KAUNAS UNIVERSITY OF TECHNOLOGY

Department of Multimedia Engineering

Experiment on frames (I, B, P) in Video compression

Digital Image and Sound Processing (P170M104)
Technical Report

Students: Muruganantham Jaisankar
(Name, Surname)

Lecturer: *Lekt. Dr. Kęstutis Jankauskas*
(Title, Name, Surname)

(Grade)

Kaunas, 2021

Table of Contents

LIST OF ABBREVIATIONS

Terms	Expansion
JPEG	Joint Photographic Experts Group
MPEG	Moving Pictures Experts Group
DCT	Discrete Cosine transform
DVD	Digital Versatile Disc
AVC	Advanced Video Coding
NTSC	National Television System Committee
SIF	Source Input Format

Introduction

Video compression techniques alter video signals to minimize bandwidth and storage requirements while preserving video quality. For developers of processors, embedded systems, and tools that target video applications, understanding the function of video coding is critical.

Still video compression is Image compression

Video clips are collection of individual images, or “frames.” Therefore, video compression algorithms contain many procedures and techniques with image compression algorithms, such as JPEG. In fact, a best way to compress video is to eliminate the unique pixels between consecutive video frames. New video compression algorithms uses methods other than image compression approaches, such as motion estimation and motion compensation. Despite, video compression algorithms make use of the correlation between video frames.

“Lossy” image compression algorithm is used in JPEG and in video compression algorithms, that is original uncompressed image can’t be reconstructed from the compressed image data, only a little part of the original image's data is discarded. Human eye doesn’t see small frequency changes in image, so reconstructed image can be visible like original image.

JPEG image compression starts from converting RGB color space to YCbCr color space, chroma down sampling, YCbCr component is divided into 8X8 blocks to do DCT, DCT result can be coefficient matrix, and DCT coefficients are quantized, while quantization, the round value will not gathered in future reconstruction of a Image, after quantization, Zigzag scheme is used to form a sequence from quantized matrix, finally, the sequence is encoded using the Huffman code technique and the Run length encoder.

In this project, every single image considered as frame, every frame can have similarity to future frame or previous frame, that similarity taken as advantage to compress video. Frames are compared with future or previous frames, by comparing frames in different positions video compression can be done better, also a technique in video compression. While compressing a video, the quality of video also should be considered for better user experience. the project focusing on compression with minimum quality reduction of a video.

The Aim

To experiment with different positions of frames (I, P, B) to reduce video size (Ex. MPEG video compression) with minimum residual metric.

The Objectives

The main objectives of the project are:

1. To predict frames with use of block matching.
2. To experiment with different positions of I, P frames to get minimum prediction error.
3. To experiment with different positions of I, B frames to get minimum predication error.
4. To experiment with different positions of I, P, B frames to get minimum prediction error.

1 Video Coding Analysis

1.1 Intraframe coding standards

In 1988, the MPEG committee of the International Organization for Standardization (ISO) released the first video compression standard for CD-based video [1]. The format's name was MPEG 1. After few years, they released the MPEG 2 standard for SD video.

1.1.1 MPEG 1

The first MPEG standard was created for SIF resolution, which is roughly half of NTSC tv resolution in each dimension. The MPEG 1 standard was designed to support bitrates of 1 to 2 megabits per second [1]. In encoder, it uses macroblock size 16x16 with block-based motion estimation and motion vectors used in decoders [2]. MPEG1 is used in computer and telecommunication network and digital storage media [3]. Figure 1.1 shows the MPEG1 digital storge media application [3].

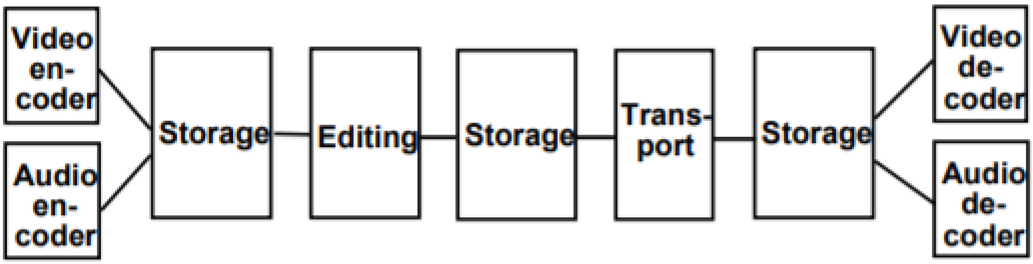

Figure 1.1: Digital storage media application

There are six layers for encoding and decoding in MPEG 1, sequence, group of pictures, pictures, slice, macroblock, block as DCT unit [3]. Figure 1.2 shows group of pictures with picture slice and macroblock [3].

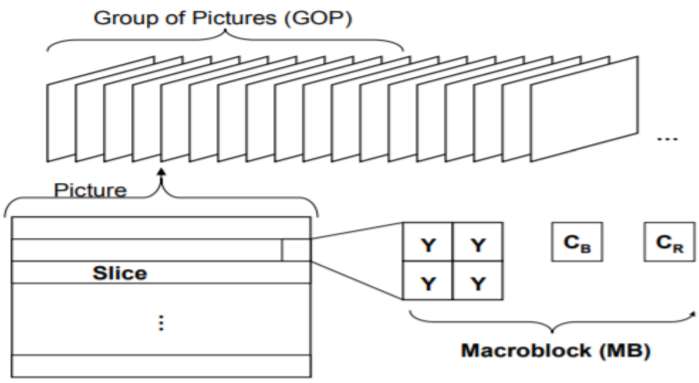

Figure1.2: MPEG 1 layers with Group of Pictures

MPEG 1 video is starting from sequence header to sequence end code. Video is starting from Sequence header and ends in sequence end code. Figure 1.3 shows the structure of MPEG 1video sequence [3].

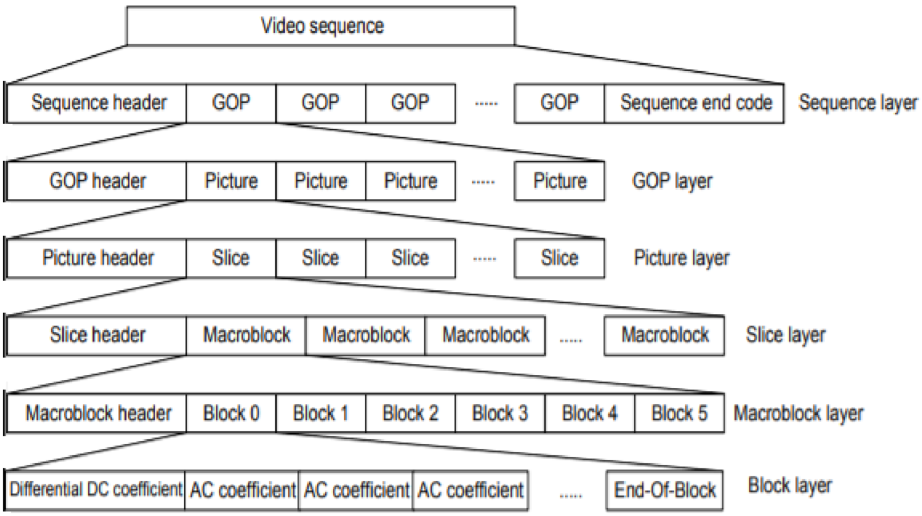

Figure 1.3: MPEG Video sequence

There are 4 types of frames in MPEG 1, I frame, P frame, B frame, D frame. B frame leads to get result as long end to end delay [1]. MPEG 2 extended with use of MPEG 1 [1]. In next part, MPEG 2 standard discussed.

