Debora Testi[1], Gordon Clapworthy[2], Stephen Aylward[4], Alejandro Frangi[3], Richard Christie[5]

[1]BioComputing Competence Centre, SCS srl, Italy
[2]University of Bedfordshire, UK
[3]Universitat Pompeu Fabra, Spain
[4]Kitware Inc, USA
[5]University of Auckland, New Zealand

ABSTRACT

Visualization of multiscale data is becoming increasingly important in all biomedical research projects. Huge amounts of data are being collected worldwide at all scale levels (from genes to body level). However, there is a clear lack of software tools to support the interactive inspection of these large data collections. A research consortium has been established to identify the challenges associated with real use case scenarios and to produce, in the next year, an open source library (MSVTK) to be used by any research project to effectively inspect and visualize multimodal, multidimensional and multiscale data. We will review here the challenges associated with the development of this visualization software library and the preliminary approaches under investigation.

KEYWORDS: multiscale, biomedical data, visualization.

INDEX TERMS: [Biomedical and Medical Visualization], [Data Fusion and Integration], [Multi-field, Multi-modal and Multi-variate Data], [Multidimensional Data]

1 INTRODUCTION

Recent years have seen an increasing trend in biomedical research towards the consideration of systemic processes. These are phenomena which are commonly observed in living organisms and which reflect the outcome of interactions between multiple sub-systems. In the past, the complexity of these problems had forced researchers to concentrate on individual sub-systems, and the most common boundary separating these has been spatiotemporal scale. However, the biomedical community has recently started to produce the first results of the so-called integrative approach, and a number of multiscale datasets have become available.

Their analysis has made evident that there is a shortage of appropriate tools for visualizing and exploring data that are defined across a broad range of spatial and/or temporal scales. At the same time, the number of biomedical problems that will demand multiscale visualization in the coming years suggests that this area should start to receive urgent attention but, surprisingly, it received almost no mention in the Visualization Research Challenges [1] document produced jointly by the USA National Institute for Health and National Science Foundation in 2006.

Multiscale visualization is not a new visualization issue and it has been investigated in other scientific contexts, such as information visualization with, perhaps, the most relevant work being undertaken in geographical data visualization, as exemplified by well-known solutions such as Google Earth. While these approaches are extremely effective within the context of their specific target problem, not all of these solutions can be generalized to other domains. In particular, many of the approaches do not translate well to the biomedical area, where datasets are usually of higher dimension and contain a greater variety of field data.

2 THE MSV PROJECT

Within this context, an international consortium has been established involving leading groups in visualization from three continents with the aim of providing solutions to these problems. The MSV[1] project, funded by the European Commission (EC) but also involving partners from the USA and New Zealand, will implement an open source software library for the multiscale visualization of, and interaction with, biomedical data, which will provide a suitable resource for the biomedical community to use in this rapidly evolving area. This library with be compatible with, and distributed alongside, the Visualization Toolkit (VTK)[2].

This objective will be achieved by:

- clearly defining the problem and its challenges;
- identifying exemplary problems and data to be used for the MSVTK testing;
- defining the best practices to address this type of problems;
- implementing the first open source version of the MSVTK library.

3 THE CHALLENGES

The identification of the challenges began with an international consensus meeting involving representatives from both research and industry. From the meeting outcomes, a review of the state of the art, and the internal discussion, it was confirmed that there is a need for a specialized software library to interact and visualize multiscale biomedical data.

To better identify real users' needs in terms of multiscale data visualization, partners have contacted EC-funded projects and NA-MIC initiatives. From the collected inputs, it has become evident that the MSV project has to address a number of challenges that have been classified and that have to be taken into account in the design and development phases (an example is given in Figure 1).

In detail, the MSV problem is defined by:

- information which will be:
  - at very different spatial and temporal scales, ranging from the molecule (nanometers and nanoseconds) up to the body level (meters and years);
  - in different forms (medical images, computer models, signals, etc.);
  - of heterogeneous dimensionality (1D, 2D, 3D, 3D+t);

[1] http://www.msv-project.eu
[2] http://www.vtk.org

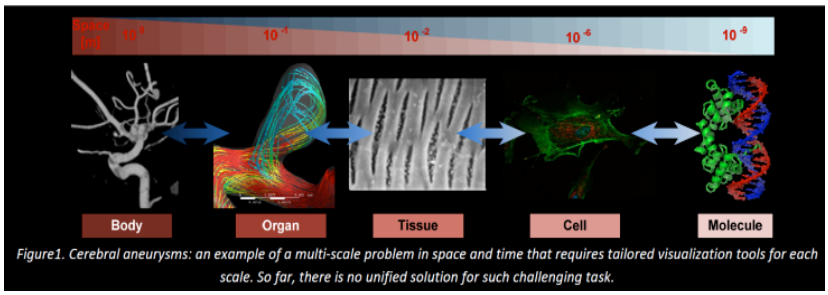

Figure1. Cerebral aneurysms: an example of a multi-scale problem in space and time that requires tailored visualization tools for each scale. So far, there is no unified solution for such challenging task.

