Sampling-based Correlation Estimation for Distributed Source Coding Under Rate and Complexity Constraints

Ngai-Man Cheung, Huisheng Wang, and Antonio Ortega

Abstract

In many practical distributed source coding (DSC) applications, correlation information has to be estimated at the encoder in order to determine the encoding rate. Coding efficiency depends strongly on the accuracy of this correlation estimation. While error in estimation is inevitable, the impact of estimation error on compression efficiency has not been sufficiently studied for the DSC problem. In this paper, we study correlation estimation subject to rate and complexity constraints, and its impact on coding efficiency in a DSC framework for practical distributed image and video applications. We focus on, in particular, applications where binary correlation models are exploited for Slepian-Wolf coding and sampling techniques are used to estimate the correlation, while extensions to other correlation models would also be briefly discussed. In the first part of this paper we investigate the compression of binary data. We first propose a model to characterize the relationship between the number of samples used in estimation and the coding rate penalty, in the case of encoding of a single binary source. The model is then extended to scenarios where multiple binary sources are compressed, and based on the model we propose an algorithm to determine the number of samples allocated to different sources so that the overall rate penalty can be minimized, subject to a constraint on the total number of samples. The second part of this paper studies compression of continuous-valued data. We propose a model-based estimation for the particular but important situations where binary bit-planes are extracted from a continuous-valued input source, and each bit-plane is compressed using DSC. The proposed model-based method first estimates the source and correlation noise models using continuous-valued samples, and then uses the models to derive the bit-plane statistics analytically. We also extend the model-based estimation to the cases when bit-planes are extracted based on the significance of the data, similar to those commonly used in wavelet-based applications. Experimental results, including some based on hyperspectral image compression, demonstrate the effectiveness of the proposed algorithms.

EDICS Category: COD-DSSC

This work was presented in part at the IEEE ICIP, Genoa, Italy, Sept. 2005, and at the IEEE ICIP, Atlanta, GA, USA, Oct. 2006 [1], [2]. The authors are with the Signal and Image Processing Institute, Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564. E-Mail: Ngai-Man Cheung (ncheung@usc.edu), Huisheng Wang (huisheng.wang@gmail.com), Antonio Ortega (antonio.ortega@sipi.usc.edu). Tel: (213) 740-4679. Fax: (213) 740-4651.

Sampling-based Correlation Estimation for Distributed Source Coding Under Rate and Complexity Constraints

I. Introduction

A. Motivation

Distributed source coding (DSC) [3], [4] studies independent encoding and joint decoding of correlated sources, for which a correlation model is known at the encoder. The Slepian-Wolf theorem states that two correlated sources can be optimally encoded (compressed at a rate approaching their joint entropy) even if the encoders only have access to the two sources separately, as long as both encoded streams are available at the decoder. Recently, DSC has attracted much attention because of its potential for a number of emerging applications including sensor networks and wireless video [5]–[7].

Central to DSC is the information about existing correlation between the source and the side-information (SI) available at the decoder. Specifically, correlation information refers to the joint p.d.f. between the source and the SI. This correlation information plays several important roles in practical distributed source coding applications. First, many applications require correlation information at the encoder to determine the encoding rate. Essentially, the encoders use the correlation information to determine the number of *cosets* for partitioning the input space, so that error-free decoding can be achieved [8], [9]. Second, for many practical Slepian-Wolf coding schemes that employ channel coding and iterative decoding, correlation information is required to initialize the decoding algorithms by providing likelihood estimates for the source bits [10]. Third, correlation information could be used at the decoder to determine the optimal reconstruction given the output of the Slepian-Wolf decoder and side information [8]. In this paper we focus on estimating the correlation for the purpose of determining the encoding rate. The results may, however, be useful for the other two cases as well.

B. Correlation Information Models in Practical DSC Applications

At the heart of practical DSC applications is a lossless Slepian-Wolf (SW) codec, which plays a role similar to that of entropy codecs in conventional image/video compression. More precisely, the Slepian-Wolf encoder compresses a discrete i.i.d. source, which could be losslessly recovered at the decoder with

the aid of the correlated side-information (SI) provided that enough compressed bits have been sent. Details on SW coding can be found in [6]–[8]. The problem to be investigated in this paper is that of determining the amount of information communicated to the decoder, i.e., the encoding rate. From [3], an ideal rate to achieve a vanishing probability of decoding error is the conditional entropy of the input source given the SI [1]. The conditional entropy, in turn, depends on the correlation information between the source and SI. Thus rate allocation in DSC can be performed by solving the associated correlation estimation problem. Various types of correlation models (e.g., binary valued p.m.f. or continuous valued p.d.f.) may be exploited to compress the data depending on the specific SW coding algorithms and applications. We illustrate some of them below.

- **SW coding with binary correlation model.** These are cases where some form of *binary* valued joint p.m.f. is used to relate the source and SI (which are not necessarily binary) for SW coding. As an example, Figure 1(a) depicts SW coding with a binary correlation structure similar to that used in [11], [12], etc. The continuous valued i.i.d. source X is mapped via scalar-quantization (or rounding) to a discrete source $\tilde{X}$. Then $\tilde{X}$ is mapped to a bit-plane representation, and each extracted bit-plane $b_{\tilde{X}}(l)$, $l = 0 \ \ldots \ L-1$, is compressed independently by the SW encoder. Here $b_{\tilde{X}}(0)$ denotes the least significant bit-plane (LSB). The SW decoder recovers $b_{\tilde{X}}(l)$ with bit-plane $b_{\tilde{Y}}(l)$ extracted from the quantized version of the correlated source Y as side information. By exploiting the joint binary p.m.f. $p(b_{\tilde{X}}(l), b_{\tilde{Y}}(l))$, each extracted bit-plane can be compressed to a rate as low as $H(b_{\tilde{X}}(l)|b_{\tilde{Y}}(l))$. To determine this encoding rate, one can estimate the joint binary p.m.f. $p(b_{\tilde{X}}(l), b_{\tilde{Y}}(l))$, and derive the conditional entropy from the estimated p.m.f. Note that independent compression of each bit-plane facilitates efficient rate scalability, which is highly desirable in some imagery applications. For some applications it is also sufficient to achieve satisfactory coding performance by exploiting the correlation between corresponding bit-planes.
- **SW coding with continuous correlation model.** These are cases where a continuous valued joint p.d.f. is exploited for SW coding. Figure 1(b) illustrates an example similar to that proposed in [9]. The continuous i.i.d. source X is first scalar-quantized, and the quantized input $\tilde{X}$ is directly compressed by the SW encoder. The decoder uses Y to recover $\tilde{X}$. By exploiting the joint p.d.f. $p(X, Y)$, the SW encoder can compress $\tilde{X}$ to a rate as low as $H(\tilde{X}|Y)$. The encoder may determine this encoding rate from $p(X, Y)$.

[1] Practical SW coders would add a small margin to this ideal rate to account for using finite-length input blocks in SW coding.

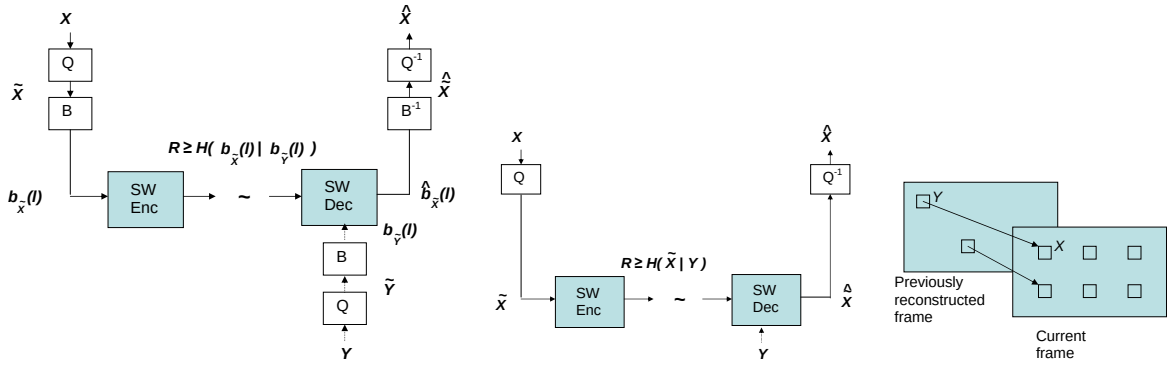


Fig. 1. (a) An example of Slepian-Wolf coding exploiting binary correlation. Boxes "Q" and "Q^{-1}" denote quantization and inverse quantization respectively. Boxes "B" and "B^{-1}" denote binarization and the inverse respectively. (b) An example of Slepian-Wolf coding exploiting continuous correlation. (c) Apply sampling to distributed video coding: when encoding the current frame, we randomly sample n macroblocks of the current frame to undergo motion estimation, with n being much smaller than the total number of macroblocks. By doing so, we would obtain n pairs of samples (X, Y), where X is the current macroblock and Y is the corresponding motion-compensated predictor. From these n sample pairs the encoding rate can be estimated. Note that here the sampling cost associated with each data sample is not primarily due to data exchange but the computation in motion search.

C. Correlation Estimation in Distributed Source Coding

While both the underlying theory and the recently proposed code construction algorithms [8], [10], [13] assume correlation information to be available *exactly* at the encoder, in many practical DSC applications, the correlation information may not be available beforehand, and one would need to estimate it during the encoding process [9] [2]. The accuracy of this correlation estimation has a direct impact on the performance of DSC-based systems. While under-estimating the correlation may result in a penalty in coding efficiency, over-estimation can cause decoding error: in this case candidate decoded values within a given coset would be too close to each other, so that it is no longer possible to guarantee that they can be disambiguated without error by using the SI, leading to degradation in reconstruction quality.

Estimating the correlation information at the encoder is a non-trivial problem due to the computational and communication constraints imposed by the target applications, i.e., often, correlation estimation in DSC has to be performed under *rate* and *complexity constraints*. For example, when DSC is applied to compress wireless sensor measurements, it is important to limit the amount of data exchanged between

[2]Note that in some cases, lack of an accurate correlation model is acceptable if there exists feedback from decoders to encoders [13], but this leads to an increase in overall delay.

nodes during correlation estimation, in order to minimize the communication cost. Similarly, in applications such as video coding, source data needed to estimate the correlation is present at the encoder, but it is desirable to limit the computation resources devoted to this estimation [9], [14].

