INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2021, VOL. 67, NO. 2, PP. 309-314
Manuscript received June 12, 2020; revised April, 2021 DOI: 10.24425/ijet.2021.135981

# Optimized Inter Prediction for H.264 Video Codec

T. Bernatin, G.Sundari, Sahaya Anselin Nisha, and M.S. Godwin Premi

***Abstract*—High definition video transmission is one of the prime demands of modern day communication. Changing needs demand diverse features to be offered by the video codec standards, H.264 fits to these requirements for video compression. In this work, an attempt has been made to optimize the inter prediction along with improved intra prediction to ensure the minimal bit rates thereby reduction in the channel bandwidth, which is required in most of the wireless applications. In intraprediction, only DC prediction mode is chosen out of 9 modes with 4*4 luma blocks that reduces the coding complexity towards optimal logic utilization in order to support typical FPGA board for hardware implementation. Most significantly, Inter prediction is carried out utilizing the M9K blocks efficiently with proper timing synchronization to reduce the latency in the encoding operation. Experimental set up comprising of two Altera DE2-115 boards connected through Ethernet cable demonstrated the video transmission. These optimized intra prediction and inter prediction stages resulted in significant improvement in the video compression possessing good subjective quality and increased video compression**

***Keywords*—H.264, Intra prediction, Inter prediction, M9K, timing synchronization , QP**

## I. Introduction

Technology always provides services, challenges and solutions. This is very much true with respect to video communications, which has made the information, knowledge and video data transfer simple and beyond the imagination of man. Besides these advantages , it has also provided newer challenges in making optimum use of this technology. Scientific community and engineers have been taking these challenges and providing solutions in ensuring the technological fruits are completely received by the common man. World is experiencing one such technological revolution with the emergence of Internet of things. Today internet and video driven services came to reality due to the evolution of distinct video coding standards and their wide range of features. Streaming video has become the rudimentary feature of modern day communication, especially in broadcasting. This feature has been explored to the maximum with the wide usage by distinct groups in the society.

The evolution of coding standardhas been active from last three decades and it has taken a leap with the introduction of H.264 AVC. The digital transmission systems should be good enough to cater the increased demand for quality and speed in video data transmission. This situation has pushed for the search of better methods to be followed , which has led to improvise the algorithms used in H.264/AVC standard. The goal of this standard is enhanced compression performance and network friendly nature. This standard also brought notable improvement in the rate distortion efficiency.

H.264 offers high quality video in 720p or 1080p resolutions, Codec scalability, Interoperability and is familiar to industry as it is possessing adherence to industry standards. The playback efficiency , highest functionality offered by H.264 based ecosystem, content management are some of the additional features.Because of the aforementioned reasons, H.264 Codec has become the dominant video streaming standard. Diverse transport protocols are required to ensure the reliable video reception; H.264 offers a range of transport protocols to ensure the reliability. Few are RTP/RTSP, Transport Stream, HTTP, TCP/RTSP etc. This offers,choices for the architects and improves the adaptability to various systems. H.264 also possess adaptive deblocking filter that reduces the artifacts. It allows storing of several video frames in the memory. It has a prediction mechanism and uses integer transform instead of DCT used by other standards.

To increase the interoperability besides containing the complexity, H.264 standard offers specification defined profiles and levels. This made the standard to be used widely and improved the deployment capability of the standard. A profile is a subset of entire bit stream syntax. All decoders compliant to a certain profile, support all the tools in the corresponding profile. A level is a specific set of constraints imposed on values of the syntax elements in the bit stream. This addresses the challenge of using decoder that is able to accommodate all the uses of a syntax in a profile. Baseline Profile, Main Profile and Extended Profile are offered by this standard. Each profile is defined for a specific purpose and in our work the baseline profile is used for simplicity and less delay for the sake of simplicity and compatability with the hardware implementation.

## II. Improved Intraprediction

H.264 standard supports intra prediction for different sizes i.e. 16x16 macro blocks as a whole or 4x4 sub blocks. There are seventeen prediction modes for a macro block, nine prediction modes for 4x4 luma sub macro blocks, four modes of 8x8 blocks are used for lumamacroblock and four modes of 8x8 blocks for chromamacroblocks.

