www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 7 July 2015, Page No. 13468-13473

Video Compression – from Fundamentals to H.264 and H.265 Standards

Sruthi S[1], Dr. Shreelekshmi R[2]

[1]Computer Science and Engineering
LBS Institute of Technology for Women,
Trivandrum.
sruthi.soman@gmail.com

[2]Computer Science and Engineering,
LBS Institute of Technology for Women,
Trivandrum.
shreelekshmir@gmail.com

Abstract: ***In this era of modern technologies and Internet, Video Compression is extremely important for systematic storing and transmission of huge sized videos. In past two decades, there was a rapid development in the area of Video Compression Technology. To attain the best equilibrium between compression efficiency and perceptual quality, a number of video compression techniques have been developed. The first part of this survey paper gives an overview about Video Compression. The latter part gives a summary about different Video Compression Standards.***

Keywords: HVS, ISO, ITU, JPEG, MPEG, H.26x, HEVC.

1. Introduction

An uncompressed video may consume a huge amount of storage, transmission time and bandwidth. So before storing or transmitting the massive amount of data involved in video, it is better to be compressed to reduce the resource utilization. A Video Compression is an application of data compression and its objective is to remove redundant information from a video and to omit those parts of the video that will not be noticed by Human Visual System (HVS). The process of reconstructing the original video from the compressed video is known as Decompression. Both the Compressor (Encoder) and Decompressor (Decoder) together constitute a Codec.

2. History

The Morse code introduced in 1832 for the use in telegraphy is the earliest known example of data compression. The actual data compression work was started in 1940, with the development of Information Theory. In 1949, Shannon and Fano introduced a method of assigning code-words according to the probabilities of blocks. The image compression was started in the mid of 1970, by using Huffman Coding invented by David Huffman. In 1974 Ahmed, Natrajan and Rao Developed Discrete Cosine Transform (DCT). Later it had become a commonly used image compression technique. In 1977, Lampel, Ziv and Welch proposed a new compression algorithm known as LZW. During the mid of 1980, this algorithm became a popular image compression technique. In 1979, Ressanen developed an advanced version of Huffman Coding called Arithmetic Coding. Discrete Wavelet Transform (DWT) was introduced in 1984 by Alfred Haar. DWT can be used in image compression for sub-band decomposition. In early 1990's a joint committee known as Joint Photographic Experts Group (JPEG) developed a compression standard for continuous tone images known as JPEG Standard. In 1992, JPEG standard was formally accepted as an International Standard for image compression. Motion JPEG is the first method used to compress video. *H.261* defined was the first popular compression standard by ITU in 1990. *MPEG-1* published in 1992 was the first video compression standard by ISO. An improved version of MPEG-1 known as *H.262/MPEG-2*, released in 1994 was jointly defined by ISO and ITU. *H.263* is an ITU standard developed after H.261. MPEG-4 released in 1998 was defined by ISO. *H.264/MPEG-4 AVC* released in 2003 was a joint project done by ITU and ISO. *HEVC (High Efficiency Video Coding) / H.265* released in 2013 is the latest and more efficient video compression standard by ITU and ISO [1] [2] [3] [5].

3. Video Compression – Fundamentals

A video is visual multimedia source that contains sequence of still images (frames) displayed over time. The smallest individual unit of a video frame is called pixel which is usually represented as positive integers. Based on the representation, video frames can be classified as Black and White frames, Grayscale frames and Color frames. Each pixel in a black and white frame is represented by a single bit which is 0 or 1. Each pixel in a grayscale image is a luminance component usually represented by an 8-bit integer whose values range from 0 to 255. In color images, each pixel has both luminance and chrominance components. Color frames can be represented in different color spaces such as RGB, YUV etc [1] [2] [3].

Video compression techniques are primarily based on the fact that most of the pixels within a frame and the pixels in adjacent frames are highly correlated. So the amount of data needed to represent a video can be reduced by removing the redundancy exists within a frame and in between the frames. The redundancy exists in a video can be broadly classified as Spatial Redundancy and Temporal Redundancy.

