Factors to Consider When Choosing an MPEG-2 Encoder

By Neil Brydon

This paper describes some important criteria to consider when selecting a video encoder. In addition to discussing features, it will outline the practical tests that users can perform to assess the relative merits of MPEG-2 encoder products. There are a variety of encoder designs but also a correspondingly wide deviation in the picture quality for a given bit rate. We are now seeing some admirable devices that go some way to measuring picture quality, but they only tell part of the story. This paper will explain how to augment objective measurements with additional tests and feature analysis. It should place the buyer in a better position to unzip the hype and fully evaluate the strengths and weaknesses of an encoder.

It is widely agreed that the encoder is one of the most difficult components to benchmark and consequently consider in the purchasing process. Integrated encoders have been on the market for nearly five years, and by now, buyers may have expected to be able to call on a strong pool of expertise and equipment to assist in sifting the hype when selecting an encoder and so reach a rational conclusion. This dream scenario is far from complete.

It is frequently forgotten that MPEG does not define the video compression process. Instead, standardization philosophy is centered on the decoder and the characteristics of the bit stream to allow it to be decoded. This successful strategy for standardization provides a high degree of freedom in the design of encoding systems. Subsequently, there are a variety of encoder designs but also a correspondingly wide deviation in the picture quality for a given bit rate, and in other words efficiency.

In most applications the efficiency of the encoder can lead directly to dollars. There are some hard facts to consider: an encoder should be standards compliant; the output bit stream must be legal and never violate any syntactical or semantic rule of MPEG-2, so that this bit stream can be decoded on an idealized decoder. Performing a feature check and determining whether an encoder can produce syntactically correct MPEG streams is relatively easy. The difficulty arises in determining whether an encoder is efficient.

The efficiency issue is even more acute when running the encoder at very low data rates where the difference between encoders can mean substantially reduced bandwidth costs. Although we are now seeing some admirable devices for measuring picture quality, they only tell part of the story. This paper will endeavor to explain the background theory and how to practically assess some of the key MPEG-2 mechanisms that have an impact on the efficiency of the process.

Total System Solution

The digital video compression core is one part of a larger system used to deliver a wide variety of services and programming to the broadcast television audience. A network management system provides the glue that binds individual hardware components. Flexibility is the key requirement as no two systems are ever the same. A modular set of products that integrate easily will allow solutions to be implemented that can support a broad range of applications. The first task is to understand the application requirements and select a short list of viable contenders.

Risks are reduced if choice vendors are restricted to those that demonstrate a family of reputable products. An established vendor will probably be on a second, third, or even fourth generation encoder design. Consider how many encoders the vendor has shipped; if the vendor is well established, there will be numerous installations that can be referenced. These references will provide a unique insight into whether the supplier is capable of delivering not just products but reliable, complete solutions for each application. It is important to consider whether there is a credible support organization able to meet local needs. Adherence to open standards offers significant benefits—don't get imprisoned by proprietary solutions.

Consider the system aspects and feature checklist. Is the system flexible: markets change quickly. Does the encoding system have sufficient configurations to support a multitude of applications such as 4:2:0 and 4:2:2? Can the encoder support analog, digital, and embedded audio? Does the product fit well into your system? Will video preprocessing and noise reduction be required? Is the user interface intuitive and easy to use? Comprehensive fault monitoring and redundancy capabilities plus a good user interface will have a positive effect on manning requirements.

Features to Consider

Before an encoder reaches the test bench, ensure that the basic encoder and system architecture can meet the application. The following checklist of hard features may help decide which systems to consider for further evaluation.

- Serial digital video with embedded audio.
- High-quality PAL/NTSC composite decoder.
- Digital and analog inputs to audio encoder.
- Optional modules for more additional audio channels.
- Prefiltering and noise reduction.
- Network management.
- Local control panel.
- VBI handling capabilities.
- 4:2:2/4:2:0 switchable.
- 576 and 480 vertical resolution plus

Presented at the 140th SMPTE Technical Conference (paper no. 140-10), in Pasadena, CA, October 28-31, 1998. Neil Brydon is with DiviCom, Milpitas, CA, 95035. An unedited version of this paper appears in *DTV Reality Check—The First Year*, SMPTE, 1998.

support for expanded window format (512 and 608).

