Wyner-Ziv Coding of Light Fields for Random Access

Anne Aaron, Prashant Ramanathan and Bernd Girod
Information Systems Laboratory
Department of Electrical Engineering
Stanford University
{amaaron,pramanat,bgirod}@Stanford.EDU

***Abstract*— Image-based rendering data sets, such as light fields, require efficient compression due to their large data size, but also easy random access when rendering from the data set. Efficient compression usually depends upon prediction between images, which creates dependencies between them, conflicting with the requirement of having easy random access. In this paper we propose to eliminate prediction at the encoder by using Wyner-Ziv coding for compressing the light field images. The images are independently encoded by a Wyner-Ziv encoder. At the receiver, previously reconstructed images are used by the Wyner-Ziv decoder as side information to exploit similarities among images. Simulation results show significant compression performance gains compared to conventional independent image coding while maintaining random access capabilities.**

I. INTRODUCTION

A *light field* [1], [2] is an image-based rendering data set that represents the outgoing radiance from a particular scene or object, at all points in 3-D space and in all directions. This 4-D data set is often parameterized as a 2-D array of images.

Efficient representation is typically a concern with light fields, due to the large amount of data involved. The most efficient compression techniques use disparity compensation, which utilizes geometry information to predict one image from one or more other images [3], [4].

One of the main uses for image-based rendering is in interactive applications, since image-based rendering involves only re-sampling the acquired image data. This is much faster than traditional approaches such as ray-tracing that synthesize scenes from light and surface shading models and scene geometry. In order to allow for re-sampling, there must be random access into the light field at the image level, and often at the pixel level.

While disparity-compensated prediction typically improves compression performance by exploiting the correlation between views, it introduces dependencies between images, which restricts random access to the data. For example, in hierarchical prediction-based schemes, it may be necessary to decode many images to decode a single image. In this case, predictive coding increases the number of images needed to encode one image and may actually increase the overall bit rate. It is shown in [5] that for some viewing scenarios, independently encoding the images is optimal.

In this work we propose Wyner-Ziv coding (source coding with decoder side information) [6], [7] of light field images to achieve both random access and compression efficiency. In the scheme, an image is encoded independent of other images, thus allowing random access. At the receiver, the Wyner-Ziv decoder exploits the similarities among images by using previously decoded images as *side information*. Information theoretic results [6]–[8] suggest that Wyner-Ziv coding of an image using other reconstructed images and geometry information as decoder side information can come close to disparity-compensated predictive coding.

In [9], Jagmohan et al. have proposed coset coding to prevent prediction mismatch in a prediction-based light field compression scheme for streaming. In their system, they store multiple residual and disparity information based on adjacent images and send additional coset bits to make the reconstruction trajectory-independent. In [10], our group also proposes multiple representation coding based on "SP-frames" to allow for random access while achieving compression efficiency.

In this work our Wyner-Ziv coding system for light fields does not require multiple representations but instead sends the appropriate number of bits depending on how good the side information is at the decoder. This minimizes storage and is potentially more robust to losses in a streaming scenario.

The remainder of the paper is organized as follows. In Section II, we describe the proposed Wyner-Ziv compression and rendering system. In Section III, we show compression results for rendering novel views and compare our scheme to conventional independent image coding.

II. WYNER-ZIV COMPRESSION SYSTEM

A. Encoding of Light Field Images

The proposed Wyner-Ziv coding scheme for light field images is shown in Fig. 1. The images of the light field are divided into key images and Wyner-Ziv images. The key images, K, are compressed using conventional image coding. For encoding the Wyner-Ziv images, W, we use a Wyner-Ziv coder similar to the coding schemes proposed for low-complexity video encoding [11], [12] and for distributed compression for large camera arrays [13].

At the encoder, a blockwise DCT is applied to the Wyner-Ziv image W to generate X. The transform coefficients are grouped together to form coefficient bands X_k, where k denotes the coefficient number. Each transform coefficient band is then encoded independently.

