MULTI-VIEW IMAGES CODING BASED ON MULTITERMINAL SOURCE CODING

Feng Yang[1], *Qionghai Dai*[1], *Senior member, IEEE, and Guiguang Ding*[2]

[1]Department of Automation, Tsinghua University, Beijing, China, 100084
[2]School of Software, Tsinghua University, Beijing, China, 100084

ABSTRACT

In this paper, we proposed a multi-view images coding method based on Multiterminal Source Coding (MSC). Due to separate encoding in MSC, our coding scheme can achieve good random access performance and the spatial redundancy can be exploited even if the encoders can not communicate with each other. Because of joint decoding, the compression performance of our scheme is promising, far better than that of separate encoding and decoding scheme. Compared to multi-view coding based on Wyner Ziv coding, our coding scheme is more flexible and can be easily extended to N views coding. There is no need to classify the view images into key images and Wyner Ziv images, images can be compressed in the same way and we can easily change the compression rate of each view to adapt to the resource conditions, like network bandwidth or storage. Experiment results show that the compression performance of our scheme is better than that of JPEG encoder and decoder scheme.

***Index Terms*—**Wyner Ziv Coding, Multiterminal Source Coding, Multi-View Coding

1. INTRODUCTION

3DTV which is very popular in today's research area is expected to significantly change people's everyday life. In 3DTV system [1], multi-view images or video of scenes are captured by a set of cameras. This resembles the Free Viewpoint Television (FTV) system [2], in which cameras are set densely on line or arc etc, so user can freely change viewpoint. Since the amount of data in both systems is too huge to store or transmit, compression is needed. Here we consider multi-view images coding, which also can be considered as compressing Intra frames in multi-view video coding. Various methods have been proposed for solving this problem. In traditional compression methods [3] cameras have to communicate with each other to remove the spatial redundancy. The flaw of these schemes is that in some cases, the cameras can not communicate with each other. Later in [4, 5] Wyner Ziv coding is used for multi-view or light fields coding to avoid the encoder communication but still exploit the spatial redundancy. This is attained through jointly decoding. Since view images in Wyner Ziv coding is not equivalent, some serve as key images, while some are Wyner Ziv images, this coding scheme is not flexible and not easily extended to N views. To avoid this problem, in this paper a multi-view images coding scheme based on MSC which means that sources are encoded separately but decoded jointly, is proposed. Due to the good properties of MSC, view images can be compressed in the same way and the compression rate of each view can be changed according to the view's capturing camera's storage or network access bandwidth. Also, our scheme can be simply extended to N views.

The rest of our paper organized as follows. Section 2 introduces some theoretical foundations about MSC and a practical MSC scheme we proposed. In section 3, the proposed multi-view images coding scheme is illustrated in great detail. Section 4 presents some experiment results and Conclusion is made in section 5.

2. THEORETICAL FOUNDATIONS AND PROPOSED MSC SCHEME

The architecture of MSC is shown in Fig.1.

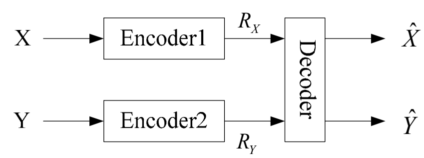

Fig.1 Multiterminal Source Coding

As shown in Fig.1, in MSC, sources are separately encoded and jointly decoded. For the lossless case, Slepian and Wolf [6] gave the theoretical bound.

$$\begin{cases} R_X \geq H(X \mid Y) \\ R_Y \geq H(Y \mid X) \\ R_X + R_Y \geq H(X, Y) \end{cases} \tag{1}$$

The loss case is first considered in [7] in 1978, they gave the inner and outer bound for rate pair (R_x, R_y) given D_x, D_y.

