Digital Video for Multimedia: Considerations for Capture, Use and Delivery

Jane Williams*, Alan Lock and Caroline Burnett

Multimedia Resources Unit
Educational Technology Service
University of Bristol
Royal Fort Annexe
Tyndall Avenue
Bristol BS8 IUJ

*e-mail: J.Willams@bristol.ac.uk
http://www.ets.bris.ac.uk/

Support Initiative for Multimedia Applications

Part of the JISC New Technologies Initiative

Contents

Introduction

The power of today's personal computers, and the recent establishment of the MPC (Multimedia Personal Computer) III standard, have increased the ease of digital video capture and the use of motion video within teaching and learning. Motion video is made up of a series of still frames played at a speed of 25 frames per second. However, although still frame or image capture is now relatively straight forward, motion video capture is more complicated than capturing individual frames and playing them together in a sequence.

Digital video can be divided into two main areas:

1. That which has been captured/digitised from an analogue video source, e.g. VHS tape, camcorder, live video feed

2. Animations created entirely in the digital domain, e.g. using software packages such as 3D Studio™ and Lightwave™

In each case the resulting video may subsequently be viewed in the analogue or digital domain.

There are, of course, many examples of video which combine these two areas. These include television adverts and programs such as Babylon 5, which makes extensive use of Lightwave, films and many CD-ROM programs.

What this Report is About

The work presented here is concerned solely with video from analogue sources and investigates considerations for its capture, use and delivery for multimedia applications. It focuses on the PC (Windows) platform as this reflects our experience. Other platforms will be mentioned where possible. Our investigations are presented in two formats: much of the theoretical information is presented here; other information including examples of digital video files, as indicated in the relevant sections, can be found at the following World Wide Web site: http://www.ets.bris.ac.uk.

The work forms part of the activities of the Multimedia Resources Unit which specialises in the digital capture, storage and network distribution of resources such as still images, video and sound, for use in electronic documents and technology enhanced teaching materials.

The report is divided into a number of sections:

Section 1: Video and learning:

Considerations and suggestions on when to use video as an aid to learning.
In some instances video may not be suitable, or the desired effect may be achieved using alternative techniques that do not place high demands on computer hardware. Some alternatives will be suggested and discussed

Section 2: Digital video: issues and choices

Digital video makes great demands on storage space and memory and there are therefore a considerable number of technical issues that need to be considered before and during the course of capturing digital video. Guidelines and tips for capturing and editing digital video are provided.

Section 3: Digital Video on Trial

A practical examination of the technical issues involved with digital video capture. A series of video clips has been prepared illustrating the results of using various compression/decompression algorithms and changing other parameters during video capture and preparation.

Section 4: Scenarios

Three examples of the use of digital video and capture techniques within teaching and research.

Section 5: The Future

This section reports on the most recent advances in digital video since the project's inception

Section 1: Video and Learning

"Throughout the twentieth century, motion pictures have been the vehicle for much of our art, information and entertainment. In the twenty-first century, it will be the way we view those motion pictures, and how we choose to view them that will further extend their scope as communications media. New levels of stereoscopic realism will be offered by virtual reality systems, new interactions of sight and sound will be discovered; and we will develop new kinds of critical awareness as we begin to take personal control of the balance between film and still, "fast forward" and "stop-motion", "close-up" and "longshot"

Cotton & Oliver, 1993

The representation of information by using the visualisation capabilities of video, whether analogue or digital, can be immediate and powerful. While this is not in doubt, there are other possibilities that digital video offers communication. As quoted above, it is the ability to choose how we view, and interact, with the content of digital video that provides new and exciting possibilities for the use of digital video in education.

It is not intended for this section to provide a complete review or understanding of the use of visualisation in learning, as this is adequately covered elsewhere (Rieber, 1994 and references therein), but merely to draw attention to some considerations of use before expanding on the technical issues.

Rieber (1994) writing on the subject of Multimedia and Interactive Video explores a number of levels of interactive video. The lowest levels include video technology only. Students and teachers can manually interrupt the video, usually stop/start, which may be delivered from videotape or videodisc. Moving up the levels, the ability to interact with the video becomes more sophisticated. Here the video and computer technologies are linked using extra hardware such as keypads, barcode readers or at a higher level, overlay facilities where the video is played onto a second monitor or the computer screen allowing text and graphic overlays. The highest level in the taxonomy is referred to as the "dare to dream" level. Traditionally, a typical system at this level will include an expensive, creative assortment of hardware with various video units, sound systems, touch screens, etc. With the advent of digital video and multimedia technologies in general, interactive video takes on a completely new definition; the ability to interact with video no longer being limited by expensive hardware.

Writing in this section, Rieber (1994) discusses the use of video as an appropriate tool to convey information about environments which can be either dangerous or too costly to consider, or recreate, in real life. Obvious examples include simulating dangerous experiments. For example: where video images are used to demonstrate particular chemical reactions without exposing students to highly volatile chemicals; medical education, where real-life situations can be better understood if represented by video. Here such rationales for video are rooted in the cognitive domain.

There are many instances where, studying particular processes, students may find themselves faced with a scenario which seems highly complex when conveyed in purely text form, or by the use of diagrams and images. In such situations the representational qualities of video are rich, and when used well, will help in placing a theoretical concept into context. Other contributory factors such as sound, and the ability to control the pace of information, help to create an environment which transcends the inanimate representation, and allows potentially real-life situations to be fully explored.

The video clips need only have a short duration; material which is only a few seconds long can be rich and powerful in content and to the point. Exploring such environments offers another dimension to the educational process - lectures, printed materials, tutorials and small group teaching, can all offer various stimuli and motivation. One of the most compelling justifications for video may be its dramatic and immediate ability to elicit an emotional response from an individual. Such a reaction can provide a strong motivational incentive to choose and persist in a task. This ability to elicit an emotional response transcends the often made assumption that technology offers a cold, and emotionally barren environment.

"Simulation techniques play an extremely important role in the development of multimedia CAL resources. They can be used in a variety of ways to make learning more exciting"

Barker, 1989

Other examples combine cognitive and motivational elements by using video to provide a meaningful context for learning. Reiber (1994) provides us with the example of students being shown segments of a movie, where, Indiana Jones replaces a golden idol with a sandbag to prevent the booby traps from being triggered as a context for understanding the relationship between volume and weight.

Visuals and Learning:

It is very tempting to use the latest computer wizardry to represent information and develop computer enhanced learning materials. However, the instructional design of these systems should be based on a careful examination and analysis of the many factors, both human and technical, relating to visual learning.

In a number of, now very famous, experiments carried out in the 1970s it was concluded that not only does the brain have an extraordinary capacity to imprint and recall, but that it can do so, with no loss of accuracy, at incredibly high speeds (Haber, 1970; Standing, 1973). Standing (1973) commented that "the capacity of recognition memory for pictures is almost limitless". Pictures have a direct route to long-term memory, each image storing its own information as a coherent 'chunk' or concept (Standing, 1973; Paivio, 1975; and Erdelyi and Stein, 1981). There is no limit to the number of "chunks" or concepts that that can be stored and retrieved (Miller, 1956; Gage and Berliner, 1988).

The reason for this is that images make use of a massive range of cortical skills; colour, form, line, dimension, texture, visual rhythm, and especially imagination (Buzan, 1990). Imagination comes from the Latin *imaginari*, meaning to 'picture mentally'. Images are generally more evocative than words, more precise and potent in triggering a wide range of associations, thereby enhancing creative thinking and memory.

However, before any image offers the potential for increased learning, a need for external aids to visualisation must be established. Reviewers (Dwyer, 1978; Levin, Anglin & Carney, 1987) have stressed the importance of first determining whether a textual passage alone elicits adequate internal imaging by students. If internal imaging does not take place, then evaluate the effectiveness of including a picture. If the use of text alone succeeds in creating adequate internal imaging, the inclusion of external visuals will probably not result in any additional learning gains. One could argue that adding such visuals may add aesthetic quality and render the material more attractive to users. Users often regard material as being of higher quality if images are present. However, there is always the potential that unnecessary visuals may distract. Even if text alone does not sufficiently induce appropriate mental imaging, visuals must be congruent, relevant, and consistent with the information presented in the text in order to be effective (Levin and Lesgold, 1978).

Care needs to be taken when using visuals for aesthetic reasons. The misuse of a single visual element can cause misrepresentation of information and become a barrier to content and impede learning, even if the program overall may, in all other aspects, follow the principles of instructional design. It is important to bear in mind the nature of the audience, especially their age group and culture mix; computer enhanced materials for education and entertainment have different objectives. Each have their own set of implications and considerations. That is not to say, however, that educational materials need not be exciting.

Literature on the educational use of video and animation is scarce. However, we believe that it is not unrealistic to apply the above to the use of video and animation.

