

MPEG: The Big Squeeze

March 1, 1996

By: Jeanette DePatie
Brian Quandt
Brian Roy
Catherine Saccente

This material was written expressly by and for HEURIS/Pulitzer. It may not be copied or duplicated in any way, whole or in part without the express, prior, written consent of HEURIS/Pulitzer or its agents.

Compliments of HEURIS/Pulitzer
314.534.1514

[This page is blank]

Table of Contents

PREFACE	5
CHAPTER 1--WHAT IS MPEG?	7
CHAPTER 2--MAKING MPEG WORK FOR YOU	9
CHAPTER 3--HOW DOES MPEG WORK?	20
FOR FURTHER READING	32
APPENDIX A -- SOME OTHER VIDEO STANDARDS	33
APPENDIX B -- THE ISO ORGANIZATION	36
APPENDIX C--SUMMARY OF THE CONSTRAINED VIDEO PARAMETERS	38
GLOSSARY OF TERMS	40

[THIS PAGE IS BLANK]

MPEG: The Big Squeeze

By now, you have probably heard the term MPEG--talking about MPEG these days is equivalent to technological “name dropping”. Unfortunately, a lot of the current talk about MPEG has resulted in confusion and misinformation.

You may be wondering why you should even bother to learn about MPEG. What makes it better than any other form of video compression? There are at least three ways in which MPEG is superior to other types of video compression:

1. **Image Quality:** MPEG is the only currently standardized, cost-effective way to get one hour of VHS quality video on a CD-ROM.
2. **Compatibility:** MPEG was designed as an international, open standard in order to ensure compatibility. MPEG gives you more choices than other video compression schemes, especially when it comes to playback.
3. **Compression Ratio:** MPEG-1 frequently offers data compression ratios of up to 200:1, yet still offers VHS quality playback. You won't get THAT anywhere else!

MPEG opens up whole new worlds in multimedia. You can encode full length movies to compact discs, create interactive video training applications, develop kiosks, send video over phone lines, and do a whole host of other projects. MPEG can be a great thing--if you know what you're doing.

In the spirit of all good compression technologists, we have tried to distill the existing factual information about MPEG to just a few pages. This little booklet will help you cut through the stuff and nonsense written and spoken about MPEG, and give you enough basic information to embark on a project of your own. Please feel free to call us at 314.534.1514 if you have questions. You can also visit our web site at <http://www.heuris.com> any time day or night.

[This page is blank]

Chapter 1--What is MPEG?

MPEG is an acronym for **M**oving **P**ictures **E**xperts **G**roup which commonly refers to the international standard for digital video and audio compression. The official name of the MPEG-1 standard is: “Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Megabits per second.” It is sometimes referred to by its ISO/IEC project number, 11172 parts 1 through 5. However, this video standard is usually just called “MPEG”.

Just as Norton Utilities is a collection of tools for maintaining computer hard drives (and more), MPEG is a collection of tools for compressing audio and video. An important note about MPEG is that it does not specify how to perform compression. It does, however, describe a set of minimum requirements that the MPEG decoder must live up to. (An MPEG decoder is the device which plays back the compressed audio and video.) In particular, it defines a fictitious MPEG decoder, that incorporates the minimum requirements which determine whether something is MPEG or not. This theoretical decoder is called the STD (Standard Target Decoder) model:

- If a product is a MPEG decoder, then it must minimally have the same functionality as the STD model.
- If a product is a MPEG encoder, then any streams it generates must be decodable by the STD model.

MPEG: A TRUE INTERNATIONAL STANDARD

Unlike DVI, TrueMotion-S, and Indeo, MPEG is an official international standard. (For a discussion of these and other video compression methods, please see Appendix A.) MPEG was developed as a joint project under the long committee description of ISO/IEC JTC1 SC29/WG11. This committee is made up of representatives from hundreds of companies throughout the world. The representative companies are very diverse, ranging from one-man consulting companies to extremely large international corporations like Philips. These companies contributed research and resources to cooperatively design a common standard method for the compression of sound and video.

To a newcomer attending a MPEG meeting, the international standard making process looks slow and cumbersome. But the committee faces many challenges both technical and political. To come to a global consensus on any technical issue takes time, and lots of it.

The general development cycle of an MPEG Standard is as follows:

1. Ideas are suggested and researched by individuals or organizations.
2. These ideas are presented to the MPEG Committee, which may form an Ad-Hoc Committee to judge the merit of the idea and to propose a recommendation to the entire MPEG Committee.
3. Proposed wordings of the idea are established and then presented to the MPEG Committee in the form of a ballot.
4. The idea is added to the MPEG Standard based upon standard voting practices.

All additions, deletions, and modifications to the particular MPEG standard proceed similarly.

One of the political challenges to the committees is that many organizations have ideas already fully developed and have actually created computer chips before agreement on a standard has been reached. Naturally, these organizations would benefit from having their ideas incorporated into an MPEG Standard. Any organization that has its ideas incorporated into a standard may benefit in many ways, including:

- Revenue from patents and royalties.
- Fame and glory, at least in the technical world.

- The ability to bring a product to market faster than anyone else, since the organization already understands some of the key technical material.

For a more complete description of the structure and organization of the ISO, please see appendix B.

MPEG-1 AND MPEG-2

There has been a lot of confusion in the industry and the media about the differences between MPEG-1 and MPEG-2. Contrary to what many people think, MPEG-1 and MPEG-2 are not competitors to one another, and MPEG-2 is not an improved version of MPEG-1.

MPEG-1 was, in fact, designed specifically for delivering video from a single speed CD-ROM drive or T1 data stream. MPEG-2 is a completely different standard from MPEG-1, and is designed for different purposes. MPEG-2 is specifically targeted at digital transmission or broadcast of video signals, and supports a much wider range of resolutions and bitrates than the MPEG-1 standard. MPEG-2 has been chosen as the standard on which US HDTV systems will be based.

Chapter 2--Making MPEG Work for You

At this point, we'll turn to a discussion of the nuts and bolts of your MPEG project. For further technical discussion about MPEG please see Chapter 3--How Does MPEG Work?

HOW TO SCHEDULE YOUR PROJECT

There are two rules to keep in mind when scheduling a project with MPEG encoded video:

- 1) These projects always take twice as long as you think.
- 2) The “final” 10% of the project always takes 90% of the time.

Many developers fall into the “almost done” trap. When it comes to multimedia projects, you must plan, plan, PLAN. Many different people with many different schedules must be coordinated together. Make sure that the people, equipment, and budget are available for each step of the project as you need them.

Some of the basic steps to keep in mind are:

X	Item
	Choose a visual style. Your video and interface should work together in an aesthetically unified fashion.
	Write a script. This script will tell the software designer what you want the buttons to do, and where they should go. This script will also provide the video director with materials to create the shot-list.
	Develop a storyboard / shot list. Develop a sketch of how you want each scene to look, what furniture, props, and people you need, as well as information about sound and lighting.
	Pick the music. Don't save this for last as it will have an impact on everything in the production, including the pacing and line readings of the “live performers” as well as the look and feel of the video post-production.
	Shoot the video. Leave plenty of time for lots of takes!

	Edit the video . Depending on how well your shots were planned, expect some time and expense in the editing suite.
	Encode the video. Remember, there are long waiting lines at many encoding service bureaus. Six weeks is common. Make sure contact is made with the encoding service early in the project and schedule appropriately.
	Develop your multimedia application. You can use commercially available authoring tools or write your own.
	Incorporate the encoded video into the application.
	Finish the software. Don't forget to test it!
	Cut the master Compact Disc . Again, watch for waiting lists!
	Replicate the Compact Disc.
	Print jackets for jewel cases. (The packaging that accompanies the CD-ROM.)

These are just a few of the major steps involved. Remember, better planning means: *lower costs, less time spent, greater efficiency, and greater control over creative aspects of the project.* Just get out a big piece of paper and start drawing up your plans (flow charts).

WHAT YOU SHOULD KNOW BEFORE YOU SHOOT YOUR VIDEO

DO opt for a component video format if available.

