

Xiaoran HAO, Anni CAI, Bojin ZHUANG

A new encoder rate-allocation method for distributed video coding

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2011

Abstract Distributed video coding (DVC) is a new video coding architecture. Compared with traditional video coding schemes, DVC has a simple encoder and complex decoder, which makes it suitable for circumstances where the encoder equipments are simple but the decoder equipments are complex. Most of the existing DVC architectures use decoder rate allocation method with the help of feedback channel. According to the results of the current decoding round, the decoder informs the encoder by feedback channel whether more parity bits are needed. The use of feedback channel not only increases the system delay, but also limits the use of DVC to instances where there is no feedback channel. In this paper, we propose a novel encoder rate allocation method. First, the simple three-step motion estimation is introduced into the encoder to estimate the side information of the decoder, and then the number of parity bits the decoder needs for each bit-plane is allocated at the encoder according to the difference of the estimated side information and the current Wyner-Ziv (WZ) frame. Experiment results indicate that the accuracy of the proposed method is 5.18%–52.93% higher than that of the method proposed by Morbee.

Keywords distributed video coding (DVC), encoder rate allocation, video coding

1 Introduction

In conventional video coding schemes, e.g., Moving Picture Experts Group (MPEG)/H.26X, the encoder is much more complex than the decoder due to heavy computations in motion estimation and compensation.

Received January 28, 2010; accepted December 1, 2010

Xiaoran HAO (✉), Anni CAI, Bojin ZHUANG
Beijing Key Laboratory of Network System and Network Culture,
Beijing University of Posts and Telecommunications, Beijing 100876,
China
E-mail: haoxiaoran@yahoo.com.cn

This asymmetry in complexity fits well scenarios where a video signal needs to be compressed once but decoded many times, e.g., in broadcasting or streaming video on demand. However, for some scenarios, for example, in wireless low-power video surveillance or wireless video camera systems, the encoders have to be simple and power restricted, but the central node, e.g., the server of surveillance center, has high computing capability and infinite power. At such circumstances a coding approach with simple encoder and complex decoder is desired. Distributed video coding (DVC) is a video coding approach for scenarios described above. In DVC paradigm, the decoder performs predictive coding to exploit the similarities of successive frames, which makes a simple encoder and a complex decoder.

The first framework of DVC was proposed by Aaron [1]. In this framework, video frames are organized into two sets: key frames and Wyner-Ziv (WZ) frames. Key frames are coded in intra-frame mode while WZ frames are coded with the aid of a feedback channel in the following way: The decoder performs motion estimation and motion compensation and decides whether the received bits of the WZ frame are enough for successfully decoding. If not, the decoder will request more parity bits from the encoder via feedback channel. To decode one WZ frame, we need repetitious feedback and Turbo decodings, which not only increase system delay and decoding complexity, but also restrict its use in scenarios without feedback channel. To resolve this problem, Ref. [2] proposed a rate allocation method to avoid use of the feedback channel. In this method the encoder decides how many parity bits the decoder needs and sends the required bits to the decoder. In encoder the residual image between the current WZ frame and the mean value of the previous and latter key frames of the WZ frame is used to estimate how many parity bits the decoder needs. However, no corresponding experiment results were given in this paper. Reference [3] proposed another encoder rate allocation method. In this method, the parameter of Laplacian distribution is first estimated at the encoder, and then the bit error rate of each bit-plane at the

decoder is predicted. Finally, the encoder determines the number of parity bits the decoder needs based on the predicted bit error rate. From the experimental results of this paper we can see that the accuracy of this rate allocation method is not satisfactory.

In this paper, we propose a new encoder rate allocation method. This method introduces simple motion estimation, the three-step search method, into the encoder, and based on the current WZ frame, the encoder can estimate the number of parity bits needed at the decoder. To enhance the error correction capability of the Turbo coder, we also added an interleaver and a de-interleaver before the first system convolution code and before the first single input single output (SISO) decoder [4] respectively.

This paper is organized as follows. Section 2 introduces the DVC scheme with feedback channel. Section 3 explains the proposed encoder rate allocation method in detail. Section 4 gives the experimental comparisons between the proposed method and Ref. [3]. Conclusions are presented in Sect. 5.

2 DVC framework based on feedback channel

The DVC framework based on feedback channel is shown in Fig. 1 [1]. X_F , X_B respectively represent the previous and the latter key frames of the current WZ frame.