Alternatives to Digital Video

Digital video offers powerful and instant representational and communication opportunities, not always offered by other forms of media[1]. When addressing the production of educational material it is important to be aware of alternative, more educationally appropriate, ways of presenting the content, even if this means challenging user assumptions. Defining the constraints of time and resource also helps to determine a ‘correct’ medium. Even though technological innovation offers endless potential and challenges for education, we must not forget the power of the inanimate, or traditional. Some questions to consider when planning the production stages of a project are “What does each medium contribute to the representation of information?” and “What does it offer in this context, that other media cannot ?”

When defining the appropriate medium to use it is vital to ‘know’ your audience and the technical specification of users’ machines. There may be technical reasons for choosing which multimedia element will best communicate certain concepts.

What ever medium is chosen, to apply a principle mentioned above to all digital media

[1] Note, in this instance media refers to different elements such as text, still images, video, sound, animation, etc.

elements, visuals must be congruent, relevant, and consistent with other information presented in order to be effective. Whatever the latest technological advance, instructional design principles apply.

Still images

We only have to look at the power of advertising in our society to see that still images can be an extremely powerful mode of communication. Still images can prove a useful alternative to a video sequence if there are technical reasons why video is not possible. A series of images run through quickly in a sequence can show the effects of the passage of time or be combined with text to make a presentation.

Animation

With the use of such literary devices as metaphors to remind us, the visual word can be a powerful and creative component to communication. However, in particular instances, the inanimate can become restricting, and concepts which can be too wordy to explain in text form can be conveyed simply using animation. Animation, by definition, provides the illusion of movement. However, there is a more subtle character of animation, which distinguishes it from other visual elements; trajectory, or the path of travel by the animated or moving object. Using animation, learning must depend on understanding either changes in an object over time (motion) or changes in the direction of motion in which the object is moving (trajectory), or both. If there is no case for the second requirement, then there is no reason why animated visuals would aid learning anymore than static visuals. Again it could be argued that these additional effects of motion and trajectory could be distracting. Static visuals would be sufficient for tasks that only require learners to visualise information. However, if a task demands that learners understand changes over time or in a certain direction, then static visuals can only hope to prompt learners to mentally construct these attributes on their own. Animation makes this cognitive task more concrete and spontaneous by providing the motion and trajectory attributes directly to the learner (Reiber, 1994). This should increase the potential for successful and accurate encoding into long-term memory. Preliminary research has shown that animation displayed with accompanying narration produces greater retention and recall than when either are presented separately or when verbal descriptions are presented before or after the animation (Meyer and Anderson, 1991)

QTVR - Quicktime Virtual Reality

Virtual Reality (VR) describes a range of experiences that enable a person to interact with, and explore, a spatial environment through a computer. These environments are typically artistic renderings of real or imagined spaces, generated on the computer. To achieve the desired effect, most VR applications require specialised hardware or accessories, a high-end graphics workstations, stereo displays, 3-D goggles, or gloves. Apple’s QTVR offers an alternative by using video techniques to depict real-world scenes. QTVR is a technique whereby still images are combined into a navigable, pseudo-3D environment, in which can be embedded other interactive objects such as video sequences, sound files, still images, other QTVR sequences, etc. Being created from still images, it currently lacks the temporal element normally associated with video sequences. The resulting files are extremely small when compared with their equivalent video counterparts.

Sound

This medium is often misused in computer enhanced learning environments and is often used as audio feedback to an action, or as a warning. Used creatively, however, it can become a stimulus to the imagination; used inappropriately it can become a hindrance or an annoyance. A script, some still images and a sound track, allow users to utilise their own power of imagination without being biased and influenced by the inappropriate use of video footage. A book is often spoilt by first watching its recreation on film.

Conclusion

The fusion of all types of media in a digital world captures the ethos of the new technological age. Multimedia: a combination of video, text, still images and sound, etc, can provide an effective learning aid. With such a hybrid form of communication it is important to understand the capabilities possessed by each medium, the knowledge of which allows choices to be made relating to the most effective way of conveying information. Each medium is a potential channel to the content and should not be fighting for representational prominence.

Section 2: Digital video: issues and choices

Moving video is made up of a series of still frames (images) played at 25 frames per second (30 frames per second in the USA). One may be excused for thinking that the capture and playback of digital video is simply a matter of capturing each frame, or image, and playing them back in a sequence at 25 frames per second. Unfortunately it is not that simple.

A single still image or frame with a window size or screen resolution of 640 x 480 pixels and 24 bit colour (16.8 million colours) occupies approximately 1MB of disc space[1]. Therefore, roughly 25 MB of disc space are needed for every second of video, 1.5 GB for every minute. Digital video files are large! Even if the storage space is available, it is not practical to play back that amount of data per second on today's personal computers.

Therefore, file size must be reduced to within the data transfer rate of the final playback system.

The three basic problems of digital video

There are three basic problems with digital video

- size of video window
- frame rate
- quality of image

It is not difficult to deal with these three problems and, although we will now take a look at these issues one at a time, they are by no means independent of one another. All of these issues can be tackled using compression techniques.

Size of video window

Digital video stores a lot of information about each pixel in each image or frame. It takes time to display those pixels on your computer screen. If the window size is small, then the time taken to draw the pixels is less. If the window size is large, there may not be enough time to display the image or single frame before it's time to start the next one. Remember, we need to draw 25 frames per second.

There are several ways to solve this problem:

- Choose an appropriate window size. It is often not necessary to have full screen video; small window sizes may be perfectly adequate for your needs, especially as with multimedia applications, there are other objects that need to be on screen at the same time. Viewing video on computer screens is not the same as watching video on television. Users will be close to the screen; in some cases a window size of 160 x

---

[1] For futher details on still image capture and related terms such as resolution, 24 bit colour, etc, please refer to Williams *et al.*, 1995

120 for a screen set to a 640 x 480 display is adequate (Hamilton *et al., 1995).*

- The reduction of the window size to anything less than the size or resolution of the original video source (in the case of the European video standard, PAL, this is 576 horizontal lines) will involve some sort of compression. This is because the same information is being represented but in less space. Reducing the window size may not always, therefore, produce desirable results and will depend upon the content of the video.

- Fast hard discs are now available enabling each frame to be read from the disc faster

- Use hardware accelerated playback. The graphic display cards now being supplied with PCs, more often than not,include facilities for Windows acceleration and video playback.

Frame Rates

The issues here are similar to those above - too many pixels and not enough time. There is not enough time to move the data from hard disc or CD to screen. One way to overcome this is to compress the data so that less data is transferred from disc to screen.

Depending on the size of video window chosen, you may also be able to reduce file size by reducing the number of frames per second to, for example, 12 frames per second. At smaller window sizes eg 160 x 120 pixels, video played at reduced frame rates is, in the majority of cases, acceptable. At larger window sizes, the video can sometimes appear jerky.

Image Quality

The image quality will depend on the quality of the original source and the degree of compression used.

During compression you will probably be asked to select a quality setting. This will be represented by an arbitrary scale of, 0-100%, 1-5, etc. A lower setting will result in greater compression and smaller file sizes but the quality of the resulting video sequence will be reduced. The relationship between file sizes and quality varies considerably for different codecs and is illustrated in Section 3: Digital Video on Trial

Compression

Compression is achieved using algorithms (mathematical formulae), which identify the information that needs to be recorded and stored. It is then 'reconstructed' during decompression. Compression can be implemented in either hardware or software. Software compression is slow compared to hardware compression.

CODEC: a contraction of **CO**mpressor **DEC**ompressor. As its name implies its main function is to: i) compress the video while digitising and ii) decompress the video during playback

Hardware codecs, as found on video capture cards, are highly optimised for compression speed. As they are hardware based they are difficult to upgrade. Software codecs are generally optimised for decompression (playback) and are necessary for playing back the digital video on the user's computer. Upgrades are usually a simple matter of installing new versions of the driver. It is common to have many different software codecs on one computer. It is essential, during playback, to match the codec used to make the particular video sequence.

When capturing video, some capture cards allow real-time compression such as Creative Labs' VideoBlaster RT 300. However, in many instances the video is captured 'raw' and compressed later in software. There is still some compression occurring at this 'raw' level, with ratios of approximately 6:1 for quarter screen (320 x 240 pixels) video, but in such a way that minimal information is lost and software compression at a later time does not cause artefacts, that is new and unwanted information, to be added.

There are two types of compression:

- lossless, in which all the data is preserved and typically will compress images 2:1

- lossy, in which the data is degraded, the more so the greater the compression ratio used.

Lossless techniques are mainly used for text-based data where compression rates are quite high due to common letter groupings, etc. For images, techniques such as run length encoding are employed within some image formats such as PCX and BMP to reduce file size. Run-length encoding takes stretches of pixels sharing the same colour and stores the information for these pixels in just two bytes; one for the colour and the other for the number of adjacent pixels. Ratios of typically 2 or 3:1 can be achieved with this technique. Large areas of the same colour are not normally encountered in moving video as information changes between frames and therefore lossy compression techniques are relied on to reduce the data to a manageable size.

