CODEX

Video compression design, analysis, consulting and research

White Paper:

**An Overview of H.264 Advanced Video Coding**

Iain Richardson

2007

1   What is H.264 ?

H.264 is an industry standard for video compression, the process of converting digital video into a format that takes up less capacity when it is stored or transmitted. Video compression (or video coding) is an essential technology for applications such as digital television, DVD-Video, mobile TV, videoconferencing and internet video streaming. Standardising video compression makes it possible for products from different manufacturers (e.g. encoders, decoders and storage media) to inter-operate. An **encoder** converts video into a compressed format and a **decoder** converts compressed video back into an uncompressed format.

Recommendation H.264: Advanced Video Coding is a document published by the international standards bodies ITU-T (International Telecommunication Union) and ISO/IEC (International Organisation for Standardisation / International Electrotechnical Commission). It defines a format (syntax) for compressed video and a method for decoding this syntax to produce a displayable video sequence. The standard document does not actually specify how to encode (compress) digital video – this is left to the manufacturer of a video encoder – but in practice the encoder is likely to mirror the steps of the decoding process. **Figure 1** shows the encoding and decoding processes and highlights the parts that are covered by the H.264 standard.

The H.264/AVC standard was first published in 2003. It builds on the concepts of earlier standards such as MPEG-2 and MPEG-4 Visual and offers the potential for better compression efficiency (i.e. better-quality compressed video) and greater flexibility in compressing, transmitting and storing video.

Figure 1 The H.264 video coding and decoding process
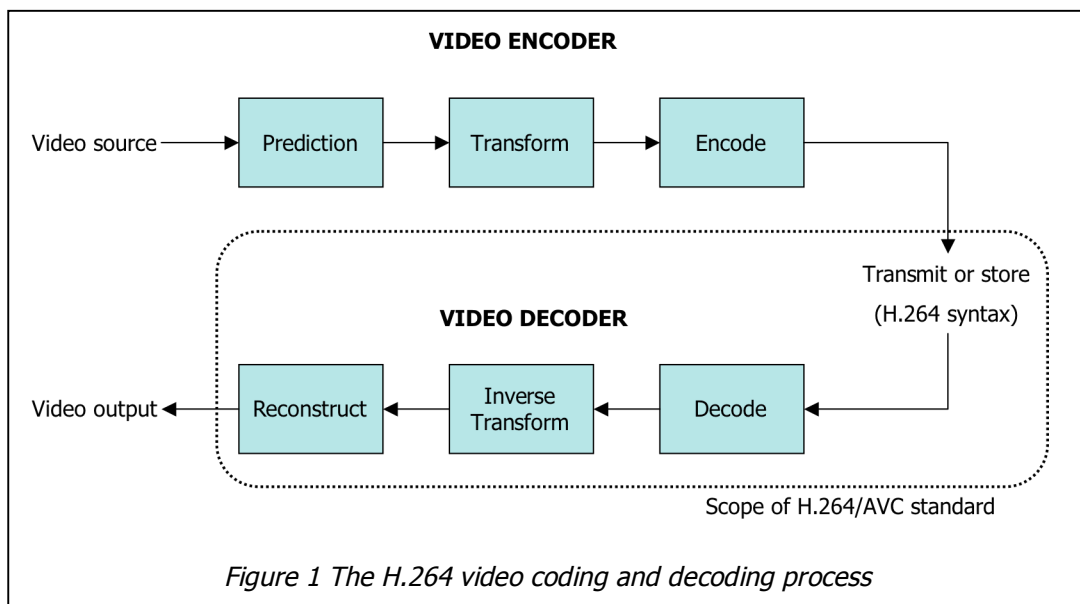
2   How does an H.264 codec work ?

An H.264 video encoder carries out prediction, transform and encoding processes (see Figure 1) to produce a compressed H.264 bitstream. An H.264 video decoder carries out the complementary processes of decoding, inverse transform and reconstruction to produce a decoded video sequence.

