Noise Reduction Preprocessing for MPEG-2 Encoding

By J. Patrick Waddell and Neil Brydon

J. Patrick Waddell

Neil Brydon

In the real world, all video content contains some noise. Even modern facilities equipped with the latest digital production equipment will inevitably import noisy content. Archival material, film, and news feeds contain high levels of noise. Noise reduction (NR) and filtering can substantially improve the image quality received by the viewer if the appropriate techniques are gracefully applied to remove the noise prior to compression. Selectively removing noise is a challenge because it shares the same bandwidth as valuable picture detail. An ideal noise reduction process will allow suppression of all types of noise while preserving clean video content. An advanced noise reduction system will use sophisticated techniques that offer powerful abilities to remove noise without introducing side effects such as motion blur or ghosting.

Harmonic Inc. has introduced an MPEG-2 encoder with third generation embedded advanced noise reduction capabilities. The architecture is unusual for MPEG-2 encoders, so the design background is discussed in this paper in some detail. The encoder has permitted users to push encoding rates for video below 2 Mbits/sec with good results. Also discussed is the challenge of signal noise in MPEG compression, MPEG-2 specifically. Assuming that not all readers are MPEG experts, a brief tutorial is in order.

The first question to address is "Why is noise an issue?" (And a major issue at that?) In analog transmission systems noise may be considered ugly, but it hardly constitutes a major performance limitation. The production community has been relying upon the ability of those analog systems to send back all kinds of ugly video to the station. "After all, it's NEWS!" is the usual excuse given. As we introduce a two-generation shift in transmission technology, this suddenly changes from being simply an annoyance to a major concern both for engineering and production staffs.

The Insidious Nature of Noise in Video

In considering the fundamental steps involved in transforming analog composite signals into digital component signals—necessary before we can even consider compression—we must ask, "how does signal noise interfere with these processes in a fundamental manner?"

Let us examine one of the worst-case situations: signals recorded in the field and played back into a digital transmission device for relay back to the station. First, the digital device must lock to the incoming signal's sync and burst. This is necessary to generate stable clocks for the digital conversion yet to come and permit the analog-to-digital converters to lock reliably to the proper levels within the signal. Given the ready availability of digital time base correctors, we tend to forget these fundamentals. However, MPEG-2 encoder designers have made different assumptions about the stability of incoming signals, which may have consequences further down the transmission chain.

While it is possible to build circuits that reliably perform the necessary sync and burst stabilization, time base correction, and analog-to-digital conversion, there may still be problems with the process. One is how signals with illegal digital values are processed: a simple clip or some sort of level compression? This is an example of a necessary design tradeoff made in building real products. A subtler problem is that the incoming signal's clocks in this case are not useful in generating the required high-precision MPEG clocks.

This raises the topic of design orientation of the MPEG-2 standard and a small fact of which many are not aware. MPEG-2 is a decoder-centric standard: the only requirements it places upon encoders is to generate a compliant transport stream (TS). There is no guidance in the standard for how to deal with noisy signals, signals with illegal levels (levels that would translate to illegal digital values), or other countless "real world" problems. The designer is left to her or his own devices, both literally and figuratively.

"A final class of noise, termed surface impairments, comes from sources such as vertical scratches present on filmstock that has been mistreated or from signal crosstalk. Once the noise is present in the signal it is very difficult to remove. Historically, techniques such as high-frequency roll-off have been employed, often doing more damage to the underlying pictures than the noise itself."

The MPEG-2 standard was designed to produce cheap decoders and it has certainly done that; but in the process, it also introduced numerous issues of interoperability. This subject is outside the scope of this discussion, but it is worth noting that two different, equally intelligent, design engineers can interpret the standards in two incompatible manners. While the worst of the interoperability problems are over, it will remain an issue for the user community to monitor.

Another major consequence of the standard being defined in decoder-centric terms is the use of high-precision clocks (27 MHz) and the ingenious manner in which periodic samples of these clocks are delivered to the decoders. This is not as simple as it sounds and has consequences throughout the transmission chain not realized by many in the RF end of the business.

These consequences translate into a requirement that all downstream equipment that changes the bitstream in any manner must make the correct and proper adjustments to these clock samples to compensate for the varying delays the device has introduced. This issue is included here because PCR jitter (to use the technical term) translates into signal noise at the decode end, just as digital jitter does in SDI video or AES/EBU audio signals.

