Fast Side Information Generation for High-Resolution Videos in Distributed Video Coding Applications

Shahzad Khursheed[1], Nasreen Badruddin[2], Varun Jeoti[3]

Department of Electrical and Electronic Engineering
Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia

Manzoor Ahmed Hashmani[4]

Department of Computer and Information Sciences
Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia

Abstract—**Distributed video coding (DVC) is an attractive and promising scheme that suits the constrained video applications, such as wireless sensor networks or wireless surveillance systems. In DVC, estimation of fast and consistent side information (S_i) is a critical issue for instant and real-time decoding. This issue becomes even more serious for high-resolution videos. Therefore, to minimise the side information estimation computational complexity, in this work, a computationally low complex DVC codec is proposed, which uses a simple phase interpolation (Phase-I) algorithm. It performs faster for all resolutions videos, and significant results are achieved for high-resolution videos with a large group of pictures (GOP). For the proposed technique, the computation time rapidly decreases with an increase in resolution. It performs 221% to 280% faster from conventional frame interpolation method for high-resolution videos and large GOP at the cost of little degradation in the visual quality of estimated side information.**

Keywords—*Fast side information algorithm; phase-based interpolation (Phase-I); DVC; DVC decoder for high-resolution videos; real-time DVC decoding; real-time side information*

I. INTRODUCTION

Wireless video sensor networks (WVSNs) are capable of capturing video at distributed video sensor nodes. Conventional video codecs are ill-suited for these nodes. Compression of the captured video has received significant interest in literature. The availability of high-resolution CMOS image sensors at low cost makes the WVSNs more trending [1], especially for real-time surveillance and environment monitoring [2, 3] and medical applications [4], etc. The new applications of WVSN are emerging very rapidly and demanding efficient pre-processing and transmission [5]. Due to challenges of being battery supported, there is a need for efficient use of storage resources and lower energy consumption in WVSNs [6]. Such video sensors, therefore, demand low encoding techniques to compress the video to the lower bit rate before storing or transmitting it [7] and to reduce the transmission delay [8]. One of the supportive coding approaches is distributed video coding (DVC) that redistributes the coding complexity in such a way that there is much low encoding computation [9] while decoding can be more complex [10].

In DVC codecs, the frames are organised in a group of pictures (GOP) of size 2, 4 or more. The key-frames are Intra-encoded first and then transmitted, and intermediate frames are WZ encoded. At the decoder, these WZ frames are estimated by the Intra-decoded key-frames. These DVC coding schemes achieved high compression while maintaining low encoding complexity by utilising the Intra-encoding at the encoder, and Inter-decoding (because the S_i estimation depends on key-frames) at the decoder [11]. The DVC encoder is deliberately kept computationally very simple but the decoder is computationally very complex since it needs to accurately estimate the replica of the WZ frames known as side information (S_i). The traditional S_i generation algorithms are computationally extensive due to the complex nature of the prediction process and take a lot of time even for a low-resolution video.

A. Motivation and Contribution

The S_i is estimated either by interpolation or extrapolation, and its quality determines the overall coding efficiency of codec [12]. Both prediction processes are considered to be time-consuming [13] activity of the decoding process. The huge computational complexity [14] is associated with these prediction processes, and it takes considerable time even for the low-resolution videos, and the decoding process slows down due to it [13]. The efforts were carried out mostly on low-resolution 1k-pixel and 176x144 pixels per frame videos [15] to improve the RD performance. The researchers are putting an effort to design the framework to achieve the low complex and real-time decoding for such low-resolution video [15] while achieving a consistent and high RD performance comparable with conventional codec.

As a low-cost standard definition (SD) and high definition (HD) mini video sensors [16] are widely available, so there is a need for real-time DVC decoding framework for such high-quality videos. However, no DVC framework for real-time or fast S_i generation is found in the literature for high-quality videos. Therefore, herein, an attempt is made to design a suitable DVC framework with low computational S_i generation algorithms for high-quality videos. In this work, a Phase interpolation (Phase-I) is incorporated for S_i generation in the DVC decoder.

The rest of the paper is organised as follows: Section II discusses the background, and Section III describes the proposed DVC model for fast S_i generation for high-resolution videos. In Section IV, results have been presented, and the performance of the proposed model in terms of computational complexity and quality analysis with (peak signal to noise ratio) PSNR is discussed. In Section V, the future directions are proposed. Finally, the conclusion is in Section VI.

II. BACKGROUND

The DVC [17] structure follows the Slepian Wolf [18] and Wyner-Ziv [19] theories proposed in the early 1970s. These theories proposed that the correlated sources can be independently encoded and jointly decoded. This way, they can still achieve the same rate as they are jointly encoded and decoded as long as the correlated S i̶ is available and used in the decoder side. In WVSN, the video sensor nodes acquire frames at some rate, and the consecutive frames are highly correlated. These correlated frames are independently encoded because the prediction loop is not involved in the DVC encoding process, and that is how the DVC provides codec-independent scalability.

