DISTRIBUTED CODING OF MULTIRESOLUTION OMNIDIRECTIONAL IMAGES

Vijayaraghavan Thirumalai, Ivana Tosic and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute, Lausanne, 1015 - Switzerland.
{vijayaraghavan.thirumalai,ivana.tosic,pascal.frossard}@epfl.ch

ABSTRACT

This paper addresses the problem of compact representation of a 3D scene, captured by distributed omnidirectional cameras. As the images from the sensors are likely to be correlated in most practical scenarios, we build a distributed algorithm based on coding with side information. A reference image is processed with a wavelet transform and progressively encoded. The Wyner-Ziv images undergo a multiresolution representation, and the generated bitplanes are channel encoded with LDPC codes. The central decoder eventually reconstructs the Wyner-Ziv images given by the syndrome bits from the channel codes using the reference omnidirectional image. It also iteratively implements motion estimation on the 2-sphere in order to improve the side information. Experimental results demonstrate that distributed coding improves the rate-distortion performance for coding a set of omnidirectional images when compared to independent coding solutions. The proposed method can further be extended to the decoding of multiple Wyner-Ziv images using one single reference omnidirectional image. Hence, it achieves a reduced overall coding rate compared to disparity-based schemes. In addition, it does not require explicit knowledge of the camera parameters nor precise calibration, which is certainly interesting in camera networks.

Index Terms— Distributed coding, omnidirectional imaging, plenoptic function

1. INTRODUCTION

Camera sensor networks are used in several applications like surveillance or object tracking to collect useful data from physical environments. These sensors capture a scene from different view points and generate a series of correlated images of overlapping views. In most scenarios, sensors hardly communicate among themselves due to power limitations. This clearly limits the possibility for joint encoding of the correlated images, which is moreover too computationally complex to be performed in the vision sensors. It becomes therefore necessary to develop distributed coding algorithms where images are encoded independently at each sensor, while the decoder tries to exploit the correlation between images for improving the quality of jointly decoded images.

Distributed source coding (DSC) relies on the theoretical background for distributed compression proposed by Slepian-Wolf [1] and Wyner-Ziv [2]. Although the theoretical foundations have been defined in the late seventies, the first constructive design of a DSC scheme using channel codes [3] has been formulated only recently. It has further motivated several researchers to build DSC solutions using different channel codes based on Turbo or LDPC codes [7]. Recently, distributed video encoders have been proposed using the Wyner-Ziv theorem [4]-[6], in an effort to reduce the complexity of the encoding process by shifting the motion estimation procedure to the decoder. Distributed coding has also been applied to multiresolution hyperspectral images [14]. However, only few studies have been reported on the application of distributed image coding to camera sensor networks. In these scenarios, researchers have reached improved coding performance by (i) exploiting the correlation between the images by disparity estimation using homography [8], (ii) motion estimation procedure [9] and (iii) coset encoding with optimal design of quantizers [10].

We consider a framework where the spherical images are captured by the omnidirectional cameras distributed arbitrarily in a 3D scene. As they give a full view of the 3D scene, omnidirectional images are very suitable for scene representation. After mapping to spherical images, omnidirectional images capture the plenoptic function in its radial form, without Euclidean discrepancies. We propose a coding scheme based on DSC principles, where the side information is generated from one reference image, contrarily to the scheme in [8] where the side information is generated using disparity estimation from two reference images. The proposed framework is therefore more robust, since it can be easily extended to multiple cameras without the need for complex calibration methods. Such an advantage becomes even more important in dynamic camera networks, where real-time camera calibration is almost impossible. The omnidirectional images undergo a wavelet transform (WT) coupled with a progressive representation implemented by the SPIHT algorithm [11]. The bit plane level statistical correlation between images is exploited using LDPC codes. The benefits of such an approach include better compression with little processing overhead, due to the compact representation offered by SPIHT. By exploiting the progressive bit plane reconstruction of images at the decoder, we iteratively refine the side information using motion estimation (ME) between approximated spherical images. Numerical results demonstrate that ME allows to efficiently exploit the correlation between images, and hence improves compression performance. Finally, we extend our scheme to scenarios with multiple cameras, where experiments demonstrate the advantage of Wyner-Ziv decoding based on a single reference image, over a scheme that rather codes the disparity computed from multiple correlated images.

2. DISTRIBUTED CODING SCHEME

We consider the scenario where omnidirectional cameras are distributed in a 3D scene. Each camera could be considered as a sampling device for the plenoptic function, which represents the intensity and chromaticity distribution of the light observed from every position and direction in the 3D space.

