

MPEG Video Compression for High Definition Digital-Television – A survey

Tejas S. Karkhanis

University of Wisconsin - Madison

Dept. of Electrical and Computer Engineering

email: karkhani@ece.wisc.edu

April 9, 2001

1 Introduction

One of the formats for High Definition Digital-Television(HDTV) in the US is 1920 pixels horizontally by 1080 pixels vertically at 30 frames per second. Additionally, there are 8 bits for each of the three primary colors per pixel. This yields a total data rate of 1.5 Gb/s . The allocated 6 MHz channel bandwidth can only support 19.2 Mb/s which is further reduced to 18 Mb/s because of the transmission of the audio and ancillary data. This restriction implies that a compression ratio of $83:1$ must be applied to the original data before transmitting the information.

DVD players, Internet video, and HDTV decoders are just some examples of products which use MPEG compression. MPEG compression is used in many products when a high quality content is to be transmitted to the end user or data is to be stored for further retrieval. Thus these applications benefit from less storage space, and lower required transmission bandwidth – here we focus on MPEG video when applied to HDTV.

MPEG stands for Motion Pictures Expert Group which worked to generate the specifications for compression under the International Organization for Standardization(ISO) and International Electrotechnical Commission(IEC). Commonly MPEG is referred to MPEG-1 and MPEG-2 – there is also a new standard MPEG-4. Although similar in basic concepts MPEG-1 and MPEG-2 are addressed towards different issues.

MPEG-1, released in 1991, is optimized for video resolutions of NTSC, 352 pixels horizontally by 240 pixels vertically at 30 frames per second, and PAL, 352 pixels horizontally by 288 pixels vertically at 25 frames per second. MPEG-1 resolution can go as high as 4095 pixels horizontally by 4095 lines vertically at 60 frames per second. A major limitation of MPEG-1 is that it is defined for progressive frames only and not for interlaced frames. Hence it is not used in broadcasting; since broadcasting uses interlaced frames.

MPEG-2 was finalized in 1994 and addresses the digital television broadcasting. It efficiently codes the field-interlaced video and scales very easily. The resulting video has a higher quality since the target bit rate was raised to 4 and 9 Mb/s from 2 Mb/s. Therefore MPEG-2 will be the topic of this paper and is implied by the use of the word *MPEG*.

MPEG compression algorithms employ discrete cosine transform(section 4.1) on an image to efficiently explore the intra-frame pel correlation. However the pels in the nearby frames have a high correlation and therefore to explore the temporal redundancy, differential pulse code modulation(DPCM) is used – DPCM uses motion compensated prediction between frames.

In the next section we first introduce the Digital Video. The layered view is presented in section 3. Section 4 examines the intra-frame mechanisms, section 5 describes the inter-frame mechanisms, and we conclude in section 6. The original intention of this survey was to present the user with the digital signal processing as well as transmission algorithms such as Coded-Orthogonal Frequency Division Multiplexing and Vestigial Sideband Modulation, MPEG-Video, and MPEG-Audio. However since the survey space was huge we explore the MPEG-Video compression in depth at the system level.

2 Digital Video

Digital video was born as a result of the limitations of the analog video such as (1) susceptible loss due to transmission noise effects, and (2) quality loss from one generation to another. Basically a digital video is just a digital representation of the analog video. In contrast to analog video, digital video's quality does not degrade from one generation to the next. Thus unlimited digital copies may be made. Moreover, since the digital video's contents can be randomly accessed the user has the luxury of indexing into any scene of the video. Frame rate, color resolution, spatial resolution, and image quality are the four important factor to consider in a digital video system. The following sub-sections discuss these in more detail.

2.1 Frame Rate

The standard for HDTV displays the video at 30 frames per second – every second of the video is made up of 30 *frames* or pictures. Moreover, these frames are split into odd lines and even lines called *fields*. Thus two fields make up a frame. An *interlaced* video displays the 30 frames per second video as 60 fields per second; the odd lines are displayed first and then the even lines for every frame. A *progressive* or *non-interlaced* video displays the frame one line at a time from the top to the bottom of the screen.