Lossy compression techniques seek a compromise between quality and quantity and rely on human ability to compensate for losses, exploiting the way we perceive. However, there are some subject areas where the use of lossy techniques demands serious attention and research, particularly in the medical field. Many of these techniques are designed to compress moving video as well as still images. Such techniques include JPEG and PhotoCD for still images, MPEG, Fractal compression (still and video), Video for Windows and Apple Quicktime. These will be explained in more detail further on in this section.

It is the various algorithms used, and the way in which they are applied, that differentiates the various codecs and gives them their relative strengths. Both lossy and lossless techniques can be, and usually are, applied both spatially and temporally.

- **Spatial** compression, more commonly known as intraframe compression, is applied to a single frame. Both lossy and lossless techniques may be used.

- **Temporal** compression, or interframe compression, looks for differences between successive frames and stores only those differences. Both lossy and lossless techniques may be used.

With lossy techniques the difficulties arise when deciding what information to discard and how best to disguise its removal. This is where the codecs show their differences. For practical examples of video sequences compressed using different codecs visit the following World Wide Web site: http://www.ets.bris.ac.uk

Decompression

It is during this process that the various codecs display their strengths and weaknesses. The compressed video file is passed into the codec which expands and reconstructs the data back to its uncompressed state. There are a number of implications here, not least the bandwidth of the bus which transfers the data to the display card. For a 320 x 240 pixel with 24 bit colour at 25 frames per second (fps) the amount of uncompressed data is 5.76 MB per second. This does not leave much time for the computer to handle any other functions, that is assuming the computer can handle this process in the first place.

The demands placed on all parts of the system at this stage are very high and it is the CODEC that is responsible for overseeing these demands. These include receiving any data, decompressing the data as fast as possible to as high a quality as possible, transferring the data to the display card and detecting whether the system is capable of handling these processes. If not, frames will be dropped during playback.

Battle of the codecs

Before providing details of the various codecs available, we will first take a look at Video for Windows and Quicktime and clear up an area of potential confusion. Neither

Video for Windows nor Quicktime are codecs in themselves, but rather, they act as containers for new video codecs

Video for Windows from Microsoft set the standard for incorporating digital video under Windows by creating a new file standard called AVI (Audio Video Interleaved). The AVI format merely defines how the video and audio will be stored on your hard disc. That is, the video and audio are laid down with frame 1 of the audio, followed by frame 1 of the video, the same for frame 2 and so on; a process referred to as Interleaving. This may appear simple but is important, as without interleaving, programs would have to jump from place to place on your hard disc to find the next bit in the sequence. This slows things down and so anything that reduces the demands made on the hard disc by video is important. What AVI does not do, however, is define how the video will be captured, compressed or played back. This means that as new technology for video is introduced, i.e. a new codec, it can be incorporated into Video for Windows. AVI files may be played using the Media Player supplied with Windows 3.1x or Windows 95 or from within applications.

Quicktime: Apple have provided their own video-plus-audio file format equivalent to Video for Windows, called Quicktime. Quicktime provides a basic set of software schemes that meet a range of compression needs for still images, animation, video and sound. A codec for playing Quicktime movies is available for Windows (although Quicktime movies play back slower on PCs than AVI files) and converters for Quicktime to the AVI and MPEG (see below) formats are available. Quicktime 2.0 will capture movies at 30fps and quarter screen resolution (320 x 240). If 15 fps is acceptable you can show movies at full screen but performance will depend on the power of your MAC.

A number of codecs which support the AVI format exist for Video for Windows and Quicktime's video-plus-audio format and include Indeo 3 (Intel), Cinepak (licensed from SuperMac) and Microsoft Video 1, offering compression ratios of up to 50 to 60:1.

Indeo: One of the most important advantages of Indeo is that compression can take place in real-time, that is 'on-the-fly'. Many video capture cards support Indeo compression. Decompression is in software only. However, if the video requires extensive editing, real-time compression should not be used. Indeo is a proprietary blend of colour sampling, vector quantization and run-length encoding. It is a 24 bit codec and dithers well to 256 colours when displayed on an 8 bit display; there is no need to reduce the number of colours as the codec copes by itself. Having said that, reducing to 256 colours helps reduce file size. Needless to say, to reap the benefits, Indeo video should be played on 16 bit or higher displays. Compression ratios of 10:1 are obtainable. Two versions of the codec in use are Indeo 3.2 and Indeo Interactive.

Indeo Interactive is an entirely new hybrid, wavelet-based software codec that enables real-time interaction and control of video and graphics imagery in multimedia/game applications. For a comparison of this with MPEG compression and other codecs visit the following World Wide Web address: http://www.ets.bris.ac.uk

Wavelet: wavelet based technology is not a codec in itself but a type of transform used during video compression. Another type of transform that is employed within codec technology is DCT (Discrete Cosine Transform) used in JPEG and MPEG compression algorithms

Cinepak: As with Indeo, Cinepak also offers very good image quality. Cinepak is a vector quantization based codec. Vector quantization stores information about differences between frames of video by quantifying the magnitude and direction of a pixel's movement. During decompression the codec uses a CLUT (colour look up table) to recreate the colour of each pixel in a frame. It is a highly asymmetric codec, taking 300 times longer to compress than decompress. Decompression is highly efficient. General opinion is that Cinepak is better than Indeo 3.2 for high action sequences. Both compression and decompression are performed in software. Compression ratios of 10-20:1 are obtainable. Cinepak was originally developed by SuperMac for integration into Apple's Quicktime (see below) but has been licensed to Microsoft for Video for Windows.

Microsoft Video 1: Is not in the same league as Indeo and Cinepak. It is a simple codec based on run length encoding and optimised for animation or cartoons.

AVI and Quicktime files can be played back in software alone and because of this the speed of playback and the size of the video window will depend on the power of the processor and graphic capability of the machine. The video will be scaled accordingly. A clip that looks perfectly acceptable on a Pentium 90 system may be barely recognisable as a piece of video on a 20 MHz 386. On a Pentium, Cinepak and Indeo can achieve 25 fps at 320 x 240 pixels and can be enlarged to 640 x 480 using graphics acceleration.

Both Video for Windows and Apple Quicktime are designed with open codec architectures. This is important for the future of digital video as it allows new codecs to be incorporated as they are developed. An example is Indeo Interactive - the most recent version of Indeo. To make use of this new codec is simply a matter of downloading the new drivers from Intel's World Wide Web site and installing them within Video for Windows. There is no need to buy any new hardware or software.

Other codecs include:

JPEG: As with still image formats, the widespread need for compression methods has resulted in the emergence of a plethora of techniques. Consequently the International Standards Organisation (ISO) set up two groups, the Joint Photographic Expert Group and the Motion Picture Expert Group (MPEG) to establish international standards for the compression/decompression of still and moving video and associated audio. JPEG is now a well-established codec for still image compression. It removes the redundancies in individual frames.

Motion JPEG (MJPEG) is a modified version of standard JPEG based on the same algorithms as JPEG to create I-frames (compressed intraframes). MJPEG capture boards are available, but beware, there are various non-compatible versions of MJPEG around. The codec is symmetrical; compression and decompression taking around one-thirtieth of a second for each I frame.

MPEG: Already we have two standards for MPEG - these are MPEG I (a sub-set of which has been defined for VideoCD/White Book CD and CD-I) and MPEG II. MPEG II is designed to offer higher quality at a bandwidth of 1.2 Mbit/second at 704 x 480 pixels and 30 frames per second and is used for images of high definition TV size. MPEG I has been developed to fit into a bandwidth of 1.5 Mbit/second to allow data retrieval from single speed CD-ROMs at 320 x 240 pixels at 30 fps. Compression ratios from 30:1 to 200:1 are obtainable.

MPEG is even more advanced than Motion JPEG and uses a process called predictive calculation. MPEG uses the same algorithms as for JPEG to create one I-frame. Information in the current I-frame is used to predict the information in following frames; the differences from its predictions being encoded only. This is known as Interframe compression, the frames being referred to a P frames.

A further standard, MPEG IV (incorporating MPEG III) is under development. Algorithms for playing MPEG I movies in software alone are available for use under Video for Windows and Apple's Quicktime but will suffer the same scalar problems. Boards for providing hardware-assisted playback are also available, allowing full-screen, full-motion video. However, MPEG is still not mainstream technology and several computer magazines are reporting on incompatibility problems between computer, CD-ROM drive and MPEG playback cards and software.

