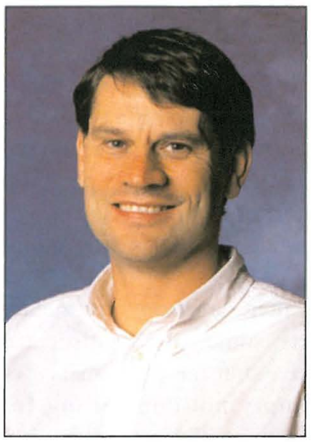
Saving Bits—The Impact of MCTF Enhanced Noise Reduction

By Neil Brydon



Motion Compensated Temporal Filtering (MCTF) has made a significant impact on the DTV industry by saving bits and allowing operators to deliver consistently higher quality video at lower rates than ever before. All content tends to contain some noise, the random and unwanted portion of a signal. Even modern facilities equipped with the latest digital production equipment will inevitably import noisy content unless the incoming material is thoroughly cleaned. Noise reduction (NR) and filtering can substantially improve the video received by a viewer if the right techniques are applied to remove noise prior to compression. Selectively removing noise is a challenge because it shares space with valuable picture detail. An ideal noise reduction process will allow powerful suppression of random noise while preserving clean video content. The major advantage of MCTF is its inherent ability to remove noise without introducing motion blur artifacts. This technique has been known for many years, but the costs associated with a high processing overhead has kept it from being commercially viable. This paper will explain how integration of the new MCTF technique with existing technologies has dramatically improved video-compression efficiency.

All content usually contains some noise, the random and unwanted portion of a signal. Modern facilities equipped with the latest digital production equipment will also encounter noise unless the incoming material is thoroughly cleaned. Some content such as archive material, film, and news feeds contain high noise levels.

The primary problem is that higher bit rates are required to encode noise and reduce picture quality. Selectively removing noise is difficult because it shares the same spectral space as valuable picture detail. However, at lower bit rates the signal complexity can be dramatically reduced, when noise is removed prior to compression. The benefit to the consumer is clearer, more consistent video quality.

A solution is needed to remove noise, yet leave clean content virtually untouched. Most NR solutions, however, are based largely on standard signal processing techniques, with the best of these deployed as external standalone solutions, which have little or no communication with the rest of the compression system. They cannot combine noise-reduction functions with the compression engine, a coupling that can significantly improve NR.

With the availability of a powerful MPEG-2 encoder silicon that can support noise-reduction processing, recent advances have been made, including the first practical application of a MCTF—a feature that provides powerful suppression of random noise while preserving clean video content. A key component in the NR process is the adaptive control of advanced filters. This paper will outline how "LookAhead" encoding statistics are used to provide forward-looking control of the filters, a process that ensures the appropriate level of filtering in the right places so content will benefit from an appropriate measure of NR.

DTV: A Solid Business Proposition

Digital television (DTV) has been a phenomenal success for obvious reasons: compared to analog, DTV allows more channels to be broadcast. The number of services depends on consumer satisfaction with the quality of service and efficiency of the compression technology. The function of a video compression encoder is to compress the video signal by extracting redundant data while minimizing removal of valuable picture information.

As bit rates drop, the compression process begins to degrade picture detail by dotting the picture with visible compression artifacts. These unwanted effects are caused by compression that applies coarse quantization and in doing so removes useful picture information.

Even though encoders are designed to minimize artifacts, compression and other processing tend to degrade the original picture quality. Consequently, the most advanced compression solu-

Presented at IBC 2000 in Amsterdam, The Netherlands, Sept. 8-12, 2000. Neil Brydon is with Harmonic Inc., Sunnyvale, CA 94089. Published by permission of the IBC.

tions have the highest value when the most compression is required. Normally the last mile to the consumer's home is where the bandwidth bottleneck is most constricted. It is also in this area that NR and pre-processing can have the greatest impact by reducing video complexity prior to compression.

