28th Picture Coding Symposium, PCS2010, December 8-10, 2010, Nagoya, Japan

BACKGROUND AIDED SURVEILLANCE-ORIENTED DISTRIBUTED VIDEO CODING

Hongbin Liu[1], Siwei Ma[2], Xiaopeng Fan[1], Debin Zhao[1] and *Wen Gao[2]*
{hbliu, swma}@jdl.ac.cn, eexp@ust.hk, {dbzhao, wgao}@jdl.ac.cn

1Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, P.R. China
2Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing, 100871, P.R. China

ABSTRACT

Distributed video coding (DVC) was proposed to meet the low complexity encoding requirement, and it was verified to work more efficiently than H.264/AVC intra coding on video sequences with low motion. This makes DVC suitable for surveillance application. This paper presents a background aided surveillance-oriented distributed video coding system. A high quality background frame is encoded for each group of pictures (GOP), which can provides high quality SI for the background parts of the Wyner-Ziv (WZ) frames. Consequently, bit rate for the WZ frames can be reduced. Experimental results demonstrate that the proposed system can decrease the bit rate by up to 67.4% when compared with traditional DVC codec.

Index Terms— surveillance, background, distributed video coding

1. INTRODUCTION

As the rapid development of mobile communication, diverse devices are equipped with cameras. As many of them are power-limited or battery-limited, low complexity encoding is required. The existing video compression standards, such as MPEG-x and H.26x, perform computationally intensive motion field estimation including motion estimation and mode decision for inter-pictures coding at encoder to efficiently exploit the temporal correlation. As a result, it makes the encoder very complicated and is inappropriate for such applications. Thus, new encoding paradigm is required. DVC is proposed recently to meet the low complexity encoding requirement.

DVC is built on the theoretic results of Slepian-Wolf and Wyner-Ziv theorems [1, 2]. Based on these two theorems, several practical DVC systems were proposed. Pradhan proposed the DISCUS [3] architecture, which used syndrome-based encoding to perform WZ video coding. Puri proposed another DVC framework, which is described as PRISM in [4]. Aaron also provided an asymmetric system with low-complexity encoder [5-6], which is known as the Stanford DVC framework. Stanford DVC framework includes two categories: pixel domain DVC (PDVC) [5] and DCT transform domain DVC (TDVC) [6]. Compared with PDVC, TDVC is more complex but it can bring significant performance improvement. In this paper, we select TDVC as the platform.

DVC has been a hot topic in video coding for providing low complexity encoding. However, it is also argued that DVC has a poor performance on sequences with high motion. This drawback makes DVC unsuitable for common video coding sceneries. Instead, it is suitable for special applications where low complexity encoding is required and the video contents changes slowly, such as surveillance system. In surveillance video coding, on one hand, low complexity encoder is preferred as the real-time encoding is usually required; on the other hand, high motion content does not appear frequently. Based on the above observation, we focus on the surveillance-oriented DVC (SDVC) in this paper.

In traditional hybrid coding, surveillance video coding has been studied by many researchers. In MPEG-4 object-based coding [7], video content are segmented into objects and compressed separately. However, accurate segmentation is still a tough task. Background modeling based algorithm, specifically, sprite coding is also proposed. In sprite coding [8] in MPEG-4, the video content is segmented into foreground and background. The background can be derived from the sprite image utilizing motion parameters and shape information [9]. Consequently, bit rate for the background can be reduced. Many background models, such as mixture Gaussian model [10] and non-parametric model [11], are also proposed to extract the background of the video content.

In this paper, we borrow the idea of the background modeling based algorithm in hybrid video coding, and propose a novel background aided SDVC system. In the proposed system, besides key frames and WZ frames, a high quality background frame is also encoded for each GOP, which is utilized to provide high quality SI for the background parts of the WZ frames. In this way, bit rate for the WZ frames is greatly reduced. For generating the background frame, existing background extraction algorithms [10, 11] are not employed because they are usually too complex for the SDVC encoder. Instead, a low complexity method, which simply averages all frames in a training set, is used to generate the background. For each GOP, the last several frames of it are used to build the training set for the next GOP.