AVOID converting your video to or from a composite format at any time during production or post-production....

BECAUSE you will suffer an irreversible quality loss and potentially introduce artifacts that will stick with you all the way to the finished product.

DO use high quality, first generation video.

AVOID using second or third generation video.....

BECAUSE the higher the quality you start with, the higher the quality of the end result. High quality video is often less "noisy". Since MPEG cannot distinguish between moving video and "noise", it will attempt to encode the noise, taking bits and quality away from your moving video.

DO use nice big fonts.

AVOID MPEG encoding text over moving video.....

BECAUSE text is high frequency video data. The moving video in the background will cause your foreground text to fade in and out. In addition, the text uses lots of bits that could be allocated to making your video look better.

DO use animation with medium amounts of detail and lines which are several pixels thick.

AVOID using computer-rendered animation with extremely fine lines (less than 3 pixels) or extremely fine detail...

BECAUSE extremely thin lines and fine details tend to “disappear” due to MPEG’s lower resolutions.

DO use fast moving video with tightly focused close-ups.

AVOID using fast moving video where background and foreground are both highly detailed and in-focus....

BECAUSE when background and foreground are both in focus, they vie with each other for bit allocation--both will require a lot of bits. This can lead to “blockiness” or “pixelization.”

DO use talking heads in video; preferably a tightly focused close-up on the face.

AVOID using talking heads that are too small.....

BECAUSE when characters on the screen talk, the viewer’s focus is drawn to the mouth. If the mouth is too small, it will not be clear and will

not look “natural”. This can be distracting.

DO use computer or hand-drawn animation.

AVOID using computer or hand-drawn animation with very sharp diagonal or vertical lines....

BECAUSE this can lead to “aliasing” which makes smooth lines look like stairsteps.

DO use scene changes and relatively quick cuts.

AVOID using extremely fast scene changes that comprise less than 2 frames or blinking or flashing screens...

BECAUSE very rapidly blinking screens and rapid scene changes are difficult for MPEG encoders to handle. Encoders need to work over multiple frames in order to achieve optimal compression.

DO use video with contrast.

AVOID using high contrasts in luminance, i.e. flames, explosions, fireworks, etc.

BECAUSE high contrasts lead to blockiness.

DO use video with lots of colors.

AVOID using monochrome scenes.....

BECAUSE while the resolution levels MPEG can handle are lower than computers, the number of colors MPEG is capable of is very high. So, if you can “say it” with color rather than “cross-hatching,” by all means do so.

These guidelines are not meant to suggest that there are “hard and fast” rules for MPEG encoding.

These are observations made from experience in doing real MPEG projects. A good MPEG encoding service bureau will be able to make most of the things in the “avoid” category look pretty good. However, if you are looking for optimum video quality, it is best to follow these guidelines as much as possible.

HOW TO JUDGE MPEG QUALITY

Image quality is subjective at best. What looks good to one person may not look good to another. This is especially frustrating when you are trying to decide which MPEG encoding house to go with. However, you can educate yourself as to what to look for in MPEG encoded material.

First of all, compare apples to apples. MPEG has a difficult time handling lots of fast motion with detailed backgrounds, areas of highly contrasting light intensity, (explosions, fireworks, lightning, etc.), and (believe it or not) simple 2-D animation sequences. Try to compare demos which display some of these difficult scenes. Just about anybody can make flowers blowing gently on the breeze or a duck gliding over the water look good.

Next--get close. All MPEG encoding looks the same from 20 feet away. Optimal viewing distance for MPEG on a standard size computer monitor is 5 feet away, at about eye level.

Finally, turn down the sound. The sound can have a strong effect on your perception of the video quality. If you're really trying to level the playing field, turn off the sound.

MPEG encoding has a host of potential quality problems all its own. Special things to look out for include:

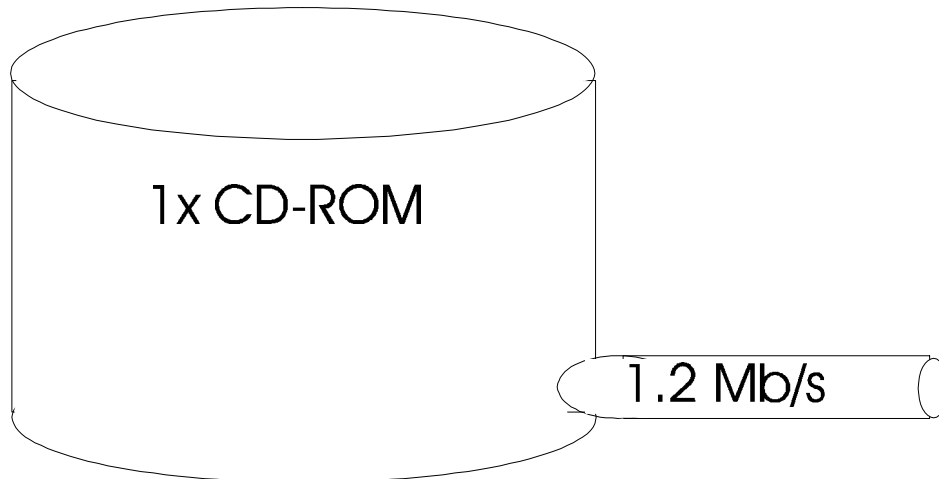
- **Blockiness:** when your picture breaks up into little squares. Especially noticeable in fast moving highly detailed sequences, and sequences with high contrasts in light intensity like explosions and fireworks.
- **Aliasing:** when lines that are supposed to be straight (especially diagonal ones) look like little “stairsteps”. Not necessarily indicative of bad encoding, but aliasing may be reduced by good encoding or extra image processing.
- **Fuzz and snow:** images that look as though your monitor is dirty or you lost a contact lens. Little gray or white flecks that intrude randomly throughout the picture.
- **Worms:** crawling dots and squirming lines. Probably the result of low quality video or bad digitizing.
- **Halos:** small area of distortion surrounding the outline of moving objects.

COMPACT DISCS: FORMATS, BIT RATES, AND COMPATIBILITY

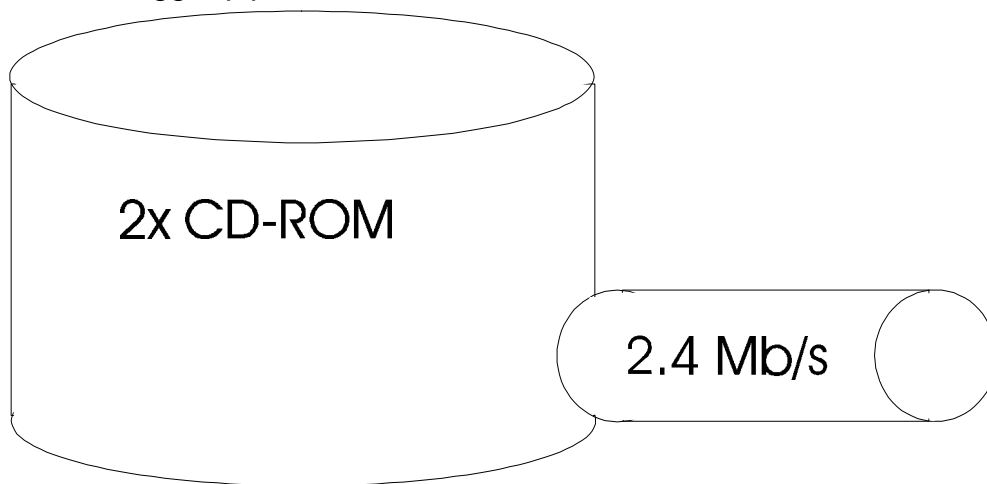
You should know BEFORE you start encoding your video which MPEG playback device you plan to use or be compatible with. Not all MPEG playback boards can play back all MPEG encoded streams, and not all MPEG encoded materials are compatible. The most common format for MPEG CDs (especially movies) is the "White Book" Standard. This standard, developed and licensed by Philips, helps to define how video information should be formatted, burned into, and read from a CD. MPEG encoding houses do not automatically provide "White Book" CDs. If the MPEG encoding house can provide them, they will probably charge extra for the service. (The systems used to develop White Book discs are proprietary and expensive.) Some MPEG playback boards cannot play back "White Book" CDs. Check the capabilities of your machine BEFORE encoding.

