International Journal of advancement in electronics and computer engineering (IJAECE)
Volume 1, Issue1, April 2012, pp.32-38, ISSN 2278 - 1412

Review of Different Standards for Digital Video Compression Technique

Rutika Joshi[1], Rajesh Kumar Rai[2], Jigar Ratnottar[3]

[1]*M. Tech Scholar, NIIST, RGTU,rutika.h.joshi@gmail.com, Bhopal (M.P.) India;*
[2] *Professor, Electronics Department, NIIST, RGTU,raj.rai1008@gmail.com, Bhopal (M.P) India;*
[3]*M.Tech Scholar, RGTU, jb.ratnottar@gmail.com, India.*

***Abstract** – Over the past decades, digital video compression technologies have become an integral part of the way we create, communicate and consume visual information. Digital video communication can be found today in many application sceneries such as broadcast services over satellite and terrestrial channels, digital video storage, wires and wireless conversational services and etc. The data quantity is very large for the digital video and the memory of the storage devices and the bandwidth of the transmission channel are not infinite, so it is not practical for us to store the full digital video without processing. For instance, we have a 720 x 480 pixels per frame, 30 frames per second, total 90 minutes full color video, then the full data quantity of this video is about 167.96 G bytes. Thus, several video compression algorithms had been developed to reduce the data quantity and provide the acceptable quality as possible as can. This paper starts with an explanation of the basic concepts of video compression algorithms and then introduces several video compression standards.*

***Keywords**: Video Compression, MPEG (Motion Picture Expert Group), H.264/AVC, Mean Square Error (MSE)*

I. Introduction

Why an image can be compressed? The reason is that the correlation between one pixel and its neighbor pixels is very high, or we can say that the values of one pixel and its adjacent pixels are very similar. This is called the intraframe correlation in video compression because it is the correlation in a single frame. Once the correlation between the pixels is reduced, we can reduce the storage quantity. The image compression method is also applied to compression video. However, there still exists temporal correlation. The video is composed of a large number of still images, and due to the images are taken at short time distance, the two neighboring images are similar. Therefore, we know that there exists high correlation between the images or frames in the time direction. The correlation in the time direction is called the interframe correlation. If we can efficiently reduce the interframe correlation, then video compression can be achieved.

Several methods to reduce the interframe and intraframe correlation will be introduced in the following sections. That paper introduces the different standards for MPEG based system also.

II. Video Quality Measure

In order to evaluate the performance of video compression coding [1], it is necessary to define a measure to compare the original video and the video after compressed. Most video compression [2] systems are designed to minimize the *mean square error (MSE)* between two video sequences Ψ1 and Ψ2 which is defined as,

$$MSE = \sigma_e^2 = \frac{1}{N}\sum_t \sum_{x,y}[\Psi_1(x,y,t) - \Psi_2(x,y,t)]^2 \qquad (1)$$

Where N is the total number of frames in either video sequences. Instead of the MSE, the *peak-signal-to-noise* *ratio (PSNR)* in decibel (dB) is more often used as a quality measure in video coding, which is defined as

$$PSNR = 20\log_{10}\frac{255}{MSE} \quad (2)$$

It is worth noting that one should compute the MSE between corresponding frames, average the resulting MSE values over all frames, and finally convert the MSE value to PSNR.

III. The Exhaustive Block-Matching Algorithm

Given a macro block in the anchor block *Bm*, the motion estimation is to determine a matching block *Bm*' in the target frame such that the error *D(s, t)* between the two blocks is minimizes. The most straight forward method is the *exhaustive block-matching algorithm (EBMA)*. The procedure of EBMA [3] is shown in Fig. 1