Key frames are coded in intra-frame mode and are used to produce the side information of WZ frames by frame interpolation. WZ frames are quantized using a 2^M -level scalar quantizer, $M \in \{1, 2, 3, 4\}$, and then the M bit-planes are encoded separately using a Turbo encoder. The parity bits produced by the Turbo encoder are stored in a buffer, and the systematic bits are discarded. To decode each bit-plane, the buffer sends a small part of parity bits of the current bit-plane to the decoder. After Turbo decoding at the decoder, if the bit error probability of this bit-plane

$P_e > 10^{-3}$ (it is assumed here that the decoder has ideal error detection capabilities), the decoder requests for more parity bits from the buffer via feedback channel. Otherwise, the Turbo decoding task of the current bit-plane is considered successful and another bit-plane starts being Turbo decoded.

We can see from the process of decoding that if there is no feedback channel, the decoding cannot be performed. To decode one WZ frame, it needs Turbo decoding many times. For example, for Foreman quarter common intermediate format (QCIF) sequence, with frame size $m \times n = 176 \times 144$ pixels and quantizer level $M = 4$, we need about 34 times of Turbo decoding [5] to decode one WZ frame of this sequence. Turbo decoding has high computational complexity which increases system delay and decoding computational complexity. If we can estimate how many parity bits the decoder needs at the encoder, the decoder only needs four times of Turbo decoding (we suppose $M = 4$) to decode one WZ frame.

3 A new encoder rate allocation method

At the decoder, how many parity bits the decoder needs mainly depends on the quality of side information. The better the side information, the less the parity bits needed. However, the encoder knows the current WZ frame, but does not know the side information. If we can estimate the side information at the encoder, we can then implement rate allocation at the encoder. Reference [6] introduced motion estimation and compensation with two-dimensional logarithmic search method into DVC, and achieved good performance at the cost of increasing the encoder complexity slightly. Enlightened by Ref. [6], in this paper, we employ motion estimation and compensation with the simple three-step search method to estimate the side information at the encoder.

At the encoder, after we obtain the estimation of the side

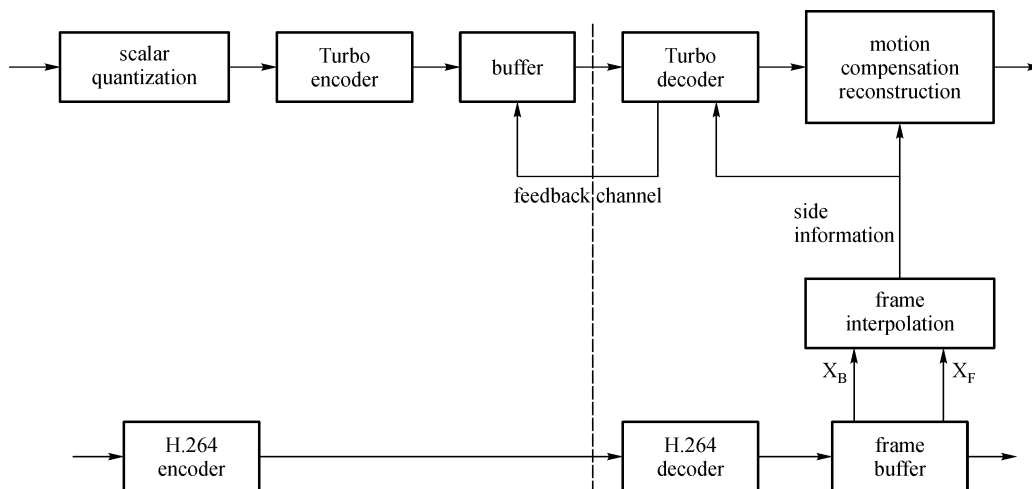


Fig. 1 DVC framework

information \widehat{S} and the parity bits of the current WZ frame, we can use the correlation between \widehat{S} and the current WZ frame to approximately estimate the difference between the current WZ frame and the side information produced at the decoder.

We first obtain the first two bit-planes of \widehat{S} and WZ frame respectively, then we count the total number of dissimilar bits between the i th ($i=1,2$) bit-plane of \widehat{S} and the i th ($i=1,2$) bit-plane of WZ frame. Also, we use this number to measure the bit-plane similarity between \widehat{S} and WZ frame. The larger the number is, the less similar the two bit-planes are.

