VISUAL INTERACTION WITH VERY LARGE SPATIAL DATA SETS

Wolfgang Walcher
Institute for Computer Graphics, Technical University Graz, Austria
E-mail: walcher@icg.tu-graz.ac.at

ISPRS - Commission I, Working Group 6

KEY WORDS: Satellite, Image, Database, Visualization, Software

ABSTRACT

The number of remote sensing instruments as well as the size and quality of the generated data is increasing continuously. Electronic catalogues produced by the different data providers are very important tools for search and retrieval of remote sensing images. The World Wide Web within the last years became the most important means of providing this kind of information to the world-wide user community. The HTTP protocol and general problems with the bandwidth on the Internet limits some functionality in these information systems. New Java™ based network software overcomes these restrictions and makes it possible to implement a new generation of tools for the interaction with image and data catalogues. This paper describes typical problems with today's stand-alone or WWW-based image and metadata information systems. It also describes what kind of data visualization is needed to ease the user's interaction with these systems and how these visualizations can be implemented using the Java programming language. To illustrate the proposed concepts a new kind of data catalogue interface for SAR radar image data was implemented to support the analysis of data from the Magellan mission to Venus. The paper also offers an outlook on how this new kind of network software can be used to integrate so far unknown functionality into upcoming WWW-based information systems for remote sensing data.

1 INTRODUCTION

1.1 Background

Within the past years an increasing number of sensors along with an also increasing geometric and radiometric resolution of the resulting images are producing an almost exploding amount of remotely sensed digital data. Ambitious programs like NASA's 'Mission to Planet Earth' are boosting this development even further. While in general users should benefit from such better datasets at greater accuracy, it becomes more difficult for that user to find the very data best suited for a given problem or application.

Electronic catalogues and information systems are being set up by the private or public organizations which provide remote sensing data. But also the number of these catalogues is increasing in a way that it has become necessary to introduce catalogues of catalogues like the International Directory Network (IDN, see CEOS, 1994) or the European Wide Service Exchange (EWSE, 1995). In general catalogues closer to the date producer contain more detailed information, even images. Systems like IDN offer mostly metadata and descriptive information about entire datasets.

Only completed missions, like those to other planets have a fixed-size data set. Otherwise perpetual growth of remote sensing image archives, requires that all catalogues be updated continuously. To make these rapid changes visible to the user, on-line access is a must. Therefore the Internet is becoming the most important medium to present remote sensing data to the world-wide user community.

1.2 Image catalogues

Special purpose software tools exit to communicate with a given remote sensing archive and are being distributed to potential customers of these data. Typically this software is executable on a single or very few hard- and software environments. But since it was written using a high level programming language any desired functionality can be implemented. Disadvantages of this concept are the limited variety of supported hardware and software environments and the need for separate software tools with different capabilities and user interface to access every single image data archive.

Providers of remote sensing data therefore are increasingly using World Wide Web based catalogue systems. The two main reasons for this development are that today's WWW browser software is inexpensive or even free of charge, and browsers are available for almost any hardware and software environment. The ease of use of a hypertext browser and the general boom of the WWW are also reasons for this trend.

Offering a WWW-based retrieval client can increase the number of accesses to a catalogue system almost exponentially. With the introduction of the WWW interface for the IDN at NASA's Global Change Master Directory (GCMD), the average number of requests per day was increasing from about 200 in August 1994 to almost 4000 in February 1996.

1.3 Standard functionality of catalogue systems

There are four basic components from which catalogue systems are usually built. Only components 1 and 4 are necessary to create a fully operational system, whereas all of components 2 and 3 or only parts are sometimes missing.

1 Query definition and retrieval of results is the basic function of all catalogue systems. This part of the system is responsible for the generation of queries to the underlying database. In WWW-based systems HTML forms are used for the definition of the query at the users site and CGI - scripts to implement the interface to a relational database.

2 Map or navigation tool. Queries for image coverage usually require the definition of a point (POI) or a region of interest (ROI). A map-like representation of the area covered by a dataset, can ease this task considerably. With the possibility to move around and zoom in and out this map regions or points can be defined by point and click operations. To visualize the content of a database or the results of a query, one may also overlay the geographical information with outlines and information about any feature with a position or some spatial extent. This can be data like images, data gaps, topographical features and names, orbit data and many others.

