DISTRIBUTED STEREO IMAGE CODING WITH IMPROVED DISPARITY AND NOISE ESTIMATION

David Chen, David Varodayan, Markus Flierl, Bernd Girod

Information Systems Laboratory, Stanford University, Stanford, CA 94305
{dmchen, varodayan, mflierl, bgirod}@stanford.edu

ABSTRACT

Distributed coding of correlated grayscale stereo images is effectively addressed by a recently proposed codec that learns block-wise disparity at the decoder. Based on the Slepian-Wolf theorem, one image can be transmitted at a rate approaching the conditional entropy if the other image is referenced as side information at the decoder. This paper improves the methods in the decoder design by refining disparity estimates to pixel resolution, generating more accurate initial disparity estimates, and modeling noise as a nonstationary random field. The new decoder enables up to an additional 9 percent bit rate savings for lossless coding. When the rate is insufficient for lossless reconstruction, the new decoder improves PSNR and significantly reduces visually unpleasant blocking artifacts.

Index Terms— image coding, data compression, estimation, stereo vision

1. INTRODUCTION

A pair of stereo images, taken from two source cameras in a camera array, share many common details. Conventionally, to transmit these images, joint source coding would be performed to achieve better compression than separate coding of the two images. The assumption, however, is that the joint distribution of the two images is available at the encoder. In a distributed coding scenario, where the source cameras perform only limited computations or avoid interacting with one another to conserve limited communication bandwidth, conventional joint coding is not practical. Fortunately, the Slepian-Wolf theorem from information theory states that lossless distributed coding can achieve the same compression ratio as lossless joint coding, provided that the joint distribution becomes available at the decoder [1].

An existing codec treats the two source images as random fields which are related through a disparity shift and additive noise, and successfully performs unsupervised learning of the disparity at the decoder [2]. It is an extension of earlier work for correlated binary images [3]. Expectation maximization (EM) is applied to iteratively estimate the disparity and progressively decode one image using the other image as side information. In this statistical framework, the efficiency of the entire system is primarily limited by the accuracy of the disparity and noise estimates.

This paper presents improved techniques in disparity and noise estimation. In Section 2, the distributed stereo image codec and the limitations of its design are reviewed. In Section 3, improvements in disparity and noise estimation are presented in three parts. First, a pixel-wise disparity field is created by bilinear interpolation of a block-wise disparity field. Second, shape-adaptive disparity initialization is shown to be beneficial in the early iterations of EM. Third, an edge-adaptive noise estimator is used to accurately predict the noise variance as a function of spatial position. Section 4 presents experimental results that show bit rate savings of up to 9 percent for lossless coding. At rates insufficient for lossless reconstruction, the improved decoder shows several dB gain in PSNR and significantly reduces blocking artifacts.

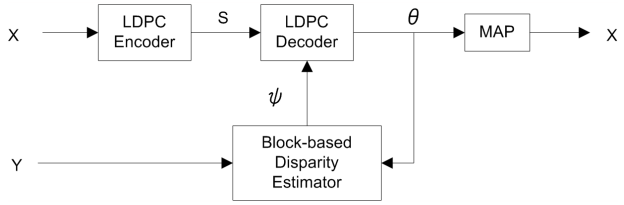

Fig. 1. Distributed stereo image codec.

2. DISTRIBUTED STEREO IMAGE CODEC

The main blocks of the distributed stereo image codec in [2] are depicted in Fig. 1. One image Y is losslessly transmitted using a conventional lossless encoder-decoder combination and is then available at the decoder as side information. The other image X is coded using a rate-adaptive low-density parity-check (LDPC) code [4]. LDPC codes have been shown to achieve good compression results for correlated sources [5]. The rate-adaptive LDPC code enables small portions of the parity sequence S to be incrementally transmitted. Then, the LDPC decoder uses the received parity bits and the side information in EM to iteratively update θ, the soft estimate of X, and ψ, the soft estimate of disparity-compensated Y. A reconstruction of X is formed from θ by maximum a posteriori (MAP) estimation. If the decoder cannot losslessly reconstruct X, the decoder requests that the encoder sends more parity bits.