One of the drawbacks of MPEG is that a great deal of processing power needs to be applied to perform the compression in the first place. Traditionally this has meant expensive, dedicated MPEG editing systems, or paying a lot of money to a bureau to do it for you. However, relatively low-cost expansion cards are beginning to emerge which claim to do the job for you. We have trialed one such card - see section 3: Digital Video on Trial for further details.

Fractal Technology: A proprietary format developed by Iterated Systems, it offers greater compression using algorithms based on fractal transforms. Still images can be compressed by up to 100:1. For video, Iterated Systems have developed “Softvideo,” providing full screen colour video at 30 frames per second on a PC using software alone. However it takes 15 hours to compress one minute but decompression is fast. A key factor of fractal compression, for both still and moving images, is scalability. That is, the video’s resolution is independent of the size of window in which it plays. As fractal images are encoded with equations, they have no inherent size and look equally as good on any size monitor.

The key question is whether developers should move to MPEG I or stick with software. Most people agree that MPEG I playback looks better and it has the added advantage of compressing audio. However, faster CPUs and buses on multimedia PCs, low cost general purpose video acceleration cards and new software codecs - such as that from Intel, Indeo Interactive - may save software codecs from being replaced entirely by MPEG. On fast Pentium and Power PCs, Cinepak and Indeo can achieve 25 fps at 320 x 240 pixels. The window size can be blown up to 640 x 480 using new graphics accelerator cards often supplied as part of the PC. The quality is perhaps not quite so good as MPEG, but close, and is a lot cheaper for encoding and playback.

Selection of CODEC

There are three primary criteria to consider when selecting a codec:

- Compression level
- Quality of compressed video
- Compression/Decompression speed

As usual there are tradeoffs within these criteria. For example, one codec may produce high levels of compression but the quality of the resulting video is poor. Another codec may produce high quality video at high levels of compression, but a powerful computer with hardware acceleration is needed to playback the video in real-time. For further guidelines on choosing a codec and examples, refer to information later in this section and section 3: Video on Trial.

Other considerations for reducing file size

There are other techniques for reducing file size. So far we have considered those that involve compression; size of window, frame rates and quality settings. Others include:

- Key frames. Many codecs work by saving a whole frame of information and removing redundant information (eg I frames) followed by only the differences between successive frames. These whole frames are referred to as key frames. Obviously, the less frequent the key frames, the less information is kept and the smaller the resultant file size. The number of key frames per second of video is one of the parameters that is set prior to compression.

- Palettes. Reduce the number of colours to 8 bit (256). Although many of the codecs handle the palette in such a way that 24 bit video can be shown on 8 bit displays with no problems, reducing the video to 256 colours may reduce the file size by two thirds.

Tip: The considerations of window size, codec, frame rates, number of key frames and quality settings are not independent of each other and there are always tradeoffs and compromises to be made. Our advice is to experiment with combinations of the above at the start of any project involving digital video, having first established your audience, and the delivery machines available.

Delivery from CD-ROM

The table below shows the data transfer rates for different speed CD-ROM drives

CD-ROM drive	Data Transfer Rate	Notes
Single speed	150 kB s^{-1}	original based on CD audio
Dual speed	300 kB s^{-1}	relative to single speed
Quad speed	600 kB s^{-1}	relative to single speed

The above transfer rates, however, are only available if the central processing unit (CPU) is not occupied with other tasks. Playing video files from CDROM, that is, loading, decompressing and transferring to the video card, requires a lot of processing power and reduces the bandwidth available.

But by how much?

Unfortunately there are no exact answers. This is because of the number of factors :

- the amount of CPU attention required by a particular CDROM drive
- CPU type/speed, caching
- amount of free and type of RAM (EDO, etc.)
- presence and settings of any caching software
- operating system

When evaluating specifications of CDROM drives, the important parameters are the transfer rates at 60% and 40% CPU usage, not the 100% figures. This gives an indication on how the drive will perform when the CPU is busy doing something else like decompressing video.

How does this translate into actual figures?

CD-ROM	Data Transfer Rate	Likely Data Transfer Rates at 40 - 60% CPU Usage
Single speed	150 kB s^{-1}	120-150 kB s^{-1}
Dual speed	300 kB s^{-1}	210-230 kB s^{-1}
Quad speed	600 kB s^{-1}	300-350 kB s^{-1}

Note: within the case of single speed CD-ROM, 150 kB s^{-1} should be achievable by nearly all CPUs.

For double speed drives compression settings with data rates of 210-230kBs should be reasonable. Moving to a quad speed drive (600kBs^{-1} raw transfer rate), for the same specifications of computer as for a dual speed CD-ROM, the processor is now decompressing more data. Therefore 300-350kBs becomes the guideline range.

When compressing for delivery from CD-ROM and if not using MPEG-1, turn on the padding option, if available, in compression settings. Data is normally written to CD-ROM with a sector size of 2 kB. Padding adds dummy data to the frames to the nearest multiple of 2 kB thus ensuring that frames always start and stop at a sector boundary. Editing software provides default settings for each codec for key frame interval, compression quality, etc, and it is probably best to accept these until you are familiar enough with them to make changes.

Sound

When discussing digital video there is, almost always, the assumption that audio is included. Of course this is not always the case, there might not be any associated sound or it might not be important. Presented here are some of the issues involved with audio.

Any audio information is going to use some of the bandwidth available. The following table gives an indication of the data rates associated with some standard settings

Sampling frequency	Mono		Stereo	
	8 bit	16 bit	8 bit	16 bit
11.025	11 kB s^{-1}	22 kB s^{-1}	22 kB s^{-1}	44 kB s^{-1}
22.05	22 kB s^{-1}	44 kB s^{-1}	44 kB s^{-1}	88 kB s^{-1}
44.1	44 kB s^{-1}	88 kB s^{-1}	88 kB s^{-1}	176 kB s^{-1}

Notes:

- The figures are for sampling rate *not* frequency response, simplistically the sampling rate should be at least twice the maximum frequency of interest. For example, if frequencies of approximately 9 kHz are important, sample at 22 kHz not 11 kHz.

- The rates are for uncompressed data.

As can be seen from the table the audio data alone can (eg stereo, 16 bit at 44.1 kHz) take more bandwidth than is available from a single speed CDROM.

What are the options available to reduce data rates?

This will, obviously, depend very much on the importance of the audio information to your video. Taking the highest data rate (176 kB s^{-1}) as a starting point consider:

i) mono instead of stereo (reduce down to 88 kB s^{-1})
ii) 22 not 44 kHz sampling (reduce down to 44 kB s^{-1})
iii) reduce to 22 kB s^{-1} by choosing either:
  a) 11 kHz sampling
  b) 8 bit resolution
iv) or both a) and b) to reduce to 11 kB s^{-1}

As usual there are other ways of tackling the problem. Again one of the most useful tools at our disposal is compression. Depending on the nature of the audio signal, there are a number of ways of compressing the data. These algorithms attempt to take account of such factors as the way we hear and the large amounts of silence that can occur in speech detected by there being no signal, etc. Fortunately there are industry standards/organisations, such as the Interactive Multimedia Association (IMA), that tackle such issues as:

- cross-platform compatibility
- compression algorithms
- mismatches in recorded bit depth and playback capability
- sample rate conversion

For example, using the DVI® (Digital Video Interactive) algorithm, which consists of 4 bit ADPCM (Advanced Differential Pulse Coded Modulation) samples, we can achieve 44kHz, 16 bit (equivalent), stereo at a data rate of 44 kB s^{-1} (25% of the uncompressed rate).

Other tips and advice

- Cross platform compatibility. As is often the case, even with 'standards', there are cross-platform compatibility problems e.g. the AVI and QT IMA implementations are not 100% compatible; AVI does not support Apples MACE (Macintosh Audio Compression Engine) format. Therefore, if authoring for more than one platform, check which formats provide the necessary compatibility and/or be prepared to undertake some audio format conversion.

- Compression. Some of the audio formats use lossy compression techniques. Therefore you should appraise the format in relation to the importance of sound quality to your material. Audio compression can enable very useful reductions in data rates while maintaining good sound quality. This is an especially important consideration for delivery from CD-ROM.

- CD Quality audio. It is worth noting here that the 44 kHz, 16 bit stereo setting is often referred to as 'CD quality'. However, do not assume that this is what you will get; it will depend very much on the quality of the original material, recorded or live, and the equipment used for digitisation. Most standard sound cards generate an audible level of noise during recording which will be interpreted into the sound file; the inside of a computer is a very noisy environment in electronic terms. If high quality recording is important to you, consider evaluating some of the higher quality sounds cards.

- Sample at the highest quality possible. It is worth sampling at the highest quality you can during capture and refrain from taking any decisions regarding delivered quality when you come to perform the final compression. This allows you to retrace and try other alternative combinations without having to recapture the sound.

