Chapter 2 Formatting Your Data Files

Chapter Overview

This chapter explains how NCSA Image reads and displays data files. The chapter contains a detailed discussion of the file formats readable by NCSA Image and mentions other software packages that will enhance your use of the program.

It is especially important that you read this chapter if you plan to load raw raster files into NCSA Image in order to ensure that they can be read by the program.

File Formats

NCSA Image reads files containing image data, scientific data, or color palette information. NCSA Image can read in datasets—including user-defined palettes created in applications such as NCSA PalEdit—from raw raster or Hierarchical Data Format (HDF) files.

Raw Raster Files

A raw raster file is a stream of raw, binary, 8-bit raster data in row-major order. Each 8-bit byte corresponds to a pixel in the image. The image is represented in *row-major order*; that is, the first raster line appears first in the file, succeeded by the next raster line, and so forth. Though raw raster files are easy to create, the raw raster file format is not very flexible and therefore, not highly recommended.

If you choose to use raw raster files to store your image data, you must specify the dimensions of the dataset before the file can be read by NCSA Image. The section entitled "Specifying the Dimensions of Raw Raster Files" discusses the procedures for specifying the dimensions of your dataset.

Raw Palette Files

A raw palette file is a stream of 768 bytes. Raw palette files can store 256 colors, which can be selected from a palette of over 16 million possible colors. Palette files are based on the red, green, and blue representation of color, the RGB color model. The files consist of, in order, 256 bytes of red, 256 bytes of green, and 256 bytes of blue. The 256 color palette entries are calculated by combining the *n*th element (red), the (*n*+256)th element (green), and the (*n*+512)th element (blue) to create the *n*th RGB component.

In other words, a palette file is a lookup table with 256 entries that tell which color to associate with each of the 256 possible pixel values. Each of the 256 palette entries in the palette is chosen from a master palette of 2^{24} RGB colors. Each palette entry consists of three bytes, one each for red, green, and blue; the first red component, the first green component, and the first blue component, for example, comprise the first palette entry.

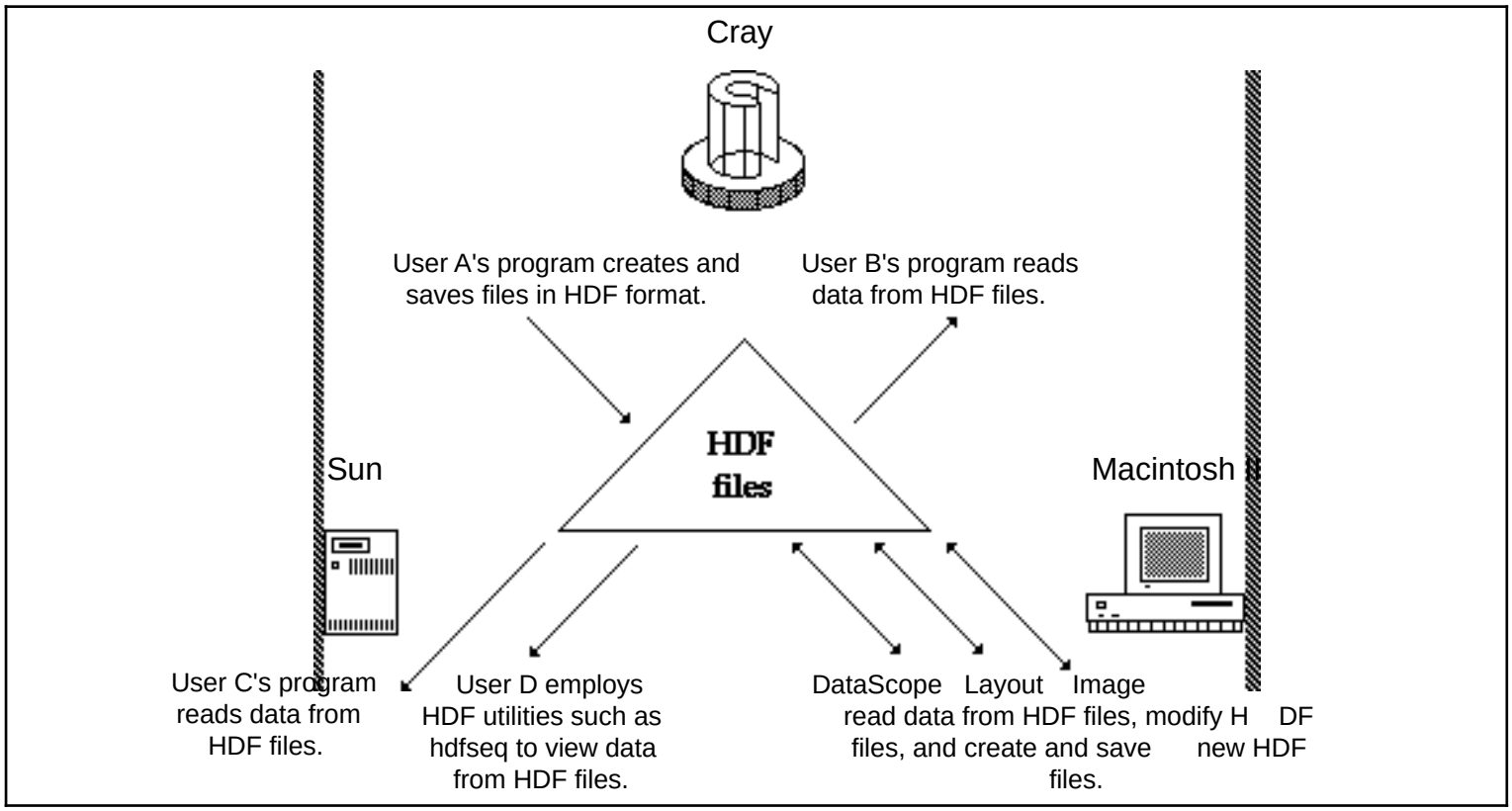
When you load a palette file, the hardware color table entries are remapped according to the new palette; that is, each color in the palette file is assigned to its corresponding entry in the hardware color table.

