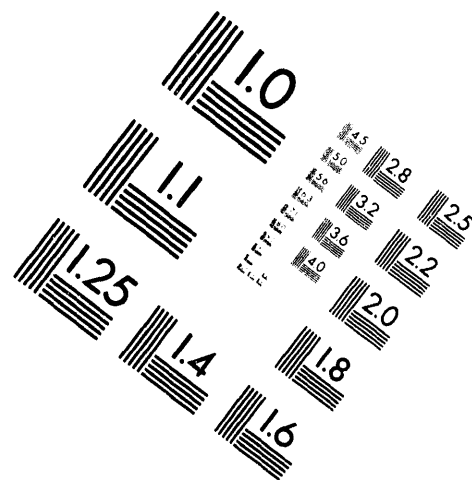
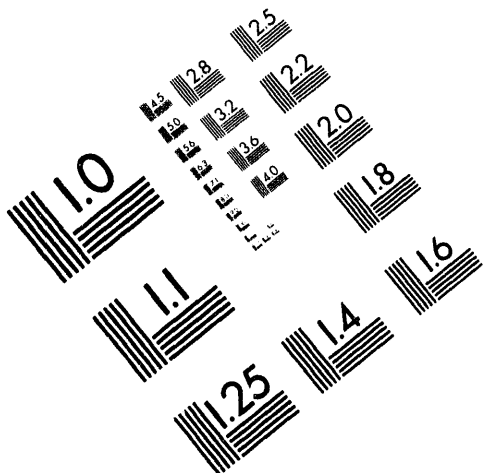




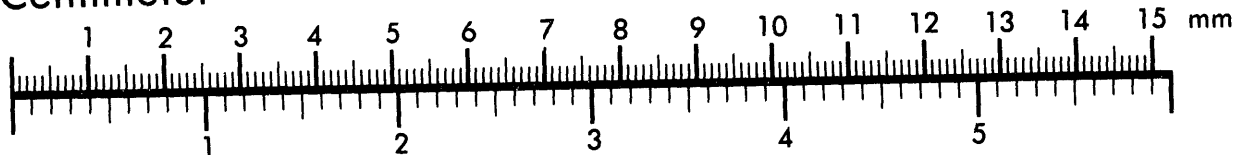
**AIM**

**Association for Information and Image Management**

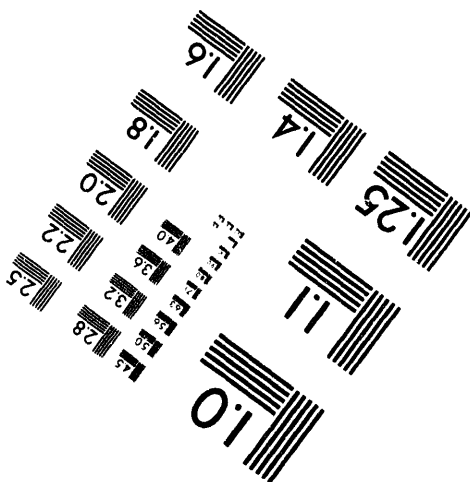
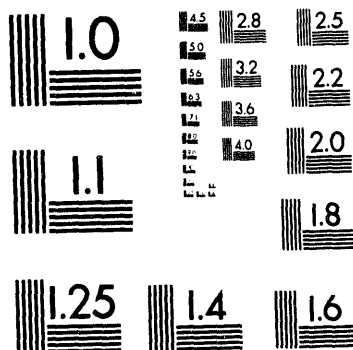
1100 Wayne Avenue, Suite 1100  
Silver Spring, Maryland 20910  
301/587-8202



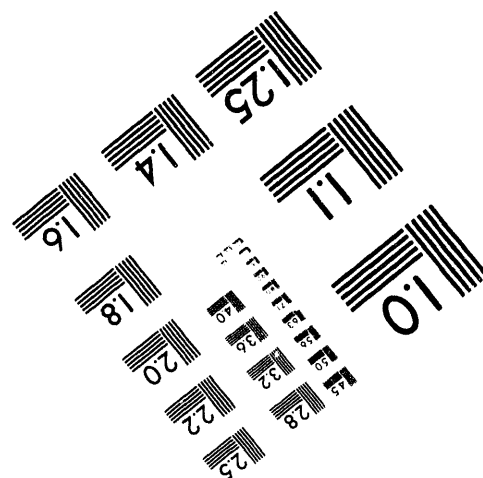
**Centimeter**



**Inches**



MANUFACTURED TO AIM STANDARDS  
BY APPLIED IMAGE, INC.