For each band X_k, the coefficients are quantized using a uniform scalar quantizer with 2^{M_k} levels. The quantized symbols, q_k, are converted to fixed-length binary codewords, and corresponding bit-planes are blocked together forming M_k bit-plane vectors. Each bit-plane vector is then sent to the Slepian-

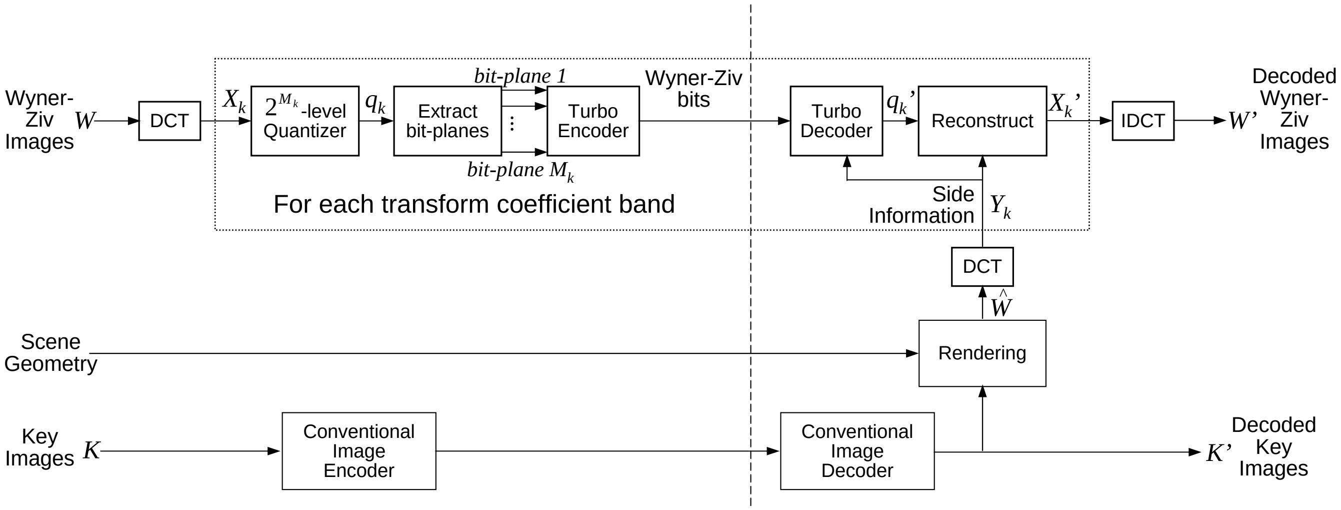


Fig. 1. Wyner-Ziv compression system for light fields

Wolf encoder [14]. The Slepian-Wolf coder is implemented using a rate-compatible punctured turbo code (RCPT) [15], [16]. The RCPT provides rate flexibility which is essential in adapting to the changing statistics between the side information and the image to be encoded. The parity bits produced by the turbo encoder are stored as the compressed version of the Wyner-Ziv image. To decode the image, a subset of these parity bits is used, with the rate dependent on the quality of the side information at the decoder. Note that since the Wyner-Ziv image is coded independent of other images, the scheme has similar random access characteristics as conventional intracoding.

B. Rendering a Novel View

To render one novel view, the receiver typically needs several reconstructed light field images and requests these from the source. The source sends the requested images, which are a combination of key images and Wyner-Ziv images. For the key images, the source sends the bits from the conventional image encoder. For the Wyner-Ziv images, the source sends the appropriate number of turbo parity bits.

At the receiver, conventional image decoding is performed on the key images. To decode a Wyner-Ziv image, the decoder needs to generate side information, that is, a good estimate of the Wyner-Ziv image. The side information image, $\hat{W}$, is generated by rendering from adjacent reconstructed images using the geometry. More adjacent images available at the decoder results in better quality of the rendered side information. The quality of the side information in turn dictates the number of parity bits required to decode the image. In a practical system, the encoder can pre-calculate the number of parity bits required given the available adjacent images at the decoder.

The Wyner-Ziv decoder applies a blockwise DCT on $\hat{W}$ to generate Y. The transform coefficients from Y are grouped together to form coefficient bands Y_k, the side information corresponding to X_k. To be able to use Y_k at the turbo decoder and reconstruction block, the decoder assumes a statistical dependence model between X_k and Y_k.

Given a coefficient band, the turbo decoder successively decodes the bit-planes starting with the most significant bit-plane. It takes the parity bits corresponding to the bit-plane and the side information Y_k to decode the current bit-plane. The probabilities generated for the current bit-plane are then used for decoding the less significant bit-planes. By using the side information Y_k and successively decoding the bit-planes, the decoder needs $R_k \leq M_k$ bits to decode which of the 2^{M_k} bins a transform coefficient belongs to and so compression is achieved.

When all the bit-planes are decoded, the bits are regrouped and the quantized symbol stream is reconstructed as q_k'. The reconstructed coefficient band X_k' is calculated as $E(X_k|q_k', Y_k)$. Assuming that q_k' is error-free, this reconstruction function has the advantage of bounding the magnitude of the reconstruction distortion to a maximum value, determined by the quantizer coarseness. This property is desirable since it eliminates large positive or negative errors for a given transform coefficient. These large errors tend to be very perceptible and annoying to the viewer. The inverse-DCT is then applied to the reconstructed coefficient bands.