In addition to White Book Video CDs, CDs can be in ISO 9660 format. However, if both an ISO 9660 disc and a Video CD are encoded for a single speed CD player, the Video CD is usually encoded at a slightly higher bit rate and therefore looks a little better. This is due to the way the two different formats handle sector sizes. ISO 9660 discs are usually formatted for playback on a double speed CD-ROM drive or faster, and may be encoded at bit-rates higher than 1.2 Megabits per second.

Now, a word or two about bit-rates. The "bit-rate" describes how much material (how many bits) can be moved from a CD-ROM or hard drive in a given amount of time (usually per second). The standard bit-rate for a single-speed CD-ROM is 1.2 Megabits per second. That means you can transfer 1.2 Megabits of data from the CD-ROM in one second. Sometimes it's helpful to think of "bit rate" as a pipeline coming from a water tank. Here's the common bit rate for a single-speed CD-ROM:



Now a higher bit rate (or larger pipe) allows you to get more information off the drive or CD (or more water out of the tank) at once. Since you can move the information faster, you don't need to compress it as much. This means that, usually, higher bit rates can produce higher quality video images. However, since each minute of video isn't compressed as much, it takes up more room on the CD or hard drive. A CD encoded at a higher bit rate can contain fewer minutes of finished video than one encoded at a lower bit-rate. What's more, in order to move the data off a CD-ROM faster, you have to spin it faster. That means, in order to play MPEG sequences encoded at bit-rates above 1.2 Megabits per second/s, you need to have a double or triple speed CD-ROM drive or a bigger pipe:



Not all playback machines or CD-ROM drives are capable of playing all bit rates. What's more, the higher bit-rates have both advantages and disadvantages. What bit-rate you choose to encode at depends on many factors, including:

- Can my chosen playback unit play at higher bit-rates?

- Do I have, or wish to be compatible with single, double, or triple speed CD-ROM drives?
- How important is image quality to my overall project?
- How much material do I need to fit on a single CD-ROM, or do I plan to play from the hard drive?
- How much money do I have to spend?
- Can my encoding service bureau provide encoding at different bit-rates?

WHAT TO LOOK FOR IN AN MPEG PLAYER

The first thing to decide is whether to use an MPEG playback board in a computer, or a multiplayer.

Playback boards work on PCs or Macs and cost between \$200 and \$1,000 depending on video quality and extra features. These work with the CD-ROM drive to play MPEG encoded video. A NOTE OF CAUTION--not all CD-ROM drives work with all playback boards. Make sure to check the product literature to ensure compatibility before purchasing. Also, playback boards often work BETTER with some CD-ROM drives than others. If you are unsure, call the technical support lines for your board.

A multiplayer is a stand alone unit with a built in MPEG playback board and CD player (The Philips CD-I player for example). These players often play audio CDs, Video CDs, and Photo CDs. The prices range from about \$500-\$1000. These units are less expensive than a computer with an MPEG playback board and make ideal engines for kiosks. However, there are some technological limitations to these machines and development is usually more expensive.

Which machine is best is in large part dependent on the application. Different degrees of interactivity and general usage will place different demands on the playback device (and the development budget!).

Things to consider include:

- Does it play back only in a window on the screen, or does it have full-screen capabilities?
- Does it provide line out stereo audio?
- What sort of video output is provided? Is it S-Video, composite, or component? Does it require a special monitor?
- At what bit-rates will it play MPEG sequences?
- How is it controlled? Does it offer "VCR" type controls? Can one select and play an individual track? Does it offer full device driver support?
- What sort of customer support is provided?
- How does video played on this machine compare to others? (Be careful to compare apples to apples. If possible, take along a reference CD-ROM and look at it on each machine. Please see the earlier discussion of judging MPEG quality.)
- Does it play video CDs?
- What does it cost?

If you require only the ability to continuously replay a video sequence without deterioration you have the most options to choose from. The easiest option is selecting one of several VideoCD multiplayers probably stocked by your local consumer electronics stores. Then select an MPEG house that can provide White Book specification VideoCD's.

If you or your software development company have a favorite authoring system like Macromedia Director for multimedia development, your choices may be limited. Find out whether the MPEG playback system you have chosen works with the authoring system you prefer.

If you are providing local movie or advertising distribution systems, you may want to consider one of the new boards that incorporates high end special effects hardware for video work directly into the playback board. Alternatively, if every client site has a decoder, you may want an inexpensive board that will merely accept MPEG sequences over a network and display full screen video.

WHAT CAN I EXPECT TO PAY FOR MPEG ENCODING?

The prices vary widely, ranging from \$30/finished minute all the way up to \$300/finished minute and beyond. You're probably wondering why the prices differ up to a factor of 10. In the world of MPEG encoding, you get what you pay for. Some

characteristics which affect the price include:

- Image quality: This is the most important factor. Don't expect the \$30/minute stuff to look as good as the \$300/minute variety.
- Turnaround time: Generally speaking, the faster you need it, the more you'll pay. Even some of the median priced companies have waiting lists of up to two months. It's good to keep that in mind when you're planning too.
- Level of customization: The more specialized the services you require--higher bit rates, special filtering, special platforms, etc.--the more you can expect to pay.

CAUTION: WATCH OUT FOR HIDDEN CHARGES!

Often the quoted price per minute is misleading. Some encoding bureaus charge extra for things like:

- Set-up
- Sequence creation
- Output media
- Sequence charges
- Shipping
- Filtering
- Multiple System Layer files
- Non-encoding machine time
- Minimum order sizes
- Multiple audio channels

Depending on the size and scope of your project, these "extras" can double the price of your job. Be certain to ask the encoding bureau about these extra charges.

WHY DON'T I JUST BUY A MACHINE AND DO IT MYSELF?

For some people, buying a hardware MPEG encoding solution and doing it in-house may be the best solution. However, there are a lot of factors to weigh in making this decision. Advantages to doing it yourself include:

- If you do A LOT of encoding, it may work out to be less expensive in the long run.
- You can better control the scheduling of the project. Some MPEG encoding service bureaus have long waiting lists.

If you DO decide to buy an encoding machine, make sure to see it in action. Also, make sure that some one-on-one instruction and reliable technical support are available. Some companies will try to tell you their product is “plug and play”. Most are not. You don't want to wind up two days from a major product deadline with a machine that doesn't work, no one to answer your questions, and no one to blame but yourself.

There are many advantages to going with an MPEG encoding service bureau. These include:

- You only pay for costs which you can pass on to your final customer. If you buy a machine, you must schedule enough jobs to pay for that machine and its upkeep.
- If you don't like one service, you can easily switch to another. If

you buy a machine, and major technological breakthroughs occur, you will likely be left with obsolete equipment.

- In the short run, you can get projects out more quickly. It takes a lot of time to develop expertise in running MPEG encoding hardware--plan on at least 6 months.
- Some MPEG service bureaus have years of experience behind them. They know how to get the MOST out of your video.
- Your MPEG service bureau may be able to offer capabilities that your machine doesn't have.
- Your MPEG service bureau may be able to offer advice, help coordinate your project, and answer technical questions.

Developing a project in MPEG may seem like a nightmarish task. It can be. However, if you plan well in advance, and go into the project armed with information, you can make your clients look great, and then you look like a pro. You can go from talking about “new media” to producing it! We hope that this informational booklet helps to answer some of your questions. If you need additional information about MPEG encoding services, give us a call at 314.534.1514. We'd like to hear from you.

[This page is blank]

Chapter 3--How does MPEG Work?

