VISUALISATION IN DIGITAL SCIENCE

New visualisation approaches are needed in digital data science to advance rapid information extraction and collaborative decision making

The size, high dimensionality and complexity of data being captured and manipulated in many scientific fields continue to increase. In particular, the modern digital era offers a deluge of complex data accessible via the internet. Understanding such data for a diverse range of applications requires progressively more advanced and powerful tools. Sometimes the scientist wants to address specific questions, either involving the statistics of large amounts of data or searching for special patterns in the ocean of data. The investigator may want to explore the data in search of unexpected phenomena, which requires human intelligence and interaction. Visual analytics, that is, the combination of automated data processing with human reasoning, creativity, and intuition, supported by interactive visualisation, is one of the prime methodologies that allow putting the human in the loop.

The Scientific Visualisation and Computer Graphics group of the Johann Bernoulli Institute is involved in several research initiatives that require advances in visual analytics and collaborative e-science environments.

Neuroimaging

In the medical area, the group's focus is on the analysis and visualisation of data obtained by structural and functional imaging. For example, it works in the area diffusion tensor imaging (DTI), an MRI-based technique enabling the visualisation of nerve fibres and structural connectivity of brain regions. In a case study, the group applied the method to study brain connectivity in relation to the phenomenon of tinnitus, which is an auditory percept in the absence of an external sound source. In the area of EEG analysis, data-driven network visualisations of high density EEG coherence were developed, which provide a quantitative measure of functional brain connectivity, and this approach was applied in studies of mental fatigue and ageing.

Recently, the group joined the GLIMPS project, a collaboration of a large number of Dutch clinical centres to collect and analyse FDG-PET scans for early-phase prediction and diagnosis of neurodegenerative brain diseases, such as Parkinson's disease. The goal is to identify distinctive structural and functional brain patterns and derived quantities like network patterns of brain activity, which can aid medical doctors in patient diagnosis.

Existing techniques are not sufficiently accurate for practical use, mostly because they are based on limited amounts of data. Many international efforts are currently underway to establish large databases of brain scans of patients with neurodegenerative diseases; the growing body of initiatives includes examples such as the International Neuroinformatics Co-ordinating Facility (INCF, http://www.incf.org) and the EU Human Brain Project (http://www.humanbrainproject.eu).

> **'Tools that combine machine learning and interactive visualisation are essential in the research stage for understanding the behaviour of machine learning algorithms…'**

The increasing availability of neuroimaging data makes a visual analytics approach based on pattern recognition, machine learning, and interactive visualisation feasible. The goal therefore is to develop and validate a more advanced system based on computer-based learning that obtains much higher accuracies for many different neurodegenerative diseases than currently possible.

Tools that combine machine learning and interactive visualisation are essential in the research stage for understanding the behaviour of machine learning algorithms, and also in the clinical stage for presentation of the predicted outcomes in an intuitive way to both doctors and patients. Therefore, this project aims to develop a self-explanatory, adaptive, system that uses the best visualisation techniques to show the doctor in a detailed and interactive way why the system outputs a particular diagnosis. Together with computer-assisted reasoning to compute confidence levels of the system for a particular diagnosis, this system would serve as a second opinion to the medical doctor.

There is a strong collaboration in this project between the University Medical Center Groningen (UMCG), the Faculty of Mathematics and Natural Sciences of the University of Groningen, the BCN-Neuroimaging Center Groningen (http://www.bcn-nic.nl), and the High Performance Computing & Visualisation centre (http://www.rug.nl/cit/hpcv), which is part of the Center for Information Technology of the University of Groningen. The database technology that is provided by the University of Groningen's Target group (http://www.rug.nl/target) is also heavily used.

Observational astronomy

Observational astronomy is a second application area in which the Scientific Visualisation and Computer Graphics group is active. This area is producing enormous data streams at an everincreasing rate. Modern astronomical surveys provide not only image data but also catalogues of millions of objects, each object with hundreds of associated parameters. New tools must be developed that can cope with the sheer data volume which has entered the petabyte regime. We have been collaborating for many years with the Kapteyn Astronomical Institute of the University of Groningen to develop new visual analytics techniques for exploring these high-dimensional data sets effectively.

Data mining, machine learning, and image processing techniques, coupled with multidimensional visualisation, are also used. Examples include high-dimensional subspace finding techniques, parallel co-ordinate and scatter plot matrix visualisation of high-dimensional data, decision tree classification and visualisation, multiscale network visualisation and analysis, perceptually enhanced search techniques, data skeletonisation, and large direct-touch interfaces such as tables and walls which support collaborative work.