**1 of 1**

ANL/DIS/CP-83606  
Conf-9408129--2

## A HYPERSPECTRAL IMAGE DATA EXPLORATION WORKBENCH FOR ENVIRONMENTAL SCIENCE APPLICATIONS

MARK A. WOYNA  
JOHN H. CHRISTIANSEN  
DAVID G. ZAWADA  
KATHY LEE SIMUNICH

ADVANCED COMPUTER APPLICATIONS CENTER  
ARGONNE NATIONAL LABORATORY

### ABSTRACT

The Hyperspectral Image Data Exploration Workbench (HIDEW) software system has been developed by Argonne National Laboratory to enable analysts at Unix workstations to conveniently access and manipulate high-resolution imagery data for analysis, mapping purposes, and input to environmental modeling applications. HIDEW is fully object-oriented, including the underlying database.

### 1. INTRODUCTION

A significant challenge to the information sciences is to provide more powerful and accessible means to exploit the enormous wealth of data available from high-resolution imaging spectrometry ("hyperspectral" imagery) for analysis, for mapping purposes, and for input to environmental modeling applications. As an initial response to this challenge, Argonne's Advanced Computer Applications Center has developed a workstation-based, prototype software workbench which employs AI techniques and other advanced approaches to deduce surface characteristics and extract features from hyperspectral images. Among its current capabilities, the prototype system can classify pixels according to surface type. The classification process employs neural network analysis of inputs which include pixel spectra and a variety of processed image metrics, including image "texture spectra" derived from fractal signatures computed for subimage tiles at each wavelength.

The prototype software system is object-based, written in the C++ language. It employs a readily comprehended graphical user interface for data navigation and visualization, and a flexible data interchange scheme to allow the system to function independently of the source of hyperspectral data. In order to explore the options for optimal access to external image and related data, we have developed and are evaluating two software system prototype variants, one interfacing with external image files in the University Corporation for Atmospheric Research's (UCAR's) netCDF standard format, and one tightly integrated with an object database management system.

HIDEW prototype is being tested in several land-use classification and feature extraction tasks as

---

The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. W-31-109-ENG-38. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

MASTER *db*  
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

an aid to U.S. Department of Energy (DOE) site characterization work, as well as in other application areas in global change research, in proof-of-principle exercises, and pilot studies using actual hyperspectral image data from the Jet Propulsion Laboratory's (JPL) Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) sensor system.

## 2. THE WORKBENCH

HIDEW prototype is being developed both as a software engineering demonstration and as an analysis tool to directly support studies. It is intended that the prototype will provide a general, easily expandable framework for hyperspectral and multispectral image analysis, allowing new tools and approaches for data visualization and analysis, and existing tools from diverse sources, to be brought to bear on the problem of exploitation of hyperspectral and multispectral data. In addition, it is planned that the prototype will be used, with application-specific enhancements, to support studies of interest to the DOE, as outlined in Section 4 of this paper.

### Software System Structure

HIDEW makes full use of the features and capabilities of the object-oriented software development paradigm. The prototype implements a C++ object-based common data representation for data transfer, manipulation, analysis, and archiving, based on the concept of an "Image Data Packet" (IDP). All image-related data manipulated by the system and displayed by its user interface are in the form of IDPs. Scores of different object methods (functions) belonging to the IDP object class have been developed to access, manipulate, and restructure IDP data. IDPs are cataloged into hierarchies of Library objects. Applications of all kinds are encapsulated within Process objects, which interact with IDPs in a consistent fashion. This object-based implementation will greatly facilitate future upgrades to the system. Figure 1 depicts the basic structure of the IDP and Process object classes.