The intra prediction is based on the correlation between the nearby pixels. For a real time image, the nearby pixels vary slightly. Here the basic process is prediction of

The inter prediction of H.264 Video codec mentioned in this paper is a part of the research project going on in Sathyabama University with the grants funded by Combact Vehicle Research Development and Establishement, India during the year 2016. We thank Sathyabama University for providing us with various resources for carrying out this work. Being a co-investigator in this work, I am grateful to my dean Dr.G.Sundari, the Principal Investigator of this research project for her unconditional support.

The Authors are with Sathyabama Institute of Science and Technology, India, (e-mail: bernatin12@gmail.com, sundariece16@gmail.com, anselinnisha.ece@sathyabama.ac.in, godwin.etc@sathyabama.ac.in).

current macroblock values from the left and top pixels of the previous macroblock(MB). The difference between the original and predicted macroblock is found.

The labeling of prediction samplesare shown in the Figure 1.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Y | M | N | O | P | Q | R | S | T |
| U | i | j | k | l | | | | |
| V | m | n | o | p | | | | |
| W | q | r | s | t | | | | |
| X | u | v | w | x | | | | |

Fig.1 Labeling of Prediction Samples

The prediction process has 9 methods or "modes" by which it can get the prediction values from the neighbour pixels. The mode, which gives minimum residual value for the given macro block is fed to transformation block. This prediction process is implemented using shift register, particularly for division that reduces the overall processing time.

Figure 2 shows the DC prediction (mode 2). Here $Pred_2$ (i,j) are the prediction arrays and T is the Temporary Register. The reference pixels are M, N, O, P, U, V, W, X and all pixels in the current 4X4 block are predicted by an average of their neighboring pixels. If all samples M, N, O, P, U, V, W, X are available, all samples are predicted by (M+N+O+P+U+V+W+X+4)>>3.

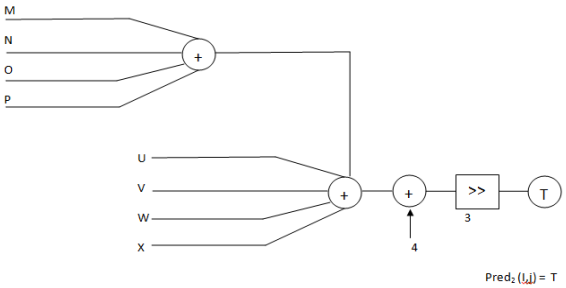


Fig.2 DC Prediction (mode2)

The economic viability of this work demands lesser complexity of video codec design. This in turn drives to reduce logical elements while implementing H.264 video codec. This can be obtained by the selection of DC prediction mode which leads to reduced complexity besides obtaining good PSNR value. Therefore DC prediction shall be used for real time application which demand higher compression ratio.

The residual data of DC prediction is fed to Integer Cosine transform (ICT) and quantization process where the lossy compression takes place and is expressed in equation 1 (Iain Richardson 2009). Figure 3 shows the process of Integer cosine transform and quantization process.

Here x(i,j) is the input matrix. Cf(i,j) is forward cosine transform matrix. The matrix multiplication is done by shift registers and adders. The temporary output matrix (T) and the transform of cosine transform matrix (CfT(i,j)) are matrix multiplied and a new temporary matrix T is created.

