IEICE TRANS. INF. & SYST., VOL.E93–D, NO.3 MARCH 2010

LETTER

Distributed Video Coding with Multiple Side Information Sets

Jung Ah PARK[†a)], **Doug Young SUH**[†], ***and*** **Gwang-Hoon PARK**[†], ***Members***

SUMMARY This letter proposes a method to retrieve the original image X out of multiple sets of SI (Side Information) in distributed video coding (DVC). Using Turbo decoding methods, the most reliable segments from the decoded Y_i's were selected for the composition of Y_∞, whose conditional entropy $H(X|Y_\infty)$ became much lower than any individual conditional entropy $H(X|Y_i)$. This proposal has improved the peak signal-to-noise ratio (PSNR) by 1.1 to 1.8 dB, compared to the conventional single SI-based approach.

key words: *DVC (Distributed Video Coding), multiple side information, Turbo coding*

1. Introduction

Pereira et al. [1] introduced emerging applications such as wireless video cameras and wireless low-power surveillance networks, disposable video cameras, certain medical applications, and sensor networks for which distributed video coding (DVC) is suitable. Video encoders should be lighter than decoders. Thus, a new paradigm is needed for video compression for these applications. DVC meets this need by removing the motion compensation process from the encoder and moving it to the decoder. This new approach is based on the Slepian-Wolf theorem [2] and the Lossy Distributed Coding Theory of Wyner and Ziv [3], which use the correlations of the original frame with a frame predicted at the decoder. This predicted frame is the side information (SI).

They used a single set of SI called Y and tried to generate Y as close to X (the original information) as possible. Other work generated SI from multiple reference frames [4]. Adikari et al. used multiple SI streams in [5], [6]. They enhanced the compression ratio by directly combining two SI streams. Based on the estimated error probability, the Turbo decoder decides which SI stream is used for decoding a given block. However, decoder-side estimation is not accurate and the block-wise approach results in discontinuities in the reconstructed image.

This letter proposes the use of multiple sets of SI and a novel method to combine information from them. In the previous papers [5], [6], the Turbo decoder requests parity bits from the encoder for the selected SI stream of the current bitstream block until it receives enough parity bits to decode the original systematic bits at a predetermined error probability (10^{-3}). However, in our proposed method, the decoder requests much less feedback from the encoder than before. Special Turbo coding is used to decide the best segments of bit sequences from multiple sets of SI.

This paper is organized as follows. Section 2 presents a brief summary of the Slepian-Wolf theorem. Section 3 introduces special Turbo coding and how to use it in deciding what are the most feasible bits of decoded SI. In Sect. 4, the performance of the proposed method is shown as results of experiments on popular video sequences. Section 5 concludes this letter and suggests ideas for future work.

2. The Foundation of Distributed Video Coding

Slepian and Wolf [2] showed that if a decoder has the SI Y that is correlated to the original information X, X can be retrieved by sending as much information as parity rate $R > H(X|Y)$, which is smaller than the entropy of the original information, $H(X)$. In a BSC (binary symmetric channel), in which the bit error rate (BER) is p, conditional entropy is

$$H(X|Y) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}. \quad (1)$$

The amount of information the BSC channel can deliver is called mutual information, where $H(X|Y)$ is information loss in the BSC. Since the decoder already has Y, X can be retrieved by sending a slightly larger R than $H(X|Y)$, as shown in Fig. 1.

In the DVC architecture shown in Fig. 2, Y is considered as X corrupted by bit errors, just as in the BSC. Instead

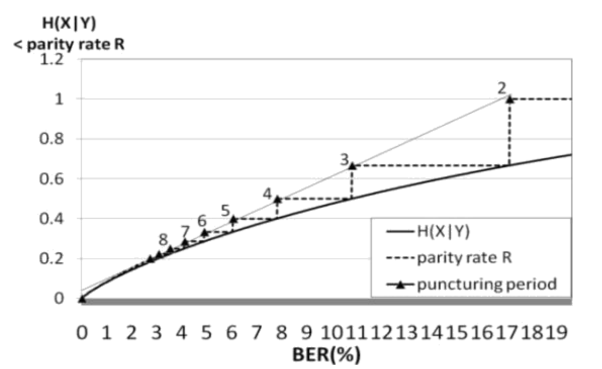


Fig. 1 Information loss $H(X|Y)$ in the BSC (M is the puncturing period of the Turbo code. For example, $M = 6$ means sending one bit at every 6 parity bits, resulting in a parity rate of 1/3.)

