Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 10, October 2015, pg.233 – 238

SURVEY ARTICLE

Different Aspect of Side Information Generation in Distributed Video Coding: A Survey

Suvendu Rup[1], Bodhisattava Dash[2], Debbrota Paul Chowdhury[3]

[1,2,3] Computer Science & Engineering, IIIT Bhubaneswar, India

suvendu@iiit-bh.ac.in, bdash.fac@gmail.com, debbrota@gmail.com

Abstract— In recent years, Distributed Video Coding (DVC) is becoming more popular due to the emerging applications like mobile camera phone, video surveillance, wireless camera. So it is a challenge for traditional video coding to fulfill the requirements of the above mentioned applications. So DVC is an ultimate choice over traditional video coding where the computational complexity has been shifted from encoder to decoder. Side information (SI) is most widely focused research area in DVC. It is one of the key information that directly influences the DVC performance. So keeping this in mind, this paper presents an extended survey on SI generation in DVC. This survey mainly focuses and highlights the contributions made by different researchers in context to SI generation and also addresses the limitations of some relevant reported schemes.

Keywords— "Distributed Video Coding, Wyner-Ziv Video Coding, Side Information, Motion Compensated Interpolation, Traditional Video Coding"

I. INTRODUCTION

The current digital video compression schemes are represented by the International Telecommunication Union-Telecommunication (ITU-T) and Motion Picture Expert Group (MPEG) standards, which rely on a combination of block-based, transform and inter-frame predictive coding to exploit the spatial and temporal redundancies within the encoded video [1]. This results in high complexity encoder due to the motion estimation task at the encoder side. On the other hand, the decoder is so simple and around five to ten times less complex than the encoder. However, this type of structure is well-suited for applications where the video is encoded once and decoded many times. It shows a one-to-many topologies for down link model applications such as broadcasting or streaming and video on demand.

In recent years, the emerging applications like mobile camera phone, video surveillance, multimedia sensor networks, wireless camera etc. where the memory requirement and computation at the encoder are scarce. The traditional video coding architecture has complex

encoder and a simple decoder. It is a challenge for traditional video coding to fulfill the requirements of the above mentioned applications. Another important goal is to achieve the coding efficiency similar to that of traditional video coding schemes i.e. shifting the complexity from encoder to the decoder should not compromise the coding efficiency [2, 3]. So, to address the challenges, distributed source coding has emerged by exploiting the source statistics partially or totally at the decoder. Thus, it enables a flexible complexity distribution between encoder and decoder [4]. To apply DSC to a video compression standard is called distributed video coding (DVC), targeting both low complexity encoding and error resilience [5, 6]. The SI generation module in DVC is considered to be a key module. As the SI generation is more accurate, the decoder of DVC requests less number of parity bits from encoder. On the other hand, the overall performance of DVC strongly depends on the SI generation. So, a significant amount of efforts have been made in SI generation schemes in DVC. The rest of the paper is organized as follows. Section II presents the Stanford based transform domain architecture in DVC which is widely adopted by different research groups. Section III elaborately reviews the different SI generation schemes in DVC. Finally, Section IV gives the concluding remarks.

II. STANFORD BASED TRANSFORM DOMAIN DVC ARCHITECTURE

Recently, major practical solutions of DVC have been proposed by two groups: Bernd Girod's group at Stanford University [5] and Ramchandran's group at the University of California, Berkeley [6]. Most of the practical solutions of DVC use Stanford based architecture. So, in this paper we have presented this architecture. In 2002, the University of Stanford put forward a proposal for pixel domain DVC, introducing two different types of frames in a video called WZ frames and key frames [7]. This framework is commonly known as Stanford DVC framework or Stanford Wyner-Ziv framework. Later, they have proposed transform domain DVC framework [8], which results good compression efficiency as compared to pixel domain coding in DVC. So for better understanding, next we have presented the architecture of transform domain Wyner-Ziv (TDWZ) video coding. Figure 1 shows the overall architecture of Stanford based TDWZ video coding. The detail stepwise procedure of TDWZ video coding is described as,

