The Berkeley-San Francisco Fine Arts Image Database

Nisha Talagala, Satoshi Asami, David Patterson
{nisha,asami,patterson}@cs.berkeley.edu
Computer Science Division
University of California at Berkeley, Berkeley, CA 94720
Tel: +1 510 642-1845

Bob Futernick, Dakin Hart
{bfuternick, dakin}@famsf.org
Fine Arts Museums of San Francisco
233 Post Street, San Francisco, CA 94108
Tel: +1 510 415 750-3508

Abstract

This paper describes the world's largest on-line art collection, provided by a collaboration between the Fine Arts Museums of San Francisco and the University of California at Berkeley. The collection contains over 70,000 images of artworks and is accessible over the web. Each image will be available at a resolution of upto 3,000 by 2,000 pixels. This paper describes the storage system and the image database that make up the web site.

1. Introduction

This paper describes a collaboration between the University of California at Berkeley and the Fine Arts Museums of San Francisco, to provide a large collection of high-resolution images over the web. Over the past few years, the Fine Arts Museums have photographed over 70,000 art objects and stored the images on CD-ROMs in PhotoCD format. Each image is 3,000 x 2,000 pixels and about 4.5 MB in size. Because of limited storage, the museum offered only low-resolution versions of the images online: 2KB thumbnails and 20-50KB JPEGs. The larger versions of the images were kept only on CD-ROMs.

Visitors to the web site can search the image database by artist, title, or description of an artwork. The image database is the largest on-line art collection in the world. Since the web site's launch in November 1996, many users have requested larger versions of the images. To satisfy this demand, the museum is collaborating with UC Berkeley to create an http server for the larger images.

The high-resolution images are stored on the Tertiary Disk prototype [2], a three terabyte disk storage system developed at UC Berkeley. The name of the prototype comes from twin goals, to achieve the cost/megabyte of tape systems and the performance of disk arrays. The Tertiary Disk prototype is a large distributed storage system created from commodity components.

The following sections describe the storage system and the image server implementation. Section 2 describes the storage system design. Section 3 describes the image database and http servers. Section 4 gives the current status and future work.

2. The Storage System

Over the past five years, disk drive prices have been falling at nearly a factor of two per year [1]. This rapid decrease makes large scale disk systems a feasible alternative to tape libraries. However, commercially available disk storage systems, like hardware RAID arrays, have a much higher cost per megabyte than the underlying disks. Finally, new disks for RAID arrays usually have to be obtained from the manufacturer, to maintain the warranty. This practice increases the cost of expansion and the lag time to incorporate new disks.

The performance of a disk array is limited by its host connection(s). In most cases the bandwidth of the host connection does not match the bandwidth of the disks. Disk bandwidths are increasing dramatically due to higher rotation speeds, higher areal densities, and better electronics [1]. If these trends continue, the host interconnect will become an increasing performance bottleneck for RAID arrays.

Tertiary Disk is a storage system design that avoids these problems. The prototype is a three terabyte disk system with a cost/megabyte close to that of the underlying disks. The basic idea is to use commodity hardware to lower cost and software to manage and monitor the system. The design is based on a group of relatively independent storage nodes. Independent nodes make the system more scalable and allow incremental expansion. Also, this design increases the number of links from the storage system to the outside world, improving bandwidth. The prototype proves by example that a large storage system can be built using off-the-shelf hardware.

The design uses PCs connected through a switched network, with each PC hosting a group of disks. The prototype contains 20 PCs and 368, eight gigabyte disks. The PCs are interconnected through 10 Mbit/s switched Ethernet. (We will soon upgrade to a 100 Mbit/s network). The disks are connected to the host machines using Fast-Wide SCSI. To improve availability, each disk is connected to two PCs. Each SCSI bus is shared, with two SCSI controllers on the bus. If one host in unable to service requests, all the disks can be accessed from the second host. In normal mode, each host accesses half the disks (i.e., there is no disk sharing). More details about this design are available in [2].

The 20 hosts are organized into 10 nodes of 2 machines each. There are two types of nodes, *light* and *heavy*. Each light and heavy node contains 32 and 56 disks, respectively. Figure 1 shows a light node. The light and heavy nodes were used to study the cost/performance trade-offs in changing the number of disks per host. A performance study of the nodes is also available in [2]. The prototype contains eight light nodes and two heavy nodes.

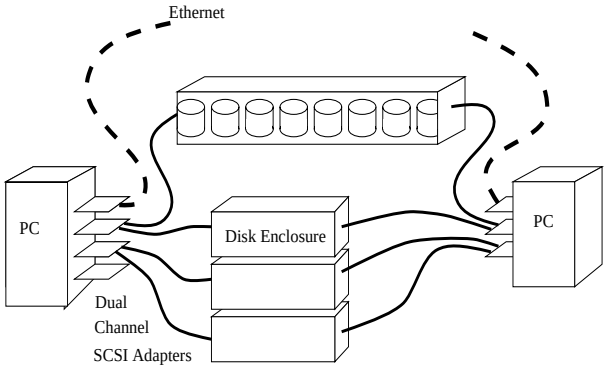

Figure 1. Design of a light node.

Management is one of the biggest challenges for a system like this. To simplify management and monitoring, we use programmable disk enclosures. Each enclosure has a serial port interface which supports a set of commands. Using this interface, the status of all enclosures can be monitored from a central location.

