Design of a Distributed Database for the Magellan Data Set from Planet Venus

Wolfgang Walcher, Herwig Rehatschek
Institute for Computer Graphics, Technical University Graz, Austria
A-8010 Münzgrabenstraße 11 Tel: ++43 316 84 1766-0 Fax: ++43 316 84 6304
walcher@icg.tu-graz.ac.at, herwig@icg.tu-graz.ac.at

Abstract
Huge archives of images and other data from missions to other planets are available from various sources. But searching for specific features on a planetary surface can be very tedious. As with other remote sensing data it is often hard to determine the exact image contents before placing an order and very often the size of the ordered data exceeds the area of interest by far because of the underlying data formats. To ease the access to the Magellan data set for the worldwide scientific community we propose the Image Server System (ISS). This system is capable to maintain an inventory of one or many data archives. This database also holds quicklooks of all images and a map of the known areas of the planet. It can be searched via Internet with a special client software that is capable to visualize the planetary surface, any search results and the quicklooks of all available images. The user can order any subset of the available data. The server retrieves the ordered data from the archives and composes a data set that has exactly the extent defined by the user. These data stacks are delivered to the user on-line or on a tape.

Introduction

The Institute for Computer Graphics (ICG) at the Technical University Graz, Austria (TUG) participates in the NASA Planetary Data System (PDS) as the European Data Node (EDN) to the Geoscience Node. The EDN has to provide access to the Magellan data set for the European scientific community. The proposed Image Server System (ISS) is designed for this very reason and will be implemented at the ICG within the next two years.

During the Magellan mission to Venus about 98% of the planets surface was mapped by the spacecraft's sensors. The data set consists of SAR images (>5200 orbits) as well as global emissivity, reflectivity, slope and topographic data with a total volume of about 400 GB. The Magellan SAR provides image strips which are about 350 x 220000 pixel in size at a resolution of about 75 meters [1]. These Full resolution Basic Image Data Records (F-BIDR) correspond to a ground size of approximately 20x17,000 km. The orbits (polar, north to south) are grouped into three cycles defined by the different look angles of the SAR sensor. To get images of rectangular shape, Full resolution Mosaic Image Data Records (F-MIDR), covering about 15 % and compressed mosaics covering the whole surface have been created.

State-of-the-art applications for earth observation data retrieval like [2] and [3], are designed to give remote access to the archives of various data suppliers. GISIS [3] supports the user with a very intuitive graphical user interface and a detailed and zoomable vector representation of the map of the earth.

Existing applications for planetary data retrieval, provided by members of the PDS, allow to search databases for named features, or for image coverage by defining either a point or a region of interest. Public accessible interfaces (WWW) to find and visualize mosaic image coverage are also available.

With the proposed ISS system users will have access to various data archives by just connecting to a single server. The retrieval process will be supported by visualizing the area of interest as well as the available images before placing an order. The system will reduce costs in terms of transmission bandwidth and media costs by cropping and reorganizing data to exactly the dimensions specified by the user.

ISS was primary designed for using the Magellan data set. But the design we propose is generic in a way, that it can be used for data of any planet or any sensor just by adding the required data (see below) to the Central Server database.

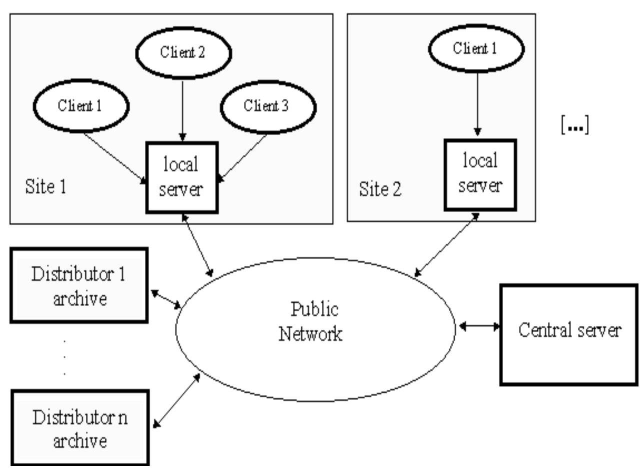

Figure 1: The general layout of ISS. The Central Server (CS) holds the inventory of all attached archives. The CS is accessed from clients via a local server at each site. The location and nature of the various archives is invisible for clients.

System Components

The system consists of a set of five different stand alone components, as illustrated in Fig. 1. Image archives are maintained by individual data providers but have to fulfill special requirements in order to become participant of ISS .