3 Preview of data. To offer the user the possibility of inspecting the data before actually placing an order, one may have quicklooks or reduced versions of images available for downloading. This feature can also save network bandwidth since it avoids the download of unneeded data. Also for copyright reasons it is often not possible to provide full resolution images before payment is guaranteed.

4 Ordering data is the most important action which data providers want users to do. Therefore it is always easy to use this function, whereas distribution of data in most cases is done the slow and traditional way by mailing a tape or CD-ROM.

2 EXISTING SYSTEMS

2.1 Stand alone software

As a marketing instrument and for user service data providers have developed interface software to their archives and catalogues. Early examples for this software were terminal based with a commandline or menu interface. Some of these products are still in use, even though newer versions with graphical user interfaces or WWW-based counterparts are available today. An example for this live cycle is the ISIS tool from DLR. The graphical version of this tool is GISIS (Graphical Intelligent Satellite Information System, Lotz-Iwen, 1995). It features all four components of such a system as well as some additional highlights, like a thesaurus for geographical names. The navigation tool, which is based on a software product form Xerox, provides a zoomable vector based map at different levels of detail. GISIS is also a multi-catalogue system and offers interfaces to different data sets and directories. Even though this is one of today's most advanced systems, the number of accesses has been almost stable over the past 15 months.

This example shows, that a user often avoids to install and maintain dedicated software products for different data catalogues, especially if one does not plan to use them frequently. Compared to the previously described access numbers of the IDN interface at GCMD, this clearly indicates, that for the user acceptance of a catalogue, the ease of access can be more important than the number of available functions.

Some other examples for special purpose systems are VISTA (Visual Interface for Space and Terrestrial Analysis, Snyder, 1994) and the Arno project (Nativi, 1995). VISTA is very special, since it supports a 3D vector representation of data like spacecraft ephemeris, sensor positions and the surface of the Earth.

2.2 WWW-based catalogue systems

Because of the reasons described above WWW-based image information systems are playing an important role today. The basic idea of the WWW was to present hypertext information containing text and images stored in HTML files to user. Build in hyperlinks allow easy navigation to other HTML documents and access to other Internet tools like ftp and e-mail. This very restricted set of functions is designed to present information to a rather passive user. To implement interaction with the user all user inputs have to be transmitted to the location of the WWW server, where a new HTML file has to be generated according to the user inputs and transmitted back to the user's site.

The example of drawing a mark at the location of a mouse click over an image displayed in an HTML document, shows the resulting intensive network traffic. First the coordinates of the mouse click have to be transmitted to the WWW server. The

Image	LL Lon	LL Lat	UR Lon	UR Lat
f00n037	33.74425676124	-2.54450642036	39.56692703798	2.545216581245
f00n042	39.36925676124	-2.54450642036	45.19192703798	2.545216581245
f00n048	44.99425676124	-2.54450642036	50.81692703798	2.545216581245
f00n054	50.61925676124	-2.54450642036	56.44192703798	2.545216581245

Fig. 1a: A few lines from the dump of a database table containing image names and positions. It is somehow easy to see that the images are forming a row. Its much harder to see the overlap between successive images.