Assuming X and Y are approximately related by a horizontal shift, a correctly shifted version of Y can aid in lowering the transmission rate for X on the basis of the Slepian-Wolf theorem. The system in Fig. 1 includes a block-based disparity estimator to calculate soft estimates of the horizontal disparity D between pixels in X and pixels in Y. The relation between X and Y is modeled as

$$X(x,y) = Y(x - D(x,y), y) + N(x,y), \qquad (1)$$

where the noise N accounts for differences remaining after disparity compensation. The disparity estimator in Fig. 1 supplies the LDPC decoder with disparity-compensated side information ψ. If this side information correctly reflects the true state of X, then the side information will help the EM algorithm to converge towards the correct solution. The results in [2] show that disparity-compensated

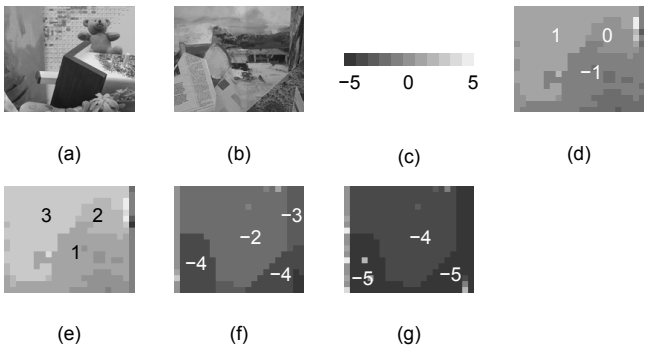

Fig. 2. Image X (8-bit, 144-by-176 pixels) for test sets (a) *Teddy* and (b) *Barn*. (c) Disparity legend. Block-wise disparity fields for *Teddy*, with side information (d) Y_1 and (e) Y_2. Block-wise disparity fields for *Barn*, with side information (f) Y_1 and (g) Y_2.

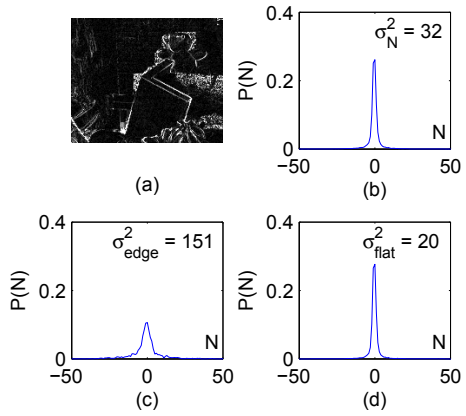

Fig. 3. (a) Residual N after optimal 8-by-8 block matching between X and Y_1 for *Teddy*. (b) $P(N)$ over the whole image. (c) $P(N)$ in edge regions only. (d) $P(N)$ in flat regions only.

side information can yield significant bit rate savings over using the uncompensated case of $D(x, y) = 0$.

The system makes a simplification in disparity estimation by calculating the distribution $P(D)$ only on a block-by-block basis. For example, the image X from two stereo test sets *Teddy* and *Barn* [6] are shown in Fig. 2, along with the optimal block-wise horizontal disparity fields between X and the corresponding side information images Y_1 and Y_2 in each test set. The block size is chosen to be 8. As can be observed from Fig. 2, the block-wise disparity field has unnatural step-like transitions along rectangular boundaries. In the improved decoder, the block-wise field is interpolated bilinearly. This procedure avoids the block boundaries and achieves smooth transitions.

Another issue with iterative disparity estimation is how to select good starting estimates for the first few iterations of EM. Incorrect disparity values can permanently bias the progression of EM towards a suboptimal local minimum, particularly at low bit rates. In Section 3, shape-based disparity initialization is introduced to provide better starting disparity estimates than those previously generated.

It has also been assumed that the additive noise N in (1) has stationary statistics and is zero-mean Laplacian distributed; that is,

$$P(N) = \frac{\lambda}{2} e^{-\lambda |N|} \tag{2}$$

where λ is a space-invariant constant by the stationarity assumption. To check this assumption, we examine Fig. 3 which shows the residual N following block matching between X and Y_1 for *Teddy*. The block size is again 8. Also shown are the statistical distributions $P(N)$ across the whole image, in edge regions only, and in flat regions only. As seen, $P(N)$ is well modeled by a zero-mean Laplacian. The assumption of stationarity, however, is not justified. Noise variance is highest along edges of objects, where disparity compensation on a block-by-block basis performs worst. In Section 3, a nonstationary edge-adaptive noise model is employed.

