TEMPORAL REDUNDANCY REDUCTION IN WAVELET BASED VIDEO COMPRESSION FOR HIGH DEFINITION VIDEOS

S. Sowmyayani[1] and P. Arockia Jansi Rani[2]

[1]Department of Computer Science, Kristu Jayanti College, India
[2]Department of Computer Science and Engineering, Manonmanian Sundaranar University, India

Abstract

Data Storage and Communication plays a significant role in every human. Digital images and videos are stored in mobile and other storage devices. More specifically, video data requires huge amount of storage space for which the storage devices are more expensive. Hence there is a necessity of reducing the storage space of the data. Video compression is more common in all researches. In this work, the role of wavelets in video compression is studied. The temporal redundant data are converted to spatial data which are then transformed to wavelet coefficients. The low frequency components are removed from these wavelet coefficients. The proposed method is tested with some video sequences. The performance of the proposed method is analyzed by comparing it with the existing recent methods and with the state-of-art H.265 video coding standard. The experimental results substantially proved that the proposed method achieves 3.8dB higher PSNR than H.265 and 1.6dB higher PSNR than recent wavelet based video codecs.

Keywords:
H264/AVC, Temporal Redundancy, Spatial Redundancy, High Definition Videos, Wavelet Transform

1. INTRODUCTION

In the past two to three decades, there has been tremendous progress in the research and development of digital video technology with significant progress in digital video compression techniques. Since the very early implementation of the H.120 video compression standard in the mid-1980s, major standards have been developed by both ISO/IEC and ITU-T and have found useful applications. Digital video finds applications in many areas, ranging from broadcast applications to interactive applications to locative applications.

The significant increase in resolution and perceptual quality may make the recent video compression standard to be unable to efficiently meet the compression demand. Every successful video coder employs Block Motion Compensation (BMC) [1], where the target frame is partitioned into blocks, and each such macroblock gets assigned the "prediction" vector which results in the smallest prediction residual between prediction residual and coding cost. Over the past few decades, video codecs have evolved from using one fixed block size (H.261), to using a large variety of blocks. The large variety of block sizes allows to better account for the fact that blocks have difficulty in representing motion in the vicinity of moving object boundaries. The recent video coding standard HEVC allows for blocks to inherit the motion of their spatial neighbors, which favors smoothness within moving objects. By eliminating artificial block boundaries to overcome the above problem, Dominic et al., describe true discontinuities in the motion flow [2]. There is a body of research on estimating piecewise-smooth motion fields with sharp

transitions at object boundaries [3-4], but this remains a challenging task.

Recently, dense motion estimation methods for video coding have been introduced which optimize for both smoothness and compressibility of the motion field [5]. Young et al. [6] explicitly handle motion discontinuities and support compression-regularized optical flow, where piecewise-smooth motion fields are estimated, and discontinuities are explicitly coded. As described in [7], such motion fields are highly scalable. In video coding, scalability refers to an encoding of a video in an embedded way such that lower qualities are embedded within higher qualities. The latest video codecs (H.264 and HEVC) have scalable extensions, H.264/SVC [8] and SHVC [9], respectively. The closed-loop nature of these codecs requires the base and enhancement layer to be decided at the encoder, limiting their scalability attributes.

A natural alternative for highly scalable video coding involves the use of wavelets. Various Wavelet-based Scalable Video Coders have been suggested in the literature. The common approach is spatial domain motion compensated temporal filtering (t+2D), where the wavelet transform is first applied in the temporal domain, followed by a 2D-DWT of the temporal subbands [10-11]. A t+2D spatio-temporal wavelet based video coder [12] is developed which exploits firstly the temporal redundancy between the different frames. The Three Step Search, an algorithm of Block Matching is used for the motion estimation. A spatial decomposition is then applied to the obtained temporal sub-bands.

Although not in most codecs, wavelet transform is used in JPEG2000 still image coding and in the VC-2 video codec. At low bit-rates and high compression, wavelet transform-based methods show better visual quality for reconstructed image and video, compared to those from block transforms. Wavelet-based schemes that handle motion discontinuities because of intrinsic properties of the setup have shown to have improved performance [13-14]. For generic video data, a more general way of handling motion discontinuities is required. Mathew et al. [15] introduced a highly scalable representation of discontinuities in both resolution and precision.

