2018 Second International Conference on Advances in Electronics, Computer and Communications (ICAECC-2018)

Transcoding of Next Generation Distributed Video Codec for Mobile Video

Vijay Kumar Kodavalla
VLSI Technology Practice Group, Product Engineering Services
Wipro Technologies
Bangalore, India
vijay.kodavalla@wipro.com

Abstract— **Presently most popular and widely adopted video coding standards include MPEGx, H.26x, DivX, AVSx and VPx. These standards are based on predictive video coding method, where encoder is designed to be more complex (about 10x) than the decoder. These predictive video coding standards are suitable for existing applications, including broadcast and video-on-demand. However, there are other evolving applications, including wireless surveillance, multi-view, one-time-use video cameras and mobile video, which demands encoder complexity being lower. This is in contrast to the existing video coding standards. To cater to these evolving applications, "Distributed Video Coding" (DVC) has progressed as an inventive solution. The uniqueness of distributed video coding solution is its deployment of channel coding methods into the source-coding problem, in reducing encoder complexity, even at the expense of increase in complexity of the decoder. Nevertheless, for a sub-set of evolving applications (for e.g., mobile video), both encoder and decoder need to be of lower complexity. For this video conference application, even the DVC is also not suitable. This paper demonstrates a method for achieving low complexity encoder and decoder from the end-user point of view, by using transcoding of DVC and H.264 Advanced Video Codec (AVC) in the network. The elaborate research and proof-of-concept models developed, over a period of several years, have become the basis for this paper.**

Keywords— DSC; DVC; WZ; GOP; AVC; MPEGx; H.26x; VPx; AVSx; DivX; QP; RD; RALDPC; DCT; PSNR

I. INTRODUCTION

In 2002, Stanford University has proposed DVC first model [1], [2]. This DVC model was built on information theory outcomes of "Slepian-Wolf" and "Wyner-Ziv" (WZ) "Distributed Source Coding" (DSC). Further, in 2004, Stanford University has proposed another transform-domain DVC model [3], which is called as Stanford model. The Berkeley University has also proposed other DVC models such as "DIstributed Source Coding Using Syndromes" (DISCUS) [4] and "Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding" (PRISM) [5], [6] in 2003 and 2007 respectively. Further, significant improvements in DVC model have been carried out in "DIStributed COding for Video SERvices" (DISCOVER) [7], [8].

The DVC idea combines channel-coding methods in the source-coding problem. In DVC encoder, the incoming video pictures or frames are grouped and each group is called "Group of Pictures" (GOP). The first picture or frame in a GOP is named as a key-frame and the remaining pictures or frames are named as "Wyner-Ziv" WZ-frames. The key-frame is encoded with intra-mode (encoding with-in a frame) of predictive video coding method. As key-frames will not go through inter-mode coding (encoding among the frames) of the predictive video coding method, motion estimation/compensation is not necessary in the DVC encoding. Further, the remaining WZ-frames in each GOP are not encoded with the predictive video coding method. Instead, only error correction codes of WZ-frames are sent to the decoder. The bit-rate of such error correction codes is very much lesser than entropy. Moreover, in the decoder, predictive intra-mode (decoding with-in a frame) decoding method is used for key-frame decoding. In addition, the remaining WZ-frames are predicted in the decoder, with formerly decoded key-frames and/or WZ-frames. The WZ-frames, which are predicted in the decoder, shall be treated as a noisy or distorted form of WZ-frames available in the encoder. Moreover, the error correction codes of WZ-frames, which are transmitted from the encoder, are used to correct the errors in the predicted WZ-frames of the decoder. With this method, a low complex encoder is achieved, which is DVC crucial objective. However, this leads to a more complex decoder, as the DVC decoder is based on the iterative soft input based channel decoding. In addition, in predictive video coding, encoder is more complex and the decoder is of very less complex, by design. However, for applications like the mobile video conference, both encoder and decoder needs to be of lower complexity, as compute resources and battery power are scarce. This is because both encoder and decoder will be there in both transmitter and receiver mobile devices in the video conference system. For this application, neither predictive video coding methods nor DVC alone is suitable. The transcoding of DVC and H.264/AVC in the network, is proposed, for solving this problem. The network can be a central station or could etc., where complexity has been pushed to.

The DVC proof-of-concept models [9], [10], [11], [12], [13], [14] have been developed, to assess the potential and suitability for practical applications. Also in this paper, transcoding of DVC and H.264/AVC is proposed, to achieve low complex encoder/decoder, which is suitable for mobile video conference applications. The overview of one of the developed DVC models is included in Section II, on which further extension is done for transcoding. The Section III details proposed DVC to H.264/AVC transcoding method along with the results and the Section IV has conclusion.

II. DEVELOPED DVC PROOF-OF-CONCEPT BASELINE-MODEL

The DVC proof-of-concept baseline-model has been developed, which is shown in Fig. 1, whose details are given in other paper [11] of the author. The DVC encoder complexity is intentionally kept very low, to meet the prime objective. The inbound video pictures or frames are divided as GOPs, with various simple histogram based motion metrics.

