WYNER-ZIV CODING OF STEREO IMAGES WITH UNSUPERVISED LEARNING OF DISPARITY

David Varodayan, Yao-Chung Lin, Aditya Mavlankar, Markus Flierl and Bernd Girod

Information System Laboratory, Stanford University, Stanford, CA 94305
{varodayan, yao-chung.lin, maditya, mflierl, bgirod}@stanford.edu

ABSTRACT

Wyner-Ziv coding can exploit the similarity of stereo images without communication among the cameras. For good compression performance, the disparity among the images should be known at the decoder. Since the Wyner-Ziv encoder has access only to one image, the disparity must be inferred from the compressed bitstream. We develop an Expectation Maximization algorithm to perform unsupervised learning of disparity at the decoder. Our experiments with natural stereo images show that the unsupervised disparity learning algorithm outperforms a system which does no disparity compensation. It is also superior to conventional compression using JPEG.

1. INTRODUCTION

Colocated pixels from pairs of stereo images are strongly statistically dependent after compensation for disparity induced by the geometry of the scene. Much of the disparity between these images can be characterized as shifts of foreground objects relative to the background. Assuming that the disparity information and occlusions can be coded compactly, joint compression is much more efficient than separate encoding and decoding. Surprisingly, distributed encoding combined with joint decoding can be just as efficient as the wholly joint system, according to the Slepian-Wolf theorem in the lossless case [1] and the Wyner-Ziv theorem in certain asymmetric lossy cases [2]. Distributed compression is preferred because it reduces communication between the stereo cameras. The difficulty, however, lies in discovering and exploiting the scene-dependent disparity at the decoder, while keeping the transmission rate low.

A similar situation arises in low-complexity Wyner-Ziv encoding of video captured by a single camera [3] [4] [5]. These systems encode frames of video separately and decode them jointly, so discovering the motion between successive frames at the decoder is helpful. A very computationally burdensome way to learn the motion is to run the decoding algorithm with every motion realization [4]. Another approach requires the encoder to transmit additional hash information, so the decoder can perform suitable motion compensation before running the decoding algorithm [6]. Since the encoder transmits the hashes at a constant rate, it wastes bits when the motion is small. On the other hand, if there is too much change between frames, the fixed-rate hash may be insufficient for reliable motion search. Due to the drawbacks of excessive computation and difficulty of rate allocation for the hash, we use neither of these approaches for compression of stereo images. Instead, we apply an Expectation Maximization (EM) algorithm [7] at the decoder to learn the disparity in an unsupervised way during decoding.

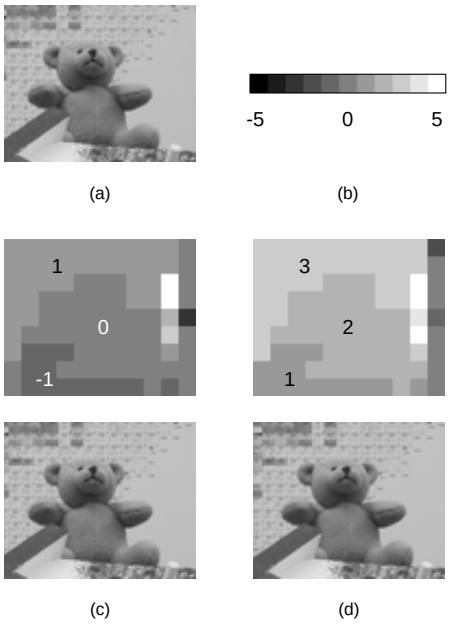

Fig. 1. (a) Source image X (72-by-88 pixels), (b) horizontal disparity legend, (c) and (d) source images Y (72-by-88 pixels) with respective 8-by-8 blockwise horizontal disparity fields D

In Section 2, we review our work on distributed lossless compression with unsupervised disparity learning for pairs of binary random dot stereograms [8] and natural grayscale stereo images [9]. In Section 3, we describe the extension to

This work has been supported, in part, by the Max Planck Center for Visual Computing and Communication and, in part, by a gift from NXP Semiconductors to the Stanford Center for Integrated Systems.

