NASA-CR-195268

Visualization Techniques to Aid in the Analysis of Multi-Spectral Astrophysical Data Sets

Final Report on grant NAGW-1902
Prepared for the National Aeronautics and Space Administration
HQ, Astrophysics (Code SZ)

September 1993

Principal Investigator: E.W. Brugel, Ph.D.
Investigators: G.O. Domik (Lead-Investigator), Ph.D., and T.R. Ayres, Ph.D.

Report prepared by

Gitta Domik[1]

Work performed at the University of Colorado at Boulder
Center for Astrophysics and Space Astronomy &
Department of Computer Science

N94-27932

Unclas

G3/90 0000450

(NASA-CR-195268) VISUALIZATION TECHNIQUES TO AID IN THE ANALYSIS OF MULTISPECTRAL ASTROPHYSICAL DATA SETS Final Report (Colorado Univ.) 51 p

1. Current address: University of Paderborn, Department of Mathematics and Computer Science, FB 17, D-33095 Paderborn, Germany
Tel. +49/5251/60-2605 Fax: 60-3836 email:domik@uni-paderborn.de

Please reference this document as:

Domik, G.O., 1993, *Visualization Techniques to Aid in the Analysis of Multi-Spectral Astrophysical Data Sets*, Final report to NASA, Astrophysics Division, on grant NAGW-1902. University of Colorado, Boulder, CO. 80309-0389.

Acknowledgements

The investigators wish to thank the Astrophysics Division of the National Aeronautics and Space Administration, specifically Dr. Günter Riegler, to give them the opportunity to investigate into the capabilities and challenges of Scientific Visualization in regards to astrophysical data.

The support of the Center for Astrophysics and Space Astronomy (CASA) in providing a stimulating scientific environment, hardware and software resources, and general administrative assistance is much appreciated. Many thanks go to Dr. Clayton Lewis, Department of Computer Science, and Dr. Theodore Snow, director of CASA, for consultancy and support during the project.

During 1989 and 1993 many contributions were made by members of the Center for Astrophysics and Space Astronomy as well as by members of the Department of Computer Science. We gratefully acknowledge the work of the following individuals:

S. Alam, G. Allison, R. Chaiken, W. Fortson, C. Joslyn, S. Merkle, K. Mickus-Miceli, P. Pinkney, W. Radinsky, R. Tarrall, A. Veale, J. Wilson.

Table of Contents

Executive Summary

The goal of this project was to support the scientific analysis of multi-spectral astrophysical data by means of scientific visualization. Scientific visualization offers its greatest value if it is not used as a method separate or alternative to other data analysis methods but rather in addition to these methods. Together with quantitative analysis of data, such as offered by statistical analysis, image or signal processing, visualization attempts to explore *all* information inherent in astrophysical data in the most effective way.

Data visualization is one aspect of data analysis. Our taxonomy as developed in Section 2 includes identification and access to existing information, preprocessing and quantitative analysis of data, visual representation and the user interface as major components to the software environment of astrophysical data analysis. In pursuing our goal to provide methods and tools for scientific visualization of multi-spectral astrophysical data, we therefore looked at scientific data analysis as one whole process, adding visualization tools to an already existing environment and integrating the various components that define a scientific data analysis environment. As long as the software development process of each component is separate from all other components, users of data analysis software are constantly interrupted in their scientific work in order to convert from one data format to another, or to move from one storage medium to another, or to switch from one user interface to another.

We also took an in-depth look at scientific visualization and its underlying concepts, current visualization systems, their contributions and their shortcomings. The role of data visualization is to stimulate mental processes different from quantitative data analysis, such as the perception of spatial relationships or the discovery of patterns or anomalies while browsing through large data sets. Visualization often leads to an intuitive understanding of the meaning of data values and their relationships by sacrificing accuracy in interpreting the data values. In order to be accurate in the interpretation, data values need to be measured, computed on, and compared to theoretical or empirical models (quantitative analysis). If visualization software hampers quantitative analysis (which happens with some commercial visualization products), its use is greatly diminished for astrophysical data analysis.

The software system STAR (Scientific Toolkit for Astrophysical Research) was developed as a prototype during the course of the project to better understand the pragmatic concerns raised in the project. STAR led to a better understanding on the importance of collaboration between astrophysicists and computer scientists.

Twenty-one examples of the use of visualization for astrophysical data are included with this report. Sixteen publications related to efforts performed during or initiated through work on this project are listed at the end of this report.

1. Introduction

1.1 Definitions[*]

"Visualization of Astrophysical Data" describes the application of graphical methods to enhance interpretation and meaning of data measured or computed to gain insight into scientific questions to be answered by astrophysicists. The goals of visualization go beyond the mere use of tools offered by computer graphics systems: visualization is directed towards the use of appropriate graphical methods to enhance current understanding of scientific data.

"Multi-spectral data" (or multi-wavelength, multi-variate data) describe celestial objects through a range of observations over the electromagnetic spectrum. This approach is often essential in the development of a complete physical model of an astronomical source. For example, to understand the energy budget of a cool star having a chromosphere and corona, it is necessary to measure broad-band fluxes in the X-ray and radio portions of the spectrum, as well as to acquire moderate-resolution emission line profiles in the ultraviolet and visible regions. To study the relationship between interstellar gas abundances and the kinematics of discrete clouds, equivalent widths from ultraviolet spectra strongly complement super-high-resolution optical spectroscopy. To properly attack the question of the central powerhouse of quasars and active galactic nuclei, it is imperative to record the energy distributions - and their variability - over virtually the entire electromagnetic spectrum.

The list of astrophysical problems best treated with multispectral analysis is nearly as long as the list of all astronomical research objectives today. The concept of multispectral astronomy is hardly new, but the tools with which to implement it have been lacking. The observational side of multispectral astronomy is being addressed with the new generation of space observatories of the 1990's and beyond; the analysis side -- software tools and environments -- is less developed. Our project sought to redress this lack by inspecting currently available software for the analysis and visualization of multispectral data and designing new software where shortcomings exist.

STAR (Scientific Toolkit for Astrophysical Research) is the software developed in the course of the project. STAR was developed as a prototype to prove the feasibility of user interface, software engineering and visualization techniques suggested in this report.

1.2 Goals of the project

The goal of this project was to support the scientific analysis of multi-spectral astrophysical data by means of scientific visualization. Scientific visualization offers its greatest value if it is not used as a method separate or alternative to other data analysis methods but rather in addition to these methods. Together with quantitative analysis of data, such as offered by statistical analysis, image or signal processing, visualization attempts to explore *all* information inherent in astrophysical data in the most effective way. Visualization is a vehicle of thinking (McKim, 1980), capable to explore spatial relationships between data items, making use of intuitive and holistic approaches to reasoning and the effortless identification of patterns and anomalies in large data sets. A scientist is in need of a multitude of methods and tools to explore *all* aspects of scientific data. It is important that all tools are integrated with each other and with the already existing environment of scientific data analysis.

In pursuing our goal to provide methods and tools for scientific visualization of multi-spectral astrophysical data, we therefore looked at scientific data analysis as one whole process, adding visualization tools to an already existing environment and integrating the various components that define a scientific data analysis environment. From this view the following work has emerged. With our work we hope to have added to a better understanding of the capabilities and challenges of Scientific Visualization in regards to astrophysical data analysis.

In the next chapter we first describe a taxonomy for data analysis. We then describe the processes involved with each analysis component in more detail. Chapter 3 lists examples of work performed under the grant. Chapter 4 recounts the history of work progress between 1989 and 1993. The final two chapters include recommendations to the sponsors of this project and a list of publications which derived from work on this project.

2. Astrophysical Data Analysis

2.1 Data analysis taxonomy

A data analysis taxonomy was developed while writing the proposal for this project (Ayres, Brugel, Domik and 1989) and is shown in Figure 1:

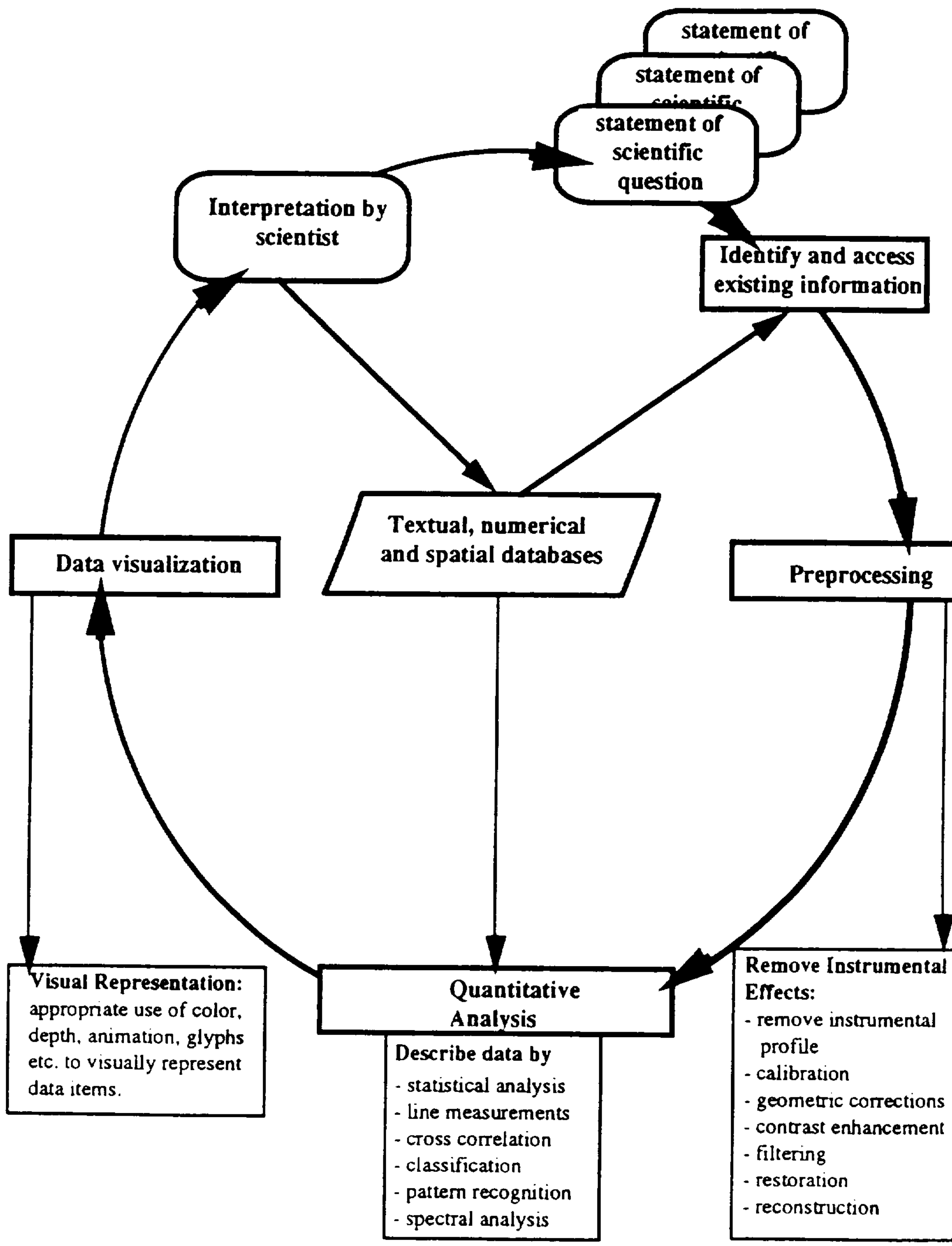


Figure 1: Taxonomy of Data Analysis.

This data analysis taxonomy provides a reference model to the environment of an astrophysicist utilizing multi-spectral data:

Besides the four domains of data analysis (identify and access existing information; preprocessing; quantitative analysis; visual representation) as defined above, we also include the user interface as a major component to the software environment of astrophysical data analysis. These five domains are explained in more detail below.

While a deeper understanding of the role of data visualization in the analysis process of scientific data evolved over the past years of project work, we are still using our original taxonomy with only minor modifications. We see this as a validation of our original model over the course of the project.

2.2 Identify and access existing information

The first task of the analysis system is to retrieve existing information and data pertinent to answering scientific questions of interest. Such information will come in various forms and structures such as images, spectra, tables or text (Figure 2). Some data will have spatial relevance, some will not. Access to different types of data bases (image, relational, textual; spatial and non-spatial) must be ensured so that a complete inventory of relevant information is presented to the scientists.

In order to improve the selection process for astrophysicists, information must be available at their fingertips. In STAR, efforts to improve the selection process were twofold: firstly, to provide access to existing data bases and secondly, to create new functionality for database selections.

2.2.1 Provide access to existing data bases

CASA's own catalog system was originally a Vms based relational data bases system containing several astrophysical catalogs, such as IRAS catalogs (point source, small scale structures, serendipitous survey, additional observations, low resolution spectra), the IUE Merged Log, SAO catalog, a cross index catalog (SAO/HD/GC/DM) and various smaller catalogs. Users could query the catalogs through "MCATS" (Multiple CATalogS, an application program developed at CASA), or directly through SQL (Structured Query Language) or RDO (DEC's proprietary query language).

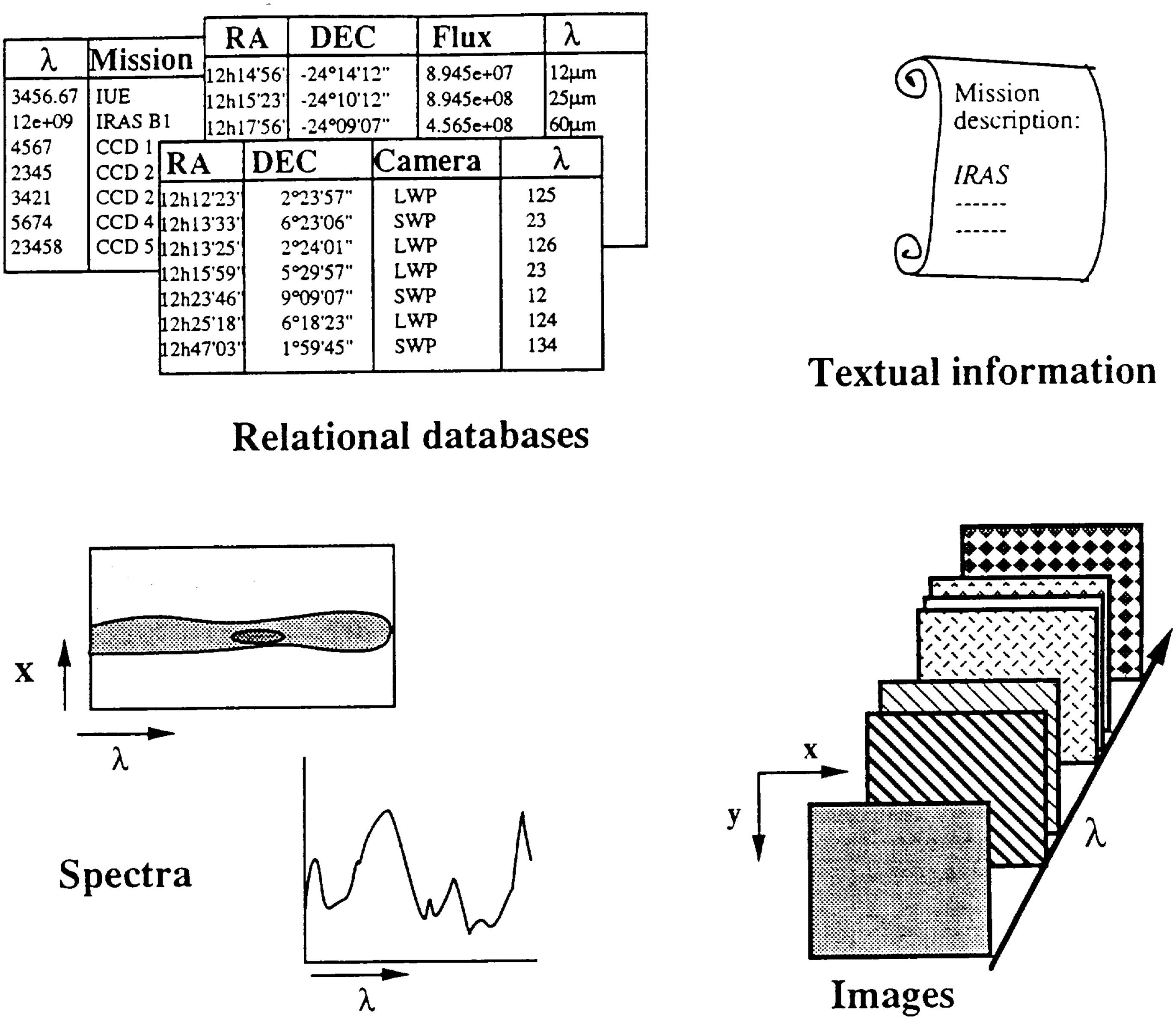


Figure 2: Images, spectra, tables or text describing astrophysical information.