- Spatial Redundancy: The redundancy exists within a frame is called Spatial Redundancy (Intra-Image or Intra-Frame Redundancy). Each frame of a video is an image which can be independently compressed. So the spatial redundancy can be removed from a video by applying digital image compression techniques on each video frame.

- Temporal Redundancy: Usually, there is a high correlation between frames of a video that were captured almost during the same time. The redundancy exists between the temporarily adjacent video frames is called Temporal Redundancy (Inter-image Redundancy or Inter-Frame redundancy).

The Video Compression can also be defined as the process of removing temporal redundancy as well as spatial redundancy within a video so as to reduce storage space, transmission time and bandwidth required for the video. The redundancy in a video can further be classified as Coding Redundancy, Inter-Pixel Redundancy and Psycho-Visual Redundancy.

- Coding Redundancy: A code-word (set of symbols used to represent a piece of information) may contain more data than actually needed. The presence of unnecessary symbols in a code-word causes coding redundancy.

- Inter-Pixel Redundancy: Due to the very high correlation between the neighboring pixels of an image, the value of one pixel can be easily predicted from the values of its neighboring pixels. So the information actually needed to be carried by a single pixel is very less and the remaining unnecessary information causes Inter-Pixel Redundancy.

- Psycho-Visual Redundancy: Human eye does not respond with equal sensitivity to all the arriving visual information. Thus some of the information may be ignored by the Human Visual System. This information are said to be psycho-visual redundant information.

Depending on the applications the Digital Video Compression algorithms can be Lossless or Lossy.

- Lossless Compression: In lossless compression, there will not be any difference between the original video and the reconstructed video. In other words, the integrity of video information is preserved in lossless compression. Lossless compression techniques are used in conditions where the data loss cannot be afforded.

- Lossy Compression: In lossy compression algorithms, the reconstructed image may contain degradation as compared to the original image. When comparing with lossless compression, much higher compression is possible in lossy compression.

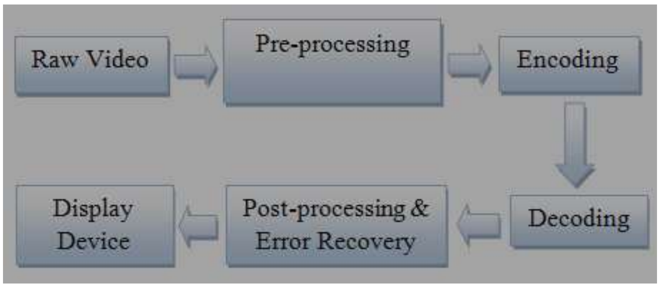

Figure 1: Video Codec - basic structure

Most of the video compression standards follow similar basic steps for encoding a video.

1. Divide each frame of the video into blocks of pixels
2. Identify and remove spatial redundancies within each frame
3. Exploit temporal redundancies between the adjacent frames and remove those redundancies.
4. Identify and remove the remaining spatial redundancies using quantization, transformation and entropy encoding

Most of the video compression standards aimed at removing correlation using motion compensation. In a video, the three major causes of motion are translation, rotation and zoom. Translation is the simple movement of objects from one position to another position. Rotation is the spinning objects about an axis. Zoom can be Zoom-In or Zoom-Out where Zoom-In causes the movement of object due to the increase in its size and Zoom-Out causes the movement due to the decrease in its size. Motion Estimation is the process of calculating the difference between the current frame and a previously reconstructed frame which is used as the reference. This difference is defined as Displacement Vector or Motion Vector (MV). Motion Compensation is the process of predicting the current frame using Motion Vector and the actual Current Frame. Compression ratio will be high if an efficient motion estimation method is used.

4. Video Compression – Constraints

While designing an efficient Video Codec, the following parameters have to be taken into consideration [2].