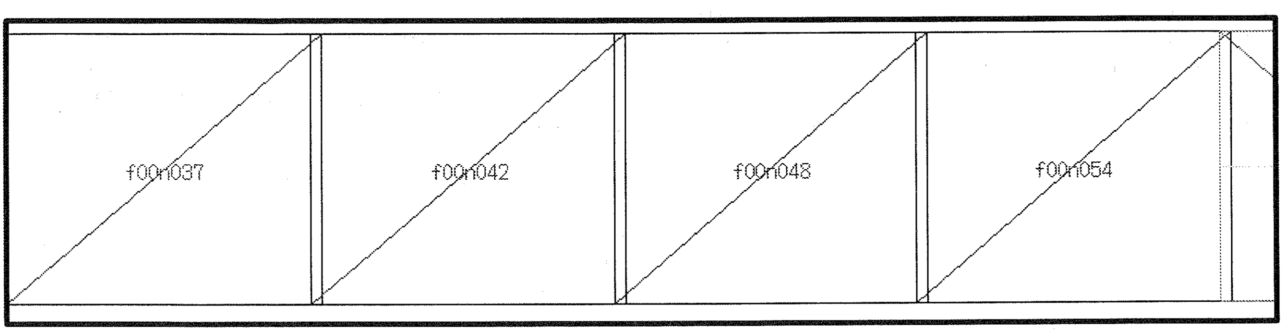

Fig. 1b: Enlarged portion of a screenshot showing the positions of the images listed in Fig. 1a. The small overlaps are clearly visible.

server has to create a new image with a mark at this point, to create a new HTML document containing the new image and to transmit it back to the user. This prohibits functions like rubberband drawing. Therefore most functions of a map or navigation tool, as introduced above, can not be efficiently implemented using standard HTML. The introduction of technologies like Java (by Sun) and Javascript (by Netscape), represents for the first time the possibility to locally process user inputs to an HTML document. These topics are covered in detail in chapter 4.

Examples for WWW-based catalogue systems are ImageNet from CORE Software Technology Inc. (CORE, 1995) that supports data access via WWW as well as via a special purpose client. The Landsat/Spot browser by the Canadian Center for Remote Sensing and Earth Observation (CEONet, 1995) and the WWW Ionia AVHRR Net Browser (Mungo, 1994) should also be mentioned.

While all these providers, and many others as well, still have different catalogues and information systems, only a single piece of software, the WWW browser, has to be maintained by the user. This single tool gives access to a variety of different catalogue systems.

3 WHY VISUAL INTERACTION

The map and navigation tool can be seen as the visual and interactive component of an image archive catalogue. It can provide two important functions: navigation and inspection. Navigation is important for the user to localize a point of region of interest, whereas the visualization of the database contents or of the results of a query can speed up the search for data dramatically. While there seems to be no doubt that the navigation tool is useful to define spatial queries, some examples are given below to prove the importance of the inspection functionality.

Even a skilled remote sensing expert has difficulties to determine the exact position or overlap of several images by looking at the coordinates of the image corners (see Fig. 1). Drawing the borders of images over a map makes it easy to understand the exact image coverage. This can be very helpful when searching for temporal series of data covering a given area, since most sensors don't have exactly overlapping orbits. In the case where more than a desired number of images covers the region of interest, one can use previewing of the images to inspect the actual image content. Depending on the resolution of the quicklooks, clouds and other problems with the data can be detected. The visualization of image extents can also be extremely helpful when searching for image pairs for stereo processing or mosaicking, or for data from different sources to perform data fusion over a given area.

For some more advanced applications navigation can be an important problem. This occurs especially over unknown areas where insufficient information is available to draw a detailed map. Datasets from other planets and moons are usually of that kind. An image map created from a mosaic of raster images can be used to visualize the planet's surface. Examples for such systems are the Mars Atlas (Batson, 1995), the Clementine Navigator (Clementine, 1996) for images from the Moon and the Interactive Venus Atlas described in Chapter 5.

Because of the restrictions of HTML, only image map based navigation tools and download of quicklooks could so far be implemented in WWW-based image retrieval systems.

4 NEW SOFTWARE TECHNOLOGIES

4.1 Java™

Java is an object oriented programming language developed by Sun Microsystems, Inc. (Java, 1996). Version 1.0 of Java was released in the 1st quarter of 1996. Java has a C++ like syntax

