# Distributed Coding of Random Dot Stereograms with Unsupervised Learning of Disparity

David Varodayan, Aditya Mavlankar, Markus Flierl and Bernd Girod
Max Planck Center for Visual Computing and Communication
Stanford University, Stanford, CA 94305
Email: {varodayan, maditya, mflierl, bgirod}@stanford.edu

***Abstract*— Distributed compression is particularly attractive for stereoscopic images since it avoids communication between cameras. Since compression performance depends on exploiting the redundancy between images, knowing the disparity is important at the decoder. Unfortunately, distributed encoders cannot calculate this disparity and communicate it. We consider a toy problem, the compression of random dot stereograms, and propose an Expectation Maximization algorithm to perform unsupervised learning of disparity during the decoding procedure. Our experiments show that this can achieve twice as efficient compression compared to a system with no disparity compensation and perform nearly as well as a system which knows the disparity through an oracle.**

## I. Introduction

Colocated pixels from pairs of stereoscopic 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 lossless compression is much more efficient than separate lossless encoding and decoding. Surprisingly, distributed lossless encoding combined with joint decoding can be just as efficient as the wholly joint system, according to the Slepian-Wolf theorem [1]. Distributed compression is preferred because it avoids 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 [2] [3]. 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 [3]. Another approach requires the encoder to transmit additional hashed information, so the decoder can select a good motion configuration before running the decoding algorithm [4]. 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 towards compression of stereoscopic images.

This work has been supported by the Max Planck Center for Visual Computing and Communication.

In Section II, we consider a toy version of the problem, involving random dot stereograms, and propose a novel decoding algorithm, which learns disparity unsupervised. We describe the algorithm formally within the framework of Expectation Maximization (EM) [5] in Section III. Section IV reports our simulation results.

## II. Random Dot Stereogram Compression

In this paper, we consider the compression of pairs of binary random dot stereograms [6]. Viewed stereoscopically as a single image, they create an illusion of depth: a rectangle appears on a different plane to the rest of the image. Although compression of random dot stereograms may be considered an academic problem with no immediate practical relevance, we are intrigued by their properties and believe that their study can yield important insights into distributed coding. Random dot stereograms have well-defined statistical properties and information theoretic bounds that can be computed easily. Each image by itself resists compression, but substantial savings are possible by exploiting the relative shift of the content. Finally, since the images are binary, coding can be performed on the pixel values directly.

For a pair of random dot stereograms $X$ and $Y$, we specify the depth illusion by disparity information $D$ as described below. We generate $Y$ by copying $X$ and horizontally shifting a rectangular region. The uncovered area is filled randomly and independent identically distributed (i.i.d.) binary noise is added (modulo 2) to $Y$. Thus, $D$ comprises the boundaries of the shifted rectangle as well as the horizontal shift value. Fig. 1 shows realizations of $X$ and $Y$ and their modulo 2 sums under different shifts.

Our compression setup is shown in Fig. 2. Images $X$ and $Y$ are encoded separately and decoded jointly. For simplicity, we assume that $Y$ is conventionally coded and is available at the decoder. The challenge is to encode $X$ efficiently in the absence of $Y$ so that it can be reliably decoded in the presence of $Y$. The Slepian-Wolf theorem states that $X$ can be communicated losslessly to the decoder using $R$ bits on average as long as $R > H(X|Y)$ [1].

Fig. 3 depicts three compression systems that can be applied to this problem. The baseline system in Fig. 3(a) is due to [7] and performs compression of $X$ with respect to the colocated

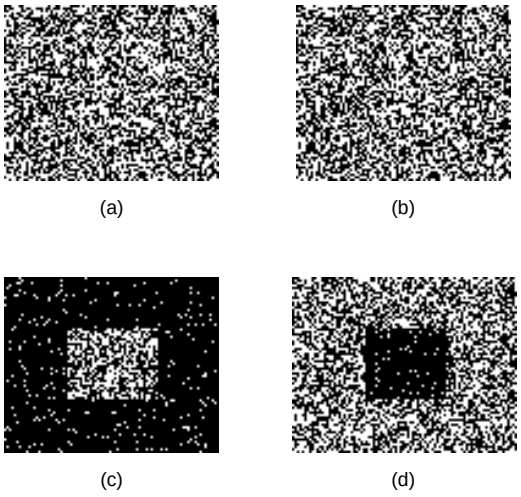


Fig. 1. (a) Source image $X$ (b) Source image $Y$ (c) Sum of $X$ and $Y$ modulo 2 (d) Sum of $X$ and $Y$ modulo 2 (shifted to realign the shifted rectangle)