1.1.2 MPEG 2 A Generic Standard

The MPEG2 coder, also known as H.262, was created in collaboration with the International Telecommunications Union's (ITU) VCEG committee and it incorporates MPEG 1 as a subset [1]. MPEG 2 also included new interlaced SD prediction modes, allowing prediction to be field-based, frame-based, or a combination of the two [3]. MPEG 2 is used as format of digital television signals that are broadcasted over the air, cable, satellite TV [4]. Figure 1.4 shows field based and frame DCT [3].

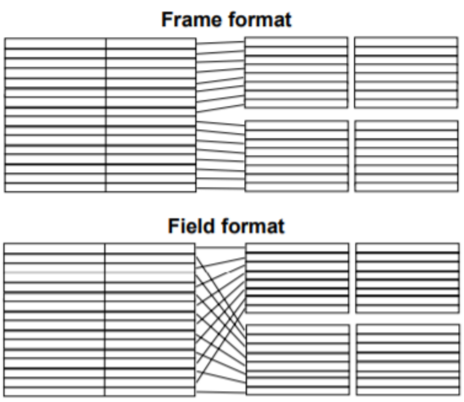

Figure 1.4: Field and Frame-based DCT

1.1.2.1 Block/Down sampling

When sending video over a low-bandwidth network, down sampling based video coding is an alternate method of achieving a low target bit rate. In the down sampling based coding strategy, a down sampling filter is employed to reduce the resolution of the video frame before to compression [4].

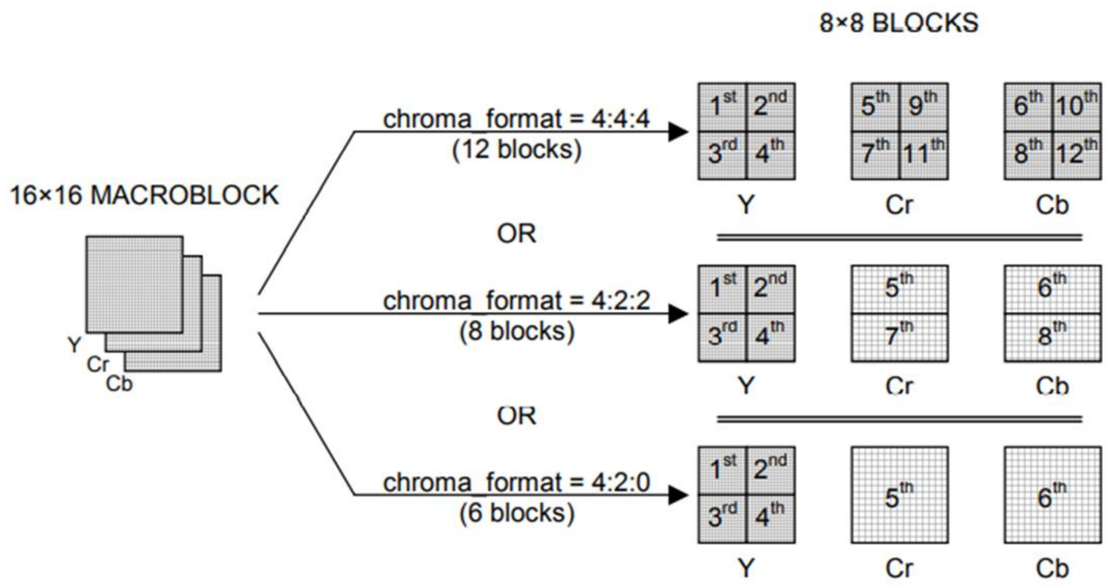

Figure 1.2 Block sampling [1]

1.1.2.2 Intra and non Intra frame encoding

Intra Encoding compresses data by using spatial redundancy and the current frame. Inter-encoding makes use of redundancy in space and time, as well as the current and prior frames [1]. Figure 1.5 shows Interlaced encoding [1]. Fiugure 1.6 shows interlaced scan and intera laced scan in MPEG 2.

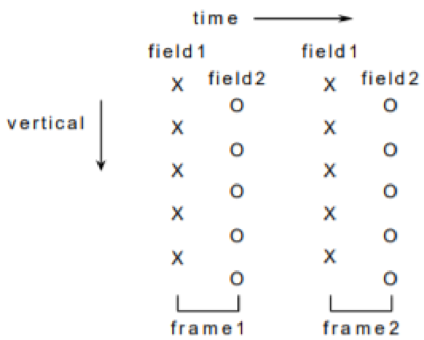

Figure 1.5: Interlaced encoding.

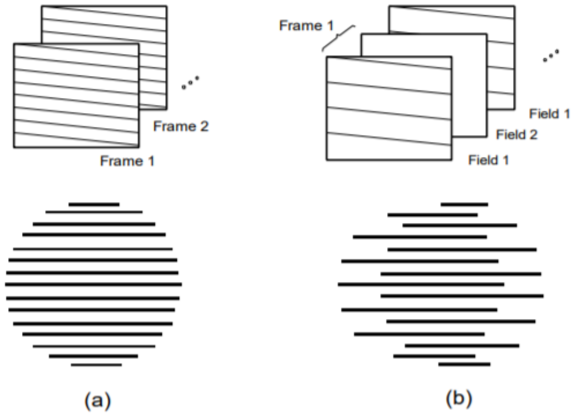

Figure 1.6: (a) Intralaced frame scan, (b) Interlaced scan

1.1.2.3 Block Matching

Block matching is useful to predict frames in a video, before moving to block matching, some important equations are needed to know. The mean absolute difference between target frame and current frame is calculated by using the difference between frames divided by the number of blocks in a target frame [5].

MAD = (difference between frames)/number of blocks in target frame (1.1)

The residual metric (RM) is an average of the sum of absolute residual values in a residual frame, and the mean residual metric is calculated with use of a residual metric and mean metric size [6].

RM = sum (absolute (difference between predicted frame ad target frame))/number of blocks (1.2)

If meanbuffer == mean metric size:

meantotal.apped(mean buffer/mean metric size)

meanbuffer.pop(0)

Mean buffer is a collection of residual metrics, and if a mean buffer is equivalent to mean metric size (mean value) then the mean total is calculated by dividing the mean buffer with mean metric size [6]. In block-matching, the anchor and target frame is processed into 8X8 blocks and each block of anchor and target frame are compared (three-step search method Figure 1.5) to get the best match of the block between the anchor and target frame [1,6].

The best match of anchor block is stored into a predicted frame, then the final result of the predicted frame is compared with the target frame, if there is a difference between predicted and target frame that can be added into the predicted frame in order to make a low mean absolute difference, the predicted frame is used instead of target frame to reduce the video size. Finally, the algorithm is used to check different frames (I, P, B) in a video compression.