To keep this discussion relatively short, focus will be on the video aspects of MPEG compression. In particular, this discussion will talk about MPEG-1 video compression, or video compression targeted at playback from computer hardware such as CD-ROMs. However, it is first necessary to understand what video is, and more importantly, what digitized video is.

THREE IMPORTANT CHARACTERISTICS OF VIDEO

Video is simply an electronic sequence of still images displayed or projected (quickly) in succession to one another. As a result, the human mind is fooled into believing that people or objects in the presented sequence move. In terms of computers, there are three important characteristics of video:

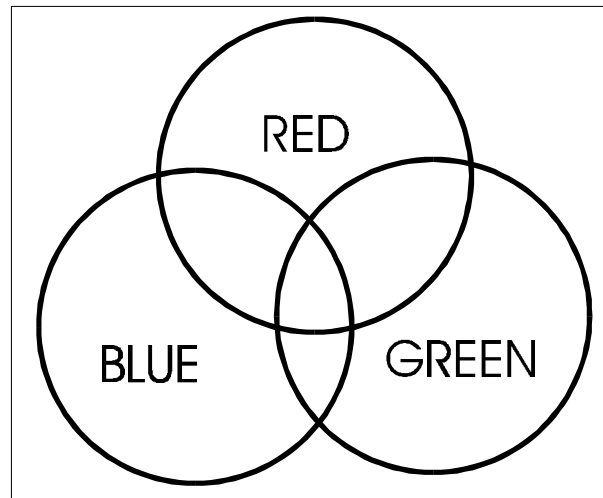
- How fast each picture is displayed (frame rate).
- How many elements create each picture in both the horizontal and vertical dimensions (frame size). This is normally given in terms of pixels.
- How many different colors the picture/pixel is made from.

Frame rate is the number of frames that are displayed to a viewer each second. For example, in motion picture film in the United States it is common to display 24 frames each second. In color television for the US home (called NTSC) 29.97 frames a second are displayed. Even though computers are not normally thought of in terms of *frame rates*, most computers “refresh” the screen by repainting every element of the screen as often as 72 times a second.

Frame size or number of picture elements is the next component of video. This is measured horizontally and vertically in pixels. “Pixels” are picture elements -- the small dots which make up the displayed picture. Some common dimensions, or resolutions numbers, in the computer world include: 640 horizontal pixels x 480 vertical pixels, 1024 horizontal x 768 vertical, and 800 horizontal x 600 vertical pixels.

The number of colors which make up each picture or frame is a third component of video. As is the case with a painter’s palette, a color can be described in terms of several “primary” colors. For instance, when playing with paints as a child, mixing equal parts of red, yellow and blue created black. By mixing these primary colors in different combinations, it is possible to produce

any other color. Color mixing works a bit differently with light than with paints, but we can still make any color from three primaries. In the video world, however, we substitute green for yellow in our “primary” color palette.



A Video Color Palette

In mixing colors of light, we vary the amount of red, green and blue light that makes up the color of a pixel. To make video practical, it is necessary to limit the number of differing shades of red, green, or blue that can be generated. This puts an upper limit on the total number of colors that video can recreate. Here is an example of a common digital color scheme. Each primary color (red, green, or blue) may have 256 different levels or shades. Since a color may be composed of the three primaries, this means we can generate 16.8 million different colors, or 256 levels of red times 256 levels of green times 256 levels of blue (16.8 million roughly equals $256 \times 256 \times 256$). The color for a pixel is normally written as follows:

pixel_color = (red_level, blue_level, green_level).

WHAT IS 24 BIT VIDEO?

The previous example just described 24 bit video, without calling it that. The term 24 bit comes from the fact that 256 shades of the primaries may be represented as an 8 bit value. Since it takes three primaries to represent a single value it takes $8+8+8$ or 24 bits to represent color for a single pixel:

8bits red 8bits green 8bits blue
RRRRRRRRGGGGGGGGBBBBBB
BB = 24 bits

As a final note about color and video, it is possible to choose different primaries or entirely different colorspace/colorsystems. The way we described colors above is not the only way to identify colors. Different “colorspaces,” or methods of describing colors have different uses. For example the common colorspace for printing is CMYK, or cyan, magenta, yellow, and black.

Another colorspace is YCrCb, or luminance (shade intensity) and

chrominance-red and chrominance-blue (chrominance components define the hue and value of the color). This last colorspace is commonly used in video, primarily because it more closely resembles the colorspace of human eyes, where rods detect luminance components and cones detect the chrominance components of color.

Now that we have discussed frame rates, frame sizes and touched the surface of color and colorspace, let's talk about MPEG's definitions of these values. For the purposes of this discussion, let's assume our video source is a typical professional digital video format called ITU-T 601 (formerly known as CCIR 601). In this format, we see the video is represented in the following fashion:

- frame rate of 30 frames a second
- picture size of one frame 720x480 (NTSC)
- color and colorspace: YCrCb 4:2:2

Luminance is sampled at full resolution; each chrominance component is sampled at full resolution one half as often. On average then, it takes 16 bits to represent each pel.

Using these values, it is easy to calculate the total disk space required to hold one second of uncompressed video in this format:

	720 horizontal Pixels
X	480 vertical Pixels
	345600 pels per frame
X	30 frames per second
	10368000 pels per second
X	2 bytes per pel
	20,736,000 total bytes per second

This means that the average 200 Megabyte hard drive could hold about 10 seconds of uncompressed video. Clearly this is not practical for most applications. This is where the incredible compression power of MPEG comes in. MPEG-1 will reduce this 20 Megabytes/second to 187,500 bytes/second (about 1.5 Megabits/second). This is a compression ratio of over 100:1!

Bits, Bytes and other Mysteries

There is a great deal of confusion surrounding these terms. One reason for this confusion is that computer programmers and engineers do not always use these terms in the same way. This booklet details these terms as understood by computer programmers, with bits being the smallest unit and Megabytes being the largest. Here is a rough conversion table of these units:

8 bits=1 byte
 8 Megabits=1 Megabyte
 1,048,576 **bytes** (or 2 to the 20th power)=1 **Megabyte**
 1,048,576 **bits** (or 2 to the 20th power)=1 **megabit**

So, if you had 23,640,251 bytes, and you wanted to know how many Megabytes you had, you would divide this number by 1,048,576 bytes:

$$\begin{array}{r} 23,640,251 \text{ bytes} \\ \text{divided by } 1,048,576 \text{ bytes} \\ \hline \text{equals } 22.54 \text{ Megabytes} \end{array}$$

However, for the purpose of rough estimates, you could just move the decimal over 6 places to the left to tell how many Megabytes (about 23 Megabytes). Then to convert to megabits, just multiply by 8.

To convert from Megabytes to bytes, simply multiply by 1,048,576 or for a rough estimate, move the decimal 6 places in the other direction.

FROM 20 MEGABYTES/SECOND TO 187,500 BYTES/SECOND

The first step of MPEG-1 is to reduce the number of pixels representing the image. It reduces the resolution of the image from 720x480 pixels to 352x240 pixels--providing more than a **4:1**

compression:

$$\begin{array}{r} 352 \text{ Horizontal Pixels} \\ \times 240 \text{ Vertical Pixels} \\ \hline 84,480 \text{ Pixels per Frame} \\ \times 30 \text{ Frames per Second} \\ \hline 2,534,400 \text{ Pixels per Second} \\ \times 2 \text{ Bytes per Pixel} \\ \hline 5,068,800 \text{ Total Bytes per Second} \end{array}$$