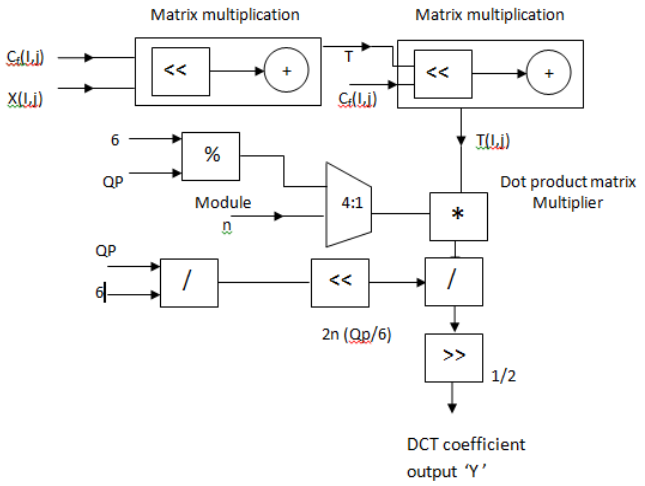


Fig.3 ICT and quantization

The equation for ICT and quantization process

$$Y = \text{round}\left([C_f].[Y].\left[C_f^T\right].m(QP\%6, n)/2^{\text{floor}(\frac{QP}{6})}\right).1/2^{15} \quad (1)$$

(3.1)

Y is the current macro block of residual values. Here % is the "modulo" operator. This finds the mod value of QP with 6.The quantization parameters (QP) can be selected with respect to the compression. This ICT process is full of matrix multiplication and power values and division process. But every power calculation and division process is in the power of 2. Hence, shift registers are viable solution for matrix multiplication.

After the prediction stage, Integer Cosine Transform (ICT) is applied that gives residual DC co-efficients. In the Quantization stage, these DC co-efficients are processed. Followed by quantization, CAVLC (entropy coding) is performed and thereby generating the bit stream. For thenext frame, data is subjected to inverse transform and de-quantization, which converts data into spatial domain for the reconstruction block to provide feedback, which is stored in SDRAM and is used in the prediction stages.

## III. INTER PREDICTION FOR HARDWARE IMPLEMENTATION

The Inter Prediction primarily compares the reference frame with the secondary frame or new frame and produces the residue, which is subjected to transform and quantization process. Effective memory management is the requirement for carrying out the aforementioned process in an efficient manner. This is being achieved by the modified intra prediction, which produces the reference frame, which is good enough to be used in the comparison process

Inter prediction process uses the previous frame of the video to compress the current frame. The first frame is not processed in inter prediction mode, which does not have a previous frame. That will be processed in the intra prediction mode and produces "I" frame. The "I" frame will be stored in the SDRAM memory module and will be the reference to the second frame.

Figure 4 explains the Flow chart for inter prediction . A 40 × 40 pixel array from the I frame or from the reference frame will load into the buffer. This will serve as the area for the particular macro block. The 40 × 40 pixel array (Search array) will be updated in the manner that, its centre position (20,20)

will be holding the macro block of same location in the reference frame.

Modified Diamond search starts with large diamond with 9 search points in first iteration with (20,20) as the centre search point. The iteration finds the residual with 9 search points, and Sum of Absolute Difference (SAD) value for these 9 search points. The search point with the minimum SAD will be fixed as the centre search point for the second iteration. Then a smaller diamond with new centre found in the first iteration will be taken with 5 new search points and 4 previous search points. The process repeats as the first iteration and finds the search point with minimum disparity.

The third and final iteration starts with the minimum SAD search point as the centre and 4 new search points will be taken. The resultant search point with minimum SAD value from the third iteration was chosen for the final residual value and the motion vector was calculated with respect to the initial location (20, 20).

An initial search area is required before starting the inter prediction process. Initial search area is filled by the macro blocks from previous frame. After that, the intra prediction module is ready to take macro block from the current frame.

A macro block from the current frame is taken with its position. A centre point is set (initial centre point for the search area is always 20x20) and search points are prepared with reference to the diamond search algorithm. Iteration 1 finds the prediction matrix from the current macro block and the search points. This process creates a prediction matrix and residual for the search points with the iteration 1 and the SAD for the each search point. The minimum SAD will be calculated from the search point and the position of the macro block with minimum SAD and the residual values will be given for next processes.
After the first iteration, the second iteration starts with the center points given by the search point with the minimum SAD value. The new search points will be selected from the search area based on the new center point.

The center point for the third iteration will be given by the minimum SAD search point position found from the second iteration. After third iteration the residual of the minimum SAD search point will be given out with the motion vector. Motion vector is calculated from the initial center point (in this case 20,20) and the final position of the minimum SAD value.

