The Relevance of Scalability: From the MPEG-2 Standard to the Proposed Novel Paradigm of Scalable HEVC Video Coding

Pooja Soni
Assistant Professor
Department of Electronics & Communication Engineering, JIET-SETG
Jodhpur, Rajasthan, India.

Ravi Soni
M.E. Scholar
Department of Electrical Engineering, M.B.M. Engineering College
Jodhpur, Rajasthan, India.

ABSTRACT

The emergence of new visual communication systems has posed the need of compression of a huge amount of information into a limited bit-rate datastream for transmission and storage purposes; with an acceptable quality of decoded data, ensured at the users end. The flexibility of new communication systems and the success of multimedia databases require interactive capabilities for content-based access and manipulation of audio-visual information. The thriving and ever-strengthening interweave of communication, computing, networking and entertainment services has assumed a prominent place in everyday lives of the human civilization.

The paper brings forth the need of a scalable video coding extension of the upcoming HEVC video coding standard for spatial and quality scalable coding. Besides coding tools known from scalable profiles of prior video coding standards, it includes new coding tools that further improve the enhancement layer coding efficiency. This paper gives a concise overview of the MPEG-2 video compression standard, with due emphasis on the pertinent coding profile i.e. SNR scalable coding profile. The Quality measure is based on the objective video metric i.e. the peak signal-to-noise ratio (PSNR) for the various frames of the employed video sequences. MATLAB has been used as the computational environment to analyse the implemented code for the MPEG-2 scalable video coding profile.

Keywords:

Video Compression, MPEG-2 video coding standard, SNR, PSNR, HEVC, SVC

1. AN INTRODUCTION TO THE VIDEO COMPRESSION STANDARD: MPEG-2

While MPEG-1 was specifically proposed for digital storage media, the idea behind MPEG-2 was to provide a generic, application-independent standard. To this end, MPEG-2 follows a "tool kit" approach, providing a number of subsets, each containing different options from the set of all possible options contained in the standard. The MPEG-2 standard comes in three main parts: Systems, Video and Audio [3]. MPEG-2 extends the functions provided by MPEG-1 to enable efficient encoding of video and associated audio at a wider range of resolutions and bit rates. MPEG-2 is suitable for coding the progressive video format as well as the interlaced video format. As for the color subsampling formats, MPEG-2 supports 4:2:0, 4:2:2, and 4:4:4. MPEG-2 uses the 4:2:0 format as in MPEG-1 except that there is a difference in the positions of the chrominance samples [1].

Basically, MPEG-2 also supports compatibility and scalability with the MPEG-1 standard. Compatibility with the MPEG-1 video standard is another prominent feature provided by MPEG-2. For example, MPEG-2 decoders should be able to decode MPEG-1 bit streams. If scalable coding is used, the base layer of MPEG-2 signals can be decoded by a MPEG-1 decoder. Finally, it should allow reasonable complexity encoders and low-cost decoders be built with mature technology. MPEG-2 syntax can support a variety of rates and formats for various applications. Similar to other video coding standards, MPEG-2 defines only syntax and semantics. It does not specify all the encoding options (preprocessing, motion estimation, quantizer, rate-quality control, and other coding options) and decoding options (postprocessing and error concealment) to allow continuing technology improvement and product differentiation.

Since there are numerous possible applications ranging from streaming video to DVD to HDTV, MPEG-2 is divided into several *profiles* and *levels*. It is neither economically viable nor necessary to implement all possible coding options. Hence, each encoder may be designed to perform a portion of the available options. To meet this need, MPEG-2 classified the groups of features for important applications into profiles. A profile is defined as a specific subset of the MPEG-2 bit-stream syntax and functionality to support a class of applications (e.g., low-delay video conferencing applications or storage media applications). Within each profile, levels are defined to support applications that have different quality requirements (e.g., different resolutions). Levels are specified as a set of restrictions on some of the parameters (or their combination) such as sampling rates, frame resolutions, and bit rates in a profile. Applications are implemented in the allowed range of values of a particular profile at a particular level.