MPEG-2 Compression Specifics

In examining the steps involved in the MPEG-2 compression process, additional ways that noise will negatively impact the results are discovered. Video compression lowers the necessary bandwidth for transmitting the picture by removing redundancy in both spatial and temporal domains. The process begins with a conversion from spatial to frequency domain via the discrete cosine transform (DCT). This transform works on square groups of pixels (blocks) and is fully reversible or lossless. Having transformed the picture from the spatial domain, the bandwidth may be further lowered by use of clever coding schemes such as variable-length coding (VLC) and run-length coding (RLC). These are also lossless techniques.

By themselves, DCT followed by VLC and RLC will not provide enough bandwidth reduction for our needs (for example, a high-definition picture in less than 18 Mbits/sec). A number of "lossy" techniques are employed to achieve that goal, including subsampling chroma information (4:2:0), horizontal subsampling, and variable length quantization of the DCT samples.

We can then examine a sequence of pictures for temporal redundancy and encode them into a group of pictures (GOP) in which redundant portions of each picture need not be retransmitted. Redundancy can even be borrowed from the future via bidirectional (or B-frame) coding. Thus the picture is transmitted once as an I-frame, then additional updates about what has moved are transmitted via predicted or P-frames and bidirectionally predicted B-frames. Remembering that the standard defines the design of the decoder, especially the sizes of its buffers, the encoder can then do its best to squeeze any redundant information from all pictures except the I-frames.

The insidious consequence of picture noise is that it adds hard-to-compress elements to the picture being coded. This translates into a loss of compression efficiency, and worse yet, for constant bit rate encoding, may require the encoder to make poor choices in encoding that translate into visible artifacts. For variable bit rate encoding in a statistical multiplex, the noise ends up robbing all of the pictures in the "pool" of encoders that comprise the multiplex of precious picture quality.

Picture Noise Specifics

Noise in video arises from a variety of sources. Most basic is the wideband or Gaussian noise that comes from pickup devices (camera tubes or CCDs), film grain, analog circuits, and so forth. For signals that have been transmitted over analog links, it is also common to see impulse noise, especially in satellite and microwave links, where it may range in intensity from a "sparkle" or two a minute to the "waterfall" of impulses seen in a satellite feed about to go into solar outage. Impulses may also come from inside a facility (e.g., the custodian plugging a vacuum cleaner into technical power).

A final class of noise, termed surface impairments, comes from sources such as vertical scratches present on filmstock that has been mistreated or from signal crosstalk. Once the noise is present in the signal it is very difficult to remove. Historically, techniques such as high-frequency roll-off have been employed, often doing more damage to the underlying pictures than the noise itself.

Noise Reduction for MPEG-2 Compression

The user community, especially the operators of direct-to-home satellite signals, has long been aware of the damage that noise does to MPEG-2 compression systems. Early adopters of statistical multiplexing technology, they began to employ "upstream" (of the compression) noise reduction equipment as soon as they realized the bandwidth consequences of noise (noisy channels took noticeably more bit rate than cleaner channels). This external equipment could never reach the level of noise reduction desired and frequently introduced artifacts or visual impairments. Additionally, it was expensive and took extra rack space and power.

Three Generations of MPEG-2 Noise Reduction Preprocessing

A nonintuitive result of the conversion to digital representation of the signals is to significantly reduce the complexity and improve the results of filter circuits. This, coupled with an understanding that certain filter aspects could more reliably be done in the frequency