Temporal redundant data is more than spatial redundant data in a video sequence. The temporal redundant data is converted or transformed to spatial redundant data. Hence each frame after conversion consists of more spatial data. The spatial redundant frame is then converted to wavelet coefficients. The low frequency components are removed as it does not have more significance in video data. High definition videos are tested with some experiments. The performance of the proposed work is compared with recent video compression standards.

The remaining of the chapter is organized as follows: section 2 describes the overall system architecture of the proposed method. Section 3 explains the detailed algorithm of the proposed method. Section 4 demonstrates the experimental results followed by conclusion in section 5.

2. SYSTEM ARCHITECTURE

The overall system architecture is shown in Fig.1. The video sequence contains more temporal redundant than spatial redundant data. Hence it is converted to video frames consisting of more spatial redundant data using T2SR transformation [16]. Now the spatially redundant data is transformed to wavelet coefficients. In the third level decomposition, the HH components are removed and it is then encoded.

The Fig.2 gives the working architecture of the proposed encoder. In the decompression phase, the encoded bitstream is decoded using SPIHT decoder. Then inverse wavelet transform is applied which is then given to Inverse T2SR transformation, the reconstructed video sequence is obtained. The Fig.3 illustrates this process. The T2SR transformation is briefly explained in section 3.
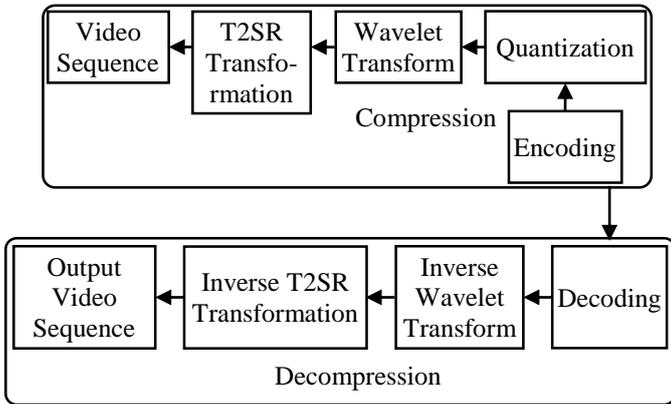

Fig.1. Proposed System Architecture

3. PROPOSED METHODOLOGY

3.1 PROPOSED COMPRESSION PROCEDURE

The design procedure for the T2SR transformation is described in [16]. It is assumed that the video sequence is having x frames of size $m \times n$. Initially, each frame in the video sequence is converted into block matrices. Any matrix A whose elements are matrices themselves is called a block matrix [17] which is given as,

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix} \quad (1)$$

where, A is a block matrix where $\{A_{i,j}\}$ are $p \times p$ matrices and p is a constant. The matrix A is called an $m \times n$ block matrix of basic dimension $p \times p$.

A frame in the input video sequence (F^x) is represented by a block matrix as,

$$F^x = \begin{bmatrix} F_{1,1}^x & F_{1,2}^x & \cdots & F_{1,m}^x \\ F_{2,1}^x & F_{2,2}^x & \cdots & F_{2,m}^x \\ \vdots & \vdots & \ddots & \vdots \\ F_{n,1}^x & F_{n,2}^x & \cdots & F_{m,n}^x \end{bmatrix} \quad (2)$$