Manuscript received June 18, 2009.
Manuscript revised November 6, 2009.

[†]The authors are with Media Lab (National Research Lab in Korea), College of Electronics and Information, Kyung Hee University, 1, Seocheon, Kiheung, Yongin, Kyunggi, 449–701, Korea.
a) E-mail: kiwilang@khu.ac.kr

DOI: 10.1587/transinf.E93.D.654

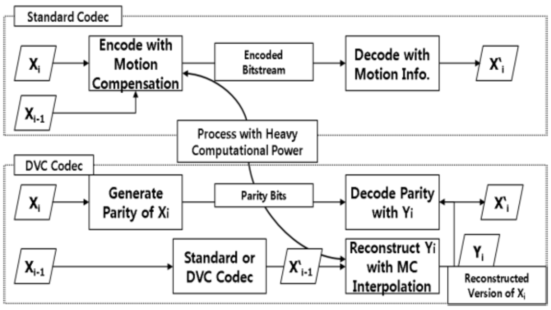


Fig. 2 DVC architecture.

of sending X, parity bits of X are sent to the decoder [7] and X is retrieved through an error correction process. Turbo code is mostly used as the channel code. SI Y is generated by using motion estimation between key frames, which are encoded by a conventional video codec. Since X is not known in the decoder, the motion estimation is unreliable.

3. DVC with Multiple Sets of SI

3.1 Multiple sets of SI

Suppose that more than one set of SI is used for decoding X. These sets will reduce the uncertainty of X so that the conditional entropy of multiple sets of SI will be less than those of fewer sets of SI. This idea is similar to the idea that a packet is sent three times through an error-prone channel and that bit errors can be corrected based on majority rule.

$$H(X|Y_1) \geq H(X|Y_1Y_2) \geq H(X|Y_1Y_2Y_3) \geq \cdots \tag{2}$$

Even though the number of transmitted bits $R < H(X|Y_i)$ for every $Y_i \in Y$, where $Y = \{Y_1, Y_2, \ldots, Y_n\}$ using X_i''s which are pre-decoded from Y_i's, one can compose Y_∞, which satisfies $R > H(X|Y_\infty)$. Without increasing transmission parity rate R, X can then be retrieved at the expense of consuming more computational power. Let X_i' be the pre-decoded Y_i. In order to make Y_∞ out of X_i''s, the most reliable bits in X_i''s should be selected in the best bit selector, shown in Fig. 3. However, because X is not known in the decoder, it is impossible to differentiate correct bits from error bits.

We noticed that during iterations of Turbo decoding, correct bits remained unchanged after a few iterations, while incorrect bits were likely to keep 'toggling.' The number of toggles of each decoded bit was roughly inversely proportional to its reliability, as shown in Fig. 4. It shows that the total number of toggles of an SI was proportional to the MSE (mean square error) between the original and predicted images. By selecting the most reliable bit from n number of candidates at every bit location, one can compose more reliable SI Y_∞, whose information loss $H(X|Y_\infty)$ would be less than any other $H(X|Y_i)$ and could be less than R, making more reliable decoding possible.

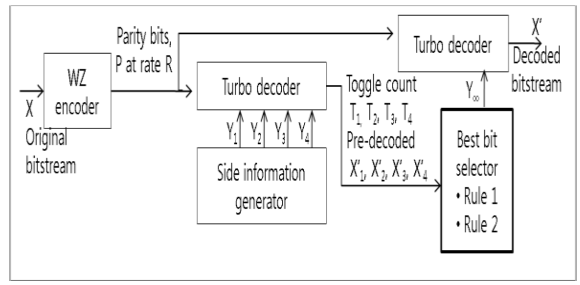


Fig. 3 Turbo decoding with multiple candidates.