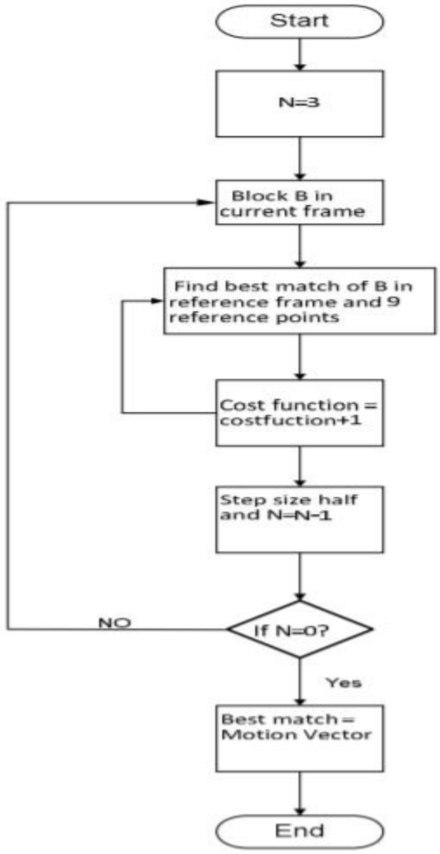

Figure 1.5 Three-step search block matching method

2 Experiment on frames

By using the block matching algorithm, the frames are predicted in different types as I-Frame, P-Frame, B-Frame. Only macroblocks that use intra-prediction that make up an I-frame, key-frame, or intra-frame. Every macroblock in an I-frame comes in different parts in video codecs, no temporal prediction is allowed in an I-Frame.

The predicted frame, or P-frame, allows macroblocks to predict spatially. P-frames utilise frames that have already been encoded to estimate motion. Every macroblock in a P-frame can be predicted either temporally or spatially, or skipped [7].

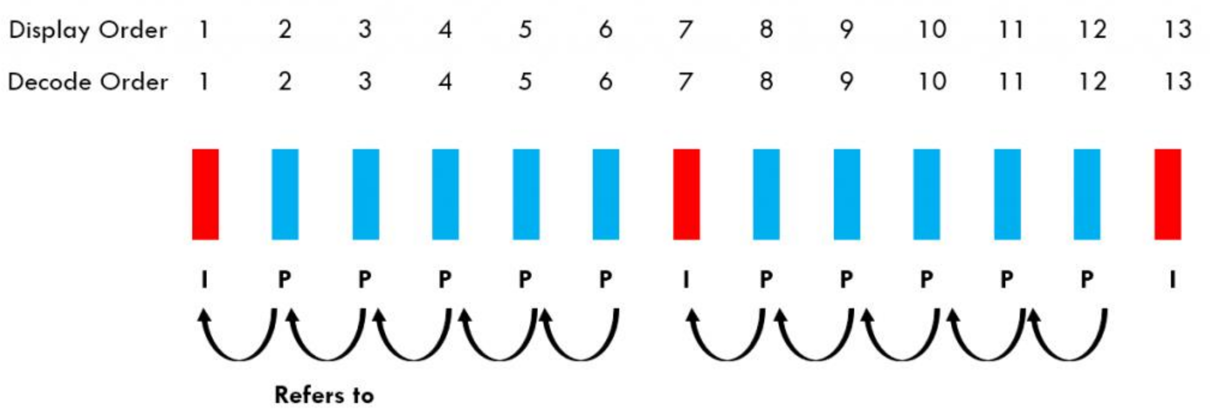

Figure:2.1-a For every 7 frames I-frame is re-established

P-frames refer to previously encoded I/P frames that are encoded and decoded in the same sequence that they are displayed to the user. Because P-frames only relate to already encoded images [1].

A B-frame refers to both preceding and following frames. For this reason, the B stands for Bi-Directional. If a video coding technique implements macroblock-based compression, backward prediction can be used to predict each macroblock in a B-frame (using frames from the past); forward prediction (using frames from the future); and totally skipped (with intra or inter prediction) [7].

B-frames can refer to two or more previous and subsequent frames, B-Frames can be highly effective at reducing frame sizes while maintaining video quality. The combination of both spatial and temporal redundancy makes B frames are most important in video compression (both in future and previous frames) [1,7].

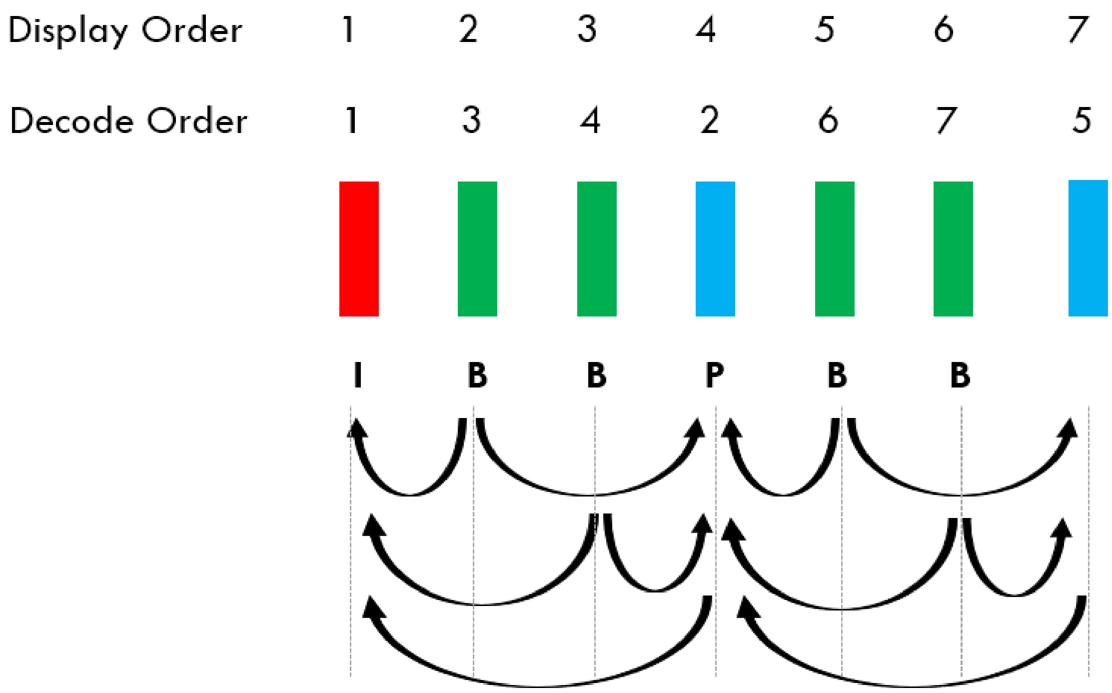

Figure:2.1-b B frame

2.1 Experiment on different position of the frames

Here P frame predicated from I-frame, I-frame established after some time in a video. The first experiment is establishing I frame in each 2 frames (figure 2.1) The second is to establish I-frame after every 5 frames (Figure 2.2)

