Comparison of Video Compression Standards

1 author:

Dr.Sabeenian R.S
Sona College of Technology
76 PUBLICATIONS **199** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Machine and deep learning in image processing platform View project

tumor segmentation View project

Comparison of Video Compression Standards

S. Ponlatha and R. S. Sabeenian

Abstract—In order to ensure compatibility among video codecs from different manufacturers and applications and to simplify the development of new applications, intensive efforts have been undertaken in recent years to define digital video standards 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 standards, techniques and algorithms had been developed to reduce the data quantity and provide the acceptable quality as possible as can. Thus they often represent an optimal compromise between performance and complexity. This paper describes the main features of video compression standards, discusses the emerging standards and presents some of its main characteristics.

Index Terms–Video compression, MPEG-1, MPEG-4, H.264, redundancies.

I. INTRODUCTION

Digital video has become main stream and is being used in a wide range of applications including DVD, digital TV, HDTV, video telephony, and teleconferencing. These digital video applications are feasible because of the advances in computing and communication technologies as well as efficient video compression algorithms. The rapid deployment and adoption of these technologies was possible primarily because of standardization and the economies of scale brought about by competition and standardization. Most of the video compression standards are based on a set of principles that reduce the redundancy in digital video.

Digital video is essentially a sequence of pictures displayed overtime [1]. Each picture of a digital video sequence is a 2D projection of the 3D world. Digital video thus is captured as a series of digital pictures or sampled in space and time from an analog video signal. A frame of digital video or a picture can be seen as a 2D array of pixels. Each pixel value represents the color and intensity values of a specific spatial location at a specific time. The Red-Green-Blue (RGB) color space is typically used to capture and display digital pictures. Each pixel is thus represented by one R, G, and B components. The 2D array of pixels that constitutes a picture is actually three 2D arrays with one array for each of the RGB components. A resolution of 8 bits per component is usually sufficient for typical consumer applications.

A. The Need for Compression

Fortunately, digital video has significant redundancies and eliminating or reducing those redundancies results in compression. Video compression can be lossy or loss less. Loss less video compression reproduces identical video after de-compression. We primarily consider lossy compression that yields perceptually equivalent, but not identical video compared to the uncompressed source. Video compression is typically achieved by exploiting four types of redundancies: 1) perceptual, 2) temporal, 3) spatial, and 4) statistical redundancies.

B. Perceptual Redundancies

Perceptual redundancies refer to the details of a picture that a human eye cannot perceive. Anything that a human eye cannot perceive can be discarded without affecting the quality of a picture [2]. The human visual system affects how both spatial and temporal details in a video sequence are perceived.

C. Temporal Redundancies

The persistence property can be exploited to select a frame rate for video display just enough to ensure a perception of continuous motion in a video sequence. Since a video is essentially a sequence of pictures sampled at a discrete frame rate, two successive frames in a video sequence look largely similar. Fig. 1. shows two successive pictures in a video. The extent of similarity between two successive frames depends on how closely they are sampled (frame interval) and the motion of the objects in the scene [3], [4]. If the frame rate is 30 frames per second, two successive frames of a news anchor video are likely to be very similar. Exploiting the temporal redundancies accounts for majority of the compression gains in video encoding.

D. Spatial Redundancies

Spatial frequencies refer to the changes in levels in a picture. The sensitivity of the eye drops as spatial frequencies increase; i.e., as the spatial frequencies increase, the ability of the eye to discriminate between the changing levels decreases. Any detail that cannot be resolved is averaged [5]. This property of the eye is called spatial integration. This property of the eye can be exploited to remove or reduce higher frequencies without affecting the perceived quality. The human visual perception thus allows exploitation of spatial, temporal, and perceptual redundancies.

E. Statistical Redundancies

The transform coefficients, motion vectors, and other data have to be encoded using binary codes in the last stage of video compression. The simplest way to code these values is by using fixed length codes; e.g., 16 bit words. However, these values do not have a uniform distribution and using fixed length codes is wasteful. Average code length can be reduced by assigning shorter code words to values with higher probability. Variable length coding is used to exploit these statistical redundancies and increase compression efficiency further.