DVC compression efficiency highly depends on S i̶ quality [20]. It is important to remember that high-quality S i̶ leads to better rate-distortion (RD) performance [21, 22] which plays a significant role in achieving low bit rate and less error correction [23], which are the main factors for the low latency optimal transmission [24]. However, high-quality S i̶ estimation is a difficult task, even for low-resolution videos. The WZ encoding and decoding are carried on either in the transform domain (TD) or the pixel domain (PD). Aaron et al. first time at Stanford proposed TD-based DVC framework [25]. In this framework, only the intra-frame statistical reliance is explored. It outperforms other codecs due to superior coding efficiency. Afterwards, the codec known as PRISM (Power-efficient, Robust, hIgh compression Syndrome based Multimedia coding) for TD was proposed by Puri et al. [26, 27]. Most of the adequate DVC codecs are found on the Stanford TD based framework. The DISCOVER [28] also complies with the Stanford framework [25].

Despite all the developments made by the DVC codecs [6, 10, 21, 29-31], consistent RD performance [32] is still an issue and, does not meet the superior performance of conventional codecs for all acute and non-uniform motion feature videos. This commonly happens due to the substandard quality of the WZ frame replica (known as S i̶) which is estimated by interpolation or extrapolation [33] at the decoder. The superior coding efficiency and even the low bit-rate are achieved by making use of a highly correlated estimated replica of WZF [34]. However, the S i̶ generation process consumes a lot of time due to the computationally complex prediction (motion estimation and compensation) activities [35]. These prediction activities are highlighted as a major source of high computational complexity at the decoder and cause latency in the decoding process [13] even for low-resolution videos. Moreover, the feedback channel for error-resilience also imposes the delay and increases the decoder complexity [36] due to the iterative requests for more bits that are required for error-correction. This, in turn, increases computation complexity and decreases the life of the video sensor because the transmission requires more resources as compared to other operations [37].

III. PROPOSED DVC MODEL WITH SIDE INFORMATION GENERATION SCHEME

In video processing, the vast amount of computation is involved in the estimation of the in-between frame due to prediction and estimation. Specifically, S i̶ estimation is the most challenging task in DVC decoding. The conventional interpolation methods require extensive computation resources and time; therefore, this extended computation complexity prolongs the decoding activity, especially for high-resolution videos.

Fig. 1 shows the proposed DVC Model for S i̶ generation with Phase-I and residual frame (Ɍ) calculation to reduce the transmission rate. The coding efficiency is associated with the transmission rate; therefore, to reduce the transmission rate, the Ɍ is calculated and encoded with WZ coding. The consecutive frames have the similarity and taking their difference will extract the motion part only which can be encoded in lesser bits than the actual WZ frame encoding, and WZ coding gives further lossy compression. The information will be encoded in fewer bits. The Ɍ can be calculated by (1).

$$Ɍ = W - K \tag{1}$$

In (1), the *W* is a current actual WZ frame, and *K* defines the previous key-frame.

Phase-I computes the pixel-wise phase modification without any extensive global optimisation and estimates pixel's motion by phase shifting of the individual pixels. In addition to this, the phase shift correction feature combines the phase information across all the levels of a multi-scale pyramid [38] in a very short time.

Avoiding the expensive global optimisation, which is a typical part of optical flow techniques, allows interpolating in-between frames in a fraction of time of the traditional interpolation methods. Therefore, in the DVC decoder, its deployment exclusively decreases S i̶ generation complexity and overall decoding time and complexity, even for the high-resolution videos.
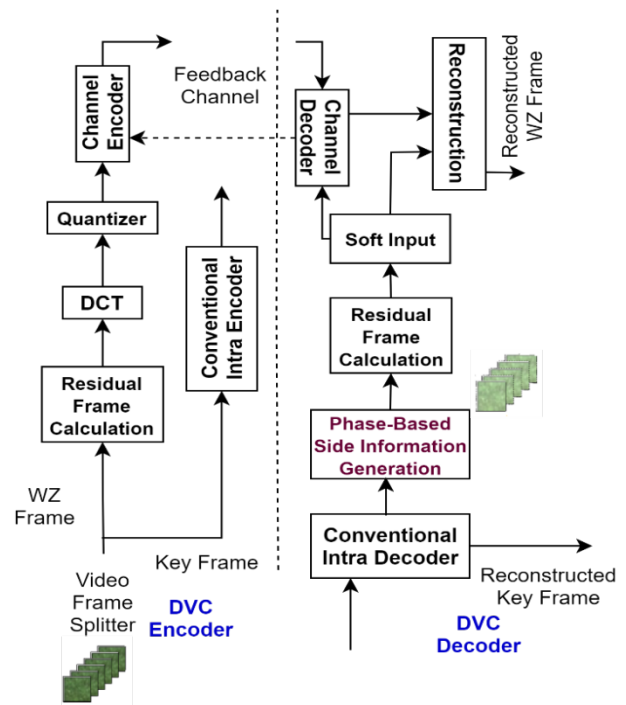

Fig. 1. Proposed DVC Model-Fast Side Information Generation for High-Resolution Videos.