- Assessing audio quality. For assessing the sound it is sensible to have audio playback equipment which is of a reasonably high standard. Audio that sounded fine through the 'multimedia' speakers supplied with your system can be actually very noisy and quite distorted when played over a high quality sound system. Sound quality is subjective; test out the quality with potential users.

- 'Lip sync'. During assessments you may find that some subjects report that speech and lip movement seem to be badly synchronised; some people are more sensitive to this than others. This 'lipsync' problem can occur during either capture or editing. Its occurrence during capture arises out of timing skews between the sound and video capture drivers. The higher quality or all-in-one capture cards are less likely to suffer from this problem. During editing, when adding or deleting, be aware that you may be altering the synchronisation between the audio and video tracks. In either case re-synchronisation can be performed in the editor before compression. Read your editors help file/manual for application specific techniques/settings.

Considerations and Guidelines for Digital Video Capture

What do you need?

- A capture card
- A computer, with fast processor and lots of RAM
- Lots of hard disc space
- Sound card
- Tape drive for backup
- Editing software for video and sound (see further on in this section)
- Time - quite a bit
- Recording equipment for video sources if not already obtained

Choosing the system

If your budget allows, choose the best option in each category. At least purchase the highest quality capture card that you can afford assuming that you do not have to obtain the analogue video material first. The latter will carve a large hole in any budget.

Capture card

As a general rule, quality, speed and cost are directly proportional to one another. If the project demands top quality digital video, you will need a top quality capture card.

If your computer has a PCI bus (v2 or later) investigate the new cards that are currently being released for this bus. Current models include the FAST AV Master and the Miro DC. These cards are able to perform a lower compression for the initial capture (typical ratios of 6:1 for full frame video) due to the throughput of the PCI bus. This provides the potential for much higher quality which is especially important if the final output is to tape.

Tips:

- If possible, evaluate the capture card in your own computer. We have experienced a number of significant problems with a capture card purchased to run on a PCI (v2) bus Pentium.
- Check that the capture card is supplied with drivers for your particular operating system.
- Read the specifications carefully, that £10 Widgets Inc. WunderCap may say 30fps and 320x240 capture – but not both at the same time!

The more sophisticated cards are also capable of capturing both fields of an analogue signal. Each frame comprises two half frames of fields which are interleaved on screen. If tape is the destination for the finished video this function is important.

Some video capture cards (e.g. FAST AV Master) also incorporate sound capture and this can alleviate the ‘lipsync’ problems referred to in the previous section on sound.

Tip:
Having difficulties with your capture driver? Try another capture program. We did with our capture card:

- The proprietary capture software would connect to the card’s driver only very erratically
- Microsoft’s VidCap would connect most of the time, unless we had used the proprietary software first – then VidCap wouldn’t connect at all.
- Adobe Premiere would connect every time

The computer

PC and RAM: Buy the fastest computer you can afford and fill it with as much RAM as possible! Our initial video work was carried out using a 486 DX2/66 MHz with 32MB RAM. A two minute video sequence would sometimes take 4-5 hours to compress in software. This machine was upgraded to a Pentium 133 MHz with 32MB EDO (Exteneded Data Out) RAM. Compression is approximately 8-10 times faster. This class of computer is a good starting point. You can, of course, use a lower specification PC but you will have to wait a long time for the results. Most capture software allows video to be digitised directly to RAM which is much faster than capturing to hard disc which may cause frames to be dropped. Large amounts of memory are therefore useful even with slow systems.

This means that short sequences (eg 10seconds of 320 x 240 pixels x 24 bit colour at 25fps with 24MB free RAM) can be captured in this way without dropping frames.

These sequences can then be saved to disc and stitched together later. Note that free RAM does not mean swapfile which refers to a space used on hard discs as a substitute for RAM.

Graphics display card: A fast graphics display card is important. Desirable features include video acceleration, 24 bit colour display to enable editing, compression, etc., to be undertaken without palette problems.

Hard disc

Add as much as possible. The price of hard discs has recently fallen dramatically. SCSI devices have several advantages over IDE/EIDE devices, including the ability to daisy-chain them (up to seven devices per controller card). They are also generally faster, available in higher capacities and, if housed externally, can be swapped between computers. This is useful if the final video needs to be taken to a bureau for writing to CD-ROM. Fast-wide SCSI controllers and high rotational speed hard discs on a PCI bus currently have the highest throughput and are certainly worth considering.

The hard discs should preferably be of the AV (Audio visual) type. These hide the thermal re-calibration that all hard discs perform periodically which could lead to dropped frames.

The introduction of new types of removable media discs (eg Iomega Jaz and Syquest SyJet) with capacities of 1GB plus are likely to prove very useful for digital video activities. They have reasonable transfer rates (approximately 4MB per second for the SCSI devices) and media costs are also reasonable (approximately £65 per GB).

Sound card

A sound card will, of course, be needed to capture the sound component of any video sequence. Most sound cards and, increasingly, those built onto motherboards, will probably suffice in the majority of cases. However, if sound quality is a significant component, you may need to budget for a higher quality card. It is also worth considering a ‘proper’ amplifier and speakers to use for assessing sound quality. Speakers supplied with computer are not always adequate. If this is not possible, purchase a pair of good quality headphones. Although previously mentioned, this last point is worth repeating.

Tape Drive for Back up

Raw video files are large and valuable. It will probably be much easier to retrieve a file from tape than locate the original material for recapture. Obtain a fast, high capacity tape drive for backups – and use it!

Cameras/Playback/Recording Equipment

As with the capture card, the quality of digitised video is directly proportional to the quality of any camera or recording equipment. File sizes will often be smaller if high quality equipment is used. Sophisticated recording equipment also produces less noise. Noise will be interpreted by the codec as content and be included in the file. If noise is a problem, investigate the use of software clean-up filters for digital video which are beginning to appear on the market.

Other tips:

- Allow a maintenance budget for equipment. Dust, dirt, head wear, etc. will all tend to generate noisy video.
- If the capture sequence is difficult and requires a number of trial runs, work from a duplicate tape and save the master for the final digitisation. Tape wear is another source of noise and something to avoid incorporating on a master tape.
- Use fresh tape for the original recording and wind through to reduce the chance of tape drop out.
- If capturing a lot of sequences it is probably worth investing in computer controllable playback equipment. Make sure that suitable drivers are available for your capture software.
- Consider whether timecode accuracy is needed. If so, this will be a requirement for any recording and playback equipment purchased. Alternatively timecode can be generated after recording by a bureau.

Preparing to shoot the video

If you are in the position of being involved before the material is actually filmed there are a few points worth considering:

- Use plenty of light. Most cameras tend to become noisier at lower light levels
- Use a tripod (or 'steadycam'). A one pixel jump may represent a large change when taken over the whole image
- Have something simple, but not plain or detailed, as a backdrop. Noise is more easily 'seen' on a plain background. See also Hamilton *et al.* (1995).
- Try to avoid excessive panning and zooming
- For material which has already been shot, but not edited, ask the editor not to use the standard fade up from/down to black. The effect may represent a change in *every* pixel for the duration of the transition; the worst case being white to black or vice versa. This will cause difficulties for the codec.
- Consider the use of pre-compression clean-up software for old/noisy video.

Tip: Imagine the situation from the point of view of the codec. Little or no change in a scene means less interframe information and lower data rates (or higher quality for the same data rate). Anything which causes a change over a large area of the image will need to be stored in the final video file.

Video Capture and Editing

- Prepare the system:

  i. Keep the hard disc de-fragmented.
  ii. Find out if your caching software slows actual transfer rates and turn it off (at least for your capture drive), if it does.
  iii. Set up a pre-allocated capture file larger than you think you'll need (most capture applications support this).
  iv. Put this file on your fastest hard drive

- Capture at the highest quality you can. It is easier to reduce quality at the compression stage than to re-capture the video. Archive your raw data.

- Capture at fractional frame rates. There is some debate concerning the importance of capturing/compressing with fractional frame rates. The term is really a bit of a misnomer. It arises from the difference between American video rates (30fps) and European (25fps). As most systems struggle to play back quarter screen video at 30 fps it is often recommended to cut this rate (and the resulting file size) by the simple expedient of dropping (at capture or, preferably, compression but not playback) every other frame i.e. ending up with a 15fps video file. However, do this to a 25fps European video and we should get a fractional, 12.5 fps video (or 25 frames per 2 seconds). Unfortunately not all capture/editing software (e.g. the current version Adobe Premiere, v4.2) supports this 'fractional' frame rate. Some options are:

  i. Try and work at 25fps. This may work on fast computers with small frame sizes, but watch data rates of finally delivering form CDROM
  ii. Ignore 12.5fps and use 12 fps. However, at some point in each second, 2 adjacent frames need to be dropped or there is some messy interpolation to do (it may not show, depending on your material)
  iii. Find some software that captures/compresses at fractional rates
  iv. Join others in complaining to the software publishers about this U.S.centric view

- Editing software. Most, if not all, capture cards come with some form of video editing software in addition to capture software. Packages such as Adobe Premiere combine the two. Unless you have specialist requirements, push the supplied software to its limits before selecting an editor suitable for your needs. You will be in a stronger position to assess the strengths and weaknesses of the alternatives. All software can appear appropriate to your needs when demonstrated by an interested third party.