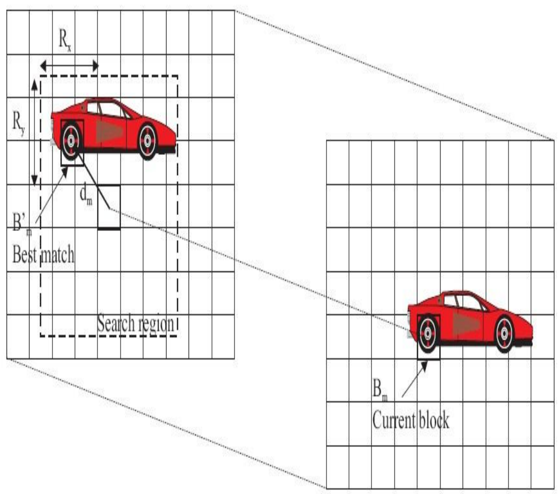

Fig.1. The search procedure of the exhaustive block-matching algorithm

In the simplest case, the step size is one pixel in both vertical and horizontal direction, which is known as *integer-pel accuracy search*. Let the block size be N×N pixels, and the search range be ±R pixels in both vertical and horizontal direction. The total number of operations for a complete frame is $M^2\ (2R+1)^2$, which is a large amount of computation when M and R is large.

III.1. Fast Block-Matching Algorithm

As shown previous, the EBMA requires a large amount of computation. To speed up the search, various fast algorithms for block matching have been developed. The key to reducing the computation is reducing the number of search candidates. Various fast algorithms differ in the ways that they skip those candidates that unlikely to have small errors. The following subsections describe two of the most well-known fast algorithms.

III.1.1 2-D Logarithm Search Method

One most well-known and simple fast algorithm is the *2-D logarithm search*, which is shown in Fig. 2 The first step is to compute the matching criteria for five points in the search window. The one corresponding to the minimum dissimilarity is picked up as the winner. In the next step, surrounding this winner, another set of five points is selected in a similar fashion to that in the first step. This procedure continues until the final step, in which a set of candidate points are located in a 3×3 2-D grid.

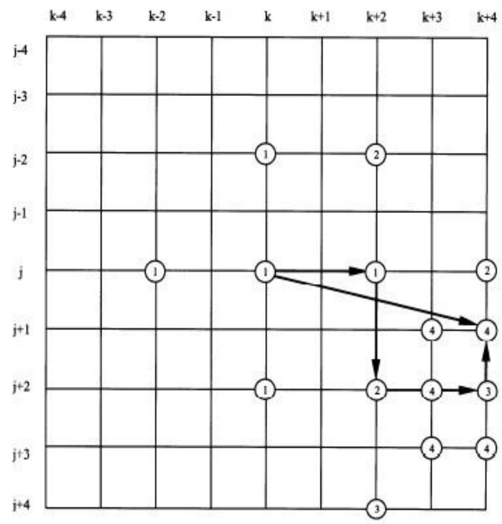

Fig.2. The 2-D logarithm search method

IV. Multi resolution Motion Estimation

The previous section describes the fixed sized block matching method. On the contrary, the variable block size method is used in this section because it gives more efficient motion estimation than the fixed sized method. One well known example is the multi resolution structure, also known as a pyramid structure, is a very powerful computational configuration for various image and video processing tasks.

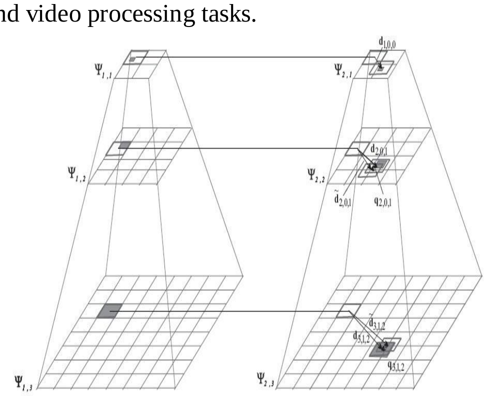

Fig.3. Illustration of the hierarchical block matching algorithm

As illustrated in Fig. 3 pyramidal representations of the two raw image frames are derived, in which each level is a reduced-resolution representation of the lower level.

IV.1. Basic Approaches to Motion Estimation

There exist two basic approaches to motion estimation

a) Pixel based motion estimation
b) Block-based motion estimation.

The pixel based motion estimation approach seeks to determine motion vectors for every pixel in the image. This is also referred to as the optical flow method, which works on the fundamental assumption of brightness constancy that is the intensity of a pixel remains constant, when it is displaced. However, no unique match for a pixel in the reference frame is found in the direction normal to the intensity gradient. It is for this reason that an additional constraint is also introduced in terms of the smoothness of velocity (or displacement) vectors in the neighborhood. The *smoothness constraint* makes the algorithm interactive and requires excessively large computation time, making it unsuitable for practical and real time implementation.