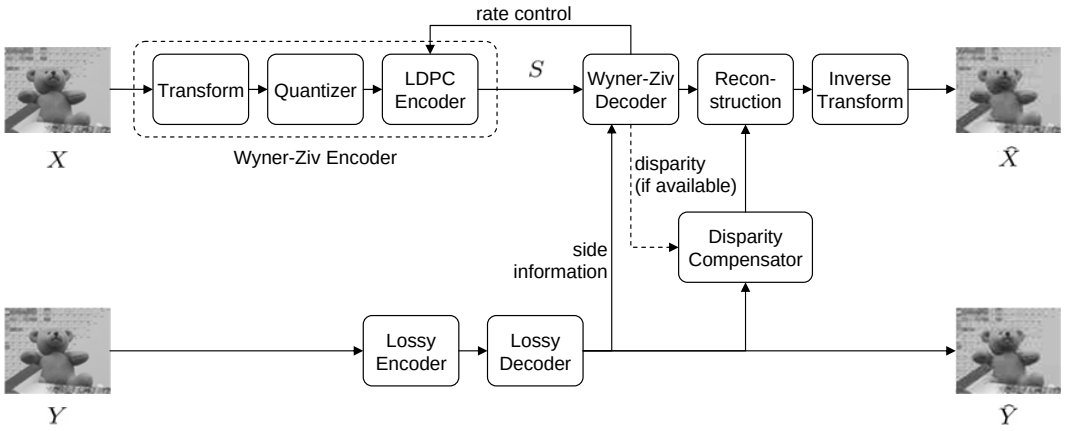

Fig. 2. Wyner-Ziv coding of source X with respect to $\hat{Y}$

lossy Wyner-Ziv coding. Section 4 reports our rate-distortion results for pairs of natural stereo images.

2. BACKGROUND

The relationship between a pair of stereo images X and Y in terms of their disparity D is illustrated in Fig. 1. A sample source image X is depicted in Fig. 1(a). Fig. 1(c) and (d) show two realizations of source image Y taken from different viewpoints [10]. For each pair, the respective block-wise horizontal disparity field D indicates which 8-by-8 block of Y (among candidates shifted up to 5 pixels horizontally) best matches each 8-by-8 block of X (in terms of mean square error). These stereo image pairs, when viewed stereoscopically, create an illusion of depth: various parts of the scene appear on different planes, according to the value of the disparity field in those parts.

For the distributed lossless compression of such stereo images [9], we developed a decoder that performed unsupervised learning of disparity and described it formally within the framework of EM. This system did not exploit the spatial redundancy, but did account for the dependency between bitplanes of individual pixels (via a construction called joint bitplane coding). The precursor work [8] tackled the same problem for pairs of synthetic binary random dot stereograms. Both papers demonstrated that the disparity learning decoder could perform almost as well as an oracle-assisted decoder, and significantly better than a decoder that did no disparity compensation at all.

3. WYNER-ZIV CODING OF STEREO IMAGES

The Wyner-Ziv stereo image coder shown in Fig. 2 extends the previous work in two ways. The system applies transforms for exploiting spatial redundancy and quantization for lossy coding. We now describe the three important components in the codec for X: Wyner-Ziv encoder, Wyner-Ziv decoder and reconstruction.

3.1. Wyner-Ziv Encoder

Just as in JPEG [11], the source image X is transformed by a blockwise 8-by-8 DCT and quantized using a midtread uniform quantizer. The resulting quantized coefficient indices are Slepian-Wolf encoded as in [9], using a low-density parity-check (LDPC) code [12], to produce the syndrome S.

