50 A high resolution digital imaging system for the photographic archive at the National Railway Museum, York

Ben Booth

50.1 INTRODUCTION

This paper describes the technical aspects of a project designed to electronically store and index the 750,000 images in the photography archive at the National Railway Museum. The purpose of this project is to make the images much more readily accessible to both public and staff. The photographs which date from about 1860, are mostly monochrome glass negatives and range in size from 35 mm to 26 inches by 21 inches, but there are also some film negatives and prints. It is expected that the archive will grow to 1.5 million images over the next two years. The project is managed by Christine Heap, Head of Library and Information Services at the National Railway Museum, with technical assistance provided by the Information Systems Section of the Science Museum.

The defined requirements for each component of the system are described. The available technology is reviewed, and the selection of the various components is documented. Whilst the guiding policy was to use well understood and widely available technology, this was not possible for the high resolution digital imaging system. This aspect of the system, and its selection are described in full, as we believe that it is the first use in a museum of a system of this type.

50.2 TECHNOLOGY REVIEW

Having decided that some form of image system was desirable, a survey of the available technology was conducted so as to characterise what was available. This survey was intended to provide sufficient information to ensure that our initial thoughts would be informed by the full range of what was possible, and would also enable us to eliminate at an early stage types of system which would be unable to provide the facilities which were required.

50.2.1 Low resolution digital
This is the type of quality of image which could be supported on a system based around an IBM PC or clone. Resolution of 640 by 480 pixels is possible, with an uncompressed image requiring approximately 300 K bytes of storage. A 600 Mb optical disk would therefore be able to store 2,000 images, or 10,000 at 5 to 1 compression. The digital technology enables images to be copied without any degradation. The technology for such a system is easily available, well understood, and relatively inexpensive. The cost for a system for image capture and archiving would be in the region of £25,000, with a display station costing about £7,000. The quality is a little below that of broadcast analogue (see below), with significantly fewer images being available at once on a single disk. The system in use at the Micro Gallery of the National Gallery in London is based on the Apple Macintosh computer, and has a resolution of 700 by 700 pixels. Higher resolutions than this would require a more powerful computer architecture. A low resolution digital system would provide a relatively inexpensive means of storing images where text and image need to be mixed on the same screen, and where image quality, and the quantity of images on–line are not important considerations.

50.2.2 Broadcast quality analogue
This is the type of image which appears on European television sets, and may be stored on video

tape or disk. Its quality is defined as "500 lines" which equates to a digital image of 768 by 575 pixels. A Philips Laservision disk is able to store 50,000 such images, whilst a Sony CRV disk can store 30,000 on each side. There is some degradation when images are copied. The cost of a system for capture and archive would be in the region of £35,000, with each display station (including a facility for indexing and text display) costing in the region of £10,000. In order to facilitate indexing, and the display of text associated with the image, a separate PC is required to control the disk device. It provides a convenient means of enabling a large number of images to be on–line at once, at an acceptable quality.

50.2.3 High definition television

Whilst at the time of this study there were no HDTV analogue devices in use, it was felt that this would become the domestic norm of the future, providing the manufacturers can agree on standards. This analogue technology at 1,125 lines is the equivalent of a digital image of 1920 by 1035 pixels. The images have an aspect ratio of 16:9, which gives it the characteristic "letterbox" shape which we are accustomed to seeing for films. It is envisaged that such systems would have similar advantages and disadvantages to the current "Broadcast Quality Analogue" systems, but at a much higher definition.

50.2.4 High resolution digital

Systems conforming to this definition are outside of what is commonly possible with a configuration based on an IBM PC or Apple Macintosh, and are typically in excess of 1,000 by 1,000 pixels. An image of 2,000 by 2,000 pixels would give a similar quality to HDTV (see above), and would be approaching the quality of a 35mm conventional photograph. An uncompressed monochrome image would require 4Mb of storage, which could be reduced to 800K at 20% compression. A 940Mb disk could hold 1,175 compressed images. The digital images could be copied without degradation. The cost of an image capture and archiving system would be of the order of £70,000, with a single display station costing £35,000. Additional display stations could be added for £8,000. Such a system would provide a high quality stable image, but such a system would be costly, and few images would be available simultaneously on–line.

