Chapter 2 Formatting Your Data Files

Chapter Overview

This chapter explains how NCSA CompositeTool reads and displays data files. The chapter contains a detailed discussion of the file formats readable by NCSA CompositeTool 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 CompositeTool in order to ensure that they can be read by the program.

File Formats

NCSA CompositeTool reads files containing image data and/or color palette information. NCSA CompositeTool 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 CompositeTool. 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.

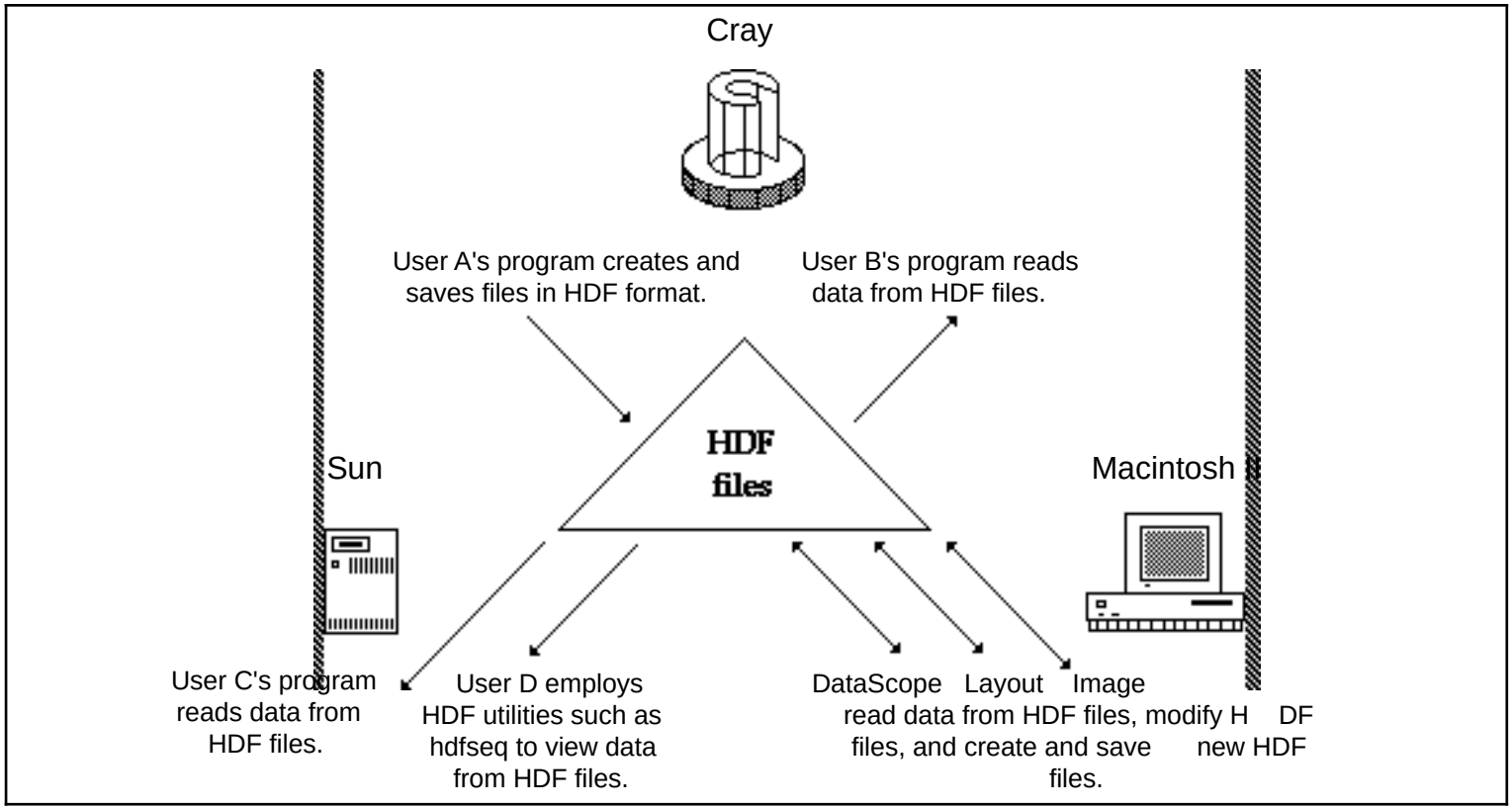
When you load a palette file, the hardware color table entries are remapped according to the new palette; that is, each color in the palette file is assigned to its corresponding entry in the hardware color table.

NOTE: The data values for user-defined palettes must be within the range of 0 to 255; however, CompositeTool currently reserves the palette values of 0 and 253–255. Consequently, you must scale your data in the range of 1–252 in order to view all your data.

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 ImageIP, NCSA Image, NCSA GelReader, NCSA DataScope, NCSA PalEdit, and NCSA Layout as well as user programs, other NCSA software, and HDF utilities. These files 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 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 CompositeTool and saves you the trouble of tracking this information externally.

What Information Goes into an HDF File?

Raster images can be stored in HDF files for use in NCSA CompositeTool.

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.

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 data. These calls can store your raw image and palette in an HDF file in proper format.

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 CompositeTool Reads and Displays Data Files

NCSA CompositeTool 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. 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.

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 the ninth color in the current color palette and is displayed accordingly.

When NCSA CompositeTool is to use a file, the application first ascertains whether the file is an HDF file. If it is, and the file contains a raster image, the image is loaded using the dimensions stored in the file. The palette must be loaded as a separate step. (See Chapter 3, the section, "Palette Frame" for more information.)

If the file is not an HDF file, the program assumes that it is a raw raster file and that the dimensions of the dataset have been typed in by the user in the appropriate field.