

# H.264 Facts and Fiction

*By Dr. Michael Korkin*



It seems everyone in the security industry is talking about the use of the H.264 compression standard for digital video, which produces high-quality video using less bandwidth than the commonly used JPEG compression. But how does H.264 differ from JPEG, and are the proposed benefits of H.264 compression too good to be true? Are there any hidden costs to using H.264 in security applications? Let us focus on the basics of the H.264 compression technology to separate facts from fiction and dispel a few myths and misconceptions.

## **H.264 and JPEG: How They Are Alike**

H.264 and JPEG are two closely related standards: Computationally they belong to the same family of compression methods. Both use similar or identical techniques to compress the video, such as transforming the video signal into frequency domain, applying quantization to the frequency-transformed signal, using variable length coding, and many other similar or identical techniques. Because the compression methods are similar, the distortion introduced into the video in the process of compression is also similar. The degree of video distortion is proportional to the degree of compression: both standards support a wide range of compression levels and, accordingly, a wide range of achievable video quality (the inverse of video distortion).

There are many metrics of video quality, some objective and some subjective. Using any measure, one can precisely demonstrate that when the compression parameters of the two standards are matched, the video quality of the same scene under like conditions is indistinguishable across a wide range of settings, with the possible exception of the extreme high-compression limit. In particular, this is easy to demonstrate using Arecont Vision's megapixel IP cameras that feature instant switching of the on-camera encoder between JPEG and H.264. In fact, if video quality was the only measure for choosing one compression standard over another it would be very difficult to make the choice.

So, if the video quality of the two standards is very much alike, then how are they different?

## **H.264 and JPEG: How They Are Different**

The main difference between H.264 and JPEG is the consumed bandwidth per given video

quality -- H.264 offers a major reduction in bandwidth relative to JPEG. Bandwidth reduction translates to a major reduction in cost of security installations: the requirements for networking equipment and disk storage are accordingly reduced.

Reduction of bandwidth is achieved at the cost of high computational complexity of the H.264 encoder. Put simply, the more computation there is, the more efficiently the data is organized and "packed." Decoding the compressed video stream is an entirely different matter, the H.264 standard is asymmetrical: all of its computational complexity is on the encoder side while the H.264 decoder is similar in complexity to a JPEG decoder.

Arecont Vision's megapixel IP cameras use a patent-pending massively parallel H.264 hardware encoder that achieves 80 billion operations per second. The high computational capacity is needed to process a large number of computational add-ons used in H.264 relative to JPEG, some of which were introduced in the earlier standards of the MPEG family to which H.264 belongs. A major departure from JPEG is that instead of encoding the video signal itself, only the inter-frame signal differences are encoded. The smaller the difference, the more economically it can be encoded into the video stream. There are two sources of inter-frame signal differences: motion in the scene and random noise.

Noise is always present, and it is notoriously difficult to compress due to its random nature. High levels of noise are typically caused by low-light conditions - they require larger bandwidth and larger disk storage space to archive.

Signal differences due to motion are much easier to compress -- the majority of computational effort is typically concentrated in estimating motion. The goal of motion estimation is to locate blocks of pixels in the current video frame that closely match blocks of pixels in the previous frame corresponding to the portions of the scene that may have moved during the interval between frames. Because the direction and the distance of such movement are unknown in advance, the motion estimator must search over hundreds of possible positions to find the best match.

The closer the match, the smaller is the signal difference to be encoded, and accordingly the smaller is the resultant video stream. Computational power of the motion estimator often determines the quality of the entire H.264 encoder: the larger the search area, the higher is the chance to find the best possible match. While many motion estimators conduct only an approximate non-exhaustive search to reduce the amount of computation, Arecont Vision's motion estimator conducts an exhaustive search over a large search area to find the best possible match.

Motion estimation and other computational components of H.264 compression explain its amazing ability to compress video into a low-bandwidth stream while maintaining high video quality. It is also the reason why H.264 is being embraced by broadcast television, by distributors of movies on DVD, and by many other industries including the professional security and surveillance market.

### **H.264 Has No 'Hidden Cost'**

A common myth about H.264 is its so-called "hidden cost", an erroneous belief that because the computational complexity of the H.264 encoder is very high, the required decoder resources must be high as well, many times higher than required for JPEG. The "hidden

cost", as the theory goes, is in the additional computer server power needed to decompress multiple H.264 video streams in a multi-camera security installation in order to display live video from multiple cameras. This "hidden cost" is alleged to be especially high for megapixel cameras.

In reality, the exact opposite is true: H.264 streams encoded by Arecont Vision cameras require less computational power to decompress than JPEG streams, a fact that has been demonstrated on brand-name and open-source H.264 software decoders, such as Intel IPP and FFMPEG used by all major NVR manufacturers.

In order to understand how this is achieved, consider that the H.264 compression standard consists of a large number of optional encoder components, each targeting its own facet of compression. Each of these optional components is capable of improving the compression by a certain amount, but every increment of improvement comes with a computational cost attached.

The computational cost is incurred mainly on the encoder side, but may affect the decoder side as well, in varying degrees. Some of these components have better cost-to-effect ratio than others. By carefully choosing the subset of optional encoder components, Arecont Vision has optimized its encoder to avoid the increase in computational load on the decoder side compared to JPEG decoder. At the same time, the H.264 video stream remains fully compliant with the standard and compatible with all compliant H.264 decoders.

As an example of computational load reduction in the decoder, consider the major computational component of the encoder -- its motion estimator. According to the H.264 standard, motion estimation could be conducted at up to quarter-pixel resolution. This means that if the encoder found the best match "in-between" the original pixels, the decoder (software running on the server) has to interpolate all the intermediate pixel values and generate 16 times more pixels than there are in the original image. This operation alone would raise the computational load of the H.264 decoder way beyond the JPEG decoder level.

The question is; How much benefit does it provide in terms of bandwidth reduction? In sub-megapixel low-resolution cameras, such as D1 format, this might be a valuable technique - there are relatively few pixels per foot of the scene, and accordingly the quarter-pixel precision of motion estimation makes a difference when motion is involved. However, when you consider a 5-megapixel camera, the number of pixels per foot is roughly 14 times higher than in D1 given identical optics and sensor size. It makes little sense to conduct quarter-pixel resolution search on top of the high resolution provided by the sensor itself.

By avoiding the unnecessary computation both in the encoder and the decoder, Arecont Vision's cameras achieve lower cost on the camera side and maintain low computational load on the server side. Motion estimation is one example of a multitude of such strategies implemented in Arecont Vision's megapixel cameras.

### **The Defacto Compression Standard**

The benefits of H.264 in terms of bandwidth utilization per given video quality and the related reduction of disk storage are obvious, the incremental costs are low, and there are no "hidden" installation costs. It is safe to predict that H.264 will become the defacto

compression standard for the security and surveillance market, especially for megapixel IP cameras where the benefits are even further multiplied. In fact, H.264 could be viewed as the silver bullet that has removed the earlier obstacles to mass penetration of megapixel IP cameras into the marketplace.

---

*Dr. Michael Korkin is Director of Engineering at Glendale, Calif.-based Arecont Vision.*