D. Our Contributions and Related Work

In this work we study correlation estimation strategies subject to rate and complexity constraints, and their impact on coding efficiency in a DSC framework. Our proposed algorithms are based on the observation that for many DSC applications side information is actually *available* at the encoder, but the encoder may not make use of this side information because of the associated communication or computational cost. As an example, in low complexity distributed video coding (DVC) [9], [13], past frames that will be used as side information are available at the encoder, but the computation cost involved in performing motion estimation may be significant. Other examples can be found in distributed multiview image/video compression (e.g., [15]), wireless sensor data compression (e.g., [5], [6]), etc. Focusing on these applications, we propose *sampling-based* algorithms to estimate the correlation information. Sampling is a well-established concept in statistics to infer the properties of a population from a small amount of individual observations [16]. To see how sampling applies to DSC consider these two examples:

- When compressing distributed sensors measurements, X, a node can request samples, Y, from the neighboring node in order to estimate the correlation $p(X, Y)$. The number of samples exchanged should be small, however, to keep the communications overhead low.
- In some DVC applications, the encoding rate depends on the joint p.d.f. between blocks in current frame (X) and the corresponding motion-compensated predictor blocks (Y) from reference frame [9], [14]. The encoder can employ a sampling-based algorithm, where only *a small portion* of current frame's blocks would undergo motion estimation, so that the joint p.d.f. can be estimated from sample pairs (X, Y), formed with a given current block and the corresponding predictor (Figure 1(c)). Since motion search is required to acquire a sample pair (X, Y) each sampling operation would require some computational cost. Therefore, the proportion of blocks undergoing motion estimation should be small.

Sampling, however, leads to estimation errors and will have an impact on coding efficiency. Analyzing this impact is a key focus of our work. Since DSC applications may exploit various types of correlation models, it is difficult to address all of them. Therefore, we focus on one particular model in this paper and briefly discuss how to analyze other models. Specifically, focusing on situations where a binary

correlation is estimated for SW coding (as discussed in Section I-B) and correlation is estimated via sampling, this paper makes the following contributions:

- **Rate penalty analysis in compression of a single binary source.** We analyze how estimation error in sampling impacts the coding performance of a DSC system when encoding a single binary source. We derive an expression to quantify how the number of samples relates to the increase in the encoding rate due to estimation error, taking into account that over-estimation can lead to significant increases in distortion in DSC applications (due to decoding error).
- **Rate penalty analysis and sample allocation in compression of multiple binary sources.** We then extend the rate penalty analysis to systems with multiple binary input sources, where each of them is compressed independently using SW coding with its corresponding side-information. Based on the analysis, we propose an algorithm to determine sampling rates to assign to each binary source so that the overall penalty in coding performance (due to estimation error) can be minimized, subject to a constraint on the total number of samples.
- **Model-based estimation in compression of a continuous source.** We consider scenarios where bit-planes are extracted from a continuous input source and each bit-plane is compressed via SW coding, e.g., as in [11], [14]. We propose a model-based method where the continuous-valued joint p.d.f. of the source and SI is first estimated via sampling of continuous valued inputs, and then the bit-plane level (binary) correlation is derived from the estimated model. This is in contrast to a direct approach where the bit-plane correlation is estimated through exchanging binary samples from the extracted bit-planes. We demonstrate that the model-based method can achieve better estimation accuracy than the direct approach provided that the continuous-valued model is sufficiently accurate. In addition, we illustrate how the same estimation framework can be applied to various scenarios when previous decoded bit-planes and continuous SI are used in joint decoding.
- **Model-based estimation for structured bit-planes.** We also describe how model-based estimation can be extended to the cases where bit-planes are extracted from continuous input data using more sophisticated methods. For example, in wavelet-based applications, bit-planes are separated into different "sub-bit-planes" depending on the magnitude (significance) of the transform coefficients. A concrete example of this, which we consider in this paper, is that of bit-planes generated by *set-partitioning* as in SPIHT [17]. This type of bit-plane generation improves coding efficiency, but complicates the model-based correlation estimation process, as will be shown. Using a practical system as an example, we demonstrate that model-based estimation can lead to an additional

advantage of efficient implementation in these types of DSC applications.

While this paper mainly focuses on situations when each bit-plane is encoded by exploring its correlation with the corresponding bit-plane of the same significance, we also briefly discuss how some of the proposed ideas may be extended to situations when several more significant bit-planes are taken as SI.

Several methods have been proposed for correlation estimation problems in DSC. In DVC, low-complexity schemes to classify macroblocks into different correlation classes have been proposed [9], while other methods use a feedback channel to convey correlation information to the encoder [18]. For robust video transmission, recursive algorithms have been proposed to estimate the correlation between the noise-free and noise-corrupted reconstructions [19]. In our prior work, correlation estimation was performed by direct bit-plane comparisons between the source and an approximation of the decoder side-information [20]. This paper proposes, however, several novel sampling-based correlation estimation algorithms applicable to a range of DSC applications, and presents a performance analysis.

A general approach for model-based estimation for DSC was first proposed in our work in [1]. The work focused on the simple cases where bit-planes are generated directly from the the binary representation of the sources. A recent work [21] has proposed a similar idea to show the advantage of Gray code representations, but does not discuss the exact algorithm to estimate the correlation.

This paper is organized as follows. In Section II we present the rate penalty analysis. In Section III we propose the sample allocation algorithm to minimize the overall rate penalty. In Section IV we propose the model-based estimation, and in Section V we extend the model-based estimation to cases where bit-planes are extracted based on the significance of the data. Section VI presents experiments with real image compression applications. Finally, Section VII concludes the work.

II. Single Binary Source: Rate Penalty Analysis

In this section we analyze how estimation error in sampling affects the compression performance of a DSC system in the case of a single binary source. Specifically, given that n samples are used to estimate the correlation, we derive the corresponding increase in coding rate due to estimation error. We focus on the cases where a binary correlation is exploited in SW coding. Consider compressing a binary source b_X with another binary SI b_Y available at the decoder. We assume $\{b_X, b_Y\}$ i.i.d. $\sim p(b_X, b_Y)$. To simplify the analysis, we assume (i) b_X is equiprobable, i.e., $Pr[b_X = 0] = 0.5$, and (ii) the correlation is symmetric, i.e., $Pr[b_Y = 1|b_X = 0] = Pr[b_Y = 0|b_X = 1] = p$, where $0 \leq p \leq 0.5$ is the *crossover* probability for the sources. With these assumptions, the lower bound in the lossless encoding rate of b_X

with b_Y available at the decoder is [3]

$$H(b_X|b_Y) = H(p). \tag{1}$$

Therefore, encoder can estimate the lossless compression rate of b_X by estimating p through n random samples pairs $\{b_X, b_Y\}$. Define the estimation error $(\triangle p)(n) = \hat{p}(n) - p$, where $\hat{p}(n)$ is some estimation of p [3]. When $\triangle p$ is negative, this could lead to a decoding error. This is because the crossover probability is under-estimated and so the number of cosets chosen for encoding may be too small. Instead, when $\triangle p \geq 0$, correct decoding and lossless recovery of b_X can be guaranteed theoretically, but there will be a penalty in compression efficiency. This difference in behavior (decoding error vs. coding penalty) will lead us to propose a biased estimator such that $\triangle p \geq 0$ with high probability (discussed in the next section). On average, the coding penalty, in bits/sample, is given by (assuming no decoding error):

$$(\triangle H)(n) \quad = \quad H(\hat{p}(n)) - H(p). \tag{2}$$

As will be discussed, $\triangle H$ is indeed a random variable, since we randomly choose samples of b_X and b_Y for estimating $H(\hat{p})$. Our focus is to derive the probability density of $(\triangle H)(n)$.

A. Correlation Estimation

For encoding b_X we need to estimate $\hat{p}$ by acquiring n random samples of b_Y. In different DSC applications, encoder may obtain the samples in different ways. For example, in a sensor application, the encoder of b_X may request samples of b_Y from a spatially-separated sensor node. In distributed video coding, the encoder may perform motion estimation to generate samples of b_Y. Common to most of the applications is that communication/computational costs will be incurred in acquiring the samples. Therefore, it is desirable to keep n small.

By inspecting the n pairs $\{b_X, b_Y\}$ now available at the encoder, an estimate of p can be computed to determine the encoding rate for b_X. A naive estimate for p is $\frac{S(n)}{n}$, where $S(n)$ is the number of inspected samples such that $b_X \neq b_Y$. Since S is essentially the summation of n i.i.d. Bernoulli random variables with success probability p, S is a binomially distributed r.v. with mean np and variance $np(1-p)$. Therefore, for sufficiently large n,

$$\frac{S(n)}{n} \sim N(p, \sigma^2) \quad , \quad \sigma = \sqrt{p(1-p)/n}. \tag{3}$$

[3]Note the dependency of $\hat{p}$ (and other quantities) on n. Precisely, the p.d.f. of $\hat{p}$ is a function of n, as will be discussed.

As a consequence, if we use $\frac{S}{n}$ as the estimator there is 50% probability of under-estimation of p. Therefore, we opt to use a *biased estimator* given by

$$\hat{p}(n) \quad = \quad \frac{S(n)}{n} + z_{\omega/2}\sigma. \tag{4}$$

That is, we bias $\frac{S}{n}$ by a factor proportional to σ. By choosing the constant $z_{\omega/2}$ we can control precisely the probability of under-estimation of p, e.g., if $z_{\omega/2} = 1.64$, $Pr[\hat{p} < p] = \omega/2 = 0.05$. We choose this biased estimator to minimize the risk of decoding failure, at the expense of some encoding rate penalty.

B. Rate Penalty Analysis

Using (4) as the estimator, we analyze how n relates to the p.d.f. of $\triangle H$. From (4),

$$\begin{aligned}(\triangle p)(n) &= \hat{p}(n) - p \\ &= \frac{S(n)}{n} + z_{\omega/2}\sigma - p \sim N(z_{\omega/2}\sigma, \sigma^2).\end{aligned} \tag{5}$$

Expanding $H(.)$ at p by a Taylor series and using the definition of $\triangle H$ in (2),

$$\begin{aligned}(\triangle H)(n) &= H'(p)\triangle p + \frac{H''(p)(\triangle p)^2}{2!} + ... \\ &\approx H'(p)(\triangle p)(n).\end{aligned} \tag{6}$$

The approximation holds when $\triangle p$ is sufficiently small. The p.d.f. of $\triangle H$ can then be derived as:

$$(\triangle H)(n) \quad \sim \quad N(H'(p)z_{\omega/2}\sigma, (H'(p))^2\sigma^2), \tag{7}$$

where $H'(p) = ln(\frac{1}{p} - 1)$ and σ is given by (3). (7) relates n to the density of $\triangle H$. Using (7), one can readily compute some statistics for $\triangle H$, e.g., $E[(\triangle H)(n)]$. Note that these statistics are functions of n.