50.2.5 Pre–press technology

This type of system is used to provide input to photo–lithography. Typically the images, which

are in digital format, are of high resolution. However such systems are very specific to the printing industry, and generally lack the database facilities which would be necessary to manage a library of more than a few hundred images. Furthermore the type of image which may be scanned are usually limited in range. Having reached the required standards for printing, the technology shows little sign of developing further, and from this point of view appears to be a "dead end".

50.2.6 Kodak's Photo–CD system

The newly announced Photo–CD system from Kodak stores images in digital form at a resolution of 2048 by 3072 pixels. Full colour images occupy 18 Mb in uncompressed form (6 Mb for monochrome), and are compressed to 2.5Mb for storage on a CD–ROM at a density of 100 per disk. Input is from 35 mm negatives and transparencies only, and images are of similar quality to conventional 35 mm photographs. The digital images may be exported, but facilities to import in digital form were not likely to be available for the foreseeable future. In the first instance Kodak are aiming this product at the domestic market, and are not offering any special developments (such as digital input, or input at other than 35 mm) for technical users. The system would be available in late 1992.

50.2.7 Initial reactions

The survey suggested that we should reject from further consideration low resolution digital imaging, as the quality was below the present threshold of domestic television, and would certainly become unacceptable to users once HDTV had become commonplace. Furthermore the storage overhead for large numbers of images was high, and a computer system based on a PC would not have the power to manipulate the potentially large amounts of data involved. Pre–press technology offered high quality images, but was also rejected, as it lacked the necessary systems for storing large numbers of images, and it seemed to have reached the limits of its development.

Whilst the HDTV standard had not yet been implemented it was of interest as it seemed likely to be commonly used in peoples homes in a few years time. In the same way that "Broadcast quality analogue" may be said to be the minimum acceptable standard for today, it seemed likely that HDTV would become such a standard for the future, and that any images which would have more than a few years life should at least conform to this standard. The Kodak Photo–CD system was not one we could use immediately, as it was

not yet available, and the lack of support for technical users was not encouraging. However if it was successful, it would be likely to become a significant standard for digital imaging, as well as potentially lowering the costs of such systems by virtue of the large numbers being produced.

"Broadcast quality analogue" technology appeared to be an economic and effective means of accessing a large number of images, at a quality which is acceptable today. Images in such a system would cease to be acceptable once HDTV became established, and would not be suitable for long term archiving, as the analogue format inevitably involves some degradation in copying. Such a system appeared to be suitable for the display of images, at a quality which would be acceptable over the next few years. High resolution digital technology appeared to be suitable for the long term archiving of images, at whatever standard (HDTV or Photo–CD) was felt to be appropriate for the project. It had the advantages of stability and quality, but was costly, and few images could be available on–line at once.

50.3 COMPONENTS OF THE SYSTEM

The initial scoping study for the project identified its size, and inferred that it was likely to have a long life, both because of the time required to capture the images, and because the resultant investment would need to be long lived to be justified. It therefore became apparent that high quality image storage with archival longevity was required. For display, images of acceptable quality but in large numbers would be required. These two were clearly contradictory, and it was therefore decided to consider the image capture and archive system separately from the image display system. Further investigation could light on an acceptable system to perform both functions, or it would be necessary for images captured and stored in one system to be transferred to the other for display. The final element to be considered was the "text database", the system for the indexing and retrieval of images.

50.4 DEFINITION OF REQUIREMENTS

Following identification of the elements of the proposed system, a formal requirement for each component was prepared. In defining the requirement, and the subsequent selection of the system, we were assisted by Peter Cheese and other staff at Cambridge Multimedia.

