Network Distribution Techniques of the Global 1-km AVHRR Data Set

Jeffery C. Eidenshink and Lyndon R. Oleson
U.S. Geological Survey, EROS Data Center Sioux Falls, SD 57198

ISPRS Commission II, Working Group II/3

KEY WORDS: Networks Distribution On_line Global Data Remote_Sensing Vegetation

ABSTRACT

Under the framework of the Earth Observing System Data and Information System (EOSDIS), which is designed to provide national and international user community access to these data, the U.S. Geological Survey's EROS Data Center (EDC) serves as a Distributed Active Archive Center (DAAC) for land processes data. One of the primary Pathfinder activities undertaken by the EDC DAAC is the global land 1-km advanced very high resolution infrared radiometer (AVHRR) data set. Two products are available from the global 1-km project: raw data and 10-day vegetation index composites. The total data volume of a single 10-band composite is approximately 10.5 gigabytes. Fifty-four 10-day composites will have been produced by the end of 1996, and the combined archive of raw data and 10-day composites will exceed 4.0 terabytes.

A data compression technique has been developed specifically for use in the storage and distribution of the global 10-day composite data. The technique results in approximately a 10:1 reduction of data. It provides for spectral, spatial, and geographical subsetting capability to the user. The EDC DAAC is providing access to the global 1-km data set's orbital segments and 10-day vegetation index composites through DAAC's World Wide Web (WWW) home page. The WWW interface offers a simple HTML form that allows users to select and request a 10-day composite or raw data.

Introduction

One of the goals of NASA's Mission To Planet Earth, and in particular the Earth Observing System (EOS) program, is to develop a comprehensive data and information system that includes the retrieval and processing of data that are critical for interdisciplinary scientific investigations of Earth system processes. Under the framework of the EOS Data and Information System (EOSDIS), which is designed to provide national and international user community access to these data, the U.S. Geological Survey's EROS Data Center (EDC) serves as a Distributed Active Archive Center (DAAC) for land processes data. The EDC DAAC carries out responsibilities for processing, archiving and distributing EOS and auxiliary land science data.

One of the greatest challenges to EOSDIS is the management of the anticipated volume of raw and processed data acquired by EOS space-borne sensors. The data

volume of the land products from the EOS Moderate Resolution Imaging Spectrometer (MODIS) sensor, for example, is estimated to total one hundred terabytes of data by the year 2000. This is in addition to hundreds of terabytes of raw data. In order to prepare the mechanisms and procedures to cope with the data volume and fulfillment of user requirements, EOS initiated a series of Pathfinder projects to assist in the development of EOSDIS capabilities using data from currently operational remote sensing systems.

One of the primary Pathfinder activities undertaken by the EDC DAAC is production of the global land 1-km advanced very high resolution infrared radiometer (AVHRR) data set (Eidenshink and Faundeen 1994). The global land 1-km data set project represents an international effort to acquire, archive, process, and distribute 1-km AVHRR data of the entire global land surface in order to meet the needs of the international science community. The partnership of the USGS, NASA, NOAA, the European Space Agency (ESA), the Commonwealth Scientific and Industrial Research Organization of Australia, and a network of nearly 30 high resolution picture transmission (HRPT) ground stations have been acquiring daily global land coverage since April 1, 1992. In addition, the International Geosphere Biosphere Program - Data and Information System (IGBP-DIS) and the Committee on Earth Observation Satellites (CEOS) have provided coordination of the international scientific community to facilitate development of processing standards (Townshend, 1992).

The scientific requirements for the data set are broad in scope and international in coverage. The IGBP-DIS land cover working group completed a study that stressed the need for a this data set for mapping global land cover. The United Nations Food and Agriculture Organization's Forest Resources Assessment 1990 Project requires 1-km AVHRR data for all the forested lands on the planet, with emphasis on the tropical zones, for their global forest inventory mandate (FAO, 1990). NOAA researchers required a 1-km AVHRR data set of the northern hemisphere to study the urban heat island effects on surface observations of temperature data (Gallo et al., 1993). The Commission of the European Communities (CEC) and the ESA have a joint requirement for global, near daily, long-term, consistent optical and thermal satellite data for tropical environments in support of the joint Tropical Ecosystem Environment Observations by Satellite (1991) project. The NASA Moderate Resolution Imaging Spectrometer (MODIS) land science team concluded that a global land 1-km AVHRR data set is crucial to develop algorithms for several land products for the Earth Observing System (EOS) (Running et al., 1993). Although there are many differences in AVHRR and MODIS sensor characteristics, the daily data volume of global AVHRR land coverage at 1.1 kilometer resolution and the pole-to-pole coverage provides a test bed for development of higher level derivative products and management and dissemination of large volumes of data.

