Performance of H.26L Video Coding

N.H. Nga, W.A.C. Fernando
Telecommunications Program, Asian Institute of Technology, P.O. Box 4, Klong Luang, Pathumthani 12120,Thailand.
Tele. - +66-2-524-5746, Fax - +66-2-524-5730
Email: fernando@ait.ac.th

Abstract: In recent years the demand for digital video and image communications has been increased tremendously. On the other hand, video communications requires very much bandwidth in comparison with other information types such as text and data. Thus to adapt with the bandwidth-limited channels, especially wireless channels, video source must be compressed extremely. A new video coding standard namely H.26L is being developed by Joint Video Team (JVT). In this paper performance of H.26L is analyzed in an AWGN environment.

1. Introduction

The past few years have witnessed a tremendous growth of mobile communications and multimedia communications. A mobile communication, which is also known as wireless communications, is shifting its focus from solely voice communications service to many new services in combating with the extremely fast development of the Internet. Among the new services, video transmission will be a major application in the up coming 3G and 4G systems and may be a key factor in their success. In the limited bandwidth environment, video compression algorithms should be taken in to account. Evidently the demand for digital video communication applications such as: video conferencing, video e-mailing and video telephony has increased considerably while the transmission rates over Public Switch Telephone Network (PSTN) and wireless networks are still very limited. The "video coding for low bit rate communications" is therefore considered to address the above question.

The success of H.261 was a milestone for "low bit rate video coding" at reasonable quality [1]. It is the first recommendation of ITU-T in the H.26x series. H.261 is designed for real-time applications and optimized for low bit rate transmission over ISDN (down to 64kbit/s). H.263 is an improvement of the preceding H.261 standard [1]. It achieves significantly better image quality and manages to compress video to even lower bit rates. An important reason for these improved results is that H.263 uses half pixel prediction when estimating motion. H.263 also has a more efficient VLC table for coding DCT-coefficients. H.263 provides better picture quality at low bit rates with little additional complexity. It also includes four optional modes aimed at improving compression performance. H.263 has been adopted in several videophone terminal standards, notably ITU-T H.324 (PSTN), H.320 (ISDN), and H.310 (B-ISDN). For near-term standardization, H.263 version 2, also known as H.263+ offers many improvements over H.263. It was officially approved as a standard in January 1998. H.263+ is an extension of H.263 providing 12 new negotiable modes and additional features [2,3]. The objective of H.263+ is to broaden the range of applications, improve compression efficiency and address error robustness and resilience problems.

For more compression enhancement, H.26L-the long term standard in H.26x series-has been paid much attention to and is the current project of two organizations: ITU-T VCEG and ISU/IEC [4]. The core codec algorithm H.26L Test Model Long Term #1 (TML-1) was defined in late 1999. The main goal of this new ITU-T standard is to reduce the bit rate by 50% compared to H.263+ for a similar degree of encoder optimization. For that comparison, the encoding algorithms that are specified in the test models for H.26L and H.263+ are used. Besides H.26L also aims to improving the network adaption and simple syntax specification. H.26L is similar to its preceding standards in its use of a block transform that is essentially an inverse discrete cosine transform (IDCT). A new feature of the H.26L design is its introduction of a conceptional separation between a Video Coding Layer (VCL), which provides the core high-compression representation of the video picture content, and a Network Adaptation Layer (NAL), which packages that representation for delivery over a particular type of network. The H.26L draft VCL design is basically similar to that of prior standards, in that it is a block-based motion-compensated hybrid transform video coder. However, in contrast with prior standard, H.26L contains a number of new features that enable it to achieve a significant improvement in coding efficiency relative to prior standard designs. H.26L draft NAL design provides the ability to customize the format of the VCL data for delivery over a variety of particular networks. Therefore, a unique packet-based interface between the VCL and the NAL is defined. For the above promising results, H.26L standard is likely to play a major role in video communications in the future and expected to bring out some significant advances in the state-of-the-art standardized video coding, including key aspects designed with mobile applications [5].

In this paper it will be investigated the performance of H.26L video coding standard for a wireless channel. Rest of the paper is organized as follows. Section

2 briefly discusses about basics of video coding and H.26L video coding. Simulation model is presented in section 3. Section 4 presents some results and section 5 includes the conclusion.