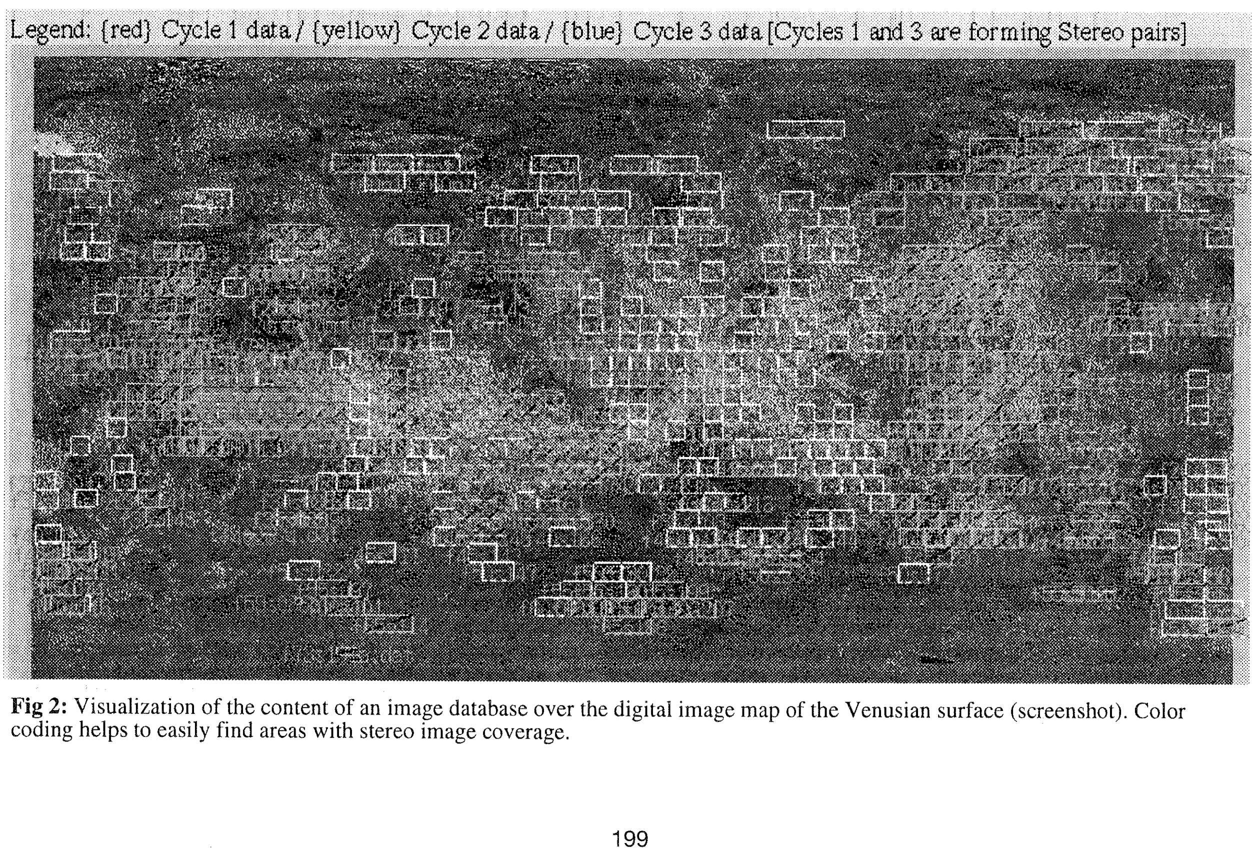

Fig 2: Visualization of the content of an image database over the digital image map of the Venusian surface (screenshot). Color coding helps to easily find areas with stereo image coverage.

but the code is compiled to an intermediate code interpreted by a virtual machine. Using a virtual machine, Java code is fully hardware independent. What makes Java so important, is the built-in support for network access via HTTP. Java programs can fetch code and data over the network at execution time. A Java program can not only execute code on the local computer but also on the computer it was loaded from, by using Java remote objects. To guarantee security a number of configurable security options was integrated into Java. Since Java is interpreted at runtime, it's execution is slower than C programs, for example. But for most applications the performance is still sufficient.

Full support for graphical user interfaces, image processing and graphic routines exists to implement state of the art search and retrieval software for image archives. With Java for the first time ever, it becomes possible to implement platform independent retrieval systems with all the functionality only stand alone software offered so far. Netscape, Microsoft, SGI and other hard and software companies have licensed Java to integrate it into their network products. A recent Java extension to interface popular relational database systems like Sybase, Oracle or Informix with native Java calls, makes Java a very useful tool to implement any WWW-based database retrieval client.