50.4.1 Text database
The requirements for the text database were defined as follows:

- To support the required data structure;
- To permit data entry, editing and retrieval;
- To readily control the display system;
- The technology should be easily obtained and understood.

50.4.2 Image display
The requirements for the image display were defined as follows:

- Quality must be at least equivalent to "Broadcast Quality Analogue";
- A large number (minimum 30,000) of images should be available at once;
- The technology should be well understood and easily obtainable;
- A single display station should be inexpensive (less than £10,000 per station);
- The technology should be appropriate for use elsewhere in the Museum;
- It should be possible to make duplicate copies of images for use on several display stations.

50.4.3 Image capture and archive
The requirements for the image capture and archiving system were defined as follows:

- The quality should be equivalent or in excess of HDTV;
- It should be possible to copy images with minimum degradation or nil degradation if regular copying was required;
- When properly stored the images should have an indefinite life;
- Images should be able to be readily transferred to other media (e.g. the display system);
- It should be practical to manage an archive of 800,000 images;
- It is desirable to be able to produce high quality prints directly from the archive copy of the image;
- Because the major cost of the project would be the staff required for image capture, it was accepted that the image capture and archiving element of the system could be relatively costly (within budgetary limits).

50.5 SELECTION OF SYSTEMS

In selecting the elements of the system, it was decided that for the text databas, and the image dis-

play system, we should go for the simplest technology, as it was apparent that the image capture and archiving system could involve considerable complexity, and would require significant resources to be successfully implemented. It became apparent that the criteria for the capture and archive and display elements of the system differed too much for the same system to suffice for both. Separate systems, between which images could be transferred would therefore be required. The systems selected, and the reasons for this selection are described below.

50.5.1 Text database

The system selected for the text database was Foxpro software, running on a PC–Clone (Opus). These are well understood hardware and software, which were adequate for what was required. They are the standard configuration in the Museum, and effectively conform to the "industry standard". They could easily be interfaced to other systems. This decision did not preclude the transfer to more sophisticated software at a later date.

50.5.2 Image display system

Because of the requirement to make available a large number of images at least of "Broadcast quality analogue" quality the choice was between Philips Laservision, or Sony CRV Technologies, both of which offered the required quality and quantity of images at an affordable price. Whilst the Philips system is less expensiv!!e, it would require laservision disks to be mastered away from the museum. It was felt preferable to use the Sony "Worm" disks, as images could be transferred to this media as required.

50.5.3 Image capture and archive system

As the first stage of selecting the appropriate technology for image capture and archiving, it was possible to reject Broadcast Quality Analogue technology on the grounds of quality and durability, low resolution digital on the grounds of quality, and Pre–press technology as it had no obvious development path, lacked database support, and lacked flexibility of input. Analogue capture and storage at HDTV resolution was not yet available, but seemed likely to become the future standard for domestic television.

Initial capture onto 35 mm film was reviewed, as this appeared to be the "orthodox" approach to this problem. It was however rejected as photographic expertise would be required to capture the images. Additionally the storage of 800,000 negatives would present logistical problems, and

the transfer from 35 mm film to electronic storage could not be achieved completely automatically.

Kodak's Photo–CD was considered, as it offered the necessary quality, and seemed set to become the standard. However it was rejected because it would involve the costly production of a 35mm intermaster. In addition it lacked the database functions necessary for a large library of images, and the initial concentration on the domestic market indicated it would not be suitable for our application in its early form. Finally, and most important, it was not available at the time of system selection. However because Photo–CD seemed likely to become the standard of the future, it was felt to be desirable to select a system which could eventually transfer its images to Photo–CD. A decision was therefore taken to raise the desired standard for the image capture and archive system from that of HDTV, to Photo–CD.

High resolution digital imaging was selected as the technology, as it was a stable medium, at high quality. The capture equipment could be operated by staff without photographic skills, as the images may be seen immediately. The images could be readily transferred to other systems, whether lower quality digital, or analogue. The production of high–quality output was possible directly from the digital image. The disadvantages of this approach would be cost and the requirements of managing an archive which was likely to grow to many hundreds of optical disks.