The Phase-I algorithm summarises the execution steps of Phase-I. The input of Phase-I are two images and interpolation parameter α, and the process starts with steerable pyramid decompositions of both images and calculation of their amplitudes. The output of this algorithm is the interpolated image.

Phase-I Algorithm

Inputs: Two input images are: I_1 and I_2

Interpolated parameter: α

Initialisation: Steerable pyramid decompositions: P_1 and P_2

Amplitudes Calculation: A_1 and A_2

Output: Output (interpolated image) : I_α

Step1: $(P_1, P_2) \leftarrow$ Decompose (I_1, I_2)

Step2: $(A_1, A_2) \leftarrow$ Amplitude (I_1, I_2)

Step3: $(\phi_1, \phi_2) \leftarrow$ Phase (P_1, P_2)

Step4: $\phi_{diff} \leftarrow$ Phase Difference (ϕ_1, ϕ_2)

Step5: $for\ all\ l = L - 1\ do$

Step6: $\tilde{\phi}_{diff}^l \leftarrow$ Shift Correction $(\tilde{\phi}_{diff}^{l+1})$

Step7: $end\ for$

Step8: $\hat{\phi}_{diff} \leftarrow$ Adjust Phase $(\phi_{diff}, \tilde{\phi}_{diff})$

Step9: $\phi_\alpha \leftarrow$ Interpolate $(\phi_1, \hat{\phi}_{diff}, \alpha)$

Step10: $A_\alpha \leftarrow$ Blend (A_1, A_2, α)

Step11: $P_\alpha \leftarrow$ Recombine (ϕ_α, A_α)

Step12: $I_\alpha \leftarrow$ Reconstruct (P_α)

In Fig. 2, the flowchart defines the basic steps for S ι interpolation with Phase-I. All mathematical notations are available in [38]. The flowchart presents the step by step process for estimation of the S ι. The Phase-I first decomposes the input images into the steerable pyramids, which is a linear multi-scale, multi-orientation image decomposition and calculates the amplitudes as well [38]. The Phase-I approach has a few intuitive parameters. The main parameters mainly used to control the number of orientations and levels corresponding to the different scales of the steerable pyramid. Better motion separation is achieved with a higher number of levels and orientations. The parameters setting used for generation of S ι is as follows; the number of orientations used is 8, and the number of levels L is determined such that the coarsest level has a minimum width of 10 pixels. For the limitation factor, we use $\tau = 0.2$. The size of the coarsest level together with this choice of τ leads to a theoretical limit of motion which can be modelled reliably as 2% of the image width.

Depending on the size of GOP, both input images can either be intra-decoded key-frames or consists of the one intra-decoded key-frame and one previously estimated WZ frame, herein called S ι. In the next step, the phases are extracted with decomposed steerable pyramids. After the phase difference calculation, it is adjusted by the shift correction process. Now, for interpolating the next frame, the new phase is estimated with interpolation parameters, previous frame phase and the new calculated phase difference. Now, for the new S ι interpolation, a new amplitude is calculated by blending the interpolation parameter and extracted amplitudes of input frames. This new amplitude is combined with the new calculated phase to reconstruct the interpolated S ι. The focus of this work is only on the fast S ι generation for high-resolution videos; therefore, the full performance of codec will be presented in future work.
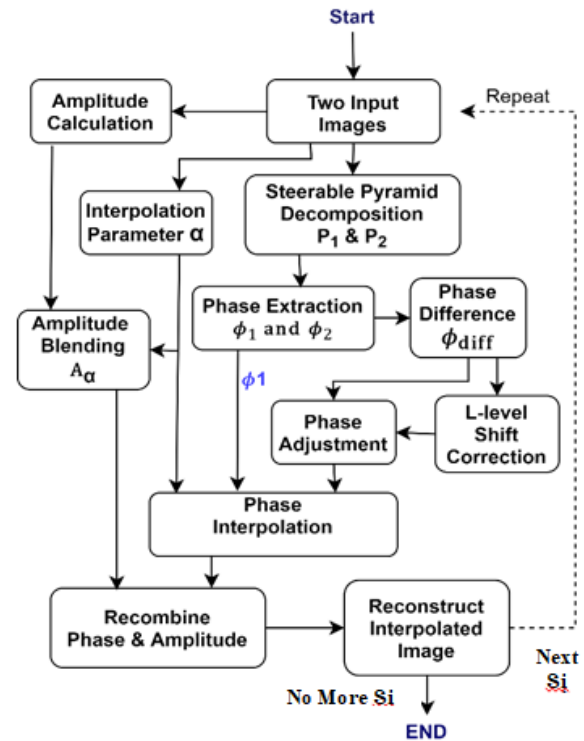

Fig. 2. Flowchart of Phase Interpolation (Phase-I) Algorithm Process for Side Information Generation.