2.2 Color Resolution

Color resolution refers to the number of colors displayed on the screen at one time. RGB, CMY, and YIQ are a few examples of different color formats. RGB color model decomposes the pixel into red, green, and blue components and is widely used in monitors and color video cameras. CMY model, used in color printers, decomposes each pixel into cyan, magenta, and yellow components. YIQ is used for broadcasting television content. It uses the properties of human visual system to prioritize information. Y is the luminance component and I and Q are the chrominance components of an image.

2.3 Spatial Resolution

Spatial resolution means "How big the picture to be displayed is". VGA computer monitors have a resolution of 640 by 480, the NTSC standard uses 768 by 484, and the HDTV standard requires a resolution of 1920 by 1080.

2.4 Image Quality

Image quality is a measure of the acceptability of the video for a given application. It is very difficult to quantify image quality since "acceptability" is qualitative. Thus the image and color resolution, along with the frame rate is used to compare the quality of the videos. For some applications 240 by 240 at 10 frames per second might be acceptable. For HDTV 1920 by 1080 at 30 frames per second is required.

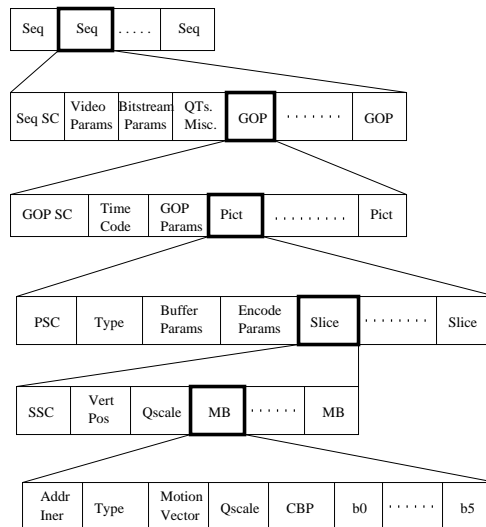


Figure 1: MPEG Video bit stream taken from [2].

3 MPEG Video: The Layered view

Figure 1 illustrates the MPEG layers. MPEG video consists of video sequences(Seq) at the highest level. Figure 2 shows a conceptual hierarchial breakdown of the data in a video sequence. A *video sequence* consists of group of pictures. The bitstream contains the start of sequence code, some additional information, group of pictures(GOP) and ends with an end of sequence code. A header, timing information, and a series of pictures make up the *group of pictures*– the group of pictures allow the random access into the sequence. The *picture* in a group of picture consists of a header, some parameters, and slices. The slices enable the error recovery since if one slice is corrupted it can be skipped. The more the slices the better is the error recovery, however more slices also add the header overhead. Each slice has a header, its position in the picture, and contiguous macroblocks. The order of the macroblocks in the slice is from left to right and top to bottom. A *macroblock* has a header, motion information, and a series of blocks – four luminance, two chrominance(one for red chrominance and the other for blue chrominance) blocks – the ratio of the number of luminance and chrominance blocks can be changed. Here we assume that the ratio is 4:1:1. *Y, U, and V* in figure 2 are examples of a macroblock. The smallest coding unit in the MPEG algorithm is a *block*. Each block consists of 8X8 pixels and can be of one of three types: luminance, red chrominance, and blue chrominance. The intra frame coding(Section 4) processes the pictures at the block level where as the inter-frame(section 5) coding does its processing at the macroblock level.