MPEG-2 aims at coding high-quality video at 4–15 Mbps for Video-on-Demand (VOD), standard definition (SD) and high-definition (HD) digital television broadcasting, and digital storage media such as digital versatile disc (DVD). The requirements from MPEG-2 applications mandate several important features of the compression algorithm. Regarding picture quality, MPEG-2 needs to be able to provide good NTSC quality video at a bit rate of about 4–6 Mbps and transparent NTSC quality video at a bit rate of about 8–10 Mbps. It also needs to provide the capability of random access and quick channel switching by means of I-pictures in GOPs. Low-delay mode is specified for delay-sensitive visual communication applications. MPEG-2 has scalable coding modes to support multiple grades of video quality, spatial resolutions, and frame rates for various applications.

Error resilience options include intramotion vector, data partitioning, and scalable coding.

2. THE SIGNIFICANCE OF SCALABILITY IN TERMS OF SNR AND SPATIAL REDUNDANCY

Scalable coding is also called layered coding. In scalable coding, the video is coded in a base layer and several enhancement layers. If only the base layer is decoded, basic video quality can be obtained. If the enhancement layers are also decoded, enhanced video quality (e.g., higher SNR, higher resolution, and higher frame rate) can be achieved [5]. Scalable coding is useful for transmission over noisy channel since the more important layers (e.g., the base layer) can be better protected and sent over a channel with better error performance. Scalable coding is also used in video transport over variable-bit rate channels. When the channel bandwidth is reduced, the less important enhancement layers may not be transmitted. It is also useful for progressive transmission, which means the users can get rough representations of the video fast with the base layer and then the video quality will be refined as more enhancement data arrive. Progressive transmission is useful for database browsing and image transmission over the Internet.

MPEG-2 supports three types of scalability modes: SNR, spatial, and temporal scalability. Different scalable modes can be combined into hybrid coding schemes such as hybrid spatial-temporal and hybrid spatial-SNR scalability. In a basic MPEG-2 scalability mode, there can be two layers of video: base and enhancement layers. The hybrid scalability allows up to three layers.

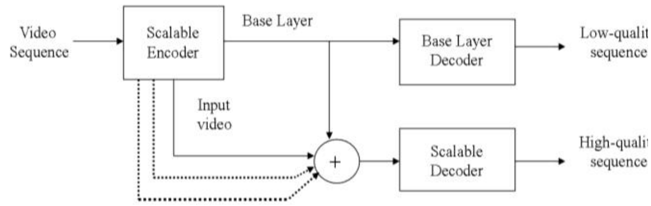

Fig. 1: Scalable or layered-coding system

3. 'SNR-SCALABLE' VIDEO CODING STRATEGY

Signal to Noise Ratio (SNR) scalability is similar to the 'Successive approximation mode' of JPEG. The picture is encoded in two layers, the lowest of which contains a "coarse" quality version of the picture and the highest of which contains enhancement information needed to decode the picture at its full quality.SNR scalability provides two (or more) different video layers of the same resolution but with different qualities. The base layer is coded by itself and provides a basic quality in terms of the SNR and PSNR. The enhancement layer is encoded to provide additional quality when added back to the base layer. The encoding is performed in two consecutive steps: first the base layer is encoded with a low quality, and then the difference between the decoded base layer and the input video is encoded with higher quality settings. At the receiver– side the base quality is obtained simply by decoding the base layer. For enhanced quality, the enhanced layer is decoded and the result is added to the base layer.

SNR scalability was introduced in the MPEG-2 standard. In this approach defined by the standard, motion compensation for a frame in the base layer requires a reference frame created by the base layer and the enhancement layer together. Error propagation may therefore take place when an enhancement layer frame is lost. Since the introduction of the MPEG-2 standard, many improved SNR scalable video coding schemes have been proposed. The main point requiring improvement in the standardized technique is the dependency of base layer frames on enhancement layer frames. No enhancement layer data participates in the encoding of the base layer. Moreover, there is no back-loop of any kind in the section that is responsible for enhancement layer encoding [9].