An alternative and faster approach is the block based motion estimation. In this method, the candidates frame is divided into non-overlapping blocks (of size 16 x 16, or 8 x8 or even 4 x 4 pixels in the recent standards) and for each such candidate block, the best motion vector is determined in the reference frame. Here, a single motion vector is computed for the entire block, whereby we make an inherent assumption that the entire block undergoes translational motion. This assumption is reasonably valid, except for the object boundaries and smaller block size leads to better motion estimation and compression.

Block based motion estimation is accepted in all the video coding standards proposed till date. It is easy to implement in hardware and real time motion estimation and prediction is possible.

V. MPEG-1

The main purpose of MPEG-1 video is to code moving image sequences or video signals. To achieve high compression ratio, both intraframe redundancy and interframe redundancy should be removed. There, the MPEG-1 video algorithm is mainly based on DCT and interframe motion compensation. The algorithm of the MPEG-1 will be described in the following subsections.

V.1. Layered Structure Based on Group of Pictures

In MPEG-1 and the video compression standards followed, the video sequence is first divided into a *group of pictures*, which is often called *GOP*. One example of GOP is shown in Fig. 4. Each GOP can consist of three types of frames:

- *Intra coded Frame (I-frame)*: I-frame is entirely coded in one frame by intraframe technique such as DCT. This type of frame no need for previous information.
- *Predictive Frame (P-frame)*: P-frame is coded using one-directional motion-compensated prediction from a previous frame, which can be either I-frame or P-frame. P-frame is generally referred to as inter-frame.
- *Bidirectional predictive frame (B-frame)*: B-frame is coded using bi-directional motion compensated prediction from a previous frame or future frame. The reference can be either I-frame or P-frame. B-frame is also referred to as inter-frame.

The distance between two nearest I-frame is denoted by N, and the distance between the nearest I-frame and P-frame is denoted by M.

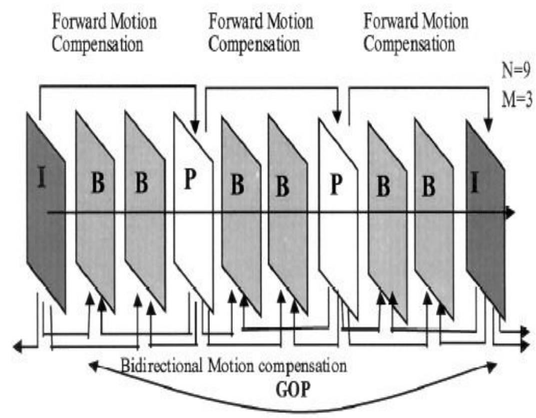

Fig. 4. A group of frames

V.2. Encoded Structure of MPEG-1

The encoder structure of MPEG-1 is shown in Fig.5

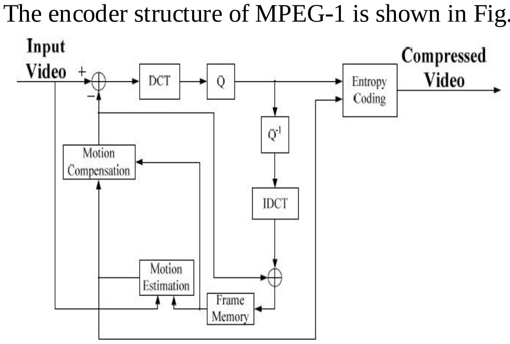

Fig.5. The Encoded Structure of MPEG-1

International Journal of advancement in electronics and computer engineering (IJAECE)
Volume 1, Issue1, April 2012, pp.32-38, ISSN 2278 - 1412

The operation of encoding can be summarized as the following steps:

Differential Coding: The input image is subtracted from the predictive image, and the differential image is generated. This can remove temporal correlation. The operation can be represented as

$$D(t) = \Psi(t) - \hat{\Psi}(t) \quad (3)$$

$\Psi(t)$ is the input image and $\hat{\Psi}(t)$ the predictive image.