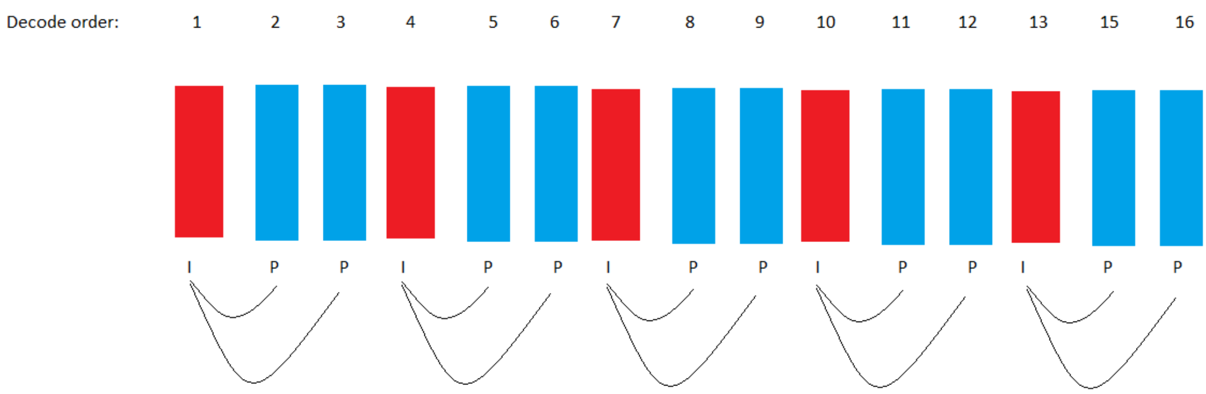

Figure 2.1 I-frame established after each 2 frame

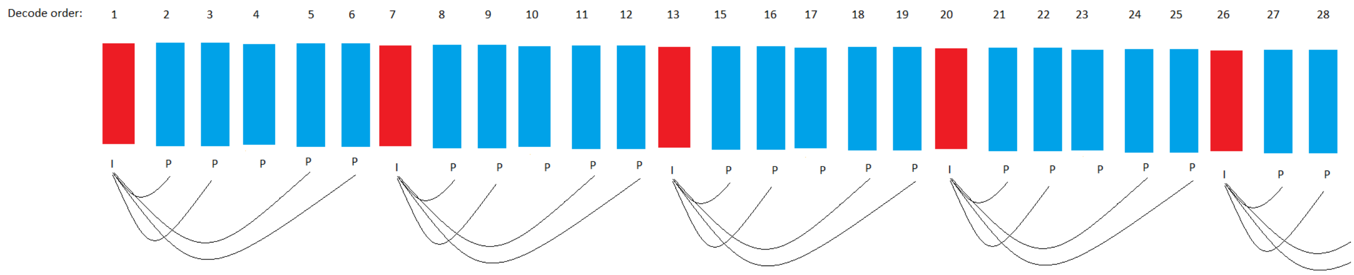

Figure 2.2 I-frame established after every 5 frames

Similarly, I-Frame is established after 10, 25, 50, 100. After the establishment of the I-Frame, each future frame is predicted with the use of I-frame. Finally, the Residual mean was calculated for each P frame with the target frame in a video.

The next experiment is with B -frame, here B-frame is referred from the anchor frame, and then I-frame is predicted to compress. I-frame is established after each 2 B-frame as shown in (Figure 2.3). Figure 2.3 shows I-frame is established after each 5 B-frame

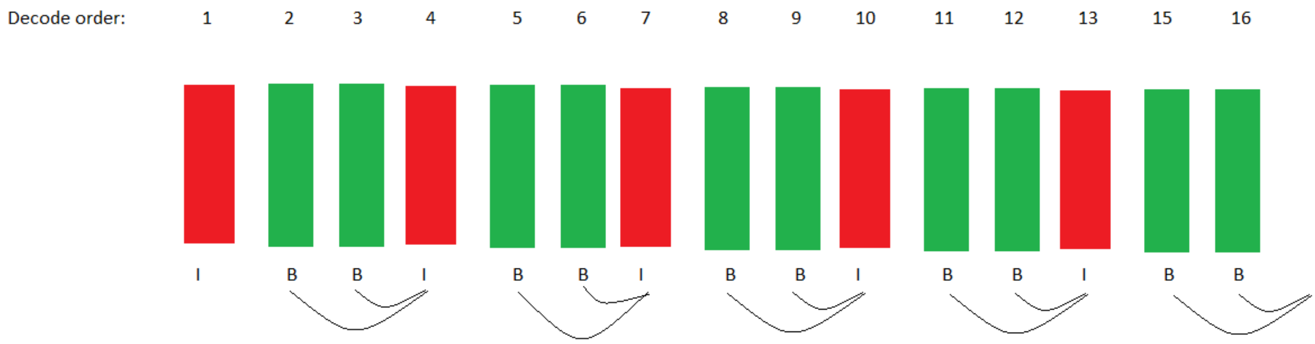

Figure 2.3 I-frame established after every 2 B-frame

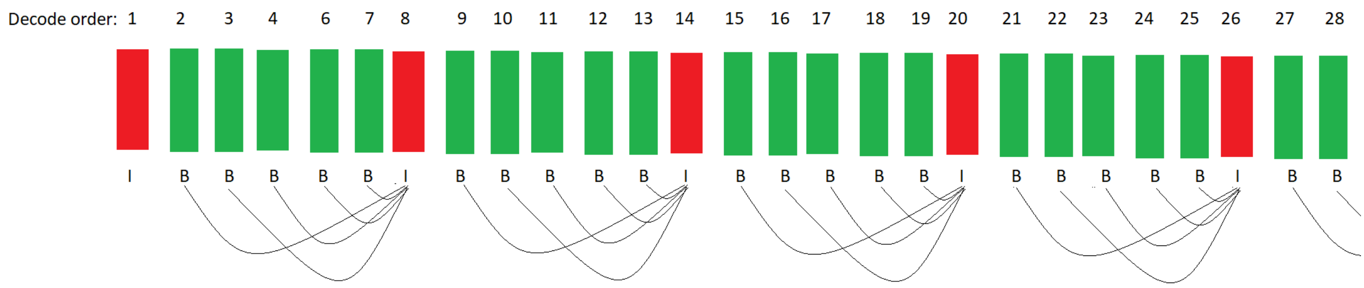

Figure 2.4 I-frame established after every 5 B-frame

From the previous experiment here P and B frames have been experimented together with I-frame (figure 2.5).