The key-frame of every GOP is encoded with H.264-Intra, which is predictive intra-mode encoder. Whereas, WZ-frame goes through block based "Discrete Cosine Transform" (DCT) first. The resultant DCT coefficients are grouped in forming the DCT bands. Further, DCT bands thus formed goes through scalar quantization (uniform) and bit-plane ordering. The bit-planes thus formed are encoded individually with "Rate Adaptive Low Density Parity Check" (RALDPC) encoder. In addition, RALDPC encoder generates a set of parity bits per encoded bit-plane, which denotes their accumulated syndromes. The parity bits thus generated are buffered in the encoder buffer and transmitted to the decoder, in chunks, based on its requests.

The decoder of DVC is more complex, compared to the predictive video decoder, in the process of simplifying its encoder. The H.264 intra-mode (H.264-Intra) decoder is used for decoding the key-frame of each GOP. In addition, the decoder utilizes formerly decoded WZ-frames and/or key-frames in predicting "Side Information" (SI), which is treated as the noisy or distorted version of WZ-frame available at the encoder. The SI is generated by applying "motion compensated temporal interpolation" (MCTI), among formerly decoded WZ-frames and/or key-frames. A Laplacian noise model is deployed to model variations between WZ-frame available in the encoder and predicted SI in the decoder. Then, DCT is applied on generated SI for calculating the WZ-frame DCT coefficients. The soft-input values of the information bits are calculated from WZ-frame DCT coefficients, by applying an online Laplacian noise model. Further, each DCT coefficient conditional probability is converted to conditional bit probabilities, by using predicted SI and/or formerly decoded bit-planes, for RALDPC decoding. If RALDPC decoding is not successful (by means of CRC check), an additional chunk(s) of parity bits are requested by the decoder from the encoder, through the feedback channel. This process of requesting the additional chunk(s) of parity bits is an iterative process and continues until decoding is successful. The bit-plane reverse ordering is done, once all the bit-planes are successfully decoded. Subsequently, inverse quantization and reconstruction is performed after the reverse ordering of the bit-planes. Next, inverse DCT is applied on WZ-frames, for converting from DCT coefficients to pixel domain.

It is evident that the developed DVC model has achieved the objective of low complex encoder, though decoder complexity is high. The transcoding of DVC and H.264/AVC is being proposed as a solution for low complexity encoder as well as the decoder. The transcoding of DVC details and its performance evaluation results are presented in Section III. For this performance evaluation purposes, various video test sequences are used with varying motion-activity, all the way from low to significantly-high. In addition, the parameters evaluated includes "Rate Distortion" (RD) (Peak Signal to Noise Ratio (PSNR) & bit-rate), the compression ratio and encoder/decoder complexities.

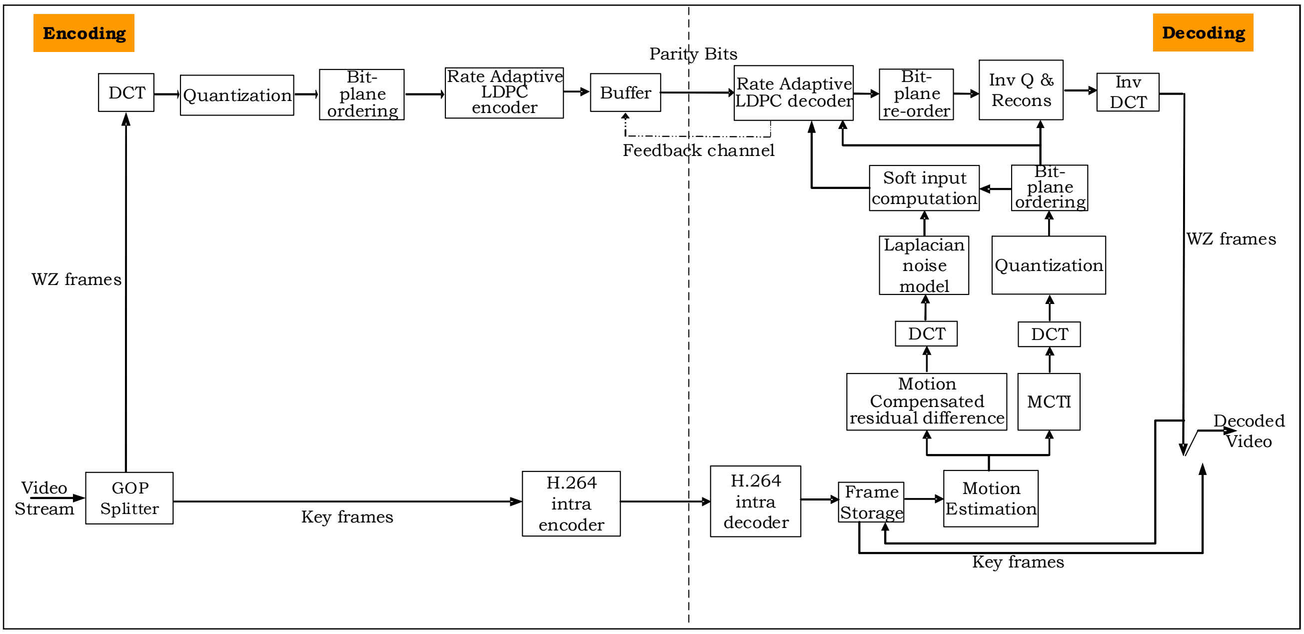

Figure 1. DVC baseline-model block diagram.