In order to provide a compact representation of the scene, without communication between cameras, we propose a scheme for distributed coding with side information. In particular, if we denote by X and Y the correlated omnidirectional images generated by two camera sensors, we are interested in the problem where a compressed version of the image Y is used as side information for decoding the image X at the joint decoder. Both images X and Y are encoded using a multiresolution transform and a progressive encoder. While the compressed image Y is sent to the central decoder after additional entropy coding, the image X is further channel encoded using an LDPC code, and only the resulting syndrome bits are sent to the central decoder. Under the assumption that images are correlated, the syndrome bits are sufficient for a central decoder to reconstruct the image X using the reference image Y. At the central decoder the side information is refined using motion estimation performed on the sphere [13], in order to compensate for the displacement of objects captured from different viewpoints. The block schemes for the independent encoders and the central decoder, are given in Fig. 1.

CS: Compressed significant bits   SS: Sign syndrome bits   RS: Refinement syndrome bits
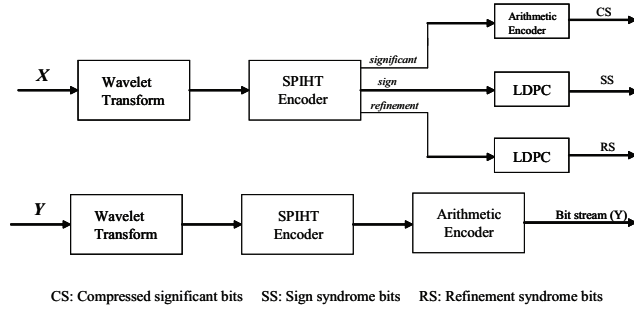
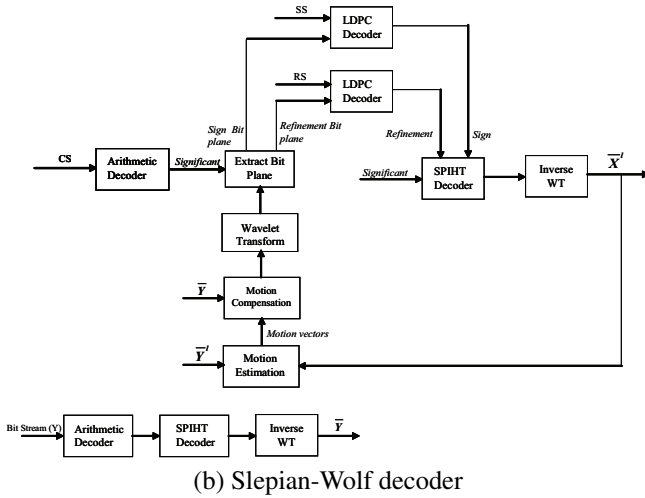(a) Slepian-Wolf encoder

(b) Slepian-Wolf decoder

Fig. 1. Block schemes of encoders and decoder.

2.1. Slepian-Wolf Encoder

The encoding strategy of the Wyner-Ziv image X is now discussed in more details. As shown in Fig.1 (a), the image X is first transformed using a wavelet decomposition, and then progressively coded using a SPIHT encoder. In the SPIHT encoding, the reconstructed image quality considerably degrades with errors in the significant bits, since they convey the significance tree information. Therefore, in the proposed Wyner-Ziv encoder significant bits are entropy coded, and they are later used by the decoder to identify the significant coefficients. On the other hand, sign and refinement bits are encoded by the LDPC code and the corresponding syndrome bits are sent to the decoder. In general, syndromes are calculated by multiplying the bit plane sequence with the parity check matrix H. In our scheme, each bit plane is encoded in multiple passes or padded with zeros to match the column length in H. The crossover probability p is used to determine the rate R of the LDPC code. Under the assumption that the channel is a binary symmetric channel (BSC), the capacity C of the channel is evaluated using the relation $1 - H(p)$. The rate $R < C$ determines which parity check matrix has to be used for a particular bit plane. In our scheme, we use irregular Gallager LDPC codes and we follow the procedure suggested by Neal [12] to construct LDPC parity check matrix.