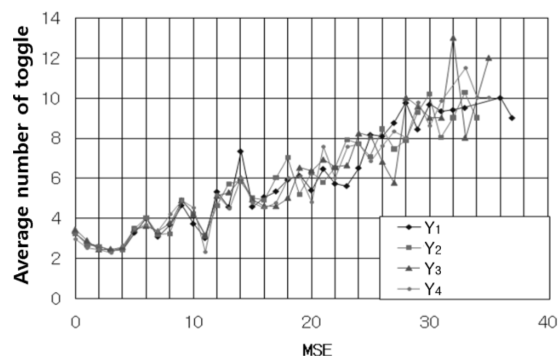


Fig. 4 Number of toggles and mean square error (MSE).

3.2 DVC with Multiple Sets of SI

The decoder initially corrected errors in Y_i's one bitstream using transmitted parity bits. During iterations of error correcting for normal and interleaved bitstreams, the number of toggles was counted at each bit. If the smallest numbers of more than one candidate are the same, two rules are used to make a decision.

- Rule 1: Scanning left to right and top to bottom, select the bit value of X_i' whose accumulated number of toggles is the smallest at each bit position.
- Rule 2 : Select the bit value of X_i' whose number of toggles in the neighborhood is the smallest. Here, the range of neighborhood is the same as the puncturing period.

After decoding the most significant coefficients, such as DC and AC coefficients close to DC, the other AC coefficients were copied from the most reliable Y_i in every 8×8 block. The most reliable Y_i in every block was selected based on the MSE between the decoded coefficients of Y_∞ and those of Y_i. We realized that DC values were so dominant that it was enough to compare just the DC values of each block.

4. Results and Discussion

The proposed approach was applied to video coding. In the experiment, the five MSB's of the DC values in the macroblocks of the Wyner-Ziv (WZ) frame were rearranged as 5 bit-planes. Since quarter common intermediate format (QCIF) video sequences were used, a bit-plane was 396 bits long, that is, one bit per each of 396 8×8 blocks. Video sequences of different characteristics were used, such as 'Foreman', 'Football', and 'Mobile'.

Two parity bitstreams, one for the normal X bitstream and the other for the interleaved X bitstream, were calculated for original data X that was 5×396 bits long using the Turbo encoder. The bitstreams were punctured by puncturing term M bits so that $R = 2 \times 5 \times 396/\mathrm{M}$ bits were sent to the decoder. The puncturing index M was determined to ensure that R was larger than $H(X|Y)$ and was large enough to decode X out of Y. As shown in Fig. 1, M was recommended to be six at a bit error rate of 5%, at which $H(X|Y) = 0.2864 < R = 0.3333$.

As shown in Table 1, the residual BER's of X_i' were high because the parity rate R was not great enough. It was even smaller than $H(X|Y)$ when $M = 7$. By selecting the most reliable bits, the residual BER of Y_∞ became quite low, as shown in Table 2. Rule 2 exploits local reliability and resulted in a better performance.

The same experiment at a BER 5% was performed, increasing R from 10% to 50% of $H(X)$. If only Y_1 was used, 40% of $H(X)$ should have been transmitted for clear decoding, as shown in Fig. 5. The ideal amount of R in Fig. 1 at a BER of 5% was more than 33% since $M = 6$. If three sets of SI, such as Y_1, Y_2, and Y_3 were used, a 1% error was corrected by Rule 1 or 2 for the three sets. Then, bitrate R, equal to as much as 25% of $H(X)$, could result in clear decoding, as shown in Fig. 5. This is compared to the ideal amount of 29% (= 2/7) at a BER of 5%, shown in Fig. 1. If four sets of SI, such as Y_1, Y_2, Y_3, and Y_4 were used, the starting BER became 3% and the required rate for clear decoding was 25% (Fig. 5) while the ideal rate was 20% at $M = 9$. Rate saving gains of the proposed method were about 25% (40%==>30%, $(40 - 30)/40 = 0.25$) and 38% (40%==>25%) for three and four sets of SI compared to a single set of SI.