2.1   Encoder processes

Prediction

The encoder processes a frame of video in units of a **Macroblock** (16x16 displayed pixels). It forms a **prediction** of the macroblock based on previously-coded data, either from the current frame (**intra** prediction) or from other frames that have already been coded and transmitted (**inter** prediction). The encoder subtracts the prediction from the current macroblock to form a **residual**[1].

The prediction methods supported by H.264 are more flexible than those in previous standards, enabling accurate predictions and hence efficient video compression. Intra prediction uses 16x16 and 4x4 block sizes to predict the macroblock from surrounding, previously-coded pixels within the same frame (*Figure 2*).

Inter prediction uses a range of block sizes (from 16x16 down to 4x4) to predict pixels in the current frame from similar regions in previously-coded frames (*Figure 3*).

Figure 2 Intra prediction


Figure 3 Inter prediction
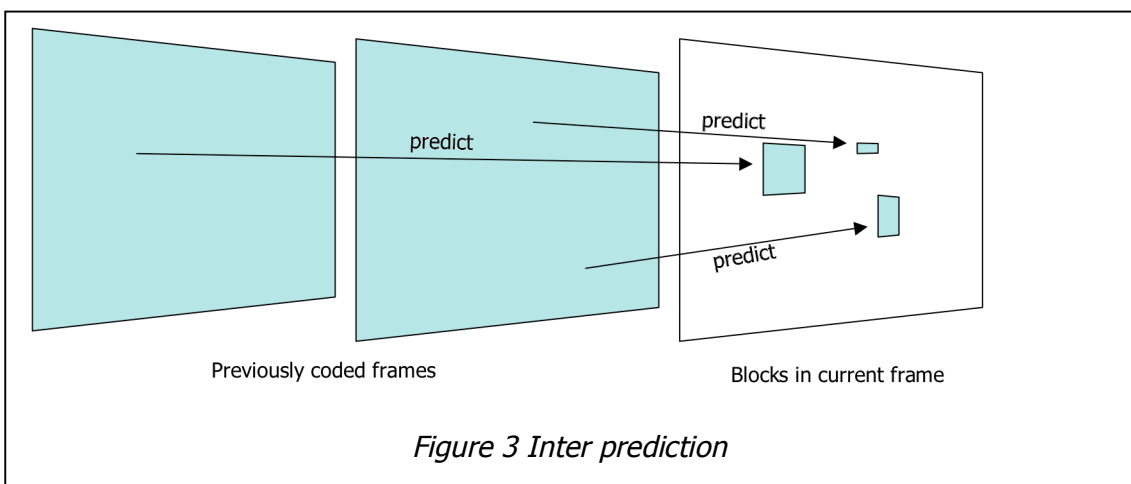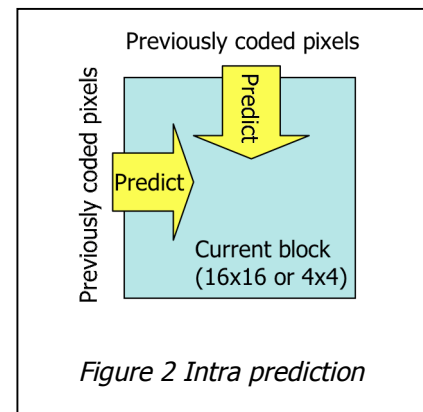
Transform and quantization

---

[1] Finding a suitable inter prediction is often described as **motion estimation**. Subtracting an inter prediction from the current macroblock is **motion compensation**.

A block of residual samples is transformed using a 4x4 or 8x8 **integer transform**, an approximate form of the Discrete Cosine Transform (**DCT**). The transform outputs a set of **coefficients**, each of which is a weighting value for a standard basis pattern. When combined, the weighted basis patterns re-create the block of residual samples. *Figure 4* shows how the inverse DCT creates an image block by weighting each basis pattern according to a coefficient value and combining the weighted basis patterns.

