How Does H.264 Work?

Understanding video compression with a focus on H.264

SALIENT SYSTEMS

Simple.Scalable.Security™

Compression Techniques and Basics

Data compression is the process of encoding information to reduce the number of bits as compared to the uncompressed version of the information. A compression process will reduce the disk space required to store the information, and bandwidth required to transmit information.

Compression can be lossless or lossy. Video Compression used in security applications is lossy compression, meaning once the original video information is compressed it can never be decompressed to restore all of the original information. In lossy compression, if information is compressed then decompressed then recompressed and so on, more and more information will be lost with each successive compression.

Video information lends itself well to lossy compression techniques. The reason being, there is a significant amount of data the human eye does not perceive in the uncompressed version of the video or image. Image and video compression takes advantage of this fact.

Specifically the human visual system perceives brightness more readily than color. If you have ever compared LCD televisions, you may have noticed displays with higher contrast ratios look better. The contrast ratio is the difference from the brightest white to the darkest black that can be displayed on the set. Since our visual system perceives differences in brightness more than color the effect is a more appealing display, which is the same effect JPEG compression takes advantage of.

Still Image Compression

At the time of this writing, Motion JEPG (MJPEG) compression is the most widely used compression technique used in Digital Video Management Systems. Motion JPEG is a series of JPEG images with a small amount of header data between them. A MJPEG stream is analogous to a film real, or flip book, containing a series of still images played back at high speed to create the effect of video.
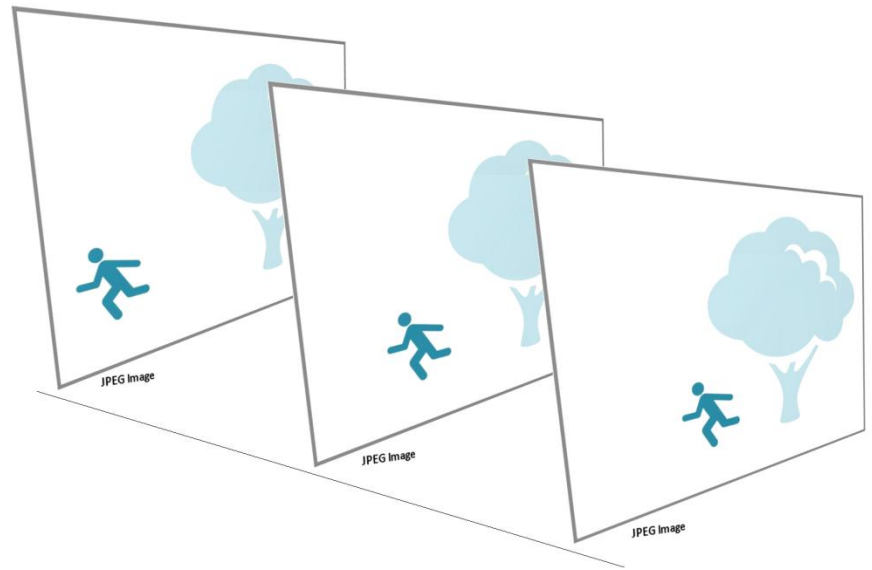
Figure 1: The Motion JPEG stream transmits a full image for each frame.

The goal of JPEG compression is to reduce Spatial Redundancies, or areas of similar color in the image. The effectiveness of JPEG compression is partly dependent on the complexity of the scene. A scene with small areas of highly contrasting colors is more difficult to compress then large areas of similar colors (see below).

Flower Garden, 14KB

White Wall, 2KB

In JPEG compression the image is broken up into squares, called macro blocks, typically 8x8 pixels or 16x16 pixels. The macro blocks go through several processes, the effect of which is to reduce color information, but maintain a level of detail for brightness information. The greater the level of compression on the JPEG image or MJPEG stream, the more pronounced the borders of the macro blocks become. This effect is called pixelization and can be seen in Figure 2.