III. PROPOSED TRANSCODING OF DVC

The proposed transcoding of DVC for mobile video is as shown in Fig. 2. The context of mobile video here is from a mobile video conference scenario, between a pair of camera phones. Both mobile phones in video conference shall encode local video and send to the other end, and at the same time shall receive the encoded video from the other end and decode. Hence, it is necessary that both encoder and decoder shall be of lower complexity. This is because in mobile devices, compute-resources and battery power is usually scarce. The decoder in predictive video coding is of very less complex by design, whereas encoder is of more complex. On the contrary, as described, encoder in DVC is of very less complex by design, whereas decoder is of much more complex. Hence, neither predictive video coding nor DVC alone is suitable for this video conference application. Ideally, a combination of DVC encoder and the predictive video coding decoder will best suit for this video conference application. This can be achieved by transcoding from DVC to predictive video coding (for e.g., H.264/AVC) in the network, such as base station or cloud. Each mobile will use the DVC encoder and the H.264 decoder. In addition, the base station or cloud will have the DVC decoder and the H.264 encoder. Hence, the complexity is pushed into a base station or cloud in this case. As a base station or cloud is not scarce on either compute-resources or battery power, pushing complexity into them is practical.

Various steps in transcoding from DVC to H.264:

(a) Video conference mobile device-1 DVC encodes captured video and transmits → In the network, DVC encoded streams are decoded and re-encoded to H.264 and transmitted to video conference mobile device-2 → Video conference mobile device-2 decodes H.264 encoded stream

(b) Video conference mobile device-2 DVC encodes captured video and transmits → In the network, DVC encoded streams are decoded and re-encoded to H.264 and transmitted to video conference mobile device-1 → Video conference mobile device-1 decodes H.264 encoded stream

Thus, mobile devices on both ends will need to have only low complex DVC encoder and H.264 decoder. This can be extended to conference applications with several users. Further, the complex functions of DVC decoder and H.264 encoder are pushed to the network (base station or cloud).

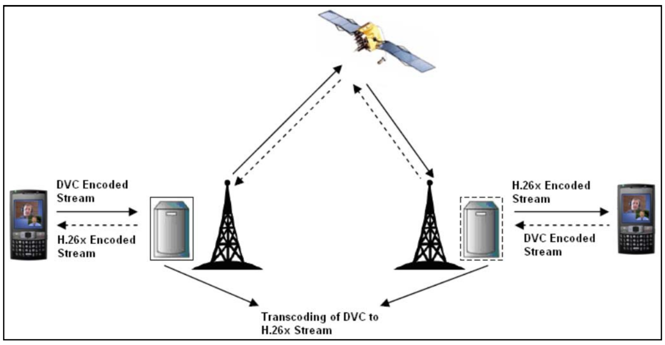

Figure 2. Proposed DVC-transcoding block diagram.

The developed DVC-transcoding proof-of-concept model is comprehensively assessed using varying motion-activity test video-sequences ("Hall-monitor", "Coast-guard", "Foreman" and "Soccer"). The test-sequences "Hall-monitor", "Coast-guard", "Foreman" and "Soccer" has low-to-medium, medium-to-high, very high and significantly high motion-activity respectively. The resolution of the test video-sequences is QCIF (176x144) with a frame rate of 15Hz and YUV 420 format. The total number of frames in these sequences is 150. Though these test video sequences are chosen for evaluation, proposed transcoding will work for any video sequences. A thorough performance evaluation of DVC-transcoding model has been carried-out and compared with H.264/AVC encoder/decoder. The performance parameters evaluated comprises of rate distortion (PSNR & bit-rates), compression ratio and encoder/decoder complexities. The DVC-transcoding is compared with DVC, H.264-Intra and H.264/AVC. This is to assess the performance of DVC-transcoding model compared to usage of DVC, H.264-Intra and H.264/AVC. The H.264-Intra uses only intra mode of coding. Further, H.264/AVC uses both intra and inter modes of the coding.

A. Bit-rate vs. PSNR - RD Performance

The rate distortion (RD) performance results of various test video-sequences are presented in Fig. 3 to Fig. 6. This is measured from one mobile device to another through, The DVC encoder in mobile device-1 → The DVC decoder in the network → The H.264/AVC encoder in the network → The H.264/AVC decoder in mobile device-2.

As shown in Fig. 3, PSNR achieved with developed DVC baseline-model with "Hall-monitor" sequence is up-to 3.6dB higher than H.264-Intra, for a specified bit-rate. Moreover, the DVC-transcoding model developed has achieved up-to 0.4dB PSNR lesser than that of DVC baseline-model, for a given bit-rate.