- Special Effects. Most editors come with a plethora of special effects. While there is no doubt about the use of these for advertising and promotional material, their use within educational software demands care. Note that special effects can demand valuable processor time during editing.

- Microsoft VidCap and VidEdit. Although no longer being sold or supported, Microsoft's VidCap and VidEdit (originally part of the Video for Windows Software Developers Kit) have a few advantages if you still have them. They are not as resource hungry as some applications and multiple copies of VidEdit can be run, facilitating the cutting and pasting of clips. VidCap supports fractional frame rates. If running Windows 95 and having difficulties running VidCap or VidEdit, contact Microsoft for the appropriate information. It can be found in their Knowledge Base, article Q131745.

- Choice of Codec.

  i. Editing software may not support all the features that are available in each codec.
  ii. Define the distribution medium and playback platform: know the bit stream the storage device can sustain and the total storage. available. For CD-ROM rates refer to the table under CD-ROM issues earlier in this section.
  iii. Check that the codec you wish to use is available on all the platforms you want your package to run on, or, there are comparable codecs that can be substituted.
  iv. Know the codec's ability to adapt the synchronised playback speed to the available hardware without user interference.
  v. Check the content of the video. Codecs differ according to type of content (talking heads, fast action scenes), whether shot outdoors or indoors, etc.
  vi. Reduce the colour depth to 256 colours if the codec supports this unless there is a good reason not to. This will also significantly reduce the file size and increase playback performance. An important consideration if using CD-ROM.

  vii. Consider developer issues. Is there enough time to use an asymmetrical codec, that is, one which takes longer to compress than to decompress? If you have not been involved with obtaining the video source, know its source and determine whether it has been previously compressed. Do not recompress already compressed video as this may cause artefacts.

- Data and key frame rates. Choose carefully. Experiment with the default settings provided for each codec. If a frame is dropped, the sequence may not pick up properly until the next key frame.

- Editing sound. Some of the video editing applications are able to edit sound as well. Most sound cards come with some form of editing software, but, depending on your requirements these may not be sufficiently powerful/flexible and you may have to budget (and evaluate) more specialist applications.

Authoring and Playback

- Artefacts. Any codec using lossy compression may introduce artefacts. The greater the compression the worse the effect will be. If the video is likely to be used in an analytical/diagnostic manner consider linking to an associated high quality still image or a shorter, less compressed, sequence. For further information on alternatives to digital video refer to section 1.

Tip: Do you need motion video?. Would the material be better served by other techniques e.g. still image sequences? ***Leave the video to do what it is good at. That is showing motion.***

- If the authoring software permits, place the top left corner of video windows on horizontal pixel co-ordinates that are divisible by 4. Some software will automatically align the video window. According to Microsoft's guidelines performance can be up to 50% worse for unaligned video.

- Codec support. As with capture/editing software, not all authoring tools will necessarily support every feature of every codec.

- Backgrounds. Nothing on the screen can be brighter than white. Using a white, or very light, background can make the video appear dim. Try a mid grey, or similar, background. Similarly a screen running at a resolution of 1024 x 768 pixels or higher is going to make even a 320 x 240 pixel video look comparatively small. Consider changing to to 800 x 600.

- Play back only at the resolution of the original file frame size. Scaling the video window is heavy on processing power at a time when there is not much to spare.

- Colour depth. Check your display card's colour depth against playback performance. An 8 bit setting is usually best, but increasingly, with the newer display cards, performance is better at 16 bit (64K colours). Obtain the most recent drivers for the cards.

- Other software. Try not to have other software running at the same time as running the application. Low resources/memory can reduce playback performance.

- Consider hardware assisted playback. However, this will mean equipping all users' machines with extra hardware.

- VIDTEST: a video and sound testing program from Microsoft is extremely informative on CD-ROM and hard disc transfer rates, CPU usage, etc. Do try and obtain a copy. Although is is not longer available from Microsoft and unsupported, we are attempting to seek permission to distribute the software via our World Wide Web site.

During our investigations into digital video for multimedia applications a series of experiments were performed to investigate the various parameters involved with video capture such as, choice of codec, frame rates, key frames, compression quality and size of video window. The results, a series of digital video clips, can be accessed and downloaded from the following World Wide Web site: http://www/.ets.bris.ac.uk. Further information expanding on the issues already discussed in this report is also available, as are links to other sites with relevant information.

Method:

A video sequence containing several scenes with differing content (scenes with little change, rapidly changing scenes, scenes with multiple content and special effects, etc) was captured 'raw' and using real-time compression at various frames rates and window sizes, and compressed using a variety of different quality settings and codecs. The sequence was captured from a non-degradable video source, an analogue write once laserdisc, to eliminate tape degradation effects and maintain quality. All sequences were captured on a PC. Two cards were used in the study:

- Creative Labs Videoblaster RT 300 - captures AVI and performs real-time Indeo compression
- Vitec Video NT - captures real-time MPEG and AVI

Results:

Table 1: Compression times and file sizes for various quality settings and codecs using MS VidEdit

Initial file: 30seconds of Intel Raw video, 25frames per second, 320 x 240, original file size 70.4MB. All files are at 24 bit colour except MSVideo 1 as this only supports 16 bit colour.

Output settings	Codec	conversion time (minutes)	size (MB)	output file
100% quality key frame every frame	MS Video 1	9:00	70.1	msvid_1.avi
	Indeo 3.2	10:45	13	ir32_1.avi
	Indeo Interactive	17:00	12.4	ir4_1.avi
	Cinepak	48:30	18.7	cnpk_1.avi
0% quality key frame every 10 frames	MS Video 1	2:50	0.61	msvid_2.avi
	Indeo 3.2	14:15	2.65	ir32_2.avi
	Indeo Interactive	43:45	1.79	ir4_2.avi
	Cinepak	33:10	18	cnpk_2.avi
Codec defaults for quality and key frame rate	MS Video 1	3:40	5.17	msvid_3.avi
	Indeo 3.2	14:05	2.89	ir32_3.avi
	Indeo Interactive	46:55	4.95	ir4_3.avi
	Cinepak	34:50	18	cnpk_3.avi

Comparing the conversions for MS VidEdit and Adobe Premiere there was no noticeable difference in either the time taken or the final file size. Adobe Premiere does afford the opportunity to output to the Apple Quicktime format, corresponding times and file sizes are given in the table below.

Table 2: Compression times and file sizes for various codecs at the highest quality setting using Adobe Premiere

All starting with the same initial file, 30seconds of Intel Raw video, 25fps capture, 320 x 240, original file size 70.4MB

Output settings	Codec	conversion time (minutes)	size (MB)	output file
quality 5 key frame every frame	MS Video 1 Indeo 3.2 Cinepak	6:35 11:40 52:05	62.8 12.6 18.2	video_1q.mov ir32_1q.mov cnpk_1q.mov

Note: At the time of writing the Quicktime format did not support the Indeo Interactive codec. Quicktime provides a quality setting range of 0-5.

Table 3: A comparison of file sizes versus frame size.

All files captured to RAM as Indeo Raw (via the Videoblaster RT 300) and recompressed at 25 fps, highest quality, key every frame using MS VidEdit

Frame size	**codec**	**file size (bytes)**
160x120	Indeo 3.2	4420298
	Indeo Interactive	4420814
192x144	Indeo 3.2	6181010
	Indeo Interactive	6516170
288x216	Indeo 3.2	11681090
	Indeo Interactive	11399920
320x240	Indeo 3.2	13280700
	Indeo Interactive	12658408
640x480*****	Indeo 3.2	1884426
	Indeo Interactive	3956890

***Note:**The low file sizes here are due to the computer and card being inable to capture at 640 x 480, captures were therefore carried out manually (i.e. frame by frame) and represent sequences of 100 frames (4 seconds).

Table 4: A comparison of file sizes versus frame rate

All files captured to RAM as Indeo Raw (via the Videoblaster RT 300) and recompressed at 320 x 240, highest quality, key every frame, using MS VidEdit

Frame rate	codec	file size (bytes)
25	Indeo 3.2 Indeo Interactive	13280700 12658408
15	Indeo 3.2 Indeo Interactive	7782486 7440252
12.5	Indeo 3.2 Indeo Interactive	6597990 6331812

Table 5: A comparison of file sizes versus codec using the Video NT card.