3.2. Wyner-Ziv Decoder

The role of the Wyner-Ziv decoder is to recover the quantized coefficient indices (denoted by Q) from the syndrome S and the lossy coded side information $\hat{Y}$. Fig. 3 depicts three Wyner-Ziv decoders that differ in their handling of disparity.

The baseline decoder in Fig. 3(a) performs no disparity compensation. It initially estimates Q statistically based on the colocated transform coefficients from side information $\hat{Y}$. An iterative belief propagation algorithm refines the estimates using S. This decoder is a concatenation of a LPDC decoder and a joint bitplane decoder (from [9]) that exploits dependency among bitplanes of Q. Since disparity exists between $\hat{X}$ and $\hat{Y}$, this scheme does not perform well because the estimates of Q from $\hat{Y}$ are poor in regions of nonzero disparity.

For comparison, Fig. 3(b) shows an impractical scheme in which the decoder is endowed with a disparity oracle. The oracle compensates $\hat{Y}$ blockwise to align with blocks of $\hat{X}$. Now the initial statistical estimates of Q use the disparity-compensated version of $\hat{Y}$ as side information and, thus, do not suffer in regions of nonzero disparity.

Finally, Fig. 3(c) depicts the practical decoder that performs unsupervised disparity learning via EM. In place of the disparity oracle, a disparity estimator maintains an *a posteriori* probability distribution on disparity D. Every iteration of LDPC/joint bitplane decoding sends the disparity estimator a

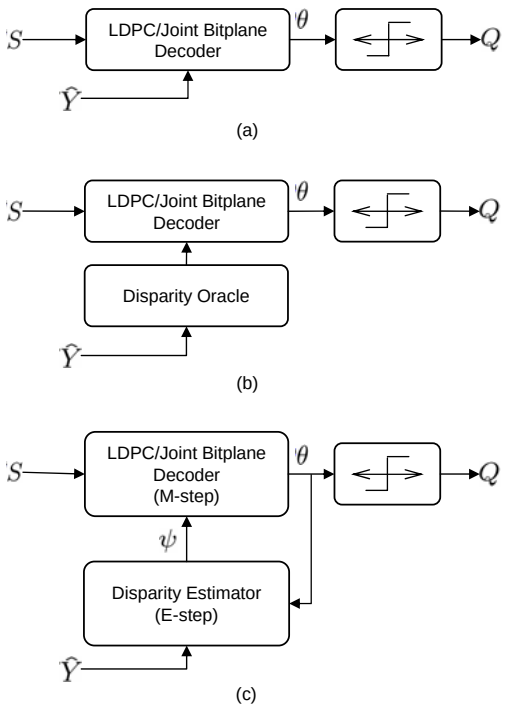

Fig. 3. Wyner-Ziv decoding of quantized coefficient indices Q with (a) no disparity compensation, (b) a disparity oracle, and (c) unsupervised learning of disparity D via EM

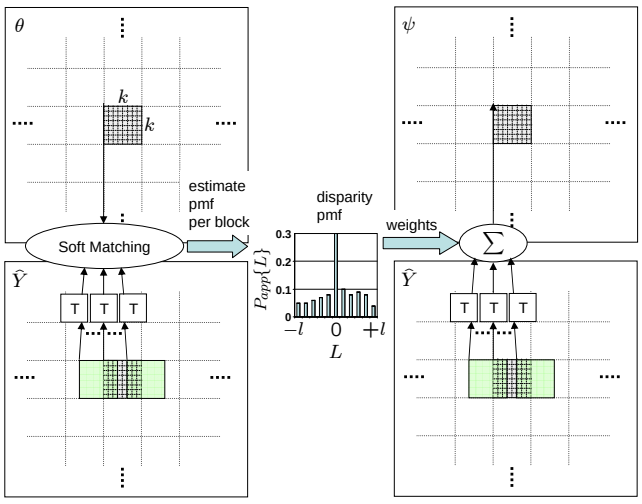

Fig. 4. Disparity field estimation (left) and side information blending (right)