The output of the transform, a block of transform coefficients, is **quantized**, i.e. each coefficient is divided by an integer value. Quantization reduces the precision of the transform coefficients according to a quantization parameter (QP). Typically, the result is a block in which most or all of the coefficients are zero, with a few non-zero coefficients. Setting QP to a high value means that more coefficients are set to zero, resulting in high compression at the expense of poor decoded image quality. Setting QP to a low value means that more non-zero coefficients remain after quantization, resulting in better decoded image quality but lower compression.

Figure 4 Inverse transform: combining weighted basis patterns to create a 4x4 image block
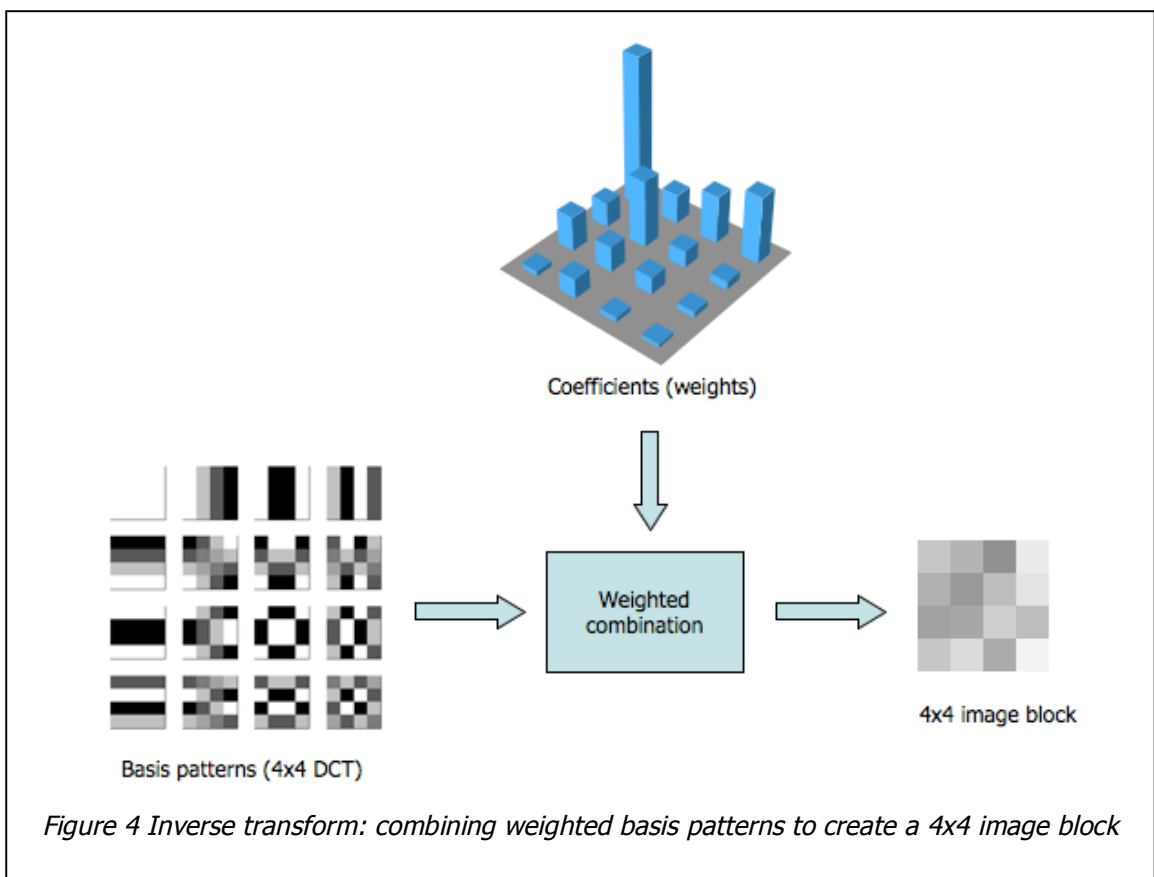
Bitstream encoding

The video coding process produces a number of values that must be **encoded** to form the compressed bitstream. These values include:
- quantized transform coefficients
- information to enable the decoder to re-create the prediction
- information about the structure of the compressed data and the compression tools used during encoding

- information about the complete video sequence.