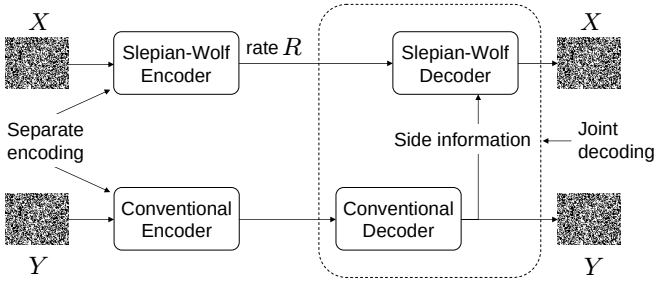


Fig. 2. Distributed compression: separate encoding and joint decoding

pixels of $Y$ without disparity compensation. The encoder computes the syndrome $S$ (of length $R$ bits) of $X$ with respect to a low-density parity-check (LDPC) code [8]. The decoder initially estimates $X$ statistically using the colocated pixels of $Y$ and refines these estimates using $S$ via an iterative belief propagation algorithm. When disparity is introduced between $X$ and $Y$, this scheme performs badly because the estimates of $X$ are poor in the shifted region. For comparison, Fig. 3(b) shows an impractical scheme in which the decoder is endowed with a disparity oracle. The oracle informs the decoder which pixels of $Y$ should be used to inform the estimates of the pixels of $X$ during LDPC decoding. Finally, Fig. 3(c) depicts our proposed practical decoder that learns disparity $D$ via EM. In place of the disparity oracle, a disparity estimator maintains an *a posteriori* probability distribution on $D$. Every iteration of LDPC decoding sends the disparity estimator a soft estimate of $X$ (denoted by $\theta$) in order to refine the distribution on $D$. In return, the disparity estimator updates the side information $\psi$ for the LDPC decoder by blending information from the pixels of $Y$ according to the refined distribution on $D$. The following section formalizes the process in terms of EM.

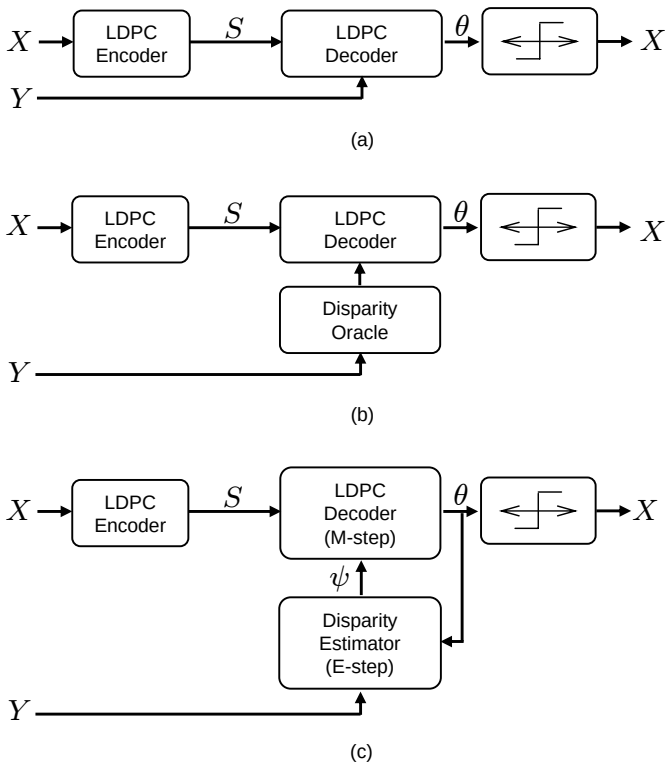


Fig. 3. (Distributed compression with (a) no disparity compensation, (b) a disparity oracle, and (c) unsupervised learning of disparity $D$ via EM

## III. Expectation Maximization Algorithm

### A. Model

Let $X$ be a binary image of size $m$-by-$n$, in which pixels $X(i,j)$ form an i.i.d. equiprobable Bernoulli random process. Let $L$ be the random horizontal shift variable with $|L| \leq l$, where $l \ll n$ is its maximum possible magnitude. Define also random indices $M_1 \leq M_2$ and $N_1 \leq N_2$ to be the vertical and horizontal boundaries of the disparity region, respectively. Thus, $D$ is the 5-tuple $(L, M_1, M_2, N_1, N_2)$. Let $R$ and $Z$ be $(M_2 - M_1 + 1)$-by-$(N_2 - N_1 + 1)$ and $m$-by-$n$ binary images, respectively, where $R(i,j)$ and $Z(i,j)$ form i.i.d. Bernoulli random processes with $P\{R(i,j) = 1\} = 0.5$ and $P\{Z(i,j) = 1\} = \epsilon \leq 0.5$. Generate the image $Y$ as follows using $R$ to fill in the uncovered area and $Z$ as modulo 2 additive noise. Notice that the pixels $Y(i,j)$ form an i.i.d. equiprobable Bernoulli random process.