2. Video Coding Basics

2.1 Overview of video compression techniques

The key idea of video compression is to exploit *spatial redundancy* and *temporal redundancy*. Spatial redundancy exists within an image due to similarities (correlations) between different areas of the image. For instance, adjacent pixels (picture elements) in an image are very likely to have similar characteristics. Temporal redundancy can be found within a sequence of frames (video) and is due to similarities between successive frames. Procedures for exploiting redundancy have different approaches depending on the redundancy being spatial or temporal.

There are two different kinds of compression: *lossless compression* and *lossy compression*, also referred as *reversible compression* and *irreversible compression*. In lossless compression technique, the information content of the original data and the compressed data is exactly the same. On the other hand, in lossy compression technique, the compressed data will contain less information than the original data, i.e., information is lost.

In addition, when dealing with discrete representation of information, data will always be *quantized*. This means that data-values, for example pixels, will always be rounded off or truncated to fit a certain number of bits. The fewer available bits, the coarser the quantization.

An important property of a compressed video signal is its *bit-rate* which is defined as the average number of bits needed to represent each second of a video sequence. In other words, the bit-rate of a video sequence is its size in bits divided by its length in seconds. In video compression in general, the goal is to obtain a lower bit-rate for the compressed video signal than for the same uncompressed signal.

It is, however, important to consider factors other than bit-rate as well. If, for instance, video is used in mobile communications, the compressed video signal needs to be insensitive to bit-errors. That is, the video signal needs to be *error resilient*. It is also preferable to reduce the complexity of the encoder and decoder and to add as little delay to the whole system as possible.

Video compression techniques therefore mainly make use of the spatial and temporal redundancy. Besides, it is also based on the insensitivity of the human eyes to loss of certain spatio-temporal visual information. Combining these two observations, a lossy compression scheme can be used to reduce the video bit rate while maintaining an acceptable image quality.

There are two different types of image coding techniques: *intraframe* and *interframe* coding. Intraframe coding only exploits the spatial redundancy while interframe coding relies on temporal redundancy as well. Still image coding techniques such as JPEG utilize only intraframe coding. For video coding such as H.261, H.263, H.26L, MPEG-1, 2 and 4, both intraframe and interframe coding are used. Moreover there is also much of redundancy between the compressed data symbols so that *entropy coding* is taken into account.

2.2 H.26L encoder and decoder

Similar to the prior standards, H.26L is also a block-based motion-compensated hybrid transform video coder. The current version of H.26L supports two different video formats: CIF and QCIF. In this model, QCIF is used as it requires less bit rate than CIF. The block transform size is 4x4 blocks instead of 8x8 blocks.

The input video sequence consists of a sequence of pictures. Pictures are divided into macroblocks of 16x16 pixels. A number of consecutive macroblocks in coding order can be organized in slices. Slices represent independent coding units in a way that they can be decoded without referencing other slices of the same frame. Efficient parallel processing of the encoder can be achieved through an effective scheduling algorithm. Figure 1 shows the H.26L video codec (encoder and decoder) structure.

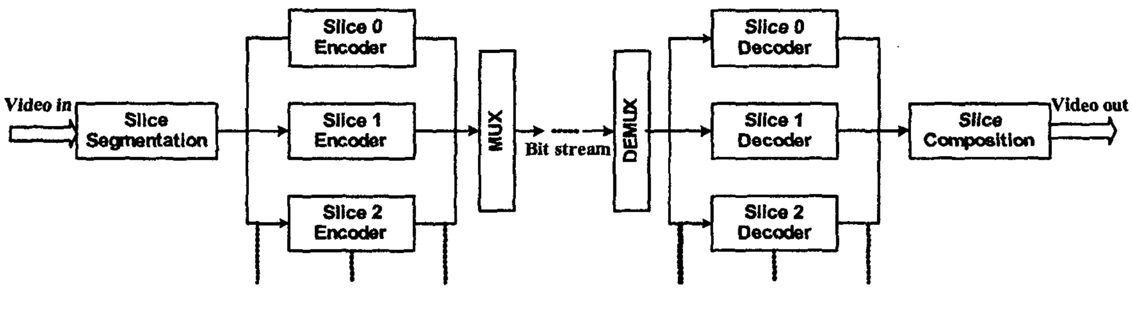

Figure 1 H.26L video codec (encoder and decoder) structure [6]

