Chapter 3 Formatting Your Data Files

Chapter Overview

File Formats

Raw Raster Files

Raw Palette Files

PICT Files

ASCII Files

HDF Files

How NCSA Layout Reads and Displays Data Files

Specifying the Dimensions of Raw Files

Chapter Overview

This chapter explains how NCSA Layout reads and displays data files, and discusses the procedures you should follow to ensure that your files can be read by the program.

File Formats

NCSA Layout reads files containing image data or color palette information. Layout can read datasets from raw raster, raw palette, PICT, ASCII, 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 order. Each 8-bit byte corresponds to a pixel in the image. The image is represented in row order, with the first raster line appearing first in the file, succeeded by the next raster line, and so forth. Though raw raster files are easy to create, 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 Layout. The section entitled “Specifying the Dimensions of Raw 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 R*GB* 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. For example, the first red component, the first green component, and the first blue component 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. The Macintosh reserves the palette values 0 and 255 for white and black, respectively. Consequently, you may have to scale your image data to the range from 1 to 254, because NCSA Layout overrides other color assignments to 0 and 255.

PICT Files

A *PICT file* is a file format that is native to Macintosh computers.NCSA Layout offers two ways to save PICT files: either as drawing commands [PICT (drawing) from the Save As...option under the File menu], which allows files to be loaded into other Macintosh applications for further editing; or in pixels [PICT (pixels) from the Save As...option under the File menu], so they can be edited by programs such as SuperPaint or Pixel Paint.

ASCII Files

ASCII files, commonly called text files, are comprised of ASCII characters. To create ASCII image data, scale the data to non-negative integers between 0 and 255, inclusive, and separate each integer with white space.

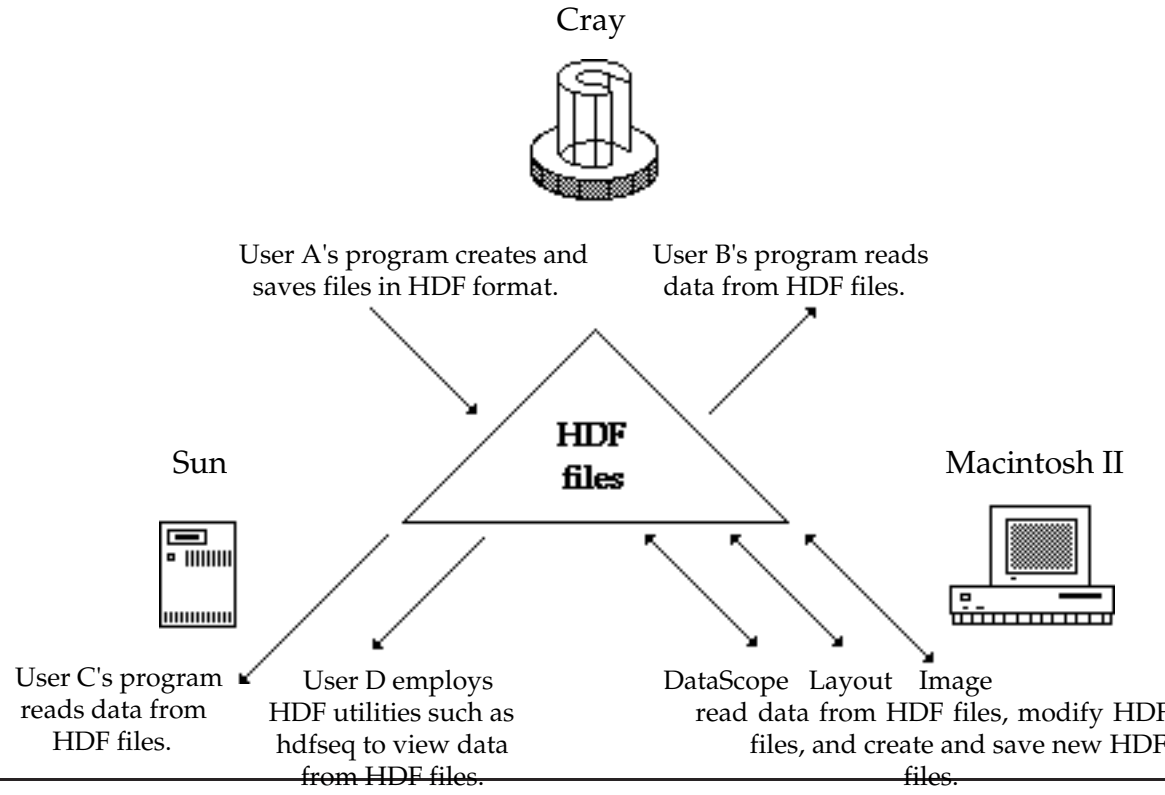
In NCSA Layout, each ASCII character is interpreted as a number which corresponds to one of the 256 palette entries. When NCSA Layout reads an ASCII file that contains integer values other than 0 through 255, it sets them to 0 and displays a message box to notify you.