3. IMPROVED DISPARITY AND NOISE ESTIMATION

Improved disparity and noise estimation is achieved with the decoder in Fig. 4. Three new blocks are introduced. First, the block-wise disparity estimator is equipped with a bilinear interpolator, so that it now generates disparity-compensated side information ψ_1 according

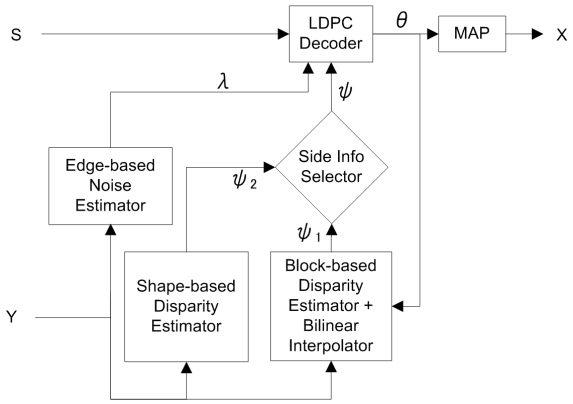

Fig. 4. Improved decoder for distributed stereo image coding.

to a smooth pixel-wise disparity field. Second, a separate disparity estimator uses shape information from Y to calculate an alternative disparity field, resulting in different side information ψ_2. Whether the LDPC decoder receives ψ_1 or ψ_2 is the decision of the side information selector. Third, a noise estimator uses edge information from Y to provide the LDPC decoder with a space-variant Laplacian parameter λ, which captures the nonstationary statistics of the noise.

3.1. Disparity Field Interpolation

A block-wise disparity field can contain unnatural step-like transitions, as explained in Section 2. Thus, the block-wise disparity field is bilinearly interpolated, which results in smooth transitions between blocks. Fig. 5 shows the pixel-wise disparity fields produced by using bilinear interpolation during the decoding process for *Teddy*, using Y_1 and Y_2 as side information. Each disparity field is recorded after the iteration in which the LDPC decoder converged to an errorless reconstruction. As desired, there are no longer any sharp rectangular boundary transitions but instead smooth transitions between the original disparity values at the block centers.

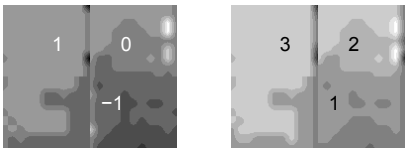

Fig. 5. Bilinear interpolations of the block-wise disparities (a) D_1 and (b) D_2 for *Teddy*, after EM convergence. Rate = 4.73 bpp.

3.2. Shape-Adaptive Disparity Initialization

Disparity estimates in the early iterations of EM can be unreliable, such as the noisy field shown in Fig. 6(a) used to relate X and Y_1 for *Teddy*. These false starting estimates can cause the EM algorithm to converge towards a suboptimal local minimum. Thus, a new shape-based disparity estimator is used in the early iterations of EM. This estimator extracts dominant shapes from the image and assigns a single disparity value for all pixels in a single shape. For example, a segmented version of Y_1 is shown in Fig. 6(b). The regions of the segmented image have been labeled so as to make this field look similar to the optimal block-based field in Fig. 2(d). This properly labeled segmented image is visibly a better early disparity estimate than the noisy field.

The heuristic currently used to choose the disparity values for each segmented region is to try different combinations of disparity values until one combination causes convergence or all combinations have been exhausted. The search begins by first assigning disparity values near zero and gradually advances to more positive and negative disparity values. This heuristic is only practical if the number of segmented regions is small, as is the case for the test images used. For n different regions and $D \in [-5, 5]$, there are at worst 11^n distinct combinations.

For the actual image segmentation, several popular algorithms have been considered: K-means clustering [7], region growing [8], and graph partitioning [9]. For our experiments, the method in [9] outperformed the other two methods in terms of correctly extracting dominant shapes.