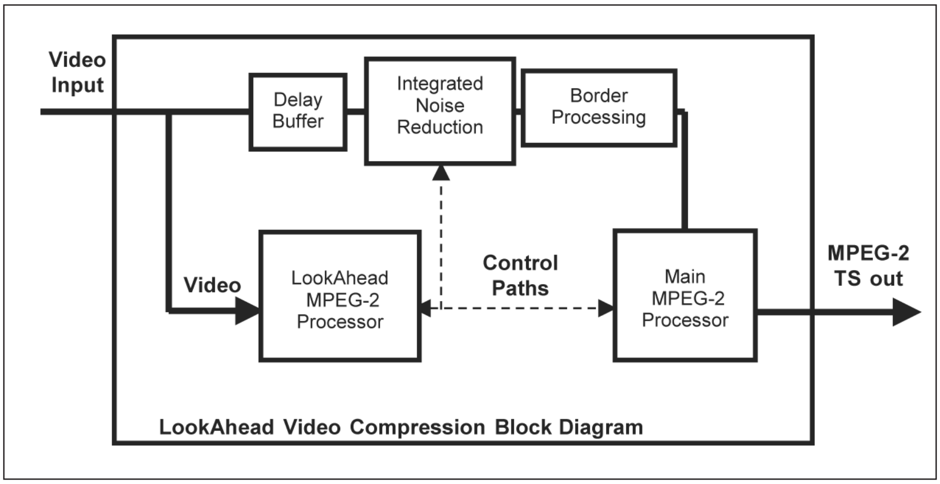

Figure 1. LookAhead two-pass architecture block diagram.

domain, led engineers to begin working with embedding the noise reduction as a preprocessing stage in MPEG-2 encoders.

First Generation Architecture

This first generation noise reduction preprocessor utilized some of the internal circuitry and instruction set of the C-Cube single-chip MPEG-2 encoder integrated circuits. The results were impressive enough to prompt the design and implementation of a second generation architecture filter set.

The filters were both temporal and spatial (the vertical filters required external hardware). Basic spatial filtering applies horizontal and vertical low-pass filtering within a frame to discard both noise and picture detail. This technique offers a tradeoff between artifacts and softer pictures, however the effect of soft pictures is easily seen.

The temporal low-pass noise reducing filter compares a pixel with other pixels at the same spatial location in different frames, offering an averaging process that cancels out random noise and is very effective with still content. In moving areas of a video scene, traditional temporal filtering will introduce artifacts that appear as ghosts or contours that follow moving objects and result from an object in one video frame being filtered with the displaced object from a previous frame. This is referred to as "motion blur."

Second Generation Architecture

Second generation filters were accompanied by a corresponding major change in the overall MPEG-2 encoder architecture: a move to two-pass encoding. This was facilitated by C-Cube's next generation single chip processor, the DVxpert-II. This processor added significant internal signal processing abilities.

The two-pass architecture, called LookAhead (Fig. 1), was an unusual system for the industry, since most encoder designers have been working diligently to minimize encoder latency. However, it was recognized that for emission (and distribution) encoding, latency does not matter. Having the ability to operate over a number of frames of video permits correct temporal and spatial filtering.

The design has other valuable characteristics. For example, in a statistical multiplexing system (with a closed loop architecture), the statistical multiplex controller has adequate time to efficiently allocate bit rates to all members of the pool, permitting the main encoder to make better coding decisions. The actual hardware implementation made use of an additional compression processor (as well as separate PLD-base coprocessors) to implement the noise reduction algorithms.

Third Generation Architecture

Second generation filters were quite successful; however, engineers realized that there was insufficient