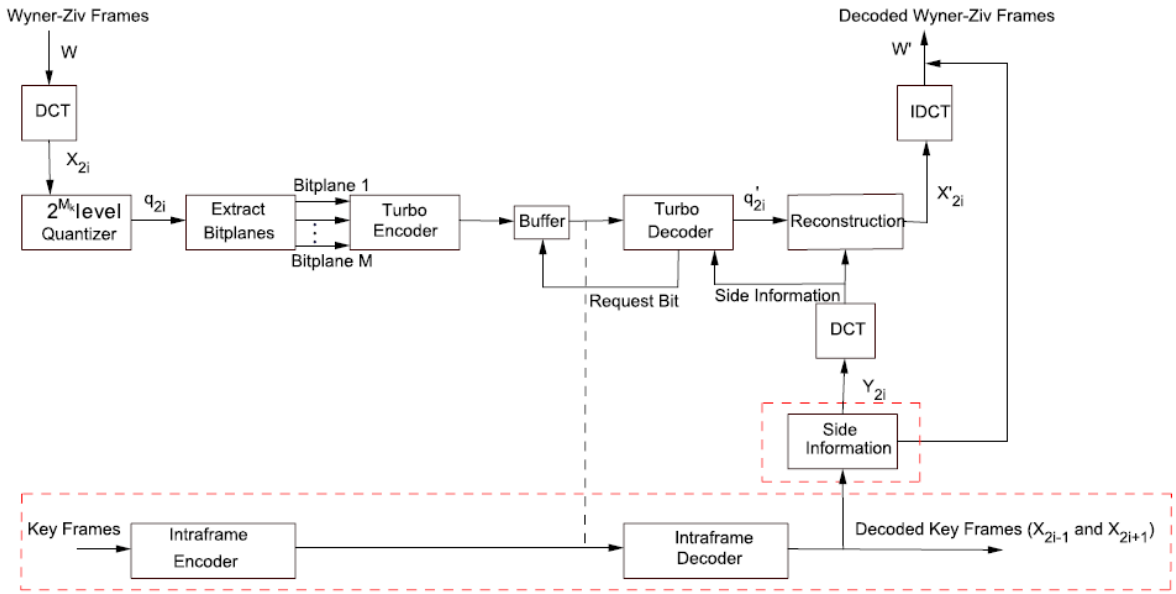

Figure 1: Stanford based transform domain architecture

I. The frames in a video sequence are divided into WZ frames and key frames.

II. For each WZ frame X2i, a 4 × 4 block wise DCT is applied.

III. The transform coefficients (DCT coefficients) of the frame X_{2i} are grouped together. The coefficients from the same position of each DCT blocks are used to compose 16 possible DCT coefficient bands.

IV. Each DCT coefficient band X_k, k = 1,2,...,16 is uniformly quantized to obtain quantized symbol stream q_{2i}.

V. The different quantized coefficients of the same band are grouped together and different bit planes are extracted. The bit planes are organized from the most significant bit (MSB) plane to least significant bit (LSB) plane.

VI. Turbo encoding is applied to each bit plane. The turbo encoder generates the parity bits for each bit plane which is saved in a buffer and sends to the decoder upon request. A pseudo random puncturing pattern is used to transmit the parity bits.

VII. Meanwhile, the key frames are encoded using the conventional intra-frame video coding.

VIII. Decoder uses the frame interpolation technique for estimation of X_{2i} frame known as side information, Y_{2i} frame from the two adjacent key frames X_{2i-1} and X_{2i+1}.

IX. The same block wise 4 × 4 DCT is applied to the interpolated frame, Y_{2i} to generate an estimate of X_{2i}. The correlation between corresponding coefficient band X_{2i} and Y_{2i} is modeled by a Laplacian distribution.