Image Archive

The requirements to integrate an archive into the ISS are:

- The archive must be on-line for 24 hours a day to handle requests from the ISS Central Server.
- All data in the archive can be accessed without any action of an operator.
- The archive must report any changes in the data set to the Central Server.
- The archive should be able to serve requests for parts of an image.
- For setting up the ISS Inventory Database, reduced quicklooks of all images in the archive must be processed. This should be done at the archives site to reduce the amount of data transmitted to the ISS.

Inventory Database

This database holds the inventories of all attached image archives, all quicklooks and a resolution pyramid of the surface map. All images are resampled to cylindrical projection. An additional color reduction to 8 bit is applied to all image data. In combination with low quality JPEG compression a data reduction of 100:1 or better is possible.

A visualization of the planetary surface is necessary to support the user interaction with the system. Because of the small number of known or named features on a planetary surface a vector representation of the surface would be very sparse. Therefore we decided to create a raster representation of a surface map. In [4] a raster map managed by a multimedia data base system was used for visualizing earth observation data. In ISS a predefined set of images that are reduced to a fourth in size and mosaiced together are forming three maps. The first map covers the region between -85° and 85° latitude. The two other maps are for visualizing the polar regions. Successive 2:1 reductions of these maps are forming pyramids. The top level of a pyramid is less than about 1200x1000 pixels in size, to fit to a regular computer screen. Each level of the map pyramid is split into small equal sized (e.g. 256x256 pixels) image tiles. This special data structure for the map is also used to code the image coverage. This allows very fast spatial search for the image coverage of any given point or region. For any planet whose data are available within the ISS a map pyramid has to be created. In the case of Venus this data set is expected to be smaller than 1 Gbyte. A relational DBMS [5] is used to handle all the data sets described so far.

Central Server (CS)

The CS is the main component of the system. The CS has to keep the Inventory database up to date. If the contents of any of the archives changes, the inventory has to be updated as well as any copy of the altered data on any of the known Local Servers. This update process is performed by use of the flood-p algorithm [6].

The Central Server is also the single point of contact for the local servers, and hides the archives from users. It handles all requests that can not be answered by a local server.

The computational most intensive task of the CS is the composition of the data sets ordered by clients. This task is

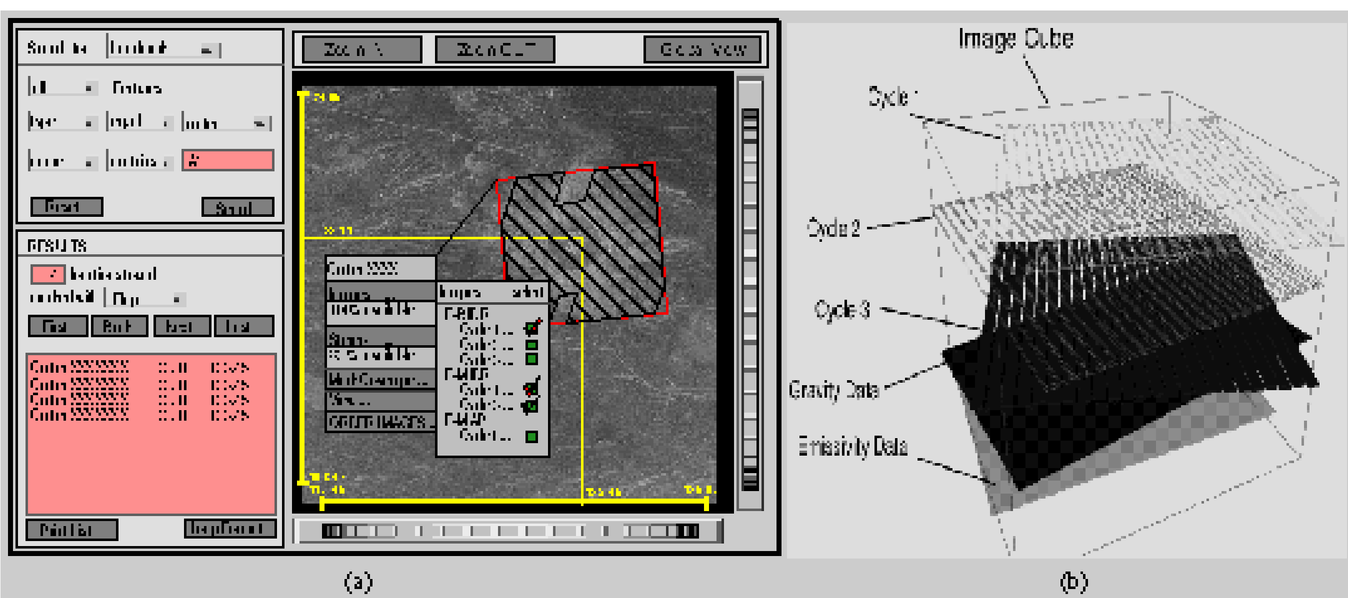

Figure 2. The user interface of the retrieval client prototype with the query dialog (left) and the interactive map browser (right) are shown in (a). Fig. 2 (b) shows how the image data are combined and formated to a data stack of the requested size.