Initial file: Approximately 17.5 seconds of Vitec 422 video (Vitec's form of raw compression), 25 frames per second, 192 x 144, original file size 25 MB, recompressed at highest quality, key frame every frame, using Vitec editing software

codec	file size (bytes)	output file
Cinepak	3792598	cnpk.avi
MS Video 1	13669916	msvid.avi
Intel Indeo 3.2	3264318	ir32.avi
Intel Interactive	3441432	ir4.avi
Vitec DCT	4484472	vitec2.avi

Note: Vitec's own DCT (Discrete Cosine Transform) codec only gave us control over some of its settings – others were greyed out.

Table 6: MPEG files.

Comparison of file size and quality settings of MPEG files created from either Indeo Raw at two different frame sizes, or compressed AVI files, using MPEG compression software (MPEG maker) supplied with the Video NT card.

Frame size	Initial codec	file size (bytes)	output file	MPEG compressor settings
192x144 file from table 3	Intel Raw Intel Raw Intel Raw	611884 1713813 2474927	hi_cmprs.mpg quality.mpg quick.mpg	default high compression default high quality compression default fast compression
320x240 original file from table 1	Intel Raw Intel Raw Intel Raw	1223695 3250012 4823189	hi_cmprs.mpg quality.mpg quick.mpg	default high compression default high quality compression default fast compression
Recompressed using files from table 1	Cinepak MS Video Intel Indeo 3.2 Intel Interactive	4370476 4245941 3812733 4073239	ex_cnpk.mpg ex_msvid.mpg ex_ir32.mpg ex_ir4.mpg	default high quality compression default high quality compression default high quality compression default high quality compression

One sequence was captured directly to MPEG using the Vitec Video NT card using the ‘high quality’ setting but – with the stuttering playback, general poor quality and time constraints – we did not consider worthwhile continuing with this approach. The file is vitec.mpg

Initial Discussions:

The above tables reveal some interesting facts about the nature of the various codecs listed above. The nature of the relationship between quality level, compression and file size varies from one codec to the next. The balance of tradeoffs is dependent upon the video material (content and quality) and its purpose. However, assessment of the various codecs is a subjective process and the above results should be considered while viewing the files. Refer to the following World Wide Web site for further comparisons and information and to download and view the video files: http://www.ets.bris.ac.uk

Section 4: Scenarios: Examples of the use of Digital Video in Teaching and Research

Department of General Medicine, University of Bristol

Medical education is experiencing a shift in balance between hospital-based services and those provided in general practice and the community With NHS reforms, teaching in traditional tertiary care teaching hospitals is becoming difficult to sustain as the number of patients being admitted to hospital is decreasing and those that are admitted stay in hospital for a shorter time.

Within the department of general medicine, especially within the area of endocrinology, virtually all patients are now seen as out patients. This means that students do not gain experience of the range of cases that is required. A solution to the problem has been to video the patients during referrals for viewing at a later date. Patients visit the consultant as normal and are captured to videotape using a small, unobtrusive Hi-8 camera and microphone. The manner in which the video is captured is low-key and informal, and does not interfere with the referral procedure. This allows patient/consultant discussions to occur naturally without altering the dynamics of the consultation. The use of video in this scenario allows important visual considerations to become more apparent. Much of our everyday communication is non-verbal, and culturally we ‘read’ such things as facial expressions, hand movements, posture, etc. The video material offers a window onto the non-verbal communication and presents a holistic approach to the consultation.

“....visuals of some sort and variety are the main vehicle of expression and communication. Consider how influential visuals such as facial gestures and other body movements (usually referred to as nonverbal communication) are in face-to-face conversations and social interactions”

Rieber 1994

Patients are filmed over a period of time, before, during and after treatment. The video is digitised using a motion JPEG capture card and Apple’s Quicktime. A typical film from a single consultation uses 3 GB of disc space. The required sequences are edited and assembled in Adobe Premiere and exported back out to tape. A typical video is 8 minutes in length. No voice-over or narration is added as this may prove influential and detract from the naturalness of the consultation.

This scenario deals with the teaching of disease manifestations that are uncommon, with symptoms gradually becoming more apparent with time. To represent the passage of time, a technique known as “Morphing” is used. Morphing is a process which merges short sequences (or still images, graphics and other visual elements) into one another. Adobe Premiere has a number of special effects of which Morphing is one. Merging these sequences in quick succession has resulted in a powerful educational tool. Text and images alone could not have had such an impact, and this example highlights benefits of the immediacy of video when combined with techniques such as morphing.

Department of Professional Legal Studies (DPLS), University of Bristol

This project involves the creation of a portfolio of scenarios for training solicitors to react rapidly and appropriately under certain conditions. The scenarios include: i) interview of suspect at police station, ii) company annual general meeting and iii) briefing of expert witnesses. Traditionally these thinking skills have been taught by the use of role play but this approach is resource intensive. The use of digital video to assist in this process combined with an authoring package to allow for student interaction is near completion.

Users (trainee solicitors) are presented with the first video sequence from one of the scenarios, eg interviewing a suspect at the police station. As the sequence plays, users interrupt the video where they think an alternative course of action should have been taken by the solicitor. Users are invited to enter the alternative course of action and are then presented with a model answer against which they can score themselves.

The scenarios are filmed and edited using Betacam SP video systems. Two specifications of delivery platform were determined at the outset of the project:

i) A mixture of 486 DX2 66 MHz and Pentium PCs present within the department's computer assisted learning laboratory (CAL Lab)
ii) Dual speed CD-ROM

Initially the video was captured at 25 frames per second at 320 x 240 pixels using the VideoBlaster RT 300 capture card. The sequences were compressed in real-time with Indeo 3.2 and with a key frame every 10 frames. This decision was reached by digitising the first two sequences of the first scenario above at different window sizes, frame rates and key frames. Ideally, the video should have been captured raw. However, our PC and card at that time was not capable of raw capture at 25 frames per second. For the next two scenarios (annual general meeting and briefing of expert witnesses), the video will be captured raw using a Pentium 133 MHz with 32 MB RAM. By capturing raw video, the video files can be re-compressed and optimised for delivery from dual-speed CD-ROM at a later date.

In this scenario there is also the sound track to be considered. The sound was captured at 22 kHz, 16 bit, momo as this gave good sound quality at the available bandwidth. The sound track will be compressed when optimising for CD-ROM delivery.

Any soundtrack has to share the available bandwidth with the video and there are often tradeoffs between the quality of the video and sound track. This can lead to problems with lip sync with the audio track often lagging behind the video.

"An experienced TV viewer (i.e. most people) will immediately spot if they are in conflict. An obvious example is lip-sync sound. Deliver lip-sync video at 10 frames a second with audio lagging behind, and the disjointed look detracts from the quality of the communication. The viewer spends more time wondering what is wrong with the picture, than listening to the words".

Designing in usability, Audio Visual, April 1992

The article, however, goes on to say that a good interactive designer would have defined this issue before it became a problem by understanding the audience, the demands of the communications required and the constraints of the technology which was being selected. In short - lip sync would not have been used.

Evidence from formative evaluations of the DPLS project, however, showed that although users were initially put off by the sound being slightly out of sync with the video, the user's attention was soon detracted from this because of the quality and completeness of the information being conveyed. What this perhaps demonstrates, is how, with good instructional design, a deviation from the 'norm' may challenge our expectations, or notions. It is also important to realise that a generation brought up with television will have certain expectations when viewing digital video. However, television and computers should not be in competition with each other. Each provides a separate channel for communication and has its own set of issues and criteria. Viewing from television and computer screen are different: for the former video is viewed at a distance, the latter, close up.

The Sebastian Diamond Mother and Baby Sleep Unit, St. Michaels Hospital, Bristol

This unit is studying some of the factors in Sudden Infant Death (SID, also known as cot death). The study method involves simultaneously videotaping and recording the physiological data of both mother and baby. As this is an overnight study it requires 3/4 videotapes.

The studies monitor the interactions of a mother and baby sleeping in two ways:

i. together
ii. in separate bed/cot

The mother and baby stay overnight in a room that is equipped with an infra-red sensitive video camera (and infra-red light). Both mother and baby are connected to equipment that measures and records various physiological data (e.g. respiration; heart rate; temperature; blood gases). Currently the physiological data is recorded onto computer, while the video output, overlayed with the physiological traces, is recorded onto video tape. A typical study will use 4 x 195 minute tapes.

The physiological data (approx. 50-100MB) stored on hard disc can be easily accessed for fast searches, reviews and reporting. Physiological data is backed up to tape. The video data is reviewed and any movements by mother or baby coded. This takes approximately 15 hours, even using accelerated playback, for a typical 10-12 hour study. The video tapes have no backup.

There are a number of problems with the data being stored in this way:

- Raw physiological and video data are stored separately
- Retrieving physiological data from backup tapes is not easy
- Accessing a specific part of a specific tape is time consuming
- Tape storage
- Sharing the data, for teaching and research means duplicating video tapes and finding a suitable method for distributing the physiological data

We considered whether or not digitising the video signal could help with some of the above together with the possibility of storing all data on CD-ROM.