This process starts with the update of the search area for the next macro block. This process updates the search area matrix of 40x40 with the new macro block position. After the update process, it can again start the process of inter prediction with new current macro block and new search area. The same process repeats with three iterations and the final residual and the motion vector values are transferred out. And the same process repeats for the next macro blocks available in the current frame.

In modified diamond search, the number of iteration are made to optimize to increase the speed of processing (with minimum number of iteration).It is observed that with three iterations, we are able to get the best match.Here totally 18 search points are used in the modified diamond search algorithm. The observation is the number of search points were reduced from 24 to 18. This inturn reduces the complexity and end-to-end latency. After the process for a particular macro block is done, the search array will be updated for the next macro block from the SDRAM. It can search for the macro blocks near to the previous one. Hence, the number of times the SDRAM accessed has been reduced.

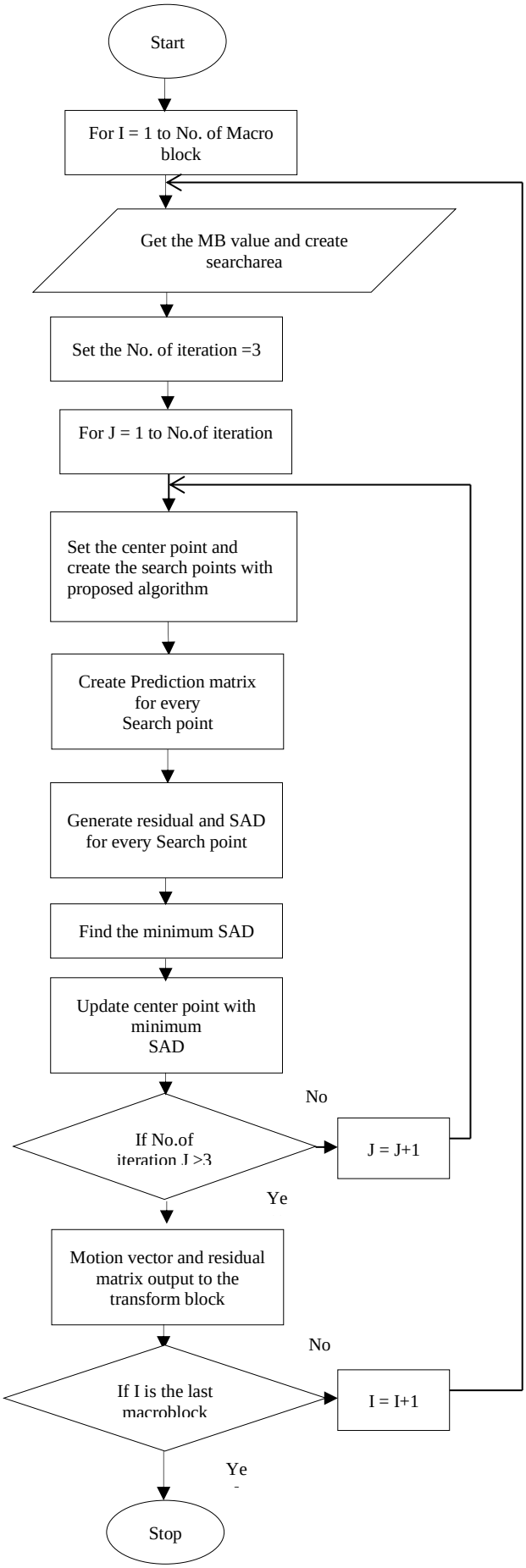


Fig.4 Flow chart for inter prediction

The above process resulted in the optimized inter prediction, which gives better compression ratio, improved latency in the board to board transmission with good quality output video.

## IV. Experimental Setup

The experimental setup requires the following rudimentary procedure for the video transmission. The video acquisition is facilitated by Pylon Camera Software Suite. This operates with Basler area scan cameras. Basler is a popular company manufacturing digital cameras for video surveillance and industrial applications.

The choice of camera interface is very important while setting up vision system. For 100 MB/s data rate and 100 meter maximum cable length Gigabit Ethernet (GigE) cameras is the ideal choice to get the best flexibility.