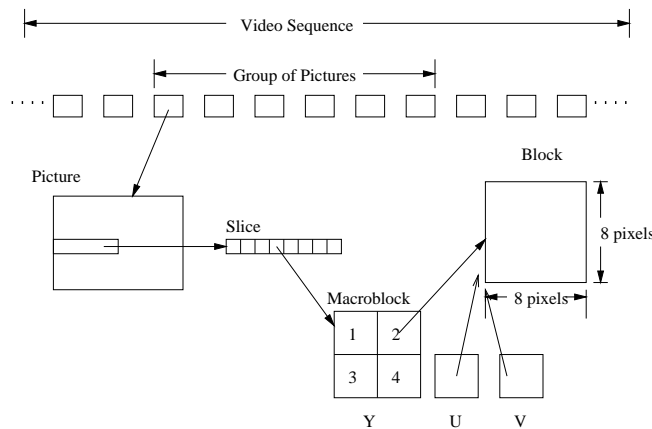


Figure 2: MPEG Data Hierarchy taken from [1].

4 MPEG Video: Intra-frame

Interpel(interpixel) correlation is the basic statistical property upon which MPEG compression techniques rely. It is assumed that the magnitude of a particular image pixel can be predicted from nearby pixels within the same frame or from pixels of a nearby frame(interframe technique – next section). In the model used, there is a high correlation between adjacent pixels and monotonic decay of correlation with increased distance between pixels. *Spatial domain* and *transform domain* are the two methods by which an image can be compressed. MPEG uses *transform coding*(see section 4.2), a reversible, linear transform, to map the image into a set of transform coefficients, which are then quantized and coded. The purpose of transform coding is to de-correlate the intraframe error image content and to encode transform coefficients rather than the original pixels of the images. Figure INTRAFRAMECODEC shows part of the MPEG codec – the part which does intraframe coding.

4.1 Video Filter

First the input is passed through a *video filter*. Studies have shown that the eye is most sensitive to changes in luminance, and less sensitive to variations in chrominance. Thus it make sense for MPEG to operate on the YUV color space which takes advantage of eye's less sensitivity to chrominance. The picture to be compressed is usually in RGB color space, thus the *video filter* transforms the RGB color space into YUV color space using the following formulas:

$$Y = 0.299R + 0.587G + 0.114B$$

$$U = B - Y$$

$$V = R - Y$$

At the decoder the YUV color space is changed back into the RGB color space with an inverse transformation.

4.2 Discrete Cosine Transform

Next *Discrete Cosine Transform(DCT)* decomposes the signal into underlying spatial frequencies. The DCT is very closely related to the Discrete Fourier Transform(DFT). DCT coefficients can be given a frequency interpretation close to the DFT; low DCT coefficients relate to low spatial frequencies and high DCT coefficients to higher frequencies within an image block. Just like a Fourier transform which decomposes a signal into weighted sumes of orthogonal sines and cosines that when added together reproduce the original signal, the DCT on an 8X8 block generates an 8X8 block of coefficients that represent a "weighting" value for each of the 64 orthogonal basis pattern that are added together to produce the original image. Equation 1 below is a forward DCT on an 8X8 block of pixels.

$$F(u, v) = \frac{1}{4}C(u)C(v) \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right) \quad (1)$$

$$C(u) = \frac{1}{\sqrt{2}} \quad \text{for } u = 0$$

$$C(u) = 1 \quad \text{for } u = 1, 2, \dots, 7$$

A major objective of DCT is to make as many tranform coefficients as possible small enough so that they are insignificant and hence need not be coded for transmission. At the same time it minimizes statistical dependencies between coefficients with the aim to reduce the amount of bits needed to encode the remaining coefficients. Coefficients with small variances are less significant for the reconstruction of the image blocks than coefficients with large variances.