Once the key images and Wyner-Ziv images are decoded, the receiver renders the novel view using the reconstructed images and the scene geometry.

III. EXPERIMENTAL RESULTS

We use the *Garfield* light field in our experiments. It consists of 288 images, captured from a hemispherical camera setting containing 8 rows and 32 columns, each at a resolution of 144×192 pixels. All experiments are conducted on the luminance component only. We obtain a reconstructed scene geometry from the original images. The object shape corresponding to each image is derived from the scene geometry. The object shape is used for shape-adaptive coding (skipping the background) in both the the Wyner-Ziv coder and the conventional image coder.

For the experiments, half of the images are designated as key images and the rest are Wyner-Ziv images. The key images are encoded using the shape-adaptive DCT coder described in [17].

For encoding the Wyner-Ziv images, we use a 4×4 DCT and each coefficient band is quantized with a uniform scalar quantizer. We use the same step size for all the coefficient bands. The number of quantizer bins coded for each band determines the bit allocation between bands. For more efficient coding, we use the derived object shape to encode only the blocks containing object pixels.

The turbo encoder is composed of two identical constituent convolutional encoders of rate $\frac{1}{2}$ and generator matrix $[1\ \frac{1+D+D^3+D^4}{1+D^3+D^4}]$ [15]. The parity bits from the convolutional encoder are stored as the compressed representation of the Wyner-Ziv image while the systematic bits are discarded. A subset of these parity bits is sent to the receiver, with the bit rate dependent on the expected quality of the side information available at the decoder. In the experiments we send enough parity bits for a given bit plane such that it can be decoded with bit error rate less than 10^{-3}. In the case where there are channel errors during transmission, the source can send more parity bits to also correct for the errors introduced by channel and achieve the target bit error rate. In the experiments, however, we assume that the parity bits are received without error.

In the current implementation, the decoder only uses adjacent available reconstructed key images to render the side information for the Wyner-Ziv images. Potentially, other reconstructed Wyner-Ziv images can also be used to improve the quality of the side information.

The turbo decoder and reconstruction block assume a Laplacian residual distribution between X_k and Y_k. Let d be the difference between corresponding elements in X_k and Y_k. We observe that the distribution of d can be approximated as $f(d) = \frac{\alpha}{2} e^{-\alpha|d|}$. Each coefficient band has a different α parameter.

To assess the performance, we render single random novel views. This represents the situation where a user wants only a snapshot of the light field. Theoretical and experimental results in [5] show that for this viewing scenario and data set, using no prediction (conventional intracoding) provides the best performance. Prediction-based schemes typically result in sending images which are not directly used in rendering the novel view but are used to decode the other images.

In Fig. 2, we compare the performance of our proposed scheme to that of independently coding all the images using the shape-adaptive DCT coder described in [17]. For each novel view, the four closest images are requested. The plot shows the average rate-distortion performance for rendering 20 random single novel views, where the distortion is calculated with respect to the novel views generated from rendering using the original images. The rate and distortion is averaged over object pixels. The rate does not include the bits required to specify the geometry information, but this rate would be identical for both encoding methods.

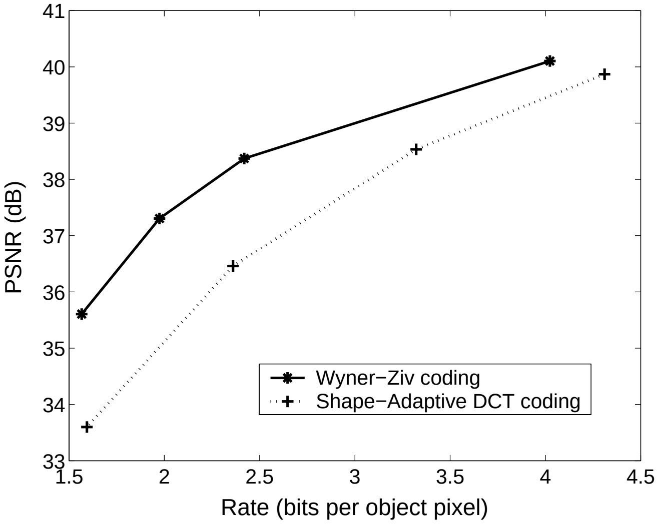


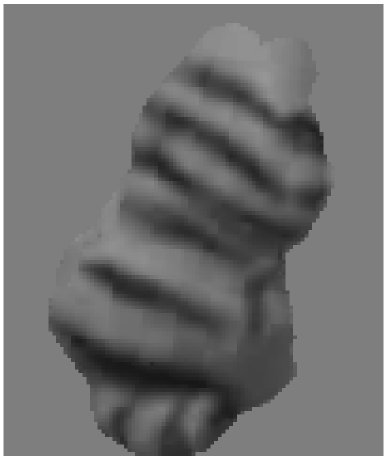
Fig. 2. Bits per rendered object pixel vs. PSNR averaged over 20 random single novel views for *Garfield* lightfield. Wyner-Ziv coding scheme shows gains above independently encoding the images using the shape-adaptive DCT coder.

