https://doi.org/10.29042/2020-10-3-01-06

A Low Complex Selective Block Correlation for Video Frame Interpolation

[*1]Madav.T.Brahmadesam, [2]Jilani. S.A.K, [3]S. Aruna Mastani
[*1,2]Jawaharlal Nehru Technological University Anantapur, Ananthapuramu, A. P, India.
[2]Department of Electronics and Communication Engineering, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India.
[3]Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University Anantapur, Ananthapuramu, Andhra Pradesh, India
Email: *madhavbt@gmail.com*, *jilani_s_a_k@yahoo.com*, *aruna_mastani@yahoo.com*

Received: 20th March 2020, Accepted: 05th May 2020, Published: 30th June 2020

Abstract

This paper presents a new frame interpolation technique based on correlation between non overlapping macro blocks. The proposed algorithm generates correlation values and identifies the median value as threshold amongst them. The blocks with correlation less than threshold are linearly interpolated. By linear interpolation between blocks reduces artifacts due to occlusions and holes caused by Motion Vector (MV) processing. Our algorithm decreases the computational complexity to minimum. Experimental results demonstrate that our proposed algorithm performs well in terms of SSIM and computational complexity compared to existing algorithms.

Keywords

Frame Interpolation, Block Sizes, Frame Rate Up Conversion, Mean Square Error, Peak Signal to Noise Ratio and Structural Similarity Index Metric.

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 FI algorithms. 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 in missing blocks interpolation.

Frame Interpolation 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 FI restores the skipped 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 order to get true FI, MVs are classified for reliability and merged using reliability map in [1] with vector median filter. The MVs are correlated in MV Processing (MVP) in the algorithm [2] used for reducing artifacts in occluded areas. The MVs at pixel level, block level and sequence level are exploited to reduce artifacts in [3]. This algorithm increases computational complexity. Using available MVs, estimating true Motion Estimation is proposed in [4] to track the projected object motion. The occlusions are reduced by using four intermediate frames thus increasing computational complexity in FRUC based on Variational Image Fusion (VIF) in [5]. The [6] proposes prediction based vector smoothing, partial average based MC and hole interpolation based on motion outliers. Based on artifacts information, MVP of MV Field (MVF) is proposed in [7] with complexity. In [9] optical flow is a part of complexity [10] and three frames are used to interpolate a single frame.

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 not used. Correlation and a selective block processing with linear interpolation stages are embedded in FI process to reduce complexity. Frame regions or blocks are matched based on correlation. This technique can be 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 enough indication for unique identification of motion.

Framework of Proposed Algorithm

In our method, we considered different video sequences. Each video sequence is converted into frames. The frame is divided into non overlapping blocks. The block size is determined by number of pixels. 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 and is Threshold value. 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 linearly interpolated. On the other hand, the correlated values greater than median value are grouped as Greater Than Threshold(GTT) values group. The intermediate frame is interpolated using the blocks, of LTT group, which are linearly interpolated pixel wise and the remaining from GTT group blocks of next frame. Our proposed method is described in the flow chart as in figure 1.

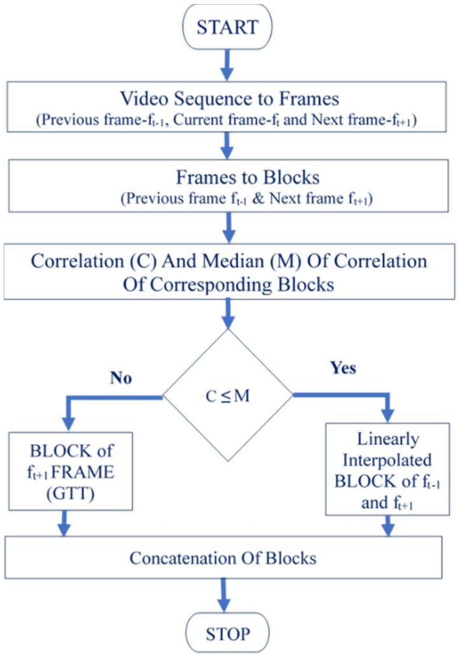
START
Video Sequence to Frames
(Previous frame-f_{t-1}, Current frame-f_t and Next frame-f_{t+1})
Frames to Blocks
(Previous frame f_{t-1} & Next frame f_{t+1})
Correlation (C) And Median (M) Of Correlation Of Corresponding Blocks
No
C ≤ M
Yes
BLOCK of f_{t+1} FRAME (GTT)
Linearly Interpolated BLOCK of f_{t-1} and f_{t+1}
Concatenation Of Blocks
STOP

Figure 1: Flowchart of Proposed Method

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 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

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

Thus, in our proposed correlation process, the similarity measure ranges between +1 and 0 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 and is designated as threshold. 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).