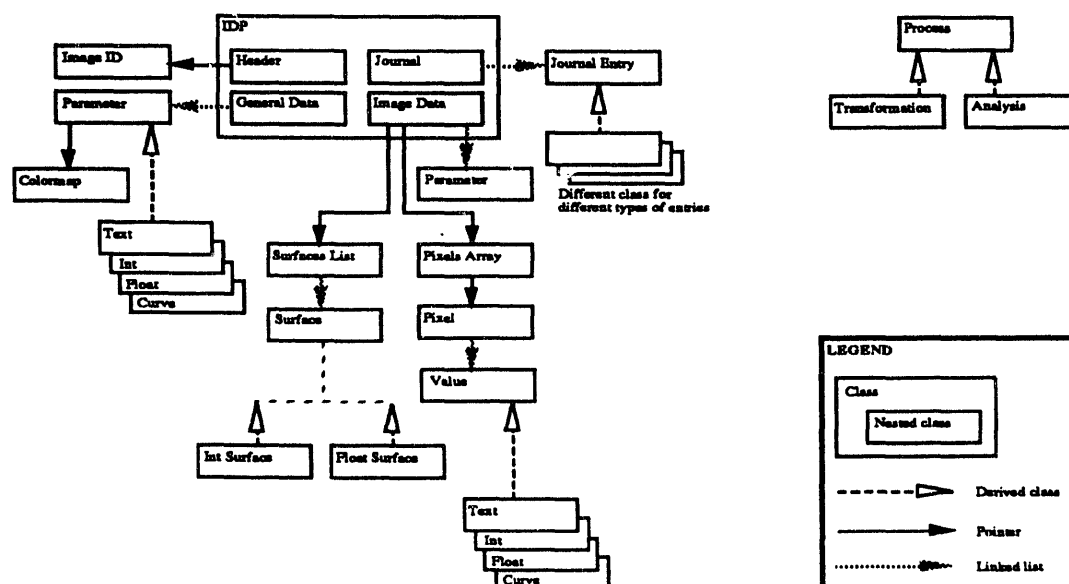


Fig. 1. Structure of the IDP and Process object classes.

## Database

Portability and expandability were critical design factors. As a software engineering experiment, we are currently comparing two versions of the workbench prototype: one using UCAR's netCDF file representation [Rew90] and the other using the Versant Object Database Management System [Versant92a].

UCAR's netCDF data access library organizes scientific data into a self-describing and network-transparent format. In general, hyperspectral image cubes (originals and manipulated versions) reside in disk files rather than in memory, due to their very large size (approximately 140 MBytes). By using the netCDF library, image cubes and processed portions of image cubes may be shared among applications and across many computer platforms.

Operations dealing with whole image cubes (e.g., in applying atmospheric corrections, taking differences between two image cubes, etc.) must be carried out by breaking up very large datasets so they can be manipulated in memory, and reconstituting the results of the manipulation on disk. In order to minimize disk access time where possible, we have implemented an "Image Memory" object which regulates the cache of image data which may be in memory at any given time. This memory management process is transparent to the user. However, this process is very computer resource-intensive, prone to synchronization problems between the inline and database representations of the same data, and represents a potential bottleneck in the image exploitation process.

Using the netCDF library with our object-based in-line data representation has helped to meet our goals of system portability and expandability, but has required a great deal of code simply to accommodate data transfers between the radically different representations of the same information in the netCDF files and in the in-line data objects. It has also introduced the need for our system to include its own memory management and file synchronization code.

We have identified the following limitations of the netCDF approach:

### Schema Support:

- The netCDF schema language was the not same as the programming language, requiring that the C++ classes be normalized to produce a set of netCDF file creation subroutine calls. This required that someone on the HIDEW development team be proficient in netCDF file design.
- Complex structures had to be converted to an artificially "flattened" array representation in the netCDF system.

### Additional Input/Output Code:

- The need to maintain two representations of data (database and application) was burdensome. Extensive code was needed to keep both representations of the object in agreement.
- Input/output code was needed to translate data from the files to the internal representation. This code was a significant fraction of the overall system code.

### Performance:

- Reconstituting complex, normalized objects required joining many files, a relatively expensive operation.
- The need to write a memory manager was burdensome, and provided little flexibility for performance tuning.
- The need to maintain two netCDF representations of the data to obtain reasonable performance for some operations doubles storage requirements.

Since the prototype is an object-based system, it is well suited for use with an object database management system (ODBMS). While most currently available ODBMSs are proprietary, the gains received from not having to write extra reformatting, memory management, and synchronization code may greatly outweigh the expense of the system.

The Versant ODBMS has its own caching mechanism which is transparent to the programmer of the system, as well as to the user. Objects, which contain all the data that the system needs, are referenced within the code as if they were in memory at all times. These same objects are “saved” as persistent objects within the ODBMS, thus encapsulating the data with the object with no need to reformat the data to and from a flat file representation.