4.2 Applets

Netscape is the first company to integrate a Java interpreter into the latest version of it's WWW browser, the Netscape Navigator 2.0. Netscape also extended the HTML language to make it possible to include Java programs into an HTML document in the same way images are included. These included Java applications are called applets.

This combination of Java and the Netscape Navigator helps to integrate search and retrieval software into HTML documents. Once this document is loaded by the browser, the applet is downloaded and executed on the local computer. Using the caching facilities of Netscape, the applet will be stored on the local filesystem. Reloading the program is not necessary until the code has changed at the server site. In the case of repeated access this will reduce the required network traffic to a minimum.

Data providers still can offer their special purpose products without the problem of porting software to different platforms or expensive distribution of software upgrades. The user has to maintain only a single software product, the Netscape Navigator so far, and benefits from the functionality of software written in a high level programming language. As long as there are no compulsory style guides like those for Motif applications, there will still be a broad variety of different looks and feels.

4.3 Javascript

Netscape also integrated a new script language called Javascript into the latest version of it's navigator. Javascript is comparable to the macro language of other applications and has nothing to do with Java besides it's C- like syntax. There is also no compiler for Javascripts. The code is embedded into the HTML document and executed when it is loaded.

5 AN EXPERIMENTAL SYSTEM

5.1 Motivation

Parts of the dataset generated from the Magellan spacecraft's mission to Venus are available at the European Magellan Data Node (EMDN) at the University of Technology in Graz, Austria. A goal of this project is to provide access to these data to the European scientific community. To reach this goal an image retrieval system has been implemented stepwise by different student projects. This also serves as a testbed for the development of a Java-based information system. It is still an evolving project where modules are replaced or improved