NOTE: The data values for user-defined palettes must be within the range of 0 to 255; however, the Macintosh reserves the palette values 0 and 255 for white and black, respectively. Consequently, you may have to scale your data to the range from 1 to 254, for NCSA Image overrides other assignments to 0 and 255 with white and black, respectively.

HDF Files

Hierarchical Data Format, or HDF, is a flexible, standard file format designed at NCSA for sharing of graphical and floating-point data among different programs and machines. This general purpose and extensible file format allows you to store raster images, their dimensions, color tables, and annotations all in the same file. In addition, you may store floating-point data, scaling information, dimensions, annotations, and units of measurement in a single file. HDF files are accessible from NCSA software for the Macintosh such as NCSA Image, NCSA DataScope, NCSA PalEdit, and NCSA Layout as well as user programs, other NCSA software, and HDF utilities. These files can be used used on such machines as the CRAY X-MP/48, CRAY-2, Sun Workstation, IBM PC, and Alliant. The portability and usefulness of HDF files is demonstrated in Figure 2.1.

Figure 2.1 HDF Environment



Why Should I Use HDF?

HDF lets you store datasets with extra file information about your data, for example, the dimensions of your image. This makes the files easier to read and manage by programs such as NCSA Image, and saves you the trouble of tracking this information externally.

What Information Goes into an HDF File?

Both raster images and scientific floating-point data can be stored in HDF files for use in NCSA Image.

A *raster image set* contains a raster image, together with its dimensions. It may also contain a palette. You may specify that the raster image be stored in compressed or non-compressed form.

NOTE: Images saved by NCSA Image will be saved in HDF files in compressed form unless you specify otherwise under the Preferences menu. The Preferences menu is described in the section entitled "Preferences."

A *scientific dataset* can store scientific data in an array of 32-bit floating-point numbers of any dimension, together with information regarding the rank and size of each dimension.

When storing your datasets in HDF files, you have the option of storing other information such as labels for the data and the axes, scales for the axes, and the maximum and minimum values of the data.

How Do I Create an HDF File?

Public domain software is available from NCSA for creating HDF files. NCSA's HDF libraries support both Fortran and C calls on any of the following machines: Cray (UNICOS), Sun (UNIX), Alliant (Concentrix), Macintosh, and IBM PC (MS-DOS).

The best way to store your data in an HDF file is to incorporate calls to the appropriate HDF library in the program that produces your image or scientific data. These calls can store your raw image, palette, scientific data, and other information in an HDF file in proper format.

If you have access to the HDF libraries, you can store floating-point data in your file, and use it directly. NCSA Image uses floating-point data rather than 8-bit binary data whenever the former is available.

NOTE: If you have a UNIX-based system, you can use the command line utility called r8tohdf to convert one or more raw raster images and palettes to HDF format.

Where Can I Obtain More Information about HDF?

If you are connected to Internet (NSFNET, ARPANET, MILNET, etc.) you can download HDF software and documentation at no charge from an anonymous file transfer protocol (FTP) server at NCSA. The steps you should follow to do so are enumerated below. If you have any questions regarding the connection or procedure, consult your local system administrator or network expert.

1. Log on to a host at your site that is connected to Internet and is running software supporting the FTP command.
2. Invoke FTP on most systems by entering the Internet address of the server:

   ```
   % ftp ftp.ncsa.uiuc.edu
   ```

   or

   ```
   % ftp 128.174.20.50
   ```

3. Log in by entering **anonymous** for the name.
4. Enter your local login name for the password.
5. Enter **get README.FIRST** to transfer the instructions file (ASCII) to your local host.

