

Extron

Video Compression Codecs

ABSTRACT

Video compression codecs have been developed to address a wide variety of applications in AV systems. They enable video conferencing, streaming video across computer networks, delivering 4K video over an existing infrastructure, extending the capabilities of a display link, and storing a full-length motion picture in HD quality on a Blu-ray™ Disc. This paper provides an overview of codecs used in AV applications.

Table of Contents

WHY IS COMPRESSION NEEDED?	2
WHAT IS COMPRESSION?	2
COMPRESSION METRICS AND IMAGE QUALITY	3
Objective Measurements of Image Quality	3
Subjective Measurements of Image Quality	4
Latency	4
COMPRESSION METHODS	4
VIDEO COMPRESSION CODECS	5
Motion JPEG	5
JPEG 2000	5
Extron PURE3® Codec	6
MPEG-2	6
H.264 / MPEG-4 AVC	7
H.265 High Efficiency Video Codec	8
VP8 and VP9 Video Codecs	8
VESA Display Stream Compression	10
VC-2 Video Codec	11
CONCLUSION	11

white paper

Why is compression needed?

The transition from analog to digital video technologies has introduced high-speed, multi-gigabit digital signals into AV systems. For example, a 1920x1200 video signal updating at 60 fps with a color depth of 24 bits per pixel requires a data rate of at least 3.3 Gbps, not including the blanking period or overhead. Emerging 4K and Ultra HD video standards are extending data rates well beyond 10 Gbps.

The demand for higher resolutions is outpacing the ability to extend, switch, and distribute multi-gigabit video signals across an AV system. Additionally, many applications require recording, storing, and sending video across computer networks. The ability to reduce the amount of data needed to reproduce video saves storage space and enables higher resolution video signals to be sent across existing infrastructures and streamed over computer networks as well as the Internet.

What is compression?

Compression is the process of reducing the amount of data required to reproduce a video signal for transmission and storage as shown in Figure 1. **Reconstruction** is the process of recovering the video signal from the compressed data. The process of compressing and reconstructing a video image can be classified as lossless, lossy, or visually lossless.

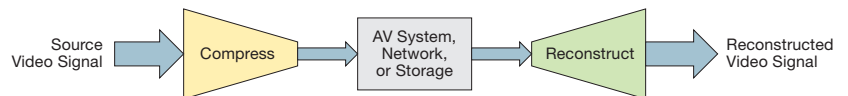


Figure 1: Video Compression and Reconstruction

Compression is considered **mathematically lossless** when 100% of the original information is recovered during reconstruction of the signal. Lossless compression achieves pixel-for-pixel performance, recovering the video in its original resolution, color depth, chroma sampling, and frame rate. It can be used for signal extension when the transmission data rate is slightly lower than the data rate of the raw video. Lossless compression is also commonly used for reducing computer file sizes into the ZIP format as well as still image files using PNG or TIF formats. It guarantees full integrity of the original data, but does not provide enough data reduction for some applications, such as streaming full motion video over a computer network or storing a feature length motion picture on a Blu-ray Disc.

Lossy compression reduces the amount of data such that some information is permanently lost through the compression process. It is used when the data reduction required exceeds the capability or cost of lossless compression. Lossy methods are typically implemented with low bit rate as a priority over image quality. In such cases,

lossy images are acknowledged to be of lower quality than the original but is acceptable for the intended application.

Visually lossless compression yields images that appear to be indistinguishable from the original to a trained observer. It uses algorithms that take advantage of an individual's visual perception abilities to discard or reduce data for parts of the image considered imperceptible or redundant to the human eye.

Compression Metrics and Image Quality

Compression enables high resolution signals to be transmitted over channels with limited bandwidth. Multiple compression methods are available for a variety of applications, and make tradeoffs between image quality, bandwidth, and latency to achieve the desired results. Image quality refers to how well a reconstructed image represents the original uncompressed image for the target application. Multiple factors affect image quality, including the compression method, compression ratio, the image content, and human visual perception.

Objective Measurements of Image Quality

The most common objective measurements are **mean squared error - MSE** and **peak signal-to-noise ratio – PSNR**. MSE is computed by averaging the squared intensity differences between the original and the compressed or impaired image, which is then used to calculate PSNR. PSNR is a ratio of the maximum intensity level to the difference value expressed in decibels – dB. A PSNR value of 35 dB or greater is typically considered a good result. Objective measurements can be used to evaluate a wide range of image distortion, including contrast changes, luminance shift, Gaussian noise, compression artifacts, blurring, scaling artifacts, spatial shifts, and rotation.

