VOL. 16, NO. 5, MARCH 2021 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A SELECTIVE BLOCK PROCESSING OF FRAMES FOR VIDEO FRAME INTERPOLATION

Brahmadesam T. Madav[1,2], S. A. K. Jilani[2] and S. Aruna Mastani[1]
[1]Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University Anantapur, Andhra Pradesh, India
[2]Department of Electronics and Communication Engineering, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India
E-Mail: madhavbt@gmail.com

ABSTRACT

In this paper, a novel selective block of different pixel sizes, between frames are processed for Frame Interpolation. The current frame is interpolated using previous and next frame. The previous and next frames are converted into blocks. Corresponding blocks are correlated, processed and concatenated after applying median of correlated values as threshold. Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics are used. The algorithm is applied on various standard video sequences with different block sizes and metrics are evaluated.

Keywords: frame interpolation, block sizes, frame rate up conversion, motion estimation, motion vectors, mean square error, peak signal to noise ratio and structural similarity index metric.

1. INTRODUCTION

Over a decade more, video communication has drawn more attention than image communication. In video communication, video compression, video encoding, video transmission and video decoding are different stages. At each stage, the basic operation is done at video frame level. Frame reconstruction or frame interpolation (FI) is done at receiver/decoder side. The FI process is based on Block Based Motion Estimation (ME) with Motion Compensation (MC) techniques. The FI is also known as Frame Rate up Conversion (FRUC). There are various FRUC algorithms [1], [2], [3], [4], [5] and [6]. These algorithms tradeoff between computational complexity and interpolated frame quality. The Motion Compensated Frame Interpolation (MCFI) algorithms uses Block Matching Algorithms (BMA) for ME to create Motion Vector (MV), which is used for estimating motion trajectory. This estimation results in translational motion between frames which helps is missing blocks interpolation.

1.1 MCFI Process

MCFI increases temporal resolution of video frames by interpolating new frames into original sequence. In band limited video temporal down sampling reduces video bit rates. At receiver MCFI restores the lost temporal frame. In MCFI process ME, MV Smoothing and MC Interpolation (MCI) are the steps. ME and MV smoothing steps provide MV. By MCI step, the pixels of the frame to be interpolated are estimated from pixels of low frame rate video. Objects in motion are categorized based on (a) Static background (b) Moving object (c) Uncovered background or region and (d) Newly covered background or region [7]. The categories (c) and (d) results in occlusions.

In this paper we propose the FI method with minimum complexity and can be implemented at the receiver. As in MCFI, block processing stage is used while ME and MC stages are discarded. Correlation and a selective block processing with linear interpolation stages are embedded to reduce complexity. Frame regions or blocks are matched based on intensity [7]. This technique is used widely because of simplicity in implementation and aptness for large motion regions. Frame is divided into regions of pixels called blocks. The size of block is varied as per technique. The hard motion constraint doesn't reflect real motion perfectly for rigid and non-rigid motion. This parameter defines the size of blocks. If large block size is selected real motion can't be represented while small blocks may not include sufficient indication for unique identification of motion.

The decoded frames suffer from blocky artefacts when MC DCT (Motion Compensation Discrete Cosine Transform) coded for low bit rate applications due to translational block-based motion model. Based on realistic structural motion models model-based coding methods are developed. This model-based coding is called as analysis - synthesis coding. In analysis-synthesis extensive use of computer vision and computer graphics is made.

The rest of the paper has Frame Work of Proposed Algorithm in Section-II, Principles in proposed method in Section-III, Video quality metrics in Section-IV, Experimental results and Analysis in Section-V and Section-VI concludes this paper.