II. MOTION JPEG AND MPEG

A. Motion JPEG

A digital video sequence can be represented as a series of JPEG pictures. The advantages are the same as with single still JPEG pictures – flexibility both in terms of quality and compression ratio. The main disadvantage of Motion JPEG (a.k.a. MJPEG) is that since it uses only a series of still pictures it makes no use of video compression techniques [4]. The result is a slightly lower compression ratio for video sequences compared to "real" video compression techniques.

B. Motion JPEG 2000

As with JPEG and Motion JPEG, JPEG 2000 can also be used to represent a video sequence. The advantages are equal to JPEG 2000, i.e., a slightly better compression ratio compared to JPEG but at the price of complexity. The disadvantage reassembles that of Motion JPEG. Since it is a still picture compression technique it doesn't take any advantages of the video sequence compression. This results in a lower compression ration compared to real video compression techniques.

C. MPEG-1

The first public standard of the MPEG committee was the MPEG-1, ISO/IEC 11172, which first parts were released in 1993. MPEG-1 video compression is based upon the same technique that is used in JPEG[6]. In addition to that it also includes techniques for efficient coding of a video sequence.
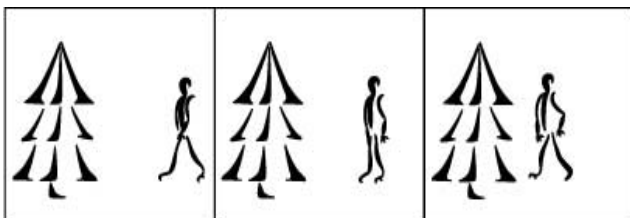

Fig. 1. A three-picture JPEG video sequence.

Consider the video sequence displayed in Fig. 2. The picture to the left is the first picture in the sequence followed by the picture in the middle and then the picture to the right. When displayed, the video sequence shows a man walking from right to left with a tree that stands still.In Motion JPEG/Motion JPEG 2000 each picture in the sequence is coded as a separate unique picture resulting in the same sequence as the original one. In MPEG video only the new parts of the video sequence is included together with information of the moving parts [7]. The video sequence of Fig. 4 will then appear as in Fig. 4. But this is only true during the transmission of the video sequence to limit the bandwidth consumption. When displayed it appears as the original video sequence again.
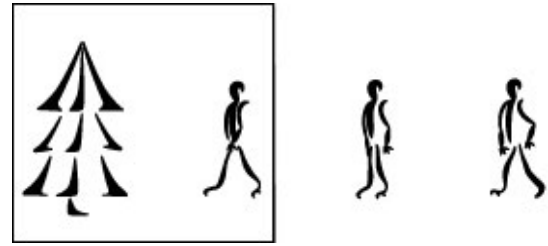

Fig. 2. A three-picture MPEG video sequence.

MPEG-1 is focused on bit-streams of about 1,5 Mbps and originally for storage of digital video on CDs. The focus is on compression ratio rather than picture quality. It can be considered as traditional VCR quality but digital instead. MPEG-1, the Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbps, is an International Standard ISO-11172, completed in October, 1992. International Standard ISO-11172, completed in October, 1992. MPEG-1 is intended primarily for stored interactive video applications (CD-ROM); with MPEG-1, one can store up to 72 minutes of VHS quality (640 x 480 s 30fps) video and audio on a single CD-ROM disk. MPEG-1 can deliver full-motion color video at 30 frames per second from CD-ROM. Because audio is usually associated with full motion video, the MPEG standard also addresses the compression of the audio information at 64, 96, 128, and 192 kbps[8] and identifies the synchronization issues between audio and video by means of time stamps. The first volume application for MPEG-1 decode chips (from C-Cube Microsystems) was a Karaoke entertainment system by JVC.