3.3. Edge-Adaptive Noise Estimation

It is observed from Fig. 3 that the power of the noise is heavily concentrated around edges of objects and close to zero in low-frequency regions. Therefore, our generalization of the statistical model of N, from stationary to nonstationary, relies on the edge information. The new noise estimator extracts a binary edge image $Y_{edge}(x, y)$ from Y using a Canny edge detector [10], where $Y_{edge}(x, y) = 1$ indicates that an edge crosses the point (x, y). Compared to simpler edge detectors, the Canny edge detector is known to perform better at identifying actual object edges and rejecting spurious noise-induced edges. From the observation that errors after disparity compensation are higher along edges, the noise estimator assigns noise variance σ^2, or equivalently the Laplacian parameter $\lambda = \sqrt{2}/\sigma$, by setting

$$\lambda(x, y) = \begin{cases} \lambda_{edge}, & \text{for } Y_{edge}(x, y) = 1 \\ \lambda_{flat}, & \text{for } Y_{edge}(x, y) = 0 \end{cases} \quad (3)$$

where $\lambda_{edge} < \lambda_{flat}$ to reflect greater noise variance along edges. This classification has the benefit of being a simple criterion to apply to any image while at the same time yielding a fairly accurate prediction of spatial variations in noise power.

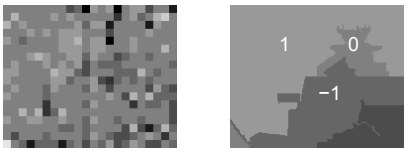

Fig. 6. (a) Noisy disparity field D_1 generated by the block-based disparity estimator after one iteration of EM for *Teddy*. Rate = 4.72 bpp. (b) Segmentation of Y_1 using graph partioning.

Table 1. Comparison of bit rates (bpp) for losslessly transmitting X. T = *Teddy* and B = *Barn*.

Side Information Image Y	T. Y_1	T. Y_2	B. Y_1	B. Y_2
Conditional Entropy	3.74	3.91	3.87	4.04
Decoder in [2]	4.45	4.55	4.45	4.64
Interpolation Only	4.33	4.33	4.24	4.30
Interp. + Noise Only	4.12	4.18	4.09	4.21
Interp. + Noise + Shape	4.09	4.12	4.09	4.21

4. EXPERIMENTAL RESULTS

Coding simulations are performed with the test images in Fig. 2. Each image is 144-by-176 pixels and has 8-bit depth. Due to memory constraints, the length of the LDPC code is limited to 50688, the same length as in [2]. For this reason, the test images used in [2] were limited to 72-by-88 pixels with 8-bit depth. To process the larger 144-by-176 image, we first subdivide X into four quadrants, each 72-by-88. Each quadrant of X can then be independently encoded and decoded using the 50688-long LDPC code, using as side information the corresponding quadrant of Y.

A simple rate control scheme is employed. For lossless coding, if after a maximum number of 100 iterations of EM the LDPC decoder still cannot reconstruct X without error, it requests additional parity bits from the LDPC encoder. Every 100 iterations of EM requires about 1 minute of simulation time. We also tested the decoding performance at rates where perfect reconstruction is not achieved, by measuring the quality of the resulting image after 100 iterations.

The initial distribution $P(D(x, y))$ for the first iteration is

$$P(D(x, y)) = \begin{cases} 0.750, & \text{for } D(x, y) = 0 \\ 0.025, & \text{for } D(x, y) \neq 0 \end{cases} \quad (4)$$

where $D \in [-5, 5]$. If shape-based disparity initialization is activated, then the side information selector uses the output of the shape-based disparity estimator for the first 10 iterations and thereafter uses the output of the block-based disparity estimator with bilinear interpolation.

4.1. Bit Rate Savings for Lossless Decoding

In Table 1, a comparison is made between the rates needed for lossless transmission of X using the discussed methods. Conditional entropy is calculated as the entropy of the difference between X and the optimal block-wise disparity-compensated version of Y, treating the flat and edge regions separately. For *Teddy*, with all three

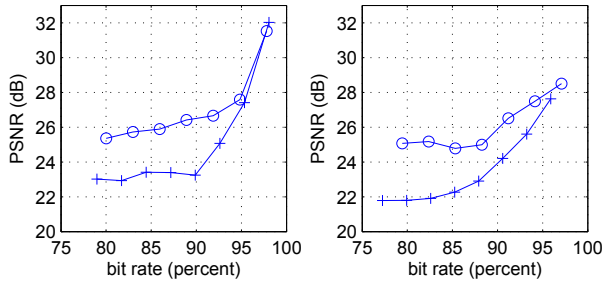

Fig. 7. Quality of reconstruction when decoding at lower rates for *Teddy*, comparing the decoder in [2] (+) and the improved coder (o), evaluated for side information sources (a) Y_1 and (b) Y_2. The rate is shown as a percentage of the minimum rate for lossless decoding.