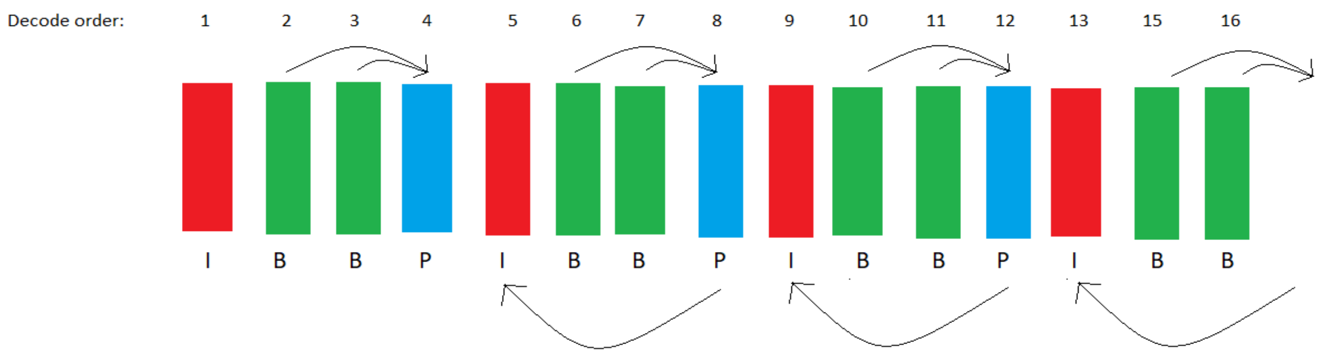

Figure 2.5 I-frame established after every 2 B-frame and 1 P-frame

In order to compress video, frames are predicted, and original frames are dropped in exchange of predicted frames. After prediction, residual means are calculated for each frame and the final results are stored to find the minimum residual mean.

Example

Following frames are sample frames from a MPEG video.

Figure 2.6: Sample frame 1

Figure 2.7: Sample frame 2

Figure 2.8: Sample frame 3

3 Results

Sample Result 1

Frame 1 and 2 (Figure 2.6 and Figure 2.7) have small number of difference, if frame 2 predicted from fram 1 then the predicted frame residual metric is 14.54(based on luma component figure 2.9). Residual metric when Frame 1 compare to Frame 2 is 66.05.

Figure 3.1: Predicted Frame 1 from Frame 2 (based on luma)

Sample Result 2

Frame 1 and 3 (Figure 2.6 and Figure 2.8) have more number of difference, if frame 2 predicted from fram 3 then the predicted frame residual metric is 75.64(based on luma component figure 2.8). Residual metric when Frame 1 compare to Frame 2 is 79.62

Figure 3.2: Predicted Frame 1 from Frame 3(based on luma)

The residual metroc is calculated and compared with every frame that is predicted with I-frame. Here figure 3.3 below shows that 2 predicated frames after establishing I frames and 5 skipped frames, finally the frames are predicated with 25 sample frame in 62 seconds. The X-axis represents the number of frames in a video, the Y-axis represents the residual metric for each frame that was predicted by using I-frame.

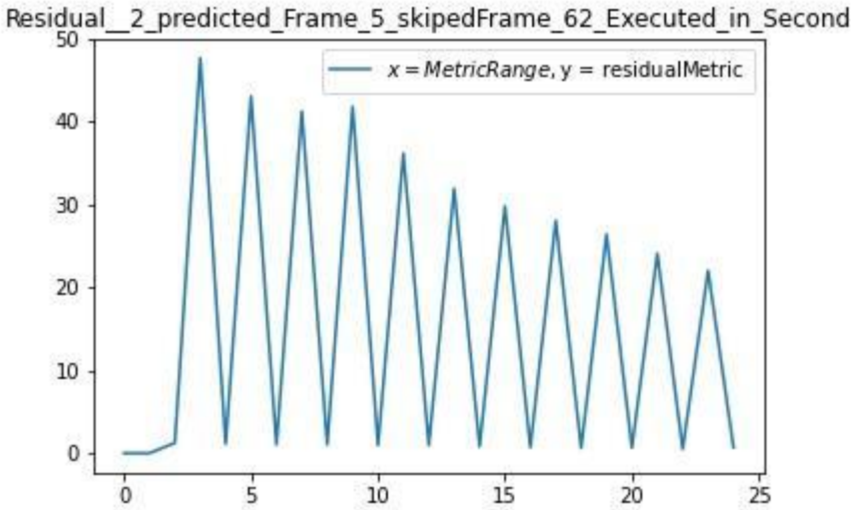

Figure 3.3 I-frame established after every 2 P-frame

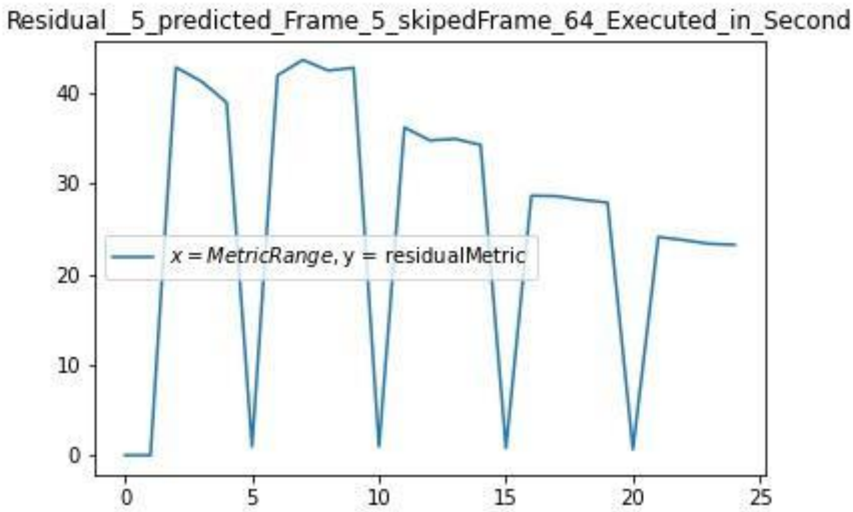

Figure 3.4 I-frame established after every 5 P-frame

The above figure 3.4 shows that 5 predicated frames after establishing I frames and 5 skipped frames, in the end, the frames are predicated with 25 frame samples in 64 seconds.

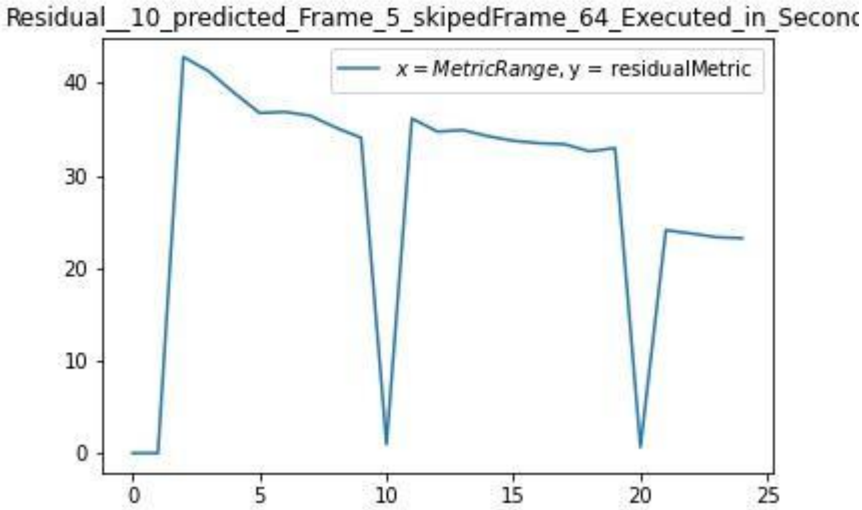

Figure 3.5 I-frame established after every 10 P-frame

The above figure 3.5 shows that 10 predicated frames after establishing I frames and 5 skipped frames, in the end, the frames are predicated with 25 frame samples in 64 seconds. Figure 3.6 shows that 20 predicated frames after establishing I frames and 5 skipped frames, lastly, the frames are predicated with 50 frame samples in 128 seconds.