Noise Wastes Money

Noise can be defined as the unwanted and spurious variation of a signal level. It can originate from many sources, including the camera that originally recorded the video, analog storage devices, transmission equipment, and other devices. Signal noise undermines video compression systems in two ways. Pictures look worse, so consumers are less likely to be satisfied with picture quality. More importantly, random noise is not easily compressed and a video compression system must spend bits to reproduce the unwanted noise component.

It is difficult to distinguish between noise and valuable picture detail since noise and picture share the same spectrum. The challenge of removing noise is compounded by the fact that it results from different processes and therefore has many characteristics. Some noise may not be strictly random in nature, but a compression system treats it as if it were. A common example is residual color subcarrier noise that arises from the composite color demodulation process.

Impulsive noise such as satellite link sparkles and dust marks from the film-to-television telecine transfer process is another common effect. This type of noise degrades picture quality but fortunately has a less wasteful impact than random noise. A negative effect is that it can interfere with other NR mechanisms. Reducing impulsive noise, therefore, helps reduce random noise.

Another type of signal noise, though rare in most modern head-end facilities, results when some signals suffer from time-base instability. Much more common is quantization noise, the unwanted but inevitable noise produced in the initial A/D process and, later, by compression. This noise will normally bear some correlation with the picture content. Visible examples are edge noise and block structure noise, produced when a picture is compressed with insufficient bits to maintain the quantization noise below a visible threshold. These problems reduce picture quality and are irreversible. The key, then, is to focus on techniques that minimize quantization effects.

Noise Inhibits Efficiency of Statistical Multiplexing Systems

The negative effects of noise are amplified in a statistical multiplexing system. In such an environment, a group of video encoders share a pool of precious bandwidth to optimize the quality of the delivered video. Video signal content and complexity vary with time: Some video scenes have little detail and are slow-moving, while others contain large amounts of detail and move quickly. More complex scenes are given a higher bit rate than simpler scenes in order to retain the same coded fidelity or quality. It is important to determine how many bits a particular frame requires, then allocate the most appropriate rate at exactly the right time.

Noise tends to spread complexity across the entire pool and undermine the system's efficiency. If the complexity across channels is consistently high due to noise, there is no flexibility to juggle the bits between channels for the greatest efficiency gains. With this complex system, channels are highly interdependent; reducing the noise on one channel will save bits across the entire pool. It is also important to prevent very noisy channels from overusing bits and diminishing the performance of the whole pool.

When noise is removed, video can be encoded at lower rates, releasing more bits into the pool for use by other channels. Streamlining bit use within the pool improves video quality and frees up bits for more channels.

Noise Reduction—Background

Noise reduction and filtering can substantially improve video reception if the right techniques are suitably applied. Inappropriate application of filtering and NR can harm video quality. Unfortunately, many NR solutions produce artifacts that outweigh the expected advantages, prompting many operators to question the effectiveness of NR techniques.

Effective NR is one of the most challenging video processing functions. There are no off-the-shelf silicon solutions. Developing an effective solution requires highly specialized DSP filter design techniques, powerful control algorithms, and thousands of hours of simulation and picture monitoring.

Even so, there are many proprietary NR products, some very sophisticated and expensive, but the performance of most is simply not compelling for many applications. For example, unsophisticated systems tend to suffer from a combination of poor filter design and limited adaptive control.

Skepticism over NR often causes users to prematurely abandon the technology or restrict it to only a few feeds. This may be short sighted, since all video feeds are likely to contain appreciable amounts of noise at certain times. Even if a studio or head-end has the latest SDI plant, operators are unlikely to have full control of all the content entering their facility. Ideally, an effective NR system should be good for all content, able to remove invisible noise from clean signals, and dramatically improve noisy feeds to provide viewers with the best possible picture quality.