VOL. 16, NO. 5, MARCH 2021 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

2. FRAME WORK OF PROPOSED ALGORITHM

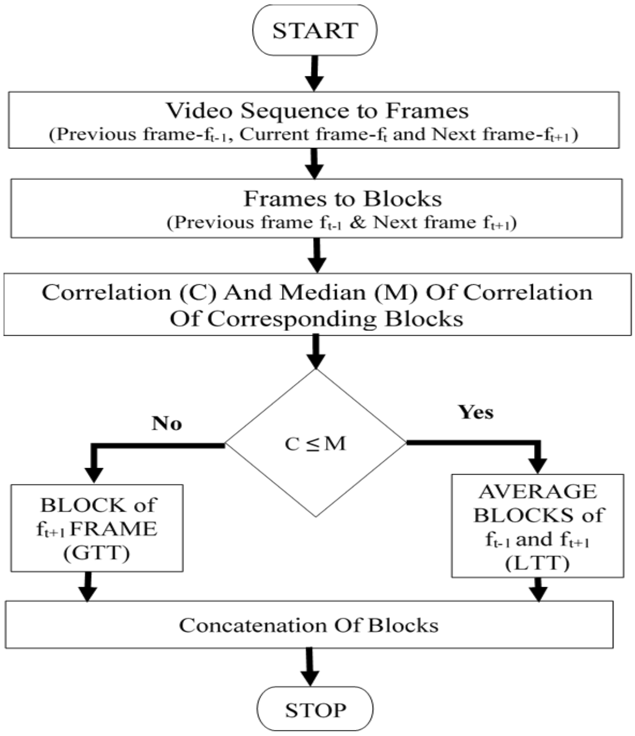

Figure-1. Flow chart of proposed algorithm.

In our method, we considered different video sequences. Each video sequence is converted into frames. The frame is divided into different blocks. The block size is determined by number of pixels considered. The number of pixels in a block is varied in terms of power of two. The block partitioning is limited by frame resolution. It is obvious that the block size can be increased from a minimum to maximum determined by height or width dimension of a frame. Thus, obtained blocks of same size of different frames are used for further processing. Let current frame be f_t, previous frame be f_{t-1} and next frame be f_{t+1}. The previous and next frames are used to interpolate the current/intermediate frame. The previous and next frames are partitioned into blocks of same size. The corresponding blocks in these frames are correlated. The median of the correlated values is generated. The group of correlation values, which are less than median value, are grouped as Less Than Threshold (LTT) values group. The correlated blocks which fall into LTT group are averaged. On the other hand, the correlated values greater than median value is grouped as Greater Than Threshold (GTT) values group. The intermediate frame is interpolated using the blocks, of LTT group, which are averaged pixel wise and the remaining from GTT group blocks of next frame.

3. PRINCIPLES IN PROPOSED METHOD

In this method, we proposed to interpolate a video frame using previous and next frames of the to be interpolated frame. A video sequence is converted into frame. Each individual frame is divided into blocks of different sizes. The size of a block is determined by number of pixels. The number of pixels considered per block are in power of two i.e., 16, 32, 64, 128, 256 and 512 pixels. The number of blocks per frame is restricted by the intrinsic frame resolution and number of pixels per block. The optimal block size is obtained by command 'bestblk' in MatlabR2015b. The steps in block generation are (1) If block size is less than or equal to row size, block size is row size (2) if block size is greater than row size, all values from minimum of row/10 and block size/2 to block size are used, so that minimum padding is required.

Correlation is used in our proposed method to find similarity measure between the blocks. Corresponding blocks in previous and next frame are correlated. The correlation operation is simple, easy to implement and powerful operation that brings out similarity measure because of linearity and shift-invariance properties. Correlation is applied on every pixel in the corresponding blocks. Correlation will be high when the blocks are perfectly matched and will be low when blocks are mismatched.

If the pixel intensity value is high, the correlation will also be high independent of pixel matching nature. This is disadvantage of simple correlation. Hence, we used normalized correlation given by

$$\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{t-1}(i,j) f_{t+1}(i,j)}{\sqrt{\sum_{i=1}^{m} f_{t-1}(i)}\sqrt{\sum_{j=1}^{n} f_{t+1}(j)}} \tag{1}$$

Thus, in our proposed correlation process, the similarity measure ranges between +1 and -1 indicating highly correlated and poorly correlated blocks respectively. All the block correlated values are stored. As a next step in our proposed method, the median amongst the correlated values is generated. If the number of correlated values is odd, after rearranging in sequential manner, the median amongst these correlated values is designated threshold. If the number of correlated values is even, after rearranging in sequential manner, the average of two correlated values centered in the sequence is the median value and is designated as threshold.