4.1 Quality

Different applications have different quality requirements. During compression, the quality of the compressed video should be taken into consideration and an acceptable level of video quality should be maintained based on the requirements. The objective measures such as Peak Signal-to-Noise Ratio (PSNR) are often used for comparing the video quality. PSNR measures the peak error between the compressed image and original image.

$$PSNR\ (dB) = [20 \log_{10} (MAX)] / (\sqrt{MSE}) \quad (1)$$

MAX represents the maximum pixel value of an image. For grey scale images, the maximum pixel value is 225. MSE (Mean Square Error) is the cumulative squared error between the compressed and reconstructed image.

$$MSE = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} [I(x,y) - I'(x,y)] \quad (2)$$

4.2 Time Complexity

Time complexity of algorithm is an important factor while executing a compression algorithm in a resource constraint system. The execution time should be optimum while running a compression algorithm on a given video. Usually it may take time to implement complex algorithms but this execution delay should not be too large.

Compression Speed = (Number of uncompressed bits) ∕ (Seconds to compress) (3)

4.3 Compression ratio

Compression ratio is defined as the ratio of the number of bits needed to represent a video before compression to the number of bits needed to represent the video after compression.

Compression Ratio = (Size of original image) ∕ (Size of compressed image) (4)

The evaluation of a compression technique can also be done using Space Savings.

Space Savings = 1 – [(Size of compressed image) ∕ (Size of original image)] (5)

Quality of image has to be compromised to obtain a better compression ratio. There will be a trade-off between quality and compression ratio. That means it is difficult to achieve both the best quality and high compression ratio simultaneously.

4.4 Bit-Rate

Bit rate is another way used for evaluating compression techniques. Bit rate is the total number of bits processed per second.

Bit Rate = (Total size of a video) ∕ (Time Duration of the video) (6)

5. Video Compression – Standards

In the past two decades, there was a quick development in Video Compression Technology. This section gives an overview about different compression standards available today. Two major organizations that define the image/video standards are: the *International Telecommunication Union (ITU)* that defines standards for telecommunication applications and the *International Standard Organization (ISO)* that defines compression standards for consumer applications. ISO cooperates with International Electro-technical Commission (IEC) for standardization in areas like Information Technology and it is often referred to as ISO/IEC. The *H.26x* standard used for video telephony was defined by ITU. The *JPEG (Joint Photographic Experts Group)* standards for image compression and *MPEG (Moving Picture Experts Group)* standards for video compression were defined by ISO.

Most of the currently available video coding standards are based on block based hybrid video coding technique. Each block is either intra-coded or inter-coded. For further compression, transformation coding is applied on this prediction error for exploiting the spatial redundancy. Even though all the video compression standards have similar basic structure, there is a significant improvement in coding efficiency from one generation to the next.

5.1 Motion JPEG (MJPEG)

Motion JPEG (MJPEG) [9] [10] is the first known method used to compress video. This method uses JPEG image compression standard to encode individual frames of a video. Here the video is represented as a sequence of JPEG images. In MJPEG, temporal redundancy is not exploited and it is not a standardized video compression method. This method was commonly used for network surveillance. In *Motion JPEG-2000 (MJPEG-2000)*, JPEG-2000 is used to represent the video frames. As compared to MJPEG, Motion JPEG-2000 is more complex, but has slightly better compression ratio.

5.2 H.261

H.261 [9] [10] defined by ITU in 1990 was the first widely accepted compression standard. This method was generally used for video conferencing and video telephony over ISDN. This is also known as *Px64*, since it was designed for ISDN where data rates are specified as multiples of 64 Kbps. It supports two picture formats: *CIF (*Common Intermediate Format) and *QCIF* (Quarter of Common Intermediate Format). The H.261 frame types are *I-frames* that are encoded without any reference to the previously coded frames and *P-frames* that are encoded with reference to a previously encoded frame. The basic steps in its coding process are prediction, block transformation, quantization and entropy encoding. H.261 uses a motion compensation algorithm which encodes the difference between neighboring frames and there by exploiting temporal redundancy. H.261 is designed for video telephony application which requires low delay and constant bit-rate.