The data bases named above included highly reduced information about data, but not the data itself. Data, such as IRAS skyflux images, or IUE spectra, were available locally at CASA in form of single files on hard disk, optical disk, or tape. Some of the database search functions would point to path- and filenames for data items when appropriate.

Parallel to the efforts of CASA, and on an immensely larger scale, NASA sponsored the development of ADS, the Astrophysical Data System, a remote network access system to data bases all across the astrophysical community.

Access to existing data bases and data base functions were made available through the CATALOG ACCESS menu in STAR. This included the use of MCATS as well as ADS, as long as these functions were supported on the workstation used to execute STAR.

2.2.2 Create new functionality for database selections

Information accessed through local or remote data bases must be visually inspected for verification of the goodness of quality or rejection and instigation of a new query. Because the visual inspection of data from data bases seemed an important aspect of the selection process, "quicklooks" of IRAS images and IUE spectra were made possible following a search through the appropriate catalogs. Quicklooks are replicas of the data at a reduced resolution, improving the scientist/machine dialogue by supporting a more discriminating selection of available data by the researchers. Information from astrophysical catalogs could be, if appropriate, overlaid on corresponding data sets.

2.3 Preprocessing and quantitative analysis

After raw data is acquired, either by new observations or from an existing archive, it must be subjected to a series of processing steps before it is suitable for visualization. We have divided the series of processing steps into two distinct classes: preprocessing and quantitative analysis. The division is based on where the major responsibility for software development historically has resided: preprocessing -- applying the necessary corrections and calibrations -- usually is the responsibility of the facility which originally acquired the raw data, the IUE Project in the case of IUE data for example; while quantitative analysis (measurements) -- extracting critical parameters from the corrected, calibrated data -- usually is the responsibility of the individual scientist, although in many cases a mission-specific facility will provide specialized tools to aid in the measurement of the primary data.

Preprocessing defines the major task of applying all known corrections and calibrations to the raw data in order to provide a final product that is as free as possible of systematic errors of instrumental origin, and which is presented in a readily usable form. An example is the geometrical correction, photometric linearization, and wavelength calibration of a ground-based CCD frame, or an IUE vidicon image. Preprocessing usually is the domain of the mission-specific processing center. However, it often is the case that the mission-specific center will release a data product that is tailored for the widest possible community of users, and is compatible with the (sometimes limited) capabilities of the mission processing system. There are many cases where sophisticated data enhancements techniques have been developed, often apart from the mission-specific center, which can significantly improve the quality of the data product. These enhancements typically are computationally intensive and can practically be applied to only small subsets of the overall data base. For the vast majority of scientific questions, the standard processing of the facility data is sufficient to provide the necessary answers. For a few applications, however, such data enhancement techniques are critical. In some cases, the necessary software is provided by the mission-center; in other cases, it is available only as custom-designed packages from the individual investigators.

Whereas the fundamental corrections to and calibrations of a particular primary data set usually are the exclusive domain of the production-processing team of a mission-specific center, the realm of quantitative analysis (measurements) is one more typically under the control of the individual scientist. In some cases, the scientist might make use of facility-developed software to expedite the measurement process; but in many cases the individual scientist will develop a specialized measurement approach that is carefully tailored to a specific research project. For example, one cool-star aficionado might desire the integrated flux of the Mg II k emission core in a particular star, whereas an interstellar-medium cognoscenti might desire to measure the strengths and radial velocities of the narrow absorption components in the k-line core of the same star. Although the scientific objectives might be quite different, and the object of attention - emission-line fluxes versus absorption-line parameters - might be quite dissimilar, the actual techniques by which one measures the relevant parameters might be essentially identical. Thus, a single generalized measurement tool - a Gaussian line fitting algorithm - might serve the purposes of a broad range of investigations based on widely differing data sets.

Expecting the scientist to choose from a variation of data sources during the first step of the data analysis process, we incorporated access to the following software packages into STAR: IUE/RDAF (IUE spectra reduction package), AIPS[2] (VLA radio map processing system), IRAF[3] (groundbased CCD reduction, as well as specialized processing modules for HST and ROSAT), and a series of inhouse-devel-

2. Astronomical Image Processing System
3. Image Reduction and Analysis Facility

oped image-enhancement procedures for the IRAS HCONs. Access to these packages provided mission-specific preprocessing tools as well as general measurement tools.

Because data formats varied with the use of different software packages, compatibility needed to be achieved between individual data items before multispectral analysis could be performed. STAR's solution to data transfer between software packages was to include import/export functions to transform between various data formats. Expecting that access to the above named packages would offer sufficient generalized preprocessing and measurement tools, only a few specialized tools were developed during the course of the project.

Newly developed tools included two semi-automatic preprocessing techniques for IRAS skyflux images that were designed and implemented as part of STAR. These preprocessing programs remove the background of IRAS skyflux images (zodiacal light) and reduce -- or remove -- the periodic stripes in skyflux data. The algorithms involved and results are explained and shown in the enclosed publication "Workstation-based Preprocessing of IRAS Sky-Flux Images", which is included as an appendix to this report. Before availability of the reprocessed skyflux data in 1992, these modules had a great value for visual and computational comparisons between different IRAS wavelengths bands.

New measurement tools were also include with some of the visualization functions, e.g. isosurface representation and flux measurements, as described in the next section.

2.4 Visual representation

2.4.1 Visual representation of data

Information extraction and preprocessing will produce corrected calibrated scientific data in highly-reduced form. The scientist now is confronted with a bulk of data for interpretation: data from individual missions; cross-correlated multi-wavelength data; and data in one or multiple dimensions. In our assumption of the importance of multi spectral data we stress the availability and use of information derived from different observation sources; consequently, however, the interpretation of results becomes more complex. The human mind is weak in making connections between large tables of numerical results. The expression "A picture is worth a thousand words" refers to the fact that the human visual system is capable of interpreting image data at an incredible faster rate than the same amount of data in tabular form. Therefore visualizing data, instead of simply reviewing large tables of numbers, will enhance the pace of data interpretation.

For example, it is much harder to interpret three single images, placed next to each other, than it is to interpret one color image incorporating these three images simultaneously. In practice one might choose a perceptive three dimensional color space, e.g. hue, lightness and saturation, that fits the principles of color perception of the human mind. Expressing multi-wavelength data characteristics by various colors at the same spatial coordinates results in a natural visual "format" for the scientist.

Other visualization techniques, besides color, also aid in the simultaneous presentation of multi-wavelength data sets: e.g. depth and animation. Both may add an additional dimension to spatial and brightness information in an image. In addition to sophisticated new techniques that aid in the process of simultaneous display of several layers of pictorial information, there are other familiar, fundamental visualization techniques that have been utilized in the past: two dimensional plots, pseudocolor displays, or graphic overlays on images and spectra.

2.4.2 Visualization concepts

The role of data visualization is to stimulate mental processes different from quantitative data analysis. Visual data analysis offers an overview of data characteristics through browsing, often leading to an intuitive understanding of data characteristics and their relationships by sacrificing accuracy in interpreting the data values. Because the human visual system emphasizes spatial relationships, up to three data characteristics can be represented in a natural, intuitive way in form of spatial dimensions. Data visualization is an indirect way of interpreting data: instead of being interpreted from their natural, usually quantitative characteristics, data are first encoded into a pictorial representation. The encoding process bears the danger of creating artifacts and therefore missing the correct interpretation: e.g. abrupt color changes may mislead by pointing to discontinuities in a data set; subjective assessments of patterns may lead to other misinterpretations.

A visual representation of data values should take into account the data characteristics as well as the interpretation intent of a scientist (Mackinlay, 1986; Wehrend and Lewis, 1990; Robertson, 1990) to encode data values into a pictorial representation from which a scientific interpretation can be derived (shown in Figure 3).

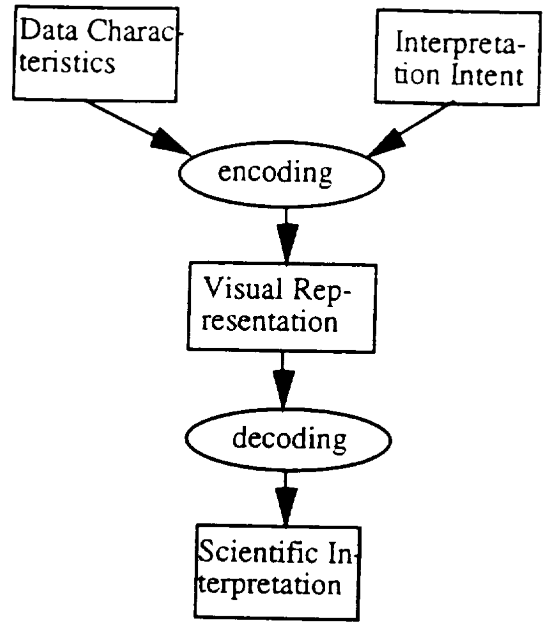


Figure 3: Process to encode data values into a visual representation.

2.4.3 Visualization tools developed for STAR

IDL[4], the software platform chosen for the development of STAR, incorporated many basic and higher-level visualization techniques (e.g. display of one and two dimensional data sets; pseudocoloring; perspective projections). Over the past years many visualization techniques that we had planned to implement, or started to implement, have recently become available, specifically in the area of three dimensional display techniques, e.g. isosurfaces or data slicers.

During our first project years we found a high demand for simple visualization tools that allowed interactive manipulation as compared to demands for complex visualization tools. Data slicers and 3-d representations of structures did not get much attention by CASA's scientists in the early stages of our inquiries. As a consequence, in the beginnings of the project, design and development of visualization tools concentrated on these early needs. As can be seen in chapter 3, latter demands were for highly dimensional and more complex visualization techniques.

Therefore simple visualization techniques, such as profiling objects, color table editors and switch boards, interactive data type conversions (the dynamic range of some astrophysical data, e.g. IRAS data, is very high compared to the 256 colors/gray values visible on a standard graphics workstation), were included into early versions of STAR. Later more complex visualization techniques followed,

4. Interactive Data Language by Research Systems, Inc.

such as interactive color coding techniques (e.g. HLS[5], HSV[6], 24-bit to 8-bit color compression), volumetric data displays (e.g. iso surface display) and data slicers.

2.4.4 Feasibility study with AVS[7] and IDL

In addition to developing new tools we studied the feasibility of well known state-of-the-art commercial visualization packages (AVS and IDL) for astrophysical data analysis. Results of these experiments are included in the example section. While we found AVS to be a sophisticated (and easy to use) graphics package, it would not fit into the astrophysical environment in its current status. Some of the short-comings we found with AVS were the lack of its use for quantitative analysis. While AVS nicely encoded data fields into pictures, e.g. showing contour lines of a 2-d array of data values, it was often impossible to retrieve the original numbers from the visual representations. The lack of data formats appropriate for the astronomical community as well as the relatively high licence fees furthermore restricted the use of AVS for astrophysical environments.

IDL has long found its place in astrophysical data analysis. Therefore a fair amount of visualizations already allow interactive, quantitative analysis. In the case of the "data slicer" and the "iso surface" displays, we have expanded the original source code to permit direct measurements and computations on the surface of arbitrary slices and inside isosurfaces. These new visualization/analysis tools are described further in the appendix.

2.4.5 3-d Interaction devices

With the exploration of volumetric data, the problem of how to interact with 3-dimensional data on a 2-d screen became obvious. Input devices (e.g. mouse) as well as output devices (e.g. 8-bit color workstation displays) that are currently available at reasonable prices are limiting the scientist.

An example might illustrate the problem in a better way: A scientist looks at an isosurface display of an astrophysical data cube, as pictured in Figure 18. By rotating the cube and its content, the 2-d screen clearly conveys the effect of a 3-d object. In order to analyze part of the cube in more detail, the scientist wishes to extract a subcube. At this moment several problems hamper the interaction:

- In order to "reach into" the cube (spatial positioning), the rotating movement has to be stopped. This results in collapsing the 3-d information into a static 2-

5. Hue-Lightness-Saturation
6. Hue-Saturation-Value
7. Application Visualization System by AVS, Inc.

d perspective projection. An output device such as a stereo-screen coupled with stereo glasses or a virtual environment would solve this problem.

- A mouse is moved over a 2-d surface to simulate 3-d movement in various ways: e.g. any 2-d coordinate plane in 3-d space can be selected consecutively to reach any spatial position in 3-d space. While 3-d spatial feedback is already poor on the 2-d screen, the separate movement is also non-intuitive to a natural gesture of grasping an object.

In view of the lack of 3-d input and output devices that are effective as well as cheap, we performed a short study on the availability, effectiveness and costs of such devices. The result of this study is included in the appendix of this report.

2.5 User interface design

In order to build a system that is liked and used by researchers, we interviewed our potential users on their likes and dislikes of analysis systems, specifically noting aspects of the user interface (Mickus, 1991; see section 6.3). The breadth of answers lends itself to a discussion that goes beyond the scope of this report. To summarize, common wishes were:

- support major application systems, such as IRAF, IUE/RDAF, IDL, AIPS;
- use on-line documentation of software;
- use windows, widgets and interactive tools;
- offer a dynamic system, e.g. easily expandable if new software is to be included, or when new application packages are to be installed.

STAR's user interface therefore incorporated all these aspects:

- In its start-up state the system allows access to IRAF, IUE/RDAF, IDL, AIPS. Customized startup files (e.g. for IRAF and IDL) can be created at one time and used for future purposes.
- On-line help and documentation is available on application software packages developed in-house. In order to keep documentation of user-built modules as updated as possible, we provided an automatic documentation tool, which extracts comments from the source code to make them available as source code documentation. Documentation for external software packages is the responsibility of the developers.
- STAR is built in a workstation environment, on top of X-windows and IDL (later versions on IDL/widgets) making therefore extensive use of windows, widgets and interactive tools.

Feedback from scientists on the use of STAR was solicited throughout the design and development period. New research results in the area of Human/Computer Interaction (HCI) were employed to encourage feedback about the scientist's desires. We were able to take advantage of a strong HCI research environment available at the University of Colorado. Under consultancy of Dr. Lewis, cognitive design techniques were used throughout the first 1.5 project years to solicit feedback from CASA's scientists about the user interface design and visualization tools.

Besides soliciting feedback at CASA, we also solicited feedback from a larger audience outside the University of Colorado about the design goals of STAR (see also section 6.5). New aspects and desires of these audiences further influenced the design of STAR.

A detailed review of design issues concerning the data analysis environment is given in the enclosed publication "Design and Development of a Data Visualization System in a Workstation Environment", which is attached as an appendix to this report.

3. Examples of Work in the Project

3.1 Data analysis cycle and user interface

Figure 4 and 5 show in a schematic and actual view the main user interface of STAR. Pull-down menus reveal the highest level of functions to perform

DATA INPUT/OUTPUT	Read in / write out data to tape or disk; conversions between various data formats; quick saving and restoring of data variables
CATALOG ACCESS	Retrieve information from databases by accessing databases/catalogs directly or by executing database programs
PREPROCESSING	Apply necessary corrections and calibrations to data
ANALYSIS	Extract, measure and mark objects
VISUALIZATION	Convert data to visual representations

Square buttons denote external software packages that may be executed by clicking on the corresponding button. Round buttons reveal different menu functions for CCD, IUE or IRAS based analysis. A "PROBLEM" button connects the user to STAR's software developers to leave complaints, demands or recommendations. A status window reports the current status of the software system, e.g. "Error"; "Waiting for user input..."; "Computing...".

Figure 6 shows one submenu of the "CATALOG ACCESS" button: MCATS, CASA's local Vms catalog access program, offers access to locally stored astrophysical catalogs.