Figure 2 and Figure 3 represent the diagram of the H.26L encoder and decoder respectively. Every input macroblock needs to be predicted in H.26L. S_o is used to select the correct prediction method for inter and intra macroblock. The intra predictions are derived from the neighboring pixels in left and top blocks. The unit size of spatial prediction is either 4x4 or 16x16. As H.26L allows more than one previous frames for prediction in inter frame, inter prediction is calculated from one of these previous frames. Seven block sizes, i.e., 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 and 4x4, are supported in H.26L. The spiral search finds the minimum cost for each block size in the given range. The cost includes signal *Sum of Absolute Difference* (SAD) and overhead bits for coding block size information and motion vectors. The optimal block size is decided based on these minimum costs.

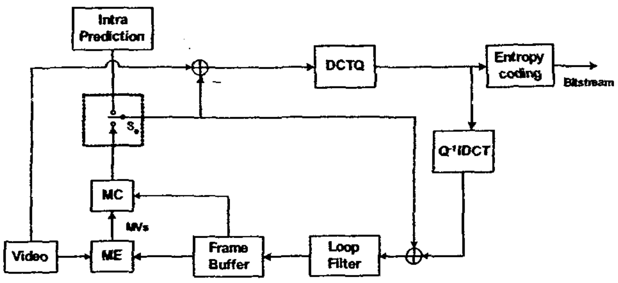

Figure 2 H.26L video encoder [7]

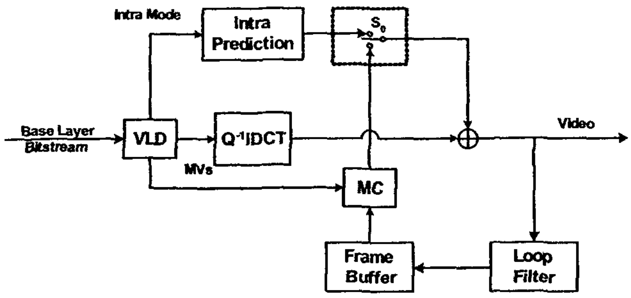

Figure 3 H.26L video decoder [7]

3. Simulation Model

Input to the H.26L encoder is a *Quarter Common Intermediate Format* (QCIF) video sequence which has 176 pixels/line and 144 lines/frame. The compressed video signal after the H.26L encoder will be modulated using a QPSK modulator and additive white Gaussian noise is added to the modulated signal. These corrupted symbols will be demodulated using a QPSK demodulator and fed into the H.26L decoder.

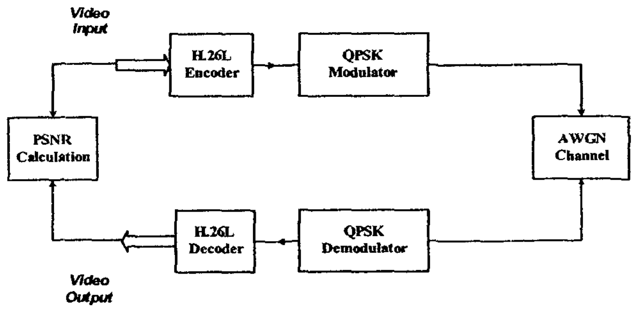

Figure 4 Simulation model.

3.1 Performance evaluation methods

There are two methods to assess the video quality: *subjective* and *objective*. Subjective test methods require human viewers, experts or non-experts and a controlled environment to rate the quality or difference in quality of two video clips. This type of assessment method consumes time and becomes costly. However, it provides the most accurate results than any objective measurement. On the contrary, objective test methods do not require human viewers or human intervention, but rather measure and analyze the video signal. The *peak signal to noise ratio* (PSNR) has traditionally been used to evaluate the relative quality of reconstructed image. It measures the distortion of an image relative to its reference image and defined as:

$$PSNR = 10\log_{10}\left[\frac{(2^n-1)^2}{MSE}\right] \quad (1)$$

where: n is the number of bits required to represent each pixel and MSE is the *mean squared error* between the distorted frame relative to the original frame. If we assume that the original frame is X, has MxN pixels, and the reconstructed frame is X'. The MSE is given as

$$MSE = \frac{\sum_{i,j}\left[X(i,j)-X'(i,j)\right]^2}{MxN} \quad (2)$$

If the number of bits required to represent each pixel is 8 bits, then PSNR is given in equation 3.