in which the block matrix elements are represented by,

$$F_{i,j}^x = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,p} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p,1} & b_{p,2} & \cdots & b_{p,p} \end{bmatrix} \quad (3)$$
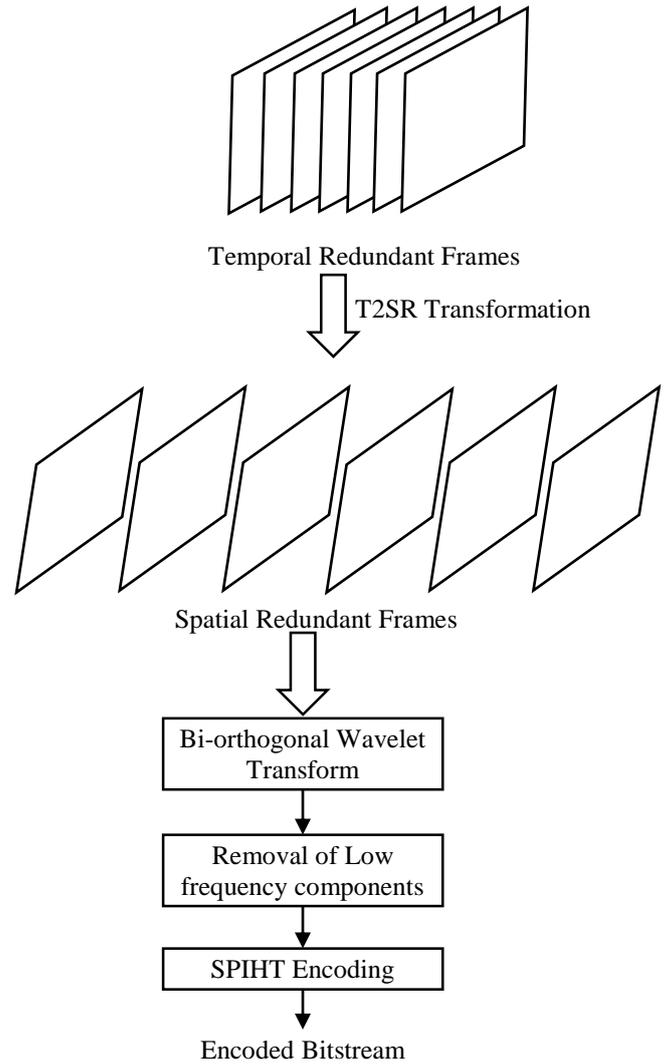

Fig.2. Proposed Encoder Architecture

Block matrix $F_{i,j}^x$ is converted into the row vector $R_{i,j}^x$ as given below

$$R_{i,j}^x = \begin{bmatrix} b_{1,1}, b_{1,2}, \ldots, b_{p,p} \end{bmatrix} \quad (4)$$

where, $i = 1$ to n/p and $j = 1$ to m/p.

The first Temporal to spatial redundancy transform matrix F_k^R is then defined as,

$$F_k^R = \begin{bmatrix} R_{i,j}^1 \\ R_{i,j}^2 \\ \vdots \\ R_{i,j}^x \end{bmatrix} \quad (5)$$

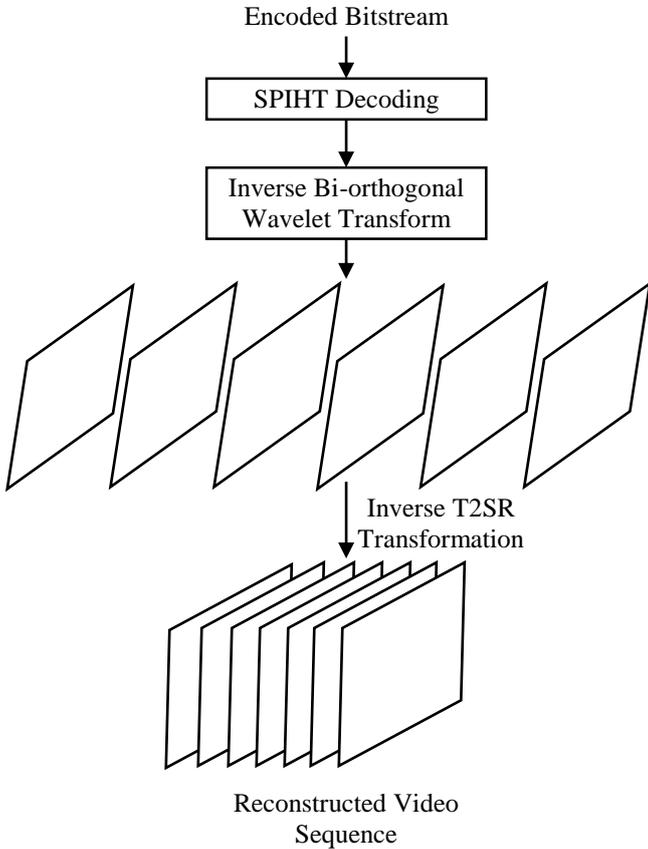where, $i = 1$ to n/p and $j=1$ to m/p and $k=p\times p$.