D. MPEG-2

MPEG-2 is the "Generic Coding of Moving Pictures and Associated Audio." The MPEG-2 standard is targeted at TV transmission and other applications capable of 4 Mbps and higher data rates. MPEG-2 features very high picture quality. MPEG-2 supports interlaced video formats, increased image quality, and other features aimed at HDTV. MPEG-2 is a compatible extension of MPEG-1, meaning that an MPEG-2 decoder can also decode MPEG-1 streams. MPEG-2 audio will supply up to five full bandwidth channels (left, right, center, and two surround channels), plus an additional low-frequency enhancement channel, or up to seven commentary channels. The MPEG-2 systems standard specifies how to combine multiple audio, video, and private-data streams into a single multiplexed stream and supports a wide range of broadcast, telecommunications, computing, and storage applications. MPEG-2, ISO/IEC 13818, also provides more advanced techniques to enhance the video quality at the same bit-rate. The expense is the need for far more complex equipment. Therefore these features are not suitable for use in real-time surveillance applications.As a note, DVD movies are compressed using the techniques of MPEG-2.

E. MPEG-4

The most important new features of MPEG-4, ISO/IEC 14496, concerning video compression are the support of even lower bandwidth consuming applications, e.g. mobile units, and on the other hand applications with extremely high quality and almost unlimited bandwidth. The making of studio movies is one such an example [9]. Most of the differences between MPEG-2 and MPEG-4 are features not related to video coding and therefore not related to surveillance applications MPEG involves fully encoding only key frames through the JPEG algorithm (described above) and estimating the motion changes between these key frames. Since minimal information is sent between every four or five frames, a significant reduction in bits required to describe the image results. Consequently, compression ratios above 100:1 [10] are common. The scheme is asymmetric; the MPEG encoder is very complex and places a very heavy computational load for motion estimation. Decoding is much simpler and can be done by today's desktop CPUs or with low cost decoder chips. MPEG-3 was merged into MPEG-2 and no longer exists.

The basic scheme is to predict motion from frame to frame in the temporal direction, and then to use DCT's (discrete cosine transforms) to organize the redundancy in the spatial directions. The DCT's are done on 8x8 blocks, and the motion prediction is done in the luminance (Y) channel on 16x16 blocks. For a 16x16 block in the current frame being compressed, the encoder looks for a close match to that block in a previous or future frame (there are backward prediction modes where later frames are sent first to allow interpolating between frames). The DCT coefficients (of either the actual data, or the difference between this block and the close match) are quantized. Many of the coefficients end up being zero. The quantization can change for every macroblock, which is 16x16 of Y and the corresponding 8x8's in both U and V. The results of all of this, which include the DCT coefficients, the motion vectors, and the quantization parameters are Huffman coded using fixed tables. The DCT coefficients have a special Huffman table that is two-dimensional that one code specifies a run-length of zeros and the non-zero value that ended the run. Also, the motion vectors and the DCT components are DPCM (subtracted from the last one) coded.

F. H.261

H.261 (last modified in 1993) is the video compression standard included under the H.320 umbrella (and others) for videoconferencing standards. H.261 is a motion compression algorithm developed specifically for videoconferencing, though it may be employed for any motion video compression task. H.261 allows for use with communication channels that are multiples of 64 kbps (P=1, 2, 3...30.), the same data structure as ISDN [11]. H.261 is sometimes called P$\times$64.

H.261 encoding is based on the discrete cosine transform (DCT) in Fig. 3. and allows for fully-encoding only certain frames (INTRA-frame) while encoding the differences between other frames (INTER-frame). The main elements of the H.261 source coder are prediction, block transformation (spatial to frequency domain translation),