For the Gaussian MSC problem with MSE distortion measure, where (X, Y) are jointly Gaussian with variances (σ_x^2, σ_y^2) and correlation coefficient $\rho = E[XY]/(\sigma_x\sigma_y)$, the inner sum-rate bound is given in [8].

$$R_X + R_Y \geq \frac{1}{2}\log^+\left[\left(1-\rho^2\right)\frac{\beta_{\max}\sigma_X^2\sigma_Y^2}{2D_X D_Y}\right] \quad (2)$$

where $\beta_{\max} = 1+\sqrt{1+\left(4\rho^2 D_X D_Y\right)/\left[\left(1-\rho^2\right)^2\sigma_X^2\sigma_Y^2\right]}$,and $\log^+ x = \max\{\log x, 0\}$.

We designed a practical lossless MSC scheme to achieve the theoretical bound. The coding scheme proposed is shown in Fig.2.

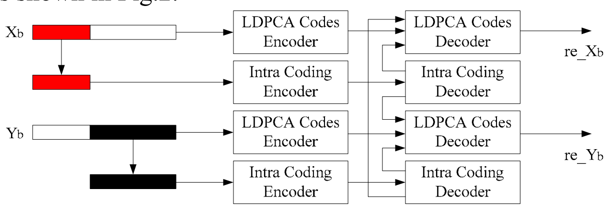

Fig.2 Practical Multiterminal Source Coding

In this paper, we add uniform quantizers before our lossless MSC scheme to achieve the loss MSC theoretical bound. X and Y are first quantized by uniform quantizers and then extracted to bit planes. Each bit plane of X, X_b as shown in Fig.2, is encoded by *Low Density Parity Check Accumulate (LDPCA) Codes Encoder* which is designed in [9] to achieve Slepian-Wolf bound. Only the syndrome bits are transmitted. Meanwhile, the red part of X_b is compressed by *Intra Coding Encoder.* Each bit plane of Y, Y_b, is also encoded by *LDPCA Codes Encoder* and the black part of Y_b is encoded by *Intra Coding Encoder.* At the decoder, *LDPCA Codes Decoder* using received syndrome bits of X_b and side information bits which are decoded red part of X_b and black part of Y_b to decode re_X_b. And re_Y_b can be got in the same way. After all bit planes of X and Y are decoded, reconstruction function is implemented to attain the estimated version of X and Y.

In the following sections, a multi-view images coding scheme is proposed based on the above MSC method.

3. PROPOSED MULTI-VIEW CODING SCHEME

In this paper we focus on the compression of two views, which can be easily extended to multiple views. In the following paragraphs, our proposed multi-view coding scheme based on the above MSC method will be introduced in great detail.

3.1. Encoder

The encoder of our coding scheme is shown in Fig.3. As shown the figure, every view is divided into two part, left part and right part. The left part of *View1* and right part of *View2* are dealt by *Intra Coding Encoder*. Note that the parts which are processed by the *Intra Coding Encoder* from the two views should be different parts. At the same time, the whole images of the two views are compressed by *Wyner-Ziv Coding Encoder.* The encoder generates four bit streams that are *bitstream1*, *bitstream2*, *bitstream3*, and *bitstream4* as shown in Fig.3.

The *Intra Coding Encoder* which is just like JPEG encoder is shown in Fig.4. Image data is first divided into 8×8 blocks and then every block is processed by *Q (quantiztion), DCT, Zigzag, Run-length Coding,* and *Huffman Encoder* to generate final encoded bits.

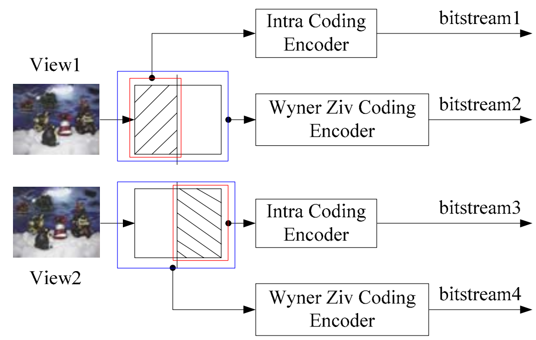

Fig.3 Proposed Encoder

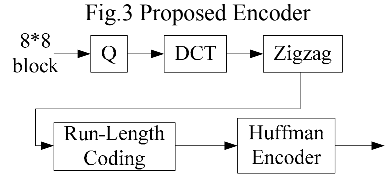

Fig.4 Intra Coding Encoder

The *Wyner Ziv Encoder* is shown in Fig.5. Image data is first divided into 8×8 blocks. After that, DCT transform is implemented for each block, and then coefficients in the same band in every block are combined together to form *Band k.* The *Band k* is then quantized and extracted to bit planes. Each bit plane is encoded by *LDPCA Codes Encoder,* only syndrome bits of each bit plane is transmitted.