$$PSNR(dB) = 10\log_{10}\left[\frac{255^2}{MSE}\right] \quad (3)$$

It should be emphasized that PSNR measure does not equate with human subjective perception accurately (i.e., higher PSNR does not always means better quality and vice versa). PSNR, from the definition of PSNR, only captures a small portion of the video quality, the important temporal information of video sequences is omitted. Besides, many studies show that PSNR does not take visual masking in to account. This means that every erroneous pixel contributes to a decrease in PSNR even if the error mainly occur in the less important bits which decrease the PSNR, but the visual quality is still maintained. Video Quality Expert Group (VQEG) has recently used PSNR as the reference model in their study to evaluate the objective methods available for video quality assessment methods. It was observed that PSNR performs statistically equally good as more complicated objective methods. It is also noted that the actual value of PSNR is not meaningful, but the comparison between two values gives a measure of quality.

Therefore PSNR measure is selected as an objective method to evaluate the quality of the picture in the simulation models.

4. Results

In this section some results are presented. Performance of H.26L is evaluated by considering PSNR between uncompressed video and compressed video as shown in Figure 4. PSNR variation is shown in Figure 5 with different SNR values. For the comparison, performance of H.263 is also presented. Results show that performance of H.26L is better than H.263 results. Figure 6 shows the visual representation of H.26L and H.263 encoded video sequence (It shows only a single picture).

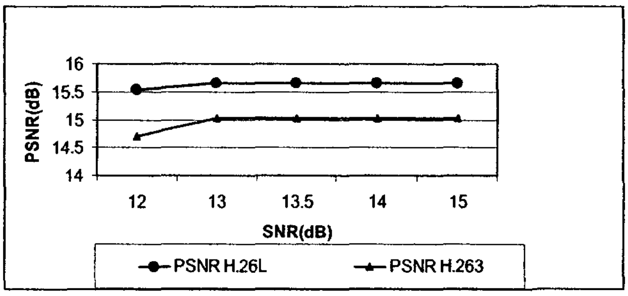

Figure 5 PSNR comparison between H.26L and H.263 standards

Figure 6 Visual representation of a sequence (shows only a single image) (Left image H.26L, Left image H.263)

5. Conclusion

In this paper performance of H.26L in an AWGN environment is considered. Results show that performance of H.26L is much better compared to the H.263. Further investigation is required to test the performance of the H.26L for a fading channel. Furthermore, this can be extended for an OFDM-CDMA system which has been under consideration for 4G mobile systems.

References

[1]Ghanbari, M. "Video Coding: An Introduction to Standard Codecs," *IEE Telecommunications Series 42*, 1999.

[2] Guy Côté, Berna Erol, Michael Gallant and Faouzi Kossentini, "H.263+: Video Coding at Low Bit Rates", *IEEE transactions on circuits and systems for video technology*, Vol.8, No.7, 1998.

[3] Berna Erol, Michael Gallant, Guy Cote and Faouzi Kossentini, "The H.263+ Video Coding Standard: Complexity and Performance", Department of Electrical Engineering, University of British Columbia.

[4] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, "Working Draft Number 2, Revision 2 (WD-2)", March 2002.

[5] Gary J. Sullivan, Thomas Wiegand, Thomas Stockhammer, "Using the draft H.26L video coding standard for mobile applications", *IEEE International Conference on Image Processing*, Vol. 3, pp. 573-576, 2001.

[6] Yadav, M. and Venkat, G., "Modeling and Simulation of H.26L Encoder", Literature Survey, March 2002.

[7] He, Y.; Wu, F.; Li, S.; Zhong, Y. and Yang, S., "H.26L-based Fine Granularity Scalable Video Coding", Microsoft Research Asia, 2001.