Moreover DCT has the property of effective energy compaction. The most significant DCT coefficients are concentrated around the upper left corner(low frequency DCT coefficients) and the significance of the coefficient decays with increased distance. The non-uniform coefficient distribution is a result of the spatial redundancy present in the original image block. Hence the coefficients which are away from the top-left corner of the DCT coefficient matrix are less important for reconstruction of the image blocks. The zigzag scan of the DCT coefficient matrix illustrated in figure ZIGZAG scan gather the important, top-left, coefficients together. Thus the bit reduction is achieved by not transmitting the lower-right, and near-zero coefficients and quantizing and coding the remaining ones. Studies have shown

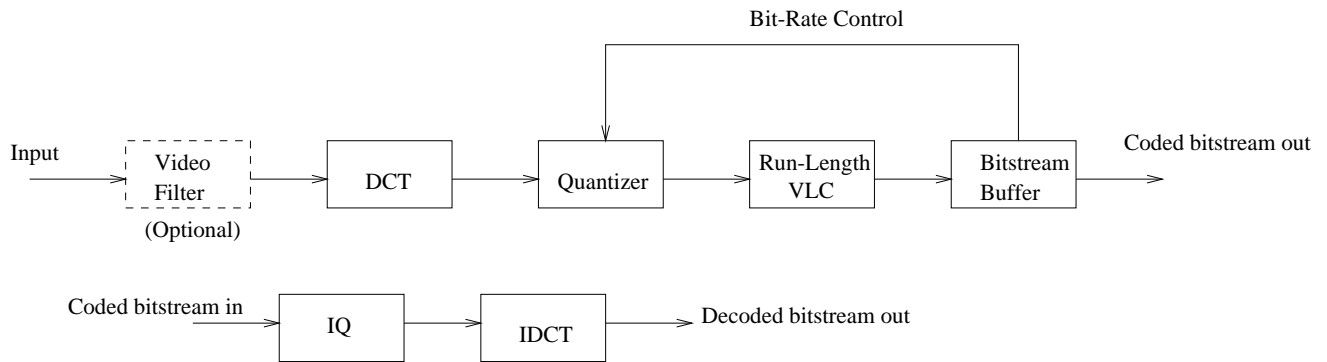


Figure 3: MPEG Intra-frame codec: high-level view [1].

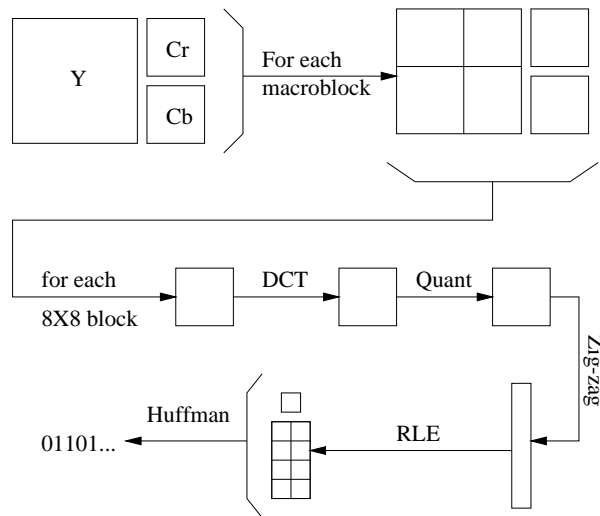


Figure 4: MPEG Intra-frame codec: detailed view [1].

4.4 Entropy Coding

Entropy coding is the last stage in the encoding process of MPEG compression. This is a lossless compression stage, hence is completely reversible. Before entropy coding, extra processing is applied to the DC coefficient of the quantized DCT coefficient matrix – the DC coefficient is at (0,0) in the matrix. The DC coefficient represents the mean value of the input block. Since there is a high correlation between DC coefficients of adjacent transform blocks, the difference between the quantized DC coefficients of adjacent blocks is computed. This value is then coded instead of the quantized value. Entropy coding consists of two major steps: run-length encoding(RLE), and variable-length coding(VLC).

4.4.1 Run Length Encoding

Run length encoding's(RLE) major objective is to compress long runs of 0's and 1's present in the bit-vector after the zig-zag scan of the quantized DCT coefficients. Since after the quantization, and even in images in general tend to have low-pass spectrum, the non-zero DCT coefficients tend to cluster at low frequencies and a large number of high frequency coefficients are likely to be zero. After the zig-zag scan the non-zero low-frequency coefficients tend to cluster towards the head of the bit-vector and the zero's at its tail. Each AC coefficient is represented by its value and its run-length[EXAMPLE]. Usually these codewords have a peak distribution and are further compressed by the variable-length coding.