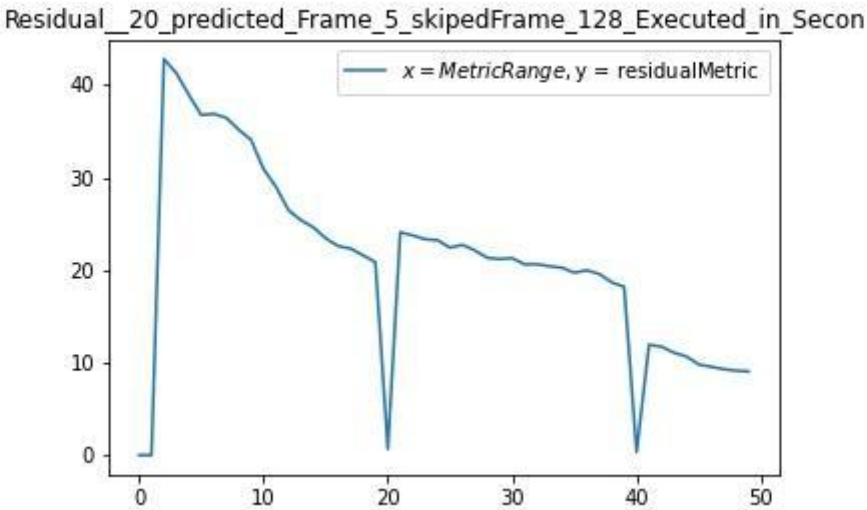

Figure 3.6 I-frame established after every 20 P-frame with 50 sample frame

From figure 3.7 and figure 3.8 residual metric increases if the establishment of the current I-frame is far away from the previous I-frame.

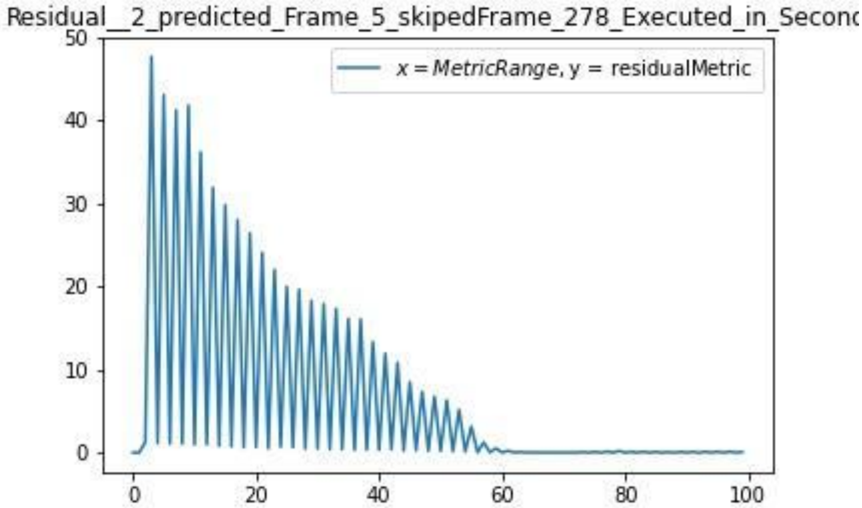

Figure 3.7 I-frame established after every 2 P-frame with 100 sample frame

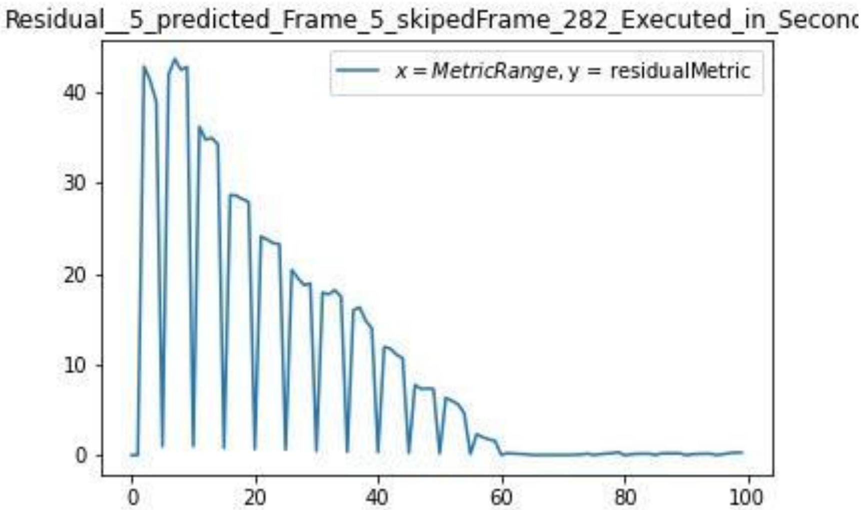

Figure 3.8 I-frame established after every 5 P-frame with 100 sample frame

From figure 3.9 and figure 3.10 residual metric gradually decreases when I-frame established after 50 frames but not if I-frame is established after 25 frames but in the end it is converse.