Temporal Filtering—The Antidote to Random Noise

Pre-processing techniques are generally categorized as spatial or temporal filtering. Spatial filtering applies horizontal and vertical low-pass filtering to discard picture detail and noise. At low bit rates this technique can offer a tradeoff between artifacts and softer pictures. The effect of soft pictures is easily seen and consumers generally have limited tolerance to the application of spatial filtering. The other approach is the application of inter-frame or temporal filtering to reduce random noise.

Significant strides have been made to enhance NR by using advanced temporal filtering. Filtering systems must be able to handle a wide variety of video scene types. Spatially, video varies widely with sharp transitions across object edges, slowly varying changes across a person's face, texture in fabric or outdoor scenes, changes due to uneven illumination, and other variances. In the temporal domain, the majority of video scene changes are caused by object or camera motion. Illumination changes are usually rare and slow, and objects do not change temporally as much as they vary spatially. Thus, reasonably strong temporal filters can remove noise more effectively than similar spatial filters, without introducing annoying visual artifacts.

In moving areas of a video scene, traditional temporal filtering will introduce artifacts, which appear as ghosts or contours that follow moving objects. This results from an object in one video frame being filtered with another object from a previous frame. To avoid objectionable motion blurring, temporal filters must incorporate an adaptive mechanism to reduce its strength when motion is detected. Some high-end products use sophisticated strategies that adaptively blend spatial and temporal filtering to find the optimal compromise to the blurring issue, a technique that produces reasonable results with analog television video but is less successful in DTV compression encoding.

The motion adaptive approach is less than ideal for a compression system since this system is most strained when a signal contains high amounts of motion. A motion adaptive temporal filter is non-optimal, since more, not less, NR is needed.

If the right techniques are subtly applied, video quality can be substantially improved. By contrast, overused or inappropriately applied filtering and NR can degrade video quality. Some types of content are sensitive to the application of NR. For instance, skin tones and complex pictures containing high motion and detail will likely show the first indication that NR is too strong. A very irritating effect that the viewer will detest is motion adaptation that causes a filter to be continually enabled and disabled. This is liable to happen with a talking head moving in and out of the motion detection threshold, so that the face on which the viewer is focused appears to continually transition between sharp and soft.

Early NR systems were intended for film transfer and off-line operations in which an operator could manually tune for the best results. Today, skilled operators who manually select the optimum settings are often a luxury. Most operators desire one or two settings that they can confidently apply throughout their facility, a configuration that will have negligible effect on clean content, yet will apply an appropriate measure of NR as required.

The weakness in a temporal mechanism is the motion blurring effect. Motion adaptation alleviates the problem but is not the ideal solution. The key is providing all the benefits of traditional temporal filtering and eliminating the drawbacks. The solution is to account for motion while applying the filtering to produce more effective and consistent results.

MCTF: The Ultimate Temporal Filter

MCTF is a technique that eliminates the drawbacks of traditional adaptive temporal filters while retaining the benefits. 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 (but different times), it is applied along "motion trajectories." If an object within a sequence moves from frame to frame, block-based motion estimation (ME) is used to

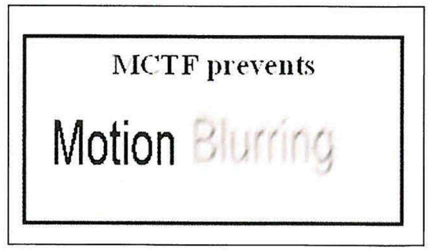


Figure 1. The unpleasant problem normally associated with standard temporal filtering.

track the direction and magnitude of the motion. Filtering is then applied along the motion trajectory using pixels that retain the same position relative to an object moving within the frame sampling structure (Fig. 1).

Motion compensation all but eliminates ghosting and trailing. As objects move, they are filtered only against displaced versions of themselves. Random components caused by noise are heavily suppressed, but picture details are preserved. With filtering adapted so artifacts are limited within the threshold of visibility, MCTF can apply stronger, more effective NR than ordinary temporal filters.