In order to evaluate the performance of our proposed approach, we used multiple sets of side information generated using hierarchical ME [8] from 28 dB to 31 dB using three sequences ('Foreman' : 10 frames, 'City' : 10 frames, 'Crew' : 10 frames). The average BER was 5%~15% and the transmission rate R was 2/7~2/4 of $H(X)$ in all sequences. Figure 6 depicts the PSNRs of Y_∞ and X' of 'City' sequence. Table 3 shows the average PSNRs of three sequences, in which (a) and (b) show PSNRs of Y_∞ and X', respectively. The final X improved compared to X_i's decoded by using single SI in all three sequences by 1.3 dB, 1.1 dB, and 1.8 dB, respectively.

Table 1 BER's of Y_i's and X_i''s in Fig. 4.

Parity rate	BER	Y_1	Y_2	Y_3	Y_4
M=6 (1/3)	BER(X,Yi)	0.062	0.045	0.049	0.043
	BER(X,Xi')	0.059	0.038	0.040	0.029
M=7 (2/7)	BER(X,Yi)	0.062	0.045	0.049	0.043
	BER(X,Xi')	0.056	0.026	0.035	0.036

Table 2 BER's of Y_∞ and X' in Fig. 4.

Parity rate	BER	No Rule	Rule 1	Rule 2
M=6 (1/3)	BER(X,Y∞)	0.032	0.032	0.018
	BER(X,X')	0	0	0
M=7 (2/7)	BER(X,Y∞)	0.021	0.013	0.011
	BER(X,X')	0.009	0.007	0.005

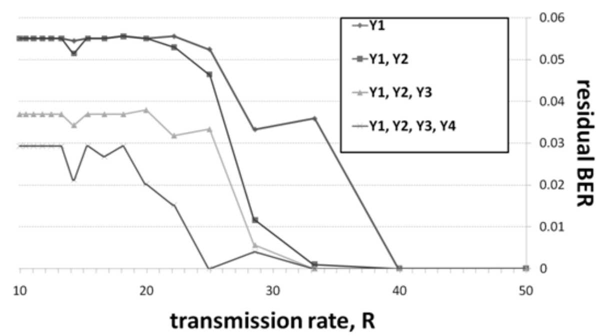
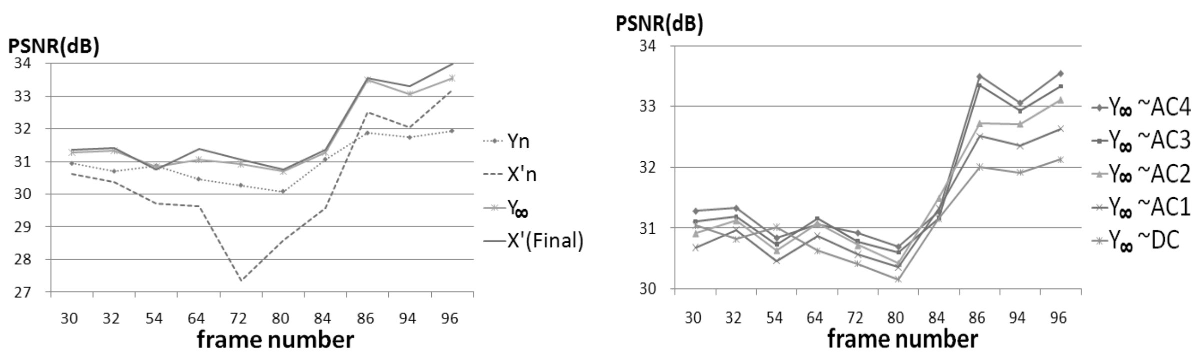


Fig. 5 Transmission rate R and residual BER of multiple sets of SI (original BER= 0.05).



(a) With proposed technique

(b) With DCT coefficient

Fig. 6 Performance comparison of the proposed technique ('City').