Figure 7 shows another "CATALOG ACCESS" function: "Quicklook" displays data files selected through database programs in reduced resolution on the screen.

Figure 8 shows a "PREPROCESSING" function to flatten IRAS skyflux images. The upper left image shows the original image, the lower right the result after flattening.

Figure 9 shows the interactive measurement of fluxes and positions in the image by moving mouse/cursor over image pixels. Corresponding horizontal and vertical profiles are plotted when clicking the mouse button.

Figure 10 shows an axonometric view of an IRAS skyflux image combined with its contour lines.

3.2 Destriping and flattening of IRAS skyflux images

Two preprocessing techniques to flatten and destripe IRAS skyflux images were designed, documented and implemented in order to enable visual and computational comparisons between different IRAS wavelength bands. The flattening process (removal of zodiacal light) is shown in Figure 8. Several examples and detailed explanations are given in Appendix A.

3.3 Simple visualization techniques

Figure 11 shows an interactive surface plot of flux values. The control window to the right defines the current view point position.

Figure 12 visually correlates three of the four IRAS skyflux bands shown in the upper picture and displays the result as a color picture on an I^2S/IVAS 24-bit color image display station. A similar color coding technique is used to correlate three IRAS skyflux bands on a (much cheaper) 8-bit color display in Figure 13. The difference in the appearance of the pictures relates to the use of different input images in the encoding algorithm.

An interactive switch board (Figure 14) facilitates the use of available color tables.

A high dynamic range of flux values (represented by the y-axis in window

"STRETCH") is being converted interactively to the available color range of the graphics workstation (Figure 15).

Figure 16 shows a flexible adjustment tool of the internal color lookup table. Linear as well as non-linear transformations can be defined interactively. Besides user-defined conversions between data values and display values, predefined transformations, e.g. statistical stretches, are available.

Figure 17 shows the interactive zooming tool.

3.4 Data cube visualizations

Collaboration with Dr. John Bally at CASA started beginning of 1992 and gave us an opportunity to present new visualization techniques as recently developed (or still under development) by the computer graphics community. Following is a short explanation of Bally's data sets and scientific goals in order to better understand the visual representations chosen.

Data is collected by a 7 m telescope dish owned by ATT Bell Labs in New Jersey. It operates at a frequency between 23 to 43 Ghz, corresponding to a wavelength of 1.3 cm to .7 cm. The collected data is in form of 2-d image tiles for each measured frequency. Processing of the collected raw data values from the heterodyne receiver results in even gridded data values defined in three dimensions (spatial, spatial, frequency). The data values correspond to a count of carbon monoxide molecules at that specific spatial location and frequency. Data values may range between -32000 and +32000.

Carbon monoxide is used to trace molecular clouds. It is important to understand changes of the molecular cloud in space as well as in frequency. Therefore, scientific tasks in interpreting data values are: observe, if the cloud is expanding; look for indications that the cloud is collapsing; in what direction is it moving; identify dense matter at each stage of frequency.

It is important to express essential data characteristics in the resulting visual representations. In the case of the astrophysical data cubes, such essential characteristics are spatial location as well as frequency and the data value itself. Leaving both spatial dimensions in their natural form and mapping frequency into a third spatial dimension created an even gridded cube ("data cube") with the data values expressed as voxels.

However, the various slices of spatial data values could also collapse into one single slice, where spatial dimensions are represented in their natural form, but various data values along one frequency dimension are expressed in one pictorial "glyph".

It is important to represent the data is an effective way, so that the decoding process from pictures to scientific interpretation is quick and accurate. The following visual representations were chosen and discussed with Dr. John Bally:

3.4.1 Iso surfaces

Data values of a certain threshold were connected to create iso surfaces. This is a well known rendering technique of a data cube representation. In this representation, the overall shape of the data can be observed as well as isolated volumes (see Figure 18). Understanding the overall distribution of the carbon monoxide in the given spatial-spectral dimensions is important in order to understand the detailed quantitative information. The iso surface representation can be enhanced by adding individual slices through the data cube (see Figure 19) or by using several transparent iso surfaces (Figures 20).

3.4.2 Translucent representations

Rays penetrate the data cube from a chosen point-of-view and accumulate values of opacity related to corresponding data values. This representation inflicts a transparent characteristic on the molecular clouds, very much like the visual form of real clouds. It allows to look <u>into</u> the cloud as opposed to observing the surface only. Because the scientist felt a natural understanding of this representation, it was favored as compared to any other representation.

Figure 21 shows a translucent rendering of the cube by looking at the data from one side: one spatial dimension increases to the right, the frequency increases from bottom up. The rapid changes of the data values in the mid-frequencies show special characteristics of carbon monoxide at these frequencies. Figure 22 shows the same data cube using the same representation looking from top down onto the cube.

Figure 23 uses colors relevant to the frequency content: high frequencies are colored blue, low frequencies are colored red. During rotation of the display, the viewer will therefore be constantly aware of the frequency range s/e is looking at.

3.4.3 Data slicer

To monitor the change of one data value in relation to its neighbor values, a data slicer was used. Even though a data slicer can only monitor the neighbors surrounding a certain data value inside a plane, flexibility in placing the slices inside the cube can monitor various changes. Figure 24 shows four slices cutting through the cube parallel to the x/y plane, enhancing the understanding of the movement of the cloud through frequency. Figure 25 shows three orthogonal slices through the data, intersecting in the center of the cube. Figure 26 shows an arbitrary slice through a data cube.

3.4.4 Glyphs

To collapse all (or a subset of) data slices along the frequency dimension into one single two dimensional image, one must map all data values along one frequency dimension into one complex "glyph". The difference of one glyph from its neighboring glyph relates to the spectral characteristics of carbon monoxide and can be interpreted accordingly.

The resulting image can also be seen as one entity, therefore allowing interpretation of the overall distribution and change of carbon monoxide in the data cube. Visual representations of collapsed data slices leave it up to the human visual system to decide if the focus is on large-scale or small-scale structures.

Figure 27 shows a glyph representation of nine consecutive slices: color slices are used inside each red square to indicate various spectral responses at each spatial location. Figure 28 encodes five slices into five characteristics of a cube: width, height, depth, color and view point.

3.5 Expansions to commercial visualization software

One major point of dissatisfaction for Dr. Bally was the fact that in many cases of commercial visualization software no interaction could take place on the visual display itself. E.g. after displaying isosurfaces of a data cube, availability for quantitative analysis, such as integrating flux values inside an isosurface, is very important to interpret data. We therefore modified the IDL modules of "data slicer" and "isosurfaces" to include quantitative analysis.

Figure 29 shows three isosurfaces and a corresponding count of flux values inside these isosurfaces.

Figure 30 shows a grid in an arbitrary plane through an isosurface. The grid aids in probing data values inside the data cube in relation to the isosurface.

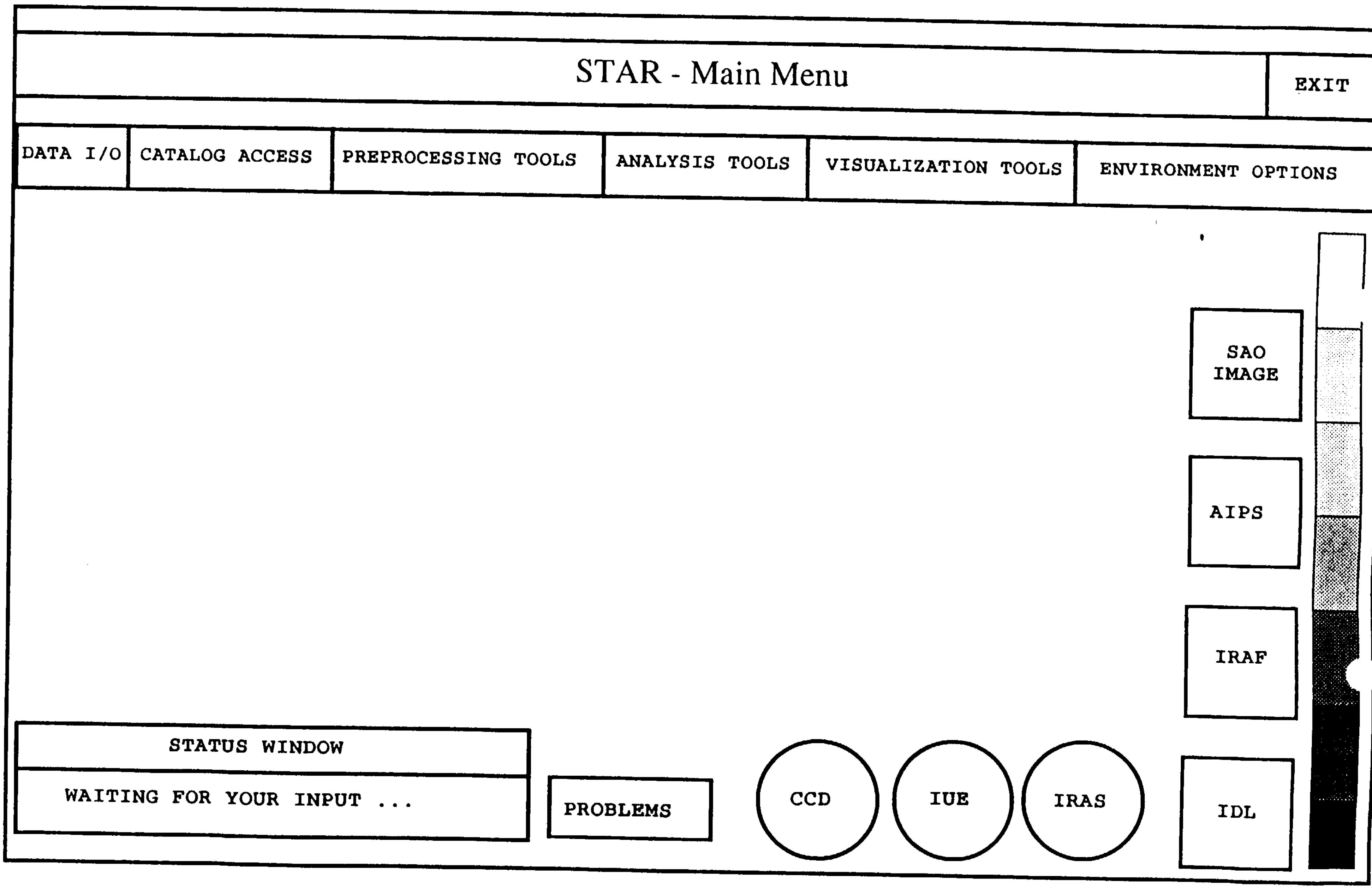
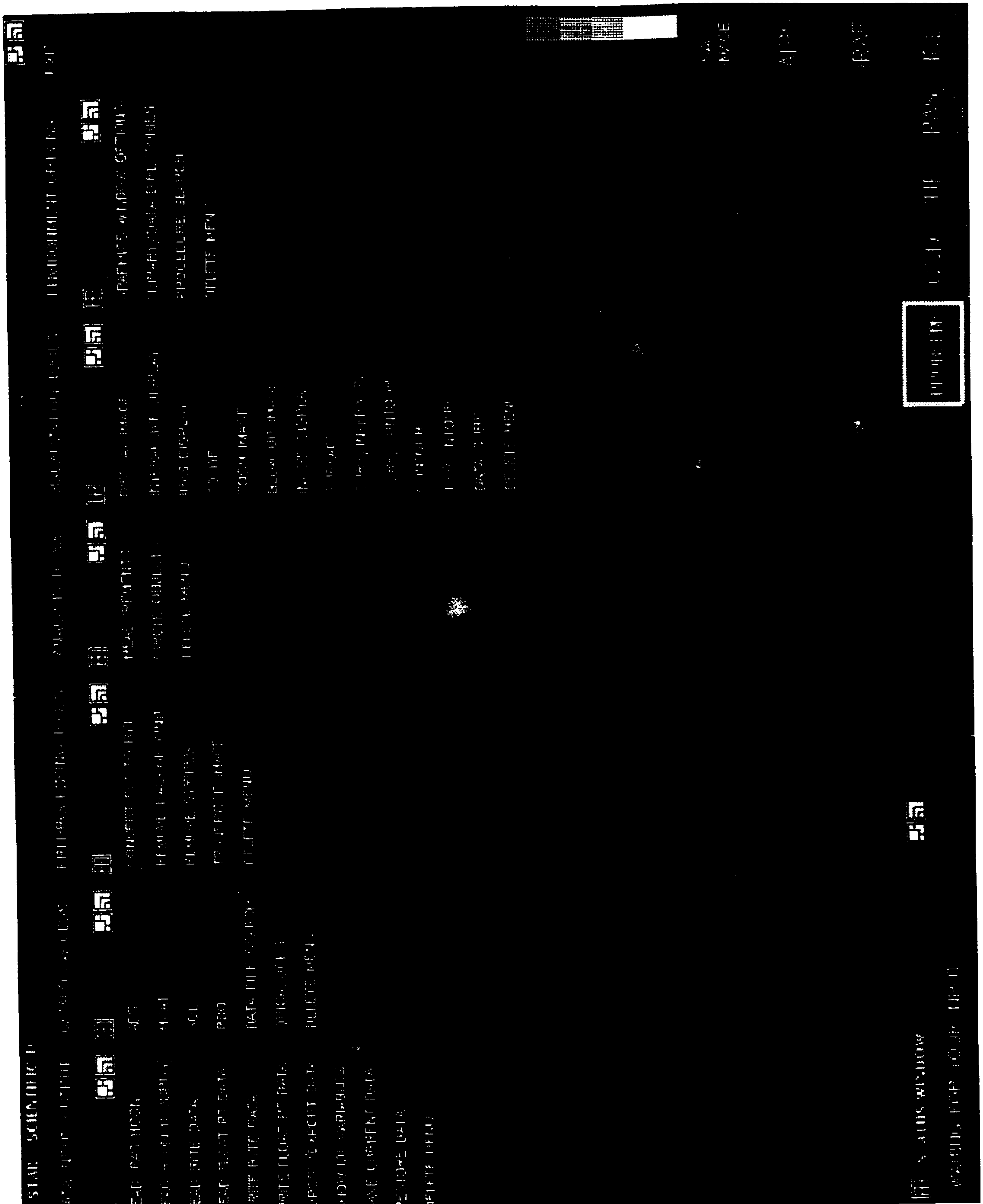


Figure 4: Schematic view of the main user interface of STAR.

Figure 5: Actual view of the main user interface of STAR.



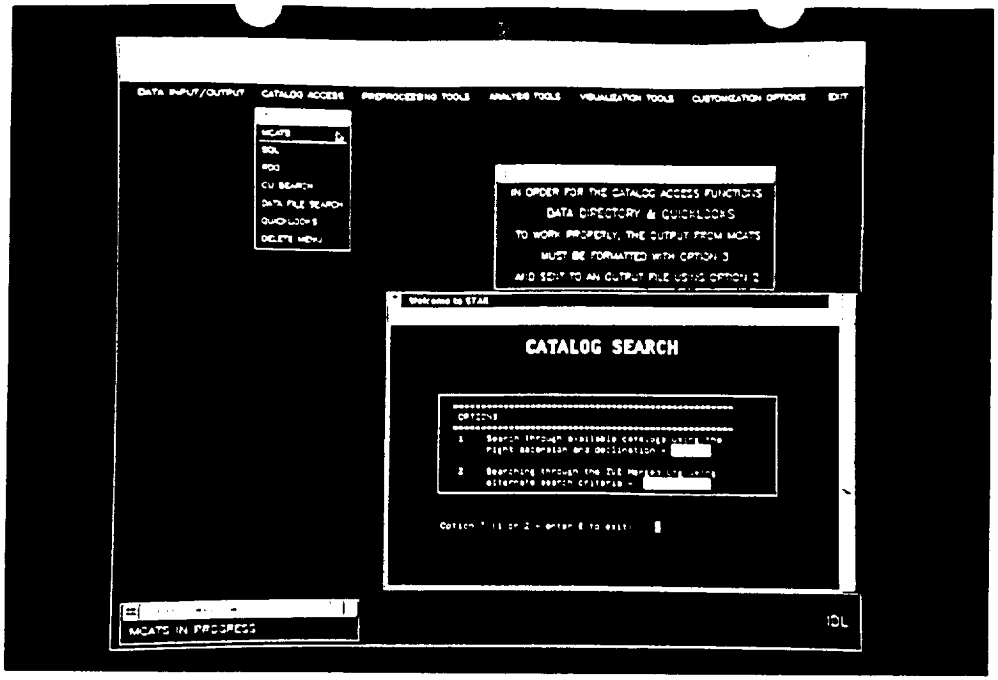


Figure 6: Submenu of "CATALOG ACCESS" button: MCATS, CASA's local catalog access program, offers access to locally stored astrophysical catalogs.