Fig.3. Proposed Decoder Architecture

In T2SR transformation, Frame resizing is done. Each frame in the video sequence is divided into blocks of size 16×16. The first block in each frame of the video sequence is placed as the first output frame as 1×256 vector. Similarly, for all other blocks, the respective output frames are created. Hence, the number of output frames is equal to the number of blocks. And, the size of the output frame is the product of number of frames in input video sequence and 256. After this conversion, each frame in the output video sequence is converted to the size in the power of two by concatenating consecutive frames. The T2SR transformation is shown pictorially in Fig.4.

The output video sequence obtained in T2SR transformation is transformed to wavelet coefficients. The wavelet used in this work is the bi-orthogonal Cohen–Daubechies–Feauveau wavelet (CDF9/7) because it has been shown that CDF9/7 has the best trade-off between complexity and speed while yielding high PSNR when the image is reconstructed [18]. Also, the CDF9/7 wavelet is widely used on the standard JPEG2000. Three levels of decomposition were used because there is no improvement on using higher decomposition levels when compressing images. In the third level of decomposition, LLHH components are removed. This does not have any significance over decoding. In Fig.5, the

three levels of decomposition are shown. The shaded region in the third level of decomposition is removed.
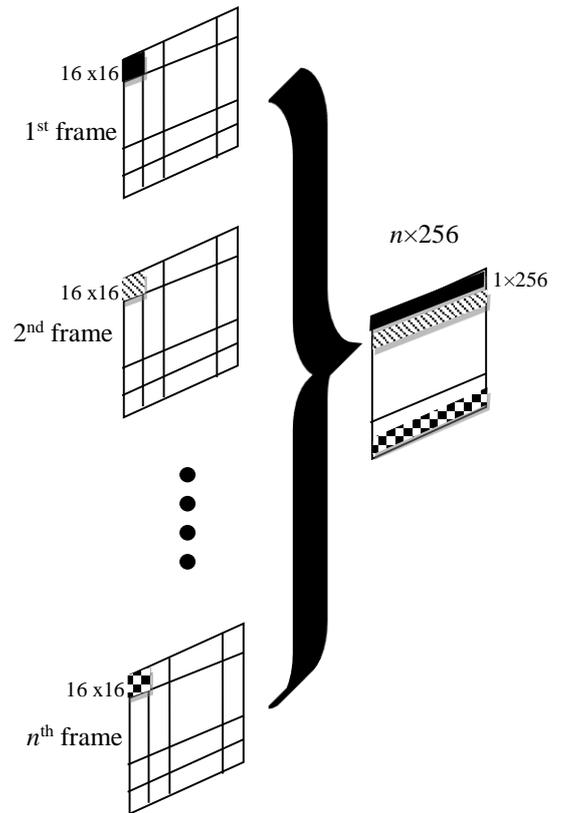

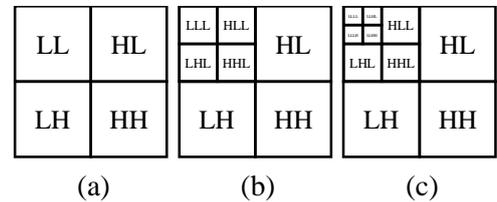Fig.4. T2SR Transformation for first output frame

Fig.5(a). First Level (b) Second Level (c) Third Levels of decomposition

The powerful wavelet-based image compression method called Set Partitioning in Hierarchical Trees (SPIHT) [19] is used to encode the transform coefficients. In SPIHT encoder, the wavelet coefficients are quantized using simple uniform quantization. Then SPIHT encoding is used for encoding.