To establish a statistical relation between the number of parity bits needed at the decoder and the number of dissimilar bits in corresponding bit-planes of \widehat{S} and WZ frame at the encoder, we performed many experiments for various QCIF video sequences. First, we obtain the actual number of parity bits N_{mm} needed at the decoder for the first two bit-planes, where m is the frame number and n is the index of the bit-plane, $n=1,2$. Then we compute the number of dissimilar bits Q_{mn} between the first two bit-planes of \widehat{S} and WZ frame at the encoder. The obtained relation between N and Q is shown in Fig. 2.

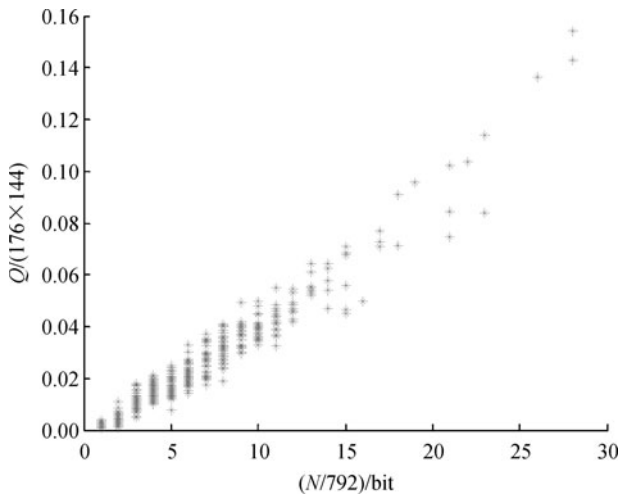


Fig. 2 Relation between N and Q

In our system the puncture period of the parity bits at the encoder is 32 bits, so that the allocated bit rate is integer multiples of 792 ($176 \times 144/32 = 792$). From Fig. 2 we can see that the larger Q is, the larger N also is. We want to determine a unique N for a given Q .

To get such a relation between N and Q , the experimental distributions of Q are examined at $N/792 = N_i$ ($N_i = 1, 2, \dots, 32$). Figures 3 and 4 show the distributions of Q at $N_i = 1, 2$ respectively.

From Figures 3 and 4 we can see that, for a certain N_i , the experimental distribution of Q approximately obeys the

Gaussian distribution. Figure 5 shows all the distributions at $N_i = 1, 2, \dots, 32$, where the x axis represents $q = Q/(176 \times 144)$, and $p_i(q)$ is the distribution of Q at N_i . Suppose that we divide q into M intervals, and assign the