soft estimate of Q (denoted by θ) in order to refine the distribution on D. In return, the disparity estimator updates the probabilistic side information estimate ψ for the LDPC/joint bitplane decoder by blending information from the transform coefficients of shifted blocks of $\hat{Y}$ according to the refined distribution on D. Fig. 4 shows the disparity field estimation and side information blending in greater detail. As depicted on the left-hand side, each block of θ (the soft estimate of Q) is matched with the candidate transformed blocks of $\hat{Y}$ to produce the *a posteriori* probability distribution on D for that block. On the right-hand side, the same candidate transformed blocks of $\hat{Y}$ are blended together according to this distribution to create the more accurate side information ψ. (For details we refer the reader to [9], noting that each coefficient band has its own Laplacian statistics.) These steps iterate with LDPC/joint bitplane decoding to jointly recover the quantized coefficient indices Q and the disparity field D.

3.3. Reconstruction

The reconstruction block recovers $\hat{X}$ from the quantized coefficient indices Q and the side information $\hat{Y}$, possibly compensated by the disparity output of the Wyner-Ziv decoder. Conventional reconstruction (with no side information) would map Q to the probabilistic centroids of the respective quantization intervals. Since side information is available, reconstruction can be improved [3]. If a value of Q matches the colocated quantized coefficient index in (possibly disparity-compensated) $\hat{Y}$, that reconstructed coefficient is set equal to the colocated coefficient. If the value of Q does not match the colocated quantized coefficient index, the coefficient is reconstructed at the boundary of its quantization interval closest to the colocated coefficient.

4. SIMULATION RESULTS

We compare the performance of coding X for the three Wyner-Ziv schemes that differ in their handling of disparity (in Fig. 3) as well as baseline JPEG. For ease of comparison, we use 8-by-8 blocks for transforms (and consequently disparity estimation) and scaled versions of the quantization matrix in Annex K of [11] with scaling factors 0.5, 1, 2 and 4. In each case, Y is JPEG-coded with the same quantization matrix. The maximum horizontal disparity shift is 5. For the disparity learning decoder, a good initialization of disparity shift distribution has a peak of 0.75 at zero and is uniform elsewhere.

Rate control is implemented by using rate-adaptive LDPC accumulate codes of length 50688 bits [13]. After 150 decoding iterations, if the recovered Q does not satisfy the syndrome condition, the decoder requests additional incremental transmission from the encoder via a feedback channel.

Figs. 5 and 6 show rate distortion curves for coding X from Fig. 1(a) with respect to $\hat{Y}$ obtained from Fig. 1(c) and (d), respectively. The four points on each curve result from the four quantization scaling factors, so that corresponding points on different curves belong to encodings with identical quantized coefficient indices Q. For each encoding, the practical unsupervised disparity learning scheme comes close in both rate and distortion to the impractical oracle-assisted scheme. The rate loss is incurred at the Wyner-Ziv decoder, and the PSNR loss at the reconstruction block. In both of

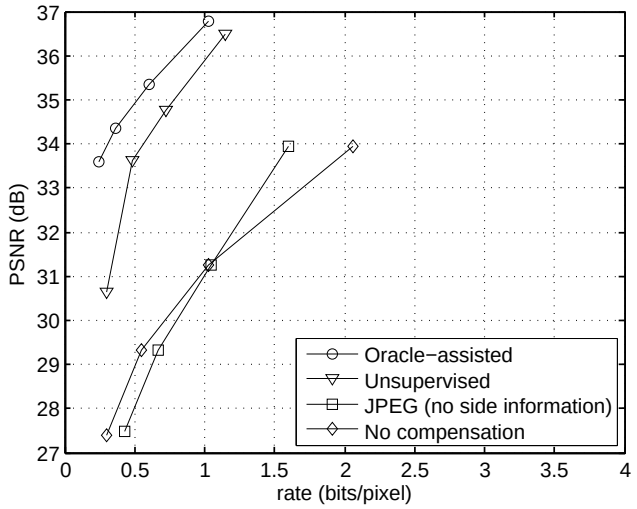

Fig. 5. Rate-distortion curves for coding X from Fig. 1(a) with respect to $\hat{Y}$ obtained from Fig. 1(c)