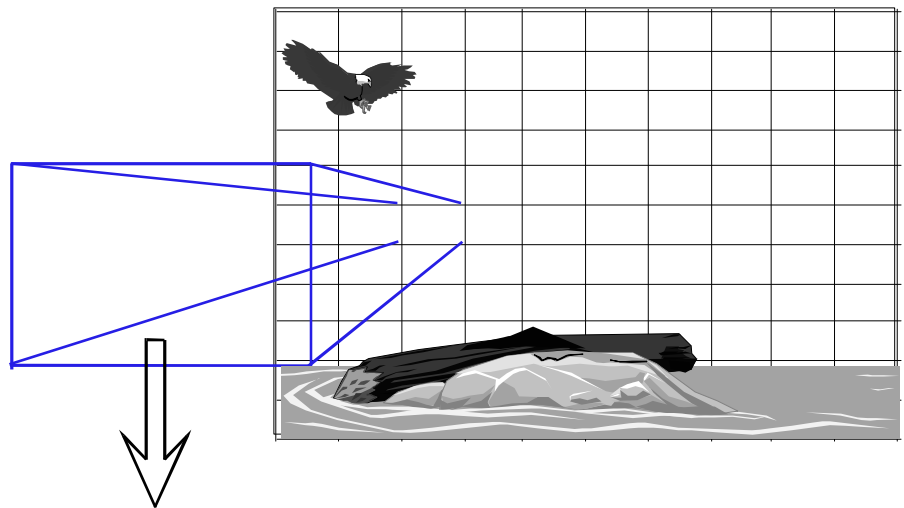
The 20 Megabytes per second is reduced to **4.83 Megabytes per second**. The 720x480 image size is considered to be professional video size. It is the image size broadcast studios use when making video. The basic premise (in analog

video) is to work with something larger than you need, and hopefully by the time it gets to the home (after video editing and retransmission) it will still contain acceptable quality. Thus MPEG reduces the picture size to 352 x 240.

The next step of video compression is to realize that over a small region of a single frame, the image does not change *much*. For example, take a picture of a field with a pond in it. If you only look at a small portion of the picture, say 8x8 block of pixels, you will probably see a block of grass, a block of blue sky, or a block that only contains the pond. In other words, most of the pixels in any given block probably look pretty much the same. MPEG takes advantage of this property of images. It breaks the image into 8x8 blocks and then applies a mathematical function to each block called a DCT. (DCT stands for Discrete Cosine Transform.) The DCT in and of itself does not provide any compression. However, the DCT does provide a means for compression. There are two properties of the DCT that make it well suited for assisting in video compression:

- It packs the data most visible to the human eye into a tighter spot, e.g., more zeros appear in the DCT encoded block (zeros are easy to compress).

It orders the data so that it is easier to throw parts away that are not particularly visible. Generally the DCT results in data in the upper left hand corner as data to keep, while the bottom right contains data that can be thrown away (possibly noise in the image).



Digitize

23	11	12	257	414	0	0
45	10	11	56	83	324	211
71	1	22	23	413	11	0
92	0	82	444	1421	0	9
80	1	264	4	4	1	0
12	11	42	2	8	0	0
0	0	0	0	11	10	31
34	0	1	0	2	1	1

DCT

63	1	2	4	4	4	1	0	0
0	1	1	3	3	4	2	1	0
1	1	2	3	4	3	1	0	1
2	0	2	4	4	4	1	0	0
0	1	2	4	4	4	1	0	0
2	1	2	2	8	0	0	0	0
0	0	0	0	0	0	1	0	0
0	1	0	2	1	1	1	0	0



Quantize

63	0	0	3	3	3	0	0	0
0	0	0	3	3	3	0	0	0
0	0	0	3	3	3	0	0	0
0	0	0	3	3	3	0	0	0
0	0	0	3	4	3	0	0	0
0	0	0	0	5	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

The key to using the DCT is quantization. The purpose of quantization is to keep only the pixels in the 8x8 block that are most likely to be visible to the human eye. Quantization also attempts to make the data more compressible by making values near each other more similar to others (the more similar the values are the higher the

compression). One example of quantization is rounding.

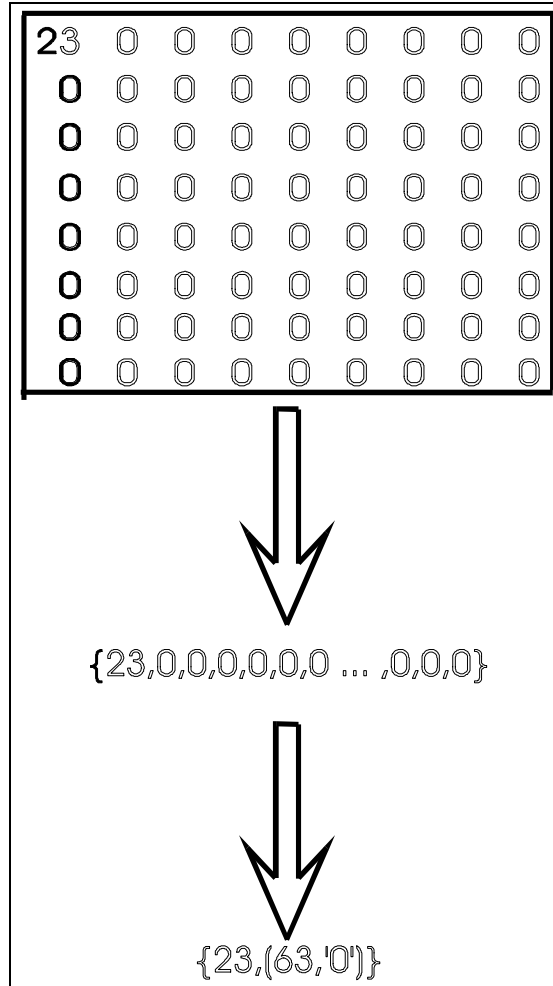
$$1.4 = 1, 2.6 = 3$$

In rounding, several digits become fewer digits. This is the only step of MPEG that is considered "lossy". Again, using the rounding example, once a number is rounded to 1, it is

impossible to determine whether the original number was 1.2 or 1.4 or any other number between .051 and 1.5.

Once the 8x8 block is DCT transformed and quantized, it can then be compressed. The

compression takes place by removing the redundancies in the data, or more specifically reducing the entropy. For example, we can see that $x+x+x+x+x$ can be represented as $5x$. One example of reducing the entropy or redundancy within the data is as follows:



This notation implies that there is a number '23' followed by 63 0's. This is clearly a compression of the original data. That is, it is shorter to write {23,(63,'0')} than (23,0,0,0,...,0,0,0) with sixty-three zeroes written out.

MPEG uses one of two types of mathematical processes for entropy encoding, either **Huffman encoding** or **Arithmetic encoding**. The DCT along with quantization can (depending on your source material) compress data from 10-20:1 without unacceptable video "artifacts" (errors or noise visible on

the screen). The amount of compression is mostly a function of the amount of quantization that takes place.

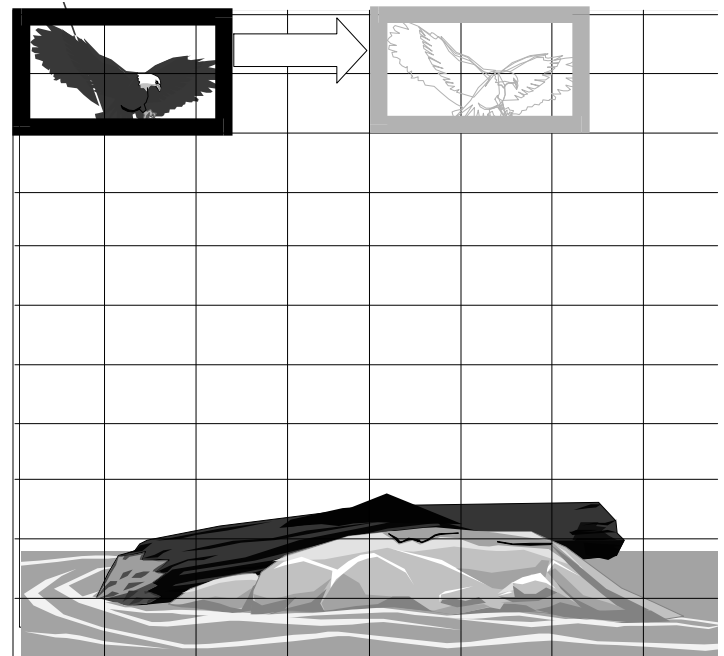
The remaining 4.83

Megabytes/second is now reduced to **0.322 Megabytes/second (15:1 compression)**.

