Converging Computing Methodologies in Astronomy: A European Science Foundation Scientific Network, 1995–1997

NETWORK COORDINATION COMMITTEE (names below)

Introduction

A "Scientific Network" is the primary funding mechanism used by the European Science Foundation, which is based in Strasbourg. The word "network" is used here to mean a group of research teams and individuals, which aims at coordinating their research work through workshops and conferences, and work visits. A few dozen such "networks" are supported by the ESF at any one time. Such collaborative "networks" have in the past been set up in many areas, spanning the natural and human sciences. They have proven very successful, as springboards for productive work, with a relatively very restricted amount of administrative overhead.

An ESF "Scientific Network" on the topic of "Converging Computing Methodologies in Astronomy" is beginning in January 1995, and will extend over 3 years. As is usual in ESF collaborative networks, the administration of the network is carried out by a committee which is geographically representative.

The increasing integration in an astronomical setting of various methodologies in pattern recognition, information retrieval and data analysis is a characteristic of modern computational astronomy. Large astronomical data-bases and archives from space-borne missions, ground-based observatories, and large wide-field surveys, provide the technological infrastructure.

Against this backdrop, boundaries between theoretical domains have been lowered.

In this project, we wish to highlight and exploit convergence of methods and techniques.

From the astronomer's viewpoint, a coordinated global view of methods and techniques allows optimal exploitation of available information.

Network Activities

- *From vision models to image information retrieval*

Methods such as wavelets and multi-resolution approaches, mathe-matical morphology, and fuzzy methods have proven their worth in the framework of accessing appropriate information from large image databases. Use of such methods in this context is not a purely engineering task (i.e. simply providing them as stand-alone commands in an information retrieval framework). Rather, the methods must be moulded together to allow semantically-driven access to data. New methods, motivated by those mentioned, are required to handle the huge quantities of data which stand ready to be analysed.

Example: the Hubble Space Telescope image database contains roughly 80,000 2-dimensional and 1-dimensional images. Certain types of survey studies could benefit from the ability to access all images satisfying certain broadly-specified characteristics: e.g. objects close in appearance to a certain morphological shape; or objects of a certain morphology within a certain distance of one another. Can multi-resolution and fuzzy characterisation methods provide the basis for such retrieval requests?

- *The data life-cycle methodological aspects*

The astronomical data life-cycle is highly digital: data capture is increasingly on CCD electronic detectors, data are subject to image processing and statistical treatment, and the final major stage in this process involves (a) data archiving, and (b) publication. Not surprisingly, the issues of electronic publishing and of digital libraries are viewed as increasingly important. Data storage and access standards, together with interface standards, must be coordinated. Data characteristics, including dynamic aspects related to the life-cycle, strongly impact on methods for analysis and treatment.

Example: Image data are nowadays stored in many versions: raw; calibrated; compressed with loss of information, for quick-viewing; perhaps iconised; and we may add subsample-selected for data samplers and image processing test suites. Not only is coordination called for, among these incarnations of database contents; but – more than that – we must ask if other derivative products are needed, and are easily produced from the afore-mentioned chains of analysis and reduction.

- *From data integration to information integration*

– Particular data integration (data fusion) problems, such as integration of data associated with different wavelength ranges, are of great relevance in the context of large space- and ground-based observing projects. In certain instances, this takes the form of co-addition in image restoration. It also includes the enhancement of image restoration and fil-tering approaches through building-in semantic information on the cosmic objects of interest. Close, complementary use of multi-million object astronomical catalogues still has some distance to go in order to be tightly coupled with analysis of images from large image databases. Complementarity of such data rests on external information (semantics), and determines the appropriate analysis methodologies. Large-scale data integration prototypes have been pursued by NASA and ESA in recent years (respectively, the Astrophysical Data System, ADS; and the European Space Information System, ESIS), and these serve to point to 'issues and concerns' which are solvable through improved methods.

– Classification of terabyte (10^{12} bytes) data collections, which includes neural networks, decision/classification trees, and machine learning approaches. Widely reused code and paradigms has not yet evolved out of extensive astronomical work done in this area. Through coordination, this could be the case.

– Long-term access to stored data, and the selection of the latter. What should be, in the phrasing of M.J. Kurtz, 'the future of memory' (i.e. humanity's memory, which includes image data and other data)? Ever greater quantities of data are collected – what should be (expensively) stored, and how should this be done? Photographic plate directories exist, and policies based on a range of theoretical issues are required for scanning and maintenance. Major projects

collect many terabytes of data, and reuse is both necessary but also problematic.

– Beyond data, astronomy is all about information. Compression is a by-product of image models. Loss of information involved in compression raises a range of questions. In its broadest sense, compression is summarisation, and therefore is part of the overall process of scientific analysis.

Example: The costs of long-term data storage are large. With greater and better ground-based and space-borne observatories, these costs cannot be ignored. With our knowledge of, and expertise in, state-of-the-art and emerging computational technologies, what is our recommendation to the world's space agencies and observatories in regard to future policy?

These thematic themes of the Network will be axed around a series of workshops.