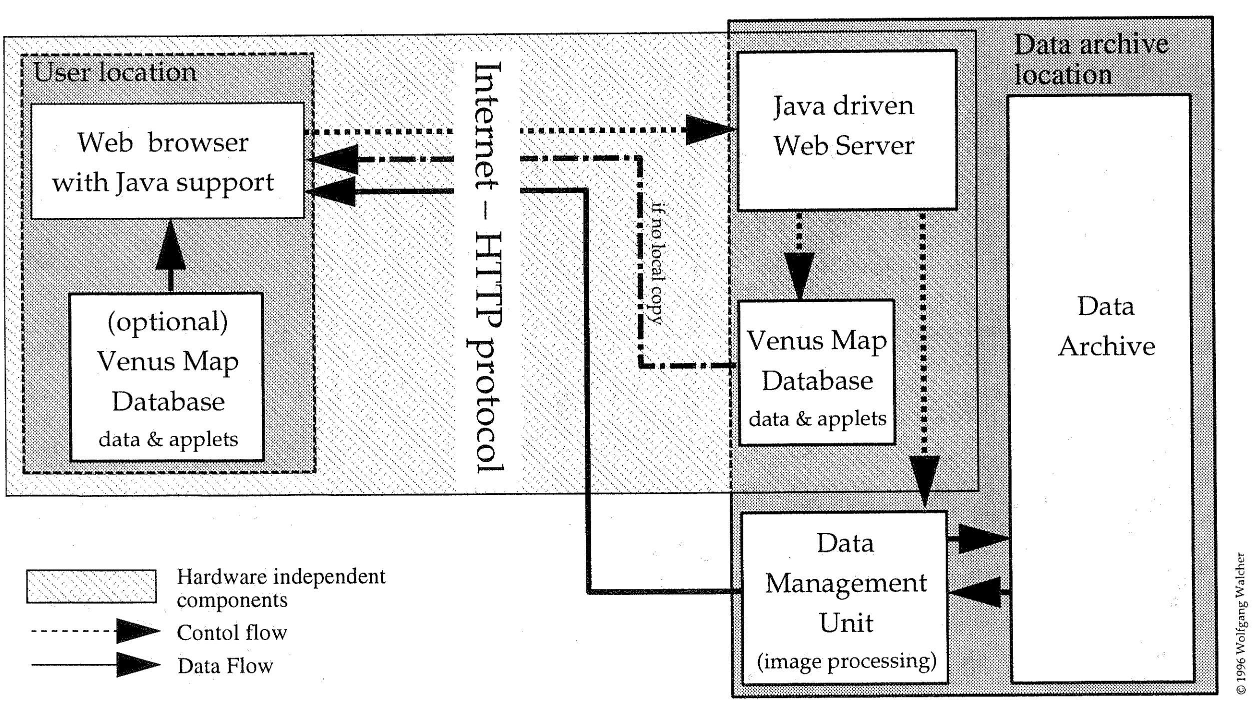

Fig 3: Layout of the experimental system.

frequently. For a schematic layout see Figure 3. Since Java is used as programming language, the entire application can be accessed via the WWW, using the Netscape Navigator without installing any additional software.

While all standard functions according to 1.3 are necessary for such a system, the navigation problem is of even more importance for extraterrestrial data. In this case Venus has about the same size as Earth. But since there is no water, the total area of land is three times bigger than on Earth. On the other hand there are only a few hundred named features, mostly craters or mountains, which are spread out irregularly on the Venus surface. This is by far not enough to draw a map that can be used to localize the content of an image. When searching for particular topographical features, a detailed representation of the planetary surface is needed. Up to now this was usually done using large paper printouts of the digital data.

The ground coverage of planetary image data sets is often much sparser and therefore it becomes an important criterion when selecting test areas. Errors and gaps in the data are also common due to sensor malfunctions or transmission problems. Visualization of image contents and position combined with data gaps and sensor characteristics, makes it easy to select test areas with 'good' image coverage or image pairs for stereo processing (see Fig. 2).

5.2 The dataset

The dataset at the EMDN consists of about 90 Gbytes of image data and some additional 120 Mbytes of metadata about the images, spacecraft and sensor characteristics as well as feature names on Venus. Image data are archived on about 150 CDs, the metadata are maintained by a relational SQL database. No spatial indexing scheme is used so far, but an R-tree based data structure (Guttman, 1984 and Samet, 1990) is under development to speed up queries for coverage and location. Details about the dataset are presented in Tables 1 and 2.

Type	Number	Total Size (Gbytes)
FBIDR	110	15
F/C-MIDR	1250	70
non SAR data	10	1

Table 1: Image data from Venus at the European Magellan Data Node. For detailed information on the data see http:/nssdc.gsfc.nasa.gov/planetary/magellan.html.

Type	Mbytes	Remarks
Image position data	15.0	also of data not at EMDN
Engineering data	110.0	orbit information
Gap Information	1.3	list of all image gaps
Gazetteer	0.1	named surface features
Crater database	0.1	named and unnamed

Table 2: Metadata at the European Magellan Data Node.

5.3 Interactive Venus Atlas

This is the main part of the system where all the user interaction takes place. The main functions are:

- Navigation,
- Definition of ROI and POI,
- Visualization of database content or retrieval results,
- Selection of data to preview or order.

The Venus atlas is able to display the entire known surface at 8 different resolution steps from 225 m/pixel to about 20 km/pixel. The image data are stored in an image pyramid which was generated from NASA/JPL C1-MIDR data with a pixel size of 225 m. To ease the user navigation the data have been resampled from sinusoidal to cylindrical projections to guarantee that north always points up and south points down. Because of the distortions of the cylindrical projection the polar regions had to be treated separately and are stored in an oblique sinusoidal projection. To reduce the size of the pyramid a radar speckle removal algorithm was applied before JPEG compression was used to store the images. The entire digital map pyramid is about 600 Mbytes and fits on a single CD-ROM.

This application takes full advantage of the graphical capabilities of Java. Only the JPEG compressed image framelets are fetched from the WWW server and assembled to a single image map. All user interaction is monitored by the local Java applets which are also responsible for generating the vector graphics overlay. As an option program threads running in the background can preload additional data in advance to speed up the assembly of the image map.