MCTF must also adapt to temporal changes that are not caused by translational motion such as crossfades. This technique is not new. However, until recently the processing overhead has been very high and the implementation cost-prohibitive. The introduction of low-cost MPEG encoding silicon with programmable ME capabilities has enabled the first viable application of MCTF. The ME operations for best compression and MCTF are not identical, and only programmable ME platforms lend themselves to useful MCTF.

An Integrated NR Architecture

Even the best filters are efficient only when effectively implemented. Figure 2 shows an architecture that tightly couples advanced NR with a dual-encoding strategy.

The MCTF filter has been carefully integrated into the LookAhead architecture to complement and enhance the existing set of filters. One encoder extracts statistics from the incoming video several frames ahead of the second encoder to preview the video content. The statistics are then used to assist the encoding process and act as control inputs to the adaptation mechanism, automatically applying the most appropriate level of filtering.

The strength of the individual filters and adaptation thresholds are controlled through the network management system. The adaptation mechanism, which automatically controls filters to suppress noise and preserve valuable picture detail such as textures and edges, enables users to confidently select a single setting that is effective with noisy content, yet still allows clean material to pass through virtually transparently. The ease of control reduces reliance on skilled operators to fine-tune and maintain system efficiency.

Historically, noise reduction and video processing equipment have required highly skilled operators to manually fine tune a system to match the right parameters to the content. Once the basic control parameters have been set through a user-friendly UI, the system automatically aims to apply the correct filters to get optimum results. NMS offers access to the filter controls that have been designed around template that make it easy to select useful settings.

LookAhead Architecture

The LookAhead architecture harnesses the power of three processors in a combination that enhances both the encoding and noise reduction, as shown in Fig. 2.

One MPEG-2 processor is used to extract statistics from the incoming video almost 1 sec ahead of the second encoding processor, while a third processor provides the tools for motion-compensated noise reduction, making video easier to encode. This proactive combination of video analysis and noise reduction helps the encoder make better encoding decisions with available buffer and filtering resources to avoid artifacts, thereby producing better pictures at lower bit rates.

Although this processing adds latency to the overall encoding process, the majority of applications benefit from greater efficiency despite the delay.

Improving the Encoding Process

Video has many characteristics: fast or slow action, film-originated, stills, scene cuts, dissolves, special effects, and so on. Algorithms, typically designed for general-purpose use, can be developed to optimize the encoding to match the incoming content. LookAhead provides effective techniques to eliminate artifacts as follows:

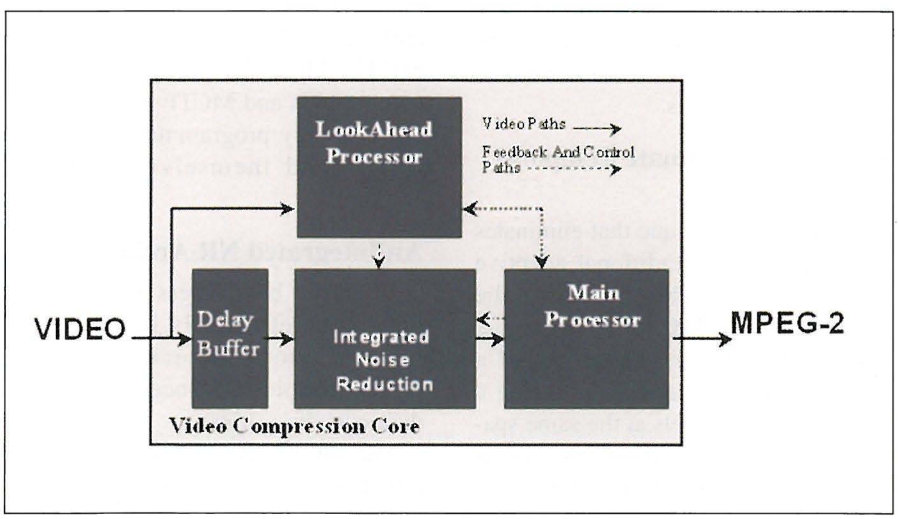


Figure 2. The architecture of the video-processing core at the heart of the MPEG-2 encoder. The noise reduction is tightly coupled with the compression core and the LookAhead processor feeds control information forward to both the main processor and noise reduction.