The use of an object database management system has provided the following benefits:

### Schema Support

- The database schema “language” is the same as the programming language (C++), and the schema is automatically captured from the application code.
- Object structures do NOT have to be “flattened” into another representation.
- No need for additional Input / Output or synchronization code. Objects are referenced as if they are in memory at all times.

### Performance

- Client-side object caching makes accessing the objects efficient (most recently referenced objects remain within the cache).
- Tuning parameters allow for fine-tuning of different operations (image loading, surface retrieval).

### Unique Features

- Support for design transactions allows transactions which can span multiple database sessions which can last several days with full commit or rollback capability.
- Support for versioning of objects.

Porting to Versant required approximately 10 person-days of effort to convert the approximately 20,000 lines of C++ code. The conversion process required the following steps [Versant92b]:

- Derive all persistent classes from the Versant PObject class to inherit persistent behavior.

- Convert standard C pointers to Versant Links, the persistent database equivalent of a transient pointer.
- Add the Versant dirty() method to all methods that update a persistent object. The dirty() method obtains exclusive locks on the object and mark the object as updated.
- Convert our transient version of a linked list to the persistent Versant VIList linked list class.
- Remove the methods which implemented persistence in the persistent classes which were no longer needed.

### Graphical User Interface

User interaction with HIDEW is facilitated by a sophisticated, yet intuitive, windowing interface. The user may navigate through the system and access the majority of its functions via mouse actions (e.g., pointing and clicking, dragging and dropping of icons). Most of the prototype's abstract concepts are represented visually as icons, with consistent appearance and behavior. The interface was designed to promote work sessions as positive, visual experiences; minimizing procedural details to maximize the user's potential for insight into the data.

### Prototype Image Analysis Tool Suite

The first major capability which was implemented in the prototype workbench was classification of image pixels according to surface types, using a neural network classifier. This analysis was performed on image data captured by JPL's AVIRIS sensor system. The sensor possesses 244 contiguous spectral channels, each having a nominal width of 10 nm, and covers the electromagnetic spectrum roughly from 400 nm to 2500 nm. A single image covers an area of approximately 10 x 11 km, with each pixel representing an approximate 20 x 20 m area on the ground. Pixel classification was accomplished using NASA's NETS backpropagation neural network simulator code [Baffes89], incorporated into the prototype as a Process object.

The next major capability to be built into the system was computation of local fractal signatures for subtiles of an image. The fractal signature module employs algorithms described by Peleg, et al. [Peleg84] and Pentland [Pentland84] to produce curves of fractal dimensions of the image surface (represented as grayscale values) as a function of image distance scale (pixel separation). We expect that these measures of local image texture at various wavelength channels will provide valuable additional context for the determination of surface type for image pixels, when the fractal signature surfaces produced by this module are submitted to surface type classifiers along with the image spectra.

A major aim of our development effort is to add and test new, AI-based tools for image analysis. In addition, we expect to add "standard" image manipulation functions as they are needed to support pilot studies. Many of these functions are readily available due to the remote sensing community's many years experience in multispectral image analysis. Examples of such functions include filtering, altering spatial and/or spectral resolution, converting images from radiance to reflectance units, etc. HIDEW will allow the user to interactively apply atmospheric corrections to images, via the LOWTRAN 7 [Kneizys88] and ARP [CSES92] approaches, among others.

### 3. PROTOTYPE TESTS AND PILOT STUDIES

Our initial development work has dealt with hyperspectral images from the AVIRIS sensor system. AVIRIS image cube data were radiometrically corrected by JPL to produce image radiance values (in microwatts / cm<sup>2</sup> / nm / steradian), and were then normalized for further analysis within the ranges defined by the extreme values for each spectral band. Initial testing has been based on AVIRIS images for the Jasper Ridge Biological Preserve and Moffett Field, both located near Stanford University in Palo Alto, California.

Based on aerial photographs, topographic maps, and limited site reconnaissance of the Jasper Ridge area, we selected a coarse set of abstract surface types: water, meadow, lawn, paved, forest, etc. We then selected a few (typically seven to ten) examples of each type from the image, and trained the net to classify all pixels in an image or subimage section.

For purposes of classifying pixels by surface type, we have employed both three-layer and four-layer backpropagation neural networks. Initially, the network has been configured with an input layer which simply had one input node for each wavelength channel (minus the channels dominated by water vapor absorption), an output layer with an output node for each defined surface class, and one or two hidden layers.