- visualization which should be interactive even if very large volumes of data can be present at each or all scales;
  - depending on the problem at hand, there may be gaps in the scales (not all levels are available) but visualization should, nevertheless, be continuous across the scales;
  - data at different scales usually have different systems of reference: proper definition of the relative position or correspondence in a common reference system is essential for an adequate interpretation of this information.

While for some of the above-mentioned issues there are specific software solutions already available both under open source or commercial licenses, so far none has tried to address them all together in a unified manner.

4 Exemplary Problems

Associated to the challenges, a series of exemplary problems have been identified to test the MSV approach. The biomedical domains, for which MSV got access to data, are very different (cardiology, cerebral aneurysm, musculoskeletal modeling, neuroimaging, oncology, virtual colonoscopy); this will guarantee that there will be data available to test all different challenges and that MSVTK will be general enough to be potentially adopted by any biomedical research project.

The available data present the challenges as in the table:

	Cardiology		Cerebral aneurisms	Musculoskeletal modelling	Neuroimaging	Oncology			Virtual colonoscopy
Challenges	ex1	ex2	ex1	ex1	ex1	ex1	ex2	ex3	ex1
different spatial scales	x	x		x		x			
registration issues	x	x		x	x	x		x	
very large data		x	x	x					x
gaps between scales	x			x		x			
multivariate data	x	x	x	x	x	x			
heterogenous dimensiona	x	x	x	x	x	x	x	x	x
high dimensionality	x	x	x		x				
interactive visualisation		x		x				x	x
time varying issues	x			x					

5 Proposed Interaction Approach

Analysis of the tools available in biomedical and other domains showed that the zoom-based approach, previously investigated during the LHDL EC-funded project [2] has proven to be very effective as in many other application domains. This approach is user friendly as the user is provided with visual cues for the positions of lower scale data with respect to the whole data, which leads to an intuitive interface for data navigation.

The MSVTK will thus start from this interaction paradigm but will use placeholders not only for the representation of lower-scale data but also for hyperlinks to provide extra information such as documentation, etc. At the same time, MSV will also investigate the possibility of conveying meaningful information about the represented data through the icon/placeholder shapes and colors, with the aim to optimize information transfer and user experience.

As mentioned above, it is necessary to provide a means by which the user can understand the relative positions of the various parts of the data, some of which may be invisible at the current visualization scale: a data tree which allows also the data selection, or a small map on the corner showing which part of the big map the user is actually looking at will be evaluated for this purpose.

However, at all scales there may be large datasets, running into many GBs, so a proper management of these data types during the interaction has to be included. The MSV project will rely on previous work carried out to manage out-of-core data using a multi-resolution and bricking approach [3]. In particular, the possibility to apply the same approach to different data types is being investigated, together with effective integration with the zoom-based navigation approach. The solution will be implemented to be valid with the data available now (around 1GB per dataset), but general enough to be used in the future with larger datasets.

6 Ongoing Work

The project partners are currently defining the best practices to be put in place when dealing with the problem described in the challenges section. The implementation of the first prototypes will start soon and, in particular, will be aimed at testing the proposed interaction approach on data coming from the exemplary problems.

As mentioned above, the MSVTK library will be implemented as an extension of the specialized Visualization ToolKit (VTK) and the first two challenges to be addressed are the zoom-based approach to navigating among different scales and the large-volume fly through.

These prototypes will be used to collect feedback on the proposed interaction paradigm, and the next year will be completely devoted to the development of the MSVTK library, which will be released as open source at the end of 2012.

References

[1] C.R. Johnson, R. Moorhead, T. Munzner, H. Pfister, P. Rheingans and T.S. Yoo. NIH-NSF Visualization Research Challenges Report. IEEE Press, ISBN 0-7695-2733-7, 2006.

[2] M. Viceconti, G.J. Clapworthy, D. Testi, F. Taddei and N.M.B. McFarlane. Multimodal fusion of biomedical data at different temporal and dimensional scales. *Computer Methods and Programs in Biomedicine*, 102(3):227-237, 2010.

[3] A. Agrawal, J. Kohout, G.J. Clapworthy, N.J.B. McFarlane, F. Dong, M. Viceconti, F. Taddei and D. Testi. Enabling the Interactive Display of Large Medical Volume Datasets by Multiresolution Bricking. *Journal of Supercomputing*, 51(1)3-19, 2010.

Acknowledgments

This work has been partially funded by the European Commission within the MSV project (FP7-IST-248032).