Figure 5 represents the acquisition of frames using pylon software from Basler camera.

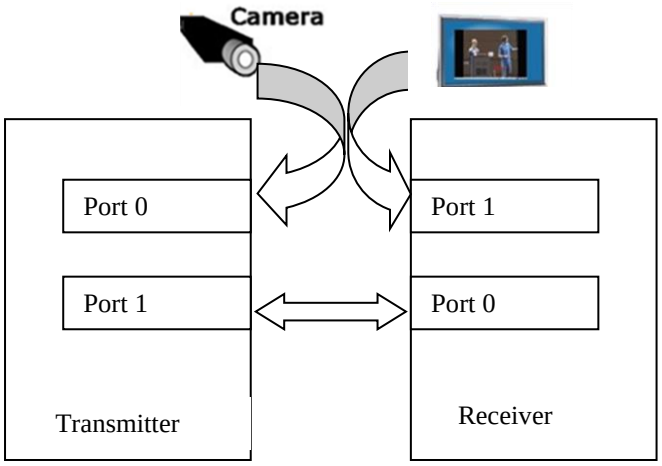


Fig.5 Experimental setup

In this work, one second video data captured at 24 frames/sec was considered. In order to establish the compatibility of the board with the camera, an interfacing dedicated GigE vision IP core was used. Figure 5 represents the transmitter to receiver board video data transfer. The camera is connected through Ethernet cable to the port 0 of the transmitter. On the other hand, the receiver board is connected to the display unit i.e. PC through port 1.

GigE vision standard allows the camera and software to integrate on the Gigabit Ethernet bus. This allows images to be transferred at higher speed over long cable lengths. This standard uses UDP protocol for communication between the device and the host. This is done using UDP packets which consist of IP header, UDP header, Ethernet header, data and Ethernet trailer (4 bits). The maximum size of the packet is of 1500 bytes and any data beyond the limit is broken into several packets.

GigE Vision control protocol (GVCP) is an application layer protocol that relies on User Defined Protocol. GVCP identifies stream of video and arranges compliant devices. This ensures integrity of the data. It permits the construction of a device and the instantiation of single or multiple control channels. The control channels are separated into main and subordinate channels. The main channel permits the request to read from and write to the device registers.

GigE Vision Stream Protocol (GVSP) is a request layer procedure that relies on the UDP transport layer protocol. It permits an application to accept image data, image information from a device. GVSP identifies the frames to be packetized and delivers devices for camera to send image data and other information to amenable receivers. GVSP offers devices to assure the reliability of packet communication and minimizes the flow control required due to the unreliability of UDP.

Figure 6 represents the image acquired in realtime. For high intensity of luminance, maximum frames can be acquired by the camera. In our working model, camera could acquire at the rate of 27 frames/sec for the set indoor illumination exposure. In general, image acquisition takes lot of memory of the hardware. To encounter this challenge, M9K block were used for the image acquisition , which reduced the logic utilization of the hardware. This M9K block is a synchronous, dual port memory block with registered inputs and optionally registered outputs. This block is available in the FPGA families, which is commonly used for the implementation of lookup tables and applications those require larger memory. These are easy to configure using FPGA software, and Cyclone IV FPGAs possess the embedded memory comprises of columns of M9K memory blocks Each M9K block is a 256X36 RAM that contains 9,216 programmable bits. The write and read signals which are independent in this block controls the power consumption , which is critical for modern day applications. As M9K block is used for storage, good amount of memory space has been saved . Amount of device utilization taken by using this M9K block has been mentioned in section V.

Fig. 6 *Board to Board Communication*

*Experimental Output:*

*Performance Analysis:* Due to large number of levels and profiles, FPGA implementation of H.264 is complicated. Every application has various levels and profiles due to variable performance, functionality and cost.

Table I shows the logic utilization summary of the Hardware implementation on various FPGAs. Therefore based on the outcome of the module,Altera Cyclone IV E is opted for this implementation.