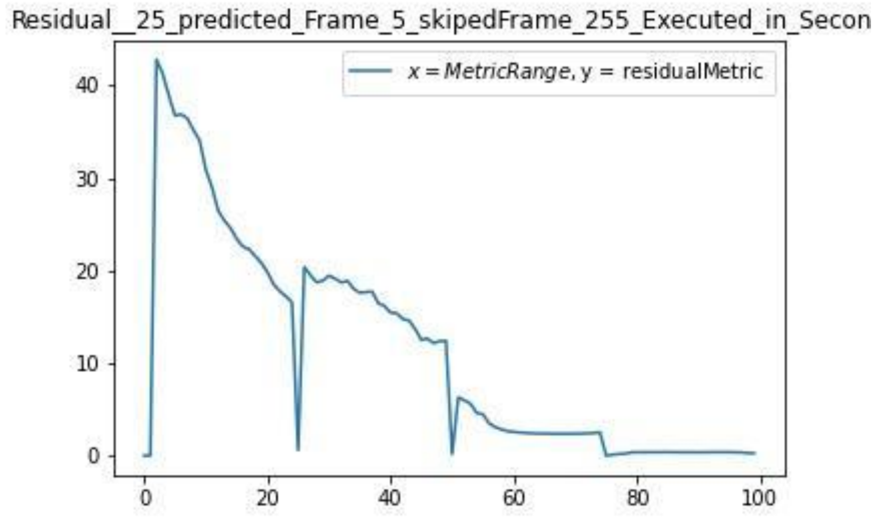

Figure 3.9 I-frame established after every 25 P-frame with 100 sample frame

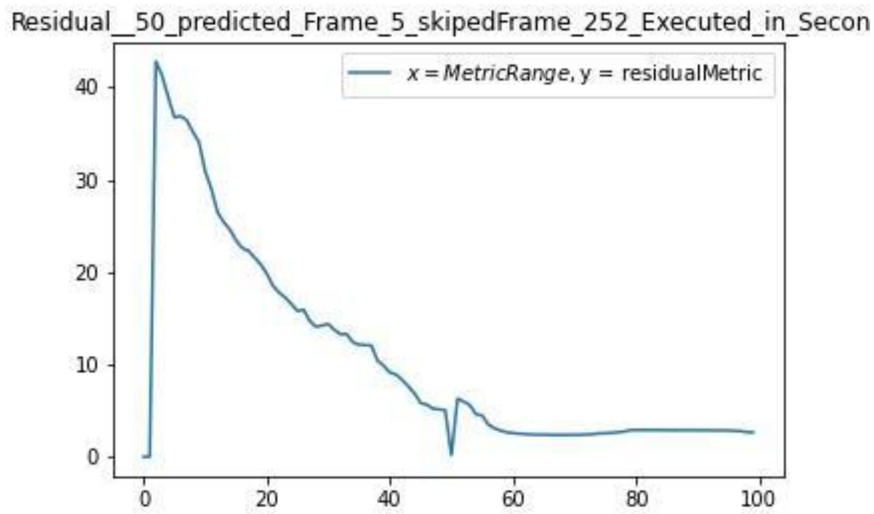

Figure 3.10 I-frame established after every 50 P-frame with 100 sample frame

Here, Mean residual metric calculated for each frame with 25 sample frames, In figure 3.11 and 3.12, X-axis refers to mean metric range, and Y-axis refers to mean residual for 10 compounded frame residual metric.

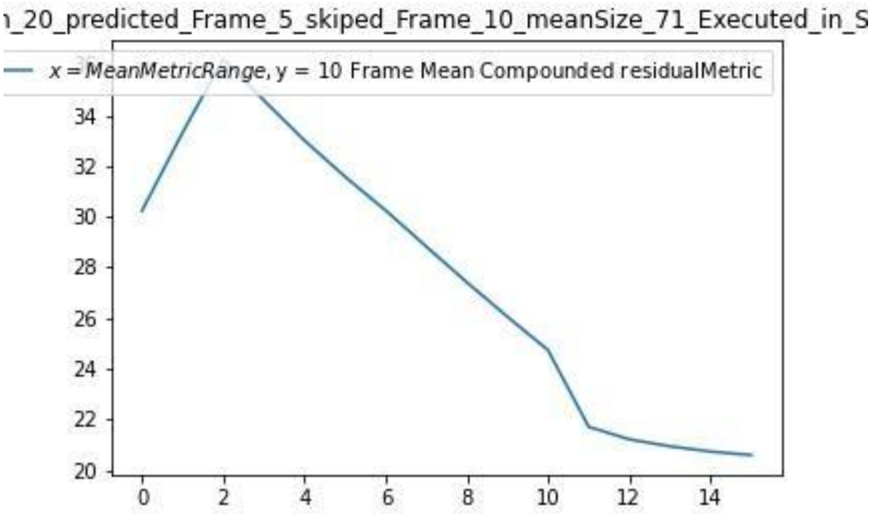

Figure 3.11 Mean residual metric when I-frame established after every 20 P-frame with mean size 10

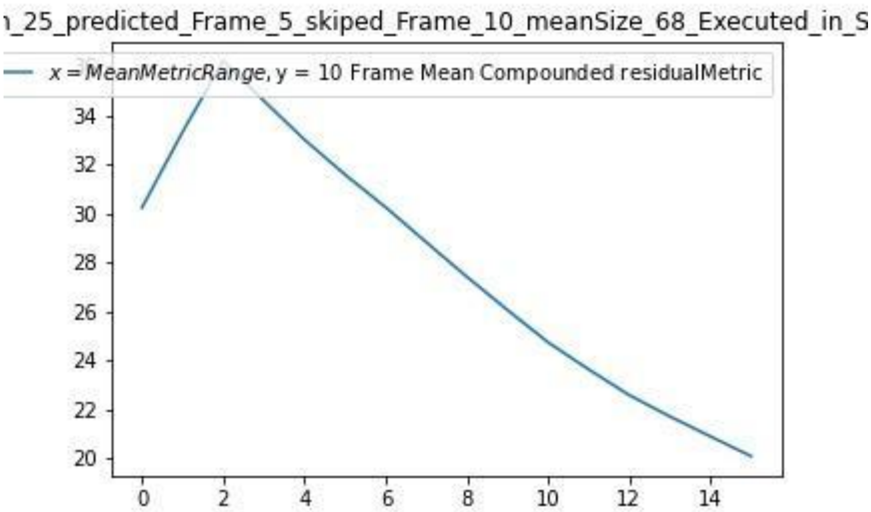

Figure 3.12 Mean residual metric when I-frame established after every 25 P-frame with mean size 10

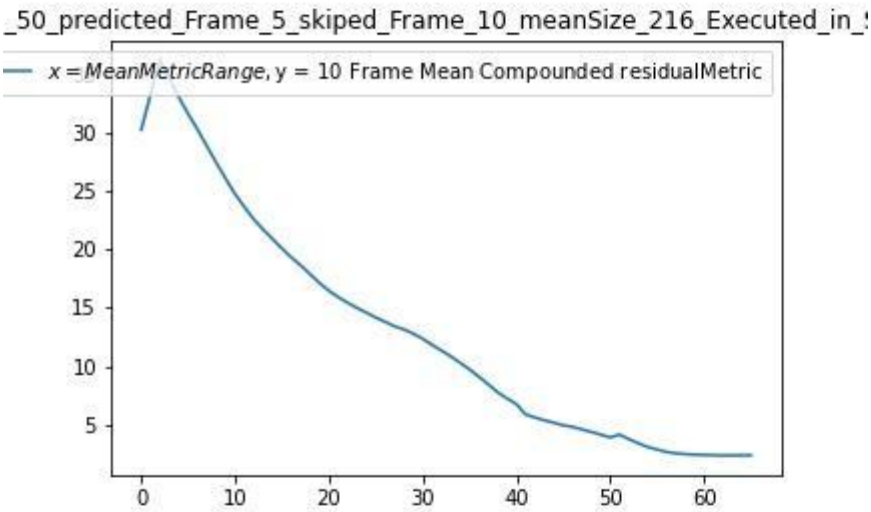

Figure 3.13 Mean residual metric when I-frame established after every 50 P-frame with mean size 10(75 sample frames)

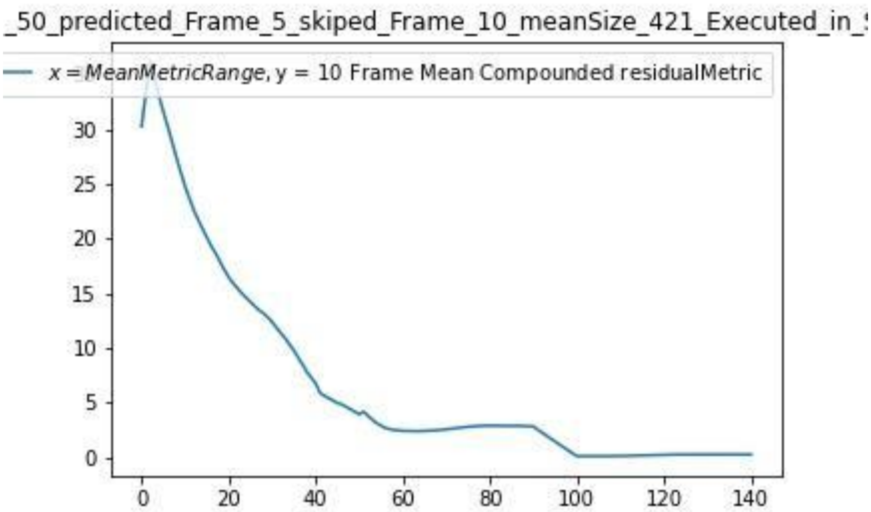

Figure 3.14 Mean residual metric when I-frame established after every 50 P-frame with mean size 10 (100 sample frames)

Here, I frame is established after some B frame predictions, and Residual values are compared as discussed in previous sections.