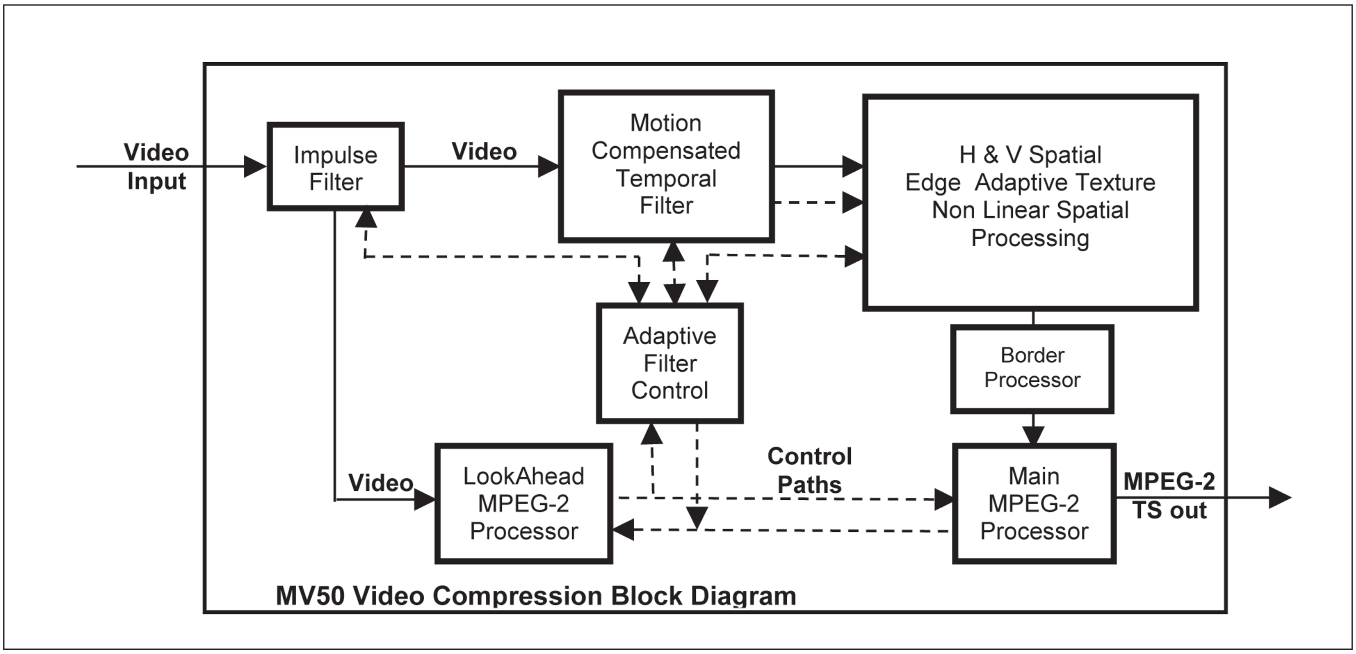

Figure 2. DiviCom MV50 encoder block diagram.

processing capability in the C-Cube processor, which after all, was designed to be primarily a compression, not a filter, engine. As a result, a new hardware platform for filtering was designed. It employs the DVXPERT-II processor to perform some specific tasks and augments it with a number of external devices that have more specific applications in digital filters.

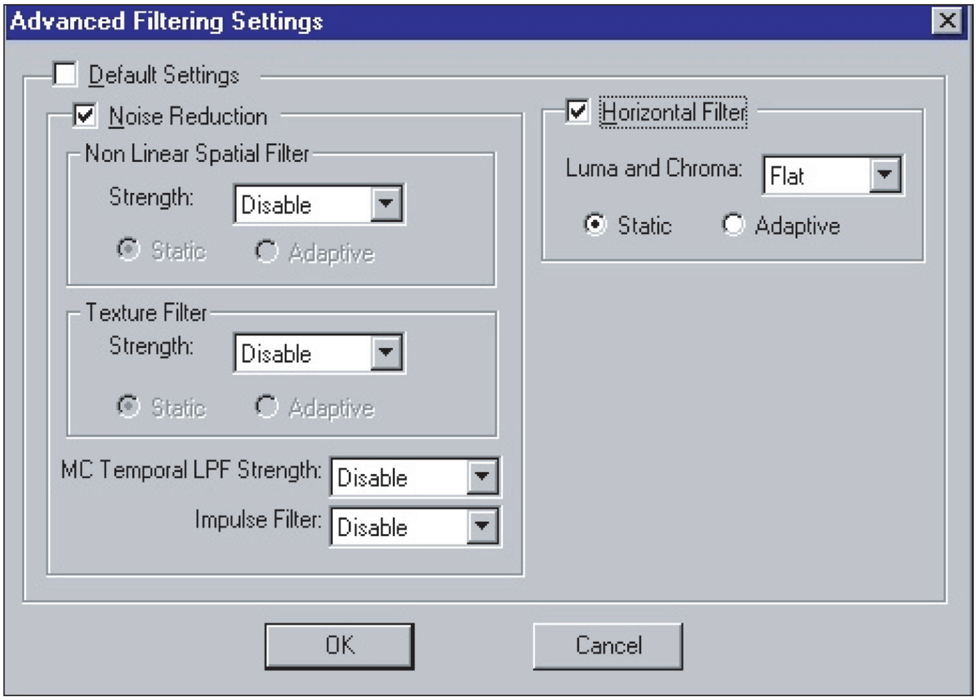

Figure 3. Filter settings provisioning GUI screen.