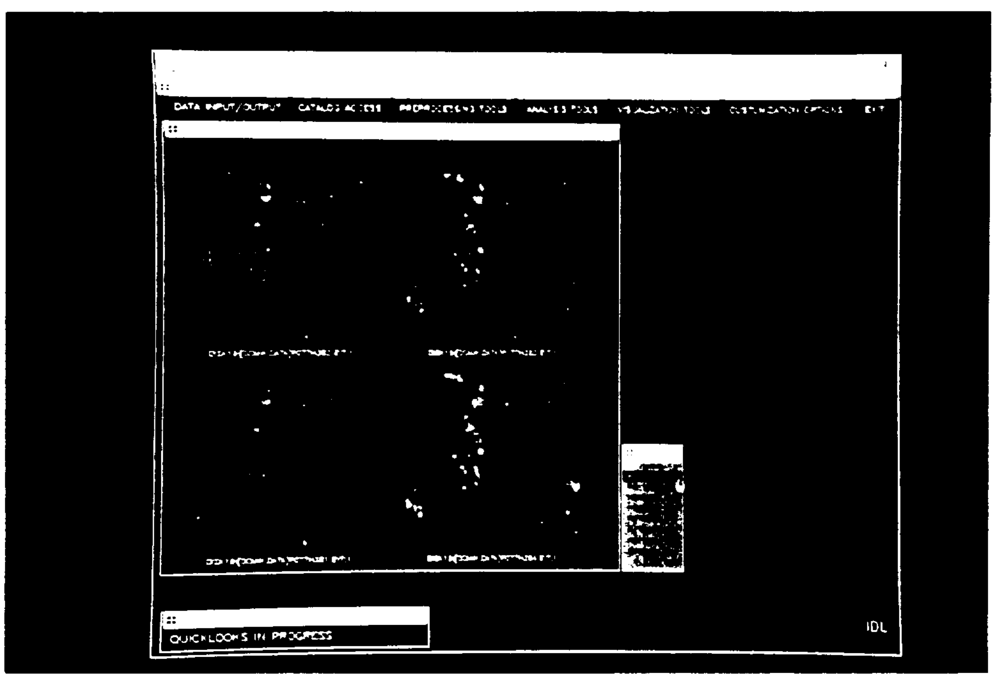


Figure 7: Another "CATALOG ACCESS" function: "Quicklook" displays data files selected through database functions in reduced resolution on the screen.

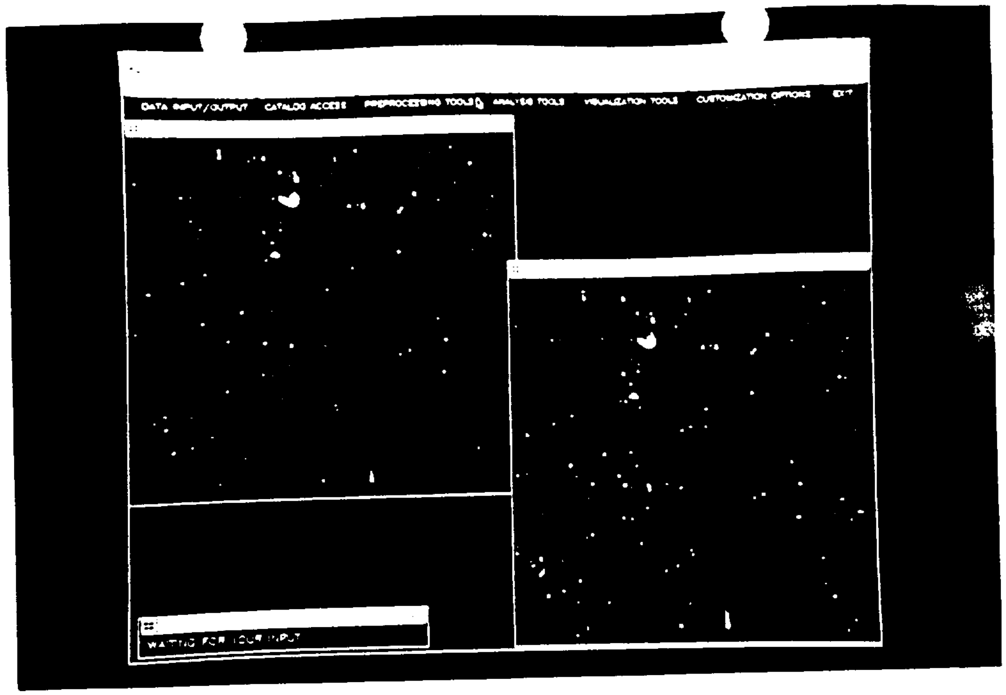


Figure 8: "PREPROCESSING" function to flatten IRAS skyflux images. The upper left image shows the original image, the lower right the result after flattening.

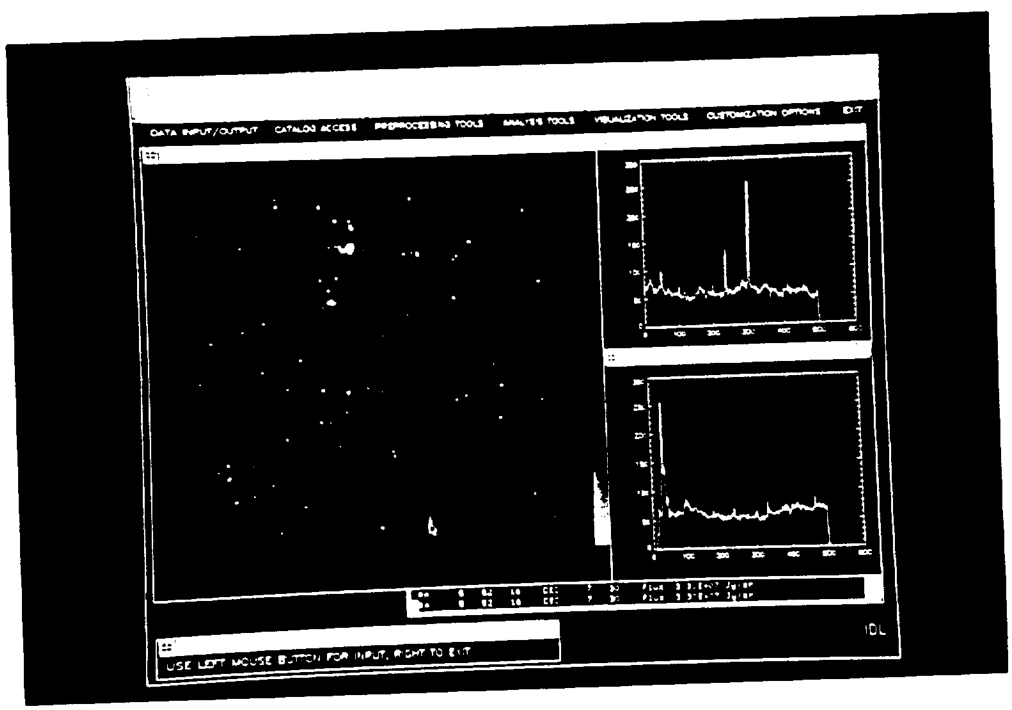


Figure 9: Interactive measurement of fluxes and positions in the image by moving mouse/cursor over image pixels. Corresponding horizontal and vertical profiles are plotted when clicking the mouse button.

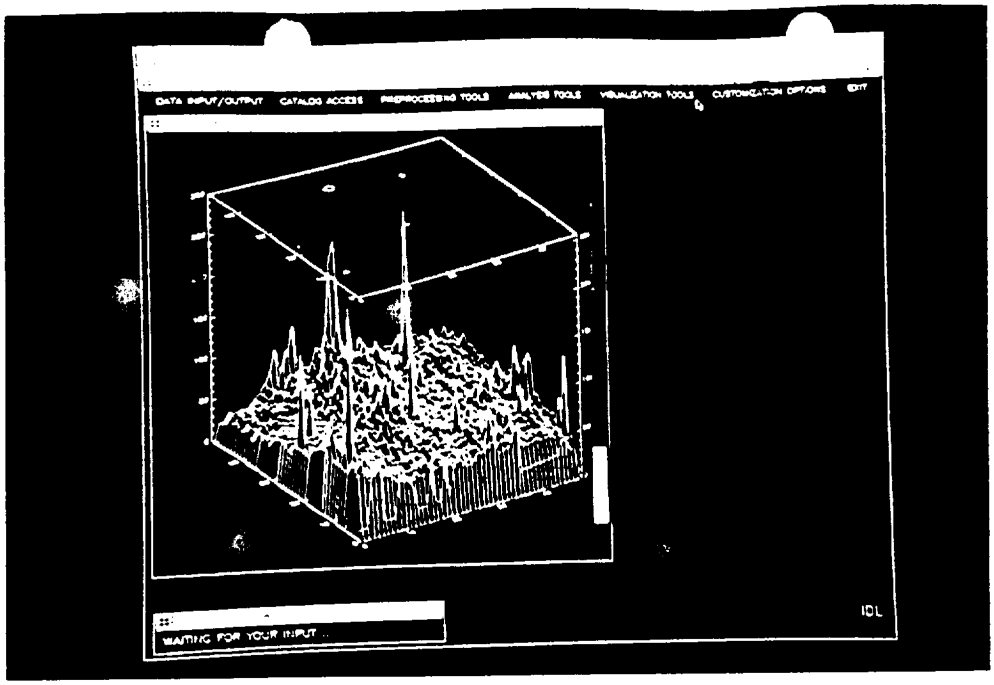


Figure 10: Axonometric view of an IRAS skyflux image combined with its contour lines.

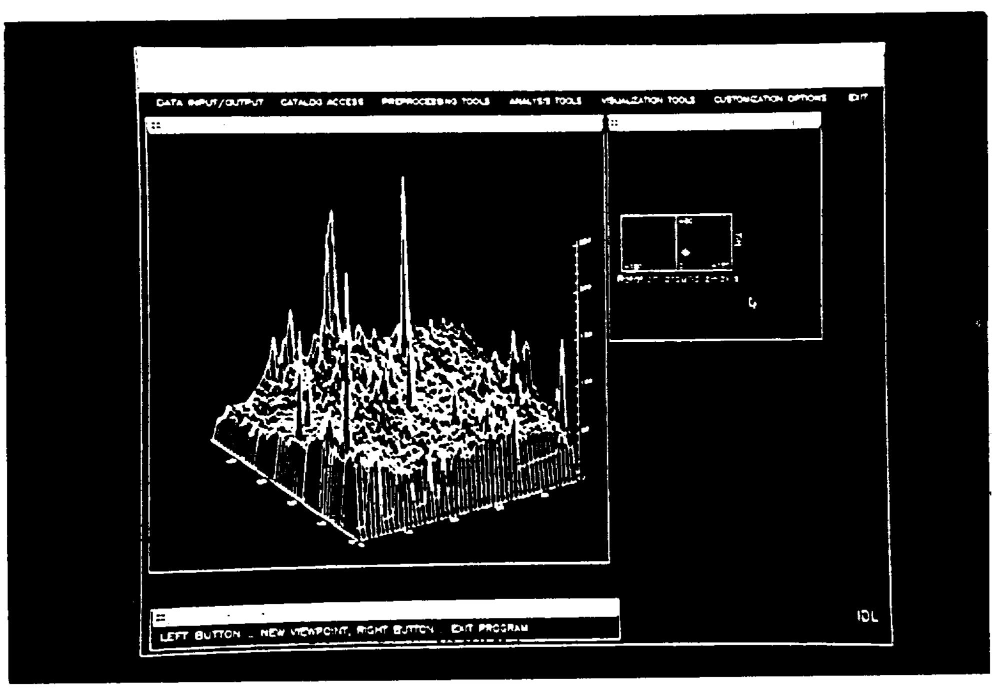


Figure 11: Interactive surface plot display of flux values. The control window to the right defines the current view point position.

Figure 12: Visually correlates three of the four IRAS skyflux bands in the upper picture and displays the result as a color picture on an I^2S/IVAS 24-bit color image display station.

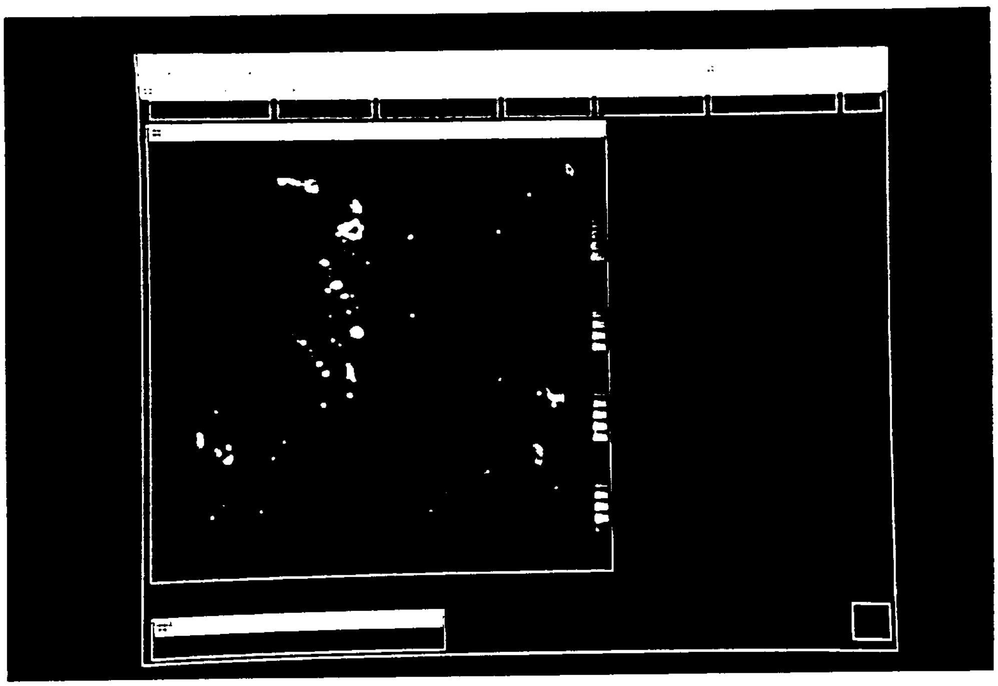


Figure 13: Visually correlates three of four IRAS skyflux bands (see Figure 12) and displays the result as a color picture on the monitor of an 8-bit graphics display.

Figure 14: An interactive switch board facilitates the use of available color tables.