The application for this storage system is the Fine Art Image Database. Over 70,000 images, in different formats, are stored on this prototype. The next section describes the image database.

3. The Image Database

The prototype is used to serve the high-resolution images of each artwork over the web. Before the images are put on-line, they are processed by a half-dozen undergraduate students. This processing removes problems that may have occurred while the images were being photographed. The images are cropped, rotated to the correct orientation, and enhanced using color correction. Since this process is very subjective, it is hard to automate, and each image is processed by hand using Adobe PhotoShop. Once the images are processed, they are converted to a layered format that is presentable over the web.

When the museum first launched its web site, they chose not to put the larger images on-line for two reasons. The first was the lack of disk space. The second was the lack of a proper user interface. It is not practical to serve the user an entire 3,000 by 2,000 pixel image. Many users do not have the network bandwidth to support such transfers (it would take a user half an hour to download a 6 MB image with a 28.8k modem). Also, no display would be large enough to hold the full-size image. Finally, there are copyright restrictions that prevent high quality images of many of the art works from being freely distributed. These problems are avoided by converting the images to a layered format that allows zoom-in.

The images are stored in the GridPix format [3], a special format developed by the Tertiary Disk group. In this format, an image is stored as a series of tiles, each tile

encoded in JPEG. A GridPix file contains header information and a tile index, followed by a series of tiles in various resolutions. The GridPix viewer is a CGI-script that generates an HTML page describing an array of tiles. A second CGI-script retrieves the requested tiles from the GridPix file. Figure 2 shows the viewer. The page contains tiles inside a fixed size window. At lower resolutions, the image contains only a few tiles and fits completely inside the window. At higher resolutions, only a part of the image fits inside the window. Hence the user can study parts of an artwork in detail. Only the part of the image that is visible to the user is transferred over the network. One of the biggest advantages of GridPix is that caching is done entirely by the browser. When a user scrolls around an image that is too large to fit entirely within the viewer's window, only the new tiles that appear need to be retrieved from the server.

Figure 2. The GridPix Viewer

GridPix was developed after considering an alternative format, FlashPix [4]. FlashPix is a proprietary format developed by Eastman Kodak, Microsoft, and others. FlashPix also allows multiple resolutions of tiled images, but has several disadvantages. First, the current implementation requires a server and browser plug-in, limiting the types of machines that can be servers and clients. Second, since FlashPix is a general format that provides many features in addition to zoom-in, the files are quite large. Each of our images occupies 4-6 MB of space in FlashPix and 1-2 MB of space in GridPix.

The images are stored in several formats. Only the GridPix images are available to the outside world. Each image is also stored in TIFF format and PhotoCD format. The GridPix and TIFF versions are mirrored, i.e. two copies of each image are kept. Only a single copy of the PhotoCD images is kept because of space considerations. The GridPix images are generated from their TIFF counterparts. Since GridPix images cannot be edited using PhotoShop, the TIFF images make it possible to make changes later if necessary.

The homepage and search engine for the museum are located at the Fine Arts Museum (http://www.famsf.org/), while the large images are served by a group of http servers at Berkeley (http://art.cs.berkeley.edu). Each host in the prototype is an http server. The images from each PhotoCD form a *folder*. Each folder contains around 100 images and has a unique ID. Each image is contained in a single GridPix file. Concatenating the ID number of the folder and the image's number within the folder creates the unique key for each image. An image's key is translated into a URL for the image.

4. Status and Future Work

We are currently processing the images and converting them into GridPix format. The processed images are available to users through a web site at Berkeley (http://art.cs.berkeley.edu/). By February 1998, this site will be linked to the museum's main web site, with approximately 30% of the collection available in the layered format. From then on, images will be added to the site as they are processed. The search index at the main web page will allow users to search for those images that have the high-resolution counterparts available.

We also plan to offer a Java applet viewer as an alternative to the GridPix viewer. Once the web site is available to all users, we will study usage patterns and load balancing problems for this workload.

5. Acknowledgements

The image processing work is being done by UC Berkeley undergraduate students Gabriela Hernandez, Nicholas Huynh, Tony Le, Shankara Lowe, Victor Wong, and Wen-Kai Zhong. The art works were photographed by Sue Grinols of the Fine Arts Museums of San Francisco.

We wish to thank IBM for donating the disk drives used in Tertiary Disk and Intel for donating the PCs. The Tertiary Disk research is performed in conjunction with the Network of Workstations Project at UC Berkeley and is funded by the DARPA N00600-93-K-2481 Roboline-Storage contract.

References

[1] Grochowshi, E. Hoyt, R. Future Trends in Hard Disk Drives. *IEEE Transactions* on Magnetics, Vol. 32, No. 3. May 1996.

[2] Talagala, N. Asami, S. Patterson, D. Anderson, T. Tertiary Disk: Large Scale Distributed Storage. University of California at Berkeley, February 1997. Available at (http://now.cs.berkeley.edu/Td/papers/).

[3] Asami, S. The San Francisco-Berkeley Fine Arts Project, October 1997. Available at (http://now.cs.berkeley.edu/Td/talks/).

[4] FlashPix Specification. Kodak Corporation, 1997.