As a first step some questions were asked

Do we need all the data?

Performing some simple calculations. If the video is captured at 25fps, 320 x 240 resolution and 24 bit colour, this gives a file size of 5.76MB per second for uncompressed data. This equates to 250GB for 12 hours. Working back from the constraints of a single CDROM of 600MB for 12 hours (allowing 50MB for the physiology data) we have 14kB s^{-1} available; a factor of more than 400:1. This is not a realistic task for current codecs in real time.

Two immediate ways of reducing file size are:

- as the video signal is greyscale, 8 bit capture is sufficient
- at smaller frame sizes the overlayed physiology data is not easy to read and poses additional problems for the codec. Given that the physiological data is being recorded separately, it was decided not to capture with the physiological data overlayed.

Should the videotape record be continued?

It was decided to continue with the videotape record. This meant that there was no need to capture the sound as a full record would be available on the videotape. In addition this reduced the need for high quality digital video and, more importantly, the need for continous video was removed. This then opened up the possibilitiy of digitising frames at intervals, that is, time lapse recording

What interval and frame size should be chosen for the 14 kB per second limit?

The content of any video affects compressibility and therefore a safety margin was built in. It was decided to make every frame a key frame as this gave the ability to enter the video sequence at any point, without having to load the preceding key frame.

A frame size of 240 x 480 at 1 frame per second was decided on. It may turn out that too great a safety margin has been built in. It may therefore be possible to increase the frame size to 320 x 240 at a later date. This would probably prove more useful than increasing the frame to 2 frames per second.

Having taken these decisions the following benefits became apparent:

- As the same computer would be performing both the video capture and data acquisition, lowering the video data rate to such an extent meant that the PC would not be overloaded by data throughput. Jitter on the physiology data is unacceptable. At the lower frame rates this is no longer noticeable.

- Playing back the video at an acceleration of 10-20fps also accelerates the review of the nights activities by the same amount.

- It becomes possible to present overviews, e.g. displaying 10 frames taken at 60 frame intervals across the screen (at reduced resolution) would give a 10 minute overview. There is also the ability to zoom in or out in time and/or pixel resolution while showing the corresponding view of the physiology data.

Other considerations and requirements:

- The capture card has to be capable of capturing/saving the video straight into an AVI file which can then be played back without the capture card being present. Compressing 35,000 plus frames is a **very** time consuming task.

- The capture must be able to do interval (frame by fame; time lapse) capture. It should also yield gracefully to other processes, in this instance the physiology data capture. Check if the capture driver supports standard Video for Windows calls (or that the suppliers will release the API information) should you need custom capture software.

- To minimise head thrashing, two external hard discs (SCSI, in a shared case) are used; one for video, the other for the physiology data. These can then be unplugged and taken to a CDROM writing PC, where they will be written to the same CD.

Section 5: The Future

Digital video is a rapidly changing field. We report here on some of the changes that have taken place since the beginning of the project.

PCI bus capture cards:

These are beginning to appear on the market and include FAST AV Master and MicroVideo DC20. Both offer much better performance in terms of frame size, frame rate and lower initial compression, than most ISA cards which are severely restricted by the bus bandwidth available. Additionally, the FAST card, with its on-board sound capture, can eliminate lipsync problems. Current costs (ex. VAT) are approximately £750 for the DC20 and approximately £1000 for the AV Master. Note that both are PCI bus cards and that the FAST card requires a computer that supports PCI bus masters.

Codecs

New codecs for Video for Windows and Quicktime from Intel: Indeo Interactive and from Horizon: Power!Video. Indeo Interactive has many features, the most important of which is its ability to scale the quality of the video depending on the capability of the computer's CPU. This will result in smoother motion videos during playback.

Iterated Systems, known for their fractal still image compression software and FIF file format, should be releasing Q3-Q4 1996, a fractal-based video codec ClearVideo. This codec will analyze the image for structures that are reducible to fractal equations. Claims are that it will give acceptable video at 15 frames per second with data rates in the tens of kilobytes per second (compared to the hundreds of kilobytes for standard codecs). It will use standard algorithms for the sound component.

Storage

We have already mentioned elsewhere the high capacity removable cartridge drives, but the high capacity 'floppy' is being resurrected again. Ultra SCSI cards are beginning to make an appearance, with transfer rates up to 40MB per second. Ultra SCSI will make moving and editing large video files a lot faster.

Built in support for multimedia

For those who use Intel processor based computers, the built-in support for multimedia-type operations in their next generation of processors should offer an all round performance increase without needing to add specialist hardware. However this will need additional support in the operating system.

The future of digital video

A new development which has already arrived; the wait for one piece of hardware to become more common; and the need for some software to be written. When the three components come together there is likely to be a huge impact on digital video production.

The component that is already here is the digital video camcorder. The camera uses a new format tape called the Digital Video Cassette (DVC). The first ‘prosumer’ cameras (from Sony and Panasonic) arrived on the market place at the end of 1995. Although much more expensive (£2.5k - £3.5k) than a standard camcorder, it offers quality comparable to Betacam. The tape format offers, amongst other things; 500 line resolution, time codes and 16 bit stereo audio (similar to DAT). Not all digital video camcorders may make full use of all DVC features. The tape format is digital, a form of MPEG2. Error correction mechanisms are built into the format meaning a much greater freedom from such defects as tape drop-out and generation loss. The Sony camcorders come equipped with Firewire (also known as P1394 or Apple’s serial SCSI). Connected to a computer with the Firewire interface, the contents of the DVC can be loaded onto hard disc without having to be digitised via a capture card. The Firewire hardware is the second link in the chain and the one that needs to become more common place.

The final link is the software for editing DVC format video. One thing we can be fairly certain of is that there are a few companies with products in development. Software codecs may well be fast enough for quarter screen playback, but hardware codecs would help the editing process. Once all this arrives we will have a video production system where the digitisation takes place just once, early on in the process at the time of shooting, and can remain in the digital domain until presented to the viewer. This should enable the quick production of high quality video, lower overall cost and without the difficulties associated with some video digitiser cards.

The future means:

- Unless digitising other sources there is no need for tape decks, capture cards, backup tape drives
- Providing you have enough disc space (this is no longer expensive) download the whole of a DVC and edit non-linearly. No awkward captures from tape
- The DVC itself should provide a robust archive format
- A very fast computer will no longer be a pre-requisite for video capture, the camera has done that job

But who knows……

References

Barker, P. (1989). Multi-Media Computer Assisted Learning. Kogan Page Ltd, London

Buzan, T. (1993). The Mind Map Book. BBC Books, London

Cotton, B.& Oliver, R. (1993). Understanding Hypermedia. Phaidon Press Ltd., London

Erdelyi, M.H. and Stein, J.B. (1981). Recognition hypernmesia: the growth of recognition memory over time with repeated testings. Cognition, **9**: 21-33.

Gage, N.L. and Berliner, D.C. (1988). Education Psychology, Houghton Mifflin Company, Boston. pp 279-310.

Haber, R.N. (1970). 'How we remember what we see'. Scientific American, **105**

Hamilton, N.M., Frade, I., Duguid, K.P., Furnace, J. & Kindley, A.D. (1995). Digital Video for Networked CAL Delivery. Journal of Audio Visual Media in Medicine, **18:** 59-63

Levin, J., Anglin, G., & Carney, R. (1987). On empirically validating functions of pictures in prose. In The Pyscology of Illustration, volume 1: Instructional Issues, pp 51-85. Eds: D. Willows & H. Houghton. Springer-Verlag, New York

Levin, J., & Lesgold, A. (1978). On pictures in prose. Educational Communications and Technology Journal, **26:** 233-234

Mayer, R.E. & Anderson, R. B (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. Journal of Educational Psychology, **19:** 30-42

Miller, G. A. (1956). The magic number seven, plus or minus two. Psychological Review, **63**: 81-97.

Paivio, A. (1975). Imagery and Long-term Memory. In: Studies in Long Term Memory. Eds. Kennedy, A and Wilkes, A. John Wiley and Sons, London.

Reiber, L. P. (1994). Computers, Graphics, & Learning. Brown & Benchmark, Madison, Wisconsin, Dubuque, Iowa. ISBN 0-697-14894-7

Standing, L. (1973). Learning 10,000 pictures. Quarterly Journal of Experimental Psychology, **25**: 207-222.

Williams, P.J., Lock, A., Crisp, J.C. & Longstaffe, J.A. (1995). The Use and Capture of Images for Computer Based Learning, Second Edition. SIMA Report Series No 10, ISSN 1356-5370

Usability in Design. Article in Audio Visual, **244:** 51. April 1992