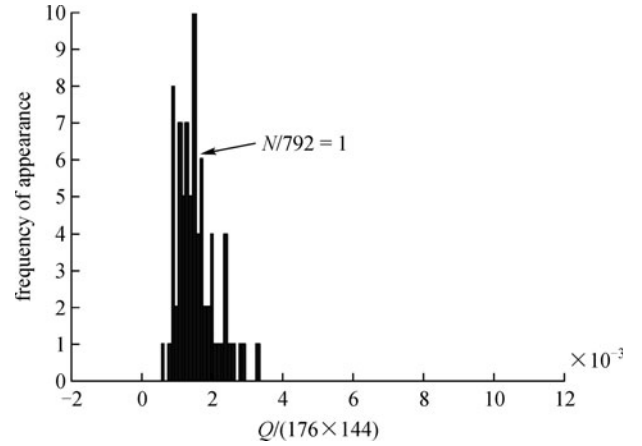


Fig. 3 $N/792 = 1$

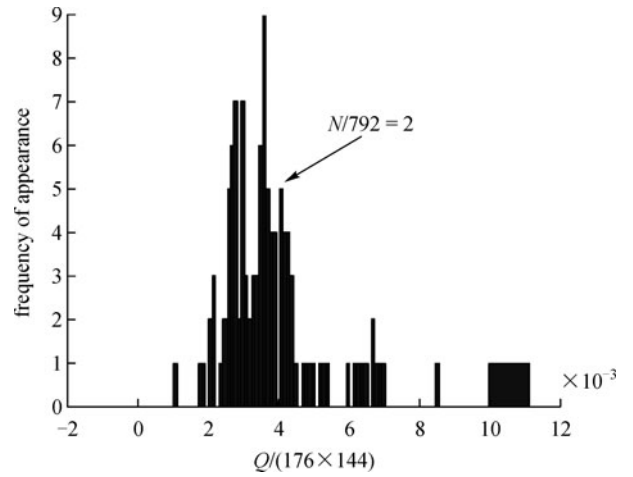


Fig. 4 $N/792 = 2$

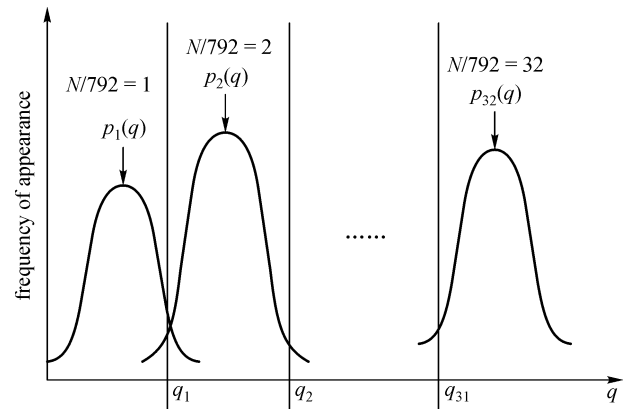


Fig. 5 Distribution of experimental data

number of allocated parity bits as N_i when q falls in interval $[q_{i-1}, q_i]$ ($i=1,2,\dots,M$). We determine the optimal edge points of the intervals q_i ($i=1,2,\dots,M$) through minimization of the total error rate:

$$\min \left[\sum_{j=i+1}^M \int_0^{q_i} p_j(q) dq + \sum_{j=1}^i \int_{q_i}^1 p_j(q) dq \right], \quad i = 1, 2, \dots, M. \quad (1)$$

In Eq. (1), the first term represents the total error rate produced by the probability distributions whose peaks are located at the left side of q_i , the second term represents the total error rate produced by the probability distributions whose peaks are located at the right side of q_i . Only $p_i(q)$ s close to q_i will dominate the precise location of q_i . Generally, q_i can be determined by 2–4 $p_i(q)$ s at both sides of q_i .

When $M=31$, the interval divisions and the needed number of parity bits of each interval are obtained and shown in Table 1.

Table 1 Interval division and the needed parity bits

$N/792$	q	$N/792$	q
1	(0.0000,0.0024]	17	(0.075,0.080]
2	(0.0024,0.0070]	18	(0.080,0.085]
3	(0.0070,0.0104]	19	(0.085,0.090]
4	(0.0104,0.0145]	20	(0.090,0.095]
5	(0.0145,0.0192]	21	(0.095,0.100]
6	(0.0192,0.0235]	22	(0.100,0.105]
7	(0.0235,0.0289]	23	(0.105,0.110]
8	(0.0289,0.0345]	24	(0.110,0.115]
9	(0.0345,0.0383]	25	(0.115,0.120]
10	(0.0383,0.0450]	26	(0.120,0.125]
11	(0.0450,0.0500]	27	(0.125,0.130]
12	(0.0500,0.0550]	28	(0.130,0.135]
13	(0.0550,0.0600]	29	(0.135,0.140]
14	(0.0600,0.0650]	30	(0.140,0.145]
15	(0.0650,0.0700]	31	(0.145,0.150]
16	(0.0700,0.0750]	32	(0.150,1.000]

Table 2 Performance comparison of two motion search methods

method	Foreman		Carphone		Salesman		Akiyo	
	PSNR	search step	PSNR	search step	PSNR	search step	PSNR	search step
three-step method	35.70	528	36.27	331	44.89	63	48.67	22
two-dimensional logarithmic method	34.75	457	36.35	325	44.79	49	48.56	16

4 Performance analysis

4.1 Performance comparison of two fast motion search methods

In this paper, we use the first 200 frames of four test sequences Foreman, Carphone, Salesman, and Akiyo. The performance comparison between three-step search method and two-dimensional logarithmic search method used in Ref. [6] is shown in Table 2.

From Table 2 we can see that the mean number of search step per frame of two-dimensional logarithmic search method is smaller than that of three-step search method for the four test sequences, and the peak signal-to-noise ratio (PSNR) of the estimated frames of three-step search method is higher than that of two-dimensional logarithmic search method for three of the four test sequences. Because the accuracy of encoder rate allocation depends on the accuracy of the estimated frames, we chose the three-step search method in this paper.

