Side Information-based distributed source coding with Low-Density Parity-Check code

DOI : 10.36909/jer.ICCSCT.19471

Hui Chen*

Department of Science and Engineering, Specialty of Electronic Information Engineering, DianChi College of Yunnan University, Kunming, 650028, China.

* Corresponding Author: yndaxuechenhui_84@163.com

ABSTRACT

Considering the shortcomings of the existing Distributed Source Coding schemes on the bit error rate and the compression ratio, a side information-based distributed source coding scheme with Low-Density Parity-Check code is proposed. In the design of source encoder, redundancy is used to provide error protection, which is similar to insert redundancy to error protection in convolution coding; Set redundancy and Side Information approach is applied in the paper. Because of the correlation along the source, Low-Density Parity-Check code based Slepian-Wolf bound is adopted to implement the distributed source coding strategy. Experimental results are shown that the proposed method in paper can reach high compression ratio when the correlation among neighboring symbols is strong. The bit error rate of the sequence can be kept low between the sources are weak. It is an efficient distributed source coding scheme and is easy to be implemented.

Keywords: Side information; Distributed source coding; Low-Density Parity-Check code; Slepian-Wolf.

INTRODUCTION

Distributed source coding is usually used on distributed audio coding, video coding, hyper-spectral image coding, media recognition, wireless sensor networks and other related fields (Xiong etal.,2004;Hong etal.,2018). Generally, the source X is performed channel coding, then the decoder is used the noise channel received signal to recover the source X, which is called channel coding. Assuming that source X1 is correlated to source X2, source X2 can be regarded as a signal polluted by channel noise through the relevant channel. For source X1, source X2 is equivalent to the noisy signal of X1. Source X1 is performed the channel coding to generate check matrix P and send it to the decoder. The decoder could be recovered source X1 with X2 and P, which is the method of distributed source coding. The correlation between sources to reduce redundancy and efficient coding is the main task of distributed source coding. The source sequence is transmitted to the receiver by the traditional source coding scheme, which is called side information. This paper is the distributed source coding based on side information.

Reference (AJ.Aljohani.,2016) shows the correlation between sources of practical examples can be regarded as the existence of a virtual channel to improve performance, while the independent coding and independent decoding scheme can not be made use of the correlation between sources and waste the information carried by the sources themselves. Therefore, in order to make use of the correlation between sources and improve the system performance, a joint coding and joint decoding scheme is proposed, but the coding efficiency is low; The low density parity check code based on the original modulus diagram is a code that integrates the characteristics of the source channel (Dong etal.,2022;F.Xia.,2018). There is residual redundancy at the output of the source encoder. With the residual redundancy to provide error protection is the similar to inserting redundancy to provide error protection in convolution coding, The method of using redundant joint code type is the design technology of joint source channel code, which not only considered the statistical characteristics of source channel, but also made use of the correlation between sources, but the compression rate is low; The statistical characteristics of sources are used to jointly decode the relevant sources with memory, and the distributed joint source channel coding is improved by global iteration (Hong etal.,2016), however, the bit error rate is high. Reference (A. D. Liveris. etal.,2002)provides the lossy extension of Slepian-Wolf coding that is implemented the application of Distributed Source Coding theory. The Distributed Source Coding-based lossy way is proposed in WT domain, named as set-partitioning in hierarchical tree with Slepian-Wolf coding (M. Cheung. etal.,2006). Moreover, it demonstrates that the presented Distributed Source Coding-based compression frameworks are very promising (E. Baccaglini. etal.,2007;E. Magli.,2009).

In practical application, Distributed Source Coding scheme is usually used to compress the source according to the higher bit plane with the certain structural characteristics, that is, there is strong correlation within the source. Although the Distributed Source Coding scheme based on channel code effectively is utilized the correlation between X and Y, the correlation between the front and rear symbols in X is not utilized. Once the correlation between sources is decreased, the compression rate of the schemes would be limited. Therefore, when the correlation between X and Y is not strong enough and there is strong correlation within X, this scheme is difficult to meet the high compression rate and the low bit error rate. Then a new scheme of low complexity side information-based distributed source coding scheme with LDPC is proposed. In the design of source encoder, redundancy is used to provide error protection; Set redundancy and Side Information are used. Because the correlation of the source, Low-Density Parity-Check code based Slepian-Wolf bound is adopted to implement the DSC strategy. It is met the high compression rate and low bit error rate.