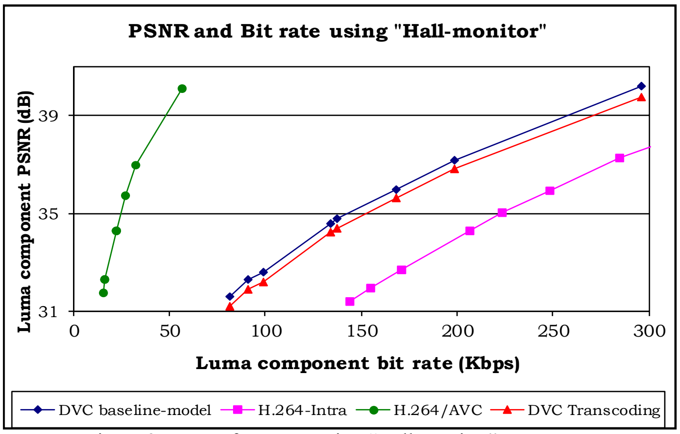

Figure 3. RD performance using "Hall-monitor" sequence.

As shown in Fig. 4, PSNR achieved with developed DVC baseline-model using "Coast-guard" sequence is up-to 2dB higher than H.264-Intra, for a specified bit-rate. In addition, the DVC-transcoding model developed has achieved up-to 1.3dB PSNR lesser than that of DVC baseline-model, for a given bit-rate.

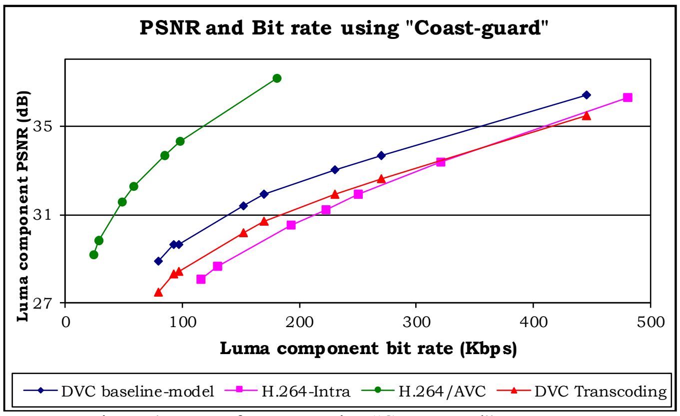

Figure 4. RD performance using "Coast-guard" sequence.

Further, as shown in Fig. 5, PSNR achieved with developed DVC baseline-model using "Foreman" sequence is up-to (-0.8)/(+1.2)dB lower/higher than H.264-Intra, for a specified bit-rate. The PSNR of DVC baseline-model is greater/lesser than H.264-Intra for low/high QP (Quantization Parameter) respectively. Also, as shown DVC-transcoding model developed has achieved up-to 1dB PSNR lesser than that of DVC baseline-model, for a given bit-rate. Moreover, as shown in Fig. 6, PSNR achieved with developed DVC baseline-model using "Soccer" sequence is up-to 2.8dB lesser than H.264-Intra, for a specified bit-rate. Also, as shown DVC-transcoding model developed has achieved up-to 1.1dB PSNR lesser than that of DVC baseline-model, for a specified bit-rate.

In summary, from the RD performance standpoint, both DVC baseline-model and DVC-transcoding are performing better than H.264-Intra, with video-sequences of up-to high motion-activity. Further, with very-high motion-activity video-sequences, DVC baseline-model is at par with H.264-Intra, though DVC-transcoding is lagging. In addition, with significantly-high motion-activity video-sequences, both DVC baseline-model and DVC-transcoding are lagging compared to H.264-Intra. In addition, in all the cases, H.264/AVC RD performance is better than the DVC, which is expected.

The RD performance comparison summary is given in TABLE I. The H.264-Intra PSNR is taken as X dB for comparison purposes. It is concluded that DVC-transcoding is yielding satisfactory results with video-sequences up-to high motion-activity.

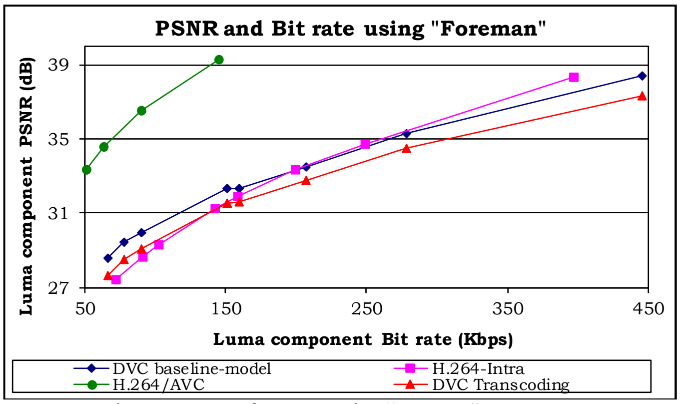

Figure 5. RD performance using "Foreman" sequence.