$$\begin{aligned} Y &:= X \\ Y(M_1 : M_2, N_1 : N_2) &:= R \\ Y(M_1 : M_2, N_1 + L : N_2 + L) &:= X(M_1 : M_2, N_1 : N_2) \\ Y &:= Y \oplus Z \end{aligned}$$

We model the decoder's *a posteriori* probability distribution of source $X$ based on parameters $\theta$ as

$$\begin{aligned} P_{ap}\{X\} &= P\{X;\theta\} \\ &= \prod_{i,j} \theta(i,j)\mathbf{1}_{[X(i,j)=1]} + (1-\theta(i,j))\mathbf{1}_{[X(i,j)=0]} \end{aligned}$$

where $\theta(i,j) = P_{ap}\{X(i,j)=1\}$ is a soft estimate of $X(i,j)$ and $\mathbf{1}_{[.]}$ denotes the indicator function.

Although this is meant to be a toy model, the restriction that the shift $L$ be small and in one dimension is reasonable for a pair of closely-spaced cameras.

### B. Problem

The decoder aims to calculate the *a posteriori* probability distribution of the disparity $D$,

$$\begin{aligned} P_{ap}\{D\} &:= P\{D|Y,S;\theta\} \\ &\propto P\{D\}P\{Y,S|D;\theta\}, \end{aligned}$$

with the second step by Bayes' Law. The form of this expression suggests an iterative EM solution. The E-step updates the disparity distribution with reference to the source model parameters, while the M-step updates the source model parameters with reference to the disparity distribution.

### C. E-step Algorithm

The E-step update (before renormalization) is written as

$$P_{ap}^{(t+1)}\{D\} := P_{ap}^{(t)}\{D\}P\{Y,S|D;\theta^{(t+1)}\}.$$

But this operation is expensive due to the large number of possible values of $D$. We simplify in two ways. First, we ignore knowledge of the syndrome $S$ since it is exploited in the M-step of LDPC decoding. Second, we permit the estimation of the horizontal shift $L$ on a block-by-block basis only. For a specified blocksize $k$, every $k$-by-$k$ block of $\theta$ is compared to the colocated block of $Y$ as well as all those shifted between $-l$ and $l$ pixels horizontally. For a block $\theta_{u,v}$ with top left pixel located at $(u,v)$, the distribution on the shift $L_{u,v}$ is updated as below and normalized:

$$P_{ap}^{(t+1)}\{L_{u,v}\} := P_{ap}^{(t)}\{L_{u,v}\}P\{Y_{u,v+L_{u,v}}|L_{u,v};\theta_{u,v}^{(t+1)}\},$$

where $Y_{u,v+L_{u,v}}$ is the $k$-by-$k$ block of $Y$ with top left pixel at $(u, v+L_{u,v})$. Note that $P\{Y_{u,v+L_{u,v}}|L_{u,v};\theta_{u,v}\}$ is the probability of observing $Y_{u,v+L_{u,v}}$ given that it was generated through shift $L_{u,v}$ from $X_{u,v}$ as parameterized by $\theta_{u,v}$. This procedure, shown in the left hand side of Fig. 4, occurs in the disparity estimator of Fig. 3(c).

### D. M-step Algorithm

The M-step updates the model parameters $\theta$ by maximizing the likelihood of $Y$ and the syndrome $S$.

$$\begin{aligned} \theta^{(t+1)} &:= \arg\max_{\theta} P\{Y,S;\theta^{(t)}\} \\ &= \arg\max_{\theta} \sum_{d} P_{ap}^{(t)}\{D=d\}P\{Y,S|D=d;\theta^{(t)}\} \end{aligned}$$