In practice, since p is unknown, σ is unknown when computing the estimator (4). A good rule of thumb is to approximate the estimator using [16]

$$\hat{p}(n) \quad \approx \quad \frac{S}{n} + z_{\omega/2}\sqrt{\frac{S}{n}(1 - \frac{S}{n})/n}, \tag{8}$$

i.e., $\frac{S}{n}$ is used to approximate p in computing the estimator. The approximation would be valid when $n \cdot \frac{S}{n} \geq 4$ and $n(1 - \frac{S}{n}) \geq 4$, as a rule of thumb [16].

The analysis can also be extended to other correlation models following the same methodology. An example could be when the inputs are non-equiprobable. In this case, assuming $Pr[b_X = 0] = \theta$, it can be shown that the lower bound in the lossless encoding rate of b_X is a function of p and θ

$$H(b_X|b_Y) = g(p, \theta) = \pi H(\frac{(1-p)\theta}{\pi}) + (1 - \pi)H(\frac{p\theta}{1-\pi}), \tag{9}$$

where $\pi = Pr[b_Y = 0] = \theta + p - 2\theta p$. Note that only a subset of b_Y's are available when encoding b_X and thus p cannot be found exactly. On the other hand, since all b_X's are available at the encoder there is no estimation error for θ. Therefore, the rate penalty can be approximated by:

$$(\triangle H)(n) \approx \frac{\partial g(p,\theta)}{\partial p}(\triangle p)(n) \quad \sim \quad N(\frac{\partial g(p,\theta)}{\partial p} z_{\omega/2}\sigma, (\frac{\partial g(p,\theta)}{\partial p})^2\sigma^2), \tag{10}$$

where

$$\begin{aligned} \frac{\partial g(p,\theta)}{\partial p} &= (1-2\theta)(H(\frac{(1-p)\theta}{\pi}) - H(\frac{p\theta}{1-\pi})) + \\ &\quad \frac{\theta(\theta-1)}{\pi}\ln(\frac{p(1-\theta)}{(1-p)\theta}) + \frac{\theta(1-\theta)}{1-\pi}\ln(\frac{(1-p)(1-\theta)}{p\theta}). \end{aligned} \tag{11}$$

Note that (10) can then be used to derive the penalty model for multiple non-equiprobable binary sources and to determine the corresponding optimum bit allocation to be discussed in Section III.

C. Experiment

In this section we assess the accuracy of the rate penalty model proposed in (7). Specifically, we perform sampling experiments and measure $\triangle H$, and compare the empirical distribution of $\triangle H$ with (7). Video data is used for the experiments. As the binary input source (b_X), we use the raw bit-planes extracted from all (quantized) DCT coefficients of a given frequency in the current frame, while as SI (b_Y) we use the raw bit-planes of same significance extracted from the corresponding (quantized) DCT coefficients of the motion-compensated predictors in the reference frame[4]. Therefore, the dimension of the source, M, is equal to the number of DCT blocks in a frame. We then sample n ($< M$) pairs of $\{b_X, b_Y\}$ randomly, and an estimation $\hat{p}$ is then computed according to (8) from the chosen pairs. With $\hat{p}$, a single $\triangle H$ can then be obtained using (2). The sampling experiment is repeated N_E times, each time with different pairs of $\{b_X, b_Y\}$ sampled and a different $\triangle H$ obtained. We compare the empirical p.d.f. of $\triangle H$ (with N_E data) with the model in (7) using Kolmogorov-Smirnov (K-S) tests [16]. Table I shows the resulting K-S statistics at different sampling rates for some bit-planes extracted from the DC coefficients quantized at $QP = 24$ (H.263 quantization) in the 2nd frame of *Mobile* (720×576, 30 fps), with $z_{\omega/2} = 1.64$ and $N_E = 100$. In particular, for the range of n and p where $n \cdot p \geq 4$, K-S tests approve the hypothesis that the empirical $\triangle H$ follows the model in (7). This results indicate that our proposed model can adequately characterize the distribution of the rate penalty for practical problems

[4]Note that it is common for DVC systems to exploit correlation between the DCT coefficients in the current frame and the corresponding coefficients in the motion-compensated predictors in the reference frame, e.g., [14].

TABLE I

KOLMOGOROV-SMIRNOV (K-S) TESTS STATISTICS FOR $\triangle H$: MOBILE DC, QP=24. NUMBERS IN PARENTHESIS INDICATE CASES THAT K-S STATISTICS ARE LARGER THAN THE 1% SIGNIFICANCE CUTOFF VALUE, 0.1608, AND THEREFORE DO NOT PASS K-S TESTS. THOSE ARE CASES WHEN $n \cdot p \geq 4$ DOES NOT HOLD. NOTE THAT BIT POSITION 6 CORRESPONDS TO A MORE SIGNIFICANT BIT-PLANE.

Bit Position	6 $p = 0.0228$ $H(p) = 0.1571$	5 $p = 0.0529$ $H(p) = 0.2987$	4 $p = 0.0961$ $H(p) = 0.4566$	3 $p = 0.2002$ $H(p) = 0.7222$
$n = 96$	(0.1912)	0.1470	0.1310	0.1510
$n = 128$	0.1580	0.1401	0.1266	0.1358
$n = 256$	0.1458	0.1343	0.1073	0.1080
$n = 512$	0.1459	0.1219	0.0943	0.0957

with sufficiently large $n \cdot p$, e.g., $n \cdot p \geq 4$. Note that a sampling size of 128 corresponds to about 2% of data for a 720×576 video. Additional results using different data lead to similar conclusions.

III. MULTIPLE BINARY SOURCES: RATE PENALTY ANALYSIS AND SAMPLE ALLOCATION

In this section we study compression of multiple binary input sources. The multiple sources scenario can arise in different applications. One example is the compression of multiple streams of sensor measurements captured in different nodes. Another important example is the compression of a continuous input source, where the source is first mapped to a bit-plane representation and then each bit-plane is compressed using DSC, so that the problem becomes one of compression of multiple binary sources.

Consider the compression of L binary sources $b_X(l)$, $l = 0$ to $L-1$. Each binary source is independently encoded using SW coding with its respective SI $b_Y(l)$ available at the decoder. We shall follow the assumptions in Section II, i.e., $\{b_X(l), b_Y(l)\}$ i.i.d. $\sim p(b_X(l), b_Y(l))$, with $Pr[b_X(l) = 0] = 0.5$ and crossover probability $Pr[b_X(l) \neq b_Y(l)] = p_l$. Let K_l be the number of binary values to be encoded for source $b_X(l)$. We follow the correlation estimation procedure in Section II, where encoding $b_X(l)$ requires observing n_l ($\leq K_l$) random samples of $b_Y(l)$ in order to compute the biased estimator $\hat{p_l}(n_l)$ according to (4) (or (8) in practice), so that the encoding rate for $b_X(l)$ can be determined. The encoding of $b_X(l)$ would suffer a rate penalty $(\triangle H_l)(n_l) = H(\hat{p_l}(n_l)) - H(p_l)$. In particular, following the discussion in Section II and using (7), $\triangle H_l$ would be normally distributed:

$$(\triangle H_l)(n_l) \quad \sim \quad N(H'(p_l) z_{\omega/2} \sigma_l, (H'(p_l))^2 {\sigma_l}^2), \tag{12}$$

where $H'(p_l) = ln(\frac{1}{p_l} - 1)$ and $\sigma_l = \sqrt{p_l(1-p_l)/n_l}$. On average, the coding penalty of the whole system, in bits/sample, is given by:

$$\triangle H = \frac{1}{K_T}\sum_{l=0}^{L-1} K_l \triangle H_l. \tag{13}$$

where $K_T = \sum_{l=0}^{L-1} K_l$. Note that in this section $\triangle H$ refers to the average coding penalty of the entire system with L sources. Since the samplings are performed independently on each source, $\triangle H_l$ are independent r.v., and therefore $\triangle H$ is normally distributed with expectation and variance given by:

$$E[\triangle H] = \frac{1}{K_T}\sum_{l=0}^{L-1} K_l H'(p_l) z_{\omega/2}\sigma_l, \tag{14}$$

$$VAR[\triangle H] = \sum_{l=0}^{L-1}\left(\frac{K_l}{K_T}\right)^2 (H'(p_l))^2 {\sigma_l}^2. \tag{15}$$

The total number of samples is limited to be n_T, i.e., $\sum_{l=0}^{L-1} n_l = n_T$, under the assumption that we would like to have $n_T \ll K_T$, because each sample would incur a communication/computational cost.

Our main goal is to minimize $E[\triangle H]$ subject to a given n_T. Note that $E[\triangle H]$ is a function of (i) $\mathbf{p} = \{p_l\}$, correlation of different sources, (ii) n_T, total number of samples used to estimate correlation, and (iii) $\mathbf{n} = \{n_l\}$, allocation of samples to different sources. In the following sections:

1) we derive an optimal sample allocation strategy, i.e., given $\mathbf{p} = \{p_l\}$, n_T, we find the optimal $\mathbf{n} = \{n_l\}$ to minimize the rate penalty $E[\triangle H]$;
2) given the optimal sample allocation, we study how $E[\triangle H]$ changes with n_T.

As will be discussed in the next section, the optimal sample allocation requires knowledge of $\{p_l\}$, which obviously is not known a priori. Several strategies will therefore be described to apply our results in practice.