- Fully flexible GOP control.
- Automatic I-frame insertion for scene cut detection.
- Field/frame adaptation modes.
- Standard interfaces (open standards).
- Film modes (repeat field detection).

Once the hard features and system options appear to offer a solution it is time to evaluate picture quality. Picture quality testing is problematic because the technology is relatively new and there are few technical standards and recommendations to follow. In 1997, a system was introduced to provide an objective figure for picture quality. This is a useful advance but it provides only a partial solution to the problem. The fact remains, subjective evaluations require significant effort. Hopefully, this paper will offer some guidelines to help ensure that a user can commit his valuable test resource wisely and ease the way to a comfortable decision.

Picture Quality Test Philosophy and Recommended Toolkit

Once the basic feature set of an encoder system is approved, the biggest differentiating factor becomes compression efficiency. This section will outline some test tools, test procedures, and good practices to follow.

A good strategy is to isolate the abilities of the compression core from all other influences. It is useful to separate static video tests from dynamic picture quality tests: static tests are easily performed to obtain objective results. Standard test stripes and an automated test tool should be used for immediate results. The tests should be performed at a high bit rate to test the actual inherent abilities of the analog-to-digital conversion segments. Dynamic picture quality tests will take significantly more effort.

Remember that compression at low rates relies heavily on human eye response. Bits are applied where they are most needed. A good encoder will have good masking control and apply the bits most profitably. No equipment presently exists that can completely supplant human eyes when performing video quality tests, so for the initial tests, use expert eyes. The test team should be alert to problem artifacts: common problems to be aware of are edge noise, blocking artifacts, and throbbing (beating) effects. Periodically, it is prudent to validate that the findings from the experts are aligned with the opinions of nonexpert viewers.

Thorough evaluation of the core abilities of a compression system requires some basic tools and rigorous methods. Try to restrict the tests to the compression system alone. If possible, use digital feeds and monitor the results from a professional decoder. For comparison tests, monitor the results on two identical high-quality monitors aligned in a good viewing environment—use Picture Line-Up Generator or something similar to set brightness and contrast.

The International Telecommunications Union (ITU) BT 500 (1) standard defines recommended test and viewing guidelines. However, here are some basic viewing environment suggestions:

- Brightness—70 to 90 cd/m^2.
- Viewing distance 4 to 6 times picture height.
- Ambient light—10% of peak white.
- Ambient background color—gray.
- Reflection-free environment.

The user will ease the workload if a portfolio of standard tests is created and readily accessed from some convenient medium such as a digital low-compression VTR. It is suggested that the portfolio should include:

- High and low-quality video—football, hockey, basketball, news acquisition, commercials, normal mix.
- High and low-quality film.
- Talking heads.
- Animation.
- Noisy material—different types of noise (film, radio frequency, random, etc.).
- Pathological tests—brutal sequences designed to find the breaking point of a system.
- Special effects—dissolves, explosions, flashes, fades, cuts.
- Synthetic—computer-generated tests, noise overload tests.

The reference to synthetic tests refers to standard test stripes, circle sweeps, computer-generated killer tests, and custom-made transition tests that include dissolves, fades, or cuts. Try to include brutal transition tests to expose weaknesses in an encoder. The video section should include normal sequences such as action sports as well as sequences that are known to be tough.

The user should become familiar with the toolkit of sequences. Drive the system hard (high resolution and low rates). Learn what to look for and where to find the problems. Be rigorous; compare like with like. Constrain the tests to a limited set of encoder parameters. Evaluate individual sections separately so that areas of strengths and weaknesses can be isolated. It may be worthwhile to record results for future reference and for future demonstrations.