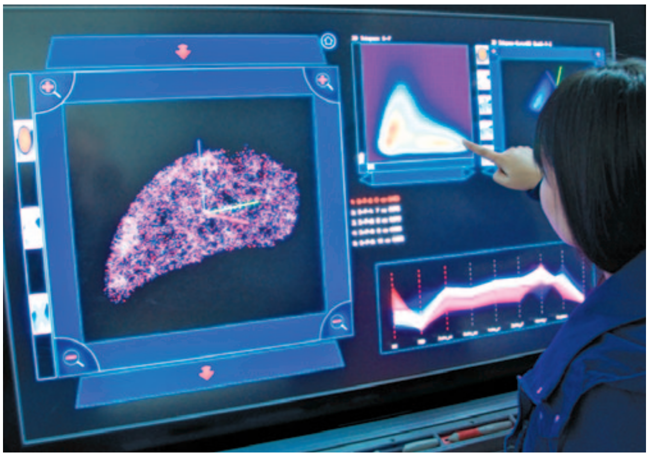
As an example, extracting information and finding unknown relations from large, high-dimensional astronomical datasets requires the identification of relevant subspaces of the data. A fully automatic algorithm applied to the full data set may miss interesting subspaces because they may be drowning in the noise. Human intelligence is needed to quickly assess potential subspaces proposed by an automatic search algorithm, and if needed adapt parameters of the search process.

e-Visualisation of Big Data

Recently, the group has embarked on a new research project, e-Visualisation of Big Data, funded by the Netherlands Organisation for Scientific Research and the Netherlands e-Science Center. Partners are the Johann Bernoulli Institute for Mathematics and Computer Science, the OmegaCEN Center for Information Technology of the Kapteyn Astronomical Institute, and the Donald Smits Center for Information Technology of the University of Groningen.

The long-term goal of this project is the creation of a distributed e-science environment, accessible through the internet to scientists and analysts from all disciplines, which allows fast evaluation, analysis, and decision making. In this project the team works with scientists from areas with specific requirements on big data visualisation. The primary two areas are medicine and astronomy in which there is already extensive experience, as outlined above. Other potential applications are real-time handling of 3D electron microscopy archives, processing sensor data for monitoring the quality and possible critical state of dikes in the Netherlands, or managing electricity grids under conditions such as power plant failures and unexpected weather conditions.

What is common to all these cases is the need for interactive query and visualisation mechanisms that can be used over the internet to support scientific discovery and rapid decision making. The project also aims to operate 'decision theatres' which allow small to large groups to participate in the decision process by utilising visualisation environments, ranging from large touch-



sensitive displays to a dome theatre (Infoversum Groningen) which is now under construction.

The Target group currently hosts the processing, mining and dissemination of two large astronomical imaging surveys using a new generation of astronomical imaging arrays at ESO's Paranal Observatory, Chile. These utilise the OmegaCAM 275 megapixel optical camera at the VLT Survey telescope, which will perform the KiDS 1,500 square degree survey, and a 67 megapixel infrared camera on the VISTA telescope, which executes the VIKING survey on the same area. The data of the KiDS and VIKING surveys are stored and processed by Target. Several petabytes of data will be produced in this imaging survey resulting in a 9-band catalogue (from UV to near infrared) with hundreds of millions of sources with dozens of parameter values. Ensuring that produced data is of scientific quality with these large data volumes and variable atmospheric conditions requires advanced and efficient visualisation tools. These tools and their results must be at the disposal of the scientist in a typical e-science environment.

The Scientific Visualisation and Computer Graphics group aims to expand its local and national e-science collaborations at the European level within the Horizon 2020 programme. Several key priorities of this programme match perfectly with its ambitions: advancing digital science through creating e-infrastructures and collaborative environments; public private partnerships; and multidisciplinary collaborations to address several societal challenges: sustainable society, energy, and healthy ageing, which are the three primary focus areas of the University of Groningen.

university of groningen

Professor Jos B T M Roerdink
Scientific Visualisation and Computer Graphics
Johann Bernoulli Institute for Mathematics and Computer Science
University of Groningen

tel: +31 50 363 3931

j.b.t.m.roerdink@rug.nl
www.cs.rug.nl/svcg/