The next major feature of MPEG compression is the realization that the interframe differences are also small (intraframe is within a frame, interframe is between frames). If you have ever looked at two frames of video side by side, you would be hard pressed to see the differences. Or, better put, the majority of the frame stays the same, while only a small portion changes. MPEG uses this property as its final tool to assist in compression.

MPEG uses the notion of **motion vectors** to track movement in an image. Similar to how the image is

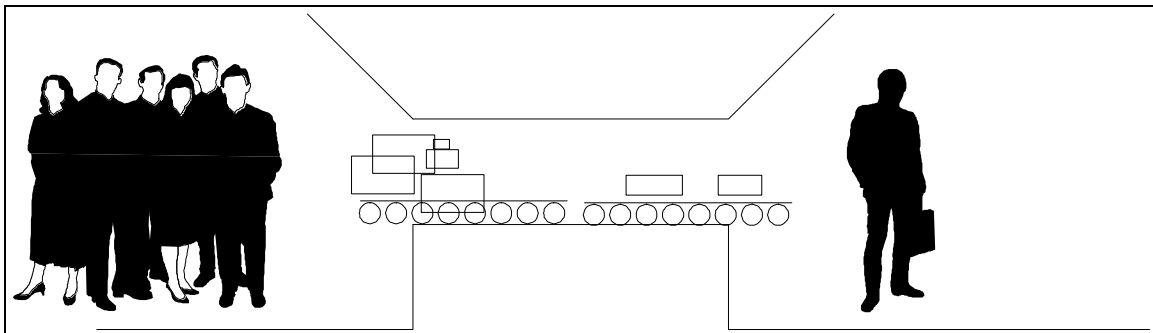
broken up into blocks for DCT compression, MPEG again breaks the picture into blocks (but this time 16x16 pixels, commonly referred to as macroblocks). Each macroblock can be encoded as its own completely separate motion vector. Therefore, if there are parts of an image that are staying still (the pond) from frame to frame, a motion vector that does not change from frame to frame can encode a portion of the pond. If there is a bird flying across the pond, a moving motion vector that tracks the bird can determine the changes in the bird's position and encode the bird from frame to frame. Using motion vectors can easily add **2-4:1 compression** so the 352,188 bytes/second easily approaches our target rate of **187,500 bytes/second (about 1.5 Megabits per second)**.



MPEG and Constant Delivery Bitrates

An often overlooked, but important part of MPEG, is that it provides a constant delivery bitrate. Although this is not important to compression, it is important to real world decoding applications. For example, in the case of MPEG-1, the target delivery rate was 187,500 bytes/second or 1.5 Megabits/second. This implies that the bitrate does not fluctuate. On average, no matter the content of the video, the delivery rate will be 1.5 Megabits/second. This is important to anyone who needs to transmit the encoded material either to a satellite, from a CD-ROM, or through a telephone switch. Without constant bitrate delivery methods it is impossible to determine (cost

effectively) the most efficient method for transmitting the data. A good analogy is a mailroom chute. When the delivery rate is constant, the packages move smoothly through the chute. Without a constant delivery rate, the chute can become overloaded or underused. The only means to handle a non-constant delivery rate is to build a mechanism into the chute that either speeds up or slows down the conveyor belt (a costly choice when that “conveyor belt” is really a satellite already orbiting the earth or transoceanic cable buried under the sea). The best way to maximize efficiency of the chute is to determine the capacity of the chute and make sure all packages are the same size, and designed to just fill the chute at all times.

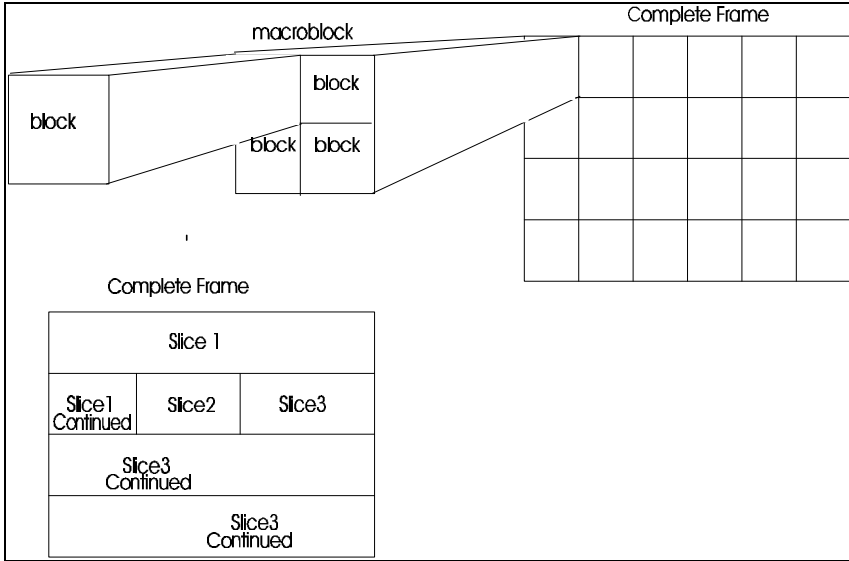


The mail chute.

I, P, AND B FRAMES. MPEG organizes video into several logical units. Each unit is logically composed of smaller units. The basic units of MPEG in order of increasing size are:

- Blocks
- Macroblocks
- Slices
- Pictures
- Group of Pictures (GOP)

GOP's are made up of pictures which are in turn made from slices which contain macroblocks which contain blocks.

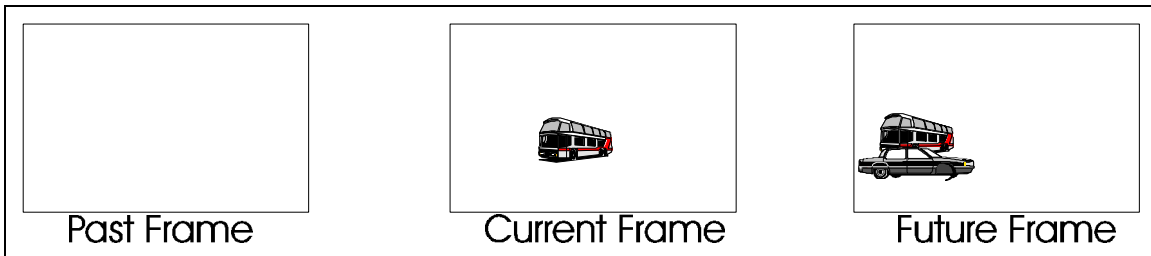


The largest data element within MPEG is the group of pictures (GOP). This is a self contained, completely decodable unit. It may consist of an arbitrary number of frames of any of the primary types (I, P or B frames).

Blocks are composed of an array of 8x8 pixels. The DCT is performed on the block level. Macroblocks are composed of blocks (in MPEG-1 a

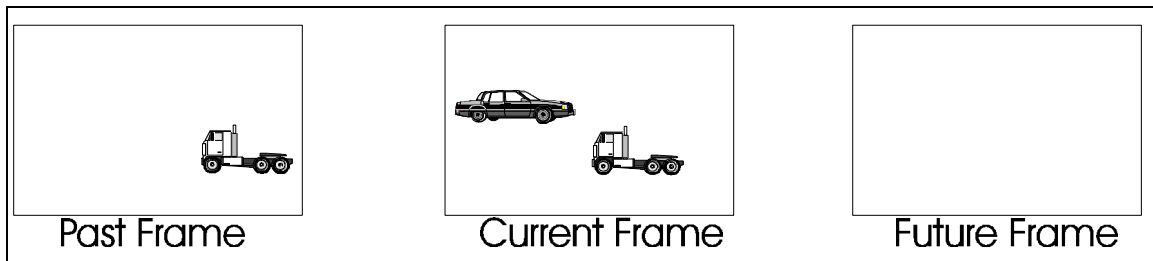
macroblock consists of 4 luminance blocks and 2 chrominance blocks). Motion prediction and interpolation are performed at the macroblock level.