DISTRIBUTED SOURCE CODING

Source coding is included fixed length code and variable length code. In lossy compression, coding based on context model, arithmetic coding, dictionary coding, Huffman coding and so on are commonly used. The essence of coding based on Context model is to describe the coding code length of the sequence $x_n \ldots x_0$ as $-\log P(x_{\mathrm{n}} ... x_0)$ of the joint probability distribution. The entropy coding is transformed the joint probability into the form of conditional probability multiplication as Equation(1).

$$L = -\sum_{i=0}^{n} \log P(x_i \mid x_{i-1} ... x_0) \tag{1}$$

After taking the negative logarithm, L is indicated the number of bits required to describe the sequence (the logarithm is based on 2), which is the length of the sequence coding. Arithmetic coding is a kind of linear coding. The previous research is mainly used in image compression, which is also an entropy coding method. Different from other entropy coding methods, other entropy coding methods usually divide the input message into symbols and then encode each symbol, while arithmetic coding directly encodes the whole input message into a decimal to meet the requirements of $0 \le n < 1$.

In 1973, Slepian and Wolf proved that the condition of the joint decoding between the sources are strong, the source side separate coding or joint coding can be achieved the same compression range of Slepian-Wolf domain (Slepian.etal.,1973). In the design of source encoder, the assumptions made on the source for simplicity or imperfect understanding are usually incorrect, which is lead to residual redundancy at the output of source encoder. The residual redundancy of the error protection is very similar to the redundancy of error protection in convolution coding. The method based on the joint design code is the idea of distributed source coding, which not only reaching the statistical characteristics of the source channel, but also making use of the correlation between the sources.

The block diagram of distributed source coding system is shown in Figure 1. It is assumed that X1 and X2 are memory-less sequences with spatial correlation, and their information bit lengths are K. The encoder of sequence X1 is used symmetric channel code coding with code rate r_1, and m_1 is represented the transmission part of check bit of sequence X1; The encoder of sequence X2 is used a symmetric channel code with a code rate of r_2, and m2 is represented the transmission part of the check bit of sequence X1. The length of encoder and the compression ratio is described as Equation(2).

$$n_i = k + m_i \qquad r_i = \frac{k}{n_i} \tag{2}$$

Where, $r = 1,2$.

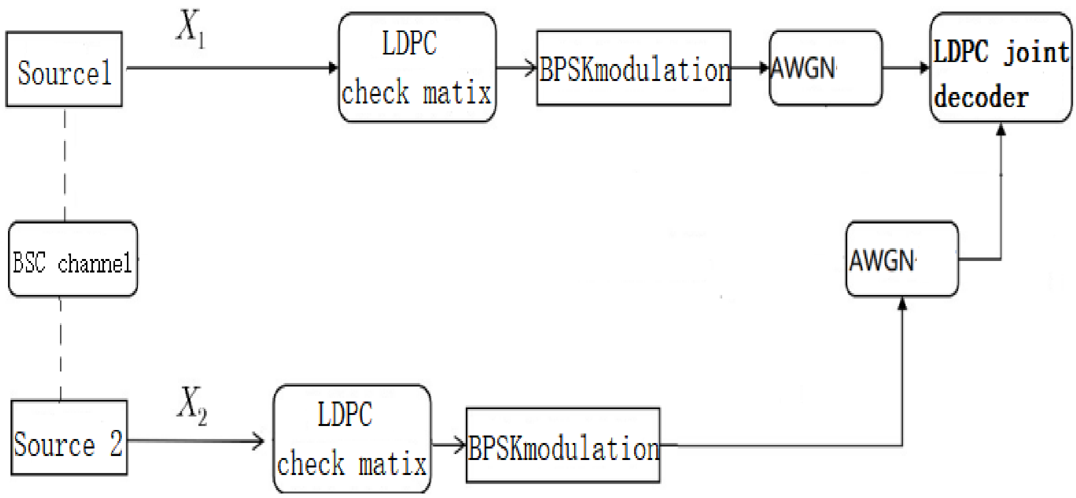

Figure 1. The diagram of distributed source coding Based on memory-less correlated source and Low-Density Parity-Check Code.