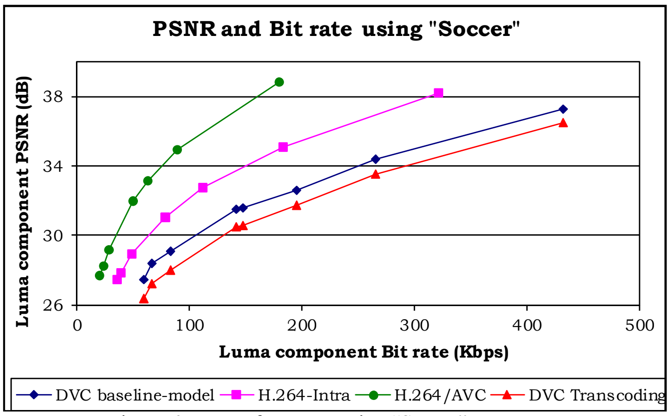

Figure 6. RD performance using "Soccer" sequence.

TABLE I. RD PERFORMANCE COMPARISON SUMMARY

		PSNR (dB)		
Video-sequence	***Motion activity***	***DVC-transcoding***	***DVC baseline-model***	***H.264-Intra***
Hall monitor	Low	X+3.6	X+4	X
Coast guard	Medium-high	X+2	X+3.3	X
Foreman	Very high	X-0.8	X+0.2	X
Soccer	Significantly high	X-2.8	X-1.7	X

B. Compression Ratio

The compression ratio results with various motion-activity video-sequences are presented in Fig. 7 to Fig. 10. This is measured from one mobile device to another device through, The DVC encoder at mobile device-1→ The DVC decoder in the network → The H.264/AVC encoder in the network → The H.264/AVC decoder in mobile device-2.

The Fig. 7 shows the compression ratio achieved using "Hall-monitor" video-sequence. The compression ratio achieved with DVC-transcoding is up-to 25% lesser than H.264/AVC. In addition, Fig. 8 shows the compression ratio achieved using "Coast-guard" video-sequence. The compression ratio achieved with DVC-transcoding is 17% lesser than that of H.264/AVC. Further, Fig. 9 shows the compression ratio achieved using "Foreman" video-sequence. The compression ratios achieved with DVC-transcoding is 20% lower than H.264/AVC. Moreover, the Fig. 10 shows the compression ratio achieved using "Soccer" video-sequence. The compression ratio achieved with DVC-transcoding is 19% lesser than that of H.264/AVC. The comparison summary of the compression ratio is given in TABLE II. The H.264/AVC compression ratio is taken as "X" for comparison purposes. It is concluded that the compression ratio of DVC-transcoding is reasonably high, though little inferior to H.264/AVC.

TABLE II. COMPRESSION RATIO COMPARISON SUMMARY

		Compression Ratio	
Video-sequence	***Motion activity***	***DVC-transcoding***	***H.264/AVC***
Hall monitor	Low	0.75X	X
Coast guard	Medium-high	0.83X	X
Foreman	Very high	0.80X	X
Soccer	Significantly high	0.81X	X

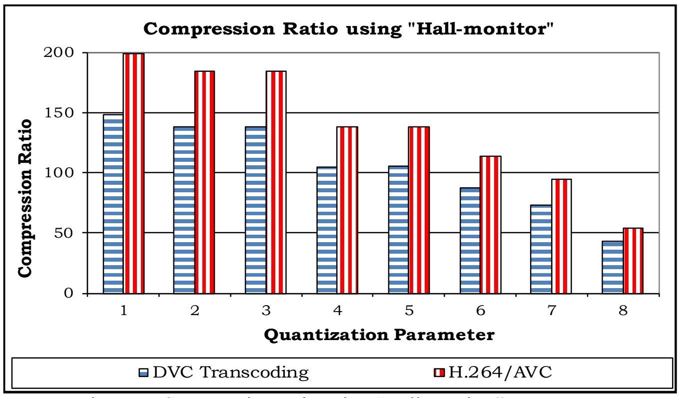

Figure 7. Compression ratio using “Hall-monitor” sequence.

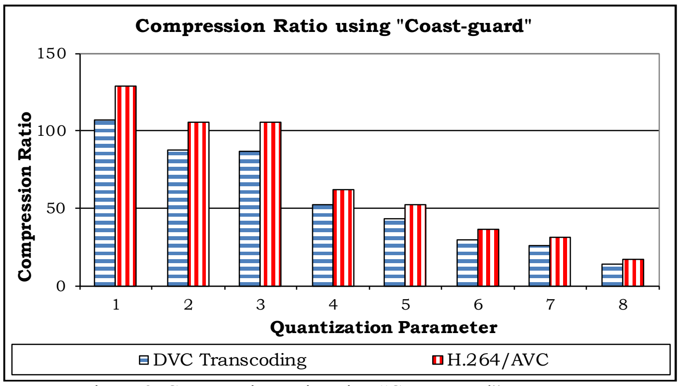

Figure 8. Compression ratio using “Coast-guard” sequence.