Figure 2, Pixelization due to high JPEG compression

Video Compression

The main difference with video compression vs. still image compression is the reduction of Temporal Redundancies. Temporal Redundancies are the similarities between images. A significant amount of storage and bandwidth, or bitrate, can be saved by not retransmitting the macro blocks which don't change from frame to frame.
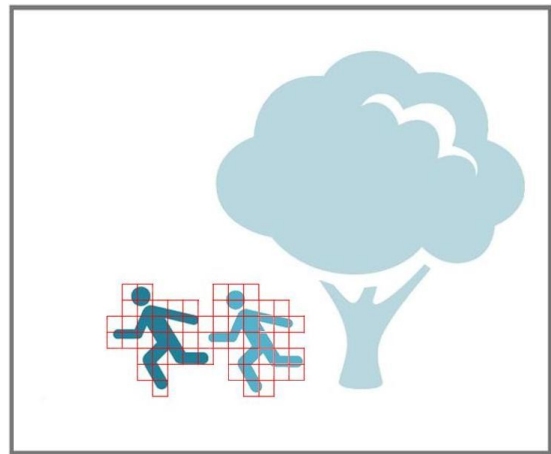
Figure 3: Only macro blocks which change are transmitted

In a security application this can refer to the background of the image. Many cameras in a video security deployment are fixed cameras. The only thing changing in the camera's field of view are vehicles passing, people walking etc. The majority of the scene from image to image is the same information (the background). True video compression techniques, like MPEG4 part 2 & H.264, take advantage of this fact and only transmit the background periodically. The result is a video stream composed of a reference frame, called an I Frame, and then changing areas of the image are transmitted and overlaid on the original reference frame to create the current image of the scene (Figure 3). The frames with changes are referred to as P or B frames.

An I Frame followed by a series of P and B frames is referred to as a Group of Pictures or GOP (Figure 4). The number of frames from one I frame to the next I frame in the stream is referred to as the GOV Length. This is usually a parameter which can be modified for the video stream. The longer the GOP Length, the fewer I frames will be present in the stream, which lowers the bitrate.
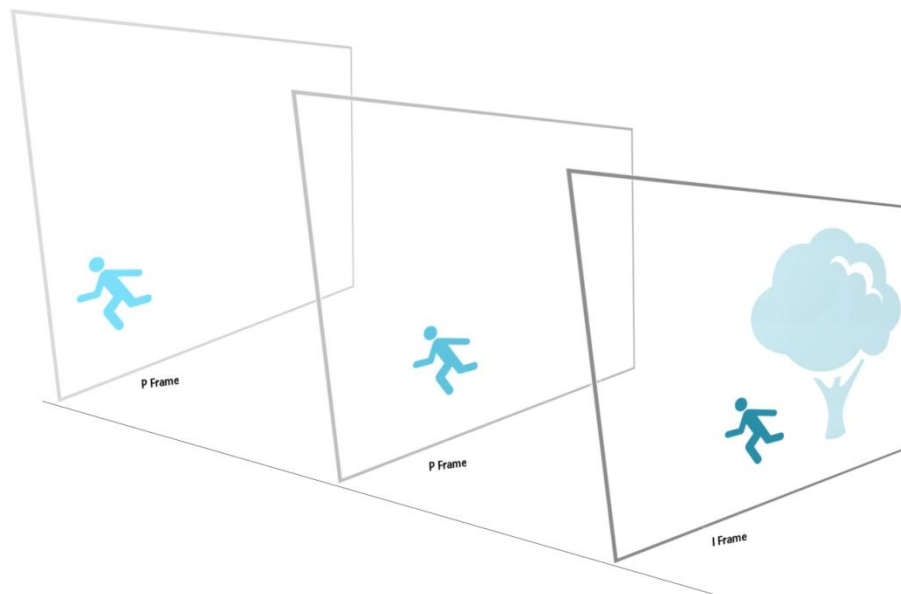

Figure 4: MPEG4 or H.264 transmits an I Frame, containing the full scene. After the I Frame only changes are sent until the next I Frame.

Advanced Features, H.264

H.264 supports advanced feature sets of video compression, such as Motion Compensation. Additionally H.264 can enhance the quality of highly compressed video using a deblocking filter.