The encoded codes are modulated by Binary Phase Shift Keying, and the information bit sequences are generated by the two sources, which is divided into two parts with the parity of the sequence. Assuming that the source X1 transmits odd part information bits and check bits through the additive white Gaussian noise channel, and the source X2 is transmit part information bits and check bits through another independent Additive White Gaussian noise channel, this processing method can be made the transmission bits of the two sources not affected by the node degree of the corresponding variable, Then the compression ratios of sequence X1 and sequence X2 should be estimate respectively by Equation(3):

$$\begin{cases} R_{X_1} = (k/2 + m_1)/k \\ R_{X_2} = (k/2 + m_2)/k \end{cases} \tag{3}$$

Where the bits transmitted by the system are called type one and the others are called type two.

DISTRIBUTED SOURCE CODING BASED ON SIDE INFORMATION

Supposed that X1 and X2 are two correlated binary sequences in DSC research, the correlation of the two sequences can be described by a virtual binary symmetric channel with transition probability P. If the channel error probability is low, the correlation is strong. Therefore, if the receiver could be completely received the source X2, the restoration of X1 can be regarded as "error correction" from X2. Therefore, in the design of distributed source coding based on LDPC, when encoding the binary source sequence X1 with length N, it is related source sequence X2 is side information, and the check matrix H is transformed into a system matrix (4) through Gaussian elimination algorithm, 2-modulation addition between rows and column exchange operation, where I is the m-order unit matrix and P is the transition probability matrix $n \times (n-m)$.

$$H' = [I \mid P] \tag{4}$$

It is calculated by the syndrome (5).

$$S^T = H'X_1^{\ T} \tag{5}$$

Then it is transmitted S^T and X2 to the receiver to compress X1, where S^T and $\boldsymbol{X}_1^{\ T}$ is the column vector; In addition, the check matrix can also be transformed into a lower triangular or approximate lower triangular matrix through column permutation, and then the check matrix can be directly used for linear iterative coding to obtain the system of Low-Density Parity-Check code, which could be made use of the sparsity of check matrix H.

In the decoding stage, the belief propagation algorithm based on syndrome is adopted, the syndrome S^T of X1 and X2 are used for error correction decoding to obtain the restored source sequence. The decoder system of X1 is shown in Figure 2.

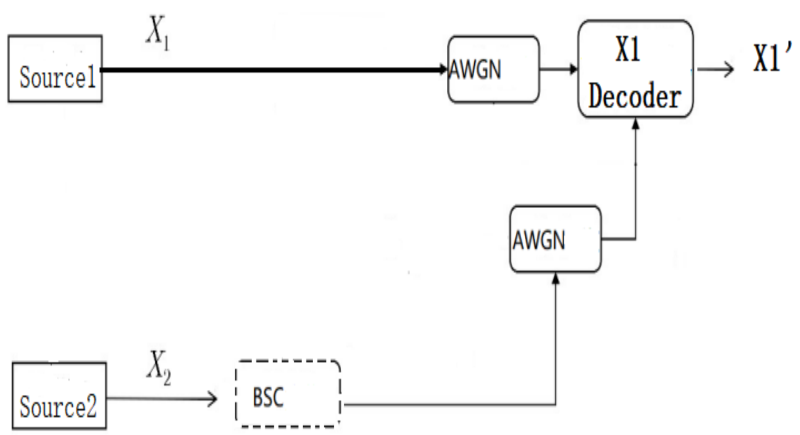

Figure 2. The diagram of the X1 decoder.

The most common decoding algorithm is log-BP algorithm, which is defined variables and sets at first:

1)$P_j(a)$ is shown the conditional probability when the decoder receives the signal y_j interfered by channel noise under the condition that the j-th transmission information bit is a;

2)$r_{ij}(a)$ is shown the probability that the sum of other variable nodes participating in the i-th check equation is a except the j-th variable node, that is, the probability that the check node c_i is passed to the variable node v_j ($c_i \rightarrow v_j$);

3)$q_{ji}(a)$ is shown that the variable node v_j=a and passed it to the verification node c_i, which is passed it to the variable node ci ($v_j \rightarrow c_i$);

4)$q_j(a)$ is shown the posteriori probability when the transmission information bit is a;

5)V(i)\j is shown the set of all other variable nodes connected to the i-th inspection node except the j-th variable node;

6)C(j)\i is shown the set of all verification nodes except the i-th verification node and the j-th variable node, where, $0 \le j \le n, 0 \le i \le m, a = \{0,1\}$.