Fig. 8. Upper right quadrants of reconstructed images for *Teddy* showing (a) blocking artifacts using the decoder in [2] and (b) showing removal of block artifacts using the improved decoder. The side information image is Y_1. Rate = 3.87 bpp.

proposed features are activated, the improved decoder achieves bit savings of 8 and 9 percent with Y_1 and Y_2 as side information, respectively, compared to the decoder in [2]. For *Barn*, the improved decoder again performs better, obtaining bit savings of 8 and 9 percent with Y_1 and Y_2 as side information, respectively.

4.2. PSNR Gains for Decoding with Errors

The proposed decoder also improves performance when the transmission rate of X is insufficient for perfect reconstruction. The quality of reconstruction for the test images is shown in Fig. 7. Increases in PSNR of several dB are obtained. Above 95 percent of the minimum rate for lossless decoding, an image with good visual quality can be reconstructed. Neither the codec in [2] nor the proposed system uses transform-based coding to exploit spatial correlation, so the rate-distortion tradeoff can be further improved. A related DCT-based codec demonstrates much better rate-distortion behavior [11].

4.3. Reduction of Blocking Artifacts

If the block-based disparity estimator chooses the wrong disparity value at rates below the minimum rate needed for lossless decoding, the reconstructed image $\hat{X}$ usually contains blocking artifacts, in which blocks from Y are incorrectly copied into regions in $\hat{X}$. The improved decoder helps to mitigate this problem in two ways. First, the bilinear interpolator smoothes the disparity field and increases disparity resolution, so that errors in $\hat{X}$ manifest as pixel-wise errors rather than block-wise errors. Second, shape-based disparity initialization helps to avoid local minima in EM that cause incorrect block matching between X and Y. Fig. 8 shows the upper right quadrants of partially decoded images for *Teddy*, comparing the decoder in [2] and the improved decoder. The improved decoder avoids generating the blocking artifacts on the bear's left arm, chest, and face created by the other decoder.

5. CONCLUSION

An improved decoder for distributed stereo image coding has been presented. It is based on an LDPC codec that learns block-wise disparity at the decoder. The proposed decoder refines the disparity field to pixel resolution using bilinear interpolation, initializes the disparity field using shape information, and estimates the nonstationary variance of the noise using edge information. For lossless coding, these three improvements result in up to 9 percent bit rate savings. At rates insufficient for lossless reconstruction, several dB gain in PSNR are obtained and visually disturbing blocking artifacts are significantly reduced.

6. REFERENCES

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

[2] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod, "Distributed grayscale stereo image coding with unsupervised learning of disparity," in *Proceedings of IEEE Data Compression Conference*, pp. 143-152, Snowbird, Utah, March 2007.

[3] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod, "Distributed coding of random dot stereograms with unsupervised learning of disparity," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing*, pp. 5-8, Victoria, BC, Canada, Oct. 2006.

[4] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," in *Proceedings of Asilomar Conference on Signals, Systems, Computers*, pp. 1203-1207, Pacific Grove, CA, Oct. 2005.

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

[6] D. Scharstein and R. Szeliski, "Middlebury stereo vision page," http://vision.middlebury.edu/stereo.

[7] Y. L. Chang and X. Li, "Adaptive image region-growing," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 868-872, Nov. 1994.

[8] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1452-1458, Nov. 2004.

[9] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167-181, Sept. 2004.

[10] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-714, Nov. 1986.

[11] D. Varodayan, Y.-C. Lin, A. Mavlankar, M. Flierl, and B. Girod, "Wyner-Ziv coding of stereo images with unsupervised learning of disparity," in *Proceedings of International Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007.