X. The decoder also determines if the current bit plane error probability, P_e exceeds 10^{-3}, the decoder requests number of parity bits; otherwise, the current bit plane is executed successfully.

XI. After all bit planes are executed successfully and the quantized symbol stream q_{2i}' is obtained, then reconstruction of each DCT coefficient band is resulted.

XII. After all DCT coefficient bands are reconstructed, an inverse discrete cosine transform (IDCT) is applied to obtain the X_{2i}' frame.

III. EXTENDED SURVEY ON SIDE INFORMATION GENERATION SCHEMES IN DISTRIBUTED VIDEO CODING

Most popular distributed video coding (DVC) solutions use the correlation between original frame with a frame predicted at the decoder. This predicted frame is known as side information (SI), which is a key function in the DVC decoder. The more accurate is the predicted SI; fewer numbers of bits need to be sent to decode the Wyner-Ziv (WZ) frame. So, SI generation is one of the most focused areas of research that directly influence the DVC performance.

The RD performance of DVC strongly depends on the quality of the decoded key frames and the accuracy of the SI generation process. In the original Stanford-based architecture, SI is generated through interpolation of two decoded key frames i.e. preceding and succeeding frames of the original WZ frames. In the recent past, several related works have been reported in the literature for SI generation which involves sophisticated framework. Aaron et al. have proposed two hierarchical frame dependency arrangements in DVC [9]. In their first approach, the SI for the current WZ frame can be extrapolated from a key frame or from a WZ frame. In their second approach, a more complex arrangement has been used with an increase temporal resolution of 2:1 with bidirectional interpolation. With the above two schemes, a poor SI estimation is resulted as group of picture (GOP) size increases and motion becomes more intense and less well behaved. The same authors have proposed solutions using motion compensated interpolation (MC-I) and motion compensated extrapolation

(MC-E) [8]. In MC-I, the SI for an even frame at time index t is generated by performing motion compensated interpolation using the two decoded key frames at time $(t - 1)$ and $(t + 1)$.
This interpolation technique involves symmetrical bidirectional block matching for the estimated motion. Since the next key frame is needed for interpolation, the frames have to be decoded out-of-order. This scheme is similar to the decoding of bidirectional frames in predictive coding which is a limitation of this framework.

In MC-E, the SI is generated by estimating the motion between the decoded WZ frame at time $(t - 2)$ and decoded key frame at time $(t - 1)$. Here, the already decoded WZ frames are used for motion estimation. So the reconstruction error from the WZ frame contributes to degradation of SI quality.

Girod et al. have proposed previous extrapolation (Prev-E) and average interpolation (AV-I) techniques to generate SI for low complexity video coding solution [10]. In Prev-E scheme, the previous key frame is used directly as SI whereas, in AV-I technique, the SI for the WZ frame is generated by averaging the pixel values from the key frames at time $(t - 1)$ and $(t + 1)$. The limitation of Prev-E scheme is that it does not employ motion compensation to generate SI. So the Prev-E scheme is better than DCT based intra-frame coding only at lower bit rates. In AV-I scheme, the pixel values from key frames $(t - 1)$ and $(t + 1)$ are averaged and it is not sufficient to generate a good quality SI. A hash based motion compensation is proposed by. J. Ascenso and F. Pereira in [11].

In 2005, Tagliasaccchi et al. have presented a motion compensated temporal filtering technique [12]. This scheme is based on pixel domain coding solution. Natario et al. have proposed a motion field smoothing algorithm to generate SI in [13]. Artigas and Torres have proposed an iterative motion compensated interpolation technique where the turbo decoder runs several times for decoding the WZ frame to estimate SI and as a result a significant delay is associated [14].