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

$$MSE = \frac{1}{N}\sum_{x,y}\sum_{t}[f_1(x,y,t_1) - f_2(x,y,t_2)]^2 \qquad (2)$$

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.

$$MAD = \frac{1}{N}\sum_{x,y}\sum_{t}|f_1(x,y,t_1) - f_2(x,y,t_2)| \qquad (3)$$

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.

$$PSNR = 10\,log_{10}\left[\frac{(Maximum\ peak\ intesntiy\ value\ of\ video\ signal)^2}{Mean\ Squared\ Error}\right] dB \qquad (4)$$

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

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

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.

Experimental Results and Discussion

The test sequences are classified into 4 types as tabulated in Table 1.

Table 1: Test Video Sequence Classification

Test Sequence Class	Characteristic	Considered Video Sequences
Class A	Low Spatial detail and Low amount of Motion.	-
Class B	Medium Spatial detail and low motion or vice versa.	Foreman, Tennis, News, Soccer
Class C	High Spatial detail and Medium amount of motion or vice versa	Highway, Stefan, Football, Garden
Class D	Stereoscopic	-
Class E	Hybrid of natural and Synthetic content	-

For evaluating our proposed method, we used 8 bench marked video sequences as shown in the table which are CIF (352 x 288) format. In generality adapted by researchers, even frames in these sequences were removed and interpolated using neighboring odd frames. In our proposed method, each frame was divided into non-overlapping blocks. For CIF format videos maximum block size is 256 x 256 pixels. In proposed method corresponding blocks in odd frames are considered for interpolating that corresponding block of even frame. The interpolated blocks are augmented in their respective locations .Objective Assessment (OA) of our proposed method was done by using the metric Structural Similarity Index Measurement (SSIM). For each even frame interpolated, original/skipped even frame was used as reference in calculating SSIM. Results of SSIM are tabulated in Table 1.

Table 2: SSIM of CIF Video Sequences. Bold indicates Maximum Value

Sl.No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Block Size	**Football CIF**	**Foreman CIF**	**Garden SIF**	**Highway CIF**	**News CIF**	**Soccer CIF**	**Stefan SIF**	**Tennis SIF**
16 Pixel	**0.9017**	0.9292	0.596	**0.9597**	0.9842	0.9067	0.7559	**0.9044**
32 Pixel	0.9004	0.9368	0.6218	**0.9597**	**0.9882**	0.9063	0.7785	0.9039
64 Pixel	0.8989	**0.9467**	0.6234	0.9539	0.9849	**0.907**	0.7768	0.9023
128 Pixel	0.8959	0.9456	**0.6246**	0.9581	0.9851	0.9042	**0.7807**	0.9032
256 Pixel	0.8959	0.9444	0.6217	-	0.9876	0.9044	-	-

From Table 2, it can be observed that Football, Highway and Tennis sequences had the highest OA at 16 pixel block size. The Highway and News sequences had the highest OA at 32 pixel block size. The 64 pixel size had the best OA for Foreman and Soccer sequences. Garden and Stefan sequences produced best OA with block size of 128 pixels.

Table 3: SSIM Comparison

	MVP for MCFI	Correlation based MVP	Multilevel VFI	Novel TME for MCFI	FRUC based on VIF	Artifact based MVP for MCFI	FRUC using OF and Patch	Triple Frame based BiDi ME for MCFI	Proposed	Δ w.r.t. Maximum SSIM
Football	0.7151	0.76	0.8105	0.746	0.8551	0.78	0.8096	-	0.9017	0.0466
Foreman	0.9647	0.96	0.9497	0.94	0.9399	0.95	0.9625	0.947	0.9467	-0.0180
Garden	-	-	-	0.976	-	-	-	-	0.6246	-0.3514
Highway	-	-	-	0.922	-	0.92	-	-	0.9597	0.0377
News	-	-	-	0.984	0.9785	-	0.9844	0.985	0.9882	0.0032
Soccer	-	-	-	-	0.8995	-	-	-	0.907	0.0075
Stefan	0.8116	0.89	0.8873	0.942	0.8777	-	0.9514	0.946	0.7807	-0.1707
Tennis	-	-	-	0.956	-	0.9	-	-	0.9044	-0.0516