Its realization by decoding algorithm is also put forward.

Step1: Set the number of iterations K to 0, and calculate the initial likelihood information of all variable nodes according to Equation(6)(Hua etal.,2008);

$$L_{ch,j} = \log \frac{P_j(0)}{P_j(1)} = \log \frac{P_r(y_j \mid x_j = 1)}{P_r(y_j \mid x_j = -1)} = \log \frac{(1 + e^{-2y_j/\sigma_n^2})^{-1}}{(1 + e^{+2y_j/\sigma_n^2})^{-1}} = \frac{2y_j}{\sigma_n^2} \tag{6}$$

Step2: After the initialization work is completed, it is used the iterative cycle process. The log likelihood information L (r_{ij}) transmitted by each check node c_i to each variable node v_j ($c_i \rightarrow v_j$) is calculated by Equation(7)(Chen etal.,2008), in which the initial information of check node c_i is 0;

$$L(r_{ij}) = \log \frac{r_{ij}(0)}{r_{ij}(1)} = 2 \tanh^{-1} \Big(\prod_{j \in V(i) \setminus j} L_{ij} \Big) \tag{7}$$

Step3: Calculate the variable node v_j through Equation(8)(B.Penna. etal,2007)and transfer the log likelihood information L (q_{ji}) of the check node c_i ($v_j \rightarrow c_i$);

$$L(q_{ji}) = \log \frac{q_{ji}(0)}{q_{ji}(1)} = L_{ch,j} + \sum_{i \in C(j) \setminus i} L_{ij} \tag{8}$$

Step4: Calculate the posteriori likelihood information when the transmission information bit is a through Equation(9)(Hou etal.,2008);

$$L(Q_j) = L_{ch,j} + \sum_{i \in C(j)} L_{ij} \tag{9}$$

Step5: Estimate any bit according to the likelihood information and the decision

Equation(10) is as follows:

$$\hat{c}_j = \begin{cases} 1, & L(Q_1) < 0 \\ 0, & L(Q_1) \geq 0 \end{cases} \tag{10}$$

Step6: Finally, the estimated sequences are obtained, then it is determined whether the decoding is successful. If the decoding is successful or it is reached the set maximum number of iterations, the decoding is over; otherwise, the number of iterations k = k+1, and return to the Step 2.

The advantages of LDPC are be described:

1)It is the linear coding complexity;

2)It is extremely low error rate;

3)The minimum distance is linear with the code length;

4) It is convenient for hardware implementation;

5) It is good of rate scalability.

MODELING AND ANALYSIS

In our simulations, the Low-Density Parity-Check codes and their check matrices are described as 4096×8192 (R=1/2), 2048×8192 (R=3/4), 1024×8192(R=7/8), where the check matrix column weight is 3. Scheme [3], Scheme [4] and the paper are used the same LDPC check matrix at different bit rates. In order to avoid the error caused by noise randomness, the experimental data are the results of repeated experiments, and the experimental object is gray image(256×256), each plane is compounded by 65536 bits as the source sequence X1. Let X1 get X2 through BSC to simulate the correlation between SI and source. In this section, comparative experiments are conducted to investigate the bit error rate and compression rate of different schemes. The information to be transmitted by distributed source coding is included the code by arithmetic coding X2 and the syndrome formula S^T of X1. For example, when the code rate of LDPC is 3/4(R = 3/4), perform arithmetic coding on X1 (code length is L), and calculate the syndrome corresponding to X1 with the code rate (total length is S), then the total code rate is (L+S)/65536. It is tested the two algorithms in the BER experiments.

Taking the third bit plane of the standard image of Man as the source, Figure 3 is shown the bit error performance of references [3], [4] and this paper. The error code of distributed source coding comes from the reconstruction process of X1, so all BER data are calculated from the number of error bits between X1 and the estimated $\boldsymbol{X}_1'$. When the correlation between sources is strong, the performance of scheme [3], [4] is better. For R=1/2, the BER is 0. When the correlation between sources becomes worse, its BER increases sharply, and the recovery ability of error bits is almost 0. When the LDPC bit rate is adjusted to 3/4, its performance is greatly reduced compared with 1/2. Therefore, under the condition of ensuring a certain BER, scheme [3] has high requirements for the correlation between sources. For discrete arithmetic coding, the BER is lower when the correlation between sources is strong, but lower when the correlation is weak. In scheme [3] and [4], when the correlation between sources is strong, the BER is relatively low. When the correlation is weak, the BER is low.