5.3 MPEG-1

MPEG-1 [9] [10], released in 1993 is the first video compression standard by ISO. It is similar to H.261 standard with some improvements. MPEG1 supports coding at a rate of about 1.5 Mbps. This standard was used for storage and retrieval of audio and video on a digital storage media. MPEG-1 uses bi-directional prediction method which reduces noise. The different types of encoded pictures in MPEG-1 are I-frames, P-frames and B-frames. I-frames (intra-frames) are treated as independent image and it is encoded without any temporal prediction. For encoding I-frames, MPEG1 uses the same coding technique as that of H.261. Like H.261, P-frames (forward Predicted frames) are encoded using motion prediction using previous I or P frames. Both the motion vector and the prediction error are encoded and transmitted. For encoding the B-frames (Bi-directionally predicted frames) interpolated motion prediction is done with reference to previous I or P frames and the next I or P frames in the sequence. Adaptive Perceptual Quantization is used where the characteristics of HVS are taken into account. Compared to H.261, the complexity of MPEG-1 is high.

5.4 H.262/MPEG-2

H.262/MPEG-2[9] [10] released in 1994 was jointly defined by ISO and ITU. MPEG-2 was created on MPEG-1 with some improvements such as wider motion compensation and support for interlaced videos. It supports both interlaced and progressive videos and for supporting high resolution videos, wider search range is used. This method is commonly used for digital TV and DVD.

5.5 H.263

H.263 [4] is an ITU standard developed in 1995 as a refinement of H.261 by adding features such as Half-Pixel Motion Estimation, Unrestricted Motion Vector Mode, Advanced Prediction Mode and 3-D Variable Length Coding of DCT coefficients. Additional features are defined in annexes such as Annex D (Unrestricted Motion vectors), Annex E (Syntax-Based arithmetic coding), Annex F (Advanced Prediction Mode) and Annex G (PB picture). H.263 supports five picture formats: Sub-QCIF, QCIF, CIF, 4-CIF and 16-CIF.The H.263 standard focused on video telephony over PSTN. Efficiency of this algorithm is higher than that of H.261. The Conversational High Compression Profile (CHC) and High Latency Profile (HLP) are two H.236 profiles which delivers better coding efficiency. CHC which shows enhanced coding efficiency can be used for low delay applications. HLP which provides B-frame support is used for applications that can tolerate high coding delay.

5.6 H.264/MPEG-4 AVC

H.264/MPEG-4 AVC [5][6][7] released in 2003 is a joint project done by ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). These standards show significant improvement in intra coding and inter coding efficiency. It shows enhanced error robustness, and increased flexibility. It has efficient motion compensation and reduced bit-rate. Different block sizes are used for performing motion compensation which results in better video quality. The basic processing unit is 16x16 pixel macro blocks. The two entropy encoding methods used are CAVLC and CABAC. For all syntax elements, Context-Adaptive Variable-Length Coding (CAVLC) uses a single codeword set. Run-Length encoding is used to code the transformation coefficient. In Context-Adaptive Binary Arithmetic coding, statistics of previously coded symbols are used for encoding and it uses arithmetic coding for transmission. HDTV broadcasting, Internet Video, Video Conference etc are some of the applications of H.264/MPEG-4 AVC.

5.7 H.265/MPEG-HEVC

HEVC (High Efficiency Video Coding) [8] [10] [11] released in 2013 is a joint standardization project done by Joint Collaborative Team on Video Coding (JCT-VC) of ITU and ISO/IEC. Compared to H.264/MPEG-4 AVC, HEVC has so many new features. In this standard, Block-based hybrid coding is used. HEVC uses a flexible coding block structure where the basic block is known as Larger Coding Units (LCU) which can be recursively split into Coding Units (CUs). The CUs further divided into Prediction Units (PU) and Transform units (TU). Both larger and smaller encoding blocks are used in HEVC where smaller blocks are used for encoding uneven surfaces. The PUs are classified as symmetric PUs for both inter-prediction and intra-prediction and asymmetric PUs for inter-prediction only. Transform and quantization are applied on the Transform Units. In HEVC, both slices and tiles can be used within the same frame for increasing flexibility. Each tile can be further divided into Entropy Tiles to support parallel processing. Block-based Intra-frame coding is used in HEVC and it has 35 luma intra-prediction modes and six chroma intra-prediction modes. For smooth areas, planar intra-prediction mode is used. Both Symmetric and Asymmetric Motion Partitioning (AMPs) are used in HEVC. By using AMPs, it can attain better coding efficiency. One dimensional 7-tap or 8-tap filters are used for generating ½ pixel and ¼ pixel luma samples. In HEVC, motion vector prediction is either spatial or temporal. Transform such as Discrete Cosine Transform is used in HEVC. The entropy encoding modes in HEVC are *High Efficiency Binarization (HEB)* which is completely grounded on *Context Adaptive Binary Arithmetic Coding (CABAC)* and *High Throughput Binarization (HTB)* which is partially grounded on *Context Adaptive Variable Length coding (CAVLC)*. The best qualities of CABAC and CAVLC such as high efficiency and low complexity respectively are used in entropy encoding.

