Frame-based skip mode decision for distributed video coding

WANG Gang,YANG Gaobo,SONG Yun
School of Information Science and Engineering
Hunan University
Changsha, China
Email: wanggang@gmail.com,yanggaobo@hnu.edu.cn

LI Leida
School of Information and Electrical Engineering
China University of Mining and Technology
Xuzhou,China
Email: 592615158@qq.com

Abstract—**Distributed video coding (DVC) is a coding paradigm that can shift the computational expensive task from the encoder to the decoder. Though in theory this reverse of encoder and decoder complexity does not cause coding performance degradation compared with traditional hybrid video codecs, the gap between the two codecs is still quite high. One reason for the performance gap is the lack of rich coding modes for current DVC codecs. This paper proposes a frame-based skip mode decision algorithm. An effective encoder-side SI quality estimation technique is introduced to facilitate an accurate skip mode decision. With the proposed technique, the encoding and decoding time are reduced by about 9% and 50% respectively and the rate-distortion (R-D) performance is improved by an average of 0.49 dB compared with the well-known DISCOVER DVC codec.**

Index Terms—**Distributed video coding ; frame-based; Skip mode ; SI quality estimation**

I. Introduction

The emerging wireless video applications such as wireless video surveillance and wireless visual communication demand simple video encoders because of their low computational capability and limited power supply. Conventional hybrid video coding paradigms like H.264/AVC [1] are not suitable for such applications because good compression performance are achieved at the expense of high encoding complexity for such codecs. Distributed video coding (DVC) [2] is very promising for such applications because the most computation-intensive task, i.e., motion estimation, is shifted to the decoder side. In many DVC implementations, an input video sequence is divided into two type of sub-sequences, namely key frames and wyner-ziv(WZ) frames. Key frames are coded using intra coding technique and two successive decoded key frames are interpolated in order to predict the WZ frame between them. Then the predicted WZ frame, which is called side information (SI), is iteratively corrected by channel codes.

Though much effort has been done to improve the coding performance of DVC, currently available DVC codecs still can't achieve the theoretical coding performance proved in [3] and [4]. The lack of a rich set of intra- and inter-prediction modes as traditional video codecs is one important reason [5]. Since low or medium-motion video sequences are most common for DVC applications, skip mode decision has attracted extensive attentions.

Current skip mode decision algorithms [5-8] usually performs skip mode decision at block level or bitplane level. For block-based skip mode decision algorithms, SAD or MSE values between collocated macroblocks of SI and the original WZ frame are estimated using successive decoded key frames or one original key frame and its next key frames. The calculated SAD or MSE values act as the distortion metric for the macroblocks in SI. Only the macroblocks with distortion value lower than a preset threshold are skipped. The method is quite simple and is very effective for video sequences with static backgrounds. However, for video sequences with global motions, the undistorted macroblocks of SI can not be efficiently skipped. For bitplane-based skip mode decision algorithms like the one in [8], the correctness of each bitplane are estimated using a rate-distortion model. The bitplane with the distortion-rate ratio lower than a preset threshold are skipped. However, skipping the entire bitplane usually can't adapt to the spatially varying quality of SI. Therefore, distortions usually happens for video sequences even with medium motion. Also, the bitplane-based skip mode decision methods are much more complex than block-based skip modes decision methods.

In this paper, a frame-based skip mode decision algorithm is proposed. Specifically, a simple quality estimation technique is introduced to provide a more accurate information of the perceptual quality of SI. With the estimated SI quality, the SI with good visual quality can be skipped without coding to save bit rate consumption. In addition, the proposed frame-based skip mode decision method is effective for video sequences with or without global motions.

The remainder of this paper is organized as follows. Section 2 briefly introduces the proposed DVC framework. Experimental results are given in Section 3. We conclude this paper in Section 4.