Motion prediction is useful when objects are just appearing in the current frame and will continue to exist in future frames.



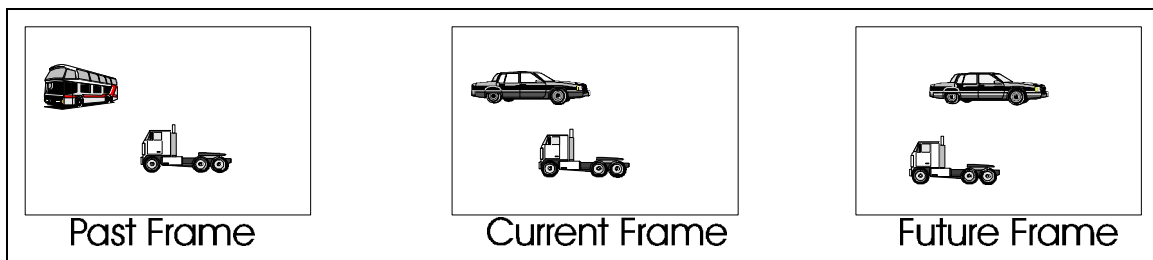
Bus is predicted.

Prediction can also be used when the object exists in the past and exists in the current frame. However, MPEG only uses forward prediction.



Truck is predicted.

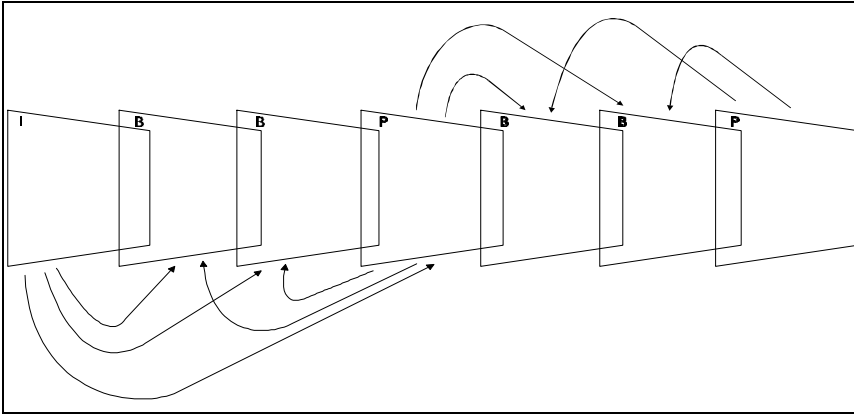
It is important to note that MPEG can also make use of interpolation. This is useful when much action is taking place in the scene. Since the truck exists in all frames, it can be reconstructed with information from both the future frame and the past frame. The truck is therefore interpolated. The average of the two will result in the final reconstructed truck. On the other hand, the car is predicted, since it only exists in the future frame. Use of both prediction and interpolation results in a better quality “current frame”. Simply put, the more information to build the “current” frame the better the image will appear.



Truck is interpolated and predicted, car is predicted.

A picture is composed of one or more slices. The “height” of a slice is one macroblock and the length can be the entire picture or any portion thereof. Slices can be used to aid in error recovery. In the case of “perfect” transmission mechanisms, such as CD-ROMs, there is normally one slice to one picture. (Compared to “imperfect” transmission mechanisms like terrestrial broadcast.)

MPEG uses three different types of frames. The primary difference among these frames is how motion vectors are used in them. Intraframes (I frames) do not use any form of motion vectors. Predictive frames (P frames) make use of predictive type motion vectors. Bi-directional frames (B frames) make use of both predictive and interpolative motion vectors.



A typical GOP illustrating information flow from I and P frames to B frames and I frames to P frames.

Conclusion:

In conclusion, MPEG is a toolset for the compression of video and associated audio. The primary tools in MPEG compression are the DCT, quantization, entropy encoding, use of motion vectors, and constant bitrate delivery mechanisms. Hopefully, enough has been said to give the reader some insight into the main concepts of MPEG without too much tedious technospeak. Much has been left out in this discussion including some key areas:

- Audio compression.
- Multiplexing of audio and video sequences into a single stream for transmission.
- Discussion of the psychovisual models and the use of colorspace.
- A discussion of groups of pictures (GOP's) and slices and how they can be effectively used in error-prone transmission environments.
- MPEG-2 and how it is targeted at professional video and high definition TV (HDTV).

For more information about these and other MPEG related issues, please see the “Sources for Further Reading” section.

FOR FURTHER READING

For a good technical introduction to JPEG, see: Wallace, Gregory K. "The JPEG Still Picture Compression Standard," Communications of the ACM, April 1991 (vol. 34 no. 4), pp. 30-44. (Adjacent articles in that issue discuss MPEG motion picture compression, applications of JPEG, and related topics.)

An alternative, more leisurely explanation of JPEG can be found in "The Data Compression Book" by Mark Nelson [Nel 1991]. This book provides excellent introductions to many data compression methods including JPEG, plus sample source code in C. The JPEG-related source code is far from industrial-strength, but it's a pretty good learning tool.

An excellent textbook about JPEG is "JPEG Still Image Data Compression Standard" by William B. Pennebaker and Joan L. Mitchell. Published by Van Nostrand Reinhold, 1993, ISBN 0-442-01272-1. 650 pages (second printing). This book includes the complete text of the ISO JPEG standards, DIS 10918-1 and draft DIS 10918-2. Review by Tom Lane: "This is by far the most complete exposition of JPEG in existence. It's written by two people who know what they are talking about: both serve on the ISO JPEG standards committee. If you want to know how JPEG works or why it works that way, this is the book to have." There are a number of errors in the first printing of the Pennebaker & Mitchell book. At last report, all were fixed in the second printing.

The official specification of JPEG is not currently available on-line. I hear that CCITT specs may be on-line sometime soon, which would change this. At the moment, your best bet is to buy the Pennebaker and Mitchell textbook.

Fritz Whittington of Texas Instruments has compiled a good slide presentation of United States, and International Standards, in "Standards 101." This is normally presented on the first day of meetings for United States X3L3 get-togethers.

Internet news (comp.compression.research, comp.compression) provides many answers to FAQ's (frequently asked questions) pertaining specifically towards JPEG and MPEG, some of which are over 40 pages in length.

International Standards Organization provides copies of the MPEG-1 standard, reference number ISO 11172, parts 1 through 5. (obtained through ANSI in Washington DC)

ISO JPEG standards, reference number ISO 10918-1.

ISO MHEG standard, ISO 13522 (DIS, Draft International Standard).

ISO MPEG-2 standard ISO 13818 parts 1-3.

"The JPEG Still Picture Compression Standard" Greg Wallace, December 1991
IEEE Transactions on Consumer Electronics.

"The computer music and digital audio series," Ken Pohlman, A-R Editions Inc.,
Madison Wisconsin.

NewMedia, March 1994, pp 48,50.

AV Video, March 1994, p112.

TV Broadcast, September 1993, pg 42.

Appendix A -- Some Other Video Standards

This is not meant to be a complete list of ALL other video standards besides MPEG, but rather a list of some of the most common ones:

Apple Video is a proprietary software-only symmetrical codec for video compression offered with QuickTime. A typical compression ratio is only about 6:1 and the resultant video quality is considerably lower than MPEG-1.

Cinepak is a proprietary software only, scalable codec for video compression licensed by Radius.

DVI is a compression scheme designed to be used with Intel's 1750 chipset. It is currently used in a few video editing systems. Intel has announced that it no longer supports this scheme.

Fractal compression is a technology aimed at achieving some of the same results as the MPEG algorithm. This technology has many proprietary implementations, and is not standardized. The basis for fractal compression is the belief that a picture can be represented by mathematical equations that are also used in representing fractals. Research suggests that this compression method may prove to have some advantages over MPEG, especially in the amount of compression. Its largest drawback

is that it is still very experimental, and is usually very expensive to implement.