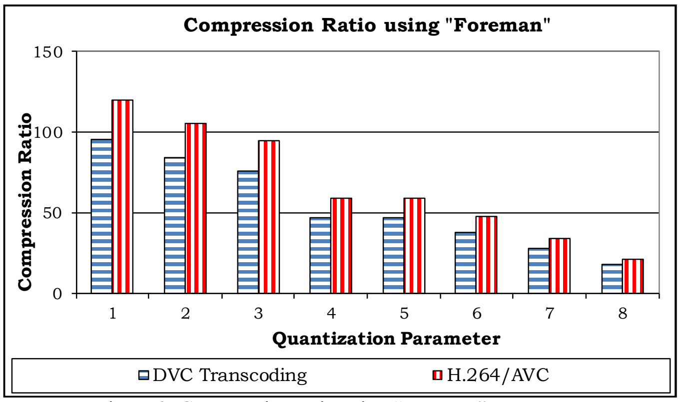

Figure 9. Compression ratio using “Foreman” sequence.

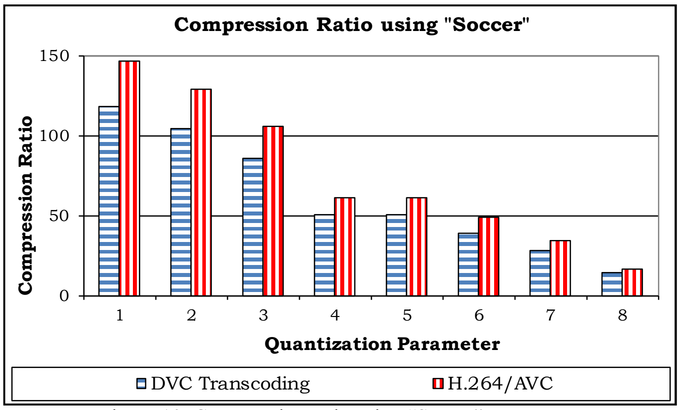

Figure 10. Compression ratio using “Soccer” sequence.

C. Encoder Complexity

The machine run time is used as a measure of encoder complexity. The encoder complexity results of various motion-activity video-sequences are presented in Fig. 11 to Fig. 14. This is measured at the mobile device-1 in the video conference application.

The Fig. 11 shows encoder run-time using “Hall-monitor” video-sequence. The DVC-transcoding encoder complexity is up-to only 5% of that of H.264/AVC. In addition, the Fig. 12 shows encoder run-time using “Coast-guard” video-sequence. The DVC-transcoding encoder complexity is up-to only 4% of that of H.264/AVC. Further, the Fig. 13 shows encoder run-time using “Foreman” sequence. The DVC-transcoding encoder complexity is up-to only 4% of that of H.264/AVC.

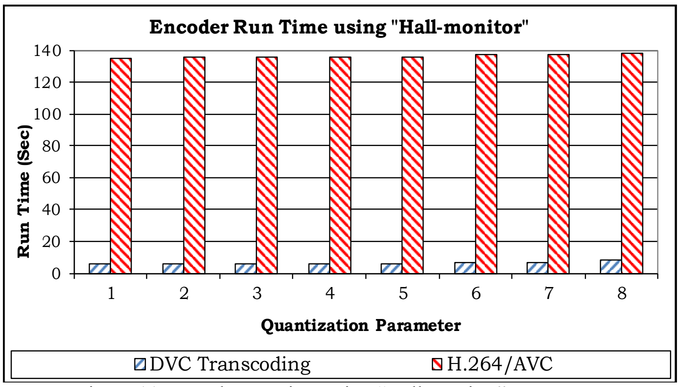

Figure 11. Encoder run-time using “Hall-monitor” sequence.

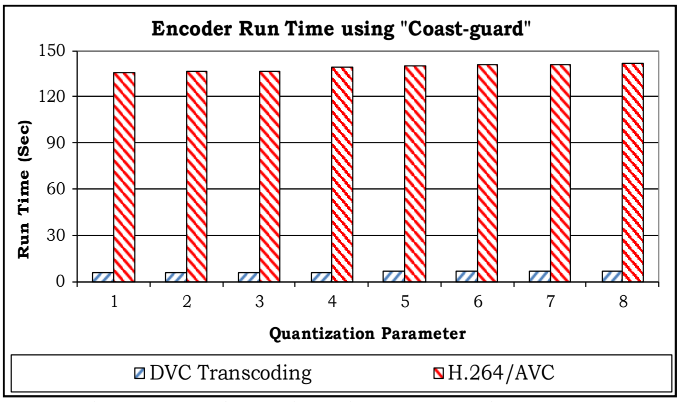

Figure 12. Encoder run-time using “Coast-guard” sequence.