The following sections describe the basic processes that affect quality and how to assess the performance. They offer some explanatory guidance on MPEG basics, motion estimation, rate control, and preprocessing tests.

MPEG Encoding—An Appreciation of the Basic Technology is Invaluable

A prospective encoder buyer is liable to make a more enlightened and comfortable decision if equipped with a rudimentary understanding of the basic principles of MPEG. Ideally, the user should be familiar with key mechanisms and their impact on the operation and efficiency of the process.

Common Terms

- CBR—constant bit rate.
- VBR—variable bit rate.
- Temporal prediction—motion estimation is used to exploit the redundancy between video pictures.
- Discrete Cosine Transform (DCT) —a math time-to-frequency transform to exploit spatial redundancy.
- Quantization—a finite buffer and rate control mechanism used to control bit rate.
- I Picture—spatially encoded picture used as a reference.
- P Picture—differentially predicted picture from I and P pictures.
- B Picture—differentially encoded from surrounding pictures.
- Group of pictures (GOP)—structured combination of I, P and B pictures.

The incoming video sequence is preprocessed using interpolation and filtering to break the picture into the basic processing units, a set of 8 x 8 pixel blocks. Next, motion estimation is applied to determine motion vectors plus error difference. This is the first process to remove inter-frame redundancy. The information is based on "this section moved over there with this set of differences" philosophy, and is done throughout the picture. The resultant data is then mathematically processed using what is called a DCT. The outcome is a stream that is naturally easy to represent with varying resolutions. This resolution reduction is termed quantization and is used to control the bit rate.

To ensure MPEG compliance, the encoder needs to control the bit rate into a buffer that replicates a standardized decoder buffer. This is termed rate control and is used to verify that the decoder will be able to manage the stream. There are also standard math processes such as Huffmann and run-length encoding to zip the data without loss. The processed data is also decoded within the encoder and fed back into the earlier processes to provide control and reference information for subsequent pictures. The encoder produces a constant stream of I, P, and B pictures, plus overhead information that a decoder must understand. In one paragraph, a simplified description of MPEG compression!

Motion Estimation

The factor that makes MPEG efficient is its ability to extract temporal (inter-frame) redundancy. The key process is the temporal prediction that exploits the intrinsic redundancy between adjacent video pictures. Normally in video sequences, adjacent pictures have a great deal of commonality.

Large parts of a picture are likely to remain significantly unchanged. Hence, by applying a "picture difference" algorithm, substantial compression can be achieved. This can be extended if picture motion is taken into account. This technique, called motion compensation, is very effective until discontinuities or transition effects reduce the effectiveness of the motion search system.

An effective encoder will have exceptional motion estimation (ME) capabilities (a multitude of ME techniques are offered). The algorithm must also recognize when this tactic is losing effectiveness and adapt to a more efficient process. Strong translation-based models are fine for pans but will find effects such as rotations, zooms, or dissolves a challenge.

Full search algorithms have limited range and are not necessarily the best choice. Limited full search is restricted in range and will tend to be good for slow material but less effective for fast motion or effects. Hierarchical search systems can offer a good compromise. This paper does not offer the scope to discuss the general merits of hierarchical versus limited full search schemes. It is suggested that the user be less concerned with the specified ME approach taken and instead evaluate whether the overall implementation is effective for all sequence types.

Good motion estimation will have a positive effect on overall quality. Use high and slow-motion test sequences to see how the system copes and how well it handles high action. Try to assess how well the system reacts to noisy pictures. High-motion areas that contain high chroma are liable to show poor edge definition, so look for ghosting, trailing, and spurious effects that may indicate weaknesses in the ME system. Action sport sequences and synthetic tests are good examples for validating the effectiveness of the ME core.

Adaptive Field/Frame Modes

Separate field and frame prediction is a feature not supported by all encoders. It offers certain advantages for some picture types but requires extra investment in terms of hardware and microcode support. Frame-based prediction is best for slow moving images, but field prediction is best for areas of higher motion. Many pictures will contain elements that require both modes to be adaptively applied.