4.4.2 Variable Length Coding

Variable length coding(VLC) further compresses the *(value,run length)* codes produced by RLE by exploiting the peak distribution usually present in the RLE-codes. This is done by assigning shorter coded to codewords having high probability of occurrence and longer codes to codewords having lower probability. The average codeword length coded by VLC will be:

$$C_{av} = \sum_k C_k P_k \quad (2)$$

4.5 Bit-rate control

In transmitting the MPEG encoded bits over a constant rate channel a good bit rate control scheme can overall improve the quality of the image/video being transmitted. The bit rate for HDTV is 18Mb/s. Unfortunately different images in the video may contain varying amounts of information, and result in different coding efficiencies. This situation may also

occur within a picture since the picture can be smooth at some places while others might contain high frequency information – due to these variations it is necessary to buffer the encoded bit-stream before transmission. The buffer must be of a limited size due to the physical constraints and thus a mechanism to control the input into the buffer must be provided. The most obvious place to control the encoded stream flowing into the buffer is at the quantizer. The DCT coefficient matrix can be changed on a picture by picture basis and the quantizer scale may be changed on a macroblock basis. Thus the underflow/overflow in the buffer may be prevented while not repeating or dropping the frames. The bit-rate control is very important in fixed bit-rate applications. MPEG-1 and MPEG-2 do not enforce a bit-rate or a bit-rate control mechanism, it is in the hands of the MPEG codec writer.

5 MPEG Video: Inter-frame

Inter-frame compression involves more than one frame. Consecutive frames are compared to remove any redundancies and arrive at “difference” information. Three types of pictures/frames are considered: I-pictures, P-pictures, and B-pictures.

I-pictures are coded without any reference to other pictures in the video sequence. They allow access points for random access and FF/FR functionality but achieve only low compression. They, however, contain data to reconstruct the whole picture since their compression is intra-frame. A GOP begins with an I-frame and ends before the next I-frame.

The inter-frame predicted, P-pictures, are coded with reference to the nearest previously coded I-picture or P-picture; motion compensation is usually used to get a higher compression. Unlike I-pictures, P-pictures do not contain the whole picture information – only the predictive information generated by looking at the difference between the present frame and the previous one – and therefore must depend upon the I-frames to reconstruct the whole picture. They provide no suitable access points for random access functionality or editability.

B-pictures or *bi-directional predicted/interpolated pictures* require both past and future pictures as references. The concept of B-pictures was introduced to further explore the advantages of motion compensation and motion interpolation based on the nearest past and future P-pictures or I-pictures. Similar to P-pictures, B-pictures do not contain information to reconstruct the whole frame. Therefore they provide no suitable access points for a random functionality or editability, either.

A series of P and B-pictures between two I-pictures forms a group-of-pictures. Figure 6 shows a 12-frame GOP with I-frames, P-frames, and B-frames – typically 12 frames occur