The correlation values which are less than threshold is grouped into LTT while the remaining into GTT. Corresponding blocks in LTT group are average and in GTT group are used for frame interpolation. The quality assessment is done in Full-Reference (FR) mode *i.e.*, interpolated frame is compared with available original frame by metrics Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

4. VIDEO QUALITY METRICS

Video Quality Measures are of Objective and Subjective types. In this paper we considered objective type criterion which gives the measure of difference between the original and the reconstructed or processed. Mean Squared Error (MSE) is the simple and basic quality measure which is given as

ARPN Journal of Engineering and Applied Sciences

$$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{x,y}} \sum_{\mathrm{t}} [\mathrm{f}_1(\mathrm{x,y,t}_1) - \mathrm{f}_2(\mathrm{x,y,t}_2)]^2 \quad (2)$$

Where 'N' is number of pixels per frame
'x' is row dimension of frame, 'y' is column dimension of frame and
't' is time or temporal dimension.

For each color component MSE is computed separately.

Another video quality metric is Mean Absolute Difference (MAD). Number of multiplications are reduced to square root times in MAD when compared to MSE as is obvious by the following equation.

$$\mathrm{MAD} = \frac{1}{\mathrm{N}} \sum_{\mathrm{x,y}} \sum_{\mathrm{t}} |\mathrm{f}_1(\mathrm{x,y,t}_1) - \mathrm{f}_2(\mathrm{x,y,t}_2)| \quad (3)$$

PSNR is one of the benchmarks for performance evaluation of objective video quality metrics. It is dependent on estimation of spatial alignment, temporal alignment, gain and level offset between interpolated frame and original frame.

$$\mathrm{PSNR} = 10 \log_{10} \left[\frac{(\text{Maximum peak intesntiy value of video signal})^2}{\text{Mean Squared Error}}\right] \mathrm{d}B \quad (4)$$

In human beings, color sensation is attributed by Luminance and Chrominance. Chrominance is attributed by Hue and Saturation. Hue is defined as color tone which is dependent on peak wavelength of the light. Saturation is defined as purity of color which is dependent on bandwidth of light spectrum.

The structural information of an image is modeled by SSIM quality metric. The structural information change in SSIM defines the image degradation. Luminance, Contrast and Structure comparison are the steps in similarity measurement. Symmetry, Boundedness and Unique maximum are properties of SSIM. SSIM is defined as

$$SSIM(\boldsymbol{x},\boldsymbol{y}) = f(l(\boldsymbol{x},\boldsymbol{y}).c(\boldsymbol{x},\boldsymbol{y}).s(\boldsymbol{x},\boldsymbol{y})) \quad (5)$$

where *l(x,y)* is the luminance at (x,y) location
c(x,y) is the contrast at (x,y) location
s(x,y) is the structural comparison at (x,y) location.

5. EXPERIMENTAL RESULTS AND ANALYSIS

The test sequences are classified into 4 types as tabulated below.

Table-1. Video test sequence classification.

Test Sequence Class	Characteristic
Class A	Low Spatial detail and Low amount of Motion.
Class B	Medium Spatial detail and low motion or vice versa.
Class C	High Spatial detail and Medium amount of motion or vice versa
Class D	Stereoscopic
Class E	Hybrid of natural and Synthetic content

We applied our proposed algorithm on various standard video sequences. The video sequences are of different resolutions from 192 x 144 to 720 x 486. As per our proposed algorithm, the video sequences are converted into frames. Out of the available frames, from the first 50 frames, every even frame is interpolated using the preceding and following odd frames.

The selected frames are partitioned into blocks of sizes 16, 32, 64, 128, 256 and 512 pixels depending on frame intrinsic resolution. The block size cannot exceed one of the dimensions of frame resolution. Corresponding blocks in odd numbered frames are correlated. Median amongst the correlated values is found and used as threshold. The even frame is interpolated using the averaged blocks of LTT group and blocks of GTT group are obtained from following frame. The interpolated frame quality assessment is done by PSNR and SSIM metrics. The PSNR values and SSIM values are averaged and tabulated for various block sizes of 16, 32, 64, 128, 256 and 512 pixels as shown below.