4.2 Side information estimation

The accuracy of encoder rate allocation depends on the accuracy of side information estimated at the encoder. If the estimated side information is close to the side information produced at the decoder, the number of allocated parity bits is close to the actually required at the decoder.

We use the four test sequences to compare the PSNRs of the estimated side information at the encoder and the side information produced at the decoder. The results are shown in Figs. 6–9 (here we suppose that group of picture (GOP) is 2 and the total number of WZ frames is 100).

From Figs. 6–9 we can see that for sequences with high motion, e.g., Foreman and Carphone, the PSNR of the estimated side information is close to that of the real side information. However, for some frames in the two sequences, the difference of PSNR between the estimated side information and the real side information is several dB because of high motion in these sequences, which brings down the accuracy of encoder rate allocation. How to improve the motion search method is the future research work of this paper. For sequences with low motion, e.g., Salesman and Akiyo, the two PSNR values are almost the same.

4.3 Encoder rate allocation

In order to compare our scheme to that in Ref. [3], we chose the same four test sequences Foreman, Carphone, Salesman, and Akiyo, the frame rate was 30 frames per second, and the WZ frame rate was 15 frames per second, to perform the experiments. The results are shown in Tables 3 and 4 for the first and second bit-planes

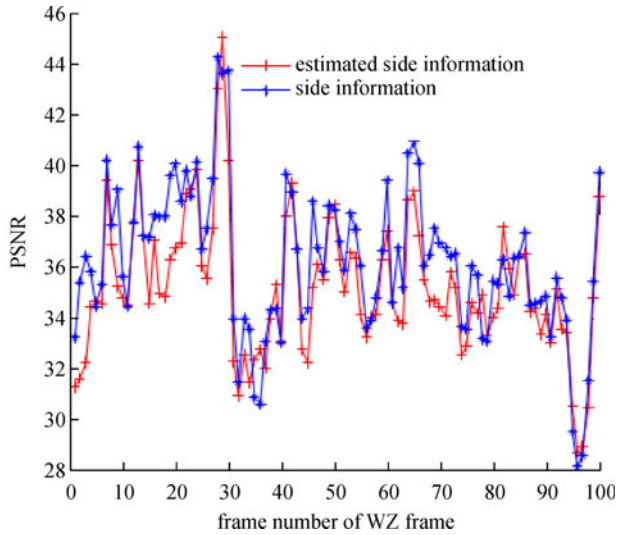


Fig. 6 Experiment result of Foreman

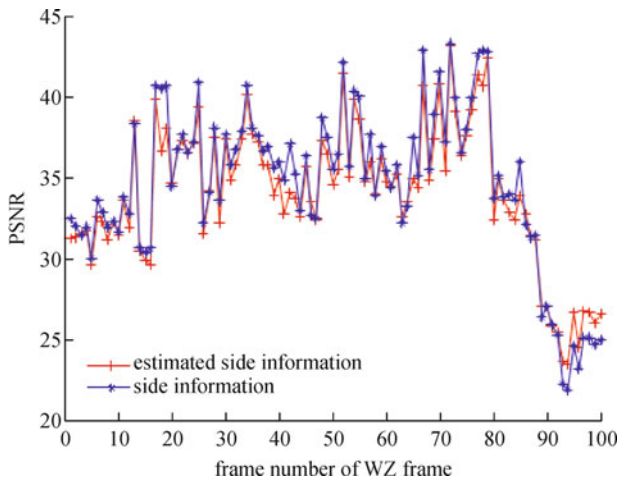


Fig. 7 Experiment result of Carphone

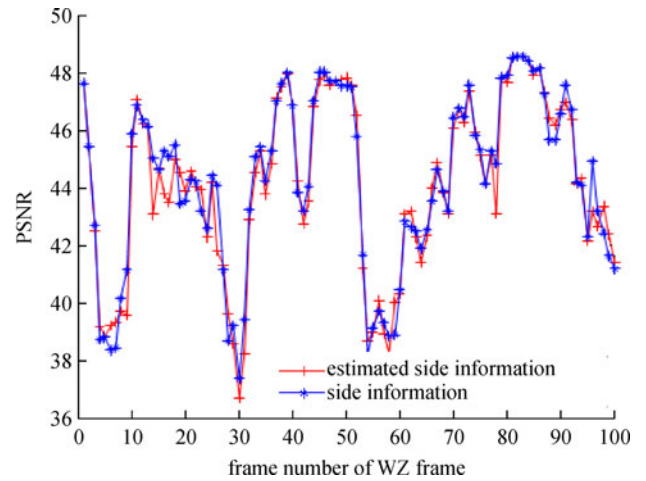


Fig. 8 Experiment result of Salesman

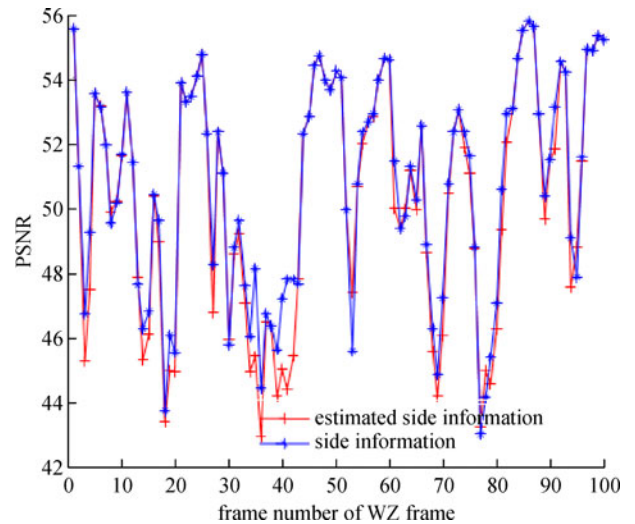


Fig. 9 Experiment result of Akiyo

Table 3 Comparison with Ref. [3] of the first bit-plane

sequence	method	frame number ratio/%				
		$C \leq -24$ kbit/s	$C = -12$ kbit/s	$C = 0$ kbit/s	$C = 12$ kbit/s	$C \geq 24$ kbit/s
Foreman	proposed method	0.00	1.01	28.28	54.55	15.15
	Ref. [3]	7.50	17.60	23.10	13.60	38.20
Carphone	proposed method	3.03	17.17	67.68	7.07	0.00