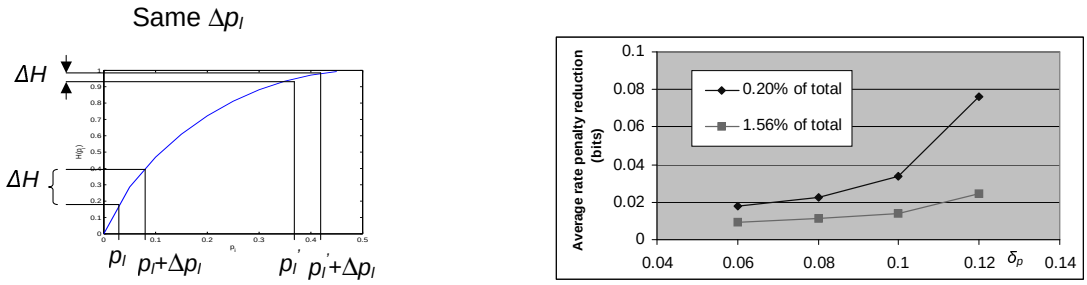
A. Optimal Sample Allocation Strategy

In this section we seek to find the optimal number of samples to allocate to different sources, $\{n_l^*\}$, so as to minimize $E[\triangle H]$. To find $\{n_l^*\}$, we solve the following constrained optimization problem:

$$\min_{\{n_l:\sum_{l=0}^{L-1} n_l = n_T; n_l \le K_l\}} E[\triangle H], \tag{16}$$

where $E[\triangle H]$ is given by (14). Applying the Lagrangian optimization method and Kuhn-Tucker conditions to deal with the inequalities constraints, we obtain (details in Appendix I):

$$n_l^* = \begin{cases} \gamma(K_l\alpha_l)^{2/3}, & \text{if } \gamma < \frac{{K_l}^{1/3}}{{\alpha_l}^{2/3}}, \\ K_l, & \text{if } \gamma \ge \frac{{K_l}^{1/3}}{{\alpha_l}^{2/3}}, \end{cases} \tag{17}$$



(a) (b)

Fig. 2. (a) Encoding rate function $H(.)$. The optimal allocation takes into account that the same $\triangle p_l$ would have a larger impact to the estimated rate if the true p_l is small. (b) Reduction in rate penalty (bits) using the optimal sample allocation, with $L = 4$, $K_l = 16384$, $n_T = 128$ or 1024 (i.e., 0.20% or 1.56% of total respectively), $N_E = 100$, and $\overline{p} = 0.25$.

where α_l is a constant that depends on $z_{\omega/2}$ and p_l:

$$\alpha_l = ln(\frac{1}{p_l} - 1) z_{\omega/2} \sqrt{p_l(1-p_l)}, \tag{18}$$

and γ is chosen so that $\sum_{l=0}^{L-1} n_l^* = n_T$. This result gives rise to a sample allocation scheme analogous to the "water-filling" results in information theory [22]. We allocate equal *weighted* number of samples to each source, until for some sources the number of samples is equal to the number of source inputs. The weighting factor $(K_l \alpha_l)^{2/3}$ is a constant that depends only on the specific characteristics of the lth source (length and crossover probability).

The intuition behind the optimal sample allocation can be understood by inspecting the rate function $H(.)$. For the same estimation error $(\triangle p_l)(n_l) = \hat{p_l}(n_l) - p_l$, the impact to the encoding rate will tend to be larger when the true p_l is small (Figure 2(a)). Since different sources have different p_l, we should allocate the samples accordingly and use more samples for those sources with small p_l, so as to minimize the overall rate penalty. This is reflected in (17) and (18).

Since we have chosen to use the same $z_{\omega/2}$ in all sources, we have the same probabilities of over-estimation for all sources. In some applications it may be more appropriate to choose different target failure probabilities for different sources, e.g., in the cases when the sources are bit-planes extracted from a continuous valued source, MSB bit-planes are more important and should have smaller failure probabilities. This can be incorporated in the sample allocation by using different $z_{\omega/2}$ for different sources in (14).

Note that the optimal sample allocation depends on the crossover probabilities $\{p_l\}$. However, $\{p_l\}$ is

obviously not known initially. In practical applications, our results can be applied as follows:

- In many applications, we may have some a priori knowledge of $\{p_l\}$. For example, in hyperspectral image compression [11], where bit-planes of spectral bands are extracted and compressed, $\{p_l\}$ of neighboring spectral bands are usually similar. Therefore, we can use some approximations of $\{p_l\}$ in the optimal sample allocation equations to determine the sample assignment. We will present experimental results in Section VI to demonstrate that this can be a viable method. Note that using the a priori knowledge directly to select the encoding rate may cause decoding error, if this a priori knowledge leads to over-estimating the correlation. Instead, by using the a priori knowledge to determine the sample allocation and (8) as the estimator to select the coding rate, we are guaranteed that $\hat{p_l}$ is larger than p_l with probability $(1-\omega/2)$, and we can bound decoding error systematically. Also this is more robust in cases where the a priori knowledge may not be a good approximation to the true $\{p_l\}$.
- We can also use an iterative approach similar to [23]. Essentially, we would allocate samples in batches of same size. For the first batch, we allocate the same number of samples to all bit-planes and obtain some initial estimates for $\{p_l\}$. For the subsequent batches, we allocate the number of samples according to the current estimates and the optimal sample allocation strategy. In this approach, we can also use any available a priori knowledge to initialize the scheme.

B. Rate Penalty Analysis - Multiple Sources

Having an expression for $E[\triangle H]$ as a function of n_T allows the encoder to select appropriate values for n_T, given that increasing n_T leads to additional overhead but also reduces the rate increase due to inaccurate estimation. We focus on the cases when $n_l < K_l$, where closed form equations can be derived. The equations relating $E[\triangle H]$ to n_T can be obtained from (14) and (17):

$$E[\triangle H] = \frac{\beta}{\sqrt{n_T}}, \tag{19}$$

where $\beta = \frac{1}{K_T}\left[\sum_{l=0}^{L-1}(K_l\alpha_l)^{2/3}\right]^{3/2}$. Note that β is a constant for a given system. Therefore, the average rate penalty is inversely proportional to the squared root of the amount of sampling overhead.

C. Experiments

In this section we assess the benefits of using the optimal sample allocation when compressing i.i.d. sources. We randomly generate L pairs of binary correlated sources $\{b_X(l), b_Y(l)\}$ each with

crossover probability p_l and dimension K_l. Then n_T samples are used to estimate the correlation. The number of samples allocated to each source is determined according to the following schemes:

- *Optimal allocation.* We use (17) to determine n_l, the number of samples allocated to the l-th source.
- *Even allocation.* We allocate the same number of samples to each source, i.e., $n_l = n_T/L$.

Note that in this section we use the true crossover probabilities to determine the optimal allocation, while a similar comparison using some a priori information to determine the optimal allocation in practical scenarios will be discussed in Section VI. The schemes are compared based on the average (overall) rate penalty after N_E sampling experiments. We compute the reduction in the average rate penalty (in bits) as: $E[\triangle H]_{even} - E[\triangle H]_{opt}$, where $E[\triangle H]_{opt}$ is the average rate penalty using the optimal sample allocation, and $E[\triangle H]_{even}$ is that of using the even sample allocation. Since there are many possible combinations of $\{p_l\}$, as an example we choose $\{p_l\}$ of the form $\{\overline{p} + k\delta_p\}$, $k = \pm 1, \pm 2, ..., \overline{p} = 0.25$. Therefore, a large δ_p corresponds to more substantial variation (standard deviation) in $\{p_l\}$. Figure 2(b) depicts the comparison results and shows that over 0.07 bits reduction in rate penalty can be achieved in the case of considerable variation in $\{p_l\}$, with this extra number of bits representing a 11.7% rate increase. In practical applications significant variations in crossover probabilities are indeed common (e.g., see [11] or data in Section II-C).

IV. Continuous Input Source: Model-based Estimation

In this section we investigate correlation estimation methods in the particular but important cases when bit-planes are extracted from a continuous input source and each bit-plane is compressed via SW coding. This situation arises in several proposed distributed image and video coding algorithms (e.g., [11], [14]). Often, in these applications, some a priori statistical model knowledge of the continuous-valued input source is available. For example, wavelet and DCT transform coefficients are typically considered to be well modeled by Laplacian distributions [24]. In what follows we propose a model-based estimation method, where the continuous-valued joint p.d.f. of the source and SI is first estimated via sampling of continuous valued inputs[5], and then the bit-plane level (binary) correlation is derived from the estimated model. This is in contrast to the direct approach studied in Sections II and III, where the bit-plane correlation is estimated through exchanging binary samples from the extracted bit-planes. We shall demonstrate that the model-based method can achieve better estimation accuracy than the direct approach provided that the continuous-valued model is sufficiently accurate.

[5]In practice, the continuous valued inputs are rounded so that the samples can be represented with a finite number of bits.

Section IV-A will outline the basic ideas of model-based estimation and its application to binary correlation model. Section IV-B will then discuss how the same framework can be applied to situations when previous decoded bit-planes and continuous SI are used in joint decoding.

A. General Approach

We first focus on the setting of Figure 1(a), where binary correlation is exploited for compression of a continuous input X using continuous side-information Y. Assume X and Y are drawn i.i.d. from $f_{XY}(x,y)$. We assume $Y = X + Z$, where Z is the correlation noise independent of X. Our proposed model-based approach starts by estimating the p.d.f.'s $f_X(x)$ and $f_Z(z)$. This can be done by choosing appropriate models for the data samples and then estimating the model parameters using one of the standard parameter estimation techniques, e.g., maximum likelihood estimation (MLE).