True maximization is intractable, so we approximate it with an iteration of LDPC decoding. The LDPC decoder's side information $\psi_{u,v}$ is created by blending estimates from each of the blocks $Y_{u,v+L_{u,v}}$ according to $P_{ap}^{(t)}\{L_{u,v}\}$, as shown in the right hand side of Fig. 4. More generally, this is

$$\begin{aligned} \psi(i,j) &= \sum_{d} P_{ap}^{(t)}\{D=d\}P\{X(i,j)=1|D=d,Y\} \\ &= E_{D}^{(t)}\left[(1-\epsilon)\mathbf{1}_{[Y(i,j+D)=1]} + \epsilon\mathbf{1}_{[Y(i,j+D)=0]}\right]. \end{aligned}$$

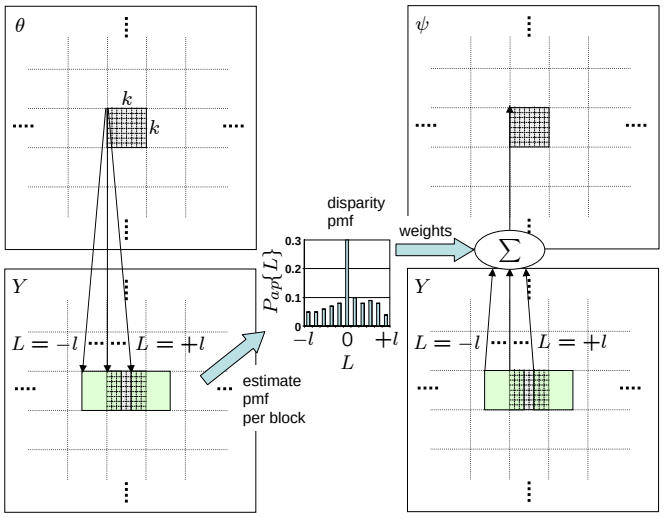


Fig. 4. E-step disparity estimation (left) and side information blending (right)

### E. Termination

Iterating between the E-step and the M-step in this way provides a coarse profile of the disparity, limited by the granularity of $k$-by-$k$ blocks. Once several contiguous blocks agree upon a value for $L$, we refine the estimate of $D$ by estimating the disparity region boundaries $\{M_1, M_2, N_1, N_2\}$. For simplicity, instead of maintaining probability distributions, we estimate a single value for each boundary variable and refine it at every iteration. This improves the quality of the compensation of side information. The decoding algorithm terminates successfully when the thresholded estimates $\hat{X}(i,j) = \mathbf{1}_{[\theta(i,j)>0.5]}$ yield syndrome equal to $S$.

## IV. Simulation Results

For our simulations, we select the following constants: image height $m = 72$, image width $n = 88$, maximum horizontal shift $l = 5$, blocksize $k = 8$. The camera noise parameter $\epsilon = P\{Z(i,j) = 1\}$ ranges between $0.01$ and $0.11$. The distributions of $L_{u,v}$ are initialized to

$$P_{ap}^{(0)}\{L_{u,v}\} := \begin{cases} 0.75, & \text{if } L_{u,v} = 0; \\ 0.025, & \text{if } L_{u,v} \neq 0. \end{cases}$$

Rate control is implemented using rate-adaptive regular degree 3 LDPC accumulate codes of length 6336 bits [9]. After 150 decoding iterations, if $\hat{X}$ still does not satisfy the syndrome condition, the decoder requests additional incremental transmission from the encoder via a feedback channel.

Figs. 5 and 6 show the compression performance of the systems in Fig. 3 and the Slepian-Wolf bound for different levels of additive noise, when the disparity region has size 32-by-32 pixels and the shift $L$ is 1 and 5, respectively. For the proposed scheme, we show results when the disparity region is aligned with the 8-by-8 block grid (best case) and when it is offset from the grid by 4 pixels horizontally and vertically (worst case). Fig. 7 shows the performance when the disparity shift $L$ varies uniformly over $-5 \leq L \leq 5$.

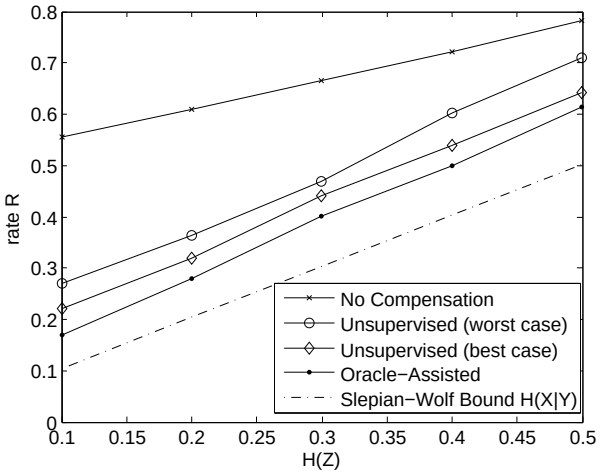


Fig. 5. Rate (in bit/pixel) required to communicate $X$ for the different systems shown in Fig. 3 when learning the disparity shift $L = 1$.