This agile adaptation can pay dividends in the crucial arena of sport. Sport is generally highly complex, hard to compress, and a very common broadcast content. Fast moving foreground action is usually best compressed by field prediction math, while the detailed but slow moving background is probably more suitable for frame prediction. Of course, not only does the encoder need to provide good modes, but it also must provide excellent adaptation control. Inappropriate adaptation can be visually intrusive.

Rate Control—Transition Response

Rate control refers to the overall control loop that manages buffer occupancy and quantization control mechanisms. Good rate control becomes most evident at lower rates where the data reduction nibbles into the natural entropy of a sequence. A balanced system will achieve better overall results and therefore require fewer bits than a system with mediocre rate control.

The quantization control function aims to constrain the incoming average bit rate through the buffer at the nominal CBR rate. The buffer is used to improve quality by absorbing instantaneous fluctuations in picture complexity. The encoder must use this buffer model to achieve optimal compression and also verify that noncompliant overflow or underflow conditions do not occur.

Underflow is easily avoided by using stuffing nulls to add "nothing messages" into the stream. This is wasteful and should be minimized. Overflows are more difficult to handle; they result when incoming content is so complex that it is overdrawing on the bit rate account. Heavier quantization to reduce the fidelity of the compressed image will resolve these overload conditions. The clever thing is for an encoder to avoid getting into difficulty in the first place and secondly to be able to gracefully manage the scenario when it occurs. The changes in quantization must be applied carefully since extremely large-step changes will be visually intrusive. Designing a really effective compression system is a real challenge, it must be all-wheel drive and able to adapt to any video terrain.

The way to test for good buffer management is to apply a sequence of material that contains many difficult transitions, such as dissolves and cuts, especially transitions between easy and extremely hard material. Make life difficult for the encoder by setting a low rate. Use your eyes and consider the decoded response: how gracefully are the cuts handled? Focus on transitions: how many frames does the encoder require after a cut before the new picture is acceptable? Apply overload tests, progressively increase the loading, see what happens. Consider how low the rate can be set before the encoder resorts to emergency panic measures. Can the encoder be provoked into a fault condition such as freeze frames or even lock up? These tests should expose weaknesses in an underdeveloped encoder, and it is possible that fatal non-compliance conditions can be found.

One rate control option that can produce significant gains is to incorporate the preprocessing function into the compression control loop. This requires close coupling between the encoder and the preprocessor, so that if the encoder starts to overload then the incoming content can be reduced in complexity by filtering to diminish the overload. Due to the time constants involved, this is a difficult function to apply successfully and requires tight control to prevent the system from becoming unstable.

Rate Control—Good IPB Balance

Part of the rate control strategy is to reflect the temporal redundancy of a picture and assign bits to each of the picture types I, B, and P. B frames cause most degradation and error accumulation. The predicted picture will decay as errors accumulate and then be cleaned up as I or P frames are inserted. If a good rate control balance between intra and predicted pictures is not maintained, annoying and visible effects may occur. The algorithm must prescribe just enough bits to each picture type to maintain efficiency and minimize refresh throbbing. This balance needs to be adaptive depending on temporal content (motion or still). Typically, negative effects will be seen as throbbing effects, either at I or IBBP rate (2 to 6 Hz).

Testing for good IPB balance is best achieved by considering some steady state (no transition) picture sequences. Set a rate that reflects the intended application then stand back and see if there are any obvious throbbing effects—hopefully not. Look more closely and observe how the encoder performs in flat areas, then in areas of high luma detail, and in noisy sequences. This is a very subjective test—ideally compare the encoder under tests with some reference system.

Statistical Multiplexing

For some applications, such as satellite broadcasting, bandwidth resource is extremely precious. The value of an extra few percent of efficiency is substantial if it allows an extra channel to be squeezed into the given bandwidth. Some encoder systems now offer statistical multiplexing options to meet this challenge. Statistical multiplexing is certainly effective, even on a limited number of video channels, and the benefits rise in conjunction with the number of channels in the pool. Approximately 25% fewer bits are required for a 10-channel system when compared to ten individual CBR channels. Does the target encoder support this?