MPEG-2 SNR scalability provides two different video qualities from a single video source while maintaining the same spatial and temporal resolutions. In the base layer, the DCT coefficients are coarsely quantized and the coded bit stream is transmitted with moderate quality, at a lower bit rate. In the enhancement layer, the difference between the non-quantized DCT coefficients and the coarsely quantized DCT coefficients from the lower layer is encoded with finer quantization step sizes. By doing this, the moderate video quality can be achieved by decoding only the lower layer bit streams while a higher video quality can be achieved by decoding both the layers.

4. THE QUALITY METRICS

Image quality metrics provide a measure of the correlation between digital images, which in the context of videos are the frames, by exploiting the differences in the statistical distribution of pixel values in the images [10]. Image compression is a measure of the amount of insignificant data that has been discarded from the digital image. However, there exists a trade off between image compression and image quality. As the compression increases, the image quality consequently degrades [3].

There are two quantitative metrics which are commonly used to evaluate the image quality of compressed images: Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR). MSE measures the image difference between the decompressed image and the original image, whereas PSNR is the ratio of the peak signal power against the average noise power (MSE). Essentially, PSNR reflects the quality of the reconstructed image and is a standard method used to gauge the image fidelity. PSNR is usually measured in decibels (dB).

In addition to the above qualitative metrics, there are two metrics that are used to quantitatively assess the compression of an image, which are compression ratio and bit rate. Compression ratio can be defined as a simple ratio of the number of bits of the original uncompressed image to its compressed reconstructed version. Bit rate is defined as the average number of bits per pixel for the entire image. Their expressions are as follows:

$$\textbf{Compression Ratio} = \frac{\text{Number of bits in Original Image}}{\text{Number of bits in Compressed Image}}$$

$$\textbf{BitRate} = \frac{\text{Number of bits in Compressed Image}}{\text{Number of pixels}}$$

Two error analysis metrics used to evaluate the transmission of the digital images via error prone channels are, Bit Error Rate (BER) and Signal-to-Noise Ratio (SNR). Bit error rate can be broadly referred to as the measure of data integrity. It is measured empirically as the ratio of the number of erroneous bits received compared to the number of bits transmitted for some duration of time and is given by:

Bit Error Rate (BER) = $\frac{\text{Number of bits in Error}}{\text{Number of bits transmitted}}$

Signal-to-noise ratio is a measure of the signal strength relative to the background noise in the channel. It is used to gauge the quality of the transmission channel. Average SNR per bit in terms of digital communication is formally represented by E_b / N_0, which is the ratio of energy per bit (E_b) to spectral noise density (N_0) given by:

Average SNR per bit = $\mathbf{E_b / N_0}$

Since E_b/N_0 is independent of modulation schemes, it is used in the plots against bit error rate and helps compare schemes. E_b/N_0 is typically expressed logarithmically in decibels (*dB*). An E_b/N_0 of zero dB cannot be represented on the logarithmic scale and is thus an indication that the signal is unreadable and impossible to interpret as the noise level severely competes with the signal. The SNR has been referred to, in the results analysis and must be interpreted as average SNR per bit.

Given two discrete, 2-D signals x (m, n) and y (m, n), each with dimensions M x N, the mean square error (MSE) of the two signals is obtained by:

$$MSE(x, y) = \frac{1}{MN}\sum_{j=1}^{N} \sum_{i=1}^{M} \left(x_{ij} - y_{ij}\right)^2$$

PSNR is defined as:

$$PSNR\ (\text{dB}) = 10 \log_{10} \{(2^n - 1)^2 / \text{MSE}\}$$

Where, ***n*** is the number of bits required to represent each pixel and ***MSE*** is the Mean Squared Error between the distorted frame and the original frame. PSNR is relatively easy to calculate and provides a "rough" approximation to the visual quality of the video frame. In general, a high PSNR indicates a "high-quality" frame. Quality degradations due to, for example, high compression and/or transmission errors, lead to a drop in PSNR. It is possible to obtain an approximate comparison of the quality of two video sequences by comparing the PSNR of the frames in each sequence, relative to the original video sequence.