2.2. Slepian-Wolf Decoder

The joint decoding algorithm proposed in this paper is illustrated in Fig.1 (b). The image Y is reconstructed independently into $\bar{Y}$, by arithmetic decoding, SPIHT decoding, and inverse wavelet transform. The bit planes of X are decoded successively, and are used to refine the side information by motion estimation on the sphere. Let assume that the first l bit planes of X are successfully decoded and the image reconstructed using the first l bit planes is denoted by $\bar{X}^l$. Similarly, the decoder can reconstruct an approximation $\bar{Y}^l$ of the image Y by decoding the first l bitplanes. The correlation between images is then estimated by ME, performed on the approximated images $\bar{X}^l$ and $\bar{Y}^l$. In more details, the ME is performed on the 2-sphere as follows. The spherical image $\bar{X}^l$ is initially divided into blocks of solid angles. For each block in image $\bar{X}^l$ a best matching block in the mean square sense is selected in the image $\bar{Y}^l$. The displacement between the blocks is represented as a motion on the sphere. The generated motion vectors are then used to compensate the motion in the reconstructed image $\bar{Y}$, and form an approximation of the image X. The motion compensated image undergoes a WT, and the $(l+1)^{th}$ bit plane is extracted using the significance tree of X, which has been lossless encoded. The extracted $(l+1)^{th}$ bit plane finally serves as side-information for decoding the $(l+1)^{th}$ bit plane of X using the LDPC sum product algorithm. After the refinement bits are successfully decoded, the decoder applies inverse wavelet transform to reconstruct the image $\bar{X}^{l+1}$, and the process iterates until all bit planes are decoded.

As the expected number of errors in the most significant bit planes is usually small, these bit planes are decoded without ME and serve as an initialization point of the proposed method. The least significant bit planes that carry more detailed scene information lead to higher number of errors, with respect to the side information provided by the reference image. In the proposed method, the refinement of side information with ME actually gives an improved correlation model between corresponding bit planes in different images

since it considers the local motion of objects, and therefore reduces the number of errors that need to be corrected with the LDPC codes.

3. EXPERIMENTAL RESULTS

We evaluated the performance of our system on two synthetic spherical images: Rhino and Room, shown in Fig.2 and Fig.3 respectively. The Daub 9/7 integer wavelet filter is used to transform images X and Y with three levels of decomposition and it is implemented by the lifting scheme on unwrapped images. Adaptive arithmetic coder is used to compress significant bits. Motion estimation is carried out on blocks of size 4. In order to save on the computational complexity, the search for the best block is limited to a window size of 64. Motion estimation is carried out after decoding first two bit planes of X, inferred directly from $\bar{Y}$.
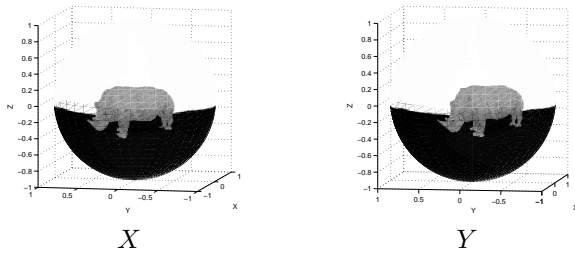

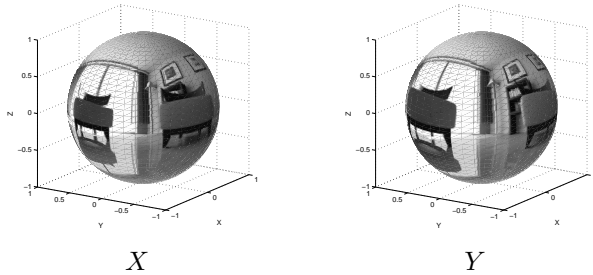**Fig. 2**. Original Rhino images.

Fig. 3. Original Room images.