Objective measurements used for video quality have physical meaning, are repeatable, and do not depend on human judgment. They are often used for debugging a design, quality assurance testing, and manufacturing testing. However, objective measurements such as MSE and PSNR do not take into account human visual perception. These measurements treat all image distortion the same, whereas human vision is more sensitive to certain types of distortion, such as blurring, and less sensitive to other types, such as a contrast shift, as shown in Figure 2. Therefore, MSE and PSNR are inadequate for evaluating image quality from the perspective of the viewer.



Figure 2: Comparison of a Contrast Shift to Blurring

Methods to measure image quality have been developed that attempt to model perceptual characteristics of human vision. A commonly used method is the **structural similarity – SSIM** index. It is based upon the concept that the human vision system extracts structural information from visual scenes. Therefore, visual structure is an important element to consider when evaluating image fidelity. The SSIM model is designed to compare the structure, as well as luminance and contrast, of an impaired image to the reference image. SSIM values are calculated on a range from 0 to +1, with +1 indicating an exact match. SSIM and similar models have shown a good correlation to human visual perception and are useful for examining the quality of lossy images. However, they are not effective for evaluating visually lossless images.

Subjective Measurements of Image Quality

Subjective testing is considered the preferred method for evaluating image quality from the perspective of the viewer and assessing visually lossless images. ISO/IEC 29170-2:2015 describes a procedure of evaluating image compression codecs uses the concept of **just noticeable difference – JND**. Test subjects are shown a series of original and compressed images, and then must identify the image with impairments in a forced-choice procedure. A JND score is assigned based upon the probability that a test subject correctly identifies the reference image as the higher quality image. If there is no difference between each set of images, the number of correct responses is very close to 50%, since test subjects are merely guessing, which corresponds to a JND score of zero. At the point where impairments are barely distinguishable, the number of correct responses starts to increase. **Visually lossless compression** is defined at the point where the number of correct responses is no more than 75%, corresponding to a JND score of one. Above 75%, image compression is considered lossy. As the image quality of lossy compression diminishes, the number of correct responses approaches 100%.

Latency

Latency is the delay from image source to its destination within an AV system. Codecs often employ complex algorithms that achieve higher compression ratios at the expense of processing time, resulting in several frames of delay. Since the amount of latency a system can tolerate depends on the application, it is important to choose an appropriate compression method. For example, television broadcast, video-on-demand, and Blu-ray applications can tolerate several seconds of latency, but a real-time application such as video conferencing or drone piloting demands extremely low latency.

Compression Methods

Compression methods reduce the amount of data required to represent still images and full motion video. Simple techniques achieve a modest reduction in the signal data rate for sending high resolution video images over an infrastructure designed for

lower resolutions. Advanced compression techniques are available that take advantage of redundant information within the image data for a more significant reduction in data while maintaining the quality of the image. Two commonly used approaches to compressing motion images are **spatial compression** and **temporal compression**. Spatial compression reduces the amount of data needed to describe a single frame or still image. Therefore, it is also referred to as **intra-frame compression**. Temporal compression reduces the amount of data needed to describe a series of frames in a sequence by taking advantage of the redundancy that occurs between frames. Therefore, it is also referred to as **inter-frame compression**.

For more information on compression methods and technologies, please refer to the Extron whitepaper *Video Compression Technologies*. It includes an in-depth discussion on spatial and temporal compression, transform coding, and other important technical aspects of image compression.

Video Compression Codecs

Motion JPEG

Motion JPEG – M-JPEG was developed by the Joint Photographic Experts Group – JPEG. It is derived from the JPEG standard for still image compression that was created in 1992. Each frame is compressed separately as a still image without the use of temporal or inter-frame compression. M-JPEG is based upon the discrete cosine transform - DCT and operates on 8x8 blocks of pixels. It is less efficient than modern codecs, producing lossy images at compression ratios up to 20:1.

M-JPEG has been widely used for video, both in semi-professional applications and in the security surveillance field. Because each frame is encoded separately, there is no frame interdependence, which facilitates video editing and random access. Consumer and semi-professional digital video operate on the same basis, but are not fully JPEG compliant.