IV. RESULT EVALUATION AND DISCUSSION

The test clips WashDc in .cif (288x352) and .sif (480x640) formats, Mobile_Claneder in NTSC (486x720) and Old_Town Cross HD (720x1280) are taken for experiments [39]. The purpose of taking the different clips with different resolutions was to evaluate the performance of traditional and proposed Phase-I S ι generation approaches for different resolutions. The S ι quality and computational performance of its generation algorithm, are evaluated with quality metric PSNR and computational complexity metric execution time, respectively. To fairly evaluate the computational or coding complexity, the data cache size, memory access bandwidth, instruction cache size, storage complexity, execution time, parallelism and pipelining, all these dimensions should be measured [40]. However, practically it is difficult to measure all these dimensions [41]. Therefore, the coding time on a computing platform is usually considered for measuring the computational complexity as it is relatively easy to measure. It not only shows the computational complexity but also

partially indicates the effects of other dimensions such as memory access, etc. in the coding process [41]. Consequently, this is a widely used metric for complexity measure. Therefore, for convenience, this is also used in this paper for coding complexity measures.

The number of calculations performed during a specific task defines the computational complexity. The total time of processor usage is directly affected by the number of calculations performed for a task; therefore, the computational complexity is always assessed and presented in processing time [42]. The performance of S_i generation algorithms is measured and compared with each other for the same video but with different high resolutions.

The performance is measured for two different GOP's of sizes 2 and 4. When GOP=2, the frames sequence will be the $I_1 W_1 I_2 W_2 I_3 W_3 I_4$, where $I_1 I_2 I_4$ are intra-encoded and decoded key-frames, and $W_1 W_2 W_3$ the Wyner-Ziv frames whose estimated replicas are called S_i. When GOP=4, the frames sequence will be the $I_1 W_1 W_2 W_3 I_2$, where $I_1 I_2$ are the key-frames, while W_1, W_2, W_3 are the Wyner-Ziv frames. In GOP=4, first, the S_{i_2} is estimated with I_1 and I_2, then S_{i_1} is estimated with I_1 and S_{i_2} and, S_{i_3} estimated with S_{i_2} and I_2.

A. Computational Complexity Measure

The simulations are carried out on the Core(TM) i7-7820HQ, CPU 2.90GHz with 64-bit OS, and RAM 32 GB. The computation complexity of S_i algorithms; conventional interpolation called motion-compensated temporal interpolation (MCTI) and Phase-I for test sequences of 288x352 and 480x640 formats are presented in Fig. 3 and 4, respectively. The computation is measured for a single GOP of size 2 and 4 in different time slots of a test sequence. Time for single GOP of 2 and 4, is measured after every 30 frames.
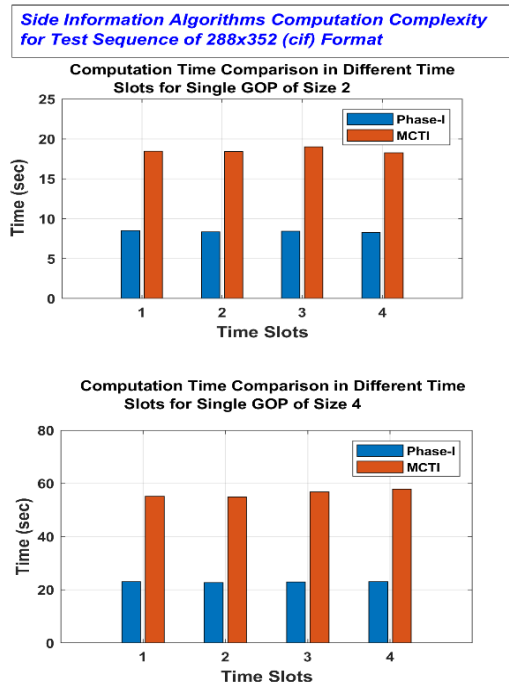

Fig. 3. Side Information Algorithm Computation Complexity for Test Sequence of 288x352 (cif) Format for GOP 2 and 4.
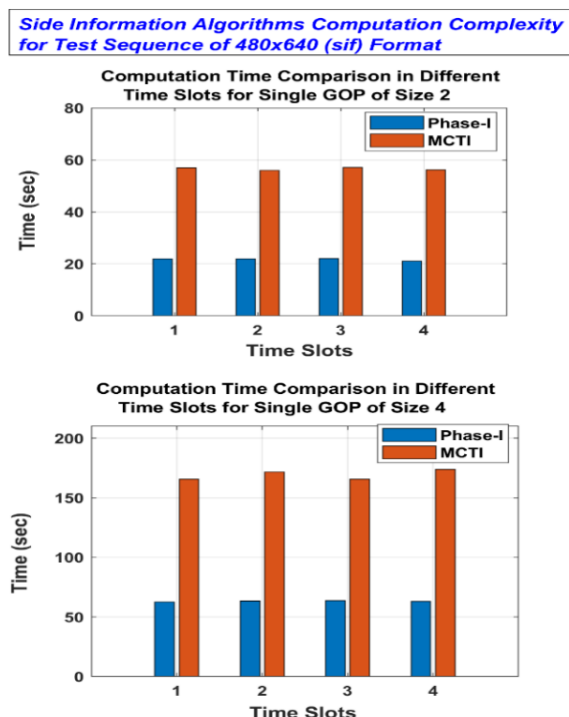

Fig. 4. Side Information Algorithm Computation Complexity for Test Sequence of 480x640 (sif) Format for GOP 2 and 4.