We first compare the performance of the proposed DSC solution to a separate SPIHT coding scheme by comparing the rate distortion (RD) behavior of coding the image X, when the Y image is coded at 1.7 bpp. Fig. 4 and Fig. 5 show the RD curve for Rhino and Room images, respectively. As expected, coding with side information outperforms the independent coding of X, and the gain can be as high as 7 or even 10 dB at low bit rate. The degradation of coding gain in the high bit rate region is due to the fact that the bit plane level correlation reduces as we move towards the least significant bit planes, which requires low rate LDPC codes. This behavior is quite expected since the fine details encoded in the least significant bit planes are usually not correlated across different views. For example, in the case of the Room images the correlation greatly decays after the third bit plane. Therefore, we encode first three bit planes with the proposed DSC scheme, while the rest of bit planes are encoded independently, without DSC. Hence, the performance of our scheme is slightly lower for the Room image than for the Rhino image. The reconstructed Rhino and Room images $\bar{X}$ are shown in Fig.6 and Fig.7, respectively.
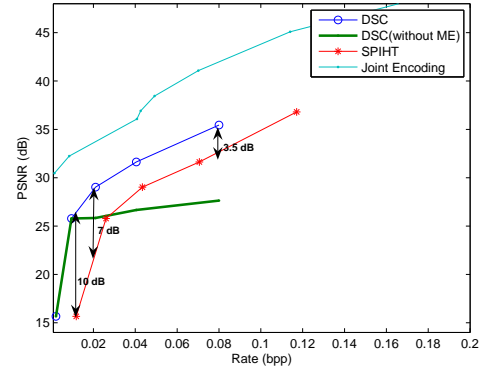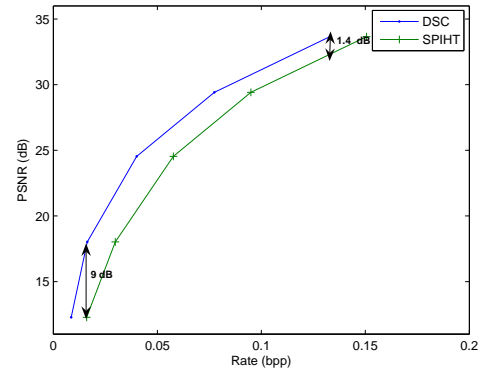

Fig. 4. RD plot of the Rhino image X

**Fig. 5**. RD plot of the Room image X

Table 1. Error in refinement bits, with and w/o motion estimation

Bit plane No.	3	4	5
With ME	67	106	277
Without ME	101	212	443

In order to evaluate the loss in performance due to separate encoding, we also compare the proposed scheme with joint encoding of X and Y images (Fig. 4). Joint encoding, with multiresolution ME at the encoder, obviously performs better than the distributed coding solution but can only be performed with increased complexity at the encoders. In the meantime, DSC performs better than the less complex independent coding solution. Thus DSC framework is a trade off between the complexity at the encoders and the performance.

Then, we compare our scheme with a DSC scheme based also on SPIHT and LDPC, but without ME at the decoder (Fig. 4). It highlights the benefits of ME at the decoder, since it is clearly mandatory for good reconstruction, especially for medium bit rates (i.e., for decoding of upper bit planes). The Table 1 clearly shows that the number of errors in the decoding of upper bit planes drastically increases when ME is not activated at the decoder.

Finally, we extend our scheme to the coding of multiple images, where the decoder uses only one reference image. We consider the scenario where three images X, Y and Z are observed and we
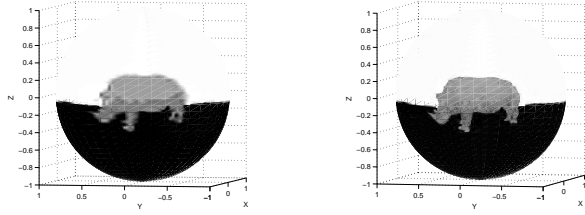
R=0.021 bpp, PSNR=25.31 dB   R=0.08 bpp, PSNR=34.50 dB

Fig. 6. Reconstructed Rhino image X.
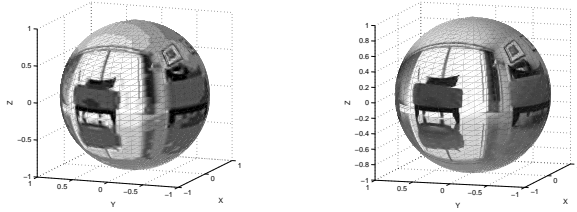
R=0.04 bpp, PSNR=24.53 dB   R=0.13 bpp, PSNR=33.66 dB

Fig. 7. Reconstructed Room image X.

compare the overall RD performance in the two following schemes. First, the image Y is independently coded and the images X and Z are Wyner-Ziv coded as described above. The side information for the images X and Z is generated from $\bar{Y}$ and $\bar{X}$ successively. Second, the images Y and Z are encoded independently as reference images, while the intermediate image X is encoded using Wyner-Ziv coder. At the decoder, the side information for decoding X is synthesized from $\bar{Y}$ and $\bar{Z}$ using motion compensated interpolation and disparity estimation based on epipolar geometry for spherical images. This case corresponds to the disparity-based approach, similar to the one proposed in [8]. Fig. 8 compares the overall RD performance for the three Rhino images, between the two considered cases. We can observe that the first scheme offers better performance since it requires smaller transmission rate due to reduction of the number of independently encoded images. The coding gain at lower rate is 8 dB and 4 dB at higher rate. The results also suggest that it is enough to transmit one view independently and other views with Wyner-Ziv coding to achieve a very good overall performance of the DSC method. Interestingly, such a solution permits to eliminate the need for complex camera calibration that is generally required for disparity estimation.