The Global Land 1-km AVHRR Data Set

The data set is composed of 5-channel, 10-bit, raw AVHRR data, at 1.1-km resolution (at nadir) for every daily afternoon orbit over all land and coastal zones. Two products are available from the global 1-km project: raw data and 10-day vegetation

index composites.
The raw data are packaged as continuous orbital segments of the afternoon AVHRR acquisitions. The orbital segments are produced by joining several consecutive observations, obtained along an orbit by the various ground stations, into a single continuous data segment for that orbit. The size of the orbital segments vary with the amount of land data covered by the orbit. Orbits beginning over Australia and continuing over Indonesia and Asia are nearly pole-to-pole. The data volume of a pole-to-pole segment is nearly 250 megabytes. The data volume of the shortest segments, such as one over the Alaska peninsula, is approximately 50 megabytes. There are 14 orbits per day with an average total daily volume of 2 gigabytes. The raw data are distributed in NOAA Level-1b format (Kidwell, 1991).

The 10-day vegetation index composites are multiband, georegistered, global data sets. The 10 bands of the composite product are comprised of: AVHRR channels 1-5, NDVI, satellite zenith, solar zenith, relative azimuth, and date of observation. The five AVHRR channel bands are stored as 16-bit. Each of which has a data volume of nearly 1.4 gigabytes. The other five bands are stored as byte data and have a volume of nearly 695 megabytes each. The total data volume of a single 10-band composite is approximately 10.5 gigabytes. Fifty-four 10-day composites will have been produced by the end of 1996, and the combined archive of orbital segment and 10-day composite data will exceed 4.0 terabytes.

Compression technique for global 10-day composites

Data compression is an important management tool for image data, especially lossless compression (Nelson, 1991). Data compression reduces the size of files to be transferred on the network, and allows for more data to be placed on physical media. In addition, it reduces the storage capacity required to maintain the data online or near-line at the distribution site and the user's site. However, decompression is usually necessary before the data can be analyzed using traditional image processing and analysis systems. The decompression step can be time consuming and require that the processing or analysis system store and manipulate the full data set.

The data volume of a multiband, 10-day global composite, at 1-km resolution, in the interrupted Goodes homolosine map projection (Steinwand, 1994) is 10.5 gigabytes. Although most users may not need or use all ten bands of a 10-day composite, the practical use of the NDVI data as a time series would require approximately 25 gigabytes for a one-year time period.

A data compression technique has been developed specifically for use in the storage and distribution of the global 10-day composite data (Kess et al., 1994).
The technique functions on a single band at a time. Each band is divided into a matrix of 128 by 128 pixel blocks. Blocks that are entirely a single value such as ocean, or the interrupted area created by the map projection, are compressed with run length encoding. The land data blocks are Huffman encoded and ordered so that each resolution level can be progressively decompressed. The geographic coordinates for each block and all of the information necessary to decompress the data in that block to different resolutions is stored in a header file.

There are many advantages of this technique to the product distribution process. The technique results in approximately a 10:1 reduction of data, reducing the near-line storage requirements. It provides for spectral, spatial, and geographical subsetting capability to the user. The spectral subsetting results from the compression of each band separately, spatial subsetting comes from the capability to decompress at selected resolutions, and geographical subsetting is possible because the image is stored as separate 128 by 128 blocks. The user can specify geographical coordinates of any rectangular subset. The decompression interface determines which array of 128 by 128 pixel blocks is required to cover the subset, and only those blocks are decompressed. This technique also minimizes the amount of data that is retrieved from the file server, and thus improves access times. The spatial subsetting is most useful if the user wants to browse the composite at coarser resolution before selecting a full resolution data set. It is also useful for many global studies where coarser resolution data would suffice.

Network Access to the Data

Network access and distribution of data has become very dynamic. New modes of access, such as the World Wide Web (WWW), provide a creative opportunity for sharing data and information. One of the ways the EDC DAAC is providing access to the global 1-km data set's orbital segments and 10-day vegetation index composites is through DAAC's WWW home page. As the orbital segment and 10-day composite products are produced, they are staged to the DAAC's Data Distribution System (DDS) for network access. The DDS includes a digital linear tape (DLT) robotic data storage library that is configured to hold 264 10-gigabyte cartridges or approximately 2.6 terabytes of storage. Access to these data are provided through a global 1-km data set WWW home page. The URL is:

(http://edcwww.cr.usgs.gov/landdaac/1KM/1kmhomepage.html).