For the 288x352 format test sequence, the average S_i estimation time taken by Phase-I is 8.38 sec, and MCTI is 18.54 sec for GOP 2. The Phase-I took an average of 22.91 sec, and MCTI an average of 56.19 sec for single GOP of size 4. Simulation results show that Phase-I is an average of 221% and 245% faster from the MCTI for GOP 2 and GOP 4, respectively. For the 480x640 format test sequence, the average S_i estimation time taken by Phase-I is 21.74 sec, and MCTI is 56.54 sec for GOP 2. The Phase-I took an average of 63.155 sec, and MCTI an average of 169.23 sec for single GOP of size 4. Simulation results show that Phase-I is an average of 260% and 268% faster from the MCTI for GOP 2 and GOP 4, respectively.

Computationally, the Phase-I is much faster than MCTI for all resolution videos and delivers optimum performance for different GOP sizes. This reduces the overall decoding complexity and hence leads to the faster decoding for high resolution of videos.

B. PSNR Performance and Discussion

The PSNR is conceived as one of the image quality measuring metric. It reflects the quality of estimation frame relative to the actual frame. Fig. 5 presents the PSNR performance as a function of frame numbers in a given sequence of frames. The comparison is made among the MCTI method and Phase-I for the GOP size of 2 (only PSNR of estimated S_i) and PSNR of relevant intra-decoded frames for both .cif and .sif formats.

The simulation results implemented for .cif format points out the average PSNR of 35.5dB and 31.2 dB for MCTI and Phase-I, respectively. The implementation results of .sif

format showed the average the PSNR 35 dB and 31 dB for MCTI and Phase-I, respectively. The proposed approach performance is approximately 4-4.3 dB poorer than that of MCTI for both formats. Although, the Phase-I lag behind the MCTI by 4.2 dB but deliver consistent performance throughout the sequence. The proposed method's efficiency falls by 11.26% of MCTI. However, the 30dB PSNR is considered as the minimum acceptable quality for a human vision [43].

Fig. 6 presents the PSNR performance graph for the MCTI and Phase-I for a GOP size of 4 (only PSNR of three estimated S_i of each GOP) and PSNR of relevant intra-decoded frames for both .cif and .sif formats. The simulation results implemented for .cif format pointed out the average PSNR of 33.24 dB and 27.95 dB for MCTI and Phase-I, respectively. The implementation results of .sif format presented the average PSNR 32.34 dB and 30.8 dB for MCTI and Phase-I, respectively. The MCTI achieved average 4.29 dB better PSNR from Phase-I but inconsistent performance for each GOP. While MCTI performance degraded for .sif format, whereas Phase-I performance improved for .sif format. Therefore, the Phase-I has great potential to deliver better for high-resolution videos with a large GOP size.

The computational complexity and quality performance evaluation of both S_i algorithms are presented in Table I. The computational complexity and quality evaluation are in the form of average time and average peak-to-signal ratio (PSNR) of S_i of the GOP respectively.

Simulation tests were also carried out on several other videos of different formats and motions, and few of them are also listed in Table I. Simulation conditions are frequently changed to analyse the visual quality and computational complexity of the proposed approach. The Phase-I based S_i visual quality is dependent on the quality of intra-decoded frames. With high visual quality intra-decoded frames, the Phase-I algorithm generates better S_i and vice versa. On the other hand, with one high quality and another low-quality intra-decoded frame, the S_i quality degraded according to the low one.

In the current scenario of GOP 4, the reference frames are changed according to the mentioned methodology, and visual quality of every estimated S_i frame varies accordingly. However, an attempt is made to generate more than one S_i frames by keeping the same reference frames, a bit low but almost consistent visual quality S_i's are generated in minimal time. In that approach, the computation time was 2-2.5 times less than the currently implemented strategy at the cost of a small degradation in the visual quality of all S_i's. This will open the door for implementation of low-delay or real-time S_i generation with large GOP size. However, the study is required to analyse its effect on transmission rate in the channel decoding step because further correction in S_i should be needed to achieve consistent high visual quality in video.