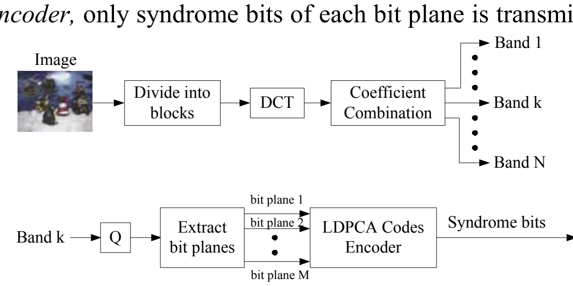

Fig.5 Wyner Ziv Encoder

From the description of our proposed encoder for multi-view coding, it can be seen that our encoder is very simple, only intra coding and channel codes' encoder. Views are encoded separately, so better random access performance can be achieved. In Fig.3 the left half part of *View1* and the right half part of *View2* are compressed by *Intra Coding Encoder*. There is no need for this. We can change the width of the left part of *View1* and the right part of *View2,* as long as the sum of the width of the left part of *View1* and the width of the right part of *View2* is equal to the width of the whole *View.* Since the bit rate of intra coding is higher than that of Wyner Ziv coding, the compression bit rate of the two views can be adjusted through changing width of intra coding parts in the two views. This is important, since in some cases, the storage or network access bandwidth of the two views' capturing cameras is not the same and may have a significant distinction. From our encoder scheme, it can be apparently seen that our scheme can be easily extended to N views. The change of encoder for N views is that, 1/N part of each view is encoded by *Intra Coding Encoder*.

3.2. Decoder

As shown in Fig.6, at the decoder, the received *bitstream1* and *bitstream3* are decoded by *Intra Coding Decoder* which is just the inverse process to *Intra Coding Encoder*. The *bitstream2* and *bitstream4* are decoded by *Wyner Ziv Coding Decoder.* The decoded *bitstream1* and decoded *bitstream3* after disparity compensation are served as side information for decoding *bitstream2,* while the decoded *bitstream3* and decoded *bitstream1* after disparity compensation are served as side information for decoding *bitstream4*.

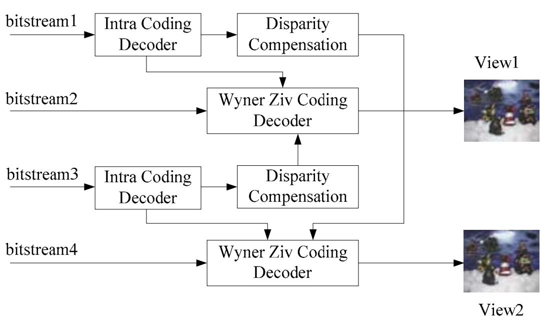

Fig.6 Proposed Decoder

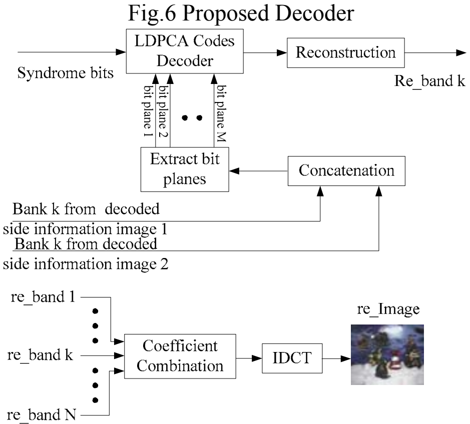

Fig.7 Wyner Ziv Decoder

The *Wyner Ziv Decoder* is shown in Fig.7. For decoding the *Bank k* of the original image data, the *Wyner Ziv Decoder* ultilizes the 'band k' of the side information image 1 and 2. The 'band k' is generated in the same way as that of the original image data. Then the two 'bank k' are concatenated together, extracted in to bit planes, and served as the side information for decoding bit planes of *Band k. The LDPCA codes Decoder* implements the decoding process using the received syndrome bits and side information bit planes. Since there is correlation between different bit planes, the Log-likelihood ratio (LLR) in *LDPCA codes Decoder* is calculated as follows.