DCT: Although the temporal correlation is removed, the spatial redundancy still exists. Therefore, MPEG-1 exploits 2-D DCT to remove the spatial correlation.

Quantization: The transform coefficients are then quantized. The function of quantization matrix is to quantize the high frequencies with coarse quantization steps because the human visual perception is less sensitive to the high frequencies. The bits are saved for coding high frequencies are used for low frequencies. MPEG-1 defines two quantization tables, the intra quantizer weighting matrix and the non intra quantization weighting matrix.

Motion Estimation and Motion Compensation: The quantization coefficients are de quantized and take the IDCT to reconstruct the approximation of the differential image. The encoder then takes motion estimation based on the input image and the reconstruction image to generate the motion vectors for each macro block. Finally, motion compensation is executed based on the reconstruction image and the motion vectors to produce the new predictive image. The motion compensation is to map the macro blocks from the previous frame to the compensated frame as shown in Fig. 6

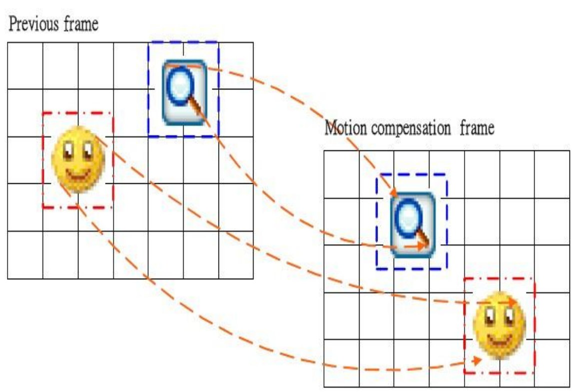

Fig. 6. The motion compensation

VI. MPEG-2

The MPEG-2 compression standard evolved to meet the needs of compressing higher quality video. MPEG-2 [4] is used in today's video DVDs and digital broadcasts via satellite and cable. It uses bit rates typically ranging from 5 to 8 Mbits/s, although MPEG-2 is not necessarily limited to a bit rate range. MPEG-2's basic compression techniques are very similar to MPEG-1 using DCT transforms but it also provides support for interlaced video (the format used by broadcast TV systems).

TABLE I

Level	Profile			
	Simple 4:2:0	**Main 4:2:0**	**Spatial scalable 4:2:0**	**High 4:2:0 or 4:2:2**
High 1920*1152 (60 frames/sec)		80 Mbps		100 Mbps For 3 layers
Main 720*576 (30 frames/sec)	15 Mbps	15 Mbps		20 Mbps For 3 layers
Low 352*288 (30 frames/sec)		4 Mbps		

PROFILES & LEVELS

Interlaced video causes less visible display flickering on a CRT monitor than non-interlaced methods by alternating between drawing the even-numbered lines and the odd-numbered lines of each picture. In contrast, a non-interlaced raster display draws every line of a picture or frame, in sequence, from top to bottom. This takes a certain amount of time, during which time the image on the CRT begins to decay, resulting in flicker.

An interlaced display reduces this flicker effect by drawing first all the even-numbered lines (forming the even field), leaving spaces between them for all the odd-numbered lines (forming the odd field), which it fills in afterwards to complete the frame. This results in the display being refreshed from top to bottom twice as frequently as in the non-interlaced case.

Although MPEG-2 video is not optimized for low bit rates (less than 1 Mbit/s), its quality outperforms MPEG-1 at 3 Mbits/s and above. MPEG-2 also introduces and defines Transport Streams, which are used in broadcast applications and are also designed to carry digital video and audio over unreliable media such as may be found in combat areas. With some enhancements, MPEG-2 is also the current standard for High Definition Television (HDTV) transmission. MPEG-2 includes additional color sub sampling, improved compression, error correction and multi-channel extensions for surround sound.

Although MPEG-2 compression excels at full broadcast television and can be used to retrieve and control streams from a server, just like MPEG-1 compression, MPEG-2 audio and video compression are still essentially linear and interactivity is limited to operations such as slow motion, frame-by frame or fast forward.