The philosophy behind statistical multiplexing is much more straightforward to grasp than actual implementation techniques that are extremely complex. The basis of the process is to marry the constant quality features of VBR encoding with the fixed channel capacity requirement of a CBR application. The fixed channel width is shared by a group of encoders that bid for bandwidth depending on the complexity of their incoming signals, then channel requests are analyzed and the available bits are distributed accordingly. It is a difficult process to determine the complexity of the incoming signal, but this analysis is required for the encoder to decide what bit budget to bid for to meet a certain quality level.

Another challenge is to do this in realtime and preferably in zero time. Most statistical multiplexing solutions are reactive, so the instances of high bit demand typically last only for a few frames. Scene cuts, for instance, require an instantaneous supply of bits over very few frames to help recreate a completely new picture. Irrespective of the reaction time, a reactive encoder is always responding to past events, and this results in compression artifacts and wasted bandwidth.

A look-ahead architecture enables the control system to proactively allocate the precise rate at the required instant, resulting in higher quality video and better bandwidth utilization. If the look-ahead architecture is configured to perform a full look-ahead encode then a wealth of statistics can be extracted to help guarantee optimal bit allocation. The look-ahead principle is not a new idea, what is new is that the technology is now available to allow practical implementations to be commercially realized.

Don't forget that a statistical multiplexing system requires a sophisticated monitoring and control interface to ensure that the full potential of the system is realized.

Preprocessing

Noise is a prime enemy of compression systems. Prefiltering and noise reduction of video inputs can offer net gains in perceived quality for the overall process. Preprocessing offers most benefits when encoding at low bit rates and with noisy or poor quality source material. A competent prefiltering system needs two things; a good set of filters and a good adaptation scheme that is able to apply the filters sensibly. The filter control needs fine control of the settings to allow graceful adaptation as filters are applied.

Users must consider the value of an internal versus an external system. The benefits of an internal system usually are reduced system cost, smaller size, and ease of redundancy. It is probable that an advanced system may have access to look-ahead and rate-control statistics to allow additional feed-forward and feed-back mechanisms that assist the adaptation process.

A preprocessing section is likely to include several different filters. Horizontal and vertical filters are spatial (intraframe) filters used to reduce high-frequency components to improve compression efficiency. Overuse of these filters can result in picture softening and reduced edge definition. Some systems may have some edge adaptation to protect the edge. Random and grainy noise is removed using interfield temporal filters (recursive). A temporal filter solution usually needs motion detection and edge adaptation control to prevent blurring in moving areas. Lastly, some systems feature a median filter to remove sparkles from satellite and RF feeds.

To analyze the prefiltering capabilities, first isolate the performance of the individual filters using familiar synthetic and reference sequences. A good test is a circle zone plate that is split so that one side is static and the other swings back and forth in an arc. This enables static and dynamic characteristics to be evaluated with one test. Set a high bit rate and apply each filter individually; look for artifacts and any unexpected behavior; analyze for frequencies response holes. Are the roll off areas smooth or abrupt; any evidence of ringing effects; does the adaptation control correctly recognize motion and edge areas?

Once this is understood, then play with the control options and verify whether the filters can be teamed to provide beneficial results. Use a variety of material; try to include a mix of clean, noisy, and grainy video; assess whether the user interface offers sufficient access to all the parameters. Do the filters perform correctly or do they introduce unwanted artifacts; how effective is the adaptation control; can the adaptation control parameters be modified to set different levels of aggression?

System Control

A system controller must provide both element and network management. The element API must be able to expand the control of most system elements to the very edge of the network. The controller ought to provide an "end-to-end" view of the digital video subsystem with an interface that is user friendly.