6. Enter **quit** to exit FTP and return to your local host.

7. Review the README.FIRST file for complete instructions concerning the organization of the FTP directories and the procedures you should follow to download the README files specific to the application you want.

Your login session should resemble the following sample, where the remote user's local login name is *smith* and user entries are indicated in boldface type.

```
harriet_51% ftp ftp.ncsa.uiuc.edu
Connected to zaphod.
220 zaphod FTP server (Version 4.173 Tue Jan 31 08:29:00 CST 1989) ready.
Name (ftp.ncsa.uiuc.edu: smith):  anonymous
331 Guest login ok, send ident as password.
Password:  smith
230 Guest login ok, access restrictions apply.
ftp>  get README.FIRST
200 PORT command successful.
150 Opening ASCII mode data connection for README.FIRST (10283 bytes).
226 Transfer complete.
local:  README.FIRST  remote:  README.FIRST
11066 bytes received in .34 seconds (32 Kbytes/s)
ftp>  quit
221 Goodbye.
harriet_52%
```

The README.FIRST file instructs you to copy the HDF README file to your directory and read it before proceeding. Your FTP session should resemble the one listed below:

```
ftp>  cd HDF
250 CWD command successful.
ftp>  get README
200 PORT command successful.
150 Opening ASCII mode data connection for README (10283 bytes)
226 Transfer complete.
local:  README      remote:  README
2080 bytes received in .14 seconds (15 Kbytes/s)
ftp>  quit
221 Goodbye.
harriet_52%
```

The HDF README file explains how to copy the contents of the HDF directory to your home directory via remote login or anonymous ftp. The precise file transfer procedure varies according to the type of operating system under which you will use HDF—UNICOS or other.

HDF software and manuals are available for purchase—either individually or as part of the anonymous FTP reel or cartridge tapes—through the NCSA *Technical Resources Catalog*. Orders can only be processed if accompanied by a check in U.S. dollars made out to the University of Illinois. To obtain a catalog, contact:

NCSA Documentation Orders
152 Computing Applications Building
605 East Springfield Avenue
Champaign, IL 61820
(217) 244-0072

How NCSA Image Reads and Displays Data Files

NCSA Image assumes that all data files are to be initially displayed as two-dimensional data, in 8-bit format, arranged in row-major order, with the origin in the upper-left corner.

Eight-bit format refers to a process whereby data is scaled onto the numerical values from 0 through 255 and is stored in single bytes, one data element per byte. When data is displayed on the screen, a byte is interpreted as a number from binary 0 through binary 255. The number represents a color from the current palette. For example, a byte that is equivalent to binary 8 is interpreted as the ninth color in the current color palette and is displayed accordingly.

NCSA Image places the first element of the dataset in the upper-left corner of the image's display window and assumes that the data is in row-major order. If the original data was created in column-major order, as in a Fortran program, it must be transposed to be read by the program. You can transpose your dataset using the Preferences option under the Edit menu. Instructions regarding this procedure are presented in the section entitled "Preferences."

When you issue the Open command and select a file to load into NCSA Image, the application first ascertains whether the file is an HDF file. If so, and if the file contains floating-point data, then NCSA Image loads in the floating-point data, the scale for the x and y direction, and the dimensions of the dataset, when these are stored in the file. If the file contains a raster image, then the raster image data, its dimensions, and a color palette, if present, are loaded in. See the section entitled "Preferences" for information regarding how to specify that a scientific dataset, when present, be loaded instead of the raster image set.

If the file is not identified as an HDF file, the program assumes that it is a raw raster file and that the dimensions of the dataset need to be determined. The procedures you should follow to ensure that the dimensions for your raw raster file are correctly ascertained are presented in the following section.

After a floating-point dataset and its scales are loaded, NCSA Image performs a bi-linear interpolation of the dataset to generate the image. A *progress box* appears on the screen during the interpolation to indicate the amount of calculation remaining.

Several options that are available to determine the manner in which the image is interpolated are described in the section entitled "Preferences."

NOTE: An image generated from floating-point data behaves differently than a raw raster image in some operations. For example, in scope mode (described in Chapter 4, "Imaging Options for Data Analysis"), a floating-point number is displayed in the text box rather than an integer. Future versions of the program will provide more extensive functionality designed for use with floating-point datasets.

Specifying the Dimensions of Raw Raster Files

To read and display an image, NCSA Image must be able to ascertain its dimensions. If you use an image stored in an HDF file, which can contain an image along with information about the image dimensions and associated color table, NCSA Image can read the dimensions from the HDF file. If the program finds the dimensions for your image in the HDF file, the image is loaded accordingly. If not, NCSA Image produces a dialog box that prompts you to specify the dimensions for your image data (see Figure 2.2).