JPEG 2000

JPEG 2000 was created to supersede the original JPEG standard. It is based on the discrete wavelet transform - DWT to achieve more efficient compression, and employs variable pixel block or tile sizes as opposed to standardized blocks. The tile size can be defined differently for each application or image within the source image creation application. Typical compression ratios are 2:1 to 3:1 for mathematically lossless images, which makes it ideal for applications requiring the preservation of the original image, such as medical, military, and archival imagery. It can also achieve visually lossless images at typical ratios up to 5:1.

JPEG 2000 has been chosen by the Digital Cinema Initiatives - DCI group within the motion picture industry as the preferred method of distributing digital cinema first-run

features to theaters. Digitally encoded films are stored, and then played back from hard disk-based storage on specialized video players where high data transfer rates can be supported internal to the player. JPEG 2000 has also been used in video streaming products.

Extron PURE3® Codec

The Extron PURE3® Codec is a unique DWT-based compression technology that exceeds many of the performance characteristics of other compression systems. The DWT, which is also used in the JPEG 2000 codec, provides an effective coding method for both natural and computer graphic images. The wavelet transform is carried out with the highest possible input image quality, maintaining 4:4:4 chroma sampling information so that image detail is not lost in the transform process. PURE3 further achieves low bit rates with the addition of a single-pass temporal compression scheme. The use of both intra-frame and inter-frame compression enables the PURE3 to exceed the performance of the JPEG 2000 codec, while maintaining excellent image quality.

The codec offers real-time performance and low latency, making it ideal for live event streaming, distance collaboration, and recording applications. Unlike codecs designed for full motion video, the PURE3 is designed to deliver high resolution video and graphics with minimal latency. PURE3 achieves low latency image compression by processing images in a single pass. This contrasts with motion-based codecs that are not designed for real-time applications, such as MPEG encoders used in video-on-demand applications, in which multiple passes or expanding the Group of Pictures - GOP are required to achieve a target bit rate. The PURE3 codec excels at streaming high resolution video or graphic data in real-time, mission critical applications where maintaining original source quality is essential to the application.

PURE3 was designed specifically for AV streaming to optimize delivery of high resolution video and graphics with minimal latency within the bit rates acceptable to commonly used IP networks. Switched IP networks often produce unpredictable errors, including out-of-order, dropped packets, or bit errors. The PURE3 codec includes an error concealment system, which makes it highly immune to network errors. Picture data from previous frames are held at the decoder as a reference. If packet errors are experienced, this data is used to conceal the missing information.

MPEG-2

MPEG-2 served as the basis for digital satellite and cable, HDTV broadcasts, and DVD-Video. It combines both temporal compression using inter-frame prediction and motion estimation along with the DCT performed on 8x8 blocks of pixels. In its most common implementation, MPEG-2 supported both progressive and interlaced standard definition video (e.g. 720x480 at 30 fps) with bit rates up to 15 Mbps. In practice, excellent results were obtained using 2 to 6 Mbps for standard definition video. “Higher

Level” variants of MPEG-2 were applied to high definition video at resolutions up to 1920×1080 and bit rates from 19 to 80 Mbps.

H.264 / MPEG-4 AVC

H.264 dramatically increased the efficiency of motion compression over MPEG-2 in many ways. It added intra-frame prediction for encoding I-frames in a method that reduces bandwidth while maintaining quality. Intra-frame prediction takes advantage of the spatial redundancy within a frame. It checks the macroblocks to the left and above the macroblock currently being encoded to determine if there is a close match. Once it finds a match, intra-frame prediction uses a vector to point to it and encode the difference as illustrated in Figure 3. For example, a frame showing a background with a large surface area benefits from intra-frame prediction by using the spatial redundancy to increase compression.