This is the DiviCom MPEG-2 encoder. The new hardware necessitated completely new filter algorithms, so considerable effort was expended to devise greatly enhanced filters, which are software based, for continued improvement in their performance. While these filters are in their first release of code, observers seem unanimous in their opinions that they are successful.

Figure 2 is a block diagram of the third generation architecture (implemented in the MV50 encoder). Note that this diagram does not attempt to show the delay buffer, which is present as it was in the second generation architecture.

Hardware Overview

Many of the hardware specifics have already been discussed, however, several aspects have not yet been considered. First, is the addition of an impulse filter on the input of both the LookAhead encoder and the embedded noise reduction filters, which were added in the third generation architecture. Impulse filters for video are simple in concept: compare several frames in succession, look for single pixels that have a major change in value only in the current frame, then safely substitute a revised pixel value based on the outer frames. The actual implementation is much more subtle; the filter must be able to interact with scene-change detection logic. A correctly functioning impulse filter relieves the downstream circuits of this extra noise.

Second, the frame delay buffer, while shown in Fig. 1 as a single direction device is actually a fully addressable memory that the filter circuits may interact with freely. As stressed earlier, this permits the filters to function correctly in both temporal and spatial domains. Finally, a number of filter operations can be more easily sensed or carried out in the frequency domain as opposed to the spatial domain. This is the reason the third C-Cube processor remains in the third generation architecture.

Filter Specifics

Second Generation Filter Set

Motion Compensated Temporal (Low Pass) Filter (MCTF). A detailed explanation of the algorithm is supplied in the description of the third generation filter below. The actual algorithm utilized is different in the second generation filter, but the results (pictorially) are the same.

Temporal Low Pass Filter. Complements the MCTF in removing random noise. LookAhead processing generates statistics that identify video frames that benefit from temporal filtering. This filter adapts similarly to MCTF to remove random noise while preserving detail.

Vertical Low Pass Filter. Applies interline processing to reduce vertical detail, which makes the video look softer but also makes it easier to encode. Edge detection is necessary to ensure edge detail is correctly preserved.

Temporal Median Filter. Removes impulse noise from the video. For scenes with large amounts of impulse noise, the temporal median filter can remove a portion of the impulses. This filter is not very good—it introduces subtle but visible artifacts.

Third Generation Filter Set

Motion Compensated Temporal Filter (MCTF). A technique used to remove/attenuate Gaussian noise from video content, it is much more powerful than temporal-only filtering. With MCTF, filtering is applied across several frames of a sequence, but rather than filtering a pixel with other pixels at the same spatial location in different frames (temporal-only filtering), filtering is applied along motion vectors. If an object within a sequence moves from frame to frame, motion estimation (ME) is used to track the direction and magnitude of the motion. Filtering is then applied using pixels that retain the same positions relative to an object moving within the frame sampling structure.

MCTF all but eliminates ghosting and trailing normally associated with temporal only filters. Random components caused by Gaussian-type noise are heavily suppressed, yet picture details are preserved. Artifacts are limited to below the threshold of visibility, while still providing a strong noise reduction capability.

Nonlinear Spatial Filter. Simple spatial filtering applies intraframe, two-dimensional, low-pass filtering to suppress high-frequency noise. The downside of this technique is less picture detail. Viewers generally have limited tolerance to the reduction in detail and edge definition; therefore basic spatial filtering should be applied sparingly. This filter employs a new adaptation technique that is edge preserving but allows significant filtering to remove noise that has Gaussian characteristics. This filter's operation is coupled to the MCTF. It is used when motion within a sequence is not well approximated by the translational block motion model, or when it has been determined that the input sequence contains a very strong Gaussian-type noise component.

"Impulse noise reduction requires a two-part process: detection and removal, with 99% of the necessary effort in the detection. The MPEG-2 encoder devotes more than two thousand million pixel comparisons per second to ascertain whether a pixel, or group of pixels, are single frame rogues."

Impulse Filter. Impulse noise is characterized as a spurious random effect that occurs for a single frame and is commonly seen as dots or scratches. This type of noise has a significant effect on perceived picture quality. Although the impulses do not waste as many bits as does random noise, they do tend to reduce the effectiveness of other noise reduction and compression mechanisms.