If you use an image stored in a raw raster file, on the other hand, you must provide the program with the dimensions for the image, according to one of the following methods.

1. Save the dimensions as part of the filename. Append the appropriate values to the filename, using the following convention:

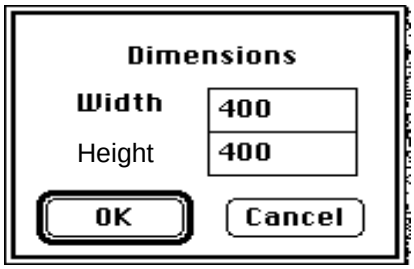
   (*horizontal size*vertical size*)

   where * is any non-numeric character. For example, to save a gas jet image that has x and y dimensions of 400 and 365, respectively, you could name the file *gasjet(400*365)* to ensure that NCSA Image can read and display the image properly.

2. Remember the dimensions and enter them when the Dimensions dialog box appears and prompts you to do so.

If the dimensions are found enclosed in parentheses in the raw raster filename, the dialog box shown in Figure 2.2 appears and reflects the appropriate horizontal and vertical values by default. If the dimensions cannot be found, the dimensions of the frontmost image window, if there is one, are shown in the dialog box as the default; otherwise, you must enter the dimensions in the text boxes labeled *Width* and *Height*. In either case, you can change or correct the dimensions that appear as the default in the dialog box.

Figure 2.2 **Dimensions Dialog Box**



NOTE: You may load part of the raster image by specifying a vertical dimension smaller than the true size of the dataset. Then only the number of rows of data specified is loaded. If you specify dimensions that are too large, an error message appears and the operation is aborted. A complete list of error messages is contained in Appendix A, "Troubleshooting."

Preferences

Under the Edit menu, NCSA Image provides a Preferences command that allows you to control certain operations of the application. When you select Preferences, a dialog box for the variable NCSA Image operations appears on the screen. This dialog box is shown in Figure 2.3. You can specify your preferences regarding these options by clicking their respective boxes on or off (checked or unchecked). Most of these options apply to loading and saving operations. These variables are described in the order in which they are presented in the dialog box.

Figure 2.3 **Preferences Dialog Box**

- **Generate image to scale**
  When generating an image from your dataset, NCSA Image allocates 400 pixels to represent the largest dimension of your dataset, x or y. In other words, the largest dimension of the generated image is always 400 pixels. The other dimension is calculated in relation to this value and the scales used.

If you enable the Generate image to scale option, NCSA Image makes the size of the image generated from your dataset proportional to the size of the true overall width and height of the entity that the dataset represents; that is, it makes the size proportional to the distances between the first and last x and y components of the floating-point matrix.

Assume, for example, that your dataset represents a piece of metal 35.7 millimeters wide and 20 millimeters high, and that you have selected Generate image to scale. The width of the image generated from this dataset is 400 pixels, because the width is the larger of the two dimensions. The height is calculated according to the following formula, where s is the smaller of the two dimensions and l is the larger.

s=(400x20/l)

In this example, 35.7 is supplied for l; and s, when calculated and truncated, translates to 244.

If you do not enable the Generate image to scale preference, NCSA Image sizes the image by making the horizontal and vertical dimensions proportional to the number of columns and rows in the floating-point dataset. That is, the number of columns and number of rows are substituted for the width and height, accordingly, in the above example and formula. Whichever is more numerous, columns or rows, is represented with 400 pixels; the remaining is calculated according to the formula above. Here, s is the less frequent of the two dimensions.

- **Transpose generated images**
  Floating-point data can be stored in either row- or column-major order in an HDF file. NCSA Image assumes that data is stored in row-major order unless the Transpose generated images option is checked, in which case, the program reads the data in column-major order. Consequently, if your data is stored in column-major order, as in a Fortran program, you must check the option Transpose generated images to ensure that your dataset is read and displayed properly.

- **Darker shades for lower magnitudes**
  NCSA Image typically draws dark shades to represent high magnitudes in the shaded and dither plots. You can reverse the representation of magnitude by selecting Darker shades for lower magnitudes.

- **Save images uncompressed**
  Images can be saved to an HDF file in a non-destructive compressed format or an uncompressed format. The program saves files in compressed HDF format by default. To save your files in uncompressed format, check Save images uncompressed.

- **Load SDS before RIS**
  Normally, when you load an image from an HDF file, NCSA Image searches first for a raster image set (RIS). If an RIS is present in the file, NCSA Image automatically loads it; otherwise, NCSA Image searches for a scientific dataset (SDS) to load. Check the option Load SDS before RIS to specify that this process be conducted in reverse order; that is, that the program first load an SDS when such is present, and then an RIS if an SDS is not found.

When you click OK, the preferences you checked are saved and remain the default until you change them again. In other words, you may quit NCSA Image and restart it, and the preferences you selected in the previous session still apply.