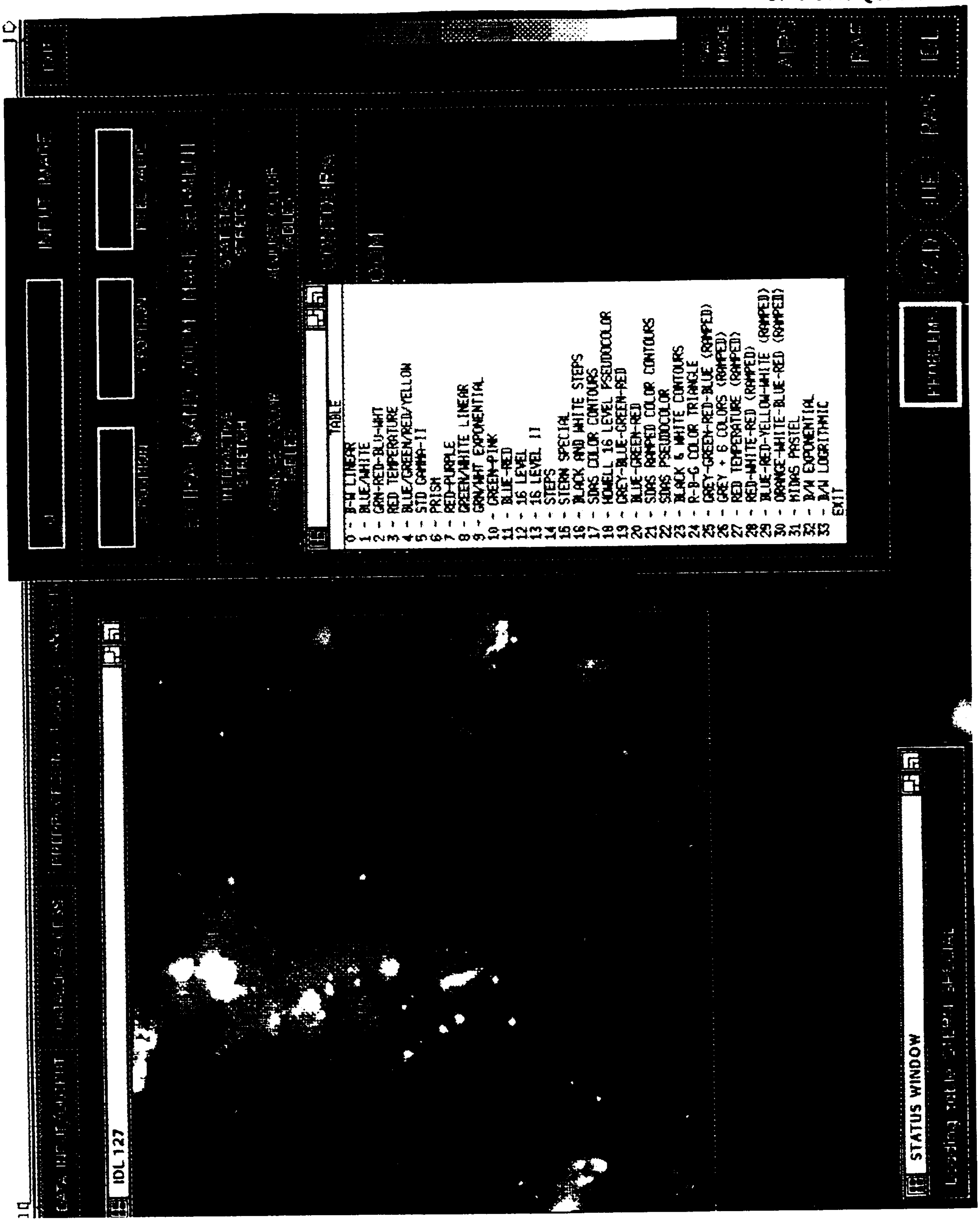


Figure 15: A high dynamic range of flux values (y-axis in window "STRETCH") is being converted interactively to the available color range of the graphics workstation.

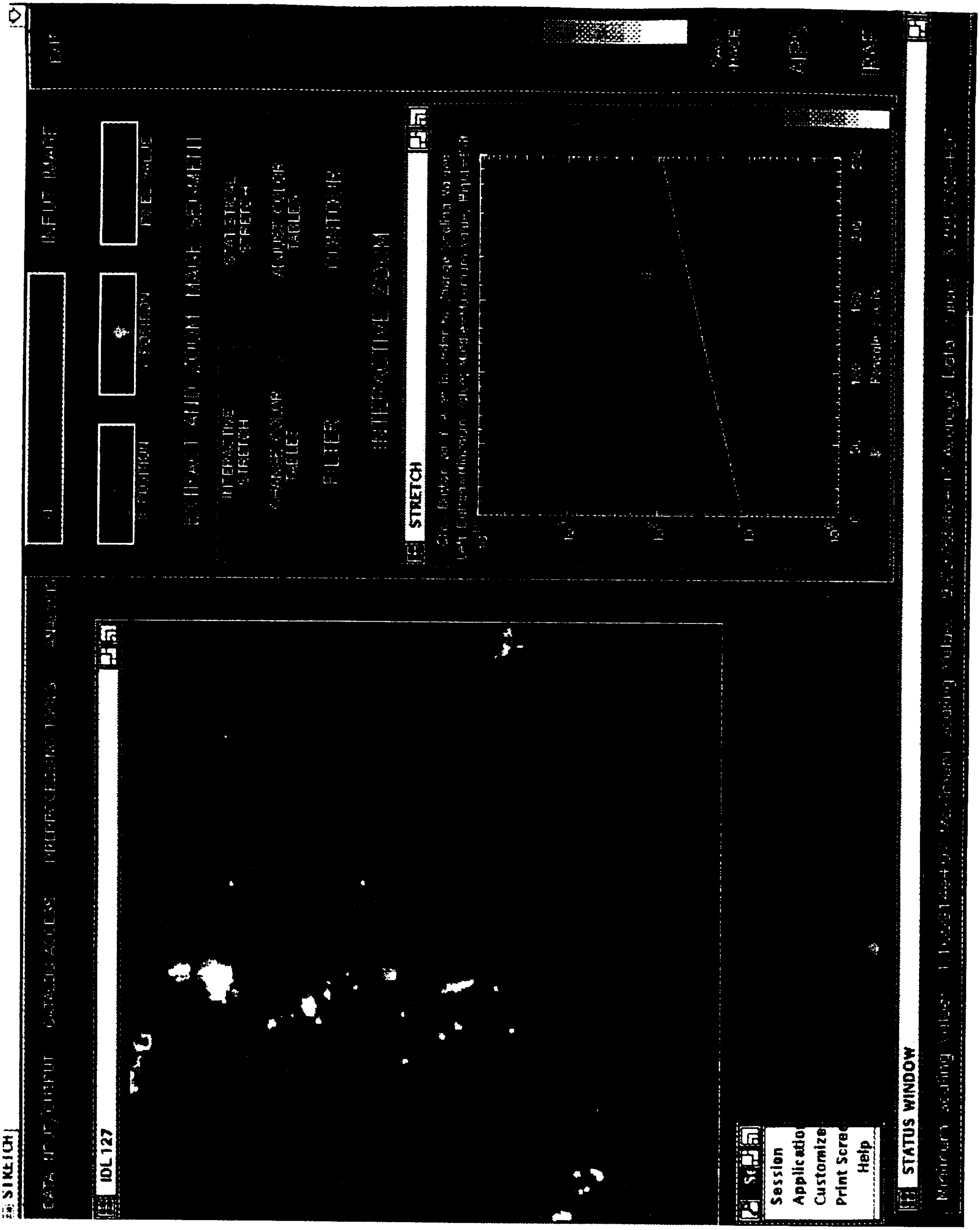
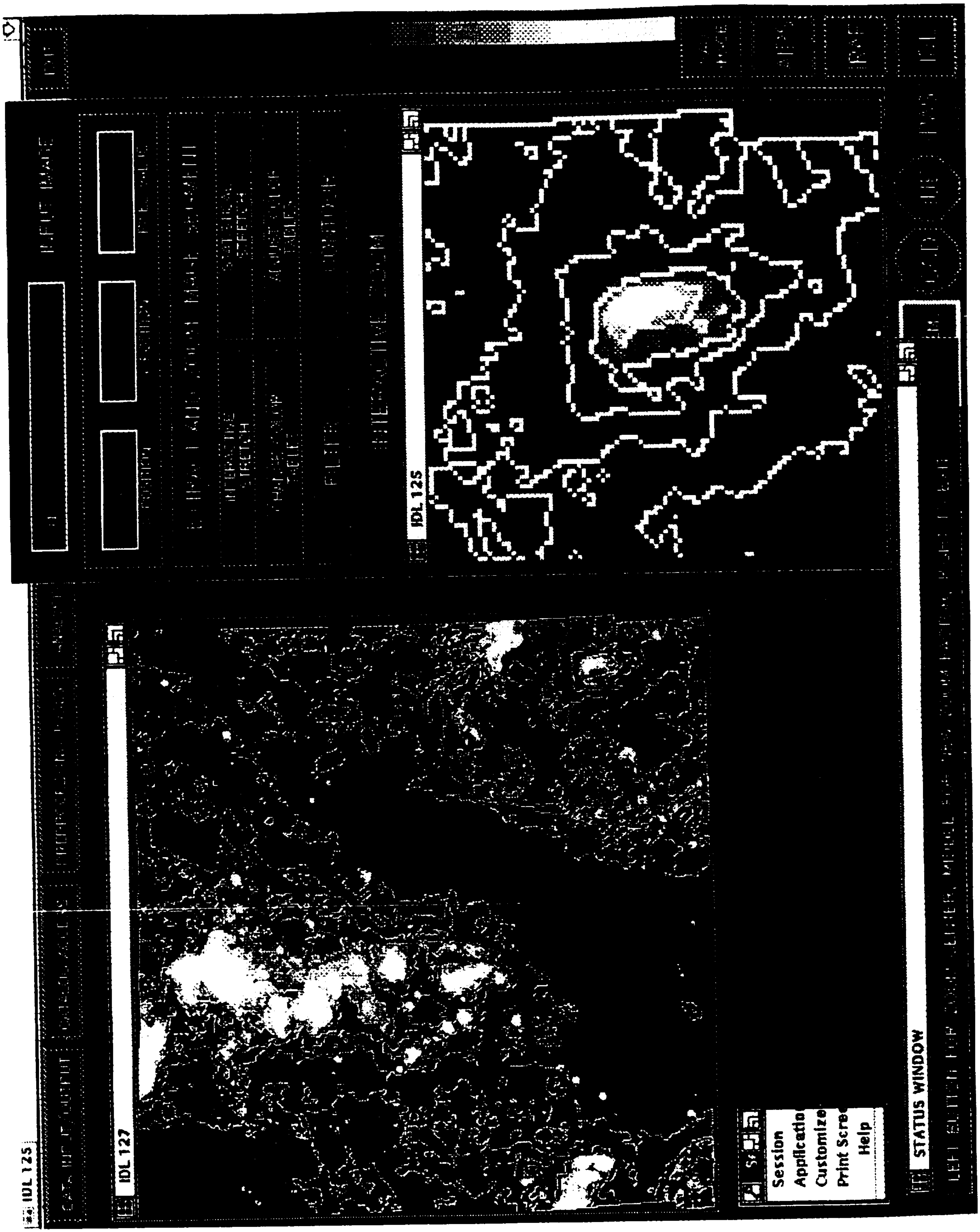


Figure 16: Adjustment tool of the internal color lookup table. Linear as well as non-linear transfc ations can be defined interactive Besides user-defined conversions between data values and display values, predefined transformations, e.g. statistical stretches, are available.

Figure 17: Interactive zooming and contouring tool.



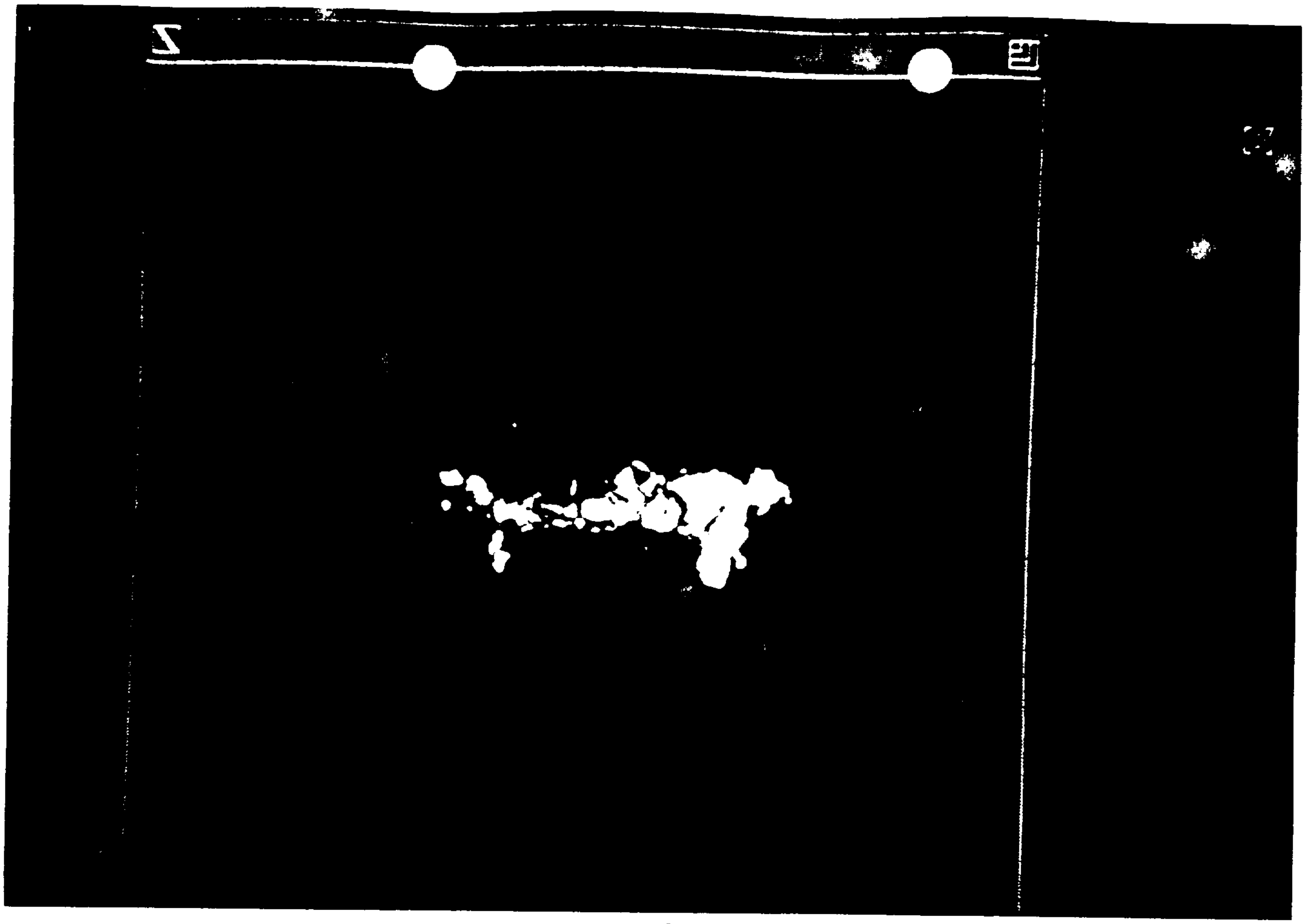


Figure 18: Isosurfaces.