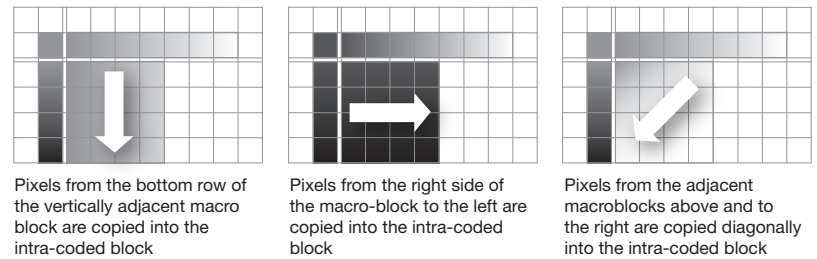


Figure 3: Intra-frame Prediction in H.264 Codec

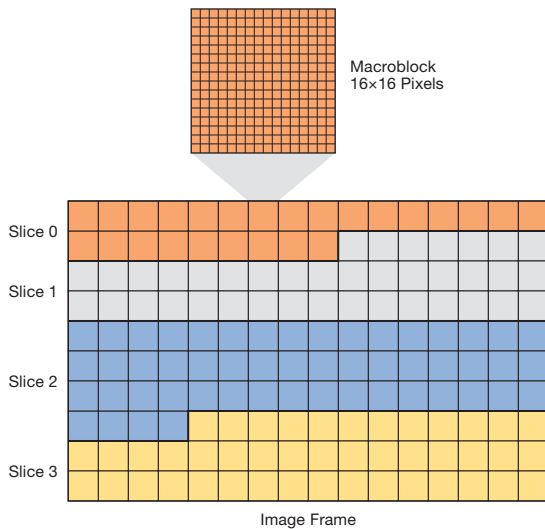
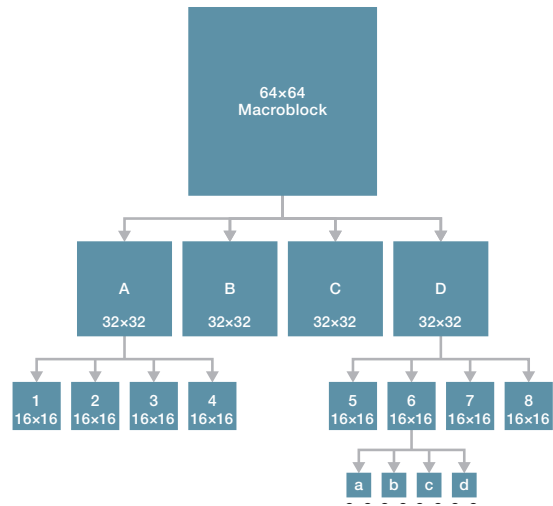


Figure 4: H.264 Divides Frames into Slices

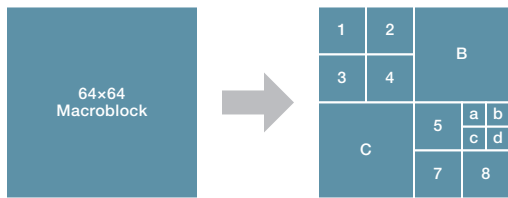
H.264 divides frames into regions called slices, each of which contains a collection of macroblocks as shown in Figure 4. There are three types of slices that are used for coding. I-slices contain only intra-coded macroblocks. P-slices contain both intra-coded macroblocks and inter-prediction macroblocks that uses one motion vector. The motion vector references a previously coded I-slice or P-slice. B-slices contain intra, inter, and bidirectional inter-predictive coded macroblocks with two motion vectors. The motion vectors reference I- or P-slices that occur in prior and subsequent frames of the video sequence. Each slice is coded independently of other slices within the frame. The use of slices can speed up compression through parallel processing but does not allow motion detection across slice boundaries..

H.264 offers motion compensation by dividing the 16x16 pixel macroblocks into smaller subblocks and tracking their individual motions. This allows P and B frame prediction based on up to 16 successive or preceding frames for reference, rather than just a single past or successive frame.

H.264 is a popular method for streaming AV signals in consumer and professional applications. It is used where high compression ratios are required at the cost of image



(a) Coding Tree Unit Applied to 64x64 Macroblock.



(b) Resulting Block Structure

Figure 5: H.265 Coding Tree Unit Partitions Image into Various Block Sizes

quality, including Blu-ray Disc, digital satellite broadcasting, video-on-demand, Web or network video streaming, and videoconferencing. It offers the ability to encode to bit rates that range from 64 Kbps to 240 Mbps, supporting many different classes of applications and achieving lossy images at ratios from 50:1 to 100:1.