Calculating the average PSNR of all the frames in a video sequence gives a measure in dB of the "quality" of the sequence. However, PSNR does not accurately reflect visual quality. For example, two frames of video with the same PSNR may appear to have very different "qualities" to the viewer. Errors or impairments in a video sequence will cause a drop in PSNR but it is not possible to accurately map this drop in PSNR to the responses of a representative group of viewers. Higher PSNR implies better quality.

From the SNR and PSNR values obtained as a result; after the execution of the codes for the two MPEG-2 scalable coding profiles, it is evident that the spatially-scalable MPEG-2 video coding approach yields an enhanced resolution, while the SNR-scalable MPEG-2 video coding profile results in an enhanced overall quality of the video; depicted by the higher PSNR values obtained. Through the simulation of the pertinent codes used for the two profiles; for each video sequence, the computed performance parameters show an orderly variation in the differences of SNR and PSNR values for base layer and enhancement layer.

From the Table 1, it is evident that the relevant parameters for each video sequence depict that in the case of SNR-scalable approach, there is a significant difference between the SNR and PSNR values for the coded base and enhancement layers, for each frame of the video sequence.

The following section gives an account of the block diagrammatic illustrations of the algorithmic approach used in the SNR-scalable coding profile, using MPEG-2 video coding standard, through Figure 2.

5. THE ALGORITHM INVOLVED

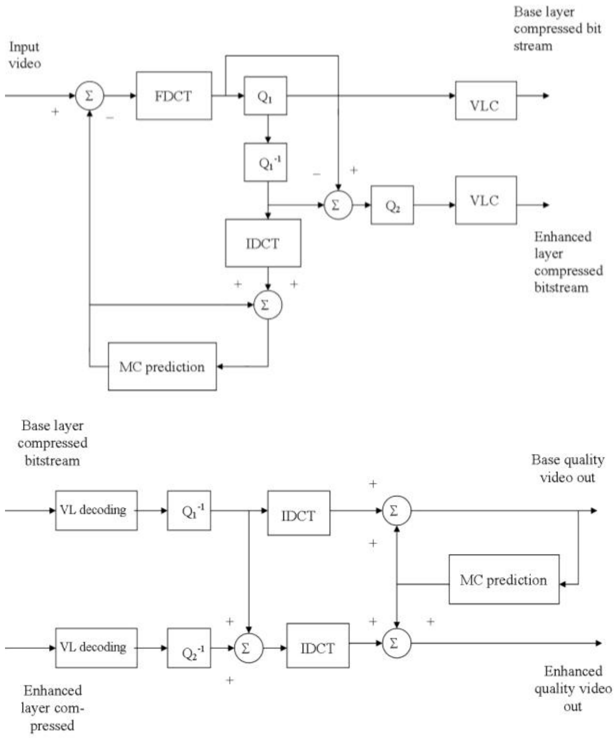

Fig. 2: Block diagram illustrating a SNR-scalable DCT-based MPEG video coding: (a) Encoder and (b) Decoder

6. THE ANALYZED SAMPLE VIDEOS

Table 1: Details of the parameters of video sequences used for analysis

Video Sequence No.	Name of the video	Resolution (H x W)	No. of frames considered for analysis	Length of the video
a.	Viplanedeparture.avi	240 x 360	300	11 sec
b.	Vipboard.avi	240 x 360	300	11 sec
c.	Vipmosaicking.avi	240 x 320	70	4 sec
d.	Vipunmarkedroad.avi	240 x 320	70	1min 23 sec
e.	Rhinos.avi	240 x 320	70	7 sec
f.	Foreman.avi	144 x 176	100	4 sec
g.	News.avi	144 x 176	100	11 sec

7. RESULT ANALYSIS

Table 2: The results obtained by the code for SNR-scalable MPEG-2 coding

Video Sequence No.	Name of Video	Average Base_snr	Average Enhanced_snr	Average Base_psnr	Average Enhanced_psnr
a.	Viplanedeparture.avi	20.41506788	34.32390448	35.77591445	49.68475104
b.	Vipboard.avi	11.61593708	32.71944512	26.5466689	47.65017694
c.	Vipmosaicking.avi	18.43265264	34.3888671	31.93771013	47.89392459
d.	Vipunmarkedroad.avi	15.55698236	30.1956579	32.51995184	47.15862738
e.	Rhinos.avi	15.08690159	30.95888673	31.71438876	47.58637391
f.	Foreman.avi	15.94233205	32.11783919	30.9725762	47.14808334
g.	News.avi	17.95900532	33.85377331	32.10002397	47.99479196