The code rate of Low-Density Parity-Check code which is the distributed source coding in this experiment is 5/8, the syndrome length is 37.5% of X1 length. When the second-order arithmetic encoder is used to compress X2, the total bit rates of scheme [3], scheme[4] and this paper are 0.22138, 0.2269 and 0.2531 respectively. Therefore, this algorithm is better than scheme [3] and scheme [4] in bit error rate.

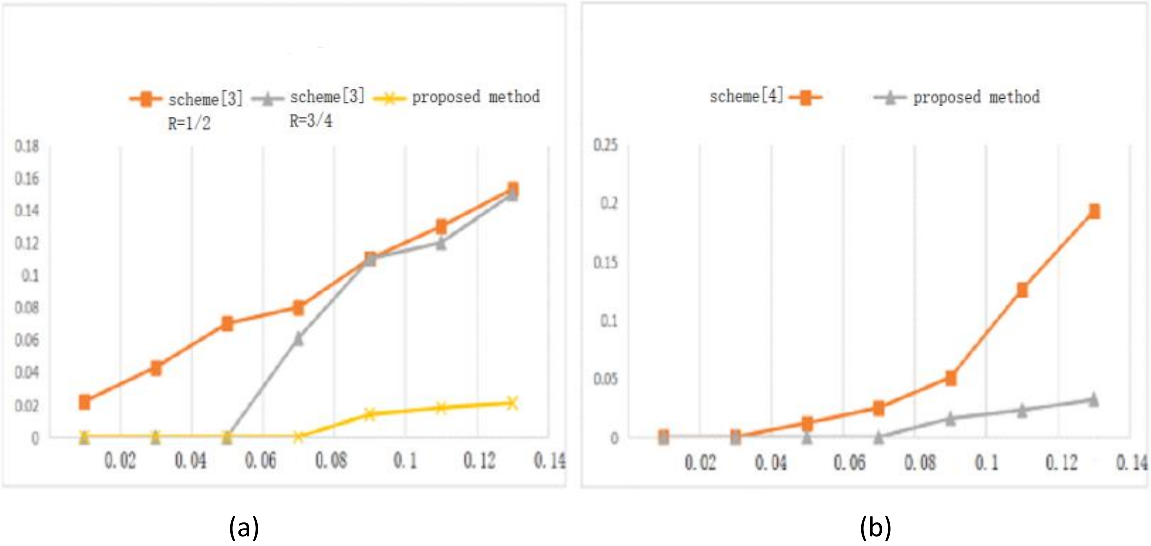

Figure 3. The comparison of bit error rate.

The compression ratio is are mentioned in Table 1. The third bit plane of standard images Lena, Boats and Woman is used as the source and the compression rate of correlation between different sources. Two experiments are carried out for each source according to the correlation strength, and the results are reported in left and right columns, with R=1/2, where C is the order of the context probability model when arithmetic coding is performed. The code rates given in TableⅠare the total code rates. It can be seen from TableⅠthat if X2 is encoded with 0-order entropy, the bit rate is about 0.6 because it can not be made use of the correlation of the sequence, which is inferior to scheme [3] and scheme [4]. However, when X2 is encoded with high-order entropy, its bit rate is much lower than that in scheme

[3] and scheme [4], and the compression rate is not varied with the change of correlation between sources, and it is very important to use the internal correlation of X1. When the correlation between the source and SI is not fully utilized, the code rate is the lowest, indicating that the utilization of the correlation between the source and the source reaches the best balance. In fact, H(X1|X2) represents the limit that the source X1 can be compressed when only the source X2 is used. In the experiment, the bit rate of scheme [3] is 1/2, and the bit rate of scheme [4] is also about 1/2. Only when H (X1|X2) is significantly less than 1/2, that is, the correlation between X1and X2 is high, can they achieve ideal coding results. When the correlation between X and Y is low and H (X1|X2) is close to 1/2, the bit error rate of these two schemes is much higher than this paper.

Table 1. The compression ratio on images.