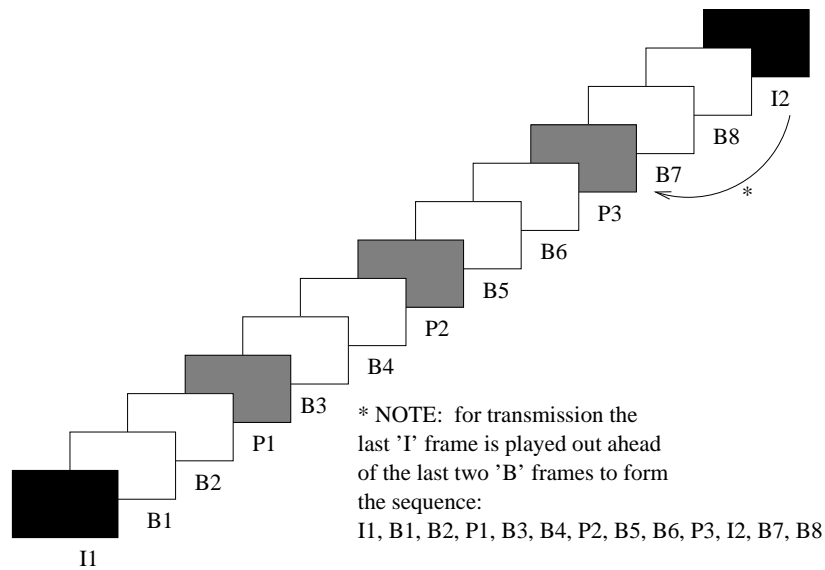


Figure 6: MPEG-2 12-frame GOP [16].

in a 25 fps signal and 15 frames in a 30 fps signal. The length of GOP can vary since a new sequence starting with an I-frame can start anytime when there is a big change at the input; for example a cut in the scene.

The kinds of pictures(B or P) used depends upon the application of the MPEG compression. A video sequence using I-pictures only (I, I, I, I, ...) allows the highest degree of random access, FF/FR, and editability, but achieves only low compression. A sequence coded with an I-pictures and P-pictures only, (I, P, P, P, P, P, I, P, P, P, P, I, ...) achieves moderate compression and a certain degree of random access and FF/FR functionality. Using I, P, and B-pictures, (I, B, B, P, B, B, P, B, B, I, B, ...), achieves high compression but a significant loss in the ability to random access and FF/FR. Moreover this also increases the coding delay. Since the channel bandwidth is of a major concern in HDTV systems, I, P, and B pictures are used.

First the I-frame is coded via intra-frame compression, then the P-frames and the B-frames are coded using motion-compensated prediction. Figure 7 shows the addition to the intra-frame MPEG codec to support the inter-frame coding.

5.1 Inverse Quantization(IQ), Inverse Discrete Cosine Transform(IDCT), and the Frame Store(FS)

Before doing the Motion Compensated Prediction(MCP) the quantized picture is inverse quantized and then passed through the inverse DCT. Although it seems intuitive the original