TABLE I
LOGIC UTILIZATION SUMMARY

| Reference Implementations | Logic Elements (LE) | Memory (KB) | Maximum Frequency (MHz) | FPGA |
|---|---|---|---|---|
| This work | 82938 | 50 | 100 | Altera Cyclone IV E |
| Tung-Chien Chen et al. (2006) | 922768 | 34 | 108 | Xilinx Virtex-5 |
| TengWang et al. (2012) | 92109 | 92 | 200 | Xilinx Virtex-6 |
| Gwo-Long Li et al. (2013) | 257 618 | 24 | 135 | Xilinx Virtex-4 |
| Kuo et al. (2013) | 265312 | 8.4 | 114 | Xilinx Virtex-5 |

Tung-Chien Chen et al. (2006) developed FPGA architectures,integer motion estimation for low-bandwidth, reconfigurable intrapredictor, fractional motion estimation, de-blocking filter and dual-buffer block-pipelined entropy coder. Using these modules, H.264 encoder is developed with 922.8 K logic utilization and 108-MHz in 34.72-KB SRAM operation frequency as mentioned in Table I.

Teng Wang et al. (2012) suggested a Hardware implementation of H.264 encoder in FPGA Dini DN-DualV6-PCIe-4 platform .FPGA is implemented on FPGA Xilinx Virtex-6, at 200 MHz.

Gwo-Long Li et al. (2013) suggested FPGA hardware and functional verification of modified intra prediction in video codec. Implementation of design was done using CMOS technology in 90-nm of 135 MHz frequency. Different modules such as variable length coding, motion estimation, NAL coding, de-blocking filter were implemented.

Kuoet al. (2013) recommended the design of optimized intra encoder for effective scalable encoding at an intermediate frequency of 135 MHz and using a SVC encoder.

Figure 7 shows screen shot of the compressed data (in hexadecimal) and latency obtained at the receiver board. From this data, compressed ratio and latency values are calculated.

Latency is the time taken from the moment the camera captures an image to the moment the image get displayed on the screen. Latency is required to compress or decompress a digital video signal at the resolution 780 x 580 in 24 frames, which corresponds to less than 100ms. VGA pixel clock frequency is given as 36 MHz (resolution 800 × 600 as per VESA standard).

Latency = Total count × Time per second
Time per second = 1/ 36 MHz
= 27 × 10-9 s
Latency = Value× 1024 × 27 × 10-9
= 256H
(obtained from the FPGA board shown in Figure 7)× 27×10-6
Latency = (598) × 27 × 10-6
= 16146× 10-6
= 16.14ms

The board to board video data transfer has been achieved using developed H.264 standard, and the compression ratio has been calculated. The size of input frame is 452400 and the resolution is 780×580.

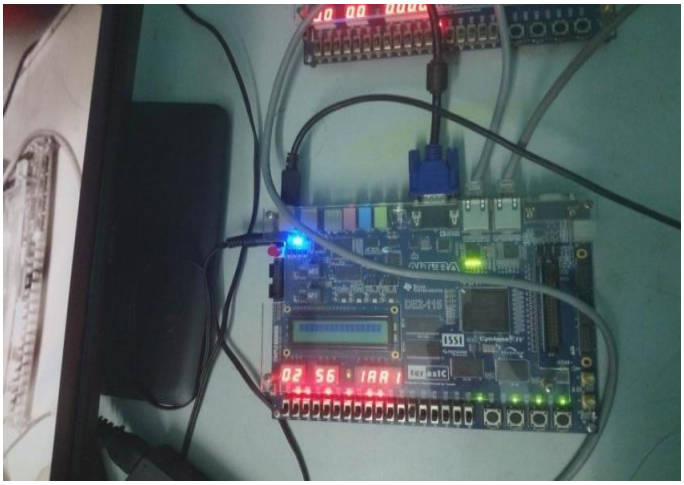

Fig. 7 Compressed data and Latency values at receiver board

Channel bandwidth required < 2 Mbps ( 2097152 bits) (compressed bits)
For a maximum of 24 frames,
Compression ratio (X) = Uncompressed / Compressed

$$X = \frac{452400 \times 24 \text{ frames} \times 8 \text{ bits}}{2097152 \text{ bits}}$$

X = 41 %

In real time,

X = 452400 bytes / ( 1AA1H) (obtained from the FPGA board shown in Figure 7)
= 452400 / 6817
CR(X) = 66 %

The compression ratio of 66% was obtained for the modified video codec implementation.