	C	**Scheme[3]**	**Scheme[4]**	**Proposed method**
Lena	0	0.5262	0.6121	0.6274
	1	0.2681	0.2787	0.3052
	2	0.2043	0.2044	0.2136
Boats	**C**	**Scheme[3]**	**Scheme[4]**	**Proposed method**
	0	0.5262	0.6121	0.6274
	1	0.3571	0.3623	0.3862
	2	0.2135	0.2217	0.2543
Woman	**C**	**Scheme[3]**	**Scheme[4]**	**Proposed method**
	0	0.5642	0.5727	0.5823
	1	0.2362	0.2411	0.2563
	2	0.1767	0.1649	0.1833

CONCLUSION

In this study, the traditional distributed source coding algorithm is an efficiency of the correlation between sources, while seldom taking the further use of the correlation within sources. The algorithm proposed in this paper can be made use of the correlation within the source sequence to achieve the lower bit error rate. Compared with the other Distributed Source Coding schemes, the compression ratio in paper can be higher and the bit error rate of the sequence can be lower. It is an efficient scheme of distributed source coding.

The new scheme has been tested on a set of scenes of images. The results further demonstrate the advantages of modification distributed source coding employed in source coding, which is characterized by close correlation. In the future, it is considered the 2-

dimensional sampling to maximize the correlation within the sequence and improve the accuracy of the symbol estimation algorithm.

ACKNOWLEDGEMENT

This work is supported by Science Foundation of Dianchi College Department.

REFERENCES

Aljohani, A.J., Ng, S.X. & Hanzo, L. 2016. Distributed source coding and its applications in relaying-based transmission[J]. IEEE Access.4: 1940-1951.

Baccaglini, E., Barni, M., Capobianco, L. & Garzelli, A. 2007. Low-complexity lossless compression of hyper-spectral images using scalar cost codes. IEEE Proceedings of Picture Coding Symposium.1: 1–4.

Cheung, N.M., Tang,C., Ortega, A. & Raghavendra,C.S. 2006. Efficient wavelet-based predictive Slepian-Wolf coding for hyper-spectral imagery. Signal Processing, 86(11): 3180–3195.

Chen, D., Varodayan, D. & Flierl, M. 2008. Winery-Viz coding of multi-view images with unsupervised learning of disparity and gray code. Signal Processing: Image Communication. 23(5):369-378.

Dong, Y.F., Dai ,J.C., Niu ,K., Wang ,S. & Yuan, Y,F. 2022. Joint Source-Channel Coding for 6G Communications[J]. China Communications.19(3):101-115.

E. Magli. 2009. Multiband lossless compression of hyperspectral images. IEEE Transactions on Geoscience Remote Sensing. 47(4): 1168-1178.

Hong, S. & Wang, L. 2016. Photograph LDPC-based distributed joint source-channel coding. Proceedings of the 2016 International Conference on Communication Systems. Shenzhen, China.

Hua, G. & Chen, C. 2008. Distributed video coding with zero motion skip and efficient DCT coefficient encoding. Proceedings of the 2008 International Conference on Multimedia and Exposition. Hannover, Germany.

Hou, Y. & Liu, G.Z. 2008. Lossy-to-lossless compression of hyper-spectral image using the improved AT-3D SPIHT algorithm. Computer Science and Software Engineering. 2(1): 963-966.

Liveris, A.D., Xiong, Z.X. & Georghiades, C.N. 2002. Compression of binary sources with side information at the decoder using LDPC codes. IEEE Communication Letters.6(1):

440–442.

Penna, B., Tillo,T., Magli, E. & Olmo, G. 2007. Transform coding techniques for lossy hyper-spectral data compression. IEEE Trans. on Geoscience and Remote Sensing. 45(5): 1408-1421.

Slepian, D. &Wolf, J. 1973. Noiseless coding of correlated information source[J]. IEEE Transactions on Information Theory.19(4): 471-480.

Xia, F. 2018. Research on distributed joint source-channel coding using photograph LDPC codes. master thesis of Xiamen University

Xiong, Z., Liveris, A.D. & Cheng, S. 2004. Distributed source coding for sensor networks[J]. IEEE Signal Process.21(5): 80-92.

Yang, H., Qing, L.B., He, X.H., Ou, X.F. & Liu, X.J. 2018. Robust distributed video coding for wireless multimedia sensor network[J]. Multimedia Tools and Application.77(4): 4453-4467.