The rest of this paper is organized as follows. In Section 2, low delay DVC system, which is chosen as the anchor

978-1-4244-7133-1/10/$26.00 ©2010 IEEE

SDVC system is reviewed. In Section 3, the proposed background aided SDVC system is described in detail. In Section 4, experimental results are provided. Finally, we conclude this paper in Section 5.

2. LOW DELAY DVC

As we know, real-time coding is usually required in surveillance application. From this point of view, low delay TDVC (LTDVC) codec [6] is a reasonable choice in SDVC. In this section, LTDVC is reviewed.

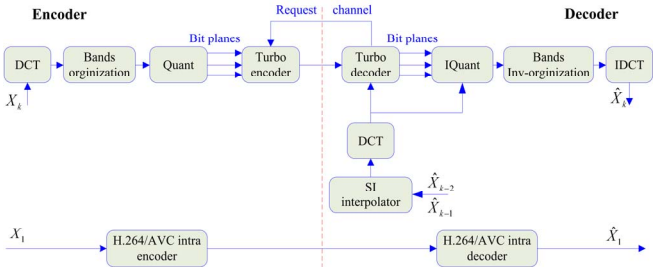

Figure 1: Block diagram of LTDVC system.

Block diagram of LTDVC is illustrated in Fig.1. At the encoder, frames are encoded into key frames and WZ frames respectively. The first frame of each GOP is encoded into key frame in H.264/AVC intra mode, while the left frames are encoded into WZ frames. As shown in Fig.1, X_1 and X_k denote the first frame the kth frame in one GOP. For each WZ frame, a non-overlapped 4×4 block based DCT is performed firstly. Then, the positions of DCT coefficients are classified into several groups according to the pre-selected quantization parameter. All coefficients in the same group are raster scanned and organized into one band, and then pipelined to one uniform quantizer. Different bands are quantized utilizing separate uniform quantizers. Finally, bitplanes of each quantized band are extracted and encoded by turbo encoder.

At the decoder, key frames are decoded by H.264/AVC intra decoder. For each WZ frame, motion compensated extrapolation (in which two previously reconstructed frames are used) is employed to generate the SI of it. Then corresponding 4×4 DCT transform, band organization and band quantization are performed on SI. After that, DCT bands are turbo decoded and reconstructed in zigzag order with the help of SI. If some bands are not encoded, the corresponding bands in SI will be used as the reconstruction. Finally, when all bands are reconstructed, they will be inv-organized and IDCT transformed to reconstruct the WZ frame.

3. PROPOSED BACKGROUND AIDED SURVEILLANCE-ORIENTED DVC SYSTEM

In this section, the proposed background aided SDVC (BSDVC) system is described in detail. Schematic block diagram of the proposed BSDVC system is shown in Fig.2, and the new introduced modules are highlighted in dashed windows.

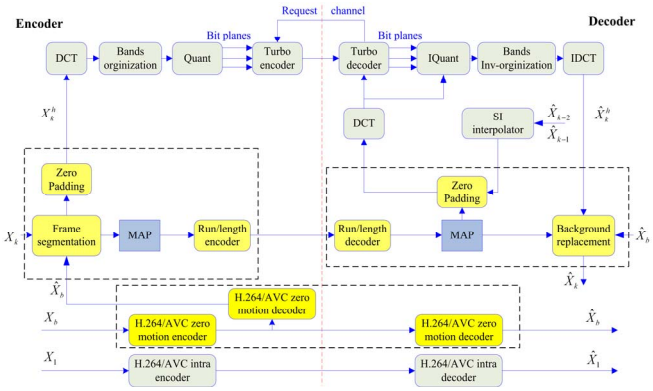

Figure 2: Schematic block diagram of the proposed background aided SDVC system.