3.2 PROPOSED DECOMPRESSION PROCEDURE

During Decoding, SPIHT decoding is used followed by de-quantization process which is followed by Inverse Wavelet transform. After the steps described in the previous subsections are completed, frame reconstruction is done using inverse T2SR Transformation to get back the original video sequence. The size of the output frame after decoding is converted to $x\times256$ where x is the number of frames in the input video sequence. The resulting intermediate video sequence is reconstructed by placing each row (1×256) in the sequence as 16×16 blocks in the corresponding output frame. That is, from the first frame, all first blocks are

placed in each output frames. Finally, the output sequence has the same size as input video sequence.

4. EXPERIMENTAL RESULTS

The proposed video coding method is tested using standard test video sequences [20]. These test video sequences are in Common Intermediate Format (CIF) which is of size 352×288. The results were evaluated using two metrics: PSNR and Structural Similarity Index (SSIM). PSNR is a classic method for evaluating image compression algorithms that evaluates how approximated the reconstructed frame is to the original frame taking into account how much of the information of the intensity of the pixels was corrupted. SSIM is another metric less commonly used for assessing image quality. SSIM aims to assess the image using a perception-based model. Images reconstructed with high quality will yield a SSIM index closer to 1, where a perfect reconstruction as in lossless compression will yield a SSIM index of 1. The results obtained by the proposed method are shown in Fig.6.

(a)   (b)   (c)

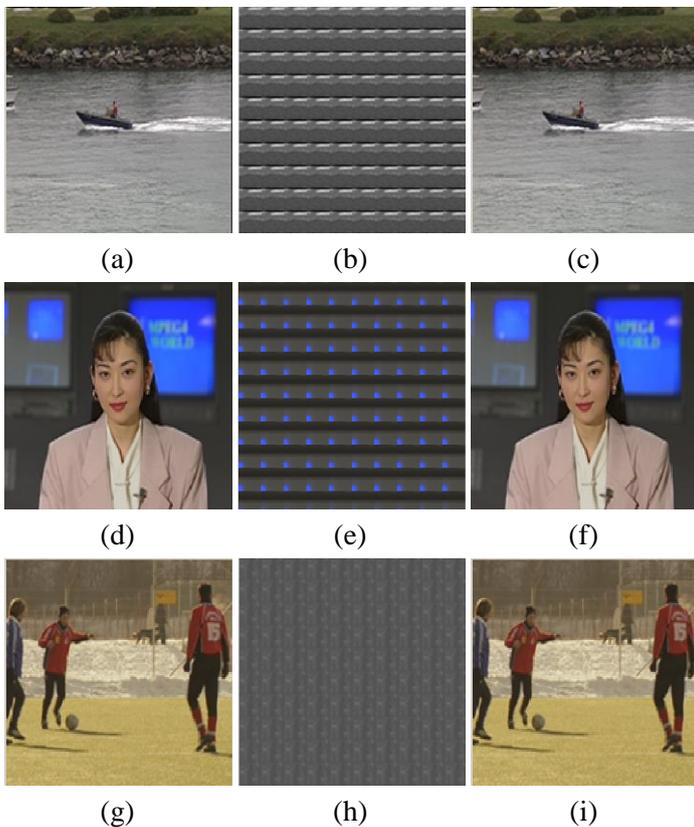(d)   (e)   (f)

(g)   (h)   (i)

Fig.6 (a) (d) (g) An input Frame (b) (e) (h) First T2SR Transformed Frame (c) (f) (i) Output Frame of coastguard, Akiyo and Soccer sequence

The performance of the proposed method is analyzed by comparing it with SPECK Codec and AFV - SPECK Codec [21] and H.265. The Table.1 and Table.2 show the PSNR and SSIM comparison of the proposed method with the above-mentioned methods respectively.

Table.1. PSNR Comparison of the proposed method with other methods and H.265

Video Sequence	PSNR (dB)				
	BPP	H.265	SPECK	AFV-SPECK	Proposed
Bowing	1.30	28.65	39.91	35.07	37.96
City	1.18	29.23	31.22	30.69	31.98
Bus	0.97	28.87	29.83	29.67	30.05
Container	1.45	29.05	33.34	31.60	32.58
Akiyo	1.19	28.96	35.42	33.12	35.8
Coastguard	1.29	28.93	30.46	29.82	30.87
Paris	1.28	28.79	30.45	29.53	31.85
Mobile	1.49	29.08	29.14	29.01	30.25
Students	1.13	29.10	32.97	31.98	32.79
Soccer	1.37	29.33	34.89	32.23	34.25