VOL. 16, NO. 5, MARCH 2021 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-2. Average Peak-Signal to Noise Ratio (PSNR) of the reconstructed first 25 even frames when first 50 frames are cosnidered for every individual video sequence.

AVERAGE PSNR								
S. No.	**VIDEO SEQUENCE**	**RESOLUTION**	**16 Pixel**	**32 Pixel**	**64 Pixel**	**128 Pixel**	**256 Pixel**	**512 Pixel**
1	Bus QCIF	192 x 144	17.514	17.467	17.535	17.410		
2	Carphone QCIF		30.639	31.189	31.284	30.623		
3	Coast Guard QCIF		29.278	29.329	29.128	29.407		
4	Container QCIF		38.789	37.874	38.912	38.947		
5	Highway QCIF		24.652	24.539	24.530	24.323		
6	Foreman QCIF	196 x 144	29.562	31.337	31.117	30.483		
7	Stefan SIF	320 x 240	19.899	21.153	21.145	21.411		
8	Tennis SIF		23.150	23.091	22.955	22.882		
9	Highway CIF	352 x 192	29.745	29.744	29.453	29.654		
10	Football CIF	352 x 240	22.353	22.211	22.102	21.780		
11	Akiyo CIF	352 x 288	39.040	40.730	39.450	40.400	39.860	
12	Foreman CIF		26.699	27.979	29.590	29.408	28.882	
13	News CIF		33.652	35.401	34.994	35.102	35.082	
14	Soccer CIF		23.837	23.705	23.683	23.285	23.320	
15	Soccer QCIF	704 x 576	23.150	23.080	23.000	22.969	22.671	22.723
16	Football QCIF	720 x 486	20.053	21.972	21.855	21.791	21.638	21.438

Table-3. Average Structural Similarity Index Metric (SSIM) of the reconstructed first 25 even frames when first 50 frames are considered for every individual video sequence.

AVERAGE SSIM								
S. No.	**VIDEO SEQUENCE**	**RESOLUTION**	**16 Pixel**	**32 Pixel**	**64 Pixel**	**128 Pixel**	**256 Pixel**	**512 Pixel**
1	Bus QCIF	192 x 144	0.4714	0.4715	0.4748	0.4757		
2	Carphone QCIF		0.9526	0.9574	0.9581	0.9542		
3	Coast Guard QCIF		0.8957	0.8954	0.8952	0.8916		
4	Container QCIF		0.9917	0.9904	0.9919	0.9918		
5	Highway QCIF		0.8806	0.8791	0.8803	0.8773		
6	Foreman QCIF	196 x 144	0.9566	0.9638	0.9626	0.9601		
7	Stefan SIF	320 x 240	0.7559	0.7785	0.7768	0.7087		
8	Tennis SIF		0.9045	0.9039	0.9023	0.9033		
9	Highway CIF	352 x 192	0.9597	0.9597	0.9539	0.9582		
10	Football CIF	352 x 240	0.9017	0.9004	0.8989	0.8959		
11	Akiyo CIF	352 x 288	0.9934	0.9943	0.9918	0.9934	0.9938	
12	Foreman CIF		0.5961	0.6218	0.6235	0.6246	0.6217	
13	News CIF		0.9842	0.9882	0.9849	0.9851	0.9876	
14	Soccer CIF		0.9067	0.9063	0.907	0.9042	0.9044	
15	Soccer QCIF	704 x 576	0.9062	0.9061	0.9063	0.9066	0.9046	0.9051
16	Football QCIF	720 x 486	0.8865	0.9067	0.9049	0.9044	0.9028	0.9016

Table-4. Block sizes of considered video sequences with best PSNR and SSIM.