The visual performance slightly depends on some parameters of Phase-I algorithm like decomposition level of steerable pyramids, phase extraction step, phase shift correction step and amplitude calculation step when only one

S_i is computed. However, the computation time rarely changed with the change of these parameters.
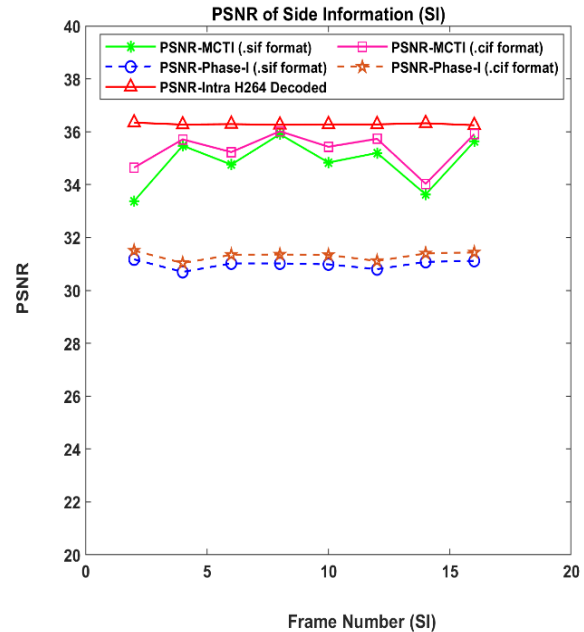
Fig. 5.   PSNR Performance Measure of Phase-I and MCTI for .cif(288x352) and .sif(480x640) for GOP=2.
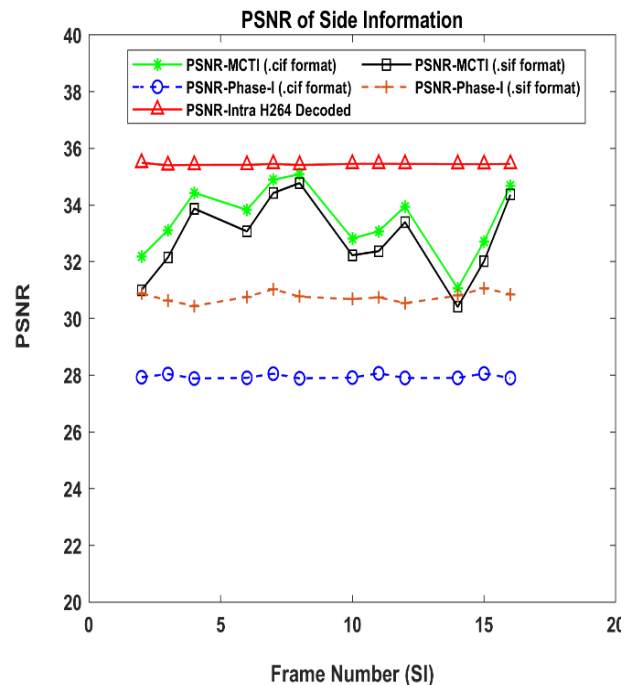

Fig. 6.   PSNR Performance Measure of Phase-I and MCTI for .cif(288x352) and .sif(480x640) for GOP=4.

TABLE I.   COMPUTATIONAL COMPLEXITY (IN AVERAGE TIME) AND QUALITY (IN AVERAGE PSNR) FOR SINGLE GOP OF SIZE 2 AND 4

Test Sequence Formats	Computational Complexity (Average Time (sec))				Quality Analysis-Average PSNR (dB)			
	GOP= 2		GOP=4		GOP=2		GOP=4	
	MCTI	Phase-I	MCTI	Phase-I	MCTI	Phase-I	MCTI	Phase-I
WashDc.cif (288x352)	18.54	8.38	56.19	22.91	34.64	31.50	33.24	31.01
WashDc.cif (480x640)	56.54	21.74	169.23	63.16	33.46	31.17	32.34	30.65
Mobile_Claneder.NTSC (486x720)	64.79	25.11	187.45	71.54	35.15	30.98	34.78	30.10
Old_Town Cross (720x1280)	147.31	52.17	386.30	139.13	34.90	31.20	34.43	30.34

V.   FUTURE DIRECTION

In current Phase-I algorithm, the visual quality changes with reference frames (frames that are used for S_i estimation) quality. Especially the performance of algorithm degrades when one of the reference frames is of low quality. Therefore, the focus can be put to design this algorithm in a way that it changes performance with respect to high-quality reference frames to get the S_i with consistent visual quality in either condition. Getting a consistent high-quality S_i will also assist in reducing the number of bits (bit rate) which are required for S_i correction in the channel decoding step. It reduces the transmission rate efficiently for both low and high-resolution video.

In large GOP size, the visual quality almost remains the same for all estimated S_i when the same reference frames are used to estimate all the intermediate S_i frames. This method of generating the intermediate S_i frames makes the Phase-I computationally very effective, but it lack-behind in visual performance. Visual quality remains a bit low from other adopted approach. Therefore, the algorithm should be designed in a way that it estimates more than one high-quality intermediate frames at once in a very short time. If this less computational complexity method estimates the consistently high-quality S_i, then it will be effective to achieve the low-delay or real-time DVC decoding for both low and high-resolution videos. Moreover, it also reduces the transmission rate efficiently for both low and high-resolution videos with large GOP size.

The visual performance slightly depends on some parameters of Phase-I algorithm like decomposition level of steerable pyramids, phase extraction step, phase shift correction step and amplitude calculation step when only one S_i is computed. But these parameters somehow put their effect when two reference frames are far away from each other and the computation time rarely affected by changing these parameters. Designing an adaptive Phase-I will be a productive step for auto-selection of these parameters, to generate a consistent high-quality S_i in small computational time and transmission rate for video applications can be controlled with it. Along with the high-quality S_i estimation in small time, the proposed residual frame calculation at the encoder further reduces the transmission rate and improves the codec coding efficiency.