	Ref. [3]	7.40	10.10	23.50	34.90	24.20
Salesman	proposed method	0.00	3.03	94.95	2.02	0.00
	Ref. [3]	8.00	10.10	45.00	26.80	10.10
Akiyo	proposed method	0.00	5.05	89.90	5.05	0.00
	Ref. [3]	12.10	14.70	59.70	10.10	3.40

Table 4 Comparison with Ref. [3] of the second bit-plane

sequence	method	frame number ratio/%				
		$C \leq -24$ kbit/s	$C = -12$ kbit/s	$C = 0$ kbit/s	$C = 12$ kbit/s	$C \geq 24$ kbit/s
Foreman	proposed method	6.06	23.23	48.48	18.18	4.04
	Ref. [3]	0.00	1.50	12.60	10.50	75.40
Carphone	proposed method	4.04	16.16	38.38	28.28	8.08
	Ref. [3]	0.00	2.00	7.40	16.10	74.50
Salesman	proposed method	2.02	35.35	59.60	3.03	2.02
	Ref. [3]	0.00	10.10	31.50	18.10	41.30
Akiyo	proposed method	0.00	2.02	70.71	27.27	0.00
	Ref. [3]	0.70	8.00	31.50	28.90	30.90

respectively (Ref. [3] only showed us the experiment results of the first two bit-planes). In the tables, ΔR represents the difference of the number of parity bits between the encoder allocated and actually needed. When the parity bits allocated at the encoder is the same as its actually needed, $\Delta R = 0$. The values shown in the two tables are proportions of the number of WZ frames with $\Delta R = C$ (C is a constant) to the total number of WZ frames.

From Tables 3 and 4 we can see that the proposed rate allocation method has an improvement of 5.18%–52.93% in accuracy compared with Ref. [3]. In Ref. [3], the probability of $|\Delta R| \leq 12$ kbit/s is 24.6%–84.5%, but in our rate allocation method, the probability of $|\Delta R| \leq 12$ kbit/s is 74.85%–100%, which is better than that of Ref. [3].

The accuracy of the proposed rate allocation method is significantly higher than that of Ref. [3]. The reasons may lie in two aspects. The first one is that for certain bit-error rate, the Turbo coder needs a certain number of parity bits, so the use of this correspondence between bit-error rate and the number of needed parity bits brings us good performance. The second one is that, at the encoder, we employ motion estimation and compensation with the simple three-step search method to estimate the side information, which makes the estimated side information closer to the real side information than that in Ref. [3].