II. The proposed DVC framework

The architecture of the proposed DVC codec is shown in Fig. 1. The proposed codec is different from DISCOVER DVC codec mainly in the frame mode decision module. In the proposed codec, the frame mode decision module determines whether a WZ frame can be skipped or not. If a WZ frame cannot be skipped, then the WZ frame will be WZ coded as
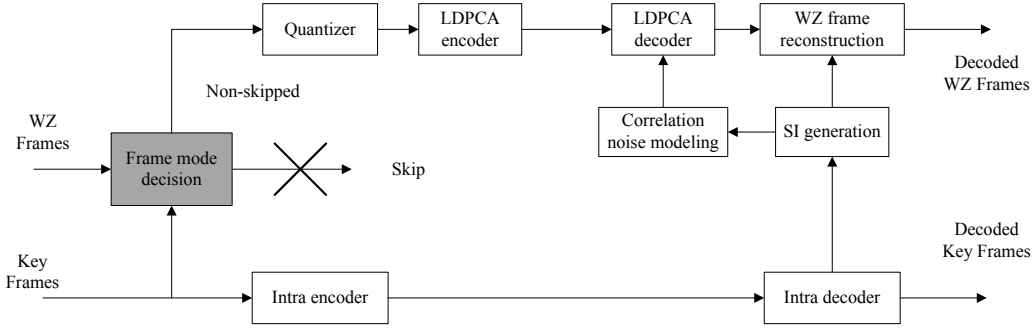
Fig. 1. Proposed DVC framework

in DISCOVER codec. The rest of this section will discuss in detail the proposed frame-based skip mode decision technique.

A. SI quality estimation

In many DVC implementations, SI is usually generated by motion-projection using two neighboring decoded key frames and interpolation errors usually happens when high motion exists. In order to reduce coding delay, the proposed SI quality estimation process is performed at the encoder side. Without the SI available, the distortion of SI is estimated using the successive key frames. Since in wireless visual communication applications where DVC may find most of its applications, the visually sensitive objects usually locates near the central regions. Therefore, a Gaussian filter that favors the distortion in central regions can be integrated to better determine visual quality os SI. The Gaussian filter used in this paper is defined as follows:

$$G(i,j) = \frac{1}{2\pi\sigma_i\sigma_j} \exp(-\frac{1}{2}(\frac{(i-u_i)^2}{\sigma_i^2} + \frac{(j-u_j)^2}{\sigma_j^2})) \quad (1)$$

Where i and j are the row and column indices of a macroblock, respectively. u_i, u_j, σ_i and σ_j are the means and standard deviations of i and j, respectively. Suppose H and W denotes the height and width of a frame measuring in macroblocks. Then σ_i and σ_j are assigned as $\frac{H}{4}$ and $\frac{W}{4}$, respectively. u_i and u_j are set to be the central coordinate of a frame, i.e., $\frac{H}{2}$ and $\frac{W}{2}$, respectively.

Suppose $M(i,j)$ denotes the MSE value between the collocated macroblock of successive keys frame at (i,j). In our experiment, we find that the sum of the mean and standard deviation of the Gaussian filtered M values can quite effective in measuring the visual quality of SI for a WZ frame. Suppose $E(i,j)$ denotes the Gaussian filtered MSE value of the macroblock at (i,j). Then $E(i,j)$ can be defined as follows:

$$E(i,j) = M(i,j) \times G(i,j) \quad (2)$$

where G(i,j) is defined in (1). Suppose $\mu(n)$ and $\sigma(n)$ denote the mean and standard deviation value of $E(.)$ over the nth frame, respectively. If we denote the sum of the mean and
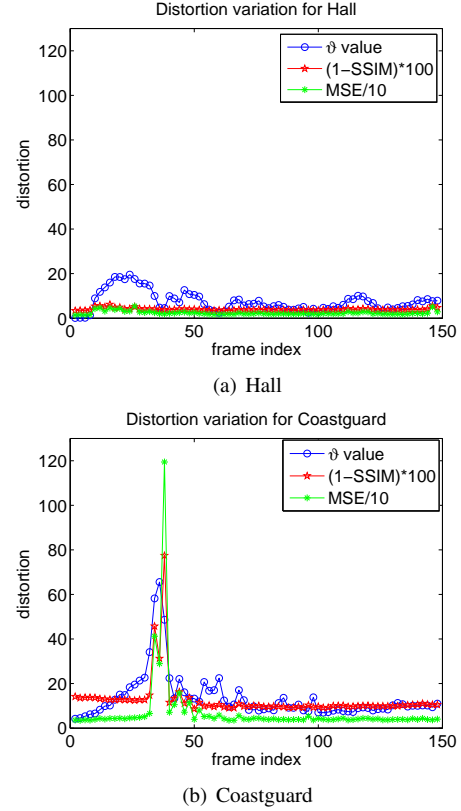