VI.   CONCLUSION

The DVC decoder faces the computational complexity while estimating the replica of the WZ frame known as side information (S_i) due to the involvement of the prediction process. The traditional S_i generation algorithms raise a high computation complexity in decoding process because of the complex and composite prediction process and even took a long time for low-resolution video. However, the emergence of high-resolution video sensor demands high-speed DVC decoder with faster S_i generation algorithms. This research work proposed the DVC model with the Phase interpolation (Phase-I) algorithm for S_i estimation. It computes the pixel-wise phase modification without any explicit correspondence estimation and pixel's motion by phase shifting of the individual pixels. In addition to this, the phase shift correction feature combines the phase information across the levels of a multi-scale pyramid in very little time. Therefore, in the DVC decoder, its deployment exclusively decreases S_i generation complexity and overall decoding time and complexity, even for the high-resolution videos. It works efficiently and even better for high-resolution videos with large GOP. It exhibits low computation complexity for both low and high-resolution videos. Moreover, it delivers significant efficiency in the computation for different GOP sizes at the cost of some degradation in the quality of estimated S_i. However, for high-resolution video with a GOP size of 4, the results were acknowledgeable because the performance of the traditional algorithm drops out rapidly, and where on the other hand, Phase-I remains stable.

REFERENCE

[1] M. Ahmad, I. Ahmed, K. Ullah, I. Khan, A. Khattak, and A. Adnan, "Energy efficient camera solution for video surveillance," International Journal of Advanced Computer Science and Applications, vol. 10, no. 3, 2019.

[2] K. Sehairi, C. Benbouchama, K. El Houari, and C. Fatima, "A Real-Time Implementation of Moving Object Action Recognition System Based on Motion Analysis," Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 5, no. 1, pp. 44-58, 2017.

[3] S. M. Jameel, A. R. Gilal, S. S. H. Rizvi, M. Rehman, and M. A. Hashmani, "Practical Implications and Challenges of Multispectral Image Analysis," in 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2020: IEEE, pp. 1-5.

[4] M. F. Ukrit and G. Suresh, "Super-Spatial Structure Prediction Compression of Medical," Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 4, no. 2, pp. 126-133, 2016.

[5] D. Xiao, M. Li, M. Wang, J. Liang, and R. Liu, "Low-cost and high-efficiency privacy-protection scheme for distributed compressive video sensing in wireless multimedia sensor networks," Journal of Network and Computer Applications, p. 102654, 2020.

[6] J. Yang, L. Qing, X. He, L. Hua, and S. Rong, "A Fast DVC to HEVC Transcoding for Mobile Video Communication," in 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2019: IEEE, pp. 505-509.

[7] A. Mahmood, L. S. Dooley, and P. Wong, "Transform domain distributed video coding using larger transform blocks," in 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2017: IEEE, pp. 1-5.

[8] A. Veeraputhiran and R. Sankararajan, "Feature based fall detection system for elders using compressed sensing in WVSN," Wireless Networks, vol. 25, no. 1, pp. 287-301, 2019.

[9] X. H. Van, J. Ascenso, and F. Pereira, "HEVC backward compatible scalability: A low encoding complexity distributed video coding based approach," Signal Processing: Image Communication, vol. 33, pp. 51-70, 2015.

[10] D. Jun, "Distributed video coding with adaptive two-step side information generation for smart and interactive media," Displays, vol. 59, pp. 21-27, 2019.

[11] C. Brites and F. Pereira, "Distributed video coding: Assessing the HEVC upgrade," Signal Processing: Image Communication, vol. 32, pp. 81-105, 2015.

[12] Y. Kuo, P. Gao, and J. Chen, "Distributed video coding with limited feedback requests," Multimedia Tools and Applications, vol. 75, no. 4, pp. 2051-2065, 2016.

[13] S. M. Venkatrav, "H. 264 Standard Based Side Information Generation In Wyner-ziv Coding," 2010.

[14] S. Acharjee, N. Dey, D. Biswas, P. Das, and S. S. Chaudhuri, "A novel Block Matching Algorithmic Approach with smaller block size for motion vector estimation in video compression," in 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), 2012: IEEE, pp. 668-672.

[15] J. Hanca, N. Deligiannis, and A. Munteanu, "Real-time distributed video coding for 1K-pixel visual sensor networks," Journal of Electronic Imaging, vol. 25, no. 4, p. 041008, 2016.

[16] X. Ren and Z. Yang, "Research on the key issue in video sensor network," in 2010 3rd International Conference on Computer Science and Information Technology, 2010, vol. 7: IEEE, pp. 423-426.

[17] B. Girod, A. Aaron, S. Rane, and D. Monedero, "Distributed video coding," in Proceedings of IEEE Special Issue on Advances in Video Coding and Delivery, 2003, vol. 93, no. 1: Citeseer, pp. 1-12.