- Maximizes the use of standardized MPEG-2 resources.
- Improves encoder decisions: selects block-encoding and scan modes.
- Customizes motion estimation.
- Adapts for brightness: accounts for human visual response to overall scene brightness that other encoders handle poorly.
- Improves I-, P-, and B-frame rate control.
- Eliminates short-term variations in quality ("I-frame beating").
- Optimizes the use of available bandwidth within the MPEG picture structure.
- Applies appropriate noise reduction: adapting to each scene, thus removing noise while retaining detail, edge definition, and sharpness.

Better Encoder Decisions

The system uses statistics to identify scene types so it can apply the proper encoding strategy to produce maximum film, video, and text quality in the following ways:

- When the encoder recognizes scene cuts and other complex effects, it overrides the regular group of pictures (GOP) structure and places I-frames at the optimal moment.
- When the encoder detects still video scenes and knows they exist for the next N frames, it encodes the still scene with proportionally larger I-frames to attain higher resolution and greater clarity, while saving bits with smaller P- and B-frames.

For example, artifacts are sometimes visible if they are using different techniques in adjacent macroblocks, resulting in a noticeable discontinuity. In this case, the encoder double-checks decisions that may be inconsistent within adjacent blocks.

Brightness Adaptation

When encoding at low rates, complex content often reveals artifacts in dark areas based on the way the human eye responds to darkness. The encoder matches the human eye response with LookAhead statistics used to identify darker regions, then adds more bits to suppress visible artifacts in those areas.

Optimal Motion Estimation

Motion estimation enables the system to extract the temporal redundancy from a sequence. The MPEG-2 processors use a multilayer, hierarchical search methodology that yields precise matches at half-Pel resolution. This provides a greater search range than alternative methods for producing exceptionally accurate motion vectors.

At low bit rates, the number of bits required to convey motion vectors becomes more significant. When two vectors are otherwise equivalent, the encoder selects the vector that requires the least number of bits to encode. This technique filters the motion vectors to replace inconsistent vectors when there are more efficient methods.

Further, the encoder analyzes the incoming video signal and modifies its motion estimation to further improve the results. For each video frame, the encoder varies where it spends its motion estimation processing power based on the amount of motion in the frame, and on the presence or absence of special effects.

Maximal Use of the MPEG-2 Buffer Resource

The MPEG-2 encoder must manage the buffer of the MPEG-2 standard target decoder (STD), which provides a buffer model that forces the encoder to tightly control output bit rates. This buffer model anchors the MPEG-2 specification around a standard decoding foundation ensuring interoperability. With a predefined buffer size, the encoder must ensure that the buffer never overflows or runs short of data (underflow).

The LookAhead technique is used to provide early analysis of video scenes so that the main processor can reduce safety margins and use the buffer more efficiently. In effect, the system can anticipate complex video sections, take precautions, and then prepare to suppress artifacts when the video sections arrive at the main processor.

Border Processing

While watching a standard television set, viewers cannot typically see the picture border. Further, picture borders often contain considerable noise. It is used to apply unseen and stronger compression on the picture's boundaries.

Applying Generous Processing Power

A wealth of processing power is used to assist various aspects of the encoding process by thoroughly analyzing options within the MPEG-2 standard to make the best decisions. One example is the adaptive scan mode, where the encoder applies zigzag and alternate scans, compares results, and selects the most efficient scan mode on a frame-by-frame basis. Another is the latest encoder, which applies more than two billion comparisons/sec in order to detect pixels that represent spurious impulses.