(a) Hall

(b) Coastguard

Fig. 2. The relations between ϑ values and SI distortions for WZ frames

standard deviation of E value for the nth WZ frame as $\vartheta(n)$, then we can obtain $\vartheta(n)$ as follows:

$$\vartheta(n) = \mu(n) + \sigma(n) \quad (3)$$

The reason for the effectiveness of ϑ values in measuring SI qualities is that high motions in the central regions could cause $\mu(n)$ to increase while camera panning causes $\sigma(n)$ to increase. Therefore, the sum of the two values can efficiently capture the high-motion frames and the frames at the camera panning where SI quality is poor. In order to show the efficiency of ϑ value in measuring SI qualities, we illustrates the relations between SI distortion and the ϑ values for all the WZ frames of Hall and Coastgaurd sequence in Fig. 2. Apart

from the ϑ values for WZ frames, both structural distortion and MSE values between the original WZ frame and SI are given in the figure. Structural distortion is denoted as $(1 - SSIM)$ and the structural similarity (SSIM) index is defined in [10]. In the figure, the $(1 - SSIM)$ value of each SI is multiplied by 100 and the MSE value of each SI is divided by 10. From Fig. 2, we can see that low ϑ values always correspond to good SI qualities while high ϑ values usually means poor SI qualities.

B. Frame-based skip mode decision

In our experiments, we find that performing SI quality estimation and skip a whole WZ frame may cause distortion for the case where part of an object suddenly appears in the scene at static video surveillance applications. When part of an object suddenly enters the scene, the SI quality is poor only in limited regions. Therefore, the overall quality of SI is quite good. However, the distortions in these limited regions are quite annoying. In order to solve this problem, video contents are analyzed using the encoder-side MSE-based skip mode decision method proposed in [5]. The difference is that successive key frames are used in this process and the threshold to perform skip mode decision are set to be 40. The generated skip map are referred to as block mode map. If the percentage of skipped macroblocks in the block mode map is higher than a pre-defined threshold θ_{skip}, performing visual distortion analysis might cause severe visual distortion. In this case, the WZ frame is not skipped unless all the macroblocks can be skipped in the block mode map.

When SI quality needs to be estimated to determine the WZ frame coding mode, the threshold θ_{low} is used to determine whether the SI quality is good enough to be skipped. Suppose $skip(n)$ denotes the coding mode of the nth WZ frame. If the ϑ value of the nth WZ frame is lower than θ_{low} or equal to θ_{low}, the WZ frame is skipped. And if $\vartheta(n)$ is higher than θ_{low}, the WZ frame is WZ coded. The process can be defined as follows:

$$\text{skip(n)} = \begin{cases} 1, & \vartheta(n) \le \theta_{low} \\ 0, & \vartheta(n) > \theta_{low} \end{cases} \quad (4)$$

where $\vartheta(n)$ is defined in (3), $skip(n)$ is the coding mode of the nth WZ frame.

III. EXPERIMENT RESULTS AND ANALYSIS

This section presents experimental results to demonstrate the efficiency of the proposed technique. Since DVC finds most of its applications in wireless surveillance and wireless visual communication, three common test sequences used in DISCOVER DVC codec [9], i.e., *Foreman*, *Hall*, *Coastguard*, are tested to represent the above-mentioned applications. All the test sequences are of QCIF format (15Hz). The first 149 frames of the test sequences are coded. The quantization matrix at different rate-distortion (R-D) points (Q1-Q7) are the same as in DISCOVER DVC codec. The experiments are are running on a personal computer with Intel Core2 2.7 GHz CPU and 2 GB RAM. The DISCOVER codec is directly obtained