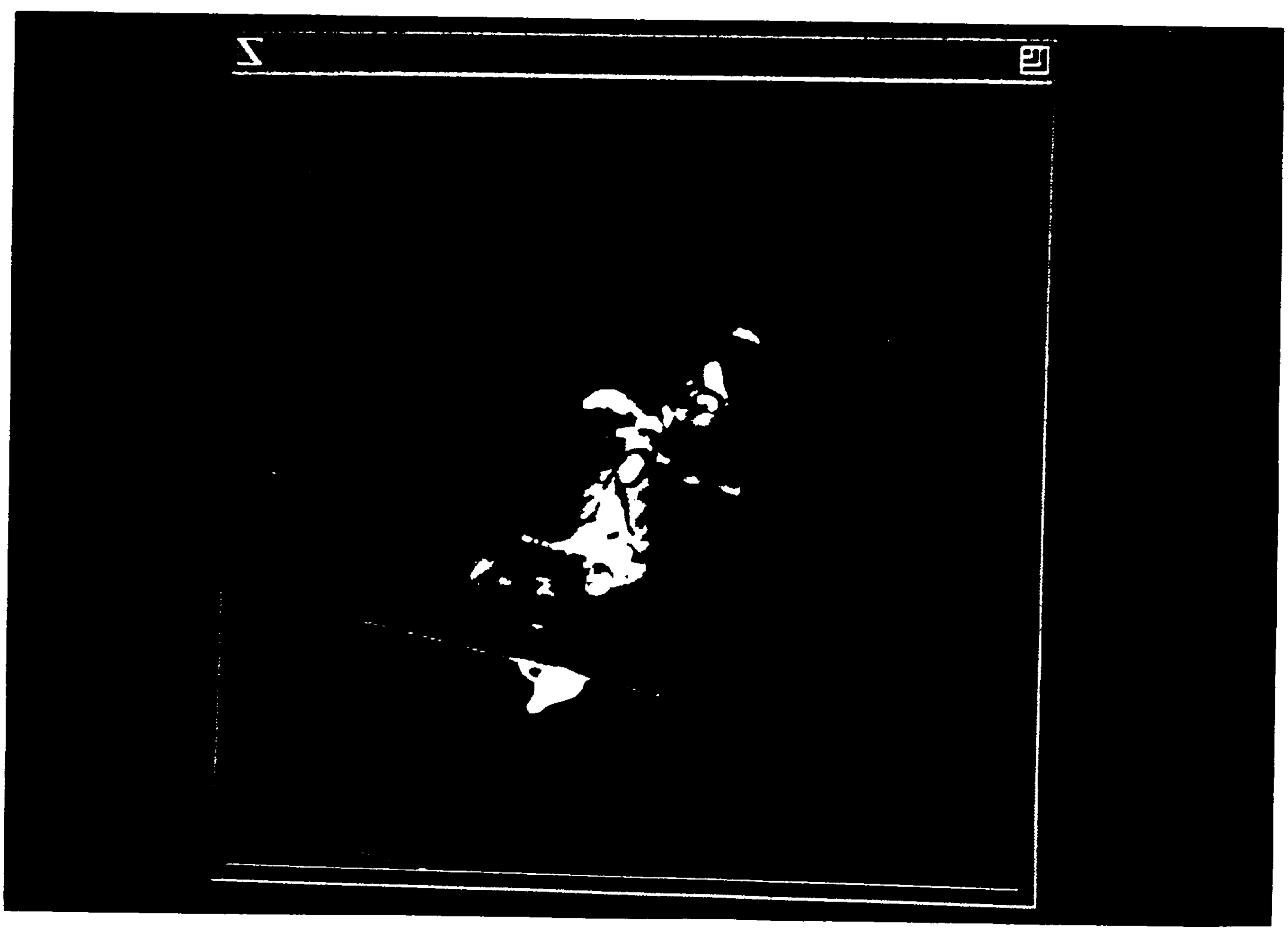


Figure 19: Isosurfaces combined with single image slice

Two transparent isosurfaces
viewing from top

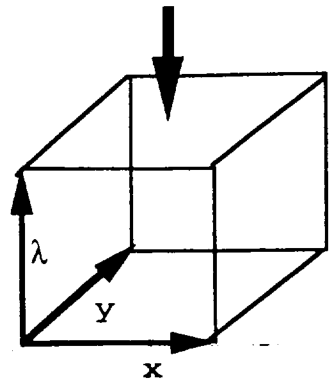
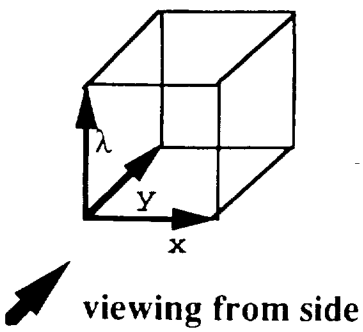


Figure 20: Transparent isosurfaces.



dark blue: low molecule count
cyan: medium molecule count
green: high molecule count
yellow: highest molecule count

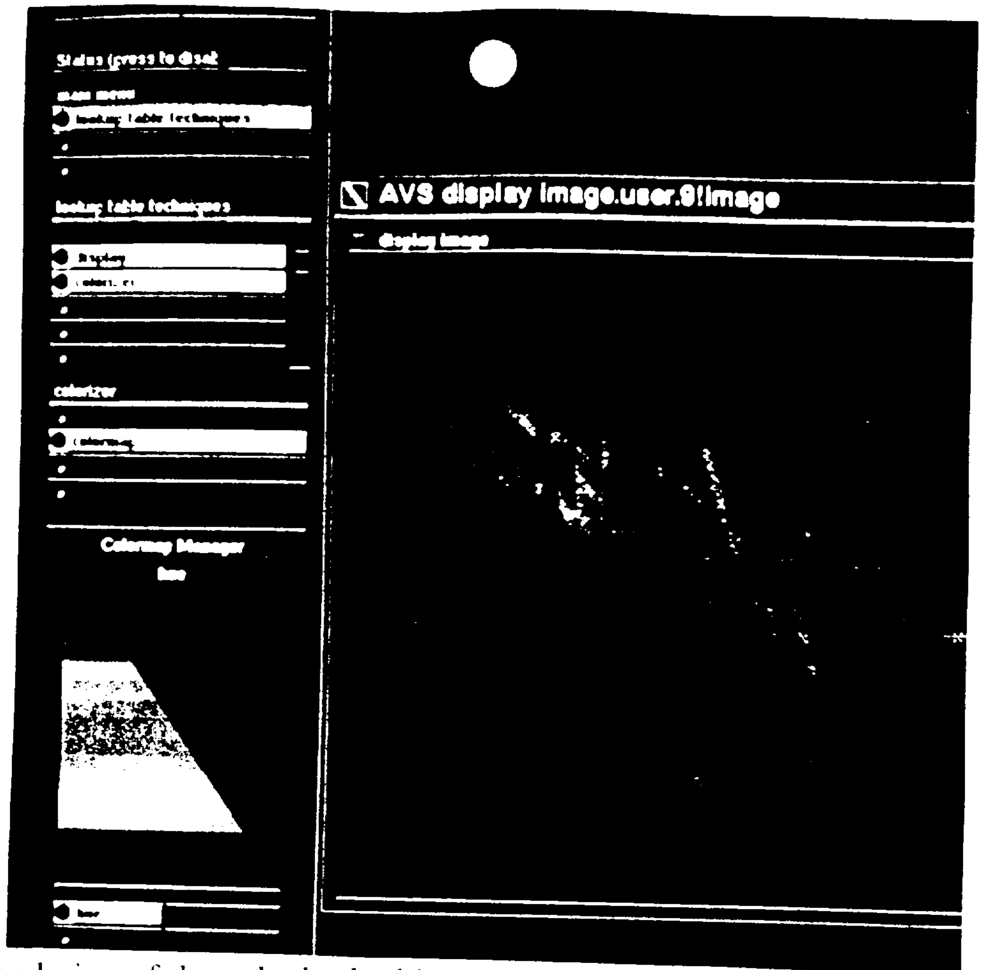
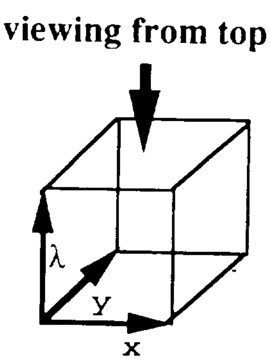


Figure 21: Translucent rendering of the cube by looking at the data from one side: one spatial dimension increases to the right, the frequency increases from bottom up.



dark blue: low molecule count
cyan: medium molecule count
green: high molecule count
yellow: highest molecule count

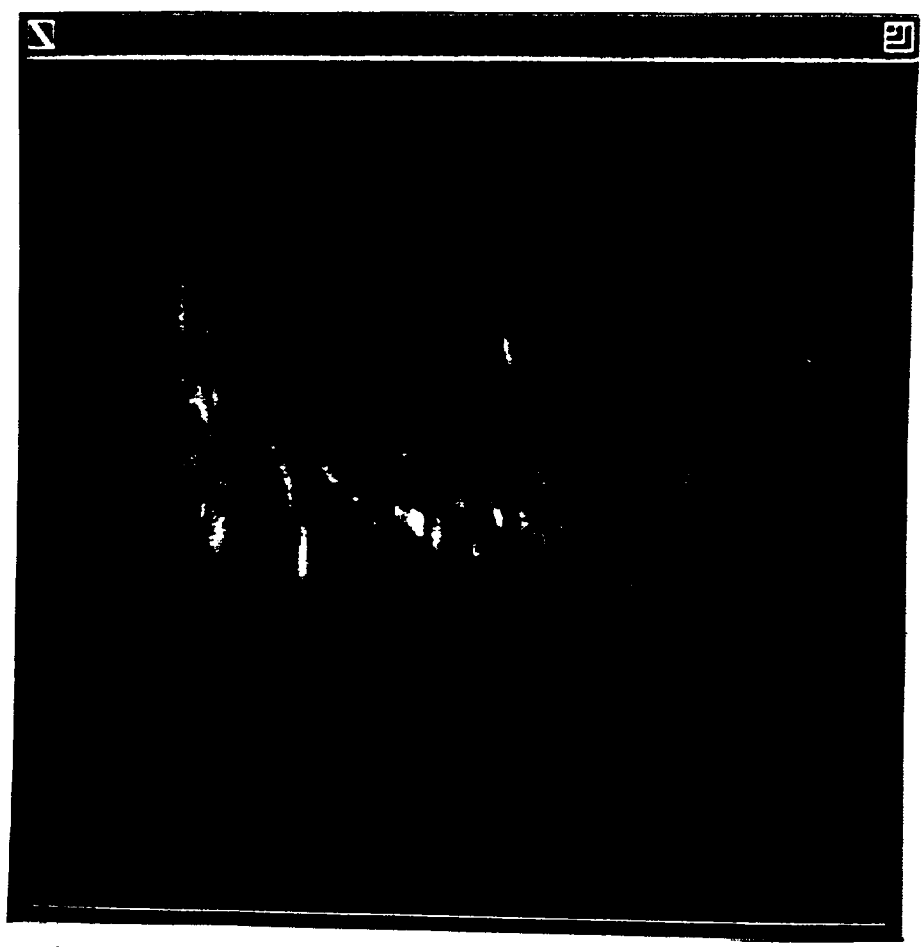


Figure 22: Same data cube as shown in Figure 21 but looking from top down onto the cube.

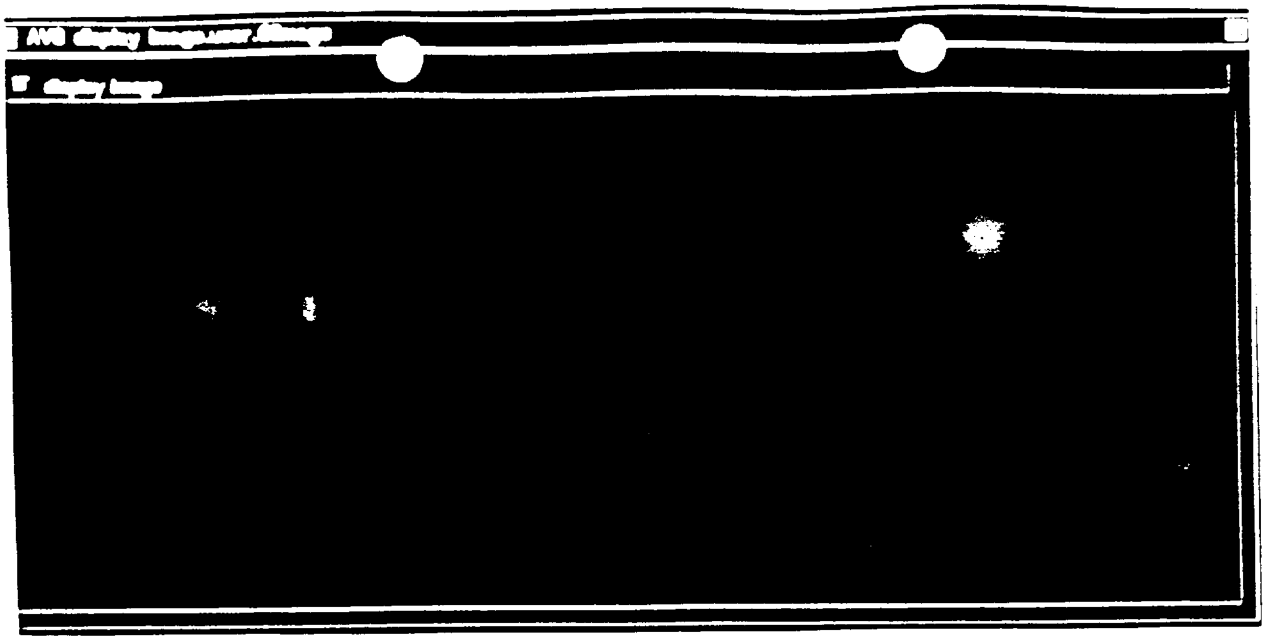


Figure 23: Use of colors relevant to the frequency content: high frequencies are colored blue, low frequencies are colored red. During rotation of the display, the viewer will therefore be constantly aware of the frequency range s/e is looking at.

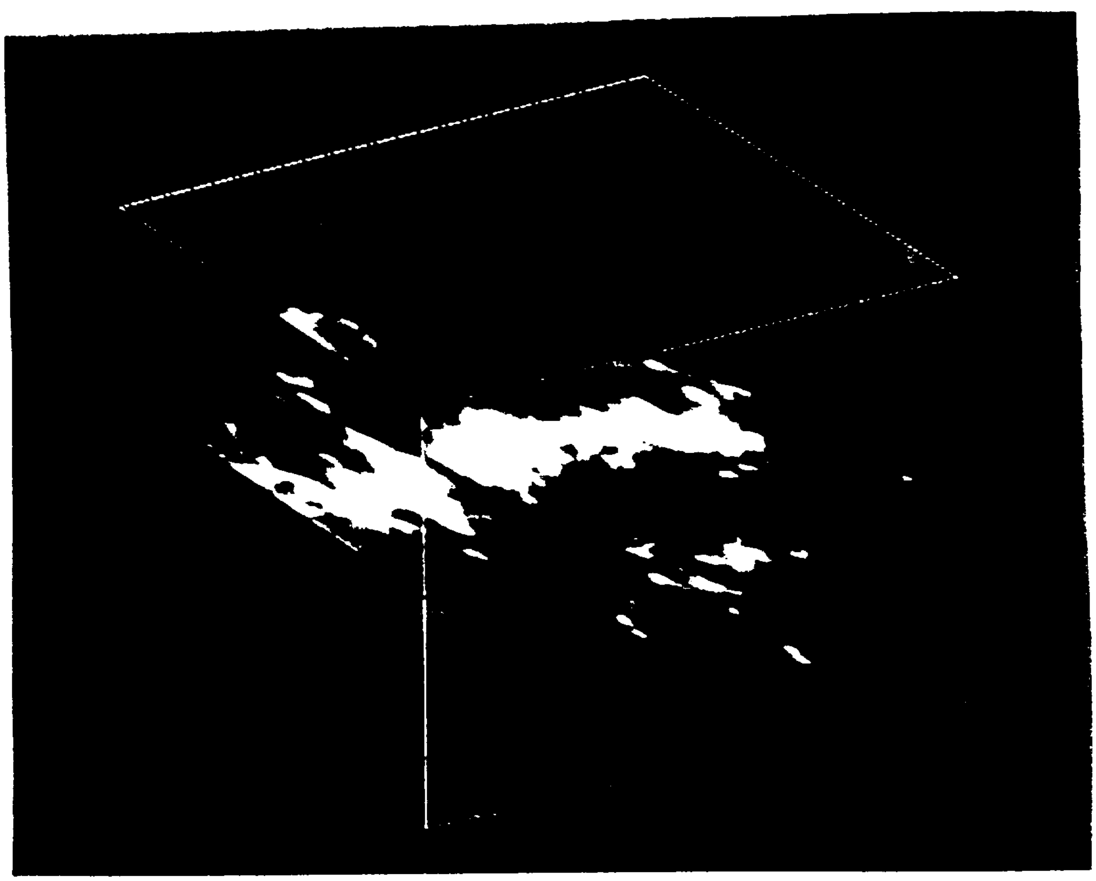


Figure 24: Four slices cutting through the cube parallel to the x/y plane, enhancing the interpretation of the movement of the cloud through frequency.