Using this WWW access path, users can read documentation and general information about the global 1-km AVHRR data set and the products that are available. The WWW interface offers a simple HTML form that allows users to select and request a 10-day composite or an orbital stitch.

The orbital segment HTML form provides documentation on the orbital segment production process and an explanation of how to identify the approximate geographic coverage of a segment. The orbital segment data are organized in directories, sorted by satellite (NOAA-11, NOAA-14), year, month, day, and hour. A user can quickly find the appropriate timeframe and orbits of interest. Since orbital segments are delivered in standard NOAA level-1b packed format (three 10-bit values in a 32-bit word), very little data compression of level-1b data can be achieved.

The global 10-day composite HTML form provides for selection of specific 10-day periods, bands, resolution and, optionally, geographic area. When the user submits the form, the DAAC's WWW server runs a program that calculates and reports the actual size and geographic coordinates of the data that will be extracted and a rough estimate of the amount of time necessary to retrieve the data. The user can then

choose to proceed with downloading the data or modifying the parameters of the request if the data volume is too large. In addition, users can optionally choose to have the data compressed just prior to transfer using standard UNIX or GNU gzip compression resulting in up to a five time reduction in the size of the data transfer. In 1995 the DAAC's WWW server and DDS handled more than 5,247 network requests for global 1-km data set products totaling more than 112 gigabytes of data.

Summary

The compression/decompression technique developed for the global land 1-km AVHRR data set provides several capabilities that are beneficial to the user and the EDC DAAC. The technique provides the user the capability to subset the data set spectrally, spatially, and geographically. The technique provides the EDC DAAC an efficient method for storage and retrieval of the data from near-line storage. The experience gained from network distribution of the data set will be used to develop the distribution techniques for NASA's Mission to Planet Earth Program.

References

Eidenshink, J. C. and Faundeen, J. L., 1994, The 1 km AVHRR global land data set: first stages of implementation. The International Journal of Remote Sensing, 15, 3443-3462.

Food and Agriculture Organization of the United Nations, 1990, Expert Consultation on Forest Resources Assessment 1990 Project, FAO Headquarters (Rome: Forest Resources Assessment).

Gallo, K. P., McNab, A. L., Karl, T. R., Brown, J. F., Hood, J. J., and Tarpley, J. D., 1993, The use of vegetation index for assessment of the urban heat island effect: *The International Journal of Remote Sensing*, 14, 2223.

Kess, B. L., Steinwand, D. R., Reichenback, S.E.,1994. Compression of the Global AVHRR 1-KM Data. IGARSS'94 Digest, Vol. 1, pp. 323-325

Kidwell, K. B., 1991, NOAA Polar Orbiter Data Users' Guide, National Oceanic and Atmospheric Administration, World Weather Building, Room 100, Washington, D.C.

Nelson, M., 1991, *The Data Compression Book* (Redwood City, California: M&T Books).

Running, S. W., Justice, C. O., Salomonson, V. V., Hall, D., Barker, J., Kaufmann, Y, J., Strahler, A. H., Huete, A. R., Muller, J.-P., Vanderbilt, V., Wan, Z. M., Teillet, P., and Carneggie, D., Terrestrial remote sensing science and algorithms planned for the Moderate Resolution Imaging Spectrometer (MODIS) of the Earth Observing System (EOS): The International Journal of Remote Sensing, 15, 3587-3620.

Steinwand, D. R., Mapping raster imagery to the interrupted Goode Homolosine Projection.The International Journal of Remote Sensing, 15, 3463-3471.

Townshend, J. T. (editor), 1992, Improved Global Data for Land Applications. IGBP Global Change Report No. 20. The International Geosphere Biosphere Programme: A Study of Global Change of the Council of Scientific Unions, Stockholm.

Townshend, J. R. G., Justice, C. O., Skole,

D., Malingreau, J.-P., Cihlar, Teillet, J., Sadowski, F., and Ruttenberg, S., The 1-km AVHRR global data set: needs of the International Geosphere Biosphere Program: The International Journal of Remote Sensing, 15, 3417-3441.

Tropical Ecosystem Environment Observations by Satellite(TREES),1991, A strategy proposal 1991-1993: Joint Research Center, TREES Series A, Ispra, Italy.