Using this configuration, we have had good success in producing high-resolution land use mappings with very little time and effort expended in "ground truthing" the site or developing complex metrics based on domain knowledge to differentiate surface types. A neural network classifier trained on the Jasper Ridge data also had good success in classifying pixels in the Moffett Field image, indicating that this relatively simple approach is at least somewhat robust.

We are presently assessing the value of local image texture spectra derived from fractal analysis as an aid to surface classification. Although the "naive" approach to surface classification with little aid from domain knowledge has worked surprisingly well in testing, extensive use of domain knowledge to aid in image exploitation is planned for future work, as outlined below in Section 4.

### 4. WORK PLANNED OR IN PROGRESS

There are at present several major areas of DOE work planned or in progress in which the Argonne Hyperspectral Image Data Exploration Workbench plays a significant role. These include global environmental change assessment, and high-resolution surface characterization, both for DOE site environmental management purposes and to support improvements in General Circulation Models for climatic change prediction.

#### Global Environmental Change Assessment

Argonne's Decision and Information Sciences Division has an ongoing effort to identify, evaluate, and test methods of quantifying environmental change, in order to develop a capability to utilize remote sensing data for research problems on vegetation and land surface changes due to human activity and/or climate change. Both multispectral (Landsat Multi-Spectral Scanner) and hyperspectral data are being utilized in this study.



To support this effort, the prototype workbench will be enhanced to compute fractal-based measures of biodiversity from classified images. Pixel classification will be based on spectral data, derived local fractal texture data, and domain knowledge. The initial prototype neural net-based surface classification schemes are fully general in that they do not depend upon *a priori* knowledge of the contents of the image. However, they are somewhat limited in that they do not take advantage of well-understood physical laws, ecosystem and landform interrelationships, etc. as aids to proper surface characterization. We, therefore, have begun to add a knowledge-based system (KBS) component which will provide a real-world context to aid the image exploitation software in its tasks. The knowledge base for this KBS, covering relevant interrelationships among surface types, will be articulated in a rules-based expert system component capable of aiding in surface identifications for individual image pixels. The classification accuracy for individual image pixels should be substantially improved by access to such domain knowledge. For example, if a single 20 m pixel is identified spectrally as "rain forest", it is highly unlikely that adjacent pixels will represent "desert".

In addition, the KBS can be used to assist in examining an image at a higher level of abstraction, to identify and characterize features extending across several pixels, and to detect more abstract features such as biomes. This process will produce an object-based interpretation of the image in terms of spatially distributed objects representing overlapping and interacting regions with distinct characteristics. This data abstraction process will feed, and be fed by, the pixel classification process, resulting in both a detailed characterization of the surface and the extraction of key features.

### Surface Characterization Studies

In developing or reviewing environmental assessments or impact analyses for remediation plans or other site projects, there is a strong need for accurate, high-resolution data characterizing the site. Until recently, such data have been prohibitively expensive to acquire, and very difficult to analyze effectively, even when available. Exploitation of hyperspectral imagery makes possible the acquisition of such high-resolution databases at relatively low cost. We are planning to try a variant of our neural network surface classifier as a front end processor to feed high-resolution surface characterization data, based on automated hyperspectral image exploitation, to an object terrain database. The database supported would be an intelligent terrain database based on a prototype currently under development at Argonne and George Mason University, in which characteristics and dynamic behavior of pixels are specified via embedded models and knowledge bases, implemented over a true object database.

Another key application area for high-resolution surface characterization via hyperspectral imagery exploitation is the Atmospheric Radiation Monitoring (ARM) Program [DOE90]. A cornerstone of the ARM program is an extensive field study to provide better characterizations of surface energy and momentum exchange for general circulation models (GCM's) used to make global climate predictions. One of the general measurement strategies of the ARM program involves measurements to support the "single column" GCM - that is, measurements made to improve the parameterization of sub-grid scale processes within a GCM. A key element in this strategy is an investigation of the appropriate scale of surface feature characterization which is necessary to adequately describe atmosphere-surface interactions. Acquisition and exploitation of hyperspectral imagery could allow us to map the entire 300 x 280 km Southern Great Plains ARM

site at 20 m resolution, for relatively low cost. Aggregations of surface characteristics from the 20 m scale on upwards can then be constructed and tested in the surface exchange models used in the single cell grid characterization.