Table 1: Comparison between H.264/AVC and H.265/HEVC

Features	*H.264/AVC*	*H.265/HEVC*
Approved Date	2003	2013
Block Structure	Macro block (maximum size 16x16)	Coding Tree Units (64x64)
Directional Modes for Intra-Prediction	9	35
Deblocking Filter	Present	Present
Bit-rate reduction	50% reduction when compared H.263	40%-50% when reduction compared H.264/AVC
Entropy Encoding	CAVLC/CABAC	CABAC
Motion Compensation	Motion Vector Prediction (MVP)	Advanced Motion Vector Prediction (AMVP)
Motion Vector	6-tap filtering of half-sample position followed by linear interpolation for quarter sample position.	7-tap or 8-tap filters are used for interpolation of fractional sample position.
Ultra High Definition (UHD)	Does not support UHD. Supports up to 4K and 59.94fps only	Supports up to 8K and 300fps

Table-1 shows the comparison between H.264/AVC and H.265/HEVC. The H.265/HEVC standard uses Enhanced Hybrid Spatial-temporal prediction model. In HEVC, the deblocking filter design is simplified and video quality identical to H.264/AVC is ensured at only half the bit-rate.

Bit-Rate	*PSNR-YUV*			
	H.265/ HEVC	H.264/ MPEG-4 AVC	H.263	MPEG -2/ H.262
2000	40.25	38.75	36.25	35.25
3000	41.35	40.10	37.9	37.25
4000	42	40.9	39.1	38.6
5000	42.4	41.5	39.9	39.5
6000	42.65	41.8	40.5	40
7000	42.9	42.15	41	40.6

Improved motion compensation is used in HEVC. These enhancements allow this standard to supports Ultra High Definition videos.

6. Video Compression Standards – Performance Comparison

References [9] [11] provides comparison of different video compression standards. In [10], a detailed comparison of HEVC with other video coding standards is given for interactive and entertainment applications.

Interactive applications such as video conferencing demand low coding delay and all frames need to be encoded in the display sequence. First frame is coded as I-frame and all succeeding frames are predicted with reference to previous frames in the sequence. The Bit Rate versus PSNR-YUV for the interactive video Kristen and Sara (1280x720, 60Hz) is presented in table-2. These results point out that, for interactive applications, the HEVC standard clearly performs better than its forerunners in terms of coding efficiency.

Table 2: [10] Bit-rate versus PSNR-YUV for interactive video Kristen and Sara

Bit-Rate	*PSNR-YUV*			
	H.265/ HEVC	H.264/ MPEG-4 AVC	H.263	MPEG-2/ H.262
1000	43	41.9	40	37.9
1500	44	43	41.75	40
2000	44.4	43.6	42.6	40.9
2500	44.6	43.9	43.2	41.6
3000	44.8	44.2	43.6	42.2
3500	45	44.4	43.9	42.7

Table 3: [10] Average Bit-Rate Savings for Equal PSNR for a set of selected interactive Applications

Encoding		*Compression Standards*		
		HEVC	H.264	H.263
Bit-Rate Savings relative to	H.264	40.3%	---	---
	H.263	67.9%	46.8%	---
	H.262	80.1%	67.0%	37.4%

The Table 3 summarizes the average bit-rate savings for uniform PSNR for a set of selected interactive applications. For interactive applications, the average bit-rate savings of the emerging standard HEVC is 35.4% higher compared to the prior popular standard H.264/MPEG-4 AVC.