8. CONCLUSION

Digital video compression is a progressive field in which fundamental technologies have been motivated and driven by practical applications, such that they have often lead to many useful advances. The need for the compression of digital video remains undiminished despite substantial increases in bandwidth and storage capabilities over the past few years. Standards have evolved significantly during the past few decades, but the basic principles of video coding remain intact. Most codecs combine two fundamental techniques: transform coding (used to compress both original and residual frames) and motion-based prediction. MPEG-2 is a widely used standard for audio and video compression. Being capable of exploiting both spatial and temporal redundancies, it can achieve compression ratios up to 200:1 and can encode a video or audio source to almost any level of quality.

MPEG-2 standard, although developed long back, finds an application in wide range of prevailing video processing applications. Through the simulation of the pertinent codes used for the profile; for each video sequence, the computed performance parameters show an orderly variation in the differences of SNR and PSNR values for base layer and enhancement layer. From the observations, it can be inferred that the SNR scalable MPEG-2 video coding profile results in an enhanced overall quality of the video; depicted by the higher PSNR values obtained. The recognition of diverse video coding strategies is a problem of considerable practical utility, but it poses formidable technological challenges because of the difficulties encountered in the processing of video data of various types. There is definitely room for improvement and the new paradigms in computing will dictate future research directions. For the forthcoming years, video compression research and standardization efforts will be centered around 3D video quality and in achieving higher video compression efficiency alongwith higher bit-rate transmission possibilities.

So far, the development of scalable extension of the HEVC coding profile is a field of intensive research and it is largely expected that the further developments will lead to emergence of better video coding and compression strategies, with enhanced applications to diverse fields. Though, the formulation of such a standard is still in progress. Thus, through investigation of the prior video coding standard, it can be said that the inclusion of 'scalability' in the HEVC standard will lead to certain developments, achieving relatively better compression efficiencies than those that persist.

9. REFERENCES

[1] G. Cermak, M. Pinson, and S. Wolf: "The Relationship among Video Quality, Screen Resolution, and Bit Rate", IEEE Transactions on Broadcasting, Vol. 57, No.2, June 2011.

[2] S. Immanuel Alex Pandian, Dr. G. Josemin Bala and Becky Alma George: "A Study on Block Matching Algorithms for Motion Estimation", International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 1, Jan 2011.

[3] S.L. Lahudkar and R.K. Prasad: "Real Time Video Compression Implemented Using Adaptive Block Transfer/Motion Compensation for Lower Bit Rates", Journal of Engineering Research and Studies (JERS), Vol. II, Issue III, July-September; pg. 147-153, 2011.

[4] Fernando Pereira; "Video Compression: Discussing the Next Steps"; Proceedings of the IEEE, 2009.

[5] Wei Yao, Zheng Guo Li and Susanto Rahardja: "Scalable Video Coding in a Nutshell", Synthesis Journal 2008.

[6] Thomas Sikora: "Trends and Perspectives in Image and Video Coding", Proceedings of the IEEE, Vol. 93, No. 1, January 2005.

[7] Eman Abdelfattah and Asif Mohiuddin: "Performance Analysis of Multimedia Compression Algorithms", International Journal Of Computer Science & Information Technology (IJCSIT), Vol.2, No.5, October 2010.

[8] "Practical Image and Video Processing Using MATLAB", By: Oge Marques; IEEE Press, John Wiley & Sons Inc. Publication, 2011.

[9] "The Essential Guide to Video Processing", By: A. L. Bovik; Academic Press, an imprint of Elsevier Inc., 2009.

[10] "Digital Image processing using MATLAB", By: R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Pearson edition, Prentice-Hall, 2004.