TABLE I
PERCENTAGE OF SKIPPED WZ FRAMES WHEN θ_{low}=10 AND THE INFLUENCE OF THE SKIPPED WZ FRAMES AT Q_6

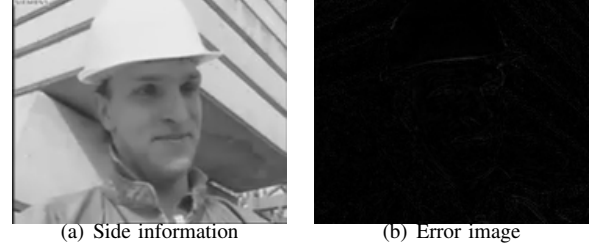| (Sequences)	Frame quality(SSIM)		%Skipped WZ frames
	Key frame	skipped SI	
Hall	0.963	0.962	62
Coastguard	0.904	0.895	49
Foreman	0.924	0.911	30

(a) Side information   (b) Error image

Fig. 4. One example of skiped WZ frame

from the DISCOVER website [11]. The threshold θ_{skip} is set to be 90% and the threshold θ_{low} is set to be 10. The rational of the threshold θ_{low} can be seen in Table 1. From the table, we can see that the SI quality of skipped WZ frames can keep roughly constant with key frames when measuring by SSIM. The coding results for the 60th frame of Foreman is given as an example to illustrate the rational to skip some WZ frames with good qualities. More experimental results for the proposed codec and the comparison results with DISCOVER codec are uploaded to [12] to facilitate readers' understanding.

A. R-D performance comparison

R-D performance is usually used to measure the coding performance of different codecs. Fig. 3 lists the R-D performance of the proposed codec and several other codecs. From the figure, we can see that the proposed codec can achieve constantly better R-D performance compared with the well-known DISCOVER DVC codec. For Foreman sequence, the BD-PSNR [13] gain of the proposed codec over DISCOVER codec is 0.27 dB. For Coastguard sequence and Hall sequence, the BD-PSNR gain of the proposed technique over DISCOVER codec are 0.38 dB and 0.81 dB, respectively. The average BD-PSNR gain of the proposed codec over DISCOVER codec is about 0.49 dB. In addition, we can see from Table 1 that the percentage of skipped WZ frames can be quite effective in reflecting the quality of SI. Hall sequence is known to be a static surveillance video sequence and it is known to have good SI quality. Therefore, the percentage of skipped WZ frames is the highest in theory. The experimental results of the proposed codec closely follows the quality of SI. Also, we can see that the higher the percentage of skipped WZ frames, the higher is the BD-PSNR gain. It is another evidence to show the efficiency of the proposed frame-based skip mode decision method in improving coding efficiency.
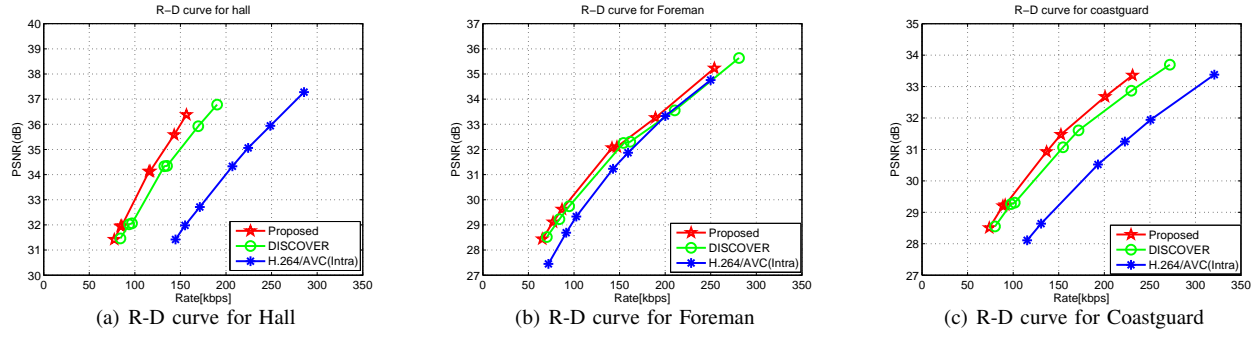
Fig. 3. R-D performance comparisons for different codecs