done by the data stack generation module. At first all necessary data are retrieved from the appropriate archives. To speed up this process an archive has to support retrieval of subimages. After that all images have to be resampled to a single cartographic projection, mosaiced or stacked and than cropped to the requested size. Depending on the size and the bandwidth of the network connection, the data can be delivered on-line or copied to any standard storage media for delivery by mail.

Local Server (LS)

One Local Server has to be set up at any user site (see Fig. 1). Main purpose of the LS is to perform data caching to keep network traffic low. For this reason clients are sending all requests for data to the LS. The Local Server contacts the Central Server only if the requested data are not available from the local cache. Data retrieved from the Central Server are stored in this local cache. The size of the cache can be configured individually. If the cache is running out of space, parts of the cache are cleared according to one of the following strategies:

- The oldest data are deleted
- The least used data are deleted
- Data of the user with lowest priority are deleted

At sites with many clients or bad network connection to the CS the LS can be configured to hold up to a full copy of the Inventory Database. On smaller sites the Retrieval Client and LS can run on a single machine and the cache size can be rather small.

The Local Server requires an authentication from any client to allow charging for data retrieval from commercial data providers.

Retrieval Client (RC)

The RC is the software that gives the user access to the ISS. It consists of two major parts, the query definition dialog and the map browser (see Fig. 2). A set of menus and forms supports the composition of user queries. The user can also specify any coordinates with the mouse pointer in the map browser window.

The map browser allows the user to visualize the surface map pyramid stored in the inventory database. The lowest resolution corresponds to the top level of the map pyramid and is displayed when the RC is started. Zooming into the map steps one level down the pyramid, zooming out steps one level up. In the worst case about 40 Kbytes or 16 image tiles have to be transferred from the Central Server to display an image of 1K x 1K pixels.

The map can be overlaid with vector or raster data. Vector data can be image borders, contour lines, names and extents of topographic features and user defined graphics like regions of interest, point marks or personal annotations. Raster data are quicklooks of available images. Color coding can be used to mark areas with a predefined image coverage (e.g. areas with stereo coverage).

Outlook

The proposed system will be implemented within the next two years. This implementation will be restricted to the Magellan data set. Also software development will be limited to a few hardware platforms. Research work will be focused on the development of parallel algorithms and methods for computational intensive programs running at Central Server. A massive parallel computer (Meiko CS-2), located at the University of Vienna, connected with a high speed ATM network [7] will be used for this purpose.

In a next step the Central Server will be expanded by a data processing server, that will be able to create new data products like digital elevation models, classification of image stacks, shaded 3D visualizations or fly-through animations on demand. This will enable the client to create these data without the need for expensive high performance hardware at the local site.

Acknowledgments

Parts of this work are 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’.

References

[1] Poehler, P.; Digital Workstation for Venus Topographic Mapping SPIE Vol. 1943; State-Of-The-Art Mapping; 1993; pp 45-56.

[2] Nativi Stefano, Giuli Dino, Pellegrini Pier Franco, „A distributed multimedia information system designed for the Arno project“, ISPRS Journal, 12-23, 1995

[3] Lotz-Iwen, H. J.; "Earth Observation User Information Services in Germany"; EEOS Workshop on European data networks and earth observation user information services; proceedings; Document refs: CEO/130/1994; March 1995, pp 359-363.

[4] Kirste Thomas, „Spacepicture - An interactive Hypermedia satellite image archival system“,ZGDV - Darmstadt, 1993

[5] Codd E.F., „A relational model of data for large shared data bases“, Comm. ACM, 13, 377-387, 1970

[6] Kappe Frank, „A Scaleable Architecture for maintaining Referential Integrity in Distributed Information Systems“, JUCS, Vol.1 No.2, Feb. 28, 1995

[7] de Prycker Martin, „Asynchronous Transfer Mode“, Prentice Hall, 1994