Impulse noise reduction requires a two-part process: detection and removal, with 99% of the necessary effort in the detection. The MPEG-2 encoder devotes more than two thousand million pixel comparisons per second to ascertain whether a pixel, or group of pixels, are single frame rogues. Once the impulses have been identified, it is relatively easy to substitute them using values based on spatiotemporal filtering operations.

Using its strongest settings, the impulse filter has proven very effective in cleaning up old film and archive material that contains significant damage. This setting allows the detection and filtering of arbitrary size impulsive spatial regions from an image sequence.

Adaptive Texture Filter. A spatiotemporal filter used to reduce bit rates, extending the encoder's performance. Filtering is performed on regions identified as "texture." The texture identification algorithm is designed so that edges defining objects, or those containing regions of texture, are left unchanged. The remaining textured areas are selectively softened in a technique similar to decimation to save bits. Filtering the textured regions can provide the benefits of improved coding efficiency without drawbacks. When this preprocessing technique is applied to sequences containing high spatial detail and motion, significant improvements in low bit rate encoding efficiency are realized.

Horizontal Filter. Permits a user to effectively utilize horizontal samples of "nonstandard" sizes (i.e., that fall between the "normal" values of 704 and 640). This filter supplies more "gentle" decimation values than may be obtained through the horizontal resolution control of the encoder.

Border Processing. Both generations of filter sets also employ "border processing" capabilities (as shown in the block diagrams). Here, the term border refers to the outermost pixels on each side of a frame. Processing provides the ability to save bits on these outermost pixels, which typically fall in the overscan area on consumer receivers. Since these blocks are normally overscanned, this provides an invisible manner to liberate a certain amount of coding bandwidth for use in the visible portions of the picture.

Software Overview

As indicated earlier, the hardware is fully programmable. C-Cube processors each run proprietary microcode for their internal processors and DSP coprocessors. Additional code is embedded for the control microprocessors, and, finally, logic masks for the programmable logic utilized. Each of these elements may be varied independently of the other. Since some of the filter algorithms may employ portions of several devices in all of the above categories, the ongoing development and debugging of these algorithms is a high priority task.

Control of each filter must also be provided to the end-user. This requires implementation in the network management system (NMS), which at present is named TheSys. An example of an NMS GUI control screen for the encoder filters is shown in Fig. 3. In each filter, the selectable settings are: Disable, Very Weak, Weak, Normal, Strong, and Very Strong.

Filter Control

Early noise reduction systems were intended for film transfer and offline operations, where an operator would manually optimize settings for the best results. In most head-end environments this is not practical and ease of application is a primary goal. Most operators require the system to automatically deliver an appropriate measure of noise reduction and to let clean content pass through with minimal adverse effect.

Overly strong or inappropriately applied filtering can degrade video quality. Some types of content are very sensitive to the application of noise reduction. For instance, skin tones and pictures that contain high motion and detail will likely show the first indications that noise reduction is too strong.

The goal of the MV50 filter control logic is to function fully in a "set and forget" manner. This implies that it must correctly sense the proper level of filtering to apply to each block. This is facilitated through use of the LookAhead architecture and the sophisticated third-generation filter algorithms outlined above.

THE AUTHORS

J. Patrick Waddell is a veteran of the broadcasting and performing arts industries with over 30 years experience in a variety of staff and freelance positions at a number of different facilities. As a second generation broadcaster, his primary area of expertise is digital video system implementation for broadcast television.

In 2000, Waddell joined Harmonic Inc., Sunnyvale, CA, where he works closely with engineering to define and deliver high-definition encoding products. He is responsible for video-quality performance as well as compliance with industry standards.

Waddell earned a BSEE degree from the University of California, Santa Barbara, and did graduate studies in technical production for live performance at San Jose State University.

Neil Brydon, currently leads the marketing team responsible for the development and delivery of Harmonic's MPEG-2 encoder family. When he first joined DiviCom (now part of Harmonic Inc., Sunnyvale, CA), he developed and managed testing processes for new video technologies at the ITV Technology Center in the U.K. Ultimately, the ITV DVB team was responsible for the design and implementation of the first digital terrestrial service broadcast facility in London

Brydon received his B.S. degree in electrical and electronic engineering from the Robert Gordon Institute of Technology, Aberdeen, Scotland.