quantization, and entropy coding. While the decoder requires prediction, motion compensation is an option. Another option inside the recommendation is loop filtering. The loop filer is applied to the prediction data to reduce large errors when using interframe coding. Loop filtering provides a noticeable improvement in video quality but demands extra processing power. The operation of the decoder allows for many H.261-compliant CODECs to provide very different levels of quality at different cost points. The H.261 standard does not specify a particular adaptive quantization method.
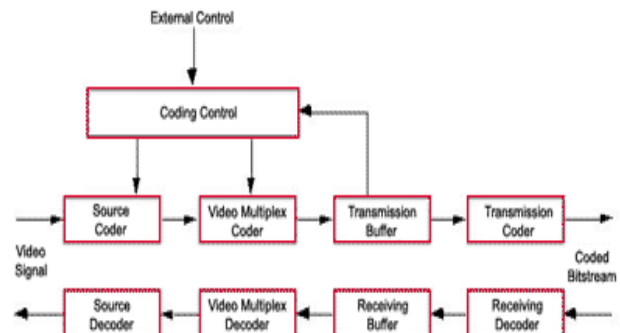

Fig. 3. H.261 block diagram from ITU recommendation.

The H.261 source code operates on non-interlaced pictures occurring 30 000/1001 (approximately 29.97) times per second. The tolerance on picture frequency is ± 50 ppm. Pictures are coded as luminance and two color difference components (Y, C_B, and C_R). These components and the codes representing their sampled values are as defined in CCIR Recommendation 601: black = 16, white = 235, zero color difference = 128, and peak color difference = 16 and 240. These values are nominal ones and the coding algorithm functions with input values of 1 through to 254.
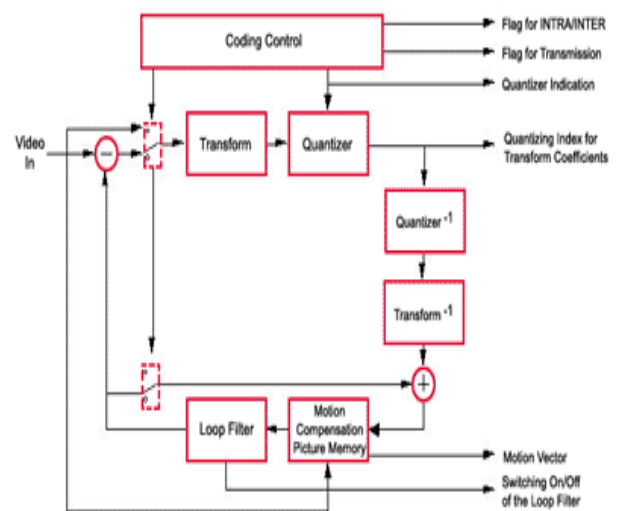

Fig. 4. H.261 source coder block diagram.

It is important to understand the hierarchical structure of video data used by H.261 in Fig. 4. At the top layer is the picture. Each picture is divided into groups of blocks (GOBs). Twelve GOBs make up a CIF image; three make up a QCIF picture in Fig. 5. A GOB relates to 176 pixels by 48 lines of Y and the spatially corresponding 88 pixels by 24 lines for each chrominance value.

Fig. 5. Arrangement of groups of blocks in an H.261 picture.

Each GOB is divided into 33 macroblocks. A macroblock relates to 16 pixels by 16 lines of Y and the spatially corresponding 8 pixels by 8 lines of each chrominance value. Macroblocks are the basic element used for many prediction and motion estimation techniques.

When an H.261 controller decides to perform an intraframe compression or an interframe compression, or when it segments data as transmitted or non-transmitted, these decisions are made on a block-by-block basis, not on a picture-by-picture basis.

G. H.263

H.263 is the video codec introduced with H.324, the ITU recommendation "Multimedia Terminal for Low Bitrate Visual Telephone Services Over the GSTN". H.324 is for videoconferencing over the analog phone network (POTS). While video is an option under H.324, any terminal supporting video must support both H.263 and H.261. H.263 is a structurally similar refinement (a five year update) to H.261 and is backward compatible with H.261. At bandwidths under 1000 kbps [12], H.263 picture quality is superior to that of H.261. Images are greatly improved by using a required 1/2 pixel new motion estimation rather than the optional integer estimation used in H.261. Half pixel techniques give better matches, and are noticeably superior with low resolution images (SQCIF).