H.265 High Efficiency Video Codec

In January 2013, the High Efficiency Video Codec – HEVC, also known as H.265, was approved by the ISO/IEC and ITU Joint Collaborative Team on Video Coding – JVT-VC. The primary objective for developing H.265 was to improve the bit rate efficiency of H.264 by 50%. It also supports use with video resolutions up to 8K – 8192x4320 and provides methods for applying parallel processing. HEVC encoder designs can trade off computational complexity, compression rate, robustness to errors, and encoding delay, depending upon the target application. The 8x8 macroblock structure used by H.264 has been replaced with coding tree units – CTU which can be applied to larger block sizes up to 64x64 as shown in Figure 5. It allows the picture to be partitioned into various structural sizes, producing greater coding efficiency. This, among other improvements to coding and compression techniques, produces bit rates that are 40% to 50% lower than those produced using H.264. Similar to H.264, specifications exist for multiple profiles and levels that provide different quality, bit rate, and coding complexity within the standard. There are two Tiers, Main and High, added to the levels that specify the maximum bit rate. For example, Level 4 Main Tier has a maximum bit rate of 12 Mbps, but High Tier has a maximum bit rate of 30 Mbps. Provisions exist for extensions to the defined profiles to support improvements to the standard and support future application requirements.

VP8 and VP9 Video Codecs

VP8 and VP9 are royalty free video compression formats owned by Google. VP8 is a compression format that was acquired by Google with its 2010 purchase of On2 Technologies. It combines inter-frame prediction and intra-frame prediction with a DCT-based encoder.

Inter-frame prediction within the VP8 codec uses three types of reference frames and prediction frames or P-frames, but does not use bidirectional prediction or B-frames. The absence of B-frames enables the compressed video to maintain the proper frame order, eliminating the need to reorder frames in the decoder. The three types of reference frames include an intra frame or I-frame, a golden frame, and an alternate reference frame. An I-frame is the most recent key frame and contains full frame information, and a golden frame is an earlier reference frame from the distant past, and an alternate reference frame is constructed from multiple source frames. Golden frames and alternate reference frames are not displayed. They are used to help improve compression efficiency, compensating for the lack of B-frames.

Intra-frame prediction within the VP8 codec operates on 4x4 luma, 16x16 luma, and 8x8 chroma blocks. There are four different prediction modes common to all block types plus six additional modes used on 4x4 luma blocks. The four common modes are DC, Horizontal, Vertical and True Motion. DC prediction fills the block with a single value that is the average of the row above and the column to the left. Horizontal prediction fills each column with a copy of the column to the left of the block and vertical prediction fills each row with a copy of the row above the block. True Motion prediction uses the column to the left, the row above, and the pixel that is above and to the left of the block to predict pixel values as shown in Figure 6. The six additional modes are prediction in different directions.

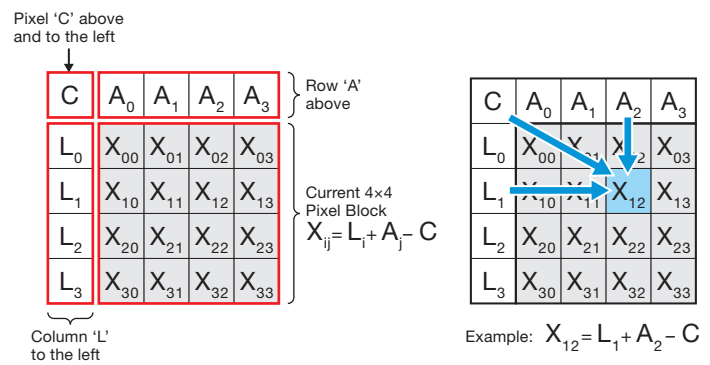


Figure 6: TrueMotion Prediction in the VP8 Codec

Since the VP8 uses blocks and the DCT, blocking artifacts are reduced by an adaptive filter in the decoder. VP8 is supported by Opera, Firefox, and Chrome web browsers. Despite being royalty free, VP8 has not found much commercial success. Head-to-head comparisons favor the image quality of the H.264 codec over VP8.