PIXELS PER BLOCK SIZE COMPARISON				
S. No.	**VIDEO SEQUENCE**	**RESOLUTION**	**Best Average PSNR @**	**Best Average SSIM @**
1	Bus QCIF	192 x 144	64	128
2	Carphone QCIF		64	64
3	Coast Guard QCIF		128	16
4	Container QCIF		128	64
5	Highway QCIF		16	16
6	Foreman QCIF	196 x 144	32	32
7	Stefan SIF	320 x 240	128	32
8	Tennis SIF		16	16
9	Highway CIF	352 x 192	16	16
10	Football CIF	352 x 240	16	16
11	Akiyo CIF	352 x 288	32	32
12	Foreman CIF		64	128
13	News CIF		32	32
14	Soccer CIF		16	64
15	Soccer QCIF	704 x 576	16	128
16	Football QCIF	720 x 486	32	32

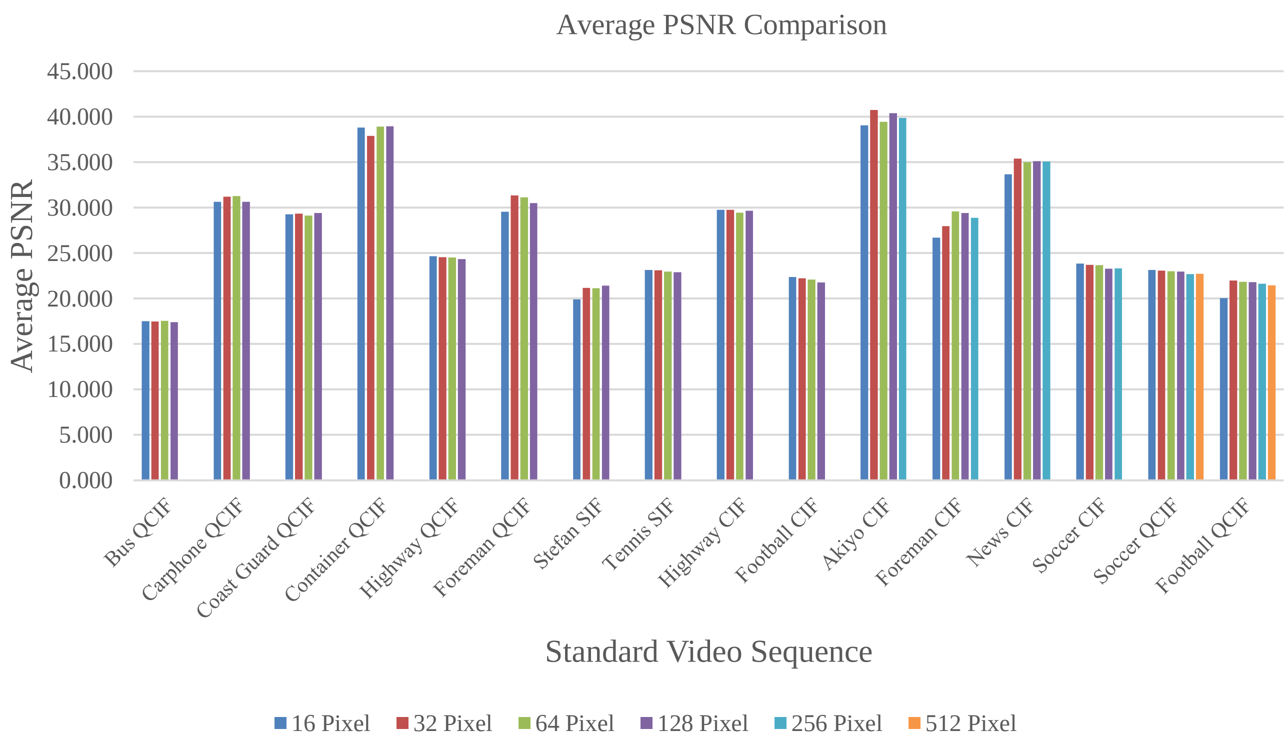

Figure-2. Average PSNRs obtained from the proposed frame interpolation algorithm for standard video sequences of different resolution and block sizes of 16, 32, 64, 128, 256 and 512-pixel size.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