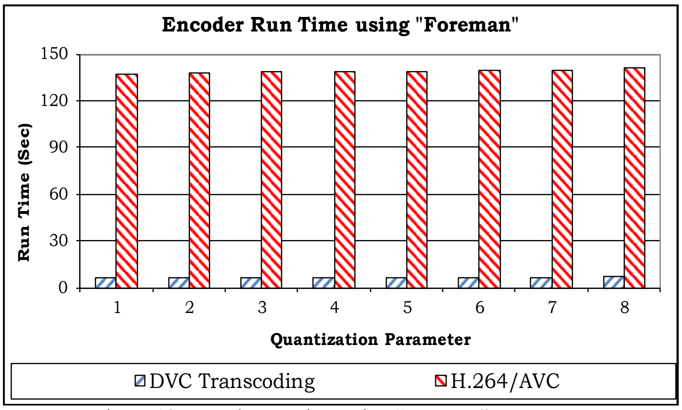

Figure 13. Encoder run-time using “Foreman” sequence.

In addition, the Fig. 14 shows encoder run-time using "Soccer" video-sequence. The DVC-transcoding encoder complexity is up-to only 4% of that of H.264/AVC.

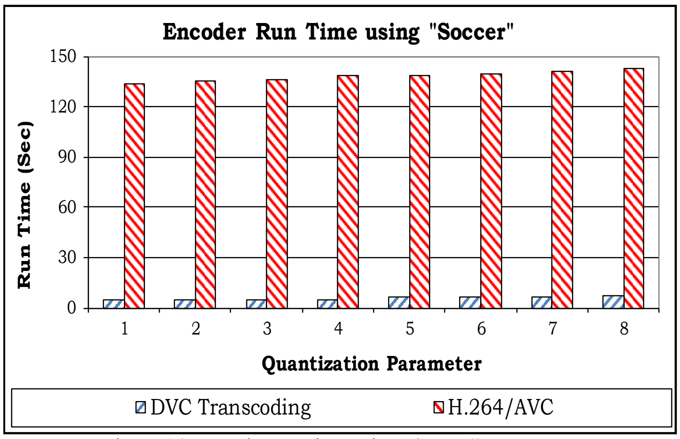

Figure 14. Encoder run-time using "Soccer" sequence.

The encoder complexity comparison summary is given in TABLE III. The H.264/AVC complexity is taken as "X" for comparison purposes. In summary, the DVC-transcoding system encoder complexity is far lesser than that of H.264/AVC, which is one of the main objectives.

TABLE III. ENCODER COMPLEXITY COMPARISON SUMMARY

		Encoder Complexity	
Video-sequence	***Motion activity***	***DVC-transcoding***	***H.264/AVC***
Hall monitor	Low	0.05X	X
Coast guard	Medium-high	0.04X	X
Foreman	Very high	0.04X	X
Soccer	Significantly high	0.04X	X

D. Decoder Complexity

The decoder complexity results of various motion-activity video-sequences are shown in Fig. 15 to Fig. 18. This is measured at the mobile device-2 in the video conference application. In summary, DVC-transcoding decoder complexity is more or less same as that of H.264/AVC, which is another key objective.

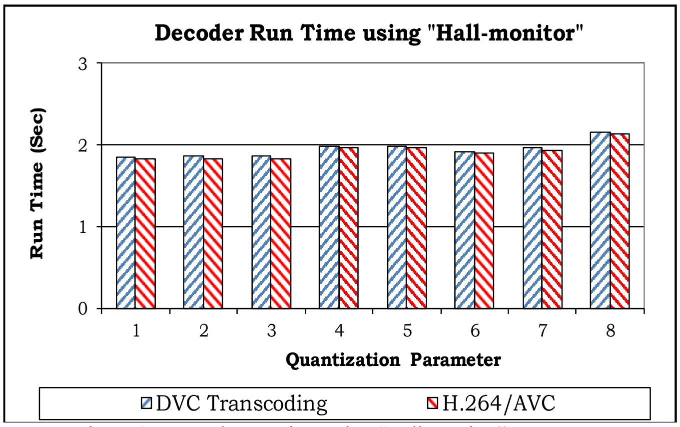

Figure 15. Decoder run-time using "Hall-monitor" sequence.

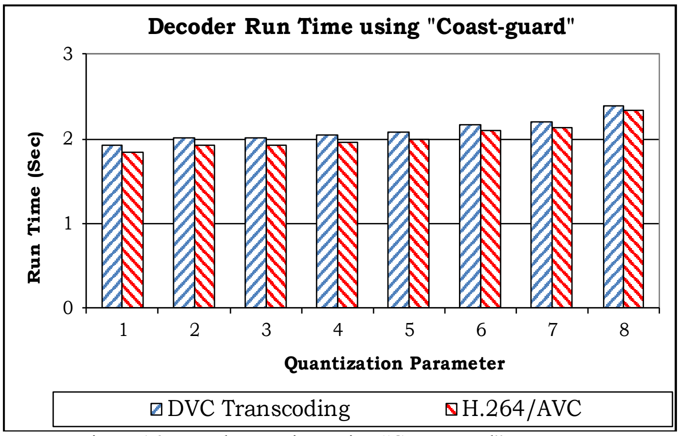

Figure 16. Decoder run-time using "Coast-guard" sequence.