The 4:3 pixel aspect ratio is the same for each of these picture formats.

TABLE I: H. 263 PICTURE FORMATS

Picture Format	pixels luminance	lines luminance	pixels chrominance	lines chrominance	H.261	H.263
sub-QCIF	128	96	64	48	optional	Required
QCIF	**176**	**144**	**88**	**72**	**required**	**Required**
CIF	352	288	176	144	optional	Optional
4CIF	704	576	352	288	NA	Optional
16CIF	1408	1152	704	576	NA	Optional

With H.263, as with H.261, each picture is divided into groups of blocks (GOBs). A group of blocks (GOB) comprises of k*16 lines, depending on the picture format (k = 1 for sub-QCIF, QCIF and CIF; k = 2 for 4CIF; k = 4 for 16CIF) and is given in Table I. The number of GOBs per picture is 6 for sub-QCIF, 9 for QCIF, and 18 for CIF, 4CIF and 16CIF.

H. H.264

H.264 is the result of a joint project between the ITUT's Video coding Experts group and the ISO/IEC Moving Picture Experts Group (MPEG). ITU-T is the sector that coordinates Telecommunication standard on behalf of the International Telecommunication Union. ISO stands for International Organization for Standardization and IEC stands for International Electrotechnical Commission, which oversees standards for all electrical, electronic and related technologies [13]. H.264 is the name used by ITU-T, while ISO/IEC has named it MPEG-4 Part 10/AVC since it is presented as a new part in its MPEG-4 suite. The MPEG-4 suite includes, for example, MPEG-4 Part 2, which is a standard that has been used by IP-based video encoders and network cameras.

Designed to address several weaknesses in previous video compression standards, H.264 delivers on its goals of supporting:
1) Implementations that deliver an average bit rate reduction of 50%, given a fixed video quality compared with any other video standard.
2) Error robustness so that transmission errors over various networks are tolerated.
3) Low latency capabilities and better quality for higher latency.
4) Straightforward syntax specification that simplifies implementations.
5) Exact match decoding, which defines exactly how numerical calculations are to be made by an encoder and a decoder to avoid errors from accumulating.

III. MPEG COMPARISON

All MPEG standards are back compatible. This means that an MPEG-1 video sequence also can be packetized as MPEG-2 or MPEG-4 video. Similarly, MPEG-2 can be packetized as an MPEG-4 video sequence. The difference between a true MPEG-4 video and an MPEG-4-packetized MPEG-1 video sequence is that the lower standard does not make use of the enhanced or new features of the higher standard[14].

The comparison of the MPEGs in Table II, contains the MPEG-1 with its most often used limitation (Constrained Parameters Bitstream, CPB), MPEG-2 with its Main Profile at Main Level (MP@ML), and MPEG-4 Main Profile at L3 Level.

TABLE II: MPEG COMPARISON

Standards	MPEG 1	MPEG 2	MPEG 4
Max bit rate (Mbps)	1,86	15	15
Picture width (pixels)	352	720	720
Picture height (pixels)	288	576	576
Picture rate (fps)	30	30	30

When comparing the performance of MPEG standards such as MPEG-4 and H.264, it is important to note that results may vary between encoders that use the same standard. This is because the designer of an encoder can

choose to implement different sets of tools defined by a standard. As long as the output of an encoder conforms to a standard's format and decoder, it is possible to make different implementations. An MPEG standard, therefore, cannot guarantee a given bit rate or quality, and comparisons cannot be properly made without first defining how the standards are implemented in an encoder. A decoder, unlike an encoder, must implement all the required parts of a standard in order to decode a compliant bit stream. A standard specifies exactly how a decompression algorithm should restore every bit of a compressed video.
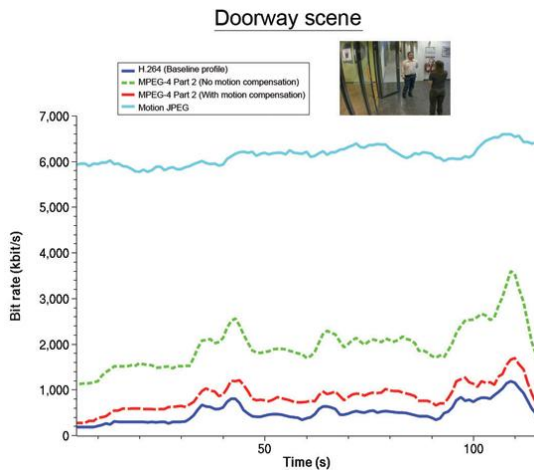