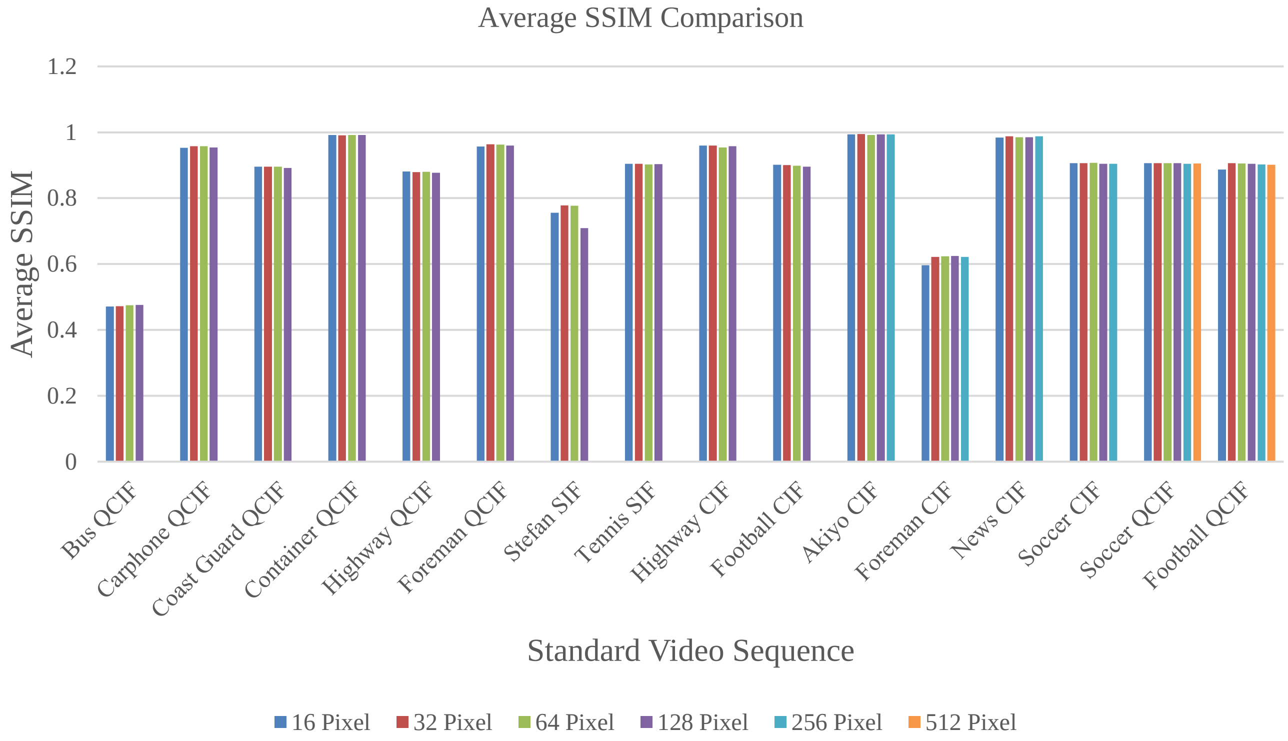

Figure-3. Average SSIMs obtained from the proposed frame interpolation algorithm for standard video sequences of different resolution and block sizes of 16, 32, 64, 128, 256 and 512-pixel size.