As it can be seen from the plot, the Wyner-Ziv coding system implemented provides up to 2 dB PSNR improvement over independent shape-adaptive DCT coding, with greater gains at the lower bit rates. This compression gain is achieved without introducing encoding dependencies among the compressed images. We therefore achieve better rate-distortion performance without affecting the random access properties of the data set.

In Fig. 3, we compare the results for a sample novel view. In Fig. 3(a), the novel view is rendered using images independently encoded at a bit rate of 1.5 bits per object pixel (bpop). Fig. 3(c) shows the same novel view rendered from images coded using the proposed Wyner-Ziv scheme at rate 1.4 bpop. For comparison, the same view is rendered using the original images as shown in Fig. 3(b). It can be seen that at similar bit rates, the novel view rendered using intracoded images exhibits more blocking and blurring artifacts compared to the novel view rendered with images compressed using the Wyner-Ziv compression scheme. The Wyner-Ziv novel view is 2.4 dB better in PSNR.

IV. CONCLUSIONS

In this paper we propose Wyner-Ziv coding for light field images to achieve both random access and good compression efficiency. The images are independently encoded using either conventional intracoding or Wyner-Ziv coding. At the decoder, reconstructed adjacent images are used to generate side information for decoding the Wyner-Ziv images. Unlike other schemes which require multiple image representations for random access capabilities, our system works with a single compressed representation at the source and the number of bits sent is determined by the other reconstructed images



(a) Intracoding, 33.8 dB

(b) Original

(c) Wyner-Ziv coding, 36.2 dB

Fig. 3. A sample novel view is rendered using (a) intracoded images at 1.5 bpop (b) the original images and (c) Wyner-Ziv coded images at 1.4 bpop.

available at the decoder. Experimental results show superior performance compared to conventional independent image coding.

ACKNOWLEDGMENT

This work is supported in part by the National Science Foundation Grant No. CCR-0310376 and ECS-0225315 and a C.V. Starr Southeast Asian Fellowship.

REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *Computer Graphics (Proceedings SIGGRAPH96)*, Aug. 1996, pp. 31–42.

[2] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Computer Graphics (Proceedings SIGGRAPH96)*, August 1996, pp. 43–54.

[3] M. Magnor, P. Ramanathan, and B. Girod, "Multi-view coding for image-based rendering using 3-D scene geometry," *IEEE Transactions on Circuits and Systems for Video*, vol. 13, no. 11, pp. 1092–1106, Nov. 2003.

[4] B. Girod, C.-L. Chang, P. Ramanathan, and X. Zhu, "Light field compression using disparity-compensated lifting," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. IV, Hong Kong, China, Apr. 2003, pp. 761–764.

[5] P. Ramanathan and B. Girod, "Rate-distortion analysis for random access for compressed light fields," in *Proc. IEEE International Conference on Image Processing*, Singapore, Oct. 2004, to appear.

[6] A. Wyner, "On source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-21, no. 3, pp. 294–300, May 1975.

[7] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[8] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.

[9] A. Jagmohan, A. Sehgal, and N. Ahuja, "Compression of lightfield rendered images," in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, CA, Nov. 2003.

[10] P. Ramanathan and B. Girod, "Random access for compressed light fields using multiple representations," in *Proc. IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, Sept. 2004, to appear.

[11] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, CA, Nov. 2002.

[12] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *Proc. SPIE Visual Communications and Image Processing*, San Jose, CA, Jan. 2004.

[13] X. Zhu, A. Aaron, and B. Girod, "Distributed compression for large camera arrays," in *Proc. IEEE Workshop on Statistical Signal Processing*, St. Louis, MO, Sept. 2003, pp. 21–24.

[14] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[15] D. Rowitch and L. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo codes," *IEEE Transactions on Communications*, vol. 48, no. 6, pp. 948–959, June 2000.

[16] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Apr. 2002, pp. 252–261.

[17] C.-L. Chang, X. Zhu, P. Ramanathan, and B. Girod, "Shape adaptation for light field compression," in *Proc. IEEE International Conference on Image Processing*, Barcelona, Spain, Sept. 2003.