Once we have estimated $f_X(x)$ and $f_Z(z)$ we can derive the bit-plane statistics as follows. Suppose we extract *raw* bit-planes from the binary representations of X and Y, and are interested in estimating p_l, the crossover probability between the bit-planes of X and Y at significance level l. Figure 3(a) depicts the events (shaded regions A_i) that lead to the occurrence of crossover between X and Y at significance level l. For example, consider $l = 2$ (i.e., the 2nd bit-plane), when X takes the values from 8 $(= 01\underline{0}00b)$ to 11 $(= 01\underline{0}11b)$, crossover occurs when Y takes the values from 4 $(= 00\underline{1}00b)$ to 7 $(= 00\underline{1}11b)$ (region A_4 in Figure 3(a)), or 12 $(= 01\underline{1}00b)$ to 15 $(= 01\underline{1}11b)$ (region A_5 in Figure 3(a)), ..., etc. Specifically, A_i is a subset of $\mathbb{R}^2$ defined as

$$\begin{aligned} A_i = \{\ (x,y) \quad | \quad & 2c \cdot 2^l \le |x| < (2c+1)\cdot 2^l,\ (2d+1)\cdot 2^l \le |y| < (2d+2)\cdot 2^l;\ \textit{or} \\ & (2g+1)\cdot 2^l \le |x| < (2g+2)\cdot 2^l,\ 2h\cdot 2^l \le |y| < (2h+1)\cdot 2^l\}, \end{aligned} \tag{20}$$

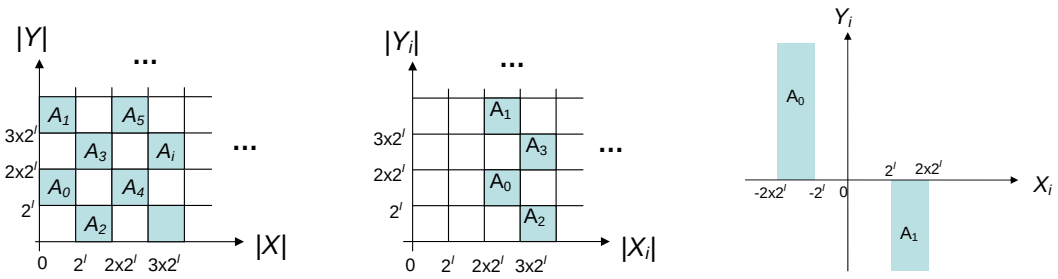
for some $c, d, g, h \in \mathbb{Z}^+$. Hence we can estimate the crossover probability at bit-plane l by

$$\begin{aligned} \hat{p_l} &= \sum_i \int\int_{A_i} f_{XY}(x,y)dxdy \\ &= \sum_i \int\int_{A_i} f_X(x) f_{Y|X}(y|x)dxdy \end{aligned} \tag{21}$$

Assuming that $Y = X + Z$ and that X, Z are independent, $f_{Y|X}(y|x)$ can be found to be equal to

$$f_{Y|X}(y|x) = f_Z(y - x) \tag{22}$$

and the integral in (21) can be readily evaluated for a variety of densities. In practice we only need to sum over a few regions, A_i, where the integrals are non-zero. Note that when l is small (i.e., least significant bit-planes) the crossover probability is close to 0.5, since in such cases A_i are small and evenly distributed throughout the sample space, and hence for most models (21) will give $\hat{p_l}$ close to 0.5.



(a) (b) (c)

Fig. 3. (a) Crossover probability estimation for raw bit-planes. A_i are the events that lead to occurrence of crossover between X and Y at significance level l. (b) Refinement bit-plane crossover probability estimation: probability of crossover and X_i is already significant. (c) Sign bit-plane crossover probability estimation: probability of sign crossover and X_i becomes significant.

B. Extensions to Other Correlation Models

1) Model-based estimation with previously decoded bit-planes as side-information: The model-based method can also be extended to estimate the correlation in the cases when the previously decoded bit-planes are used as SI. Specifically, we consider the cases when bit-planes are extracted from a continuous input source, and each bit-plane $b_X(l)$ is compressed with both the previously decoded bit-planes of the same source $b_X(l+1), ..., b_X(l+m)$ and that of the correlated source $b_Y(l), b_Y(l+1), ..., b_Y(l+m)$ as SI (Figure 4(a)) [6]. We assume bit-planes are communicated to the decoder starting from the MSB, while the reverse order can be addressed similarly. For encoding $b_X(l)$ we need to estimate the coding rate

$$H(b_X(l)|b_X(l+1), ..., b_X(l+m), b_Y(l), b_Y(l+1), ..., b_Y(l+m)). \tag{23}$$

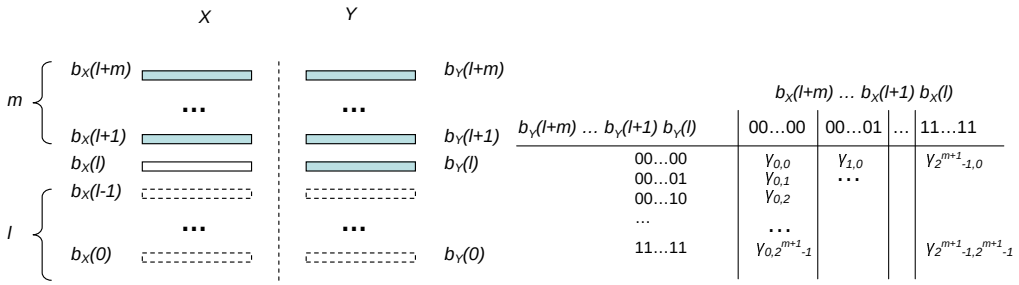
To determine (23), we need the joint p.d.f. between the input and all the SI:

$$p(b_X(l), b_X(l+1), ..., b_X(l+m), b_Y(l), b_Y(l+1), ..., b_Y(l+m)), \tag{24}$$

which has $2^{(2m+2)} - 1$ free parameters. While estimating the joint p.d.f. may appear complex, the model-based estimation for (24) in fact exhibits regular structure, which greatly simplifies the estimation process. In addition, in some applications further improvement from using $m > 1$ may be negligible (Figure 5(a))[7], and estimation with $m = 1$ requires only modest complexity.

[6]In Section IV-B.2 we will consider the case where the continuous SI Y is used in joint decoding.

[7]Note that the improvement of using previous bit-planes would depend on the statistics of the data, and hence is application-dependent [25].



(a) (b)

Fig. 4. (a) Encoding of bit-plane $b_X(l)$ with both the previously decoded bit-planes $b_X(l+1), ..., b_X(l+m)$ and that of the correlated source $b_Y(l), b_Y(l+1), ..., b_Y(l+m)$ as SI. (b) Joint p.d.f. between the input and all the SI.

We denote $\gamma_{i,j}$ the joint probability when the binary representation of i is $b_X(l+m)...b_X(l+1)b_X(l)$ and that of j is $b_Y(l+m)...b_Y(l+1)b_Y(l)$ (Figure 4(b)), i.e.,

$$\gamma_{i,j} = p(\langle b_X(l+m)...b_X(l+1)b_X(l)\rangle = i, \langle b_Y(l+m)...b_Y(l+1)b_Y(l)\rangle = j),$$

and $\langle b(l+m)...b(l+1)b(l)\rangle$ denotes the numerical value of the concatenation of the sequence of the bits $b(l+m), ..., b(l+1), b(l)$, i.e., $\sum_{i=0}^{m} b(l+i) \times 2^i$. It can be shown that by tracing the binary representations of X and Y the events leading to the occurrence of $\langle b_X(l+m)...b_X(l+1)b_X(l)\rangle = i$ and $\langle b_Y(l+m)...b_Y(l+1)b_Y(l)\rangle = j$ correspond to the region $A_{i,j}$ in the sample space of X and Y in Figure 5(b). Therefore,

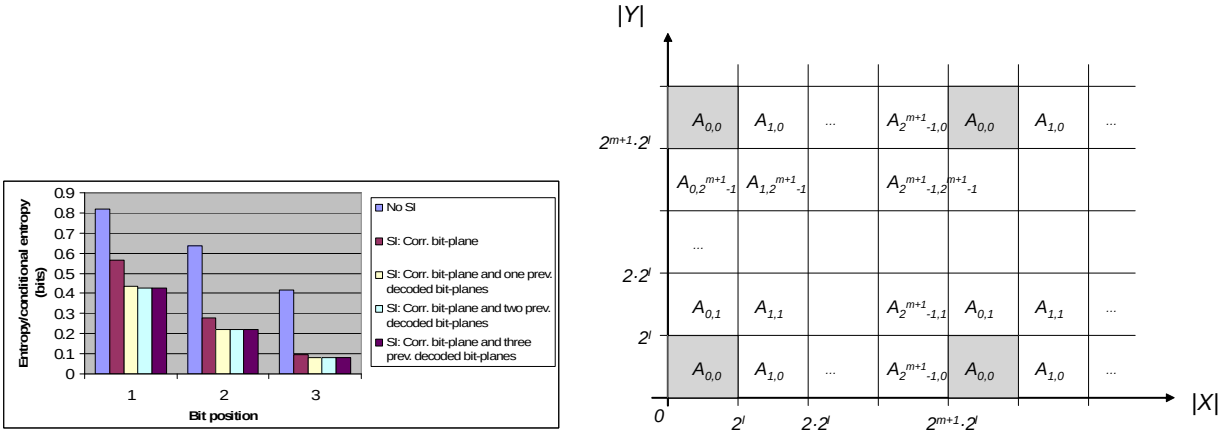
$$\gamma_{i,j} \quad = \quad \int\int_{A_{i,j}} f_{XY}(x,y)dxdy \tag{25}$$

where

$$\begin{aligned} A_{i,j} = \{\ (x,y) \quad | \quad & c \cdot 2^{l+m+1} + i \cdot 2^l \le |x| \le c \cdot 2^{l+m+1} + i \cdot 2^l + 2^l - 1, \\ & d \cdot 2^{l+m+1} + j \cdot 2^l \le |y| \le d \cdot 2^{l+m+1} + j \cdot 2^l + 2^l - 1,\ c, d \in \mathbb{Z}^+\}. \end{aligned} \tag{26}$$

(25) can be readily computed by factorizing f_{XY} and estimating f_X and f_Z as discussed in Section IV-A. In practice, we only need to sum over a few regions where the integrals of f_{XY} are practically non-zero. Note that we can extend this to estimate the encoding rate for structured bit-planes (i.e., sign/refinement bit-planes) as will be discussed in Section V.