VP9 is a compression codec developed by Google as the successor to VP8. Its goals were to reduce the bit rate by 50% compared to VP8 at the same video quality, and to offer better compression efficiency than H.265/HEVC. One of the improvements over VP8 is the use of superblocks of 64x64 pixels with a quadtree coding structure. Each superblock can be recursively divided into square or rectangular blocks. It also uses a hybrid transform using the DCT and the asymmetric discrete sine transform – ADST. The ADST uses a sine basis function that diminishes at known block boundaries reducing the need for a deblocking filter in the decoder. The result is that the VP9 offers significant improvements over VP8 and comparable performance when compared to H.265/HEVC. The codec is supported in Opera, Chrome, and Firefox Web browsers. Youtube also uses VP9 to enable efficient streaming of 4K content and reduce the bandwidth of HD content.

VESA Display Stream Compression

VESA Display Stream Compression – DSC is an open standard developed for a video compression method to be used within a display link such as DisplayPort, enabling built-in support within sources and displays. It is currently optional for use in Embedded DisplayPort – eDP v1.4 and MIPI DSI v1.2. It is open for use in other display interface standards, as well. DSC provides visually lossless compression with extremely low latency for real-time performance. Supported video formats include RGB or YCbCr with chroma sampling of 4:4:4 and 4:2:2 with 8, 10, or 12 bit color depth. For desktop applications, DSC enables DisplayPort 1.3 to support 8K video, extending the life of the interface. For mobile applications, it reduces power consumption and lowers cost by enabling a lower bandwidth channel to support higher resolution video. Typical compression ratios range from 2:1 to 3:1 with visually lossless performance.

A DSC encoder is based upon Differential Pulse Code Modulation - DPCM predictive coding with an Indexed Color History – ICH as shown in Figure 7. The DSC codec includes three different prediction modes. The first prediction mode is modified median-adaptive prediction – MMAP, which is similar to the median-adaptive prediction used in JPEG-LS. The modification enables parallel prediction of three consecutive sample values. Using MMAP, the predicted value is derived from previously coded values to the left and above the current sample. The second mode is block prediction, which uses previously reconstructed pixels to the left on the same scan line. The previous scan line is used to identify the pixel with the smallest difference value, and uses the pixel below it, which is on the current scan line. Block prediction is optional in decoders to allow for reduced hardware costs. Midpoint prediction - MPP is the third mode, and uses the midpoint of the component range. The encoder and decoder automatically select the appropriate mode using the same algorithm, eliminating the need for signaling in the bit stream.

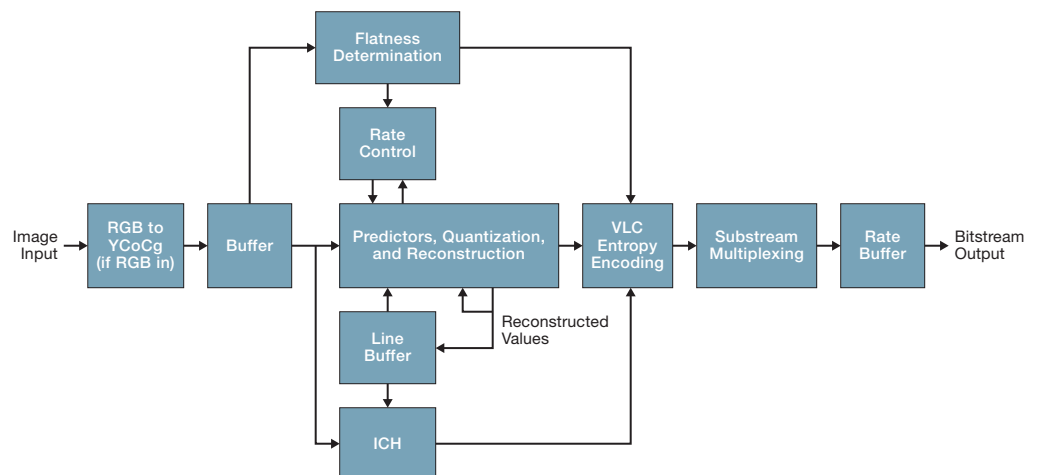


Figure 7: VESA Display Stream Compression Encoder

In content, such as computer graphics and text, similar pixel values occur in close proximity to each other. The ICH is used to store recently used pixel values, such that both the encoder and decoder store identical values. In ICH mode, the encoder selects index values corresponding to pixel values within the ICH, and sends the index value to the decoder. The decoder uses the index value to retrieve the corresponding pixel value from its ICH.