A conference on the Network's overall theme will be held shortly before the end of the Network, with the aim of bringing the results obtained to a wide audience. A selected set of papers will be published in book form.

The Coordination Committee and Contact Information

The Coordination Committee consists of:

A. Bijaoui, Observatoire de la Côte d'Azur, Nice (bijaoui@corelli.obs-nice.fr)

V. Di Gesù, Department of Mathematics and Applications, University of Palermo (digesu@ipamat.cres.it)

A. Heck, Observatoire Astronomique, Strasbourg (heck@cdsxb6.u-strasbg.fr)

M.J. Kurtz, Center for Astrophysics, Harvard (kurtz@cfa.harvard.edu)

P. Linde, Lund Observatory, Lund (peter@astro.lu.se)

M.C. Maccarone, IFCAI, Palermo (cettina@ifcai.pa.cnr.it)

R. McMahon, Institute of Astronomy, Cambridge (rgm@mail.ast.cam.ac.uk)

R. Molina, University of Granada (rms@ugr.es)

F. Murtagh, European Southern Observatory, Garching (fmurtagh@eso.org)

E. Raimond, Netherlands Foundation for Research in Astronomy, Dwingeloo (exr@nfra.nl), ESF representative

The Coordination Committee decides on administrative matters, but needless to say, many others will be involved. The Network Chair is

M.C. Maccarone (cettina@ifcai.pa.cnr.it),

and the Secretary is F. Murtagh(fmurtagh@eso.org).

Information on the Network can be obtained on the WWW at the address: http://www.hq.eso.org/conv-comp.html.

ESO on CompuServe

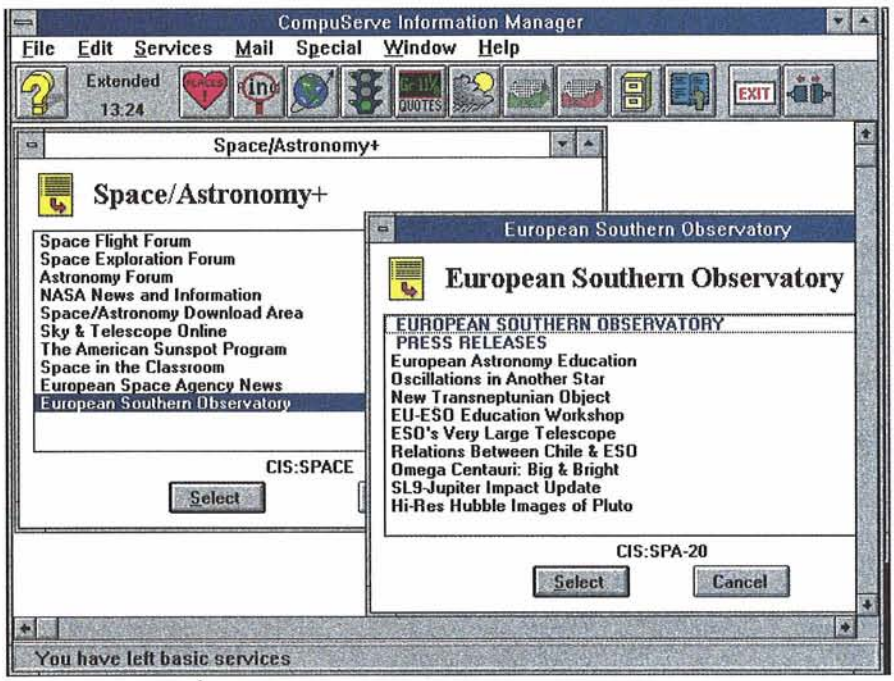
F. MURTAGH, ST-ECF, Garching

ESO press releases and a great deal of other ESO information are available on the World-Wide Web (URL: http://www.hq.eso.org/). For over a year, press releases including images are also available on CompuServe. They may be accessed in the Space/Astronomy area (GO SPACE).

CompuServe is a commercial information service, offering local access at many locations worldwide. Membership is about 2.5 million worldwide, of which around 170,000 are in Europe. A very large number of discussion groups and information libraries are available which span all subject domains. Other areas cater for weather images and forecasts, news-feeds from major agencies, a range of databases, and so forth. A number of user interfaces (browsers) are available which support access in a convenient manner, display of images and videos, e-mail, file management of download areas, and navigation through the varied services on offer. Access is often by modem, using dial-up to a local telephone number.

In some respects this is like what the Internet offers, with newsgroups, anonymous ftp and WWW. In fact, it can be claimed with justice that the Internet is slowly groping towards the type of service which has long been available on systems like CompuServe. Only recently, with the emergence of new, reliable and user-friendly toolsets, has the Internet become easily accessible to those who are not networking enthusiasts. The Internet has still some way to go, in terms of ease of access to relevant information.

Especially in Europe, where home and personal connectivity to the Internet is not very high as yet, CompuServe offers an excellent medium to allow for wide



A view of CompuServe, as seen on a Windows notebook, showing the Space/Astronomy area and ESO press releases. Items are clicked on, to access contents. ESO images are available in the area's libraries.