2) Model-based estimation with continuous side-information in joint decoding: In this section we consider situations where each bit-plane $b_X(l)$ is compressed with both the previously decoded bit-planes of the same source $b_X(l+1), ..., b_X(l+m)$ and the continuous SI Y available to be used in joint



(a) (b)

Fig. 5. (a) Entropy/conditional entropy of the video bit-planes data used in the experiment in Section II-C, i.e., X and Y are the quantized DCT coefficients in a current frame and the corresponding quantized coefficients in the motion-compensated predictors in the reference frame, respectively, using the 2nd AC coefficients of *Mobile* (720×576, 30 fps), at QP$=$ 12: (i) Without SI, i.e., intra coding ("No SI"); (ii) Using only the corresponding bit-plane as SI, i.e., $m = 0$ ("SI: Corr. bit-plane"); (iii), (iv), (v): Using corresponding and one, two or three previously decoded bit-planes as SI, i.e., $m = 1, 2$ or 3 respectively. (b) Model based estimation: The events leading to the occurrence of $\langle b_X(l+m)...b_X(l+1)b_X(l)\rangle = i$ and $\langle b_Y(l+m)...b_Y(l+1)b_Y(l)\rangle = j$ correspond to the region $A_{i,j}$ in the sample space of X and Y.

decoding[8]. As will be shown, this could be addressed with minor extensions of the techniques we have discussed. For encoding $b_X(l)$ we need to estimate the coding rate

$$H(b_X(l)|b_X(l+1),...,b_X(l+m),Y). \tag{27}$$

To determine (27), we need the joint p.d.f.:

$$p(b_X(l),b_X(l+1),...,b_X(l+m),Y). \tag{28}$$

We denote $p(\langle b_X(l+m)...b_X(l+1)b_X(l)\rangle = i, Y = y)$ by $\gamma_i(y)$. Following the discussion in Section IV-B.1 it can be shown that

$$\gamma_i(y) \quad = \quad \int_{A_i} f_{X,Y}(x,y)dx \tag{29}$$

where A_i are subsets of X (Figure 6):

$$A_i = \{\ x \quad | \quad c\cdot 2^{l+m+1} + i\cdot 2^l \le |x| \le c\cdot 2^{l+m+1} + i\cdot 2^l + 2^l - 1,\ c \in \mathbb{Z}^+\}. \tag{30}$$

[8]Note that while this model is commonly used in distributed video applications (e.g., [14]), many lossy image compression applications may not communicate the LSB bit-planes and therefore a full resolution version of Y would not be available to be used for joint decoding.

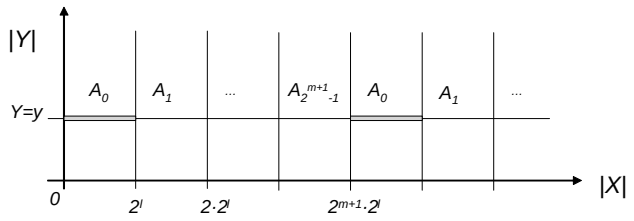


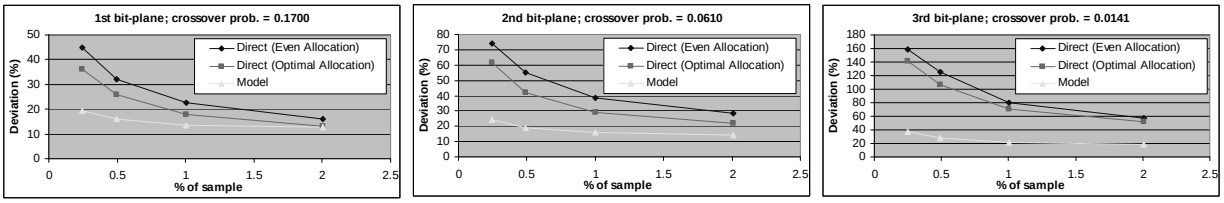
Fig. 6. Model based estimation: The events leading to the occurrence of $\langle b_X(l+m)...b_X(l+1)b_X(l)\rangle = i$ correspond to regions A_i, subset of X.

C. Experiments

We now compare the accuracy of direct estimation and model-based estimation as discussed in Section IV-A. We generate i.i.d. Laplacian sources X and Z of dimension M with model parameters β and α respectively, i.e., $f_X(x) = \frac{1}{2}\beta e^{-\beta|x|}$, $f_Z(z) = \frac{1}{2}\alpha e^{-\alpha|z|}$. Then the crossover probability p_l of X and $Y(= X + Z)$ at significance level $l, 0 \leq l \leq L-1$, is estimated using the following approaches:

- *Direct estimation with even sample allocation.* This is similar to the estimation method in Section II, where n binary samples of the l-th bit-planes are exchanged. Since there are L bit-planes in total, the total amount of exchanged data is $L \cdot n$ bits.
- *Direct estimation with optimal allocation.* This is similar to the aforementioned approach except that the optimal sample allocation (17) is used to distribute the $L \cdot n$ binary samples among bit-planes.
- *Model-based estimation.* Here n L-bits random samples of Y are sent to the encoder of X, where the model parameters β and α are estimated from the n samples of X and $Z(= Y - X)$ respectively, using MLE [16]. Then the estimate of $p_l, 0 \leq l \leq L-1$, can be derived analytically from $\hat{\beta}$ and $\hat{\alpha}$ using (21). Therefore, the model-based approach also incurs $L \cdot n$ bits to estimate the crossover probabilities of all the bit-planes.

Note that in the direct approach we do not include offset in the estimator, i.e., the estimator is $\frac{S(n)}{n}$ following the notations in Section II. Therefore, it is fair to compare with (21). Practical applications may choose to bias both the direct and model-based estimators as in (8) (in model-based estimation we would replace $\frac{S}{n}$ by that calculated in (21)). The approaches are compared based on the deviation of the estimates from the true (empirical) crossover probability: $|\hat{p}_l - p_l|/p_l$. The deviation is measured for different bit-planes using different percentage of exchanged samples, $\frac{n}{M}$, with $\beta = 0.3$, $\alpha = 2.5$, $M = 6480$. These parameters are similar to those observed in the video data used in the experiments in Section II-C, i.e., X and Y are the quantized DCT coefficients in a current frame and the corresponding quantized



(a) The 1st bit-plane. (b) The 2nd bit-plane. (c) The 3rd bit-plane.

Fig. 7. Comparing estimation accuracy between the direct approaches and the model-based approach.

coefficients in the motion-compensated predictors in the reference frame, respectively. Figure 7 depicts the comparison results, where the deviations are obtained by averaging over $N_E = 1000$ experiments. As shown in the figure, the model-based estimation can achieve considerable improvements in the estimation accuracy, especially when only a small number of samples are used or crossover probability is small. Note that model-based estimation utilizes the information that the bits to be encoded have been extracted from continuous-valued data following certain distributions, and therefore would tend to perform better than direct estimation, which does not use such information. However, model-based estimation is applicable only to bit-planes extracted from continuous sources, and obviously its performance depends on how accurately the continuous-valued data can be modeled. Additional experiments assessing the performance in terms of coding rate and distortion using a real application will be presented in Section VI.

V. Model-based Estimation on Structured Bit-planes

In this section we discuss how to extend the model-based estimation to the cases when bit-planes are extracted using more sophisticated methods. For example, in the cases when X and Y are wavelet transform coefficients[9], bit-planes are usually partitioned depending on the magnitude of the transform coefficients to improve coding efficiency, as in the *set-partitioning* algorithm used in SPIHT [17]. Specifically, in these "significance coding" techniques, the encoder first signals the *significance* of each of the components at a given bit-plane. After a component becomes significant, *sign* information is conveyed and then further *refinement* bits are transmitted. While different wavelet systems may use different techniques to encode the bits (e.g., context coding can be used as in JPEG2000 or alternatively zerotree coding can be used as in SPIHT to represent significance maps), the definitions of the (uncoded) sign/refinement bits

[9]A concrete application scenario can be X and Y are collocated wavelet transform coefficients of two correlated images.

and significance maps are largely the same, so the techniques we propose in the context of SPIHT in this section would also be applicable to other wavelet-based compression schemes that use bit-plane encoding. We shall consider systems where SW coding is applied to compress the sign/refinement bit-planes[10], and propose extension of model-based approach to estimate the corresponding crossover probabilities, so that encoding rate can be determined. The extended approach can also be applied to facilitate adaptive combinations of SW/entropy coding to improve coding performance [26].

In what follows we will first discuss how to extend model-based approach to estimate crossover probabilities of sign/refinement bit-planes. Since significance coding is usually used in wavelet-based applications (e.g., [17], [27]), we will also discuss how to address some of the issues when applying model-based estimation in wavelet-based DSC applications.

A. Model-based Estimation for Refinement/Sign Bit-planes

Given an input source X_i to be compressed using side information Y_i $(= X_i + Z_i)$, and following the same assumptions in Section IV-A, our goal is to estimate the crossover probability of the refinement and sign bit-planes of significance level l, denoted as $p_{ref}(l,i)$ and $p_{sgn}(l,i)$ respectively[11]. Following from the definition of refinement bits, the refinement bit-plane of significance level l includes only coefficients that are *already* significant [17], i.e., $|X_i| >= 2^{l+1}$. Therefore, the crossover probability of the l-th refinement bit-plane for source X_i is

$$p_{ref}(l,i) = \frac{Pr(R \cap |X_i| >= 2^{l+1})}{Pr(|X_i| >= 2^{l+1})} \tag{31}$$

where R denotes the event of crossover in magnitude bits, i.e., $R = \bigcup A_i$, with A_i defined in (20). Following the discussion in Section IV-A, we can calculate $Pr(R \cap |X_i| \geq 2^{l+1})$ by integrating the joint p.d.f. of X_i and Y_i, $f_{X_iY_i}$, over the shaded regions in Figure 3(b), similar to (21). Moreover, $f_{X_iY_i}$ can be factorized as in (22). Therefore, $p_{ref}(l,i)$ can be readily calculated.

The crossover probability of sign bit-planes can be derived in a similar fashion. The difference here is we need to integrate different regions in the sample space of X_i and Y_i. The l-th sign bit-plane

[10]Since significance map carries structural information, a single decoding error in the significance map would cause decoding failure of all the subsequent bits, and therefore SW coding may not be suitable for compressing significance map.

[11]We introduce the subscript i in this section to facilitate the discussion of wavelet-based applications in the next section. Specifically, in Section V-B, X_i will be used to denote the wavelet transform coefficient in the i-th subband. We use separate models for different subbands in order to take into account different statistics in different subbands (e.g., variances tend to decrease when going from high level subbands to low level subbands).

includes only the sign bits of the coefficients that *become* significant at significance level l [17], i.e., $2^{l+1} > |X_i| \geq 2^l$. Hence the crossover probability of the l-th sign bit-plane in the source X_i is

$$p_{sgn}(l,i) = \frac{Pr(S \ \cap \ 2^{l+1} > |X_i| \geq 2^l)}{Pr(2^{l+1} > |X_i| \geq 2^l)} \tag{32}$$

where $S = \{ \ (x_i, y_i) \mid x_i > 0, \ y_i < 0\} \cup \{ \ (x_i, y_i) \mid x_i < 0, \ y_i > 0\}$ denote events of crossover in sign bits. $Pr(S \ \cap \ 2^{l+1} > |X_i| \geq 2^l)$ can be calculated by integrating the joint p.d.f. of X_i and Y_i over the shaded regions in Figure 3(c), similar to (21), and factoring the p.d.f. as in (22). However, estimating $f_{X_i}(x)$ and $f_{Z_i}(z)$ may not be necessary, since this has been done in refinement crossover estimation.