B. Coding time comparison

Since a high percentage of WZ frames is skipped, the encoding time and decoding time are also reduced. Table 2 lists the encoding and decoding time of both the proposed codec and DISCOVER codec at different R-D points. In the table, QP denotes quantization parameter for key frames and QM denotes quantization matrix for WZ frames defined in DISCOVER codec. The encoding and decoding time for the proposed codec and that of DISCOVER are collected by coding 149 frames of Hall sequence. From the table we can see that,with the proposed frame-based skip mode decision algorithm, the encoding time reduces by about 9% and the decoding time reduces by around 50% for Hall sequence compared with the well-known DISCOVER DVC codec. The frame-based skip mode decision process only cost about 1 ms to deal with a single frame and it can be negligible when compared with the encoding time reduction. Note that the time cost to perform the frame-based skip mode decision process is also included in the coding time.

IV. CONCLUSIONS

In this paper, we proposed a frame-based skip mode decision algorithm for DVC. In the proposed framework, an effective encoder-side SI quality estimation technique is introduced. The WZ frames with good perceptual quality are skipped. With the proposed frame-based skip mode decision method, the R-D distortion performance is improved and both the encoding and decoding time are reduced. This makes the proposed codec more suitable for wireless transmission applications.

ACKNOWLEDGEMENTS

TABLE II
CODING TIME COMPARISON WITH DISCOVER DVC CODEC

(QP,QM)	Encoding time(ms)		Decoding time(ms)	
	Proposed	DISCOVER	Proposed	DISCOVER
(37,1)	8420	8859	1.2×10^5	2.5×10^5
(36,3)	8434	8899	1.5×10^5	3.1×10^5
(33,5)	8441	8917	1.9×10^5	3.9×10^5

[2] Dufaux F., Gao W., Tubaro S., Vetro A., "Distributed video coding: trends and perspectives", Journal on Image and Video Processing, 10, pp. 1-10, 2009.

[3] SLEPIAN D., WOLF J., "Noiseless coding of correlated information sources", IEEE Transactions on Information Theory, 19(4), pp. 471-480, 1973

[4] WYNER A., ZIV J., "The rate-distortion function for source coding with side information at the decoder", IEEE Transactions on Information Theory, 22(1), pp. 1-10, 1976.

[5] Mys S., Slowack J., Škorupa J., Deligiannis N., Lambert P., Munteanu A., Van de Walle R., "Decoder-driven mode decision in a block-based distributed video codec", Multimedia tools and applications, 58(1), pp. 239-266, 2012.

[6] Chiou B. R., Shen Y. C., Cheng H. P., Wu J. L., "Performance improvement of distributed video coding by using block mode selection", ACM Proceedings of the international conference on Multimedia, pp. 1207-1210, October 2010.

[7] Feng Y., Li Y., Wu C., Song R., "Coding scheme with skip mode based on motion filed detection for dvc", International Society for Optics and Photonics, pp. 70840M-70840M, August 2008.

[8] Chien W. J., Karam L. J., "BLAST-DVC: BitpLAne SelecTive distributed video coding", Multimedia Tools and Applications, 48(3), pp. 437-456, 2010

[9] Artigas X., Ascenso J., Dalai M., Klomp S., Kubasov D., Ouaret M., "The DISCOVER Codec: architecture, techniques and evaluation", Picture Coding Symposium, Lisbon, Portugal, pp. 1-6, November 2007.

[10] Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P., "Image quality assessment: From error visibility to structural similarity", IEEE Transactions on Image Processing, 10(4), pp. 600-612, 2004.

[11] DISCOVER codec, http://www.discoverdvc.org/, accessed 24 june, 2013.

[12] Proposed codec, https://www.dropbox.com/sh/hg2905d71u0ger5/M788E1XrBZ, accessed 24 April, 2014.

[13] Bjøntegaard G., "Improvements of the BD-PSNR model", ITU-T SG16 Q.6 Document, VCEG-AI11, Berlin, July 2008.

REFERENCES

[1] Wiegand T., Sullivan G. J., Bjontegaard G., Luthra A., "Overview of the H. 264/AVC video coding standard", IEEE Transactions on Circuits and Systems for Video Technology, 10(7), pp. 560-576, 2003.