Other functions to improve performance are also included within the encoder. Rate control adjusts the quantization value to maximize image quality, maintain a constant bit rate, and prevent the rate buffer from overflowing. Flatness detection reduces quantization artifacts for transitions from non-flat to flat areas that are signaled in the bit stream. Substream multiplexing enables three entropy encoders to run in parallel for decoding three pixels per clock.

VC-2 Video Codec

VC-2 is a royalty-free compression codec that was developed by the British Broadcasting Corporation - BBC and standardized as SMPTE 2042. It uses a DWT-based intra-frame only compression algorithm and is designed for applications that require light compression and low latency. The wavelet transform is applied to the entire frame to eliminate blocking artifacts. The lack of inter-frame compression simplifies applications such as switching streams and video editing. VC-2 can achieve visually lossless compression at ratios up to 4-to-1 and lossy compression up to 25-to-1.

The codec includes a Low Delay, Simple, or Main profile. The Low Delay profile features very low latency and a constant bit rate. After the DWT is applied, the transform coefficients are divided into slices so that each slice can be independently scaled, quantized, and encoded using variable length coding. The Low Delay profile enables compression ratios up to 8-to-1. The Simple profile operates on the entire frame rather than using slices to achieve a latency of 1 to 2 frames and compression ratios up to 12-to-1. The Main profile also operates on the entire frame and adds arithmetic entropy coding for compression ratios up to 25:1 with a latency of 2 to 3 frames.

Conclusion

Video compression provides unique benefits to an AV system, enabling new applications. Lossless compression with low latency enables delivering 4K video over an existing infrastructure and extending the capabilities of display links. For example, the VESA DSC codec enables visually lossless video in a display link, such as DisplayPort. Additionally, the Extron PURE3 codec enables visually lossless compression in real-time applications. It is optimized for AV applications and designed to deliver high resolution video and computer graphics with minimal latency.

Low latency encoders enable real-time applications for streaming, transmitting, and recording motion video. For example, the PURE3 codec is designed for streaming video and graphics over managed IP networks. It enables streaming high quality video and computer graphic images in mission-critical applications that require very low delay to support interactive, real-time communications. The PURE3 codec's extensive bit rate management and error concealment make it adaptable to different network conditions, with flexibility to support a wide range of streaming and recording applications. Additionally, specialized versions of H.264 codecs have been developed to target real-time applications. They take advantage of slice encoding to achieve low latency, may also include an intra-only mode where only I-frames are used, and often include error concealment.

Video encoders designed for computer graphics must maintain 4:4:4 chroma sampling without converting video to a 4:2:2 or 4:2:0 format. Video encoders designed to enable high compression ratios with an acceptable image quality for full motion video produce artifacts in computer images due to chroma subsampling. Image details can become distorted, particularly when these compression systems are applied to single pixel lines, small characters, and fine graphic details.

Lossy compression enables video conferencing, streaming video across computer networks, and storing full-length motion pictures in HD quality on a Blu-ray Disc. A variety of codecs are available to deliver the performance needed for each of these applications. The H.264/MPEG-4 AVC has become the most widely used codec for streaming HD video across computer networks and storing full-length motion pictures on Blu-ray Discs with HD quality. The H.265/HEVC codec improves upon the H.264 to enable streaming 4K and 8K video. The VP9 codec provides a royalty-free solution that might compete head-to-head with the H.265/HEVC codec.

AV professionals can benefit from understanding compression codecs and the tradeoffs between image quality, bandwidth, and latency to ensure they select the appropriate equipment for the target application. The technical information presented enables an understanding of compression algorithms and codecs, and how they impact AV system design.

Extron Electronics, headquartered in Anaheim, CA, is a leading manufacturer of professional AV system integration products. Extron products are used to integrate video and audio into presentation systems in a wide variety of locations, including classrooms and auditoriums in schools and colleges, corporate boardrooms, houses of worship, command-and-control centers, sports stadiums, airports, broadcast studios, restaurants, malls, and museums.

www.extron.com
© 2017 All rights reserved.

WORLDWIDE SALES OFFICES

Anaheim • Raleigh • Silicon Valley • Dallas • New York • Washington, DC • Toronto • Mexico City • Paris • London
Frankfurt • Madrid • Stockholm • Amersfoort • Moscow • Dubai • Johannesburg • Tel Aviv • Sydney • Melbourne
Bangalore • Mumbai • New Delhi • Singapore • Seoul • Shanghai • Beijing • Hong Kong • Tokyo

www.extron.com