Table.2. SSIM Comparison of the proposed method with other methods and H.265

Video Sequence	SSIM				
	BPP	H.265	SPECK	AFV-SPECK	Proposed Method
Bowing	1.30	0.85	0.97	0.96	0.88
City	1.18	0.63	0.84	0.84	0.86
Bus	0.97	0.66	0.76	0.75	0.8
Container	1.45	0.76	0.87	0.86	0.85
Akiyo	1.19	0.78	0.94	0.92	0.93
Coastguard	1.29	0.66	0.81	0.80	0.82
Paris	1.28	0.76	0.85	0.83	0.84
Mobile	1.49	0.74	0.82	0.81	0.83
Students	1.13	0.71	0.89	0.89	0.9
Soccer	1.37	0.74	0.93	0.92	0.93

From Table.1, it is clear that the proposed method achieves 3.8dB higher PSNR than H.265. It is also observed that the proposed wavelet based method achieves 1.6dB higher PSNR than recent wavelet based SPECK methods. The Table.2 shows that the proposed method achieves better SSIM than H.265 and recent wavelet based video codecs. The Fig.7 shows the bar chart displaying PSNR comparison of the proposed method with H.265 and other methods.
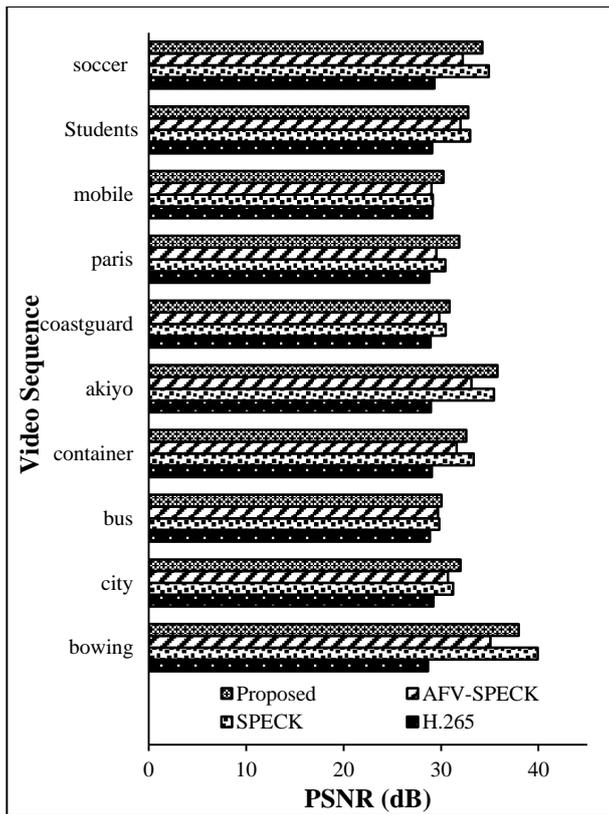
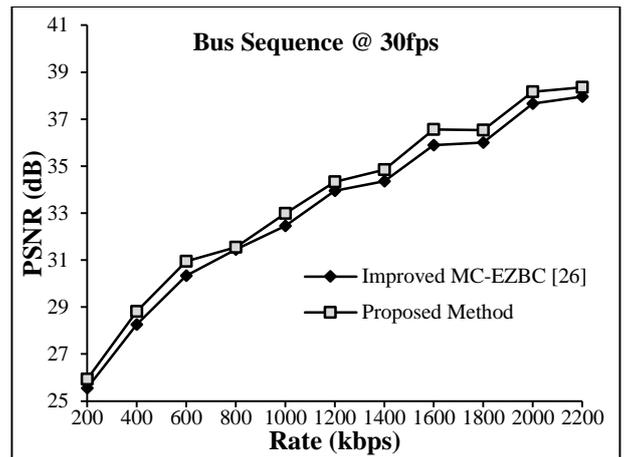Fig.7. PSNR comparison of the proposed method with other recent methods