5.4 Database interface

The interface to the relational database for metadata is also written in Java and is integrated into the user interface of the Interactive Venus Atlas. It's main purpose is to create SQL queries and to transmit them to the database engine. Retrieval results are treated in two different ways. The data are first transformed to a table of human readable text and than all geometric information is extracted and transformed to a vector data set for display over the Venus atlas. This is no general purpose tool and can not be used for other datasets so far.

6 CONCLUSIONS

Traditional query and retrieval programs and interfaces for access to remote sensing data archives and catalogues have different disadvantages. Stand-alone programs are usually dedicated developments for a single archive and are also available only for a few hardware platforms. HTTP based systems accessible via a WWW browser do not provide all the needed functionality due to the restrictions of HTML.

With the recent advent of Java and it's integration into a WWW browser like the Netscape Navigator, the above problems can be avoided. A range of interactive programs is accessible from any platform via WWW. This concept was used to implement an Magellan data catalogue and an Interactive Venus Atlas.

In future applications the remote processing capabilities of Java could also be used to perform arbitrary manipulation of data at the archive site before transmitting them to the user. This would extend the database to a 'method-base', where on- demand generation of data products can be invoked by remote (and paying) users.

The first official version of Java was released in the first quarter of 1996, yet there are good reasons that this technology will become a standard for the development of any information system on the Internet within a very short period of time.

7 ACKNOWLEDGMENTS

Parts of this work were funded by Project 1/task 4 (Parallel Processing Strategies For Large Image Data Sets) of the Austrian 'Fond zur Förderung wissenschaftlicher Forschung' (FWF) research program 'Theory and applications of digital image processing and pattern recognition' and by the Space Research Program of the Austrian Academy of Sciences.

8 REFERENCES

Batson, R. M., 1995; Digital Maps of Mars; Photogrammetric Engineering & Remote Sensing. Vol 61, No. 12, Dec 1995, pp. 1499-1507

CEONet, 1995. Canada Center of Remote Sensing and Earth Observation, Landsat/Spot Browser, http://www.ccrs.nrcan.gc.ca/ccrs/imgprod/imgprode.html

CLEMENTINE, 1995; NASA Planetary Data System Clementine Navigator Development Team; http://cdwings.jpl.nasa.gov/cgi-bin/Clem.pl

CORE, 1995. CORE Software Technology, Inc.,"ImageNet" - Testsite and Information Page, http://coresw.com

EMDN, 1995. European Magellan Data Node at the ICG (Institute for Computergraphics) University of Technology, Graz, Resources on Magellan Mission to Planet Venus, http://www.icg.tu-graz.ac.at/PDS/emdn.html

EWSE, 1995. European Wide Service Exchange, Clearinghouse for Remote Sensing Data in Europe, http://ewse.ceo.org

Guttman, A. ; 1984; R-trees: A dynamic index structure for spatial searching; Proc. of the ACM SIGMOD; 1984/06; pp. 47-57

IPA, 1995. Interactive Planetary Atlas, http://cdwings.nasa.gov /PDS/public/IPA.html

Java, 1996; Java Documentation; http://java.sun.com/doc.html

Lotz-Iwen H.J., 1995. Earth Observation User Information Services in Germany. EEOS Workshop on European Data Networks and Earth Observation User Information Services, Proc. Document refs: CEO/130/1994; pp 359-363.

Mungo F., Arino O., 1994. AVHRR CD-Browser Ionia 1km net browser. Proc., 6th International Symposium on Physical Measurements and Signatures in Remote Sensing, ISPRS, January 17-21 1994, Val d'Isere.

Nativi S., Giuli D., Pellegrini P.F., 1995. A distributed multimedia information system designed for the Arno Project. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 50 Nr.1 1995, pp. 12-22.

PIA, 1995. Planetary Image Access, NASA service to provide users with publicly released images from various Solar System exploration programs, http://acheron.jpl.nasa.gov/PIA/PIA.html

Samet, H., 1990. Applications of Spatial Data Structures. ISBN: 0-201-50300-X, Addison-Wesley.

Snyder W., 1994. Metadata Management and the VISTA System. HICSS-27 Conference of System Sciences, DSS/Knowledge-Based Systems, Vol.III, pp. 418-427..