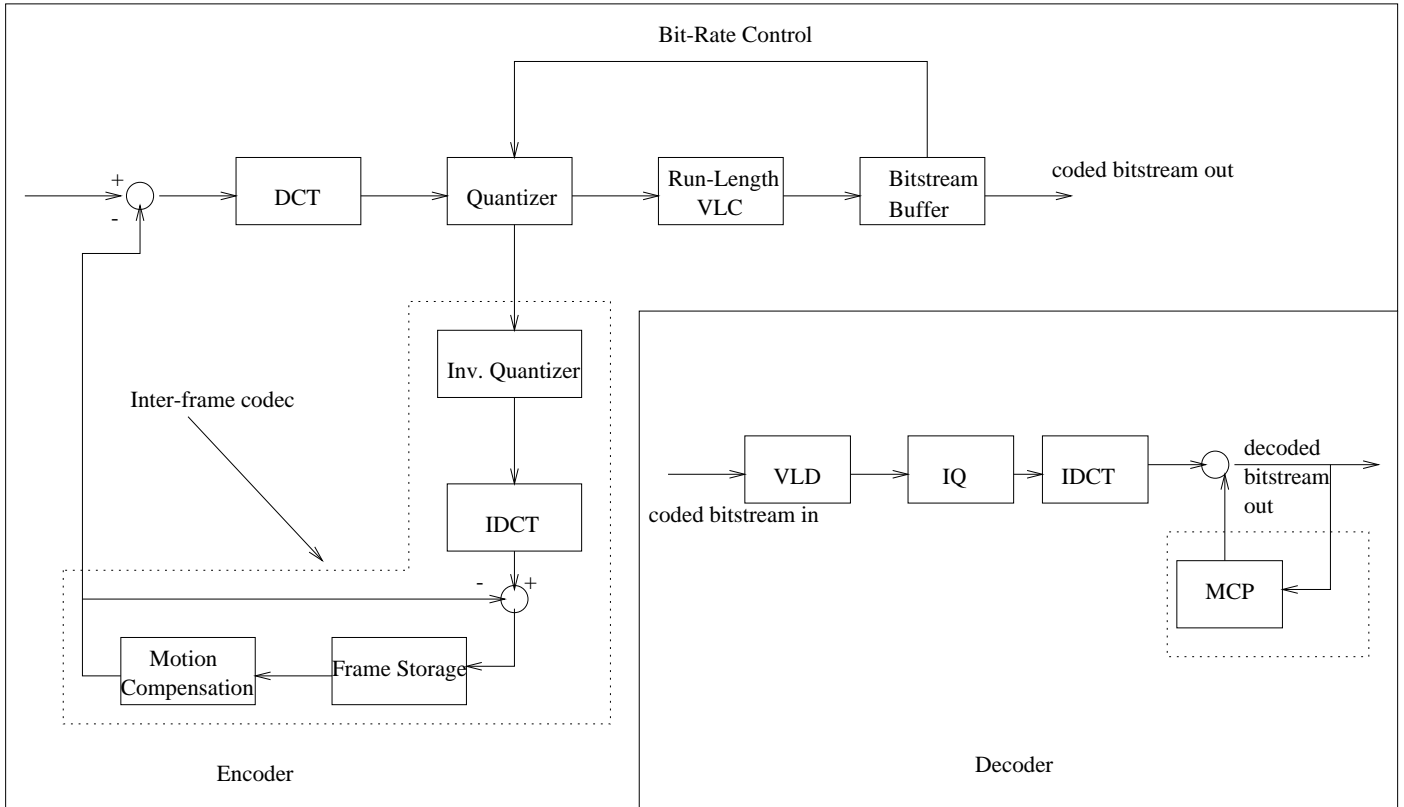


Figure 7: Interframe codec(the part in the dotted box) added to the intraframe codec.

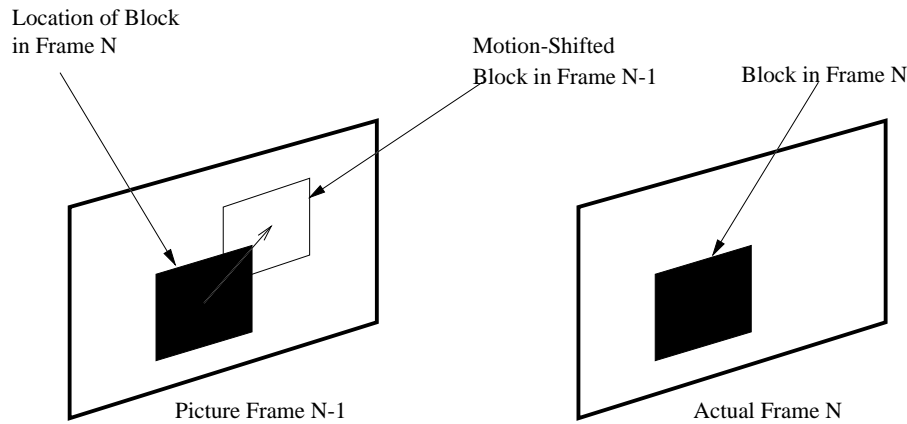


Figure 8: Block matching in Motion-Compensated Prediction

image is not used because there are differences in the original image and the one after being passed through inverse quantizer and inverse DCT at the decoder. Therefore the encoder has a part of the decoder in it and performs MCP on the image decoded by the decoder; basing all images on decoded copies of encoded images reduces error in the encoding process. This inverse quantized, and inverse discrete cosine transformed image is then stored in the *frame store (FS)*. The *motion compensated prediction (MCP)* unit then uses the image stored in the FS to remove temporal redundancy.