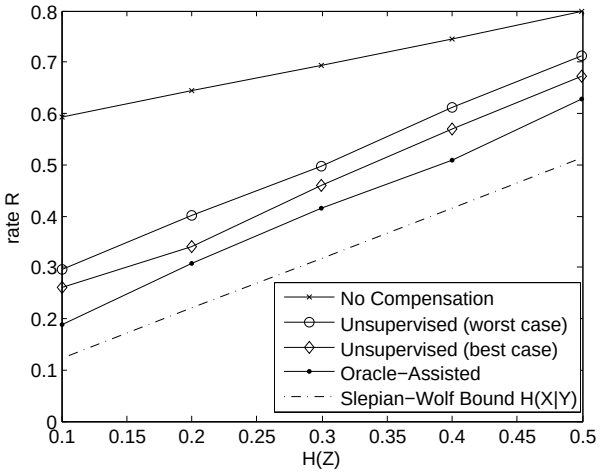


Fig. 6. Rate (in bit/pixel) required to communicate $X$ for the different systems shown in Fig. 3 when learning the disparity shift $L = 5$.

Figs. 5 and 6 demonstrate that a larger shift $L$ moves all the rate curves and the Slepian-Wolf bound upward, since it uncovers a larger region of the source image $X$, but the performance gaps stay approximately constant. The results in Fig. 7 show that unsupervised learning of disparity can achieve performance twice as good as the system that allows no disparity compensation, for a 32-by-32 disparity region in a 72-by-88 image. Moreover, it only incurs a small performance loss with respect to the impractical oracle-assisted scheme. The further gap to the Slepian-Wolf bound is due to the inefficiency of short length regular LDPC codes.

To illustrate the progress of the unsupervised learning decoding algorithm, we show how the disparity probability distribution evolves for a sample 8-by-8 block in Fig. 8.

## V. Conclusions

We address the problem of distributed stereoscopic compression with disparity learning at the decoder. For distributed

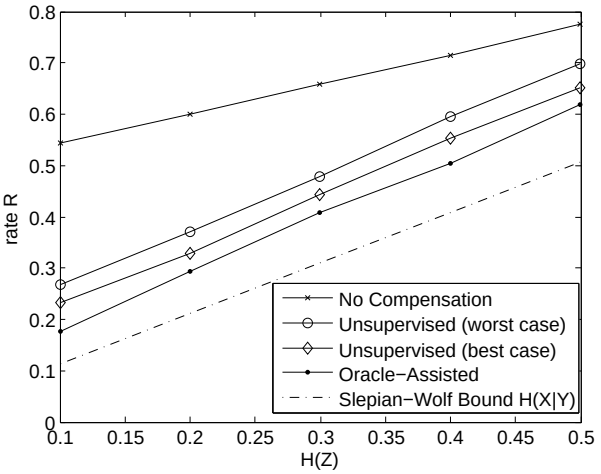


Fig. 7. Rate (in bit/pixel) required to communicate $X$ for the different systems shown in Fig. 3, for disparity shift uniform over $-5 \leq L \leq 5$.

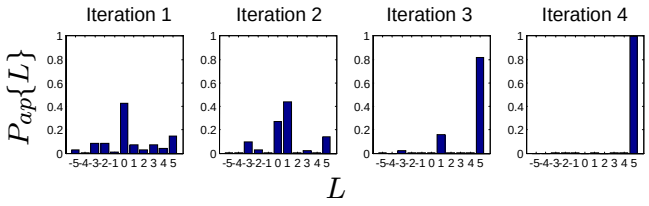


Fig. 8. Evolution of a disparity probability distribution for a sample 8-by-8 block with true disparity shift $L = 5$.

compression of random dot stereograms, we develop an iterative EM algorithm that alternates between updating the disparity distribution and the source model. We show that unsupervised learning of disparity is a practical method that can achieve twice as efficient compression compared to a system that does not compensate for disparity.

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

[3] 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.

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

[5] 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.

[6] B. Julesz, "Binocular depth perception of computer generated patterns," *Bell Sys. Tech. J*, vol. 38, pp. 1001–1020, 1960.

[7] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, Oct. 2002.

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

[9] 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.