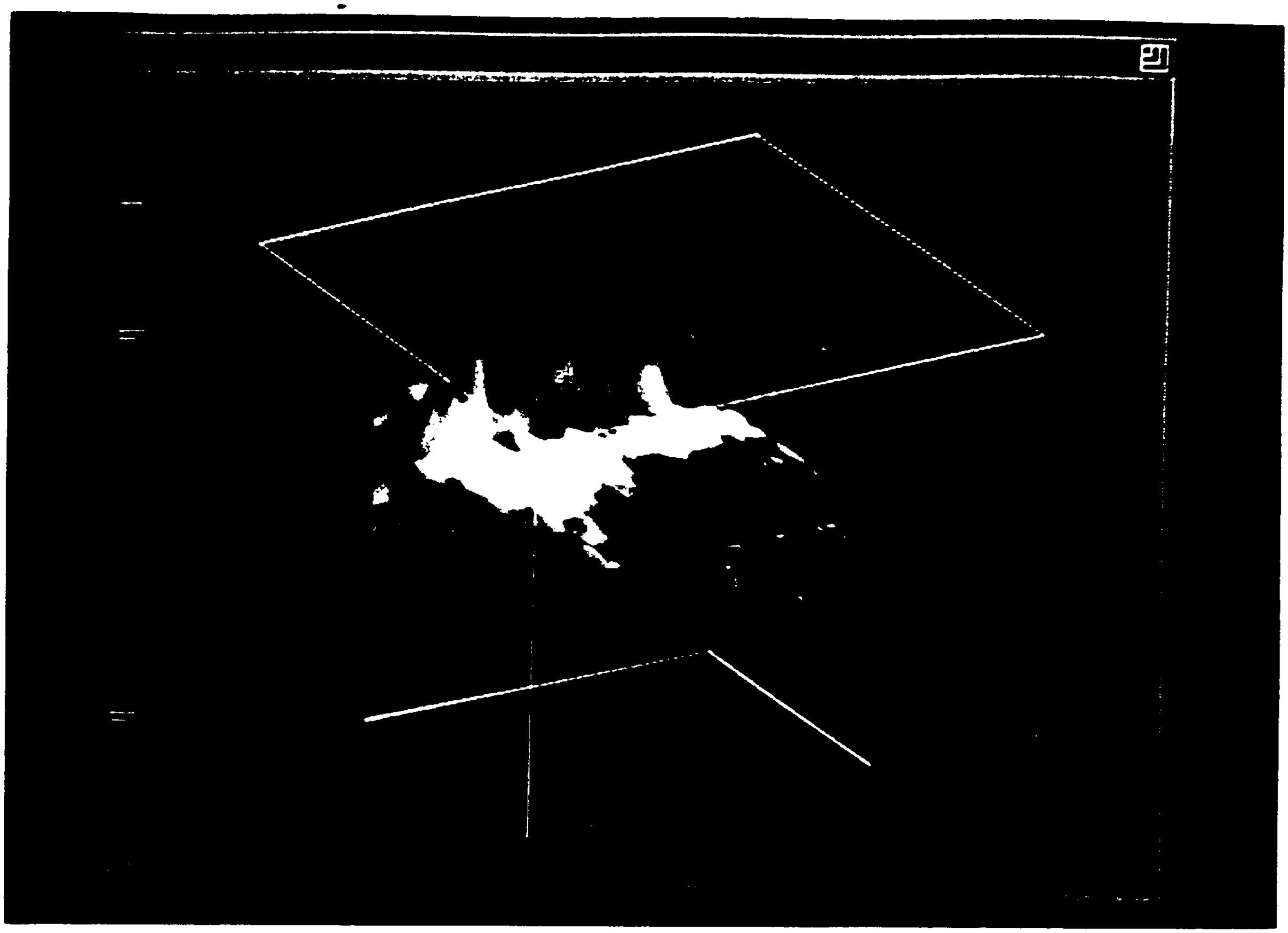


Figure 25: Three orthogonal slices through the data, intersecting in the center of the cube.

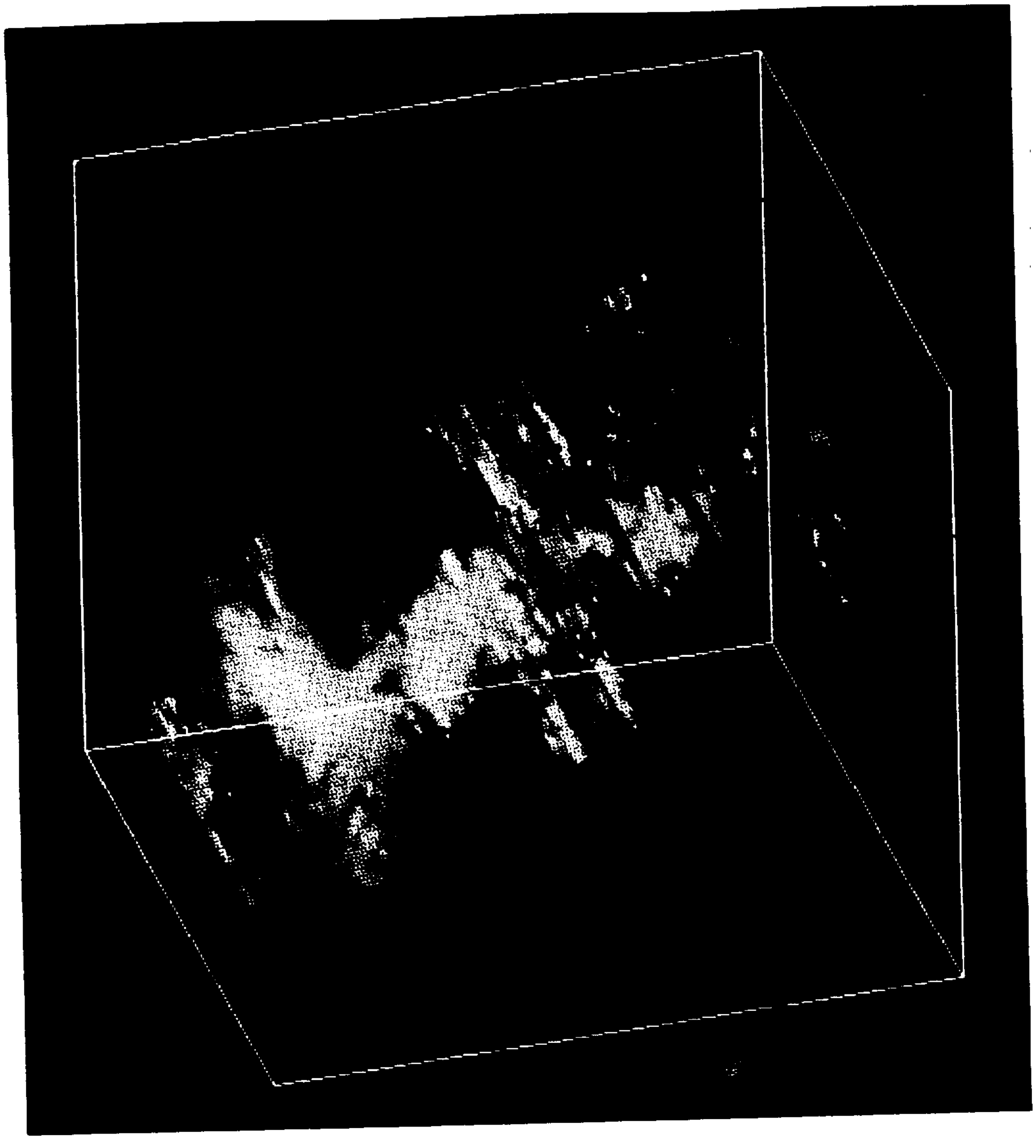


Figure 26: Arbitrary slice through a data cube.

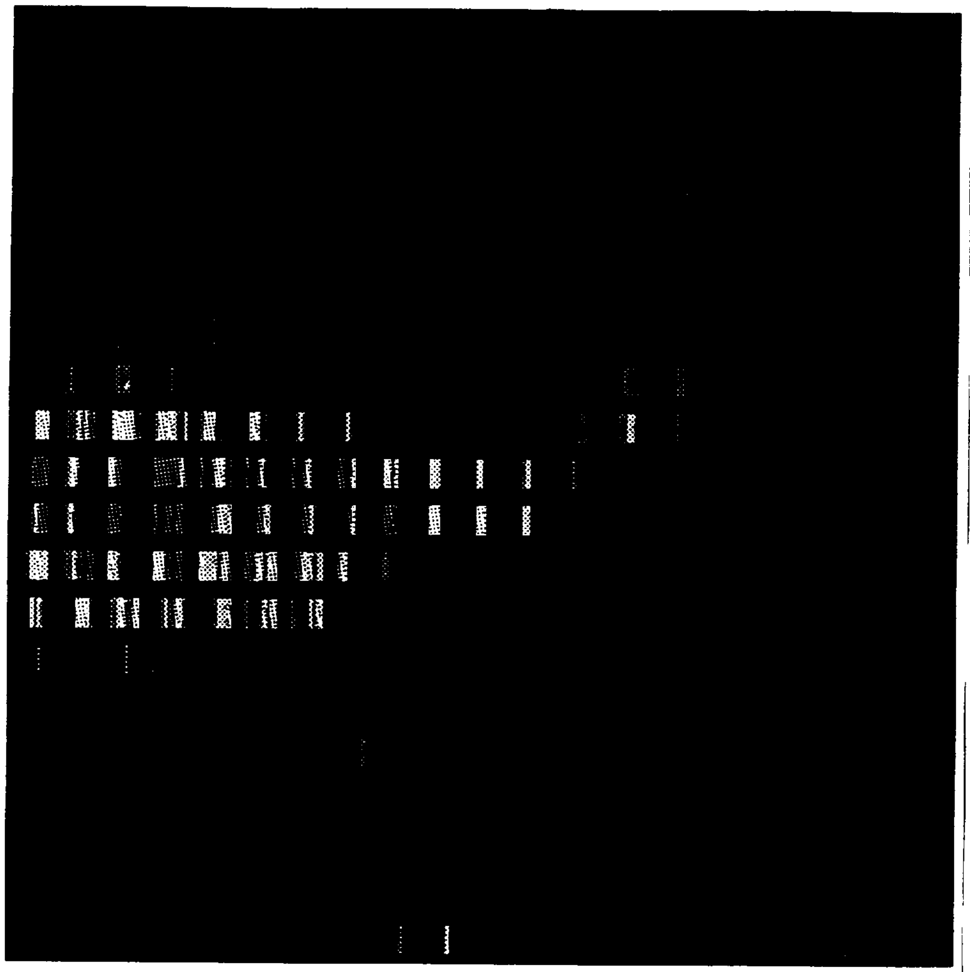


Figure 27: Glyph representation of nine consecutive slices: color slices are used inside each red square to indicate various spectral responses at each spatial location.

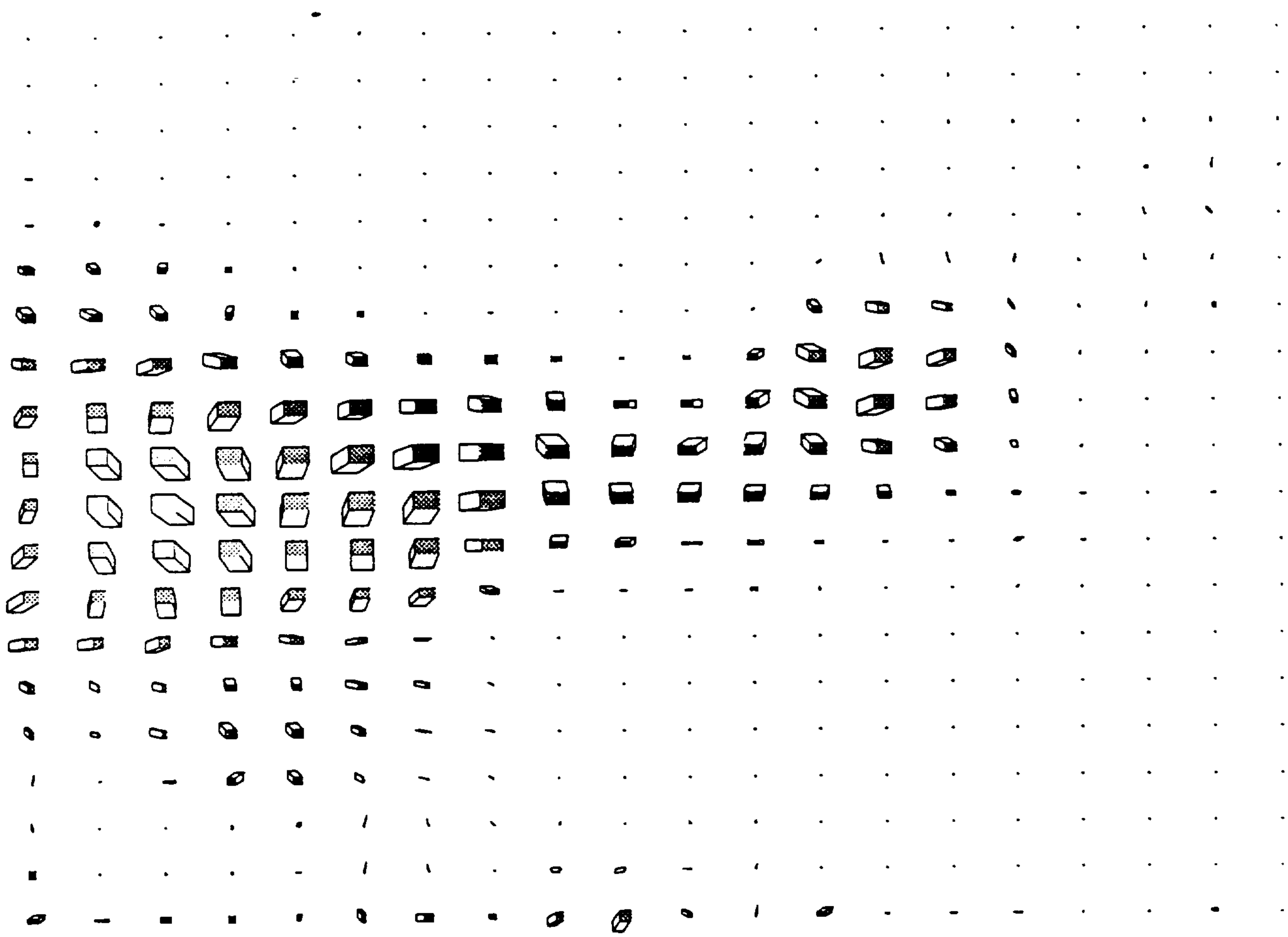
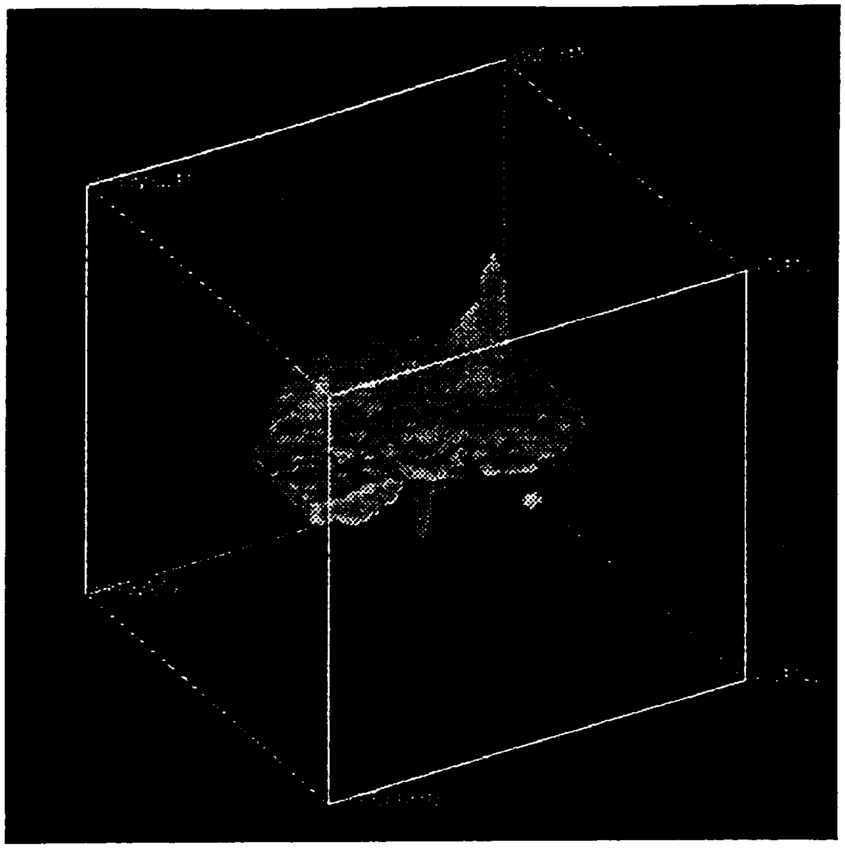


Figure 28: Each five data values are being encoded into five characteristics of a cube: width, height, depth, color and view point.



Objects enclosing values higher than 2.00000

Object 1: 45004.1
Object 2: 2.00966
Object 3: 46.4203

Total number of objects: 3
Total sum for all objects: 45052.5

Figure 29: Three isosurfaces are displayed and a corresponding count of flux values inside these isosurfaces has been calculated.