Usually entertainment applications have diminished delay constraints. The Table 4 shows the Bit Rate versus PSNR-YUV for the entertainment video Kimomo1 (1920x1080, 24Hz).

Table 4: [10] Bit-rate versus PSNR-YUV for entertainment video Kimono1

Table 5: [10] Average Bit-Rate Savings for Equal PSNR for a set of selected entertainment Applications

Encoding		*Compression Standards*		
		HEVC	H.264	H.263
Bit-Rate Savings relative to	H.264	35.4%	---	---
	H.263	65.1%	46.6%	---
	H.262	70.8%	55.4%	16.2%

The Table 5 gives a summary of the average bit-rate savings for identical PSNR for a set of selected entertainment applications. For entertainment applications, the average bit-rate savings of the emerging standard HEVC is 35.4% higher compared to the prior popular standard H.264/MPEG-4 AVC.

These results show that, in terms of coding efficiency, the performance of HEVC standard in both interactive and entertainment applications are better compared to its forerunners.

7. Conclusion

From this survey, it can be concluded that due to high bit rate requirements, H.264 AVC is not practical for the distribution of UHD (Ultra High Definition) contents. The emerging standard HEVC provides a remarkable increase in coding efficiency and can supports up to 8K UHD videos. Because of larger Prediction units, costly Motion Estimation and improved flexibility, HEVC is computationally expensive compared to its prior standards. Most of the currently available video compression standards are too complex and time consuming and this leaves room for improvement in the area of Video Compression.

References

[1] Gonzalez, R. C., Woods, R. E. “Digital Image Processing”. Prentice Hall, Upper Saddle River, NJ , 3rd edition,2008

[2] Khalid Sayood, “Introduction to Data Compression”, 3rd edition, Morgan Kaufmann Series in Multimedia Information and Systems, Elsevier 2006, pp 1-680, 2006.

[3] Guy Cote, Lowell Winger, “Recent Advances in Video Compression Standards”, IEEE Canadian Review Spring / Printemps 2002, pp 21-24, 2002

[4] Karel Rijkse, “H.263: Video Coding for Low Bit-Rate Communication”, IEEE Communications Magazine, December 1996, , pp 42-48, 1996

[5] Jorn Ostermann, Jan Bormans, Peter List, Detlev Marpe, Matthias Narroschke, Fernando Pereira, Thomas Stockhammer, and Thomas Wedi, “Video coding with H.264/AVC: Tools, Performance, and Complexity", IEEE Circuits and Systems Magazine 2004, pp 7-28, 2004

[6] Gary A. Sullivan and Thomas Wiegand, ―Video compression—from concepts to the H.264 /AVC standard‖, IEEE International Conference on image processing, Vol. 93, pp. 521-524, January 2005.

[7] Detlev Marpe, Thomas Wiegand, Gary J. Sullivan "The H.264/MPEG4 Advanced Video Coding Standard and its Applications", IEEE Communicationss Magazine, August 2006, , pp 134-143, 2006

[8] Mahsa T. Pourazad, Colin Doutre, Maryam Azimi, and Panos Nasiopoulos "HEVC: The new Gold Standard for Video Compression", IEEE Consumer Electronics Magazine July 2012, pp 36-46, 2012

[9] Thomas Wiegand, Heiko Schwarz, Anthony Joch, Faouzi Kossentini, Gary J, Sullivan, "Rate-Constrained Coder Control and Comparison of Video Coding Standards". IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, July 2003, pp.688-703, 2003

[10] Jens Reiner Ohm, Gary A. Sullivan, Heiko Schwarz, Thiow Keng Tan, Thomas Wiegand."Comparison of Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)". IEEE transactions on circuits and systems for video technology, Vol. 22, No. 12, December 2012, pp.1669-1684, 2012

[11] Shivam Bindal, Udit Khanna, Manoj Sharma, "Trends in Video Compression Technologies and Detailed Performance Comparison of H.264/MPEG-AVC and H.265/MPEG-HEVC", International Journal of Engineering Research & Technology (IJERT) Vol. 3 IJERT Issue 12, pp 748-754, December-2014