In the proposed system, besides key frames and WZ frames, a high quality background frame is also encoded for each GOP. Key frames are encoded using H.264/AVC intra encoder, while background frames are encoded into H.264/AVC zero motion mode (in which motion estimation is not required, thus, low complexity requirement is still fulfilled) to save bits. Besides the traditional SI which can only be accessed at decoder, the decoded background frame is used as the second SI at both the encoder and the decoder.

In surveillance application, background does not change frequently. Consequently, the high quality background frame is expected to provide a high quality SI for the background part of the WZ frames. Following this idea, the "background", actually blocks which do not change much from their collocated blocks in the background frame, of the WZ frame can be omitted during the encoding. Based on the above analysis, the encoding of WZ frame is done in two steps, which is different from LTDVC encoder. Firstly, the WZ frame X_k is segmented into two parts based on 16×16 block, namely WZ part (which is WZ encoded) and skip part (which is not encoded), with the help of the background frame. The skip part of X_k is padded with zero to generate a new hybrid frame X_k^h. Next, X_k^h is encoded by the LTDVC encoder, in which the turbo encoder is modified to adapt to the specific characteristic of X_k^h. Meanwhile, the segmentation map for the WZ frame is also entropy encoded and transmitted to the decoder. Here, a simple run-length encoder is used to encode the map. At the decoder, the segmentation map is firstly decoded and used to identify the skip part of the WZ frame. Then, the corresponding skip part of the SI is padded with zero. After this, the hybrid frame is decoded by the LTDVC decoder, and the reconstructed X_k^h is denoted as $\hat{X}_k^h$ in Fig.2. Finally, the skip part of $\hat{X}_k^h$ is replaced by its collocated part of the background frame to reconstruct the WZ frame.

Important issues such as turbo encoder modification, sequence structure, background frame generation method and WZ frame segmentation method are depicted in detail.

Meanwhile, a special case: WZ frame with zero motion skip is also described.

3.1 Modified turbo encoder

The frequently used turbo encoder in DVC consisted of two RSC (recursive systematic convolutional) encoders with parallel concatenation, and an interleaver is placed before the second RSC encoder. When errors in the input are randomly distributed, this is reasonable; however, if the errors tend to be aggregated, another interleaver is necessary to be placed before the first RSC encoder.

In the specific hybrid frame condition, because the skip parts of both the hybrid frame and the SI are padding with zero, the errors locate solely in the WZ part. Meanwhile, WZ part tends to be consecutive in a local region. Therefore, errors (between the SI and the hybrid frame) are aggregated rather than randomly distributed for the first encoder, which may deteriorate performance of the turbo code greatly. To tackle this problem, we add another interleaver (interleaver 1) in the turbo encoder as shown in Fig.3. The turbo decoder is modified correspondingly.

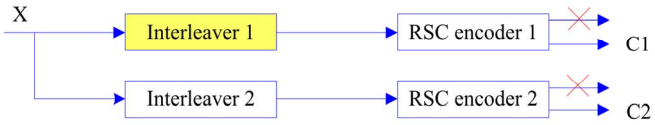

Figure 3: Modified turbo encoder.

3.2 Sequence Structure

As the background frame is used in the proposed system, the sequence structure is modified accordingly as shown in Fig.4. A predecessor training set (training set 0 in the Fig.4) is defined to build the background frame of the first GOP, followed by the GOPs with equal length. The last several frames of each GOP serve as the training set which is used for generating the background frame of the next GOP.

Figure 4: Proposed sequence structure.

3.3 Background Frame Generation

The background frame is generated as the average of all frames in the training set as follows:

$$X_b(i,j)=\frac{\sum_{i=1}^{k} S_i(i,j)}{k} \quad . \qquad (1)$$

Here, X_b refers to the background frame, and the S_i denotes the ith frame in the training set. k is the size of the training set.