5.2 Motion Compensated Prediction

Motion-compensated prediction (MCP) exploits the temporal redundancies between frames to reduce the information coded. The concept of motion compensation is based on the estimation of motion between video frames – if all objects in a scene are more or less displaced by the same amount then the motion can be described by the original frame and some motion vectors. MCP works at the macroblock level i.e. 16X16 pels. A macroblock in the current frame is tried to map to locations on the previous frame (figure 8); the very first frame is the I-picture. This frame is stored in the frame store and then subtracted from the next frame. The matches between the predicted and the original frame are evaluated by the amount of difference between the original block; the match does not have to be perfect, however. By only storing and transmitting the difference, the coordinates of the B and P-frames do not have to be stored or transmitted. Thus less information has to be transmitted and also less memory is used.

On the receiving end the decoder executes the same process. The first frame, I-frame, is received, then the error vectors are subtracted from the first frame yielding the second frame

and so on the whole video is received.

6 Conclusions

MPEG is a generic compression standard which gives its implementators the flexibility to design their own decoders. This yields high performance application specific implementations. DCT and MCP are used to exploit the spatial and temporal redundancies in the video. Because higher bandwidths will be needed by future user applications further more there will be digital video libraries. These advances will need compression like MPEG. Therefore MPEG-2 is not the final MPEG standard. MPEG-4 is in its development stage. MPEG-4 is aimed at mobile applications supporting 5-64kbits/s. It is based on the concept of objects. Unfortunately MPEG-4 is not mature enough to include it in this survey. However, with the understanding of MPEG the reader should be able to appreciate MPEG-4.

7 Acknowledgements

The author thanks Prof. Hu for his guidance in surveying the papers, and the development of this survey.

References

- [1] S. A. Basith, and S. R. Done, "Digital Video, MPEG and Associated Artifacts", June 1996, http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/sab/report.html
- [2] "A Beginners Guide for MPEG-2 Standard", <http://www.fh-friedberg.de/fachbereiche/e2/telekom-labor/zinke/mk/mpeg2beg/beginnzi.htm>
- [3] A. B. Watson, "Image Compression Using the Discrete Cosine Transform", *Mathematica Journal*, 4(1), 1994, p. 81-88
- [4] "Color in Image and Video", <http://www.cs.sfu.ca/CourseCentral/365/li.material/notes/Chap3/Chap3.3/Chap3.3.html>
- [5] P. N. Tundor, "MPEG-2 Video Compression", December 1995, *Electronics & Communication Engineering Journal*

- [6] S. F. Chang and D. G. Messerschmitt, "Designing high throughput VLC decoder Part I - Concurrent VLSI architecture", IEEE Trans. On Circuits and Systems for Video Technology, June 1992
- [7] S. F. Chang and D. G. Messerschmitt, "Designing high throughput VLC decoder Part II - Parallel decoding method", IEEE Trans. On Circuits and Systems for Video Technology, June 1992
- [8] E. Cooper, "Minimizing Quantization Effects using the TMS320 Digital Signal Processor Family", Texas Instruments Application Report, 1994.
- [9] A. Dekker, "Kohonen neural networks for optimal colour quantization", Network: Computation in Neural Systems, Volume 5, pp 351-367, 1994.
- [10] O. Verevka, and J. Buchanan, "Local K-means Algorithm for Color Image Quantization",
- [11] O. Verevka, "Color Image quantization in Windows Systems with Local K-means Algorithm",
- [12] R. Gonzalez, and R. Woods, *Digital Image Processing*, Addison Wesley Publishing Company, Inc., 1993.
- [13] A. Campos, "Run Length Encoding, Copyright(c) A. Campos, 1999.
- [14] *RLE - Run Length Encoding*, Data Compression Reference Center, 1997-2000.
- [15] Entropy Coding, Hani Sorial, 1999.
- [16] *Digital Fact Book*, 1998 at <http://www.quantel.com>