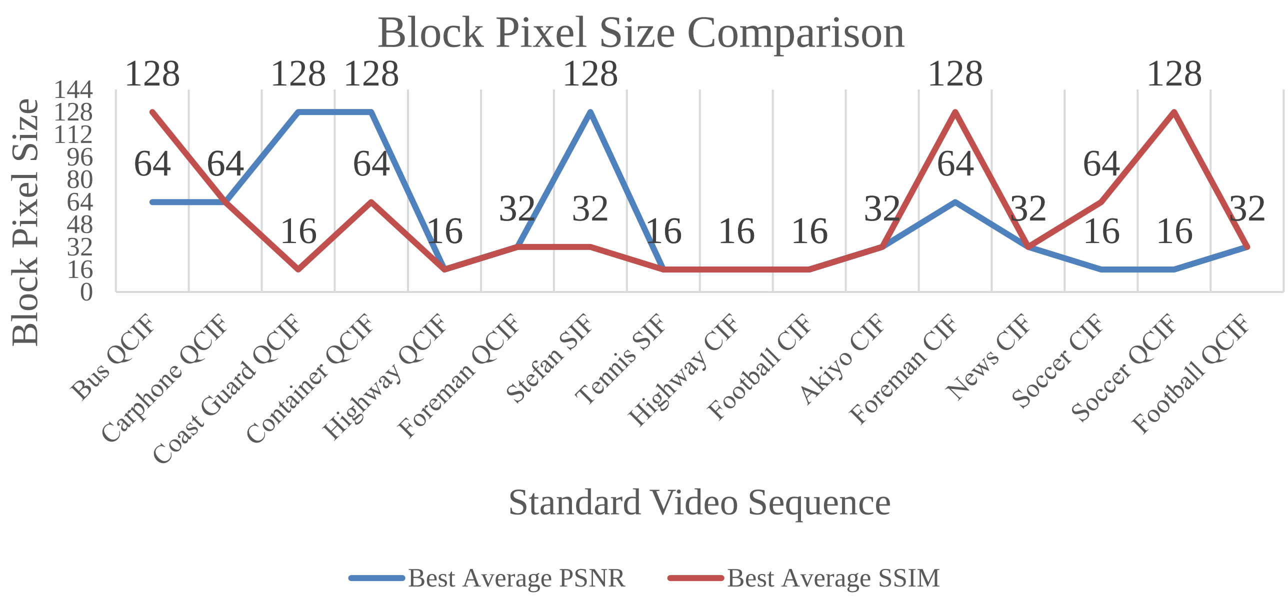

Figure-4. Block sizes of best average PSNRs and SSIMs of standard video sequences obtained by applying proposed algorithm.

The results can be analyzed from Table-4. The video sequences Carphone, Highway and Foreman have best averaged PSNR and averaged SSIM at 64, 16 and 32 block pixel sizes respectively when QCIF (192 x 144) resolution is considered. In case of Tennis, Highway, Football and Akiyo video sequences,16,16,16 and 32pixel block sizes produced best averaged PSNR and best averaged SSIM with SIF (320 x 240) and CIF (352 x 192, 352 x 240) resolution, respectively. The Football video sequence of QCIF (720 x 486) resolution produced best averaged PSNR and SSIM at 32-pixel block size. For the remaining video sequences, the block sizes vary for PSNR and SSIM values.

6. CONCLUSIONS

In this paper we proposed Frame Interpolation method based on pixel correlation values. In this way, complex calculations for ME and MV generation is avoided. Median of the correlated values is used as threshold used in interpolating. Occlusion problem is addressed by this simple method. Experiments showed that PSNR with different block sizes vary significantly while SSIM, which measures Human Visual System approximation, is utmost similar. Thus, SSIM similarity supports our algorithm.

ACKNOWLEDGEMENT

The authors are thankful to 'Xiph.org', an Open Source Community for the availability of standard video sequences. Also, express gratefulness to Madanapalle Institute of Technology and Science for the grants under TEQIP-II.

REFERENCES

[1] Hoda Rezaee Kaviani and Shahram Shirani. 2016. Frame Rate Upconversion Using Optical Flow and Patch-Based Reconstruction. IEEE Trans. Circuits and Sys. For Video Technology. 26(9): 1581-1594.

[2] Nidhi Kunal Parmar and Myung Hoon Sumwoo. 2015. Recent Progress on Block-Based Motion Estimation Techniques. IETE Technical Review. 32(5): 356-363.

[3] Yehuda Dar and Alfred M. Bruckstein. 2015. Motion-Compensated Coding and Frame Rate Up-Conversion: Models and Analysis. IEEE Trans. on Image Processing. 24(7): 2051-2056.

[4] Un Seob Kim and Myung Hoon Sunwoo. 2014. New Frame Rate Up-Conversion Algorithms with Low Computational Complexity. IEEE Trans. on Circuits and Systems for Video Technology. 24(3): 384-393.

[5] Suk-Ju Kang, Sungjoo Yoo and Young Hwan Kim. 2010. Dual Motion Estimation for Frame Rate up Conversion. IEEE Trans. Video Technology. 20(12): 1909-1914.

[6] Ai-Mei Huang and Truong Nguyen. 2009. Correlation-Based Motion Vector Processing with Adaptive Interpolation Scheme for Motion-Compensated Frame Interpolation. IEEE Trans. Image Processing. 18(4): 740-752.

[7] Sumana Gupta. 2016. Digital Video Signal Processing. IIT Kanpur, Web Based Course.