4.4 Quality and bit rate of reconstructed images

At the decoder, we also use the method in Ref. [3] to decide whether the decoding is successful or not. The PSNR of the

reconstructed image and the bit rate of WZ frame are shown in Tables 5 and 6 respectively. In Table 5, “the first bit-plane” represents the PSNR of the reconstructed frames after decoding the first bit-plane, “the second bit-plane” represents the PSNR of the reconstructed frames after decoding the second bit-plane. In Table 6, “the first bit-plane” and “the second bit-plane” respectively represent the bit rates of the first and second bit-plane.

Tables 5 and 6 show us that the quality of the reconstructed image is higher than that of Ref. [3], and the bit rate used by our method is lower than that of Ref. [3].

In Ref. [3], the estimated Laplacian parameter $\hat{\sigma}^2$ at the encoder is an overestimate of the real σ^2 since it is expected that the motion compensated interpolation performed at the decoder to obtain the side information will be more accurate than the simple averaging of the two closest K -frames. Therefore, in many frames, an overestimation of the rate is observed. However, in our proposed rate allocation method, because of using fast motion estimation and compensation method, the overestimation goes down. Therefore, the bit rate of the proposed method is lower than that in Ref. [3].

4.5 Encoder complexity analysis

Reference [6] introduced the two-dimensional logarithmic motion vector search method into DVC, which achieve good performance. After comparing the amounts of

Table 5 PSNR of reconstructed image of this paper and Ref. [3]

sequence	PSNR			
	Ref. [3]		proposed method	
	the first bit-plane	the second bit-plane	the first bit-plane	the second bit-plane
Foreman	36.65	37.08	36.7094	37.1714
Carphone	33.43	34.10	35.0205	35.6506
Salesman	44.05	44.35	44.4332	44.5134
Akiyo	49.95	49.99	50.7789	51.0410

Table 6 First two bit-planes encoding rate comparison with Ref. [3]

sequence	bit rate			
	Ref. [3]		proposed method	
	the first bit-plane	the second bit-plane	the first bit-plane	the second bit-plane
Foreman	72	111	61.2891	98.6719
Carphone	80	138	59.0625	123.7500
Salesman	33	62	23.8191	51.2109
Akiyo	30	31	14.1797	18.3984

computations of several fast motion vector search methods, we chose the three-step search method which has the least amount of computations. For details of this search method, please refer to Ref. [7]. The amount of computations of the three-step search method is lower than that in Ref. [6] but slightly higher than that in Ref. [3] where no motion estimation is involved.

5 Conclusions

In this paper, we proposed a new encoder rate allocation method for DVC. In this method we introduced simple three-step motion vector search method into the encoder to help with estimating the side information at the encoder, and based on this estimated side information the encoder determines how many parity bits the decoder needs. Compared with Ref. [3], the proposed rate allocation method improves the accuracy of rate allocation up to 52.93% at the cost of a slight increase of the encoder complexity. Our encoder rate allocation method decreases the system bit rate, and improves the quality of reconstructed images.

Acknowledgements This work was supported by the Fundamental Research Funds for the Central Universities.

References

1. Aaron A, Zhang R, Girod B. Wyner-Ziv coding of motion video. In: Proceedings of IEEE International Conference on Signals Systems and Computers. 2002, 1: 240–244
2. Artigas X, Torres L. Improved signal reconstruction and return channel suppression in distributed video coding systems. In: Proceedings of the 47th International Symposium ELMAR-2005. 2005, 53–56
3. Morbee M, Prades-Nebot J, Pizurica A, Philips W. Rate allocation algorithm for pixel-domain distributed video coding without feedback channel. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. 2007, 1: 521–524
4. Dalai M, Leonardi R, Pereira F. Improving turbo codec integration in pixel-domain distributed video coding. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. 2006, 2: II
5. Belkoura Z M, Sikora T. Towards rate-decoder complexity optimization in Turbo-coder based distributed video coding. In: Proceedings of International Picture Coding Symposium. 2006, 1–5
6. Chen H, Steinbach E. Fast motion estimation-based reference frame generation in Wyner-Ziv residual video coding. In: Proceedings of IEEE International Conference on Multimedia and Expo. 2007, 168–171
7. Cai A, Sun J. Foundation of Multimedia Communication Technology. Beijing: Publishing House of Electronics Industry, 2000, 58–63 (in Chinese)