B. Model-based Estimation for Wavelet-based Applications

Since significance coding is used mostly for wavelet-based compressions, in this section we will discuss the particular scenarios of applying model-based estimation for wavelet-based DSC applications. We denote X the wavelet transform coefficients of the input data, with the i-th subband denoted $X_i, 0 \leq i \leq N_B - 1$, where N_B is the total number of subbands. A main issue to extend model-based approach to wavelet-based applications is that in some subbands there may not be enough coefficients to obtain reliable estimates of the model parameters (e.g., high level subbands in the case of dyadic decomposition). We will discuss how to address the issue in the following.

1) Estimation with adequate number of samples: Following the discussion in Section V-A, to estimate the crossover probabilities of sign/refinement bit-planes, we need to estimate the models $f_{X_i}(x)$ and $f_{Z_i}(z)$. Experimental results on real image data suggest a single model $f_Z(z)$ can be used for all Z_i without any noticeable impact on coding performance. Estimation of $f_Z(z)$ (at the encoder of X) involves samples of Y and therefore is subject to communication/computational constraints, and only a small amount of samples $Z \ (= Y - X)$ should be used to estimate $f_Z(z)$ (as illustrated in the experiments in Section VI). On the other hand, estimation of $f_{X_i}(x)$ involves samples of X_i and results are affected by the number of coefficients in the i-th subband, N_i. Since some subbands may not have enough coefficients to obtain reliable estimate of $f_{X_i}(x)$, we partition the set of subbands $\{i \mid 0 \leq i \leq N_B - 1\}$ into $\mathbb{L}$ and $\mathbb{H}$, where $\mathbb{L}$ denotes the subset of subbands (low level subbands) which have enough coefficients for reliable estimation of the models, and $\mathbb{H}$ denotes the set of remaining subbands (high level subbands). The partition of all subbands into $\mathbb{L}$ and $\mathbb{H}$ is determined by N_i. In particular, if Laplacian model $f_{X_i}(x) = \frac{1}{2}\beta_i e^{-\beta_i |x|}$ is chosen of X_i, and MLE is used to estimate β_i, then the MLE estimator $\hat{\beta}_i$ has a *percentage deviation* $D = (\hat{\beta}_i - \beta_i)/\beta_i$. It can be shown that $D \sim N(0, 1/N_i)$, i.e., it depends on N_i only. Therefore, we

can select a threshold to apply to N_i in order to classify a subband into $\mathbb{L}$ or $\mathbb{H}$ according to a desired distribution of D.

Estimation of $p_{ref}(l,i)$ and $p_{sgn}(l,i)$, where $i \in \mathbb{L}$, can be performed using the algorithms discussed in Section V-A, with models $f_{X_i}(x)$ and $f_Z(z)$ estimated from transform coefficients samples using standard methods, e.g. MLE. Alternatively, the correlation noise model $f_Z(z)$ can be estimated from statistics in the raw data domain (e.g. pixel data) calculated using raw domain samples, and this can lead to some implementation advantage as transformation of the side-information is no longer required. For example, if Laplacian model $f_Z(z) = \frac{1}{2}\alpha e^{-\alpha|z|}$ is chosen for Z, then we can estimate α by calculating the standard deviation of Z, σ_Z, using raw domain samples, and using the relationship between standard deviation and model parameter in Laplacian distribution, $\alpha = \sqrt{2}/\sigma_Z$. This is viable since the variance of the correlation noise would be the same in the raw and transform domains if orthogonal filters are used. For some bi-orthogonal filters, e.g. 9/7, the variance of the correlation noise in the raw data domain would also be very close to that in the transform domain [28], and we can estimate α using similar procedures. For other bi-orthogonal filters, e.g. 5/3, the raw domain variance would need to be properly normalized, following the discussions in [28], so that α can be accurately estimated.

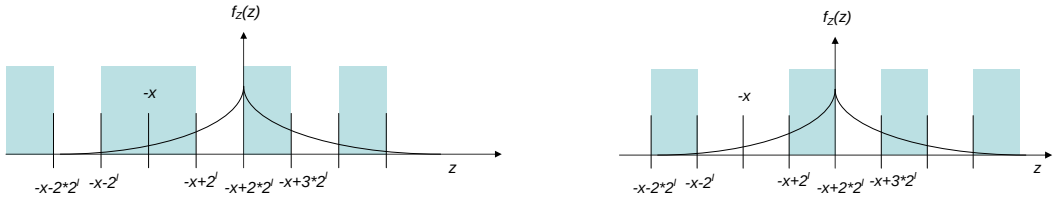
2) Estimation without adequate number of samples: Subbands in $\mathbb{H}$ do not have enough coefficients to estimate $f_{X_i}(x)$ reliably. Instead, we use the empirical p.m.f. $Pr(X_i = x)$ of subbands in $\mathbb{H}$ along with the correlation noise $f_Z(z)$ to estimate sign/refinement crossover probabilities. Specifically, we derive the average crossover probability for the refinement bits consisting of i-th subband coefficients, $i \in \mathbb{H}$, by

$$p_{ref}(l,i) = \sum Pr(U(l,x))Pr(X_i = x) \tag{33}$$

where $U(l,x)$ denotes the events of l-th refinement bits crossover when $X_i = x$, and the summation is taken over all the possible values of X_i where $Pr(X_i = x)$ is non-zero. We can determine $Pr(X_i = x)$ empirically during encoding by binning the coefficients in the i-th subband. To determine $Pr(U(l,x))$, we assume $Y_i = x + Z$ (note that here x is a constant instead of a random variable), and $U(l,x)$ is a subset of the sample space of Z and can be found to be equal to

$$U(l,x) = \begin{cases} \{z \mid & -x + (2k)2^l \le z \le -x + (2k+1)2^l \text{ , or} \\ & -x - (2k+1)2^l \le z \le -x - (2k)2^l\}, & \text{if } \left\lfloor \frac{|x|}{2^l} \right\rfloor \text{ is odd}, \\ \{z \mid & -x + (2k+1)2^l \le z \le -x + (2k+2)2^l \text{ , or} \\ & -x - (2k+2)2^l \le z \le -x - (2k+1)2^l\}, & \text{if } \left\lfloor \frac{|x|}{2^l} \right\rfloor \text{ is even}, \end{cases} \tag{34}$$

where $k \in \mathbb{Z}^+$. Therefore, $Pr(U(l,x))$ can be derived by summing the integrals of $f_Z(z)$ over the shaded regions as depicted in Figure 8. In practice we only need to sum over a few regions where the integrals



(a) (b)

Fig. 8. $Pr(U(l,x))$ when (a) $\lfloor \frac{|x|}{2^l} \rfloor$ is odd ; (b)$\lfloor \frac{|x|}{2^l} \rfloor$ is even.

are non-zero (e.g., around $Z = 0$, if Z is Laplacian distributed). Note that computing $Pr(X_i = x)$ by binning the coefficients may not incur much complexity as the subbands in $\mathbb{H}$ have only a small number of coefficients.

Similarly, we can derive $p_{sgn}(l,i)$, $i \in \mathbb{H}$, by

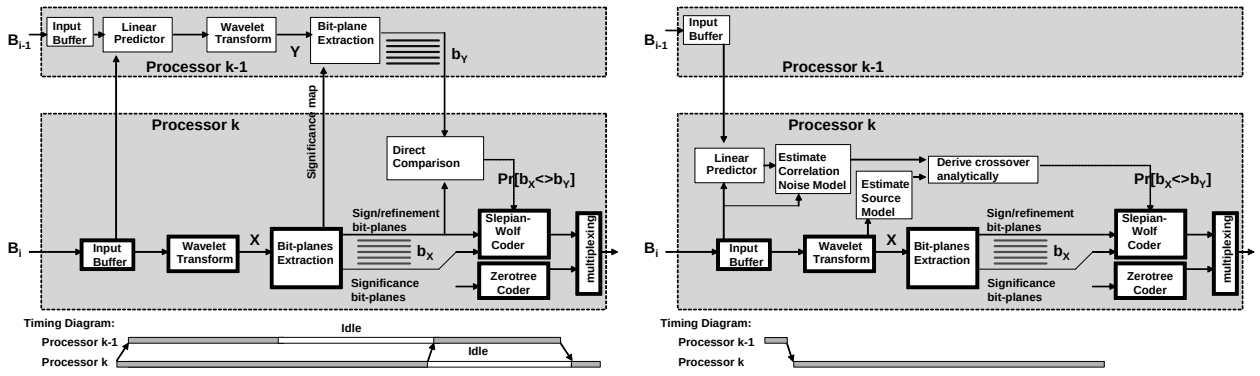
$$p_{sgn}(l,i) = \sum Pr(V(l,x))Pr(X_i = x) \tag{35}$$

with $V(l,x)$ denotes the event of the l-th sign bits crossover when $X_i = x$. It can be shown that $Pr(V(l,x)) = \int_{-\infty}^{-|x|} f_Z(z)dz$ when $f_Z(z)$ is symmetric. Note that we use (31) and (32) to estimate the crossover probabilities when there are enough samples in a subband to allow reliable estimation of $f_{X_i}(x)$, and (33) and (35) when there are insufficient samples in a subband and empirical p.m.f. $Pr(X_i = x)$ is used to characterize the data.

VI. Hyperspectral Image Compression Experiments

In this section we describe several additional experiments on the proposed algorithms using real image compression applications. In particular, we use a DSC-based hyperspectral image compression proposed in [20], [26] as a test-bed[12] (Figure 9(a)). To compress the current spectral band B_i, the sign and magnitude bits of the wavelet transform coefficients are extracted using an algorithm similar to the standard SPIHT [17], with modifications such that at some significance levels the magnitude bits are extracted as raw bit-planes (instead of separating them into significance and refinement bit-planes) in order to improve coding performance. Then LDPC-based Slepian-Wolf coding is employed to compress sign/refinement/raw bit-planes, using as side information the sign/refinement/raw bit-planes of same significance extracted from $a\hat{B}_{i-1} + b$, where $\hat{B}_{i-1}$ is the previous adjacent reconstructed band available only at the decoder, and a

[12]We choose the hyperspectral image applications for experiments mainly because of the availability of the system.