Let $\bar{X} = [x_1, x_2, \cdots, x_N]$ denotes quantized version of *Band k* of the original image data, the corresponding side information is $\bar{Y} = [y_1, y_2, \cdots, y_N]$, the bit planes for $\bar{X}$ is $\bar{U}^r = [u_1^r, u_2^r, \cdots, u_N^r], r = 1, \cdots, M$ and the bit planes for $\bar{Y}$ is $\bar{V}^r = [v_1^r, v_2^r, \cdots, v_N^r], r = 1, \cdots, M$ .When decoding V*iew1*, we have $x_i = y_i$ for $i = 1, \cdots, N/2$. If decoding *View2* we have $x_i = y_i$ for $i = N/2+1, \cdots, N$. Suppose decoding *View1*, then

$$\begin{cases} LLR\left(u_i^r \mid u_i^s, v_i^t, s=1,\cdots,r-1, t=1,\cdots,r\right) & \\ = \log\dfrac{P\left(u_i^r = 0 \mid u_i^s, v_i^t, s=1,\cdots,r-1, t=1,\cdots,r\right)}{P\left(u_i^r = 1 \mid u_i^s, v_i^t, s=1,\cdots,r-1, t=1,\cdots,r\right)} & i = N/2+1,\cdots,N \\ LLR\left(u_i^r \mid u_i^s, v_i^t, s=1,\cdots,r-1, t=1,\cdots,r\right) = Infinity & v_i^r = 0 \; and \; i = 1,\cdots,N/2 \\ LLR\left(u_i^r \mid u_i^s, v_i^t, s=1,\cdots,r-1, t=1,\cdots,r\right) = -Infinity & v_i^r = 1 \; and \; i = 1,\cdots,N/2 \end{cases} \tag{3}$$

The reconstruction function used here is just the inverse of the quantization function. After all the *Band k* are decoded, the coefficients from the same block are put together, and then inverse DCT transform is implemented to get the reconstructed image data.

From the description of the decoder, it can be easily seen that the decoding of the two views are a joint job. That is to say, the decoding of one view needs to decode half part of the other view first. Through this method, the correlation between the two views is exploited at the decoder. The compression performance is better than separate decoding. The higher the correlation between the two views, the high the compression ratio. Since the disparity compensation and other complexity techniques such as channel decoding are used at the decoder, the decoder is more complex than separate decoding. But we can sacrifice compression performance for simple decoder. The change of the decoder for N views is just the change of the calculation of LLR.

4. SIMULATION RESULTS

We test our coding system on multi-view sequences *Xmas sequence* and *Breakdancer&ballet sequence*. The multi-view images chosen in *Xmas sequence* are xmas_t028_000 and xmas_t028_001 which are the captured images of camera 0 and 1 at time 28. The images chosen in *Breakdancer&ballet sequence* are the first image captured by camera 2 and 3. The resolution of the images used here is 176×144. The block size is 8×8. The LDPCA codes' codeword length is 396. The degree distribution for LDPCA codes are the same as that in [10]. That is $\delta_2 = 0.321$, $\delta_3 = 0.456$, $\delta_6 = 0.010$, $\delta_7 = 0.174$, $\delta_8 = 0.039$, where

δ_r is the proportion of nodes of degree r. For convenience, the calculation of the LLR just depends on two upper bit planes. The quantization table is just the product of the JPEG quantization table and a coefficient. Compression rate is changed through changing the coefficient. The compression performance is compared with separate encoding and decoding, in which the two views are encoded and decoded by JPEG codec separately.