VII. MPEG-3

MPEG-3 is the compression standard that never was. While the MPEG committee originally intended that an MPEG-3 standard would evolve to support HDTV, it turned out that this could be done with minor changes to MPEG-2. So MPEG-3 never happened, and now there are Profiles of MPEG-2 that support HDTV as well as Standard Definition Television (SDTV).

VIII. MPEG-4

MPEG-4 has emerged as much more than a video and audio compression and decompression standard. The MPEG committee designed MPEG-4 to be a single standard covering the entire digital media workflow from capture, authoring and editing to encoding, distribution, and playback and archiving. It is a container for all types of items called media objects beyond audio and video. Media objects can be text, still images, graphic animation, buttons, Web links and so on. These media objects can be combined to create polished interactive presentations. The MPEG-4 architecture is shown in Figure 7.

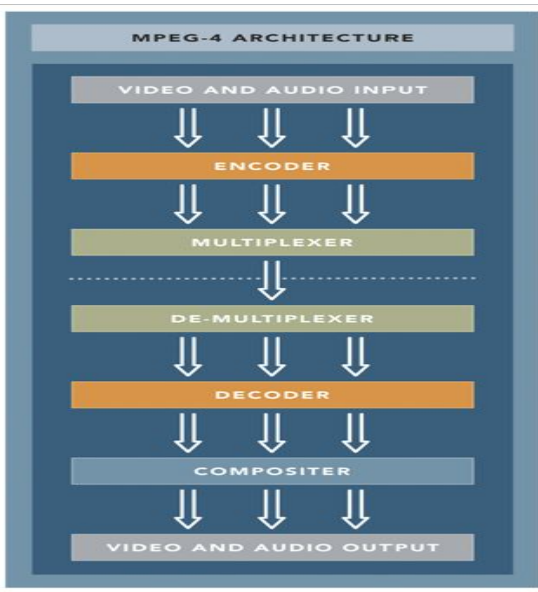

Fig.7. MPEG-4 Encoding & Decoding Architecture

The MPEG-4 file format, based on Apple Computer's QuickTime technology, was developed by the MPEG committee as a standard designed to deliver interactive multimedia and graphics applications over networks and to guarantee seamless delivery of high-quality audio and video over IP-based networks and the Internet. A major goal of the MPEG-4 standard was to try to solve two video transport problems: sending video over low-bandwidth channels such as the Internet and video cell phones, and achieving better compression than MPEG-2 for broadcast signals.

MPEG-4 functions well in terms of compression and it is used in a wide range of bit rates, from 64 Kbits/s to 1,800 Mbits/s. However, it had limited success in achieving dramatically better compression than MPEG-2 for broadcast signals, and although it is in the range of 15% better at compressing video data than MPEG-2, this has not been enough of an advantage to convert the whole broadcast industry to MPEG-4. So, MPEG-4's role will likely remain in lower-bandwidth applications in the desktop computer, Internet and cell phone worlds, plus new applications where a 15% compression improvement over MPEG-2 is desired and MPEG-2 compliance is not an issue.

MPEG-4 is actually a super-set of MPEG-2, so MPEG-4 players, which decompress the video stream, can theoretically play both MPEG-2 and MPEG-4 formats. However, for secondary reasons, this may not always be true in practice.

IX. H.264/AVC

With MPEG-4 failing to considerably improve compression performance for full broadcast signaling, another effort was initiated late in the 1990s. This new effort, H.264, is able to achieve a 2:1 improvement over MPEG-2 on full-quality SDTV and HDTV, and it is expected to come into wide use in satellite and cable TV over the next decade.

H.264/MPEG4-AVC [5] is a jointly developed standard by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) and has been standardized by the ITU under the H.264 name. It is also called MPEG-4 Part 10 AVC (Advanced Video Compression), even though it is unrelated in operation to MPEG-4.

The main goals of the H.264/MPEG-4 AVC standardization effort are to provide significantly enhanced compression performance and provision of a network-friendly packet-based video representation addressing conversational (video telephony) and non conversational (storage, broadcast or streaming) applications.