In addition to PSNR and SSIM, the proposed method is also evaluated using computation time. To compare the proposed method, HVSBM [22], RWTH-MC-EZBC [23], RPI-MC-EZBC [24], ENH-MC-EZBC [25] and Improved MC-EZBC [26] are used. The average time saving $\Delta T(\%)$ in measurement of encoding time is defined by:

$$\Delta T = \frac{\left(T_{\text{Pro}} - T_{\text{Ref}}\right)}{T_{\text{Ref}}} \times 100\% \qquad (6)$$

where, T_{Pro} and T_{Ref} are the encoding time of the proposed algorithm and reference algorithm, respectively; "+" means increase; and "-" means decrease. The Fig.8 shows the RD performance comparison of improved MC-EZBC with the proposed method for Bus and Soccer sequence.

Table.3. Encoding Time Comparison of Proposed Method with other Methods

Sequence/ Method	HVSBM	RPI-MC-EZBC	RWTH-MC-EZBC	ENH-MC-EZBC	Improved MC-EZBC	Proposed
Bus	-31.01	+21.56	+15.74	-13.92	-8.99	-5.08
Soccer	-29.43	+28.52	+23.96	-16.45	-11.42	-7.25

(a) Bus Sequence @ 30fps
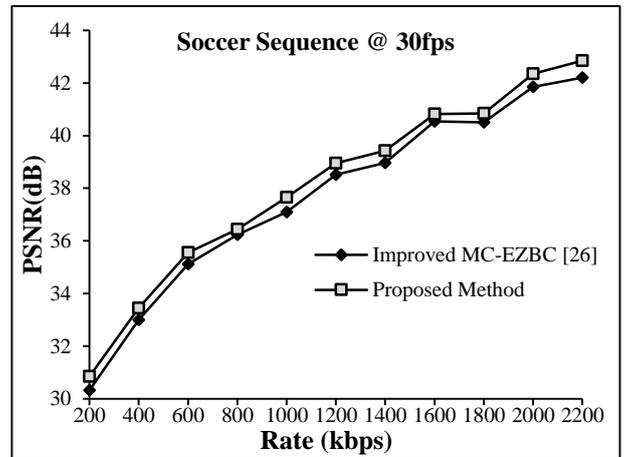
(b) Soccer Sequence @ 30fps

Fig.8. RD Comparison of the Proposed Method with Improved MC-EZBC [26] Method

From Fig.8, it is clear that the proposed method achieves little higher PSNR gain when compared to improved MC-EZBC Method.

5. CONCLUSIONS

Video Compression technology is essential in all digital storage devices. In this work, the role of wavelet in video compression is studied. The temporal redundant data present in the video sequence are converted to spatial redundant data. This data is again transformed into wavelet coefficients. The HH components of third level of wavelet decomposition is removed which does not have significant change in the quality of the video. The proposed method is tested on some video sequences. The performance of the proposed method is analyzed by comparing it with recent wavelet based video compression methods. From the results, it is proved that the proposed method achieves better PSNR than H.265 and other recent wavelet based video codecs.

REFERENCES

[1] J. Jain and A. Jain, "Displacement Measurement and its Application in Interframe Image Coding", *IEEE*

Transactions on Communications, Vol. 29, No. 12, pp. 1799-1808, 1981.

[2] Dominic Rufenacht, Reji Mathew and David Taubman, "Novel Motion Field Anchoring Paradigm for Highly Scalable Wavelet-Based Video Coding", *IEEE Transactions On Image Processing*, Vol. 25, No. 1, pp. 39-52, 2016.

[3] L. Xu, J. Jia and Y. Matsushita, "Motion Detail Preserving Optical Flow Estimation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 9, pp. 1744-1757, 2012.

[4] J. Wulff and M.J. Black, "Modeling Blurred Video with Layers", Available at: http://files.is.tue.mpg.de/black/papers/WulffECCV2014.pdf.

[5] G. Ottaviano and P. Kohli, "Compressible Motion Fields", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2251-2258, 2013.

[6] S.I. Young, R.K. Mathew and D.S. Taubman, "Joint Estimation of Motion and Arc Breakpoints for Scalable Compression", *Proceedings of IEEE Conference on Signal and Information Processing*, pp. 479-482, 2013.