Our proposed method was compared with 8 benchmark methods developed as shown in table 3. The change in proposed method SSIM with the method that produced maximum SSIM was represented in last column of the table. For Football and Soccer video sequences, our proposed method improves 0.0466 and 0.0075 SSIM respectively when compared to FRUC based on VIF. For Foreman sequence, MVP for MCFI has 0.018 better SSIM than our method. For Garden sequence Novel TME has 0.0377 better SSIM than proposed while in News sequence SSIM is improved by 0.0042. For Stefan sequence proposed method has 01653 better SSIM than Artifact Based MVP. For Tennis sequence proposed method SSIM lies between Novel TME by 0.0516 and Artifact based MVP by 0.0044 as represented graphically in fig.2

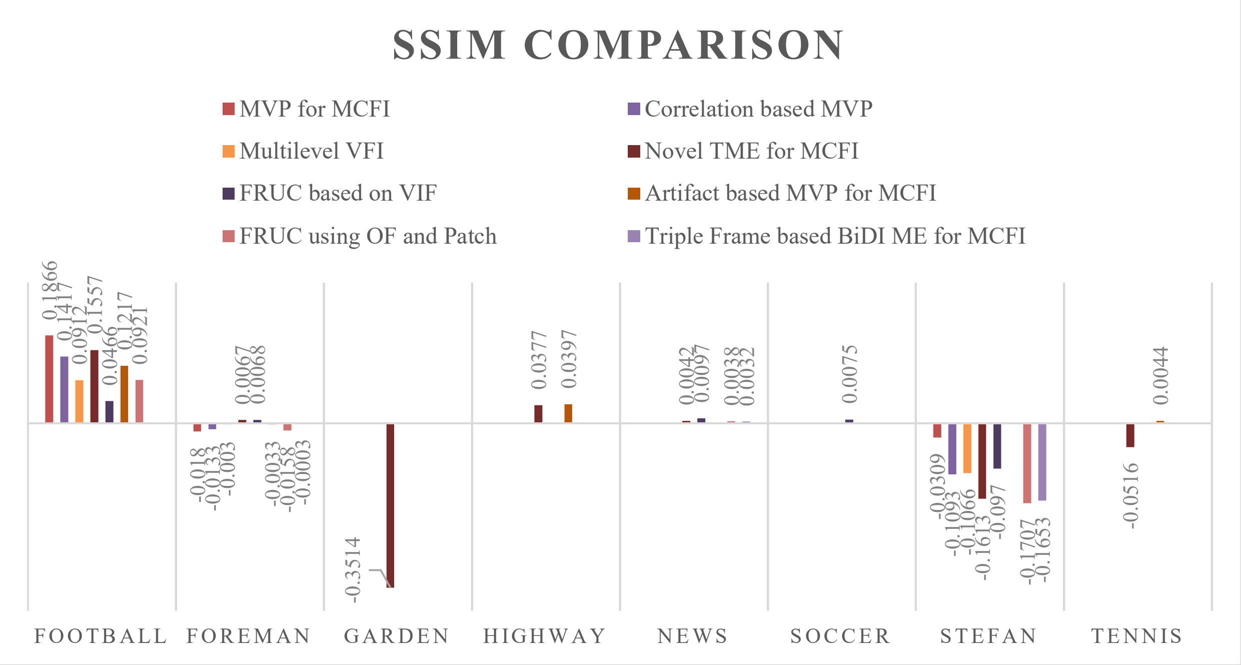

Figure 2: Change in SSIM with respect to Benchmark methods

Computational Complexity

To maintain generality, the frame dimensions were considered as N×N pixels. Each frame is divided into B sized blocks. The correlation had Sum of Absolute Difference(SAD), as shown in MSE. The number of SAD operations determined the computational complexity of the algorithm. As MVP was not used in our proposed method the search window (W) parameter was discarded. The patch (p) size was included in calculation by [9]. The benchmark FI algorithm complexity is tabulated in table 4.

Table 4: Computational Complexity of Various Algorithms

Method	Number of SAD
Multi-level Video Frame Interpolation[3]	N^2 klog(N^2k)
Dual ME [6]	$2(1.5N)^2 (2W+1)^2 - \{(1.5N)^4 – 24(1.5N)^2 + 144\}$
Trilateral filter	$2\{N^2(2W+1)^2\}$
VS-BMC	$N^2 (2W+1)^2+(N^2+4N)(2W+1)^2$
Partial Average based MV Smoothing [6]	$2\{N^2(2W+1)^2\}+(N^2×4+18)$
FRUC using Optical Flow and Patch base [9]	$N^2p^2+N^2\log N$
Proposed	N^2/B

From Table 4, it can be understood that our proposed method has the least computational complexity. The number of computations for the video sequences were graphically represented as in figure3.