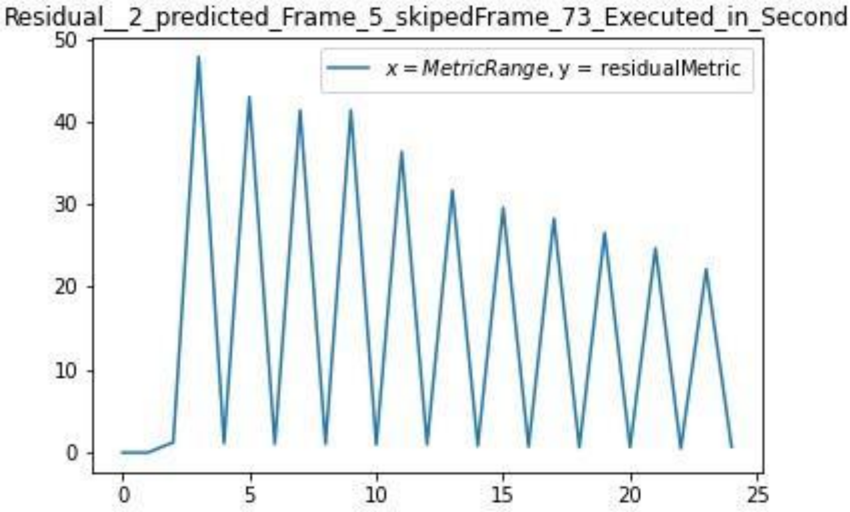

Figure 3.15 I-frame established after every 2 B-frame(25 sample frame)

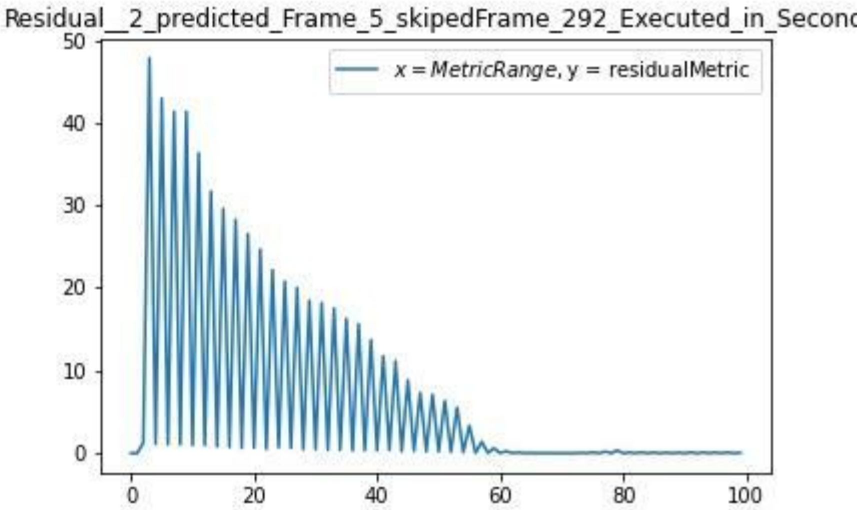

Figure 3.16 I-frame established after every 2 B-frame(100 sample frame)

The below figure 3.17 frames are established as IBBPIBBP, this type of frame format reduces the time when compared to figure 3.16

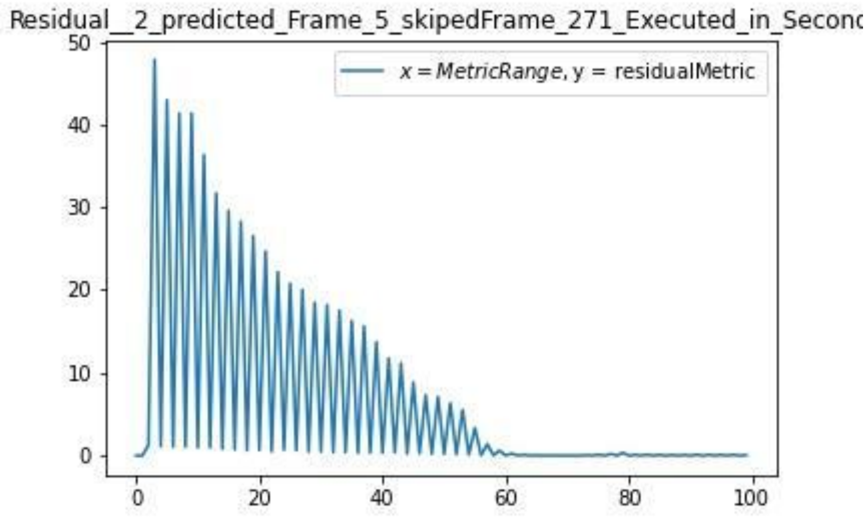

Figure 3.17: Frames are established as IBBPIBBP

In the experiment, if the established frame is nearest to the target frame then there is a minimum difference, in converse if the established frame is far away from the target frame and there is no similar pixels then there is maximum difference.

4 Conclusions

1. The implementation of Block matching algorithm used to predict frames and calculation of residual metric done with target frame and predicted frame.
2. P frame is useful to get minimum residual metric if a frame has a smaller number of differences when compared to target-frame
3. B frame is useful to get minimum residual metric if a frame has a greater number of differences when compared to target-frame.
4. Combination of I, P, B frames can be useful to get lowest residual metric for compressing a video, however it is depending on videos. One particular positions of frames in a video cannot be useful to reduce residual metric for every other videos.

References

[1] Aramvith, S.; Sun, M., T. MPEG-1 AND MPEG-2 Video Standards. Image and Video Processing Handbook, 1999.

[2] MPEG 1 standards https://en.wikipedia.org/wiki/MPEG-1 visited on 21.01.2021

[3] MPEG 1 and MPEG2 standards https://www.cs.ccu.edu.tw/~cwlin/courses/ncku/mpg12c.pdf

[4] MPEG2 standards https://en.wikipedia.org/wiki/MPEG-2 visited on 21.01.2021

[5] Block matching https://en.wikipedia.org/wiki/Block-matching_algorithm visited on 21.01.2021

[6] Sihan Hao; Sumei Li presented a paper as Weighted Mean Absolute Error Metric for Image Quality Assessment https://ieeexplore.ieee.org/document/9301889

[7] Video compression picture types https://en.wikipedia.org/wiki/Video_compression_picture_types visited on 21.01.2021