In a P or B frame, only moved or changed macro blocks are transmitted. Motion Compensations allows the encoder to transmit motion vector information about

the macro blocks that move but which have the same pixel level information

Figure 5: Macro blocks already transmitted are moved using motion vector information.

(Figure 5). The decoder will move the macro blocks, which have already been transmitted, to their new location using the motion vector data. This saves a significant amount of bitrate as compared to retransmitting the pixel information which makes up the macro block. Some implementations of MPEG4 also have this feature.

The deblocking filter is a default feature of H.264. It removes artifacts associated with very high compression (as seen in Figure 2). Smoothing the edges between macro blocks makes a considerable difference in the perception of video quality. The deblocking feature can obviously improve image quality, but also it can allow the user to configure the video stream for a higher compression that would otherwise be considered unacceptable for the user's purposes. This feature can be considered a way to further decrease the bitrate of the stream as compared to equivalent quality Motion JPEG or MPEG4.

As compression increases, the colors of each macro block become more averaged, making the pixels within macro blocks closer to the same color. As this effect increases, the borders of the macro blocks become more obvious. This effect causes pixelization, or the clear display of squares in the compressed image. The deblocking filter samples pixels on each side of two bordering macro blocks. Based on the sample pixels, the filter decides the best average color and recolor's the borders on each side of the macro block. This creates a smoother transition between macro blocks which can be seen in Figure 6.
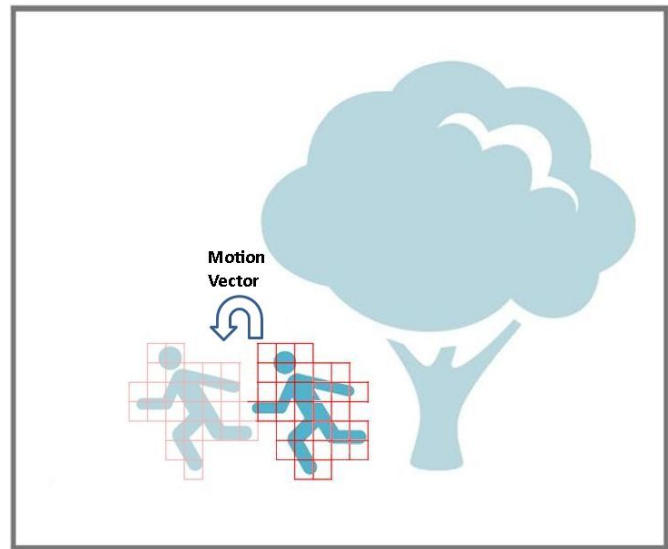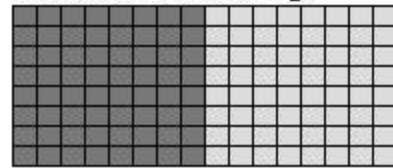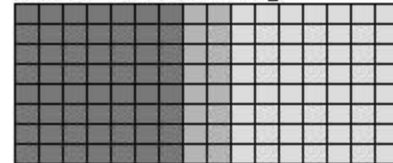
Figure 6: The image on the right is the result of the highly compressed image on the left after the deblocking filter has been applied.

ABOUT SALIENT SYSTEMS

Salient Systems offers network friendly, comprehensive IP and analog video surveillance management systems (VMS) built on open architecture. As the recognized transition leader from analog to digital video, Salient Systems' VMS, CompleteView™, is scalable and provides everything needed to manage a multi-server enterprise from a single desktop. Salient delivers simple and scalable security today…and tomorrow. For more information about Salient Systems and CompleteView, visit www.salientsys.com.

ABOUT THE AUTHOR

Brian Carle is the Director of Product Strategy for Salient Systems Corporation. Prior to Salient he worked as the ADP Program Manager for Axis Communications.

For information about this white paper or CompleteView, email info@salientsys.com.

Salient Systems
10801 N. MoPac Expy.
Building 3, Suite 700
Austin, TX 78759
512.617.4800
512.617.4801 Fax
www.salientsys.com