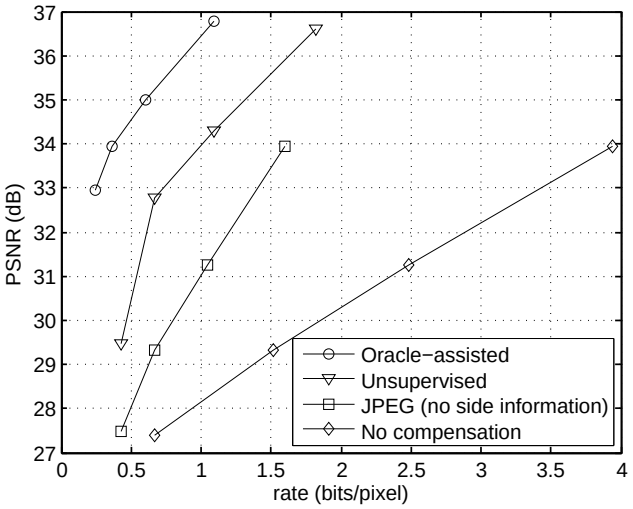

Fig. 6. Rate-distortion curves for coding X from Fig. 1(a) with respect to $\hat{Y}$ obtained from Fig. 1(d)

these ways, the unsupervised disparity learning scheme is superior to the scheme with no disparity compensation. In fact, the uncompensated side information $\hat{Y}$ is so poor for reconstruction that our plots show conventional centroidal reconstruction for the scheme with no compensation. The rate-distortion curves for JPEG are identical in both figures because the source X is the same, but in both cases worse than the curves for the unsupervised disparity learning scheme by 2 to 5 dB. The three Wyner-Ziv schemes perform worse in Fig. 6 than in Fig. 5 because less of the disparity field matches the initialization peaked at zero. But the scheme with no compensation suffers most dramatically because it cannot compensate for the greater disparity mismatch.

5. CONCLUSION

We extend distributed compression with unsupervised learning of disparity at the decoder to the lossy transform-domain case. For natural stereo images, our proposed practical Wyner-Ziv scheme shows rate-distortion performance between 2 to 5 dB superior to conventional JPEG compression, and similar or greater performance gaps relative to a Wyner-Ziv scheme with no disparity compensation. Future work should explore test images with greater scene complexity and disparity range, and investigate reduction of computational load.

6. REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 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. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[3] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conf. on Signals, Syst., Comput.*, Pacific Grove, CA, 2002.

[4] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Proc. Allerton Conf. Commun., Contr. and Comput.*, Allerton, IL, 2002.

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

[6] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver," in *Proc. IEEE Internat. Conf. Image Processing*, Singapore, 2004.

[7] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc., Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[8] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod, "Distributed coding of random dot stereograms with unsupervised learning of disparity," in *Proc. IEEE Internat. Workshop Multimedia Signal Processing*, Victoria, BC, Canada, 2006.

[9] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod, "Distributed grayscale stereo image coding with unsupervised learning of disparity," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, 2007.

[10] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. Comput. Vision and Pattern Recog.*, Madison, WI, 2003.

[11] ITU-T and ISO/IEC JTC1, "Digital compression and coding of continuous-tone still images," *ISO/IEC 10918-1 — ITU-T Recommendation T.81 (JPEG)*, Sept. 1992.

[12] R. G. Gallager, "Low-density parity-check codes," *Cambridge MA: MIT Press*, 1963.

[13] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," in *Proc. Asilomar Conf. on Signals, Syst., Comput.*, Pacific Grove, CA, 2005.