[18] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions on information Theory, vol. 19, no. 4, pp. 471-480, 1973.

[19] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Transactions on information Theory, vol. 22, no. 1, pp. 1-10, 1976.

[20] H. Yang, L. Qing, X. He, X. Ou, and X. Liu, "Robust distributed video coding for wireless multimedia sensor networks," Multimedia Tools and Applications, vol. 77, no. 4, pp. 4453-4475, 2018.

[21] H. Yang and J. Yang, "DVC Decoding Algorithm Based on Wireless Transmission," in 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2019: IEEE, pp. 261-265.

[22] S. Zheng, X.-P. Zhang, J. Chen, and Y. Kuo, "A New Multihypothesis Prediction Scheme for Compressed Video Sensing Reconstruction," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020: IEEE, pp. 4337-4341.

[23] R. Chen, Y. Tong, J. Yang, and M. Wu, "Residual reconstruction algorithm based on sub-pixel multi-hypothesis prediction for distributed compressive video sensing," in Conference on Complex, Intelligent, and Software Intensive Systems, 2018: Springer, pp. 599-605.

[24] L. Van Hoesel, T. Nieberg, J. Wu, and P. J. Havinga, "Prolonging the lifetime of wireless sensor networks by cross-layer interaction," IEEE Wireless Communications, vol. 11, no. 6, pp. 78-86, 2004.

[25] A. Aaron, S. D. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in Visual Communications and Image Processing 2004, 2004, vol. 5308: International Society for Optics and Photonics, pp. 520-528.

[26] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in Proceedings of the annual allerton conference on communication control and computing, 2002, vol. 40, no. 1: Citeseer, pp. 586-595.

[27] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," IEEE transactions on image processing, vol. 16, no. 10, pp. 2436-2448, 2007.

[28] R. Halloush, K. Misra, and H. Radha, "Practical distributed video coding over visual sensors," in 2009 Picture Coding Symposium, 2009: IEEE, pp. 1-4.

[29] J. Zhou, Y. Fu, Y. Yang, and A. T. Ho, "Distributed video coding using interval overlapped arithmetic coding," Signal Processing: Image Communication, vol. 76, pp. 118-124, 2019.

[30] Y.-C. Shen, H.-P. Cheng, J.-C. Luo, Y.-H. Lin, and J.-L. Wu, "Efficient real-time distributed video coding by parallel progressive side information regeneration," IEEE Sensors Journal, vol. 17, no. 6, pp. 1872-1883, 2017.

[31] Y. Cao, L. Sun, C. Han, and J. Guo, "Improved side information generation algorithm based on naive Bayesian theory for distributed video coding," IET Image Processing, vol. 12, no. 3, pp. 354-360, 2017.

[32] P. C. BURILEANU, "Navigarea Autonomă utilizând Codarea Video Distribuită Free Navigation with Distributed Video Coding (DVC)," University "Politehnica" of Bucharest, 2016.

[33] Z. Gong and Z. Yang, "Video frame interpolation and extrapolation," ed: Stanford University, Tech. Rep., 2017.[Online]. Available: http://cs231n …, 2017.

[34] B. Dash, S. Rup, A. Mohapatra, B. Majhi, and M. Swamy, "Decoder driven side information generation using ensemble of MLP networks for distributed video coding," Multimedia Tools and Applications, vol. 77, no. 12, pp. 15221-15250, 2018.

[35] W. Bao, W.-S. Lai, X. Zhang, Z. Gao, and M.-H. Yang, "Memc-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement," IEEE transactions on pattern analysis and machine intelligence, 2019.

[36] M. Nikzad, A. Bohlooli, and K. Jamshidi, "Video quality analysis of distributed video coding in wireless multimedia sensor networks," International Journal of Information Technology & Computer Science, vol. 7, no. 1, pp. 12-20, 2014.

[37] X. Li, M. Wien, and J.-R. Ohm, "Rate-complexity-distortion optimisation for hybrid video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 7, pp. 957-970, 2011.

[38] S. Meyer, O. Wang, H. Zimmer, M. Grosse, and A. Sorkine-Hornung, "Phase-based frame interpolation for video," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1410-1418.

[39] "Xiph.org Video Test Media [derf's collection]." https://media.xiph.org/video/derf/ (accessed.

[40] M. Horowitz, "Toward useful complexity evaluation methods," in Proceedings of VCEG Meeting ITU-T SG16 Q. 6, 33rd Meeting, 2007.

[41] X. Li, M. Wien, and J.-R. Ohm, "Rate-complexity-distortion evaluation for hybrid video coding," in 2010 IEEE International Conference on Multimedia and Expo, 2010: IEEE, pp. 685-690.

[42] G. Correa, "Computational Complexity Reduction and Scaling for High Efficiency Video Encoders," 2015.

[43] D. Younis Rasool Wani, S. Sharma, and S. Chopra, "Video Quality Evaluation of MPEG-4 Using (MOS) Mean Opinion Score in NS-2," 2017.