H.264 uses techniques fairly different from MPEG-2 and can match the best MPEG-2 quality at up to half the data rate. H.264 also delivers excellent video quality across the entire bandwidth spectrum from 3G to HDTV and everything in-between (from 40 Kbits/s to upwards of 10 Mbits/s). Efficient encoders and decoders for H.264 are just coming into use in 2005.

The H.264 design incorporates a Video Coding Layer (VCL), which provides the core high compression of the video content, and a Network Abstraction Layer (NAL), which packages that compressed content for delivery over networks. The VCL design has achieved a significant improvement in rate-distortion efficiency providing nearly a factor of two in bit rate savings against existing standards. The NAL designs are being developed to transport the coded video data over existing and future networks such as circuit-switched wired networks, MPEG-2/H.222.0 transport streams, IP networks and 3G wireless systems.

H.264 contains a number of features that allow it to compress video much more effectively than older codecs. Key features of H.264 are shown in the accompanying sidebar, Features of the H.264 Video Compression Standard.

X. Conclusion

In this paper, we have introduced the fundamental concepts of video compression and the characteristics of various video compression standards. Although the existed video compression standards can compress the video effectively, it still leaves room for improvement. For example, for reducing the temporal and spatial redundancy, Block Matching Algorithm and DCT has been exploited. On the other hand, the Exhausted BMA needs a large amount of computation. In H.264, the concept of variable block size has been proposed, which increases the computation cost more.

These are a few of the many challenges in video compression to be fully resolved and may affect the compression performance in the years to come.

References

[1] Yun Q.Shi and Huifang Sun, "*Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards*", CRC press, 2000.

[2] G. Sullivan and T. Wiegand, "Video Compression - From Concepts to the, H.264/AVC Standard", *Proceedings of the IEEE*, Special Issue on Advances in Video Coding and Delivery, December 2004.

[3] Yao Wand, Jorn Ostermann and Ya-Qin Zhang, "*Video Processing and Communications*", Prentice Hall, 2007.

[4] Barry G, Haskell, Atul Puri and Arun N. Netravali, "*Digital Video: An Introduction to MPEG-2*", Boston : Kluwer Academic, 1999.

[5] G. Sullivan and T. Wiegand, "Video Compression - From Concepts to the H.264/AVC Standard", *Proceedings of the IEEE*, Special Issue on Advances in Video Coding and Delivery, December 2004.

Author's Profile

Rutika H. Joshi, *born on 27th December,1986 at Rajkot,Gujarat,India. She has completed his SSC & HSC studies in recently living city Rajkot. She received her B.E degree in Electronics & communication from SAURASHTRA UNIVERSITY. Rajkot. (GUJRAT) and Research scholar from NRI, institute of science & technology, Bhopal in Digital Communication. She has already published her paper in AES Journal, IJETAE journal and also IEEE conference will be organized in Rajkot, May-2012.*

Rajesh Kumar Rai *received M. E. (Elect) Degree with specialization in Digital Techniques & Instrumentation from S.G.S.I.T.S. Indore in June 2008. His Research interests are Image Processing, Embedded System & Communication. He is Ph.D scholar in JJT University, Rajasthan. He has worked as a Assistance Professor & Head of Electronics Department in Siddhant College of Engineering, Pune, affiliated to University of Pune, Pune (India).*
Presently he is associated with NIIST, RGTU, Bhopal as a Associate professor in Department of Electronics & Communication.
Life time member of IEEE & ISTE.Published 8 international papers

Jigar Ratnottar, *born on 26th December, 1984 at Limbdi, Dist. Surendranagar. Recently lived in Rajkot, India.*
He completed his SSC from one of the well known institute of Rajkot "Virani Vividhlakshi Vidhyalay", also he completed his HSC level education at R.C.Patel school at Gandhinagar and got the admission in Electronics & Communication branch in Govt. Engg. College, Modasa, Gujarat, India. So in 2007 he got his graduation degree. Currently he pursuing in final semester of his M.Tech from RGPV University.
Jigar already presented his paper in National conference held in Nagpur, India -2012. He also publish his paper on " Three step search motion estimation algorithm" in AES international journal at Sangali, Maharastra. Recently his paper on Fat Block-Matching Algorithms for Motion estimation also selected in IEEE Conference will be organized in Rajkot in May 2012.