Video quality relates directly to the efficiency of the algorithms and the number of services that these algorithms support. Evaluating video quality is important, but unfortunately, there are no metrics that can consistently characterize the measurement of video quality as it is perceived by a viewer. The exceptional ability of the human eye to judge video quality is difficult to model. Although objective measurement tools are available, the results are often inconsistent and can produce misleading conclusions. The most astute operators rely heavily on human

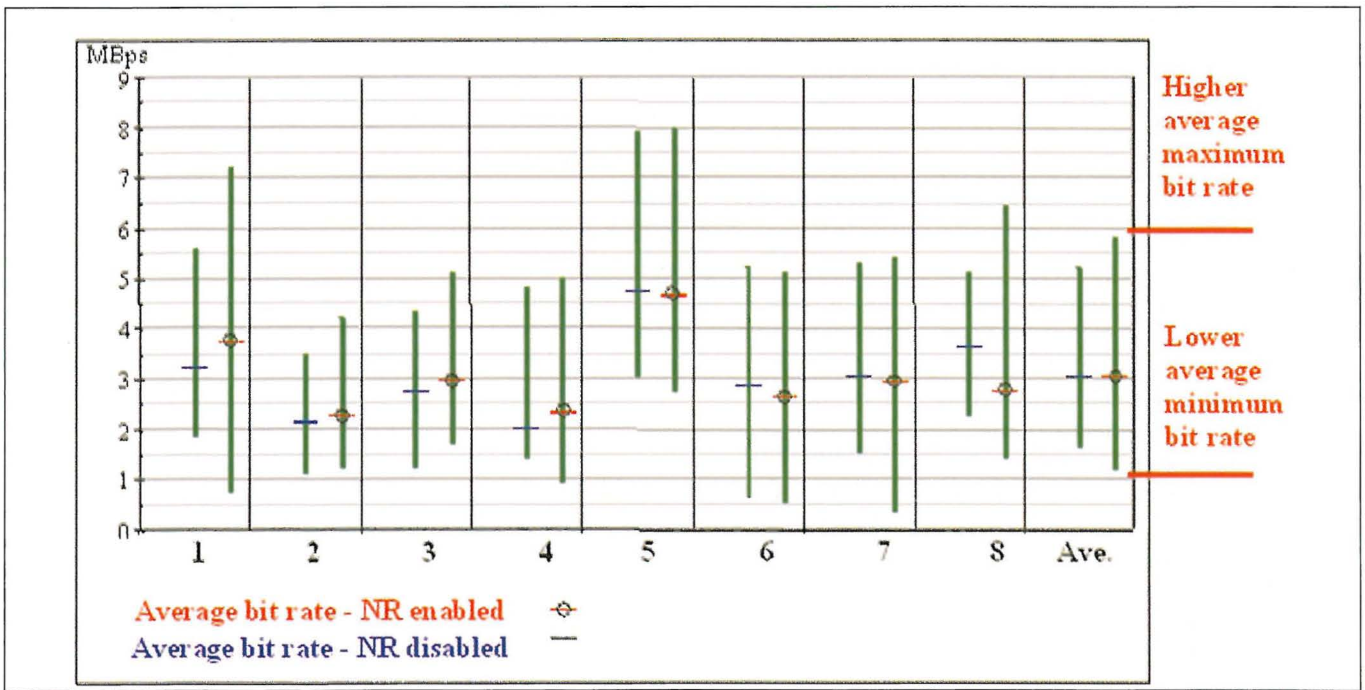


Figure 3. A bit rate allocation test applied to an eight-channel statistical test pool.

eye testing and side-by-side comparisons.

Judging quality when changing parameters results in a tradeoff between resolution and artifacts. Whether a change is an improvement or not depends on the video source, the bit rate, and the viewer's preference. The most refined systems that perform at low bit rates provide transition and background quality, ability to deal with noise, and fast action motion.