[7] S.I. Young, R.K. Mathew and D.S. Taubman, "Embedded Coding of Optical Flow Fields for Scalable Video Compression", *Proceedings of 16th IEEE Workshop on Multimedia and Signal Processing*, pp. 1-6, 2014.

[8] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 9, pp. 1103-1120, 2007.

[9] P. Helle et al., "A Scalable Video Coding Extension of HEVC", *Proceedings of IEEE Data Compression Conference*, pp. 201-210, 2013.

[10] A. Secker and D. Taubman, "Motion-Compensated Highly Scalable Video Compression using an Adaptive 3D Wavelet Transform based on Lifting", *Proceedings of IEEE International Conference on Image Processing*, pp. 1029-1032, 2001.

[11] B. Pesquet-Popescu and V. Bottreau, "Three-Dimensional Lifting Schemes for Motion Compensated Video Compression", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1793-1796, 2001.

[12] I. Charfi and M. Atri, "Spatio-Temporal Wavelet Based Video Compression: a Simulink Implementation for Acceleration", *International Review on Computers and Software*, Vol. 10, No. 5, pp. 1-6, 2015.

[13] H.G. Lalgudi, M.W. Marcellin, A. Bilgin, H. Oh and M.S. Nadar, "View Compensated Compression of Volume Rendered Images for Remote Visualization", *IEEE Transactions on Image Processing*, Vol. 18, No. 7, pp. 1501-1511, 2009.

[14] J.U. Garbas, B. Pesquet-Popescu and A. Kaup, "Methods and Tools for Wavelet-based Scalable Multiview Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 21, No. 2, pp. 113-126, 2011.

[15] R. Mathew, D. Taubman and P. Zanuttigh, "Scalable Coding of Depth Maps with R-D Optimized Embedding", *IEEE Transactions on Image Processing*, Vol. 22, No. 5, pp. 1982-1995, 2013.

[16] S. Sowmyayani and P. Arockia Jansi Rani, "An Efficient Temporal Redundancy Transformation for Wavelet based Video Compression", *International Journal of Image and Graphics*, Vol. 16, No. 3, pp.1-6, 2016.

[17] Anil. K. Jain, "*Fundamental of Digital Image Processing*", PHI Publication, 2014.

[18] S. Mallat, "*A Wavelet Tour of Signal Processing*", 3rd Edition, Academic Press, 2008.

[19] Amir Said and William A. Pearlman, "A New, Fast and Efficient Image Codec based on Set Partitioning in Hierarchical Trees", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, No. 3, pp. 231-247, 1996.

[20] Xiph.org Foundation, Available at:https://media.xiph.org/video/derf/.

[21] J.C. Galan-Hernandez, V. Alarcon-Aquino, O. Starostenko, J.M. Ramirez-Cortes and Pilar Gomez-Gil, "Wavelet-based Frame Video Coding Algorithms using FOVEA and SPECK", *Engineering Applications of Artificial Intelligence*, Vol. 69, pp. 127-136, 2018.

[22] S.J. Choi and J.W Woods, "Motion-Compensated 3-D Subband Coding of Video", *IEEE Transactions on Image Processing*, Vol. 8, No. 2, pp. 155-167, 1999.

[23] M. Wien, T. Rusert and K. Hanke, "RWTH proposal for Scalable Video Coding Technology", Technical Report, ISO/IEC/JTC1/SC29/WG11/MPEG2004/M10569/S16, Munich, Germany, 2004.

[24] Y. Wu, "Fully Scalable Subband/Wavelet Video Coding System", PhD Dissertation, Department of Computer Science, Rensselaer Polytechnic Institute, 2005.

[25] Y. Wu, K. Hanke, T. Rusert and J.W. Woods, "Enhanced MC-EZBC Scalable Video Coder", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, No. 10, pp. 1432-1436, 2008.

[26] Ying Chen, Guizhong Liu and Juncai Yao, "An Improved 3D Wavelet-based Scalable Video Coding Codec for MC-EZBC", *Multimedia Tools and Applications*, Vol. 76, No. 6, pp. 7595-7632, 2017.