Systems that are intuitive and easy to use help minimize training requirements and human investment. Quick start guides and context sensitive help buttons will assist the user in every operation. Additionally, it may be beneficial if the system software can auto detect hardware configurations to prevent users from making unnecessary connections and reconfigurations.

One way to manage the elements in the network is to use a local area network/wide area network (LAN/WAN) subsystems approach. A protocol such as Simple Network Management Protocol (SNMP) combined with TCP/IP can provide communication to the devices in a standardized and open fashion. By using standard interfaces and software protocols it is possible to easily create software drivers to integrate third-party systems into the overall network. The LAN/WAN approach means that control could be applied remotely. In addition, there is the option to consider distributed multiple management platforms that can localize faults as well as distribute processing power throughout the network.

Service Management Considerations

Automated fault detection, isolation, and redundancy options can minimize system downtime and reduce the need for human intervention. A compression system should be capable of maintaining service despite complete failure of the network management application.

Although there is some hardware-intensive processing involved in the system, it is probable that overall functionality can largely be software configurable. This philosophy reduces hardware inventory and field service overheads and also enables upgrades to be wired directly to customers or configured through flashcards. The business environment is often difficult to predict, and flexible upgrade opportunities offer longer product life and increased ROI. Consider whether software upgrades can be applied as a background task and implemented quickly, with minimum intrusion.

More rack space equals more racks and therefore higher facility costs. Physical size, weight, and power dissipation have ultimate relevance in mobile units but can be significant considerations in other applications. Many purchasers of digital video systems fail to accurately estimate the real costs and value associated with rack space. Consider the system as a complete proposal. Is there enough space for now and for the future? The implications of providing extra space and cooling can be daunting if not foreseen.

MPEG-2 Synchronization

The MPEG-2 transport stream relies on presentation timestamps in order to reinstate the correct timing of the decoder. This requires that the encoder accurately lock the audio and video to the program clock (PCR). This step is often missed when amateur system designers try to patch a system together from separate audio and video encoders; at best, they discover they have audio and video synchronization problems. A worse scenario is that the decoder fails to track the audio because the presentation time stamp (PTS) is out of range for some of the audio packet identifiers (PIDS).

Since the PCR underpins the system timing, it has a strong onus to be stable. Unwanted fluctuation or jitter can cause problems downstream and perhaps destabilize the decoder timing recovery loop. Other issues may arise if the stream has to undergo some later multiplexing function or transmission function.

Commercial Issues

Like most things in life, you get what you pay for. Be wary of manufacturers who may try to recover costs with expensive additional items. Consider the longer view: Is the vendor wholeheartedly committed to this segment and able to demonstrate an attractive roadmap and upgrade path; are the installation and service contracts reasonable; is the vendor able to demonstrate credible support in your geographical area?

An allegiance to open standards and nonproprietary solutions offers a customer the greatest flexibility and strongest opportunity to find the most competitive solution. This extends to the whole solution and should include consideration for conditional access and the set-top box solution. Open standards will naturally tend to offer greater economies of scale and lower risk than any custom proprietary solution.

Conclusion

At this point, making the final choice should be easy. If the technical research is done properly, one or two compression system solutions will probably have emerged as compelling solutions. Negotiations with the vendor will have resulted in relationships that will help to seal the deal. The biggest problem will be if the preferred technical choice is barred through political or commercial issues, and if this happens, at least you will know which vendor to approach when things go wrong!

THE AUTHOR

Neil Brydon's primary area of expertise is in video quality. Since joining DiviCom, he has focused on marketing encoding systems, including DiviTrack, a statistical multiplexing feature of the MediaView MV40 program encoder.

Previously, Brydon worked for the ITV Network Technology Centre in the U.K., where he led the hardware approval tests for the ITV Synchronous Digital Hierarchy inter-studio network. He also developed and approved picture quality tests for PAL decoders, aspect ratio converters, videotape recorders, disk servers, and noise reduction and satellite news gathering equipment.

Brydon received a B.S. degree in electrical engineering from the Robert Gordon Institute of Technology in Scotland.