50.6 THE DIGITAL IMAGE CAPTURE AND ARCHIVING SYSTEM

Having selected the technology, a detailed request for proposal was prepared, and sent out to suppliers who had been identified as being able to provide a system broadly conforming to our requirements. Selection of potential suppliers was problematic, as whilst there are a number who produce the various hardware and software components of such a system, there were few who could supply a complete system. The request for proposals was prepared according to the UK Government recommendations for such documents (CCTA 1989).

A Varsity system from Primagraphics of Royston, England was selected and is due for installation in June 1992. The system is a monochrome imaging archive system, which captures and stores images at 2500 by 3500 pixels. It is capable of enhancement to 5000 by 7000 pixels and

can be upgraded to full colour. With the addition of an appropriate interface and driver, a colour printer may be added.

The Varsity system incorporates an MC68030 based host processor and has both a VME bus and Primabus, a proprietary high speed bus capable of data transfer speeds of up to 60Mbytes per second. Specialist graphics hardware within the system, which links to both busses, provides facilities for image capture, display, handling and processing, compression (to the JPEG standard) and storage. The operating system is Unix version 5.3, which is used in conjunction with Primagraphics' own image processing and graphics library called Primalib. The system is also equipped with a 638 MbWinchester disc drive, a 940 Mb optical disk drive unit, and a 5.24inch floppy disk drive. The monitor is a Sony 19 inch high resolution colour display monitor. Scanning is carried out by Primascan, a 5 K by 7 K pixel scanning camera developed by Primagraphics, mounted on a specially constructed operator console incorporating the camera stand and light box.

To capture an image a negative is placed on the light box, and scanned by the camera. The scanned image is converted from negative to positive, and is then displayed on the high resolution monitor. At this point the operator has the opportunity to adjust the brightness and contrast, and perform any required cropping. The image may then be saved to the optical disk, either at a predetermined, or selected compression ratio. During the storage process the operator is prompted for the negative number of the image, which is then stored in the database used to retrieve images. Once stored images may be retrieved either by negative number, or the optical disk and frame numbers. With the addition to the system of a colour dye sublimation printer, printed output of "photographic quality" at A4 size may be produced in a few minutes.

Images stored in the high resolution system may be transferred when required to the Sony CRV device which will be used to display the images. The Sony system, which can store up to 60,000 images at "broadcast quality" is linked to an Opus PC, on which is a textual database which is used to retrieve images from the Sony system. This database, which uses the FoxPro software, facilitates retrieval according to a variety of criteria.

50.7 CONCLUSION

Our investigation of the technical options for the National Railway Museum's image archive has shown that there are at present a range of systems for electronic imaging, many of which would be applicable in appropriate circumstances. It is argued that the selection of an effective system depends on a thorough definition of requirements and comprehensive assessment of the available systems against this requirement. Where a collection of photographs is to be treated in this way, the process of data capture is likely to take many years and will inevitably be costly. The investment in the image data will be far greater than the hardware and software costs of the system, and it is this image archive which must be conserved. It is therefore essential that the data is collected in a form which will be durable and is of sufficient quality to satisfy users in the distant future. In addition to the need for a high quality and stable archive, it is also essential to be able to transfer the image data to other systems as technology develops. The study conducted for the National Railway Museum's image archive suggests that such an archive should be in digital form, and should be of at least HDTV quality. The likely primacy of Kodak's Photo–CD suggests that the Photo–CD is the standard to which should be aspired.

Reference

CCTA
1989 *The Information Systems Guides: B5 The Operational Requirement.* John Wiley & Sons, Chichester, England.

Author's address

Ben Booth
The Science Museum
Exhibition Road
GB–SW7 2DD London
Tel 071–938–8210
Fax 071–938–9778