3.4 WZ Frame Segmentation

At the encoder, the decoded background frame of one GOP is used to segment all WZ frames in the same GOP. The cost function defined in (2) is used to segment the WZ frame based on 16×16 block.

$$\text{cost}=\frac{\sum_{i=1}^{16}\sum_{j=1}^{16}\left(X_{wz}(i_0+i,j_0+j)-\hat{X}_b(i_0+i,j_0+j)\right)^2}{256} \quad . \qquad (2)$$

Here, X_{wz} is the current WZ frame and $\hat{X}_b$ is the reconstructed background frame. (i_0, j_0) is the coordinate of the upleft pixel of the current block. If the cost is larger than a threshold, the current block is marked as WZ block; otherwise, it will be marked as skip block.

The segmentation map is generated as follows: if the block is a WZ block, a '1' is allocated to it; otherwise, a '0' is allocated to it. Therefore, the segmentation map is a binary map and the length of it is 1/256 of the frame size.

3.5 A Special Case: WZ Frame with Zero Motion Skip

If the training set window slides continuously, i.e., the frames with constant distances from the current WZ frame are used to build the training set for it, specifically, when the size of the training set is 1, the proposed BSDVC deteriorates to the zero motion skip method which is also used in [12]. In this special case, as the background frames change continuously, it is bits wasteful to encode them as aided information. Instead, at the encoder, the original previous frame is used as the background frame to segment the current WZ frame. While at the decoder, the previously reconstructed frame is used as the background frame to pad the skip part of the WZ frame. In this way, no bits are needed for the background frames.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed BSDVC system, simulation results and performance comparisons are provided in this section.

LTDVC in [6], the special case where zero motion skip is used for WZ frames (ZMDVC), and high quality background aided SDVC (BSDVC) are compared. Meanwhile, H.264/AVC intra coding and H.264/AVC inter coding are also chosen as the benchmarks. Two sequences with different resolutions: *Hall Monitor* QCIF@30Hz and *Bank* CIF@30Hz (*Bank* sequence is recorded by our lab and can be downloaded from ftp://159.226.42.57/, and the first frame of it is shown in Fig.5) are selected as the test sequences in the experiment, and the first 300 frames of them are encoded. GOP size is set to 20. The first frame in each GOP is encoded as the key frame and the other frames are encoded as WZ frames. QP = 24, 28, 32, and 36 are used for encoding key frames while QP = 18 is used for encoding background frames. QP of WZ frames are selected accordingly. Size of the training set in subsection 3.2 is set to 5. *Threshold* used for the cost function (2) is set to 25 and 20 for BSDVC and ZMDVC, a smaller threshold is set for ZMDVC because original background frame is used for calculating the cost in it.

RD curves of the five algorithms are shown in Fig.6. To compare the RD performance, the tool proposed in [13] is used. It is found that BSDVC can decrease the bit rate by up to **42.7%** and **67.4%** on *Hall Monitor* and *Bank* respectively when compared with [6], and it can still decrease the bit rate by up to **26.6%** and **66%** on *Hall Monitor* and *Bank* respectively when compared with

ZMDVC. The superior performance of BSDVC mainly attributes to the high quality background frame used in it. When compared with [6] and ZMDVC, the high quality background frame can dramatically improve the SI quality of the background parts of the WZ frames. Therefore, bit rate of WZ frames can be significantly reduced.

Figure 5: First frame of *Bank* CIF.