Table 3 Performance comparison of the proposed technique.

Average	Y_n	X'_n	Y_∞	X'	$X'-X'_n$
DC	31.0	30.3	31.1	31.3	**+1.0**
DC~AC1		30.2	31.3	31.4	**+1.2**
DC~AC2		30.2	32.5	31.6	**+1.4**
DC~AC3		30.3	31.6	31.8	**+1.5**
DC~AC4		30.4	31.8	31.9	**+1.5**

(a) City

Average	Y_n	X'_n	Y_∞	X'	$X'-X'_n$
DC	31.4	31.0	31.7	31.8	**+0.8**
DC~AC1		31.0	31.8	32.0	**+1.0**
DC~AC2		31.0	31.8	32.0	**+1.0**
DC~AC3		30.6	31.8	32.0	**+1.4**
DC~AC4		30.6	31.9	32.0	**+1.4**

(b) Foreman

Average	Y_n	X'_n	Y_∞	X'	$X'-X'_n$
DC	30.6	29.1	30.4	30.6	**+1.5**
DC~AC1		28.9	30.4	30.5	**+1.6**
DC~AC2		28.8	30.6	30.7	**+1.9**
DC~AC3		28.8	30.8	30.9	**+2.1**
DC~AC4		28.8	30.8	30.9	**+2.1**

(c) Crew

5. Conclusion

This letter proposed a method for improving the performance of DVC without increasing the transmission rate. Instead of using a single set of SI, multiple sets of SI were used. These could be generated using various kinds of image interpolation techniques. In conventional video codecs, multiple reference frames are used for motion estimation, and the closest block in one of the frames is selected for each block of the image to be encoded.

Similarly, instead of calculating the SAD (sum of absolute differences), the number of toggles at each bit is counted as iterating normal decoding and interleaved decoding in the Turbo decoding process. It is shown that the number of toggles is highly correlated to the correctness of every decoded bit. The ultimate information loss $H(X|Y_\infty)$ and the transmission rate R will get smaller and smaller as the diversity of SI increases.

The proposed method improved the quality of the decoded image by $1-2$ dB compared to conventional DVC that uses a single set of SI. If the resulting quality is set to be the same, the proposed method would reduce the transmission rate by 38%.

In this letter, we applied our approach to DC values. The approach may also be applied to AC values. In that case, the knowledge accumulated during the decoding of DC values may be used to select reliable bits. It is important to develop rules for the selection. Some reasoning techniques such as the fuzzy theory may be helpful for that purpose.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) through the National Research Lab (No. R0A-2005-000-10061-0 (2009)).

References

[1] F. Pereira, P. Correia, E. Acosta, L. Torres, C. Guillemot, M. Ouaret, F. Dufaux, T. Ebrahimi, R. Leonardi, M. Dalai, and S. Klomp, "Distributed coding for video services, d-4," project of the IST FET program of the European Union within the FP6/2002/IST/C Call, March 2006.

[2] J.D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. Inf. Theory, vol.IT-19, no.4, pp.471–480, July 1973.

[3] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inf. Theory, vol.IT-22, no.1, pp.1–10, 1976.

[4] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/MPEG-4 AVC standard," IEEE Trans. Circuits Syst. Video Technol., vol.17, no.9, pp.1174–1185, Sept. 2007.

[5] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, and H.K. Arachchi, "Sequential motion estimation using luminance and chrominance information for distributed video coding of Wyner-Ziv frames." Electron. Lett., vol.42, no.7, pp.398–399, March 2006.

[6] A.B.B. Adikari, W.A.C. Fernando, and W.A.R.J. Weerakkody, "Multiple side information streams for distributed video coding," IET Electron. Lett., vol.42, no.25, pp.1447–1449, March 2006.

[7] R. Puri and K. Ramchandran, "PRISM: An uplink-friendly multimedia coding paradigm," International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, April 2003.

[8] K. Min, S. Park, and D. Sim, "Side information generating using adaptive search range for distributed video coding," 2009 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Canada, Aug. 2009.