Early trials of the system were extremely encouraging, but also highlighted the challenge involved in ensuring optimal internal code settings for all content, bit rates, resolutions, and formats. Further, users frequently ask for recommended standard settings, which is a challenge, given that preferences can vary considerably. Some users are strongly biased against softening effects, while others are happy to tolerate slight softening for reduced artifacts. The solution is a system that could easily be configured to allow the user to tune it to match expectations.

It has always been clear that motion-compensated NR would bring significant benefits. However, the magnitude of the benefits has significantly exceeded expectations. Objectively, quantifying the benefits precisely is hard, as results depend on the source content and other factors. Certainly the greatest benefits will be accrued at facilities with the noisiest sources or on services with noisy old films or material archived on analog format.

Some larger-than-expected gains can be attributed to imperfections with the composite decoding process, which will always tend to leave a residue of the color sub-carrier. The MCTF process is an effective countermeasure, preventing bits from being wasted on this noise component in the signal.

Constant bit rate applications have also benefited dramatically by removing noise from the signal, significantly reducing the chance of overload and subsequent artifacts. This allows the operator to run the channel at a lower bit rate or improve the outgoing picture quality.

In a statistical multiplexing system, reducing noise in any channel will enhance the overall efficiency by allowing additional bits to be applied to more complex channels.

Figure 3 illustrates a bit rate allocation test applied to an eight-channel statistical multiplexing test pool. This was performed repetitively, using a server to play out loops of real content. The pool was set to run at an average rate of 3 Mbits/sec and a resolution of 704 pixels/line. Two test samples can be compared: one taken with NR enabled, the other disabled. The NR was set at a strength that would have little effect on clean material. Samples with NR enabled have larger dynamic range, and the minimum bit rate allocations are lower. Channel one, for instance, has had its allocation drop as low as 0.8 Mbits/sec, compared to 1.8 Mbits/sec without NR. A consequence of reducing bits spent on noise is that the dynamic range of bit rate allocation is greater, and illustrated by the average bar, which also indicates the difference between the average minimum rate of 0.42 Mbits/sec. This implies that an average of about 0.42 Mbits/sec was spent on noise in each channel. In an eight-channel system this equates to an extra 3.3 Mbits/sec of useful capacity and represents a saving of approximately 20% pool bandwidth.

Development Potential

While the platform has proven to be very good for advanced temporal filtering, it is very clear that further innovation with spatial filtering can also be deployed to reinforce the overall performance.

Conclusion

The introduction of MCTF has already had an impact on the broadcast industry, by saving bits and allowing operators to broadcast higher quality video at the lowest rates. Further developments can be expected.

A new NR feature was introduced along with significant algorithmic enhancements. First-rate encoding at low bit rates requires a combination of refined algorithms and powerful noise reduction. LookAhead processing is a technique that dedicates an MPEG-2 processor to fully analyze incoming video content before the video meets the main processor. With this technique, the encoder is able to proactively apply the most efficient encoding and noise reduction strategies.

An MPEG-2 encoder is a complex mathematical model developed to match human eye responses. MPEG-2 standards, developed from a decoding perspective, offer the freedom to build more efficient performance on the encoding side.

Acknowledgment

The author would like to acknowledge Paul Haskell, director of advanced developments, for his leadership during development of this new technology and for his assistance in writing this article.

THE AUTHOR

Neil Brydon currently leads the marketing team responsible for development and delivery of Harmonic's MPEG-2 encoder family. When he first joined DiviCom (now part of Harmonic Inc.) in 1997, he focused on video compression and pre-processing, a crucial feature of successful digital television system design.

Prior to joining Harmonic, Brydon developed and managed evaluation and selection processes of new video technologies for the ITV Technology Center in the U.K. Ultimately, the ITV DVB team was responsible for the design and implementation of the world's first digital terrestrial service.

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