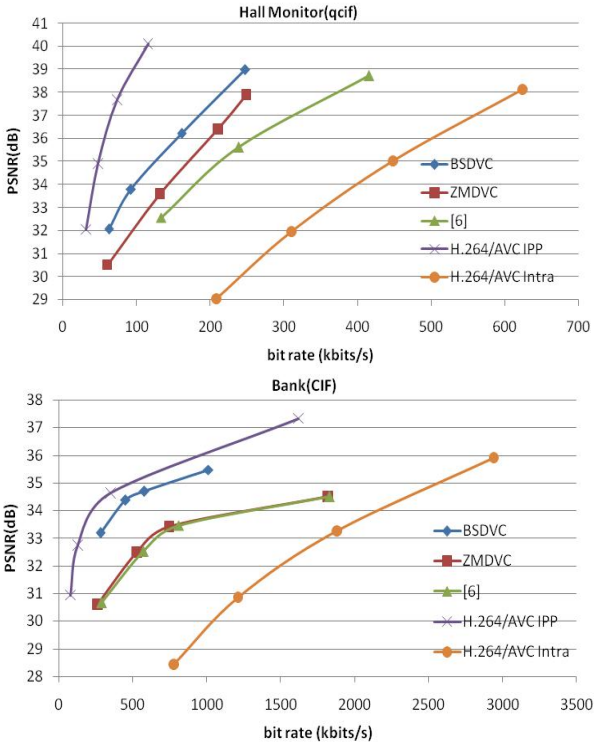

Figure 6: RD performance comparisons of *Hall monitor* and *Bank*.

4. CONCLUSION

A background aided surveillance-oriented DVC system is proposed in this paper, in which a background frame is encoded for each GOP. The background frames are encoded with small QP and of high quality, and can provide high quality SI for the background parts of the WZ frames. Consequently, bit rate of WZ frames are significantly decreased.

ACKNOWLEDGEMENTS

This work was supported in part by National Science Foundation (60736043 and 61073083) and National Basic Research Program of China (973 Program, 2009CB320905).

REFERENCES

[1] J.D. Slepian and J.K. Wolf, "Noiseless Coding of Correlated Information Sources", IEEE Transaction on *Information Theory*, Vol. 22, pp.471-480, July. 1973.

[2] A.D. Wyner and J. Ziv, "The Rate-distortion Function for Source Coding with side information at the decoder", IEEE Transaction on *Information Theory*, Vol. 22, pp.1-10, Jan. 1976.

[3] S. S. Pradhan and K. Ramchandran, "Distributed Source Coding using Syndromes (DISCUS): Design and Construction", in Proc. IEEE *Data Compression Conference*, Snowbird, Utah, USA, Mar. 1999.

[4] R. Puri and K. Ramchandran, "PRISM : A New Robust Video Coding Architecture Based on Distributed Compression Principles", 40th Allerton Conference on *Communication, Control and Computing*, Allerton, IL, Oct. 2002

[5] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv Coding of Motion Video," in Proc. Asilomar Conference on *Signals and Systems*, Pacific Grove, CA, Nov. 2002.

[6] A. Aaron, S. Rane, E. Setton, and B. Girod, "Tansform-Domain Wyner-Ziv Codec for Video," in Proc. SPIE *Visual Communications and Image Processing*, San Jose, CA, Jan. 2004.

[7] T. Sikora, "The MPEG-4 video standard verification model" , IEEE Transaction on *Circuits and Systems for Video Technology*, Vol. 7, pp. 19-31, Feb. 1997.

[8] MPEG Video Group, "Committee Draft," ISO/IEC JTC1/SC29/WG11, N2202, Tokyo, Japan, Mar. 1998.

[9] Y. Lu, W. Gao and F. Wu, "Efficient background video coding with Static Sprite Generation and Arbitrary-Shape Spatial Prediction Techniques", IEEE Transaction on *Circuits and Systems for Video Technology*, Vol. 13, pp. 394-405, May. 2003.

[10] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking", IEEE Computer Society Conference on *Computer Vision and Pattern Recognition*, Ft. Collins, CO, USA, Jun. 1999.

[11] A. Elgammal, D. Harwood and L. Davis, "Non-parametric Model for Background Subtraction", 6th European Conference on Computer Vision, Dublin, Ireland, June 26 - July 1, 2000.

[12] G.G. Hua and C.W. Chen, "Distributed Video Coding with Zero Motion Skip and Efficient DCT Coefficient Encoding", International conference on *Multimedia & Expo*, Hannover, Germany, Jun. 2008.

[13] S. Pateux and J. Jung, "An excel add-in for computing Bjøntegaard metric and its evolution", ITU-T SG16 Q.6 Document, VCEG-AE07, Marrakech, Jan. 2007.