Adikari et al. have proposed a multiple SI stream for DVC [15]. It uses two SI streams which are generated using motion extrapolation and compensation (ME-C). The first SI stream (SS-1) is predicted by extrapolating the motion from the previous two closest key frames. The second SI stream (SS-2) is predicted using the immediate key frame and the closest WZ frame. Fernando et al. have proposed a SI scheme using sequential motion compensation, using both luminance and chrominance information to improve the decoding performance of DVC [16]. This work has been extended by Weerakkody et al. in [17]. Here, a spatio-temporal refinement algorithm is used to improve the SI resulting from motion extrapolation.

In 2006, Kubasov et al. have presented a mesh-based MC-I approach to resolve the problems occurred in the SI interpolation. Their approach is based on block translational motion model [18]. In this scheme, it addresses the problem of motion discontinuities and occlusions. The overall increasing accuracy of SI leads to the improvement in WZ coding. However, this scheme suffers from a high computational complexity burden due to the mesh based structure adopted in the framework. A technique to improve the performance by sending additional hash information to help SI has been presented in [19]. However, this framework is suitable for PRISM based architecture rather than Stanford-based architecture. Tagliasaccchi et al. have used a Kalman filter to generate improved SI at the decoder [20]. This scheme shows better coding performance. However, it increases the encoder complexity due to motion estimation task being performed at the encoder.

Badem et al. have proposed a novel SI refinement technique based on motion estimation in DCT domain for DVC [21]. Varodayan et al. have proposed an unsupervised motion vector learning algorithm for SI generation [22]. This method applies an expectation maximization (EM) algorithm for unsupervised learning of motion vectors. The authors have claimed a better RD performance.

Brites et al. have proposed a frame interpolation framework with forward and bi-directional motion estimation with spatial smoothing [23]. This framework is commonly known as Instituto Superior Tecnico Transform Domain Wyner-Ziv (IST-TDWZ) video coding and used by many DVC researchers. However, the most promising SI refinement framework has been adopted and extended by the same group in their other contribution in [24]. Here, first both the key frames are low-pass filtered. Then a block matching algorithm is used to make the motion estimation between two adjacent key frames. The bi-directional motion estimation is performed followed by the forward motion estimation. Once the final motion vectors are obtained, the bidirectional motion compensation is performed. As so many modules are associated to create SI, it affects the decoding time complexity. The distributed coding for video services (DISCOVER) codec which also uses the same SI generation framework proposed by IST group [25]. IST-TDWZ video codec shows similar performance as DISCOVER video codec. The DISCOVER codec uses lower density parity check (LDPC) code whereas IST group uses turbo code.

IV. CONCLUSIONS

This paper presents an extended literature survey based on SI generation schemes in DVC. It is observed from the literature that DVC is a prominent area of research due to its vast applications in handheld devices with less memory and computing power. In addition, DVC performance is associated with mostly on side information generation, which in turn dependent on reconstructed key frames. From the thorough investigation from the reported literature it has been observed that there exists a scope for improvement of DVC performance through better quality side Information.

REFERENCES

1. T. Sikora. MPEG digital video coding standards. Signal Processing Magazine, 14(5):82–100, September 1997.
2. S. Park, Y. Lee, C. Kim, and S. Lee. CDV-DVC: Transform domain distributed video coding with multiple channel division. Journal of Visual Communication and Image Representation, 24(4):534–543, April 2013.
3. Catarina Brites. Adavances on distributed video coding. Master's thesis, IST Portugal, December 2005.
4. C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann. Distributed monoview and multiview video coding: basics, problems and recent advances. IEEE Signal Processing Magazine, Special Issue on Signal Processing for Multi terminal Communication Systems, pages 67–76, September 2007.
5. B. Girod, A. Aaron, S. Rane, and D. Monedero. Distributed video coding. Special Issue on Advances in Video Coding and Delivery, 93(1):71–83, January 2005.
6. R. Puri and K. Ramchandran. PRISM: a new robust video coding architecture based on distributed compression principles. In Allerton Conference on Communication, Control and Computing, pages 1–10,October, 2002
7. R. Puri, A. Majumdar, and K. Ramchandran. PRISM: A Video Coding Paradigm with Motion Estimation at the Decoder. IEEE Transactions on Image Processing, 16:2436–2448, October 2007.
8. Aaron, R. Zhang, and B. Girod. Wyner Ziv coding of motion video. In Proceedings of Asilomar Conference on Signals and Systems, pages 240–244, November 2002.
9. Aaron, S. Rane, E. Setton, and B. Girod. Tansform domainWyner Ziv for video. In Proceedings of SPIE Visual Communication and Image Processing, pages 520–528, Sanjose,Carlifornia, January 2004.