Fig. 6. H.264 encoder was at least three times more efficient than an MPEG-4 encoder and at least six times more efficient than with Motion JPEG.

The graph on the above page provides a bit rate comparison, given the same level of image quality[14], among the following video standards: Motion JPEG, MPEG-4 Part 2 (no motion compensation), MPEG-4 Part 2 (with motion compensation) and H.264 (baseline profile) in Fig. 6.

TABLE III: MPEG Comparison with Pros & Cons

Standards/ Formats	Compression Factor	Pros	Cons
M-JPEG	1 is to 20	1.Low CPU utilisation 2.Clearer images at lower frame rates, compared to MPEG-4 3.Not sensitive to motion complexity, i.e. highly random motion	1.Nowhere near as efficient as MPEG-4 and H.264 2.Quality deteriorates for frames with complex textures, lines, and curves
MPEG-4 Part 2	1 is to 50	1.Good for video streaming and television broadcasting 2.Compatibility with a variety of digital and mobile devices	1.Sensitive to motion complexity (compression not as efficient) 2.High CPU utilisation
H.264	1 is to 100	1.Most efficient 2.Extremely efficient for low-motion video content	1.Highest CPU utilisation 2.Sensitive to motion complexity (compression not as efficient)

Since the H.261/H.263 recommendations are neither international standards nor offers any compression enhancements compared to MPEG, they are not of any real interest. There are two approaches to achieving video compression, viz. intra-frame and inter-frame. Intra-frame compression uses the current video frame for compression: essentially image compression. Inter-frame compression uses one or more preceding and/or succeeding frames in a sequence, to compress the contents of the current frame. An example of intra-frame compression is the Motion JPEG (M-JPEG) standard [15]. The MPEG-1 (CD, VCD), MPEG-2 (DVD), MPEG-4, and H.264 standards are examples of inter-frame compression.

The popular video compression standards in the IP video surveillance market are M-JPEG, MPEG-4, and H.264are in Table III.

IV. CONCLUSION

There is a constant improvement in video compression factors, thanks to new techniques and technology, and some new formats in the horizon are H.265 and VP8:

1) H.265 is still in the process of being formulated, and aims to achieve a 25% improvement in the compression factor while lowering computational overhead by 50%: for the same perceived video quality.

2) VP8 is a codec from On2 Technologies (which recently agreed to be acquired by Google), who claims that the codec brings bandwidth savings and uses less data than H.264: to the extent of 40%. There is currently a fight over the standard to be chosen for Web video (fuelled by the upcoming HTML5 standard), and VP8 is slugging it out with H.264.