If your image is stored in ASCII format, you must specify the dimensions of the dataset before the file can be read by NCSA Layout. The section entitled “Specifying the Dimensions of Raw Files” discusses the procedures for specifying the dimensions of your dataset. Also, you must inform Layout that you are loading an ASCII file as instructed in the section entitled “Loading Images” in Chapter 3.

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 Layout, NCSA DataScope, NCSA Image, and NCSA PalEdit as well as user programs, other NCSA software, and HDF utilities. They can be 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 3.1.

Figure 3.1 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 for programs such as NCSA Layout, 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.

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.

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.

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 following:

```
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 Layout Reads and Displays Data Files

NCSA Layout assumes that all data files are to be displayed as two-dimensional data, in 8-bit format, arranged in row-major order, with the origin in the upper-left corner. If the original data was created in column-major order, as in a Fortran program, it must be transposed prior to being read by NCSA Layout.

The term *8-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 eight entries from the first color in the current color palette and is displayed accordingly.

NCSA Layout displays images that have been saved in raw raster, PICT, ASCII, or HDF files. To load images from raw files, which do not store information regarding the dimensions of the data, you must specify the dimensions of the data.

In addition, to load a non-HDF ASCII file, you must first inform NCSA Layout that the file you are loading is an ASCII file, as instructed in the section “Loading Images” in Chapter 4.

NOTE: You can scale HDF floating-point data to 8-bit data using NCSA Image.

Specifying the Dimensions of Raw Files

There are two ways to specify the dimensions for non-HDF ASCII or raw raster images:

- save the dimensions as part of the filename
- enter the dimensions of the image when you load it

To save the dimensions as part of the filename, affix 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 shockwave image that has x and y dimensions of 400 and 365, respectively, you could name the file *shockwave(400*365)* to ensure that Layout can read and

display the image properly.

If NCSA Layout does not find the dimensions for your image enclosed within parentheses in the raster filename, the dialog box shown in Figure 3.2 appears on the screen and prompts you to specify the appropriate horizontal and vertical values.

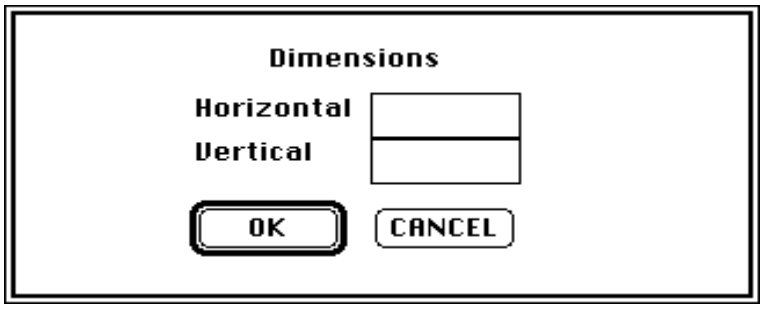


Figure 3.2 Specifying Dimensions for Raw Raster Image Data