Indeo is a proprietary symmetrical compression method which has much lower color resolution than MPEG, and generally results in lower video quality than JPEG or MPEG compression methods. This standard is generally not well suited to low bit-rate situations like CD-ROM delivery. However, this proprietary method has been designed in such a way that no special hardware is required to encode or decode the video.

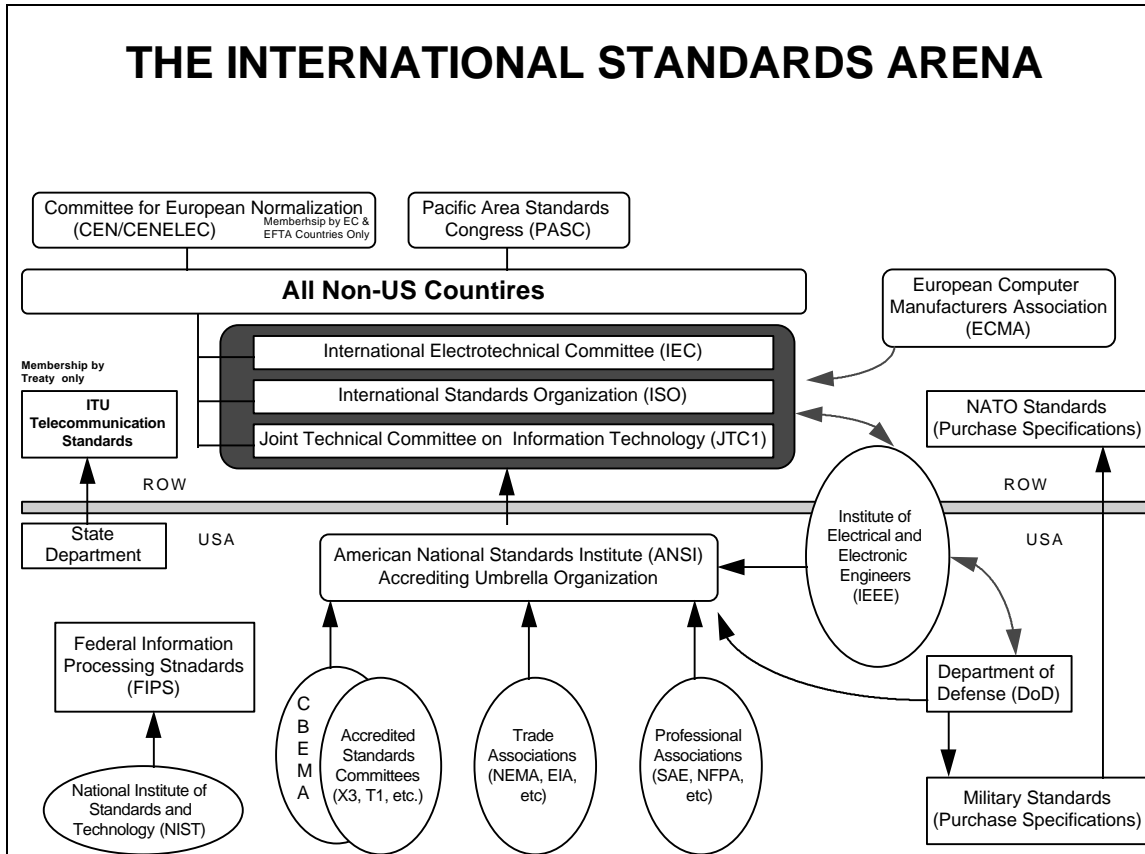
JPEG is a standard designed for the compression of continuous tone still pictures.

Microsoft Video1 is a proprietary symmetrical compression method which has scaleable playback depending on hardware capabilities. It was developed for Video for Windows. Resultant video quality is significantly lower than MPEG, but encode and decode can be handled in software without special dedicated hardware.

Motion JPEG refers to a method of utilizing JPEG still image compression over a sequence of frames. Motion JPEG is not a standard, and many different implementations exist. Motion JPEG usually allows maximum compression ratios of 20:1 in contrast to the 200:1 compression usually associated with MPEG-1.

TrueMotion-S is a proprietary method of video compression. It still requires a custom capture/encoder card, which may not be needed in the future. What separates this system from the rest is that its compression algorithm enables full-frame-size, full-frame-rate motion video to be played back through a computer or video game system without any additional hardware. Video quality compares roughly to MPEG-1 encoded at 1.15 Megabits per second.

Appendix B -- The ISO Organization



Courtesy of Texas Instruments

APPENDIX C--Summary of the Constrained Video Parameters:

Horizontal picture size	Less than or equal to 768 pels
Vertical picture size	Less than or equal to 576 lines
Picture area	Less than or equal to 396 macroblocks
Pixel rate	Less than or equal to 396 x 25 macroblocks per second
Picture rate	Less than or equal to 30 Hz
Motion vector range	Less than - 64.+ 63.5 pels (using half-pel vectors) [Vector scale code <=4]
Input buffer size (in VBV model)	Less than or equal to 327 680 bits
Bitrate	Less than or equal to 1 856 000 bits/second (constant bitrate)

Glossary of Terms

Arithmetic encoding: a mathematical entropy encoding process for the compression of data. May provide better compression than Huffman Encoding.

CD-I: "Compact Disc - Interactive," Philips' entry into the home video multi-player market. In common usage for industrial video CD applications.

CD32: The CD32 is Commodore Business Machines' entry into the home video machine market. Built around the Amiga computer hardware and software, the CD32 is an inexpensive and popular delivery mechanism for video CD playback. This product is currently not manufactured.

Colorspace: system for measuring and describing color. For example, CMYK (cyan, magenta, yellow, and black) is a common colorspace for printing.

Entropy coding: a class of lossless compression methods characterized by their orientation toward finding the shortest expression of data.

HDTV: high definition television.

Huffman encoding: a mathematical entropy encoding process for compression of data.

IEC: The International Electrotechnical Commission. The IEC is composed of 41 member nations. Two types of committees exist within IEC: standards writing committees and advisory committees. Within the standards writing committees are over 80 technical committees, 100 subcommittees and 700 working groups (in which the real work takes place). There are three advisory committees within IEC: ACET (Telecom), ACoS (Safety), ACEC (EMC).

ISO: An acronym for International Standards Organization.

ISO/IEC JTC1 SC29/WG11: (Formal name of the MPEG committee.)
Short for "ISO/IEC Joint Technical Committee 1, Subcommittee 29, working group 11" which defines data compression techniques and standards for motion pictures with sound.

Lossless: Converting a black picture using a lossless process and then applying the inverse process would yield an exact duplicate of the original black image. Mathematically, lossless algorithms are one to one functions.

Lossy: Converting a picture using a lossy process and then applying the inverse process may result in a picture slightly different from the original. Mathematically, lossy algorithms are many to one functions.

MPEG: An acronym for "Moving Pictures Experts Group", MPEG is the collective noun often used to refer to ISO working group WG11 (ISO/IEC JTC1 SC29/WG11), which defines data compression techniques and standards for motion pictures with sound.

MPEG-1: Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Megabits per second (International Standard IS - 11172, 1992). MPEG-1 is used for interactive multimedia and video CDs. The normal parameters for MPEG-1 are frames of 352x240 pixels in the YCbCr color space delivered at a rate of 30 frames per second.

Pixel: A picture element, the smallest dot on the screen. Also called pel.

Quantization: A method of renumbering values to aid in compression. An example of quantization is rounding: $1.4=1$, $2.6=3$, etc.

SMPTE: An acronym for the Society of Motion Picture and Television Engineers, a professional organization concerned with the technical aspects of movies and television.

Standard target decoder: the hypothetical model of an MPEG decoder. Encoded streams must be decodable by the standard target decoder (STD) or they cannot be labeled MPEG.

STD model: standard target decoder; see definition above.

Symmetry: In the context of data compression, symmetry is used to describe the relative complexity of the encoding and decoding algorithms. Because MPEG requires a great deal of complexity to encode and relatively little to decode, MPEG algorithms are described as "asymmetric".