These values and parameters (**syntax elements**) are converted into binary codes using **variable length coding** and/or **arithmetic coding**. Each of these encoding methods produces an efficient, compact binary representation of the information. The encoded bitstream can then be stored and/or transmitted.

2.2   Decoder processes

Bitstream decoding

A video decoder receives the compressed H.264 bitstream, decodes each of the syntax elements and extracts the information described above (quantized transform coefficients, prediction information, etc). This information is then used to reverse the coding process and recreate a sequence of video images.

Rescaling and inverse transform

The quantized transform coefficients are **re-scaled**. Each coefficient is multiplied by an integer value to restore its original scale[2]. An inverse transform combines the standard basis patterns, weighted by the re-scaled coefficients, to re-create each block of residual data. These blocks are combined together to form a residual macroblock.

Reconstruction

For each macroblock, the decoder forms an identical prediction to the one created by the encoder. The decoder adds the prediction to the decoded residual to **reconstruct** a decoded macroblock which can then be displayed as part of a video frame.

3   H.264 in practice

3.1   Performance

Perhaps the biggest advantage of H.264 over previous standards is its compression performance. Compared with standards such as MPEG-2 and MPEG-4 Visual, H.264 can deliver:
- Better image quality at the same compressed bitrate, or
- A lower compressed bitrate for the same image quality.

For example, a single-layer DVD can store a movie of around 2 hours' length in MPEG-2 format. Using H.264, it should be possible to store 4 hours or more of movie-quality video on the same disk (i.e. lower bitrate for the same quality). Alternatively, the H.264 compression format can deliver better quality at the same bitrate compared with MPEG-2 and MPEG-4 Visual (*Figure 5*).

---

[2] This is often described as **inverse quantization** but it is important to note that quantization is not a fully-reversible process. Information removed during quantization cannot be restored during re-scaling.

The improved compression performance of H.264 comes at the price of greater computational cost. H.264 is more sophisticated than earlier compression methods and this means that it can take significantly more processing power to compress and decompress H.264 video.

Figure 5 A video frame compressed at the same bitrate using MPEG-2 (left), MPEG-4 Visual (centre) and H.264 compression (right)

3.2   Applications

As well as its improved compression performance, H.264 offers greater flexibility in terms of compression options and transmission support. An H.264 encoder can select from a wide variety of compression tools, making it suitable for applications ranging from low-bitrate, low-delay mobile transmission through high definition consumer TV to professional television production. The standard provides integrated support for transmission or storage, including a packetised compressed format and features that help to minimise the effect of transmission errors.

H.264/AVC is being adopted for an increasing range of applications, including:
- High Definition DVDs (HD-DVD and Blu-Ray formats)
- High Definition TV broadcasting in Europe
- Apple products including iTunes video downloads, iPod video and MacOS
- NATO and US DoD video applications
- Mobile TV broadcasting
- Internet video
- Videoconferencing

Further reading

See http://www.vcodex.com/links.html for links to further resources on H.264 and video compression.

About the author

As a researcher, consultant and author working in the field of video compression (video coding), my books on video codec design and the MPEG-4 and H.264 standards are widely read by engineers, academics and managers. I advise companies on video coding standards, design and intellectual property and lead the Centre for Video Communications Research at The Robert Gordon University in Aberdeen, UK.

Iain Richardson
iain@vcodex.com
Telephone 020 7193 6500 (UK)