10. Aaron, E. Setton, and B. Girod. Towards practical Wyner-Ziv coding of video. In Proceedings of International Conference in Image Processing, pages 869–872, September 2003
11. J. Ascenso and F. Pereira. Adaptive hash-based side information exploitation for efficient Wyner-Ziv video coding. In International Conference on Image Processing, pages 29–32, San Antonio, September 2007.
12. M. Tagliasaccchi and S. Tubaro. Combining MCTF with distributed source coding. In Proceedings of Visual Communication and Image Processing, pages 797–800, Beijing, China, July 2005.
13. L. Natario, C. Brites, J. Ascenso, and F. Pereira. Extrapolating side information for low-delay pixel domain distributed video coding. In Lecture Notes in Computer Science, pages 16–21, April 2006.
14. X. Artigas and L. Torres. Iterative generation of motion-compensated side information for distributed video coding. In IEEE International Conference on Image Processing, pages 833–836, September 2005.
15. Adikari, W. Fernando, and K. Arachchi. Multiple side information streams for distributed video coding. IEEE Electronics letter, 42(25):1447–1449, December 2006.
16. Adikari, W. Fernando, H. Arachchi, and W. Weerakkody. Sequential motion estimation using luminance and chrominance information for distributed video coding of Wyner-Ziv frames. IEEE Electronics letter, 42(7):398–399, March 2006.
17. W.Weerakkody, W. Fernando, J. Martinez, and F. Quiles P. Cuenca. An iterative refinement technique for side information generation in DVC. In IEEE International Conference on Multimedia and Expo, pages 164–167, July 2007.
18. D. Kubasov and C. Guillemot. Mesh-based motion-compensated interpolation for side information extraction in distributed video coding. In IEEE International Conference on Image Processing, pages 261–264, October 2006.
19. T.N Dinh, G. Lee, J. Y. Chang, and H. Cho. Side information generation using extra information in distributed video coding. In Proceedings of IEEE International Symposium on Signal Processing and Information Technology, pages 138–143, December 2007.
20. M. Tagliasacchi, S. Tubaro, and A. Sarti. On the modeling of motion in Wyner-Ziv video coding. In International Conference on Image Processing, pages 593–596, October 2006.
21. M. Badem, M. Mark, and W. Fernando. Side information refinement using motion estimation in dc domain for transform-based distributed video coding. IEEE Electronics letter, 44(16):965–966, July 2008.
22. D. Varodayan, D. Chen, M. Flierl, and B. Girod. Wyner-Ziv coding of video with unsupervised motion vector learning. Signal Processing Image Communication, 23(8):369–378, June 2008.
23. J. Ascenso, C. Brites, and F. Pereira. Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding. In The 5th EURASIP Conf. Speech and Image Processing, Multimedia Communications and Services, 2005.
24. Brites, J. Ascenso, J. Pedro, and F. Pereira. Evaluating a feedback channel based transform domain Wyner-Ziv video codec. Signal Processing Image communication, 23(4):269–297, April 2008.
25. X. Artigas1, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret. The DISCOVER codec: architecture, techniques and evaluation. In Proceedings of Picture Coding Symposium, pages 1–4, Lisbon, Portugal, November 2007.