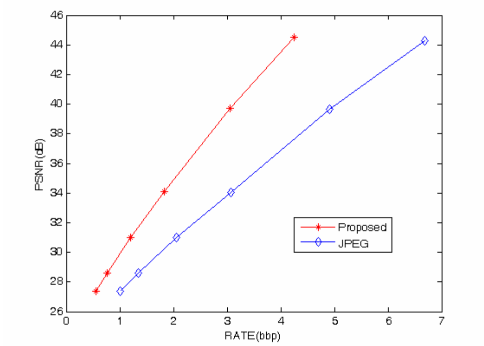

Fig.8 Compression performance for images from *Xmas sequence*

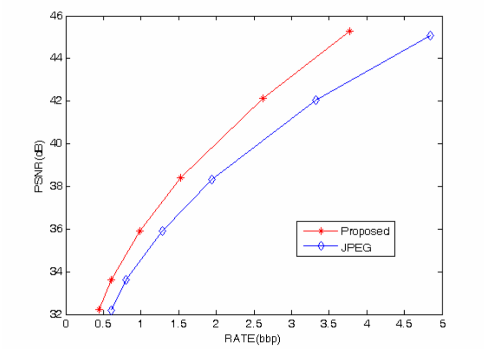

Fig.9 Compression performance for images from *Breakdancer&ballet sequence*

Fig.8 and 9 show the compression performance of our proposed coding scheme and JPEG. It is can be apparently seen that our proposed method is better that of JPEG. The rate here is the total rate for compressing both views and the PSNR is the mean of the PSNR of the two views. From the figure we can see that the performance gap between our proposed scheme and JPEG become larger with the increasing of the total rate. The reason for this is that with the growth of the rate, more gains can be got from jointly decoding. Also we can see that the gap between our proposed system and JPEG for images from *Xmas sequence* is larger than that for images from *Breakdancer&ballet sequence.* It is because the correlation between views in *Xmas sequence* is bigger than that in *Breakdancer&ballet sequence.*

5. CONCLUSIONS

In this paper, a multi-view images coding scheme based on MSC is proposed. Due to the good properties of MSC, our scheme can attain better random access and compression performance. Compared to multi-view coding based on Wyner Ziv coding, our scheme can be more flexible and easily extended to N views. Here, only multi-view images coding is considered, future research on multi-view video coding based on MSC can be fruitful.

6. ACKNOWLEDGEMENTS

The authors acknowledge the support received from the Distinguished Young Scholars of NSFC (No.60525111), the National Natural Science Foundation of China (Project 60502014), and the key project of National Natural Science Foundation of China (Project 60432030).

7. REFERENCES

[1] B.Javidi and F.Okano, Eds., *Three-Dimensional Television, Video, and Display Technologies*, Springer-Verlag, 2002

[2] P.Na Bandchang, T.Fujii, M.Tanimoto, "Experimental System of free viewpoint television," *Proc. of IS&T/SPIE Symposium on Electronic Imaging*, Santa Clara, CA, USA, Vol. 5006-66, pp.554-563, Jan. 2003

[3] M.Magnor, and B.Girod, "Data Compression for Light-Field Rendering," *IEEE Trans. On Circuit Systems for Video Technology*, Vol.10, No.3, pp.338-343, April 2000

[4] A.Aaron,P.Ramanathan and B.Girod, "Wyner- Ziv coding of light fields for random access,"*IEEE 6th workshop on Multimedia Signal Processing*, p.323-326, 29 Sept-1 Oct. 2004

[5] M.P.Tehrani, T.Fujii, M.Tanimoto, "Distributed Source Coding of Multiview Images," *Proc. of SPIE Visual Communications and Image Processing*, Vol.5308, pp.300-309, Jan. 2004

[6] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources", *IEEE Transactions on Information Theory*, vol. 19, pp. 471-480, 1973

[7] S.Y.Tung, "Multiterminal Source Coding," *PhD thesis*, Cornell University, 1978

[8] Y.Oohama, "Gaussian multiterminal source coding,"*IEEE Trans. Inform. Theory*, vol.43, pp.1912-1923, Nov.1997

[9] D. Varodayan, A. Aaron and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, 1203- 1207, Nov. 2005

[10] D.Varodayan, "http://www.stanford.edu/%7Edivad/software.html", 2006