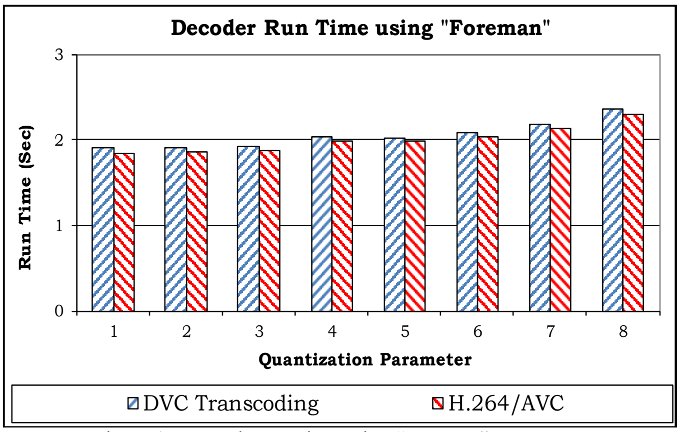

Figure 17. Decoder run-time using "Foreman" sequence.

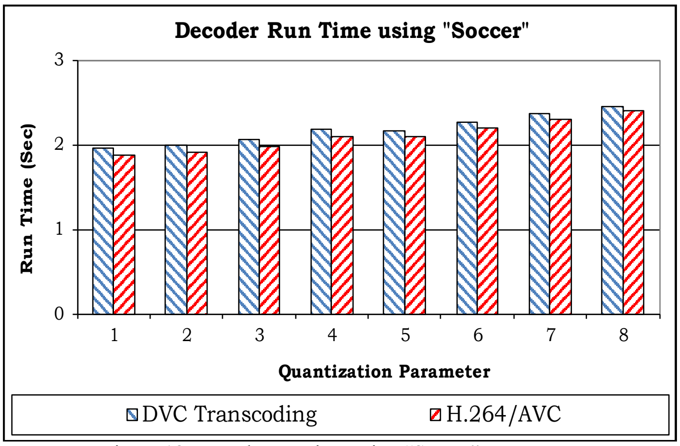

Figure 18. Decoder run-time using "Soccer" sequence.

IV. CONCLUSION

The proposed DVC-transcoding method has met its key objectives of low complexity encoder and the decoder at both mobile devices of video conference application. The performance results of DVC-transcoding presented in this paper demonstrates that superiority of RD performance depends on motion-activity in the input video-sequence and QP. From DVC-transcoding model performance evaluations, it is concluded that its RD performance is superior to H.264-Intra with up-to high motion-activity video-sequences. In addition, the DVC-transcoding compression ratio achieved is reasonably high, though marginally inferior to H.264/AVC. Further, the encoder complexity of DVC-transcoding is much lower than H.264/AVC, which is one of key objectives. Similarly, DVC-transcoding decoder complexity is more or less same as that of H.264/AVC, which is another key objective. For DVC-transcoding to gain popularity and wide usability, further enhancements are needed in terms of extra hash bits from the encoder and successive refinement of the predicted SI at the decoder, to push the DVC performance levels up. This includes a complex fusion for the SI generation and more effective error correction methods.

REFERENCES

[1] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv coding of motion video," in Proceedings of Asilomar Conference on Signals and Systems, Pacific Grove, USA, November 2002.

[2] A. Aaron, E. Setton, and B. Girod, "Towards practical Wyner-Ziv coding of video," in Proceedings of IEEE International Conference on Image Processing, Barcelona, Spain, September 2003.

[3] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform domain Wyner-Ziv codec for video", in Proceedings of SPIE Visual Communications and Image Processing, San Jose, USA, January 2004.

[4] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," IEEE Transactions on Information Theory, vol. 49, no. 3, March 2003.

[5] R. Puri and K. Ramchandran, "PRISM: a new "reversed" multimedia coding paradigm," in Proceedings of International Conference on Image Processing (ICIP), Barcelona, Spain, September 2003.

[6] 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, October 2007.

[7] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The Discover Codec: Architecture, Techniques and Evaluation", in Proceedings of Picture Coding Symposium (PCS), Lisboa, Portugal, November 2007.

[8] J. Ascenso et al., "The VISNET II DVC codec: architecture, tools and performance," in European Signal Processing Conference, 2010.

[9] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Chroma components coding in feedback-free distributed video coding," in Proceedings of 5th IEEE International Conference on Internet Multimedia Systems Architecture and Applications (IMSAA), Bangalore, India, December 2011.

[10] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Distributed video coding: feedback-free architecture and implementation," International Journal of Image and Graphics (IJIG), vol. 12, no. 2, April 2012.

[11] Vijay Kumar Kodavalla, and P.G. Krishna Mohan, "Distributed Video Coding: Codec Architecture and Implementation," Signal & Image Processing: An International Journal (SIPIJ), vol. 2, no. 1, March 2011.

[12] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Chroma components coding method in distributed video coding," in Proceedings of IEEE International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, March 2012.

[13] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Multi-view distributed video coding," in Proceedings of IEEE International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, March 2012.

[14] Vijay Kumar Kodavalla and P.G. Krishna Mohan, "Chroma components coding in distributed video coding," International Journal of Computer Science and Communication (IJCSC), vol. 3, no.1, June 2012.