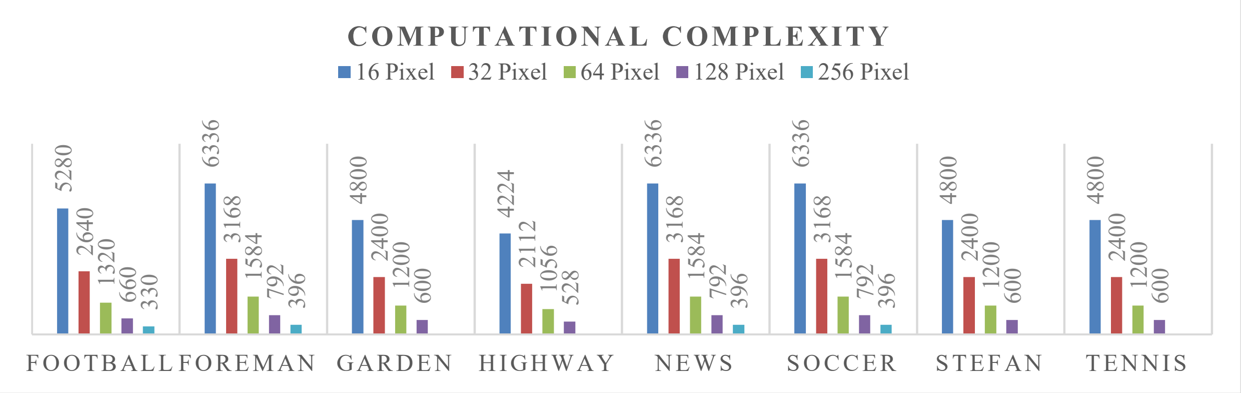

Figure 3: Number of SAD Operations in Proposed Method

Conclusion

In this paper we proposed, correlation of blocks in frames for frame interpolation. Since the motion between frames was linear and correlated highly, interpolating a frame using both end frames had minimum complexity. Our method could interpolate an even ordered frame using neighboring odd ordered frames with minimum complexity. Our experimental results demonstrated that our proposed method performed better in terms of objective assessment SSIM and with minimum computational complexity than other methods.

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, MITS, India for the grants to carry out the research work.

References

1. A.M. Huang and T.Q. Nguyen, "A Multistage Motion Vector Processing Method for Motion-Compensated Frame Interpolation," *IEEE Trans. on Image Processing* , Vol.17, No.5, pp. 694-708, May 2008.
2. A.M. Huang and T.Q. Nguyen, "Correlation-Based Motion Vector Processing with Adaptive Interpolation Scheme for Motion Compensated Frame Interpolation," *IEEE Trans. on Image Processing* , Vol.18, No.4, pp. 740-753, Apr 2009
3. Z.Yu, H.Li, et.al., "Multilevel Video Frame Interpolation: Exploiting the Interaction Among Different Levels", *IEEE Trans. on Circuits and Systems for Video Technology,* Vol.23, No.7, pp.1235-1248, Jul. 2013.
4. S. Dikbas and Y. Altunbasak, "Novel True-motion Estimation Algorithm and its Application to Motion-Compensated Temporal Frame Interpolation", *IEEE Trans. on Image Processing*, Vol.22, No.8, pp.2931-2945,Aug. 2013.
5. W.H.Lee, K.Choi, et.al., "Frame Rate Up Conversion Based on Variational Image Fusion", *IEEE Trans. on Image Processing*, Vol.23, No.1, pp.399-412,Jan. 2014
6. Un Seob Kim and Myung Hoon Sunwoo, "New Frame Rate Up-Conversion Algorithms with Low Computational Complexity", *IEEE Trans. on Circuits and Systems for Video Technology"*, Vol.24, No.3, pp. 384-393, Mar. 2014.
7. Q. Lu, Y.Wang and X. Fang, "An Artifact Information Based Motion Vector Processing Method for Motion Compensated Frame Interpolation," *Journal of Display Technology,* Vol.10, No.9, pp. 775-784, Sep. 2014.
8. Y. Dar and A.M. Bruckstein, "Motion-Compensated Coding and Frame Rate Up-Conversion: Models and Analysis," *IEEE Trans. Image Processing,* Vol.24, No.7, pp. 2051-2066, Jul. 2015.
9. H.R. Kaviani and S. Shirani, "Frame Rate Up Conversion using Optical Flow and Patch-Based Reconstruction", *IEEE Trans. on Circuits and Systems for Video Technology"*, Vol.26, No.9, pp. 1581-1594, Sep. 2016.
10. G.Choi, P. Heo and H. Park " Triple-Frame-Based Bi-Directional Motion Estimation for Motion-Compensated Frame Interpolation", *IEEE Trans. on Circuits and Systems for Video Technology"*, Vol.29, No.5, pp. 1251-1258, May 2019.