## **5. SUMMARY**

The work which has been done to date on the Hyperspectral Image Analysis Workbench prototype has demonstrated the feasibility of hyperspectral image exploitation using an object-oriented data model, software design, and implementation paradigm coupled with state of the art analysis and visualization techniques. By designing the prototype system to be easily expandable, we have made it possible for the system to support a wide range of image analysis applications. We have also demonstrated the feasibility of utilizing an object database management system in the support of large, complex, scientific data sets.

## **6. ACKNOWLEDGEMENT**

The work discussed in this paper was supported by the U. S. Department of Energy under contract W-31-109-Eng-38.

## **7. BIBLIOGRAPHY**

Mark Woyna is the Section Manager for the Information Systems Section in the Advanced Computer Application Center of the Decision and Information Sciences Division of Argonne National Laboratory, Argonne, Illinois, 60439 (708-252-2557) (Fax: 708-252-5128) He received his Master of Science in Electrical Engineering and Computer Science from the University of Illinois at Chicago, Chicago, Illinois. (woyna@anl.gov)

John Christiansen is the Section Manager of the Environmental Modeling Section in the Advanced Computer Application Center of the Decision and Information Sciences Division of Argonne National Laboratory, Argonne, Illinois, 60439 (708-252-3291) (Fax: 708-252-5128) He received his Master of Science in Astronomy from the University of Texas, Austin, Texas. (jhc@dis.anl.gov)

David Zawada is an Assistant Scientist in the Environmental Modeling Section in the Advanced Computer Application Center of the Decision and Information Sciences Division of Argonne National Laboratory, Argonne, Illinois, 60439 (708-252-5072) (Fax: 708-252-5128) He received his Master of Science in Ocean Engineering from the Florida Institute of Technology, Melbourne, Florida. (zawada@dis.anl.gov)

Kathy Lee Simunich is an Assistant Scientist in the Environmental Modeling Section in the Advanced Computer Application Center of the Decision and Information Sciences Division of Argonne National Laboratory, Argonne, Illinois, 60439 (708-252-3285) (Fax: 708-252-5128) She received her Master of Science in Computer Science from North Central College, Naperville, Illinois. (simunich@dis.anl.gov)

## 8. REFERENCES

- [Baffes89] Baffes, P.T., 1989: NETS User's Guide, Version 2.0, Software Technology Branch, Lyndon B. Johnson Space Center, Publ. JSC-23366, September 1989.
- [Gao92] Gao, B.-C., K.B. Heidebrecht, and A.F.H. Goetz, 1992: Atmospheric Removal Program (ARP) User's Guide, Version 1.0, Center for the Study of Earth From Space (CSES), Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, May 1992.
- [Kneizys88] Kneizys, F. X., E.P. Shettle, L.W. Abreu, J.H. Chetwynd, G.P. Anderson, W.O. Gallery, J.E.A. Selby, and S.A. Cough, 1988: User's Guide to LOWTRAN 7, Optical/Infrared Technology Division, Air Force Geophysics Laboratory, Hanscom AFB, Publ. AFGL-TR-88-0177, 16 August 1988.
- [Peleg84] Peleg, S., J. Naor, R. Hartley, and D. Avnir, 1984: Multiple Resolution Texture Analysis and Classification, IEEE Computer Society Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No. 4, July 1984, pp. 518-523.
- [Pentland84] Pentland, Alex P., 1984: Fractal-Based Description of Natural Scenes, IEEE Computer Society Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No. 6, November 1984, pp. 661-674.
- [Rew90] Rew, R.K., and G.P. Davis, 1990: The UNIDATA netCDF: Software for Scientific Data Access, Preprints of Sixth International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, Anaheim, CA, February 1990, pp. 33-40.
- [DOE90] U.S. Department of Energy, 1990: Atmospheric Radiation Measurement Program Plan, Office of Energy Research, Office of Health and Environmental Research, Atmospheric and Climate Research Division, Publ. DOE/ER-0441, February 1990.
- [Versant92a] Versant Object Technology, Inc. Versant DBMS, 1992.
- [Versant92b] Versant Object Technology Corporation, 1992: Versant System Reference Manual, Release 1.7, Menlo Park, CA, January 1992.

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

**DATE  
FILMED**

**10/14/94**

**END**