TABLE II
COMPARISON OF PERFORMANCE MATRICES WITH DIFFERENT QP VALUES FOR INTRA PREDICTION

| S.No | QP | PSNR(Y) in dB | PSNR (U) in dB | PSNR (V) in dB |
|---|---|---|---|---|
| 1 | 10 | 52.2 | 51.7 | 52.0 |
| 2 | 15 | 50.2 | 49.8 | 50.2 |
| 3 | 20 | 47.4 | 48.4 | 48.6 |
| 4 | 25 | 45.3 | 46.2 | 46.7 |
| 5 | 27 | 42:8 | 46.1 | 45.4 |
| 6 | 30 | 40.8 | 45.5 | 43.9 |
| 7 | 35 | 38.5 | 39.8 | 39.5 |
| 8 | 40 | 33.9 | 44.0 | 37.6 |
| 9 | 45 | 32.4 | 40.7 | 38.9 |
| 10 | 50 | 30.4 | 44.0 | 35.0 |

It has been observed that for the QP value of 27, the compression ratio is better as a result of optimized inter prediction.

## V. Conclusion and Futurescope

This work has been carried out using Verilog programming for the simulation and Altera DE2-115 Cyclone IV E boards for video compression. The simulated response obtained for optimized intra prediction and inter predication is in coherence with the theoretical computations. The image quality is good at the quantization parameter value of 27, where the compression ratio is obtained as 66%. The experimental setup successfully transferred video data at 24 frames/sec. 72% of reduction in device utilization is achieved with the usage of block in the FPGA.

The board to board video data transfer has been achieved using H.264 standard, and the Compression ratio, latency has been calculated. The data rate obtained is within 2 Mbps, with minimum latency. This work shall be used for any frame rate and also can be used for the different cameras with the existing hardware setup.

The future scope of the research work is that the implemented H.264 video codec can be applied to a variety of real time problems such as video conferencing, internet, telecommunication, multimedia streaming, and military application. This work will pave the way for the exploration of motion estimation blocks in the H.264 standard further for the other profiles also. The extension can also be possible with the higher data rate transmission and an integrated model shall be designed for all types of camera input.

This work shall be extended for obtaining better latency values and higher subjective video quality.As the range of video compression requirements and the scope of different video compression standards are ever changing, this research problem always throws open challenges to research community.

## References

[1] Bernatin.T ,Sundari.G, " Video compression based on Hybrid transform and quantization with Huffman coding for video codec", International conference on control, instrumentation, communication and computational technologies (ICCICCT), pp 476-480,IEEE.

[2] The Evolution of H.264 From Codec to System Architecture, White Paper , VBrick Systems, December, 2010

[3] Arun Kumar Pradhan, Lalit Kumar Kanoje and BiswaRanjan Swain, 2013 "FPGA based High Performance CAVLC Implementation for H.264 Video Coding" International Journal of Computer Applications (0975 – 8887) Volume69– No.10.

[4] Teng Wang, Chih-Kuang Chen, Qi-Hua Yang and Xin-An Wang (2012), "FPGA Implementation and Verification System of H.264/AVC Encoder for HDTV Applications", Advances in CSIE, Springer-Verlag Berlin Heidelberg, Vol. 2, AISC 169, pp. 345-352

[5] Gwo-Long Li et al. (2013), "135-MHz258-K gates VLSI design for all-intra H.264/AVC scalable video encoder", IEEE Trans. Very Large Scale Integr.(VLSI) Syst., Vol. 21, No. 4, pp. 636-647

[6] Kuo, H.-C., Wu, L.-C., Huang, H.-T., Hsu, S.-T. and Lin, Y.-L. (2013), "A low power high-performance H.264/AVC intra-frame encoder for 1080pHD video", IEEE Trans. Very Large Scale Integr. (VLSI) Syst., Vol. 19, No. 6, pp. 925-938