4. CONCLUSIONS

In this paper we have presented a transform-based DSC scheme for compression of correlated images in omni-directional camera networks. The proposed scheme combines the efficiency of progressive encoding and the correcting power of LDPC codes. The ME step introduced at the decoder to refine the side information significantly improves the performance of our scheme. The proposed method can be easily extended to any number of cameras and does not need camera calibration. Only one reference image is required for decoding of multiple Wyner-Ziv images, leading to improved rate-distortion performance compared to disparity-based method. Our future work includes improvement of the statistical correlation model in order to facilitate the estimation of LDPC parameters.
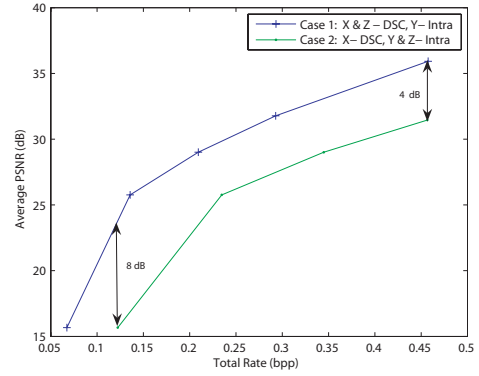
Fig. 8. Overall RD performance for three Rhino images X, Y and Z

5. REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources", IEEE Trans. on Inform. Theory, vol. 19(4), pp. 471-480, July 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side-information at the decoder", IEEE Trans. on Inform. Theory, vol. 22(1), pp. 1-10, Jan. 1976.

[3] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS)", IEEE Trans. on Inform. Theory, vol. 49(3), pp. 626-643, Mar. 2003.

[4] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed video coding", Proceedings of the IEEE, vol. 93(1), pp. 71 - 83, Jan. 2005.

[5] X. Artigas, L. Torres, "Iterative Generation of Motion-Compensated Side Information for Distributed Video Coding", Proc. IEEE ICIP 2005, vol. 1, pp. 833-836, Sept. 2005.

[6] J. Ascenso, C. Brites, F. Pereira, "Motion Compensated Refinement for Low Complexity Pixel Based Distributed Video Coding", Proc. IEEE AVSS 2005, pp. 593-598, Sept. 2005.

[7] A. D. Liveris, Z. Xiong and C. N. Georghiades, "Compression of binary sources with side-information at the decoder using LDPC codes", IEEE Commun. Lett., vol. 6(10), pp. 440-442, Oct. 2002.

[8] X. Zhu, A. Aaron and B. Girod, "Distributed compression for large camera arrays", Proc. IEEE SSP 2003, pp. 30-33, Sept. 2003.

[9] T. Tillo, B. Penna, P. Frossard and P. Vandergheynst, "Distributed Coding of Spherical Images with Jointly Refined Decoding", Proc. IEEE MMSP 2005, pp. 1-4, Nov. 2005.

[10] J. Kusuma, L. Doherty and K. Ramachandran, "Distributed Compression for Sensor networks", Proc. IEEE ICIP 2001, vol. 1, pp. 82-85, Oct. 2001.

[11] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec using set partitioning in hierarchical trees", IEEE Trans. on Cir. and Sys. for Video Tech., vol. 6(3), pp. 243-250, June 1996.

[12] Methods for constructing LDPC codes: Available in URL http://www.cs.utoronto.ca/pub/radford/LDPC-2001-05-04/pchk.html.

[13] I. Tosic, I. Bogdanova, P. Frossard and P. Vandergheynst, "Multiresolution Motion Estimation for Omnidirectional Images", Proc. of EUSIPCO 2005, Sept. 2005.

[14] N.M. Cheung, C. Tang, A. Ortega, and C.S. Raghavendra, "Efficient waveletbased predictive Slepian-Wolf coding for hyperspectral imagery," Eurasip Journal on Applied Signal Processing, vol. 86(11), pp. 3180-3195, Nov. 2005.