REFERENCES

[1] ITU-T and ISO/IEC JTC 1, "Generic coding of moving pictures and associated audio information – Part 2: Video," ISO/IEC 13818-2 (MPEG-2), 1994.
[2] ISO/IEC JTC1/SC29, "Coding of Audio-Visual Objects," ISO/IEC 14496-2, *International Standard: 1999/Amd1*, 2000.
[3] A. Puri, X. Chen, and A. Luthra, "Video Coding Using the H.264/MPEG-4 AVC Compression Standard," *Signal Processing: Image Communication,* September 2004 issue.
[4] ISO/IEC JTC 1, "Advanced video coding," *ISO/IEC FDIS 14496-10, International Standard*, 2003.
[5] M. Horowitz, A. Joch, F. Kossentini, and A. Hallapuro, "H.264/AVC Baseline Profile Decoder Complexity Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 704-716, 2003.
[6] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video CodingStandard," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 13, no. 7, pp. 560-576, 2003.
[7] I. E. G. Richardson, H.264 and MPEG-4 Video Compression, UK Wiley, 2003.
[8] M. Flierl, T. Wiegand, and B. Girod, "Multihypothesis Pictures for H.26L," *IEEE ICIP 2001*, Greece,2001
[9] M. Flierl and B. Girod, "Generalized B Pictures and the Draft H.264/AVC Video-Compression Standard," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 13, no. 7, pp. 587-597, 2003.
[10] D. Marpe, H. Schwarz, and T. Wiegand, "Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 13, no. 7, pp. 620-636, 2003.
[11] Z. Zhou, M. T. Sun, and S. Hsu, "Fast Variable Block-size Motion Estimation Algorithms Based on Merge and Split Procedure for H.264/MPEG-4 AVC," *IEEE ISCAS 2004 Conference*.

[12] Z. Zhou and M. T. Sun, "Fast Macroblock Inter Mode Decision and Motion Estimation for H.264/MPEG-4 AVC," *IEEE ICIP 2004 Conference.*

[13] P. Chen and J. W. Woods, "Improved MC-EZBC with quarter-pixel motion vectors," *ISO/IEC JTC1/SC29/WG11, MPEG2002/m8366,* 2002.

[14] J. Xu, R. Xiong, B. Feng, G. Sullivan, M. Lee, F. Wu, and S. Li, "3D Sub-band Video Coding using Barbell lifting," *ISO/IEC JTC/WG11 M10569.*

[15] M. Karczewicz and R. Kurceren, "The SP- and SI-Frames Design for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 637-644, 2003.

R. S. Sabeenian is currently working as a professor in ECE Department in Sona College of Technology, Salem, Tamil Nadu, and India. He received his Bachelors in Engineering from Madras University and his Masters in Engineering in Communication Systems from Madurai Kamaraj University. He received his Ph.D. Degree from Anna University, Chennai in the year 2009 in the area of Digital Image processing. He is currently heading the research group named Sona SIPRO (SONA Signal and Image PROcessing Research Centre) centre located at the Advanced Research Centre in Sona College of Technology, Salem. He has published more than 65 research papers in various International, National Journals and Conferences. He has also published around seven books. He is a reviewer for the journals of IET, UK and ACTA Press Singapore. He received the "Best Faculty Award" among Tamil Nadu, Karnataka and Kerala states for the year 2009 given by the Nehru Group of Institutions, Coimbatore and the "Best Innovative Project Award" from the Indian National Academy of Engineering, New Delhi for the year 2009 and "ISTE Rajarambapu Patil National Award" for Promising Engineering Teacher for Creative Work done in Technical Education for the year 2010 from ISTE. He has also received a Project Grant from the All India Council for Technical Education and Tamil Nadu State Council for Science and Technology, for carrying out research. He received two "Best Research Paper Awards" from Springer International Conference and IEEE International Conference in the year 2010.He was also awarded the IETE Biman Behari Sen Memorial National Award for outstanding contributions in the emerging areas of Electronics and Telecommunication with emphasis on R&D for the year 2011.The Award was given by Institution of Electronics and Telecommunication Engineers (IETE), New Delhi. He is the Editor of 6 International Research Journals Research Journal of Information Technology, Asian Journal of Scientific Research, Journal of Artificial Intelligence, Singapore Journal of Scientific Research, International Journal of Manufacturing Systems and ICTACT Journal of Image Processing. He is also associated with the Image Processing Payload of the PESIT.

S. Ponlatha is currently working as an associate professor in ECE Department in AVS Engineering College ,Salem. She received her Bachelors in Engineering from Madras University and her Masters in Engineering in Communication Systems from Anna University and She is Pursuing Ph.D. Degree under Anna University, Chennai in the area of Digital Image processing.She published papers in International Journals & Conferences.She is a Member of ISTE, IEEE, IETE.