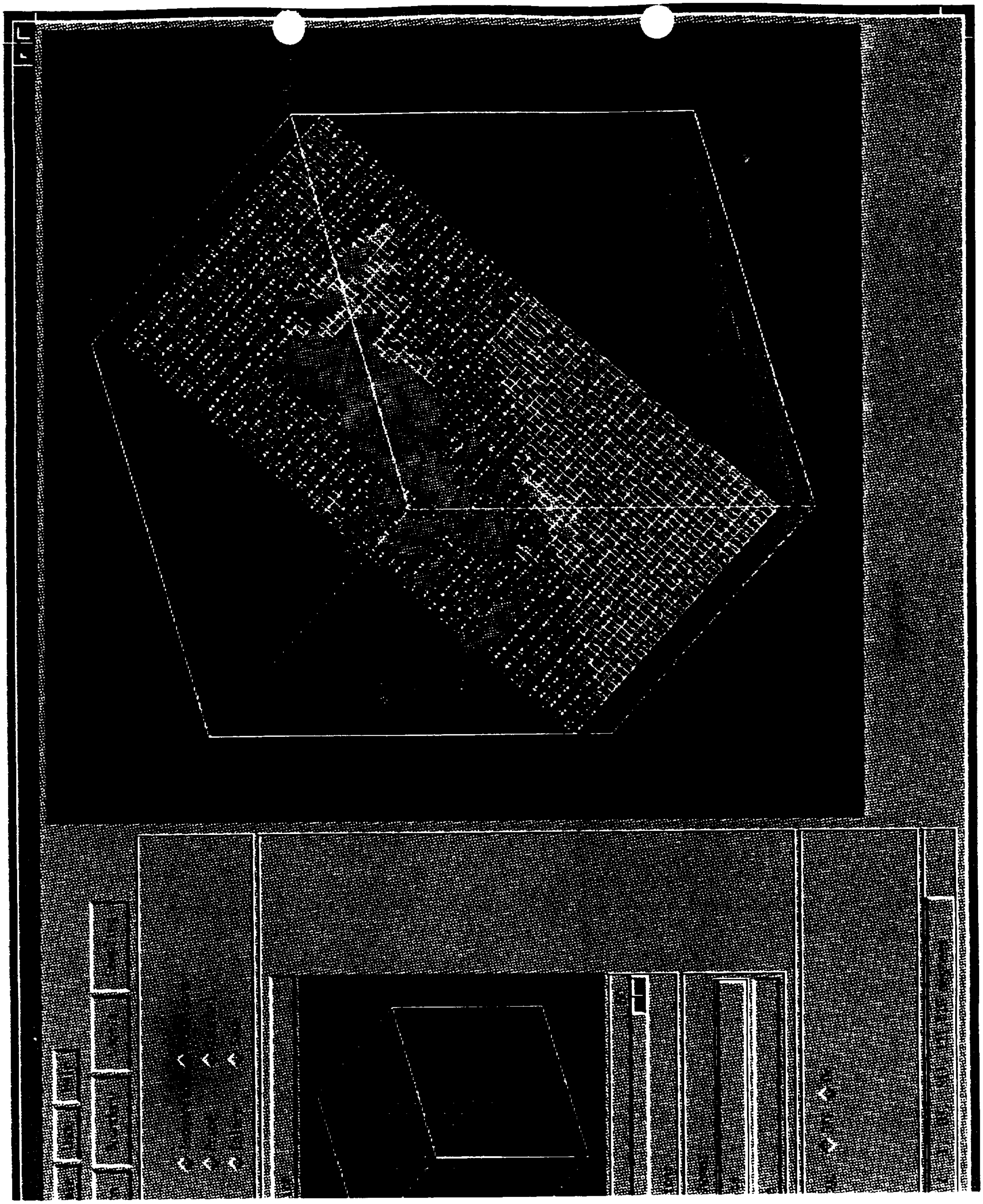


Figure 30: Grid in an arbitrary plane through an isosurface. The grid aids in probing data values inside the data cube in relation to the isosurface.

4. History of Work Progress (1989 to 1993)

The project started in Fall 1989. At that time the computer environment at the Center for Astrophysics and Space Astronomy (CASA) was mainly based on Microvaxes and VAXstations (Vms operating system), even though two DECstations 3100 under the operating system Ultrix had recently been purchased. IDL with the operating system Vms provided the main software development environment at CASA.

At the start of the project major changes in the hardware and software choices of the future could be predicted: scientific computing environments were (slowly) changing from the Vms environment to RISC/Unix workstations. The Unix workstations provided a desirable price-performance ratio, high-resolution and color on their monitors, affordable peripherals, and much room for improvement in each of these areas. While the astrophysical community in general agreed that RISC/Unix was here for the benefit of scientific computing, "porting software" to new platforms was dreaded. It was the lack of funding for adapting software to a different operating system (and different programming languages) that slowed the switch from Vms to Unix. CASA faced the same problems as other scientific centers of the same size: while the body of software was not as extensive as in larger centers, the necessary programming work on top of ongoing software development was a challenge. Current scientific analysis could not be put on hold for six months to a year while the software was adapted to a new platform. Until June '89, when two DECstations 3100 (Ultrix) were acquired, CASA had strictly operated with DEC equipment and under the Vms operating system. Starting 1989, the policy at CASA was set towards a slow move into the Unix operating domain, in order to take advantage of the superior hardware of RISC architecture, while not interrupting the ongoing scientific work at CASA.

At the start of this project therefore, when all in-house developed and most public domain and commercial software used at CASA still operated only under Vms, the choice was made to acquire a VAXstation 3100/ Model 38 (Vms). Development work was to be done under IDL. The VAXstation joined the existing VAXcluster consisting of two Microvaxes, one VAXserver, 10 VAXstations and two image display stations as node "COSMOS". Cosmos was configured to run Vms and DECwindows (DECs flavor of X-windows). The choice of Vms allowed us to work in an alive and active software environment, even though a future move towards Unix was expected during the project. In participating the later porting to Unix, a platform-independent IDL version was used. Even so, slight differences in the IDL/Vms and IDL/Unix codes were expected, e.g. the use of directory names. Code was therefore carefully developed to use features common to both operating systems; indispensable differences were solved by CASE statements specifying a different solution for each operating system.

The first version of STAR (Scientific Toolkit for Astrophysical Research) was developed, tested and used in the Vms environment between 1990 and 1991. During 1991 a new version of IDL was developed by Research Systems, Inc. to allow a more flexible and extensive use of widgets. Therefore during 1992, STAR was rewritten to take advantage of this new IDL version, improving the appearance and functionality of the user interface, and at the same time making the final switch to Unix. Unfortunately, the local data bases made available under IDL/Vms were not compatible with the Unix environment, so that we lost some of our database functionality with the newest versions of our software. Additionally, and through various unrelated causes, we have lost some of our main users during the last two years (1992/1993). This interrupted the growth of STAR and left STAR in its prototyping phase, giving us a platform to experiment and learn from as well as validate our design goals. STAR in its current form can be ftp'd from the anonymous account at "cetus.colorado.edu" (ip address: 128.138.141.22). Path/filenames to ftp from are "pub/star/star.tar.Z and README.FIRST"[8].

Since 1990 various scientists at CASA and outside the University of Colorado made use of STAR in different ways:

a) Use of STAR: During the course of the project we supported specific needs of scientists by developing new visualization tools and an environment that would benefit individual research demands. As an example, Dr. Edward Brugel, Dr. Robert Stencel, and Dr. John Bally at CASA and their students were strongly supported by STAR's broad functionality (see sections 3 and 6). Outside visitors, like Muriel Taylor from Goddard Space Flight Center, came to CASA to test STAR with their own data.

b) Use of STAR's software: Several scientists and graduate students at CASA (e.g. John Saken) as well as centers outside the University of Colorado (e.g. COBE software development group) have adapted the user interface design or analysis software modules into their own environment. We have allowed this way of using our software as we realized that many scientists design, program and maintain their own code and feel flexible in their own environment. Basing STAR on IDL, a much used software platform in astrophysics, has allowed a widespread use of STAR's modules.

c) Design principles of STAR: More than on software development itself we focused on making general software development guidelines available to the astrophysical community: e.g. how to build user interfaces for astrophysical systems; how to decrease the complexity of large software environments; how to improve

8. In case of problems please contact Janet Shaw (jes@qso.colorado.edu).

on available visualization modules. The multitude of resulting publications (see section 6) proves our emphasis on this issue.

5. Recommendations to the Sponsors

In order to underline some of the lessons learned during the project, the author would like to emphasize two inherent propositions made in this report:

- the integration of visualization into the data analysis process; and
- the collaboration between astrophysicists and computer scientists.

Integration of visualization into the data analysis process

When we talk about performing data analysis, we actually mean the execution of a series of processing steps. In this report we have divided data analysis into five main components (see section 2): identification and access of existing information; preprocessing; quantitative analysis; visual representation; user interface. Designing and developing software for any of these components should be done in view of the whole process instead of the individual component. The reason for emphasizing the integration of data analysis components is that astrophysicists need to use *all* aspects of data analysis to answer a scientific question. As long as the software development process of each component is separate from the other components, the user of data analysis software is constantly interrupted in their scientific work in order to convert from one data format to another, or to move from one storage medium to another one, or to switch from one user interface to another. Much frustration is spent this way and time used up that could otherwise be used for scientific work.

Because this report was mainly concerned with integrating visualization software into the data analysis process, a smooth integration of visualization functions into the currently existing analysis environment was stressed. However, visualization of scientific data is meaningless on its own: data has to be documented, calibrated and measured together with their visualization in order to reveal their meaning in respect to a scientific question.

Collaboration between astrophysics and computer sciences

Closer collaboration of astrophysicists and computer scientists will increase productivity and quality in astrophysical research. The competitiveness of research, current funding structures, and historical separation is hampering the collaboration between these groups. A certain indifference of each others needs and contributions leads to commercially available visualization systems that do not cover the needs of astrophysicists (or other scientists); an ignorance of new methods available for multi-dimensional visualizations; or misleading assumptions in introducing new technologies.

Recommendations to support necessary changes include appropriate actions in the research funding structure, the sponsorship of computer scientists to spend time in the environment of astrophysicists, or the sponsorship of computer scientists to develop and teach tutorials on new methods and technologies at NASA's facilities.

Closer collaboration is to the benefit of both parties, and efforts such as the ones instigated through the Center for Excellence of Space Data Information Sciences (CESDIS) are expected to make a difference in the future.

6. Dissemination of Project Activities Outside the University of Colorado

6.1 Publications in journals or books

Domik, G., Brugel, E.W., Stencel, R.E., Vasudevan, S., Pang, J., 1990, *Workstation based Preprocessing of IRAS Skyflux Images*, Publications of the Astronomical Society of the Pacific, October 1990.

Domik, G. O. and Mickus-Miceli, K.D. 1992, *Design and Development of a Data Visualization System in a Workstation Environment*, J. Microcomputer Applications, Vol. 15, pp. 81-88.

Domik, G.O. and Mickus-Miceli, K.D., 1992, *Software Design and Development in a Scientific Environment: Lessons Learned During the Development of STAR, an Astrophysical Analysis and Visualization Package*, Astronomical Society of the Pacific Conference Series, Volume 25, Astronomical Data Analysis Software and Systems I, ed. by D.M. Worrall, Ch. Biemesderfer, and J. Barnes; pp. 95-99.

E.W. Brugel and R. Fesen, *Optical Imaging and Spectroscopy of Three New Northern Hemisphere Herbig-Haro Complexes*, in preparation.

6.2 Publications in proceedings

Domik, G., Vasudevan, S., Pang, J., 1990, Brugel, E.W., Stencel, R.E., *Applications of IRAS Preprocessing at the Workstation*, 176th meeting of the American Astronomical Society, BAAS Vol. 22, No. 2, 1990.

Mickus, K., Domik, G., Brugel, E., Ayres, T. 1990, *STAR-- A Scientific Toolkit for*

Astrophysical Research, 176th meeting of the American Astronomical Society, BAAS Vol. 22, No. 2, 1990.

K. Mickus, E. Brugel, G. Domik, T. Ayres, 1991, *A Case Study: Multi-Sensor Data Analysis of HH Objects via STAR*, BAAS Vol. 22, No. 4, 1991.

G. Domik and K. D. Mickus, 1991, *Visualization in the Analysis Cycle of Observational Data*, Proceedings of the Gesellschaft für Klassifikation, Salzburg, Feb. 25-27, 1991.

E.W. Brugel and R. Fesen, 1991, *Discovery of Three New Herbig-Haro Objects*, BAAS, Vol. 23, No. 2.

E. Overgard, R. Stencel and K. Mickus, 1991, *Workstation Based Analysis of IRAS Views of OB Star Associations*, 103rd ASP meeting, Laramy, Wy..

Domik, G., 1992, *Designing User-Centered Interfaces for Astrophysical Software*, Workshop Proceedings on "User Interfaces for Astrophysical Software", April 14-15, 1992, Goddard Space Flight Center, Greenbelt, MD, sponsored by NASA Headquarters, Astrophysics Division Workshop.

Domik, G. , 1992, *Visualization of Multi-Dimensional Arrays in Astronomy*, Conference on "Astronomy from Large Databases II", Haguenau, France, September 14-16, 1992.

6.3 Reports

Ruiz, J. and Domik, G., 1990, *Use of Color Coding Techniques to Search for Herbig-Haro Objects in the Infrared*, internal REU (Research Experience for Undergraduate Program of the National Science Foundation) report.

K. D. Mickus, 1991, *Participatory User Interface Design for Scientific Visualization Systems*, Masters Thesis, University of Colorado, Department of Computer Science, Boulder, Colorado, 80309-0430.

Domik, G., 1992, *A Case Study in Astrophysical Data Visualization*, Technical report CU-CS-622-92, University of Colorado, Department of Computer Science, Boulder, CO. 80309-0430.

Kahn, Bredekamp, Cheeseman, Domik, Knighton, Treinish, and Weir, 1992, *How Best to Create Interactive Data Analysis Tools That Allow the User to Go from the Pictures to the Numbers*. Special Report to the Communications and Information Systems Division, NASA, March, 1992.

6.4 Participation at workshops relating to project work

User Interfaces for Astrophysical Software, a workshop sponsored by the Astrophysics division of NASA, April 14-15, 1992:

Going from the pictures to the numbers, a workshop sponsored by the Communications and Information Division Systems of NASA, February 1992.

6.5 Talks

A series of about 20 talks were given over the past four years to report on project work and to solicit feedback from scientists about the goals and results of the project. About an equal share of presentations were given in front of scientists (e.g. at the Jet Propulsion Laboratory; at NASA Goddard Space Flight Center; at the Department of Meteorology at the University of Innsbruck, Austria) and computer scientists/engineers (e.g. Department of Informatics, University of Linz, Austria; Storage Tech, Louisville, Colorado; Department of Computer Science, University of Illinois, Urbana-Champaign and National Center for Supercomputing Applications (NCSA)). Consequently a better understanding of the issues surrounding the software environment of data analysis, and specifically visualization software, emerged.

References

Ayres, T., Brugël, E., Domik, G., 1989, *Visualization Techniques to Aid in the Analysis of Multi-Spectral Astrophysical Data Sets,* Proposal to NASA, Astrophysics Division.

Mackinlay, J., 1986, *Automating the Design of Graphical Presentations of Relational Information*, ACM Trans. Graphics, Vol. 5, No. 2, April 1986, pp. 110-141.

McKim, R.H., 1980, *Experiences in Visual Thinking*, PWS Publishers, second edition.

K. D. Mickus,1991, *Participatory User Interface Design for Scientific Visualization Systems*, Masters Thesis, University of Colorado, Department of Computer Science, May 1991.

Robertson, P.K., 1990, *A Methodology for Scientific Data Visualization: Choosing Representations Based on a Natural Scene Paradigm*, Proceedings of the Conference on Visualization '90, October 23-26, San Francisco, pp. 114-123.

Wehrend, S., and Lewis, C., 1990, *A Problem-Oriented Classification of Visualization Techniques*, Proceedings of the Conference on Visualization '90, October 23-26, San Francisco, pp. 139-143.