(a) (b)

Fig. 9. DSC-based hyperspectral image compression with (a) direct correlation estimation; (b) model-based estimation.

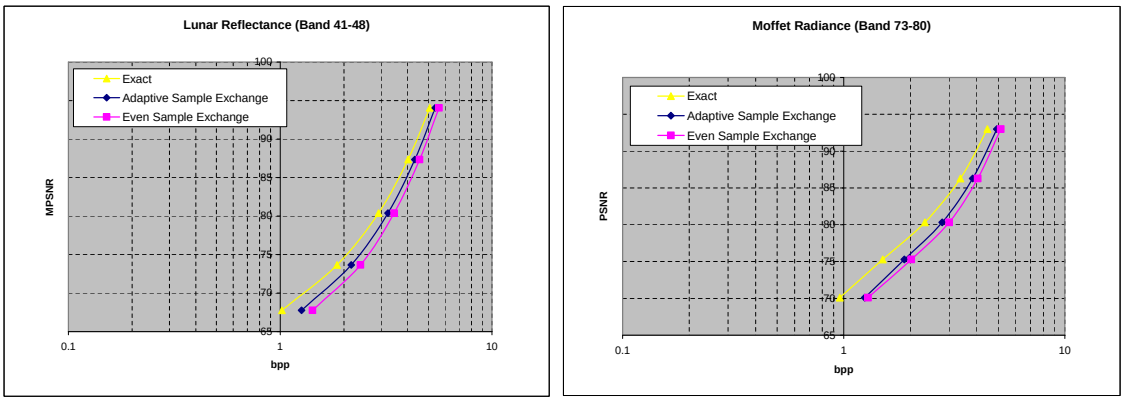
and b are some linear prediction coefficients. To determine the coding rate, the *original* previous band B_{i-1} is used to approximate $\hat{B}_{i-1}$ at the encoder for high fidelity applications, and sign/refinement/raw bit-planes are explicitly extracted from transform coefficients of $aB_{i-1}+b$ to estimate the crossover probabilities using the direct estimation approach discussed in Section II. The amount of information exchanged needs to be kept small so that the algorithm can be used in parallel encoding scenarios, where each band is assigned to a different processor. To ensure source and SI bit-planes are formed with wavelet coefficients at the same locations, we need to apply significance trees of B_i when extracting bit-planes from B_{i-1} [20].

A. Sample Allocation Experiments

Given that n_T binary samples can be used to estimate the crossover probabilities when compressing B_i, we compare two strategies to determine how to allocate the samples to different bit-planes:

- *Adaptive sample allocation.* We use (17), i.e., the optimal sample allocation, to decide the numbers of samples allocated to different bit-planes. However, since the crossover probabilities of B_i are unknown, we use as a priori information the crossover probabilities of B_{i-1} in (17) (which have been estimated during the compression of B_{i-1}). When compressing the first DSC-coded band, since a priori information is not available we allocate the same number of samples to each bit-plane.
- *Even sample allocation.* We allocate the same number of samples to each bit-planes for all the bands.

The NASA AVIRIS image data-sets [29] are used in the experiment. The original image consists of 224 spectral bands, and each spectral band consists of 614×512 16-bits pixels. In the experiment, we



(a) (b)

Fig. 10. Sample allocation experiments - (a) *Lunar* (reflectance data), (b) *Moffet* (radiance data), using 0.25% total sample. An adaptive sample allocation scheme using the proposed optimal sample allocation strategy in Section III with a priori information from previous encoded band is compared with the even sample allocation. Here "Exact" denote the cases when the exact correlation information is used to determine the coding rate, i.e., no rate penalty.

compress 512×512 pixels in each band. Figures 10(a) and 10(b) depict the RD performances of the system under different sample allocation strategies. Here MPSNR $= 10\log_{10}(65535^2/\text{MSE})$, where MSE is the mean squared error between all the original and reconstructed bands. Also shown in the figures are the RD performances when the exact crossover probabilities are used to determine the coding rate. As shown in the figures, in the situations with small number of samples exchanged (e.g., 0.25% of total), the adaptive sample allocation can reduce the rate penalty by about 1dB, as compared to the even sample allocation. Note that the adaptive sample allocation requires negligible overhead: it simply uses (17) to determine a more efficient sample allocation across bit-planes based on any available a priori information.

B. Model-based Estimation Experiments

Model-based estimation can be applied to the hyperspectral image system following the algorithms outlined in Sections IV and V, with X and Y being the transform coefficients of B_i and $aB_{i-1}+b$ respectively. As discussed, continuous-valued source samples are used to estimate the models in model-based estimation, so bit-plane extraction from SI is no longer necessary. In addition, the correlation noise model can be estimated in the pixel domain in this case following the discussion in Section V-B. Therefore, the model-based estimation can result in a more efficient system as depicted in Figure 9(b). Here the core compression module is the same as that in the original system (with direct estimation) in

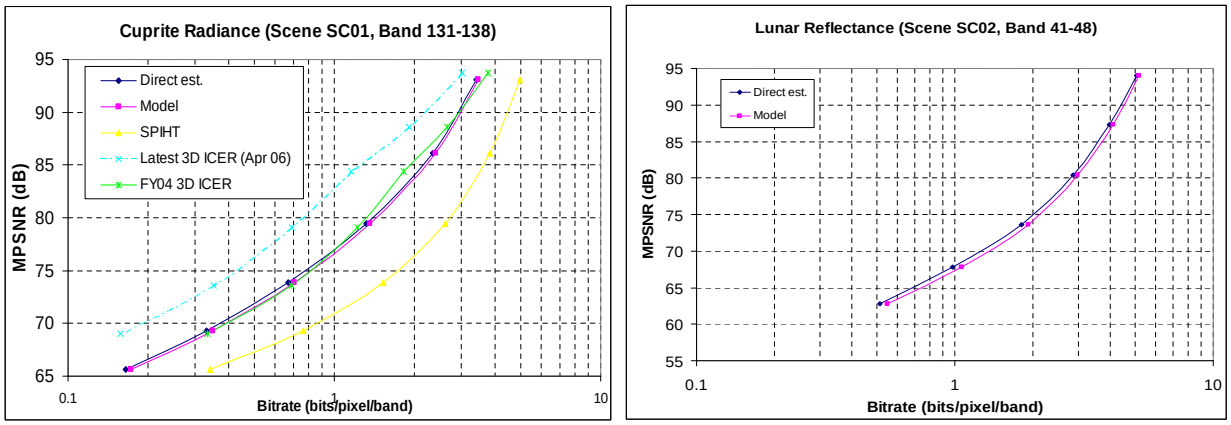
Figure 9(a), while the correlation estimation algorithm is modified following the model-based approach, leading to the following implementation advantages in this application:

- First, the model-based system requires less computation. This is evident when comparing Figures 9(a) and 9(b): Wavelet transform on and bit-planes extraction from SI as in the original system[13] are no longer required, while estimating the model (using MLE) and the crossover probability (using analytical equations) require only small amounts of computation in the model-based system.
- Moreover, in parallel encoding scenarios, the model-based system requires less data traffic between processors. This is also evident when comparing Figures 9(a) and 9(b): In the model-based system, encoder of B_i only needs to request pixel domain samples from encoder of B_{i-1} at the beginning of processing to compute the linear prediction coefficients and correlation noise model, whereas in the original system additional traffic is incurred for exchanging significance tree and SI bit-planes.
- Furthermore, the model-based system can achieve better parallel encoding, as there are only small amounts of dependency between encoders of different bands at the beginning of processing (Figure 9(b)), and processors can proceed without further synchronization.

The performance of model-based estimation is assessed by comparing with the original system, where SI bit-planes are explicitly extracted and *exact* crossover probabilities are used to estimate the encoding rate, i.e., there is no rate penalty. In the model-based approach, samples of B_{i-1} are obtained by downsampling the image by factors of four and eight horizontally and vertically respectively. Therefore, 3.125% of image data are used to estimate the correlation noise model. To prevent decoding error due to under-estimating the crossover probability, we allow a larger margin to determine the encoding rate, at the expense of coding efficiency. Specifically, a 0.15-bit margin is added to the estimated Slepian-Wolf bound, so that there is no decoding error occurred in the testing data-sets. Figures 11 depicts the RD performance. As shown in the figures, model-based estimation incurs only small degradation in coding efficiency. In most cases, the difference is less than 0.5dB when compared to the direct estimation with exact crossover probabilities used to determine the coding rate.

For reference we also present the coding performance of several 3D wavelet systems (3D-ICER) developed in NASA-JPL [29]. As shown in Figure 11(a), the DSC-based systems are comparable to a version of 3D wavelet system (FY04-3D-ICER) in terms of coding efficiency. FY04-3D-ICER uses the standard dyadic wavelet decomposition and a context-adaptive entropy coding scheme to compress

[13]It may be possible to avoid wavelet transform of SI by storing and re-using the coefficients during the compression of B_{i-1}. But this would significantly increase the memory requirements. And bit-plane extraction from SI is always required if generating SI bit-planes explicitly, since the SI sign/refinement bits need to extracted based on the significance tree of B_i.



(a) (b)

Fig. 11. Coding efficiency comparison: (a) *Cuprite* (radiance data); (b) *Lunar* (reflectance data). The model-based system is compared with the original system, which uses all samples in direct estimation and exact crossover probabilities to determine the encoding rate, i.e. no rate penalty. Coding performances of several other wavelet systems are also presented for reference.

coefficients bits. However, there is still performance gap when comparing the DSC-based systems to a more recent and sophisticated version of 3D wavelet (Latest-3D-ICER). Latest-3D-ICER exploits the spatial correlation remaining in the correlation noise [29], while the DSC-based systems use simple i.i.d. model for correlation noise and do not utilize the spatial correlation in the noise, leading to some coding inefficiency. We have also compared the DSC-based systems with 2D-SPIHT, and the DSC-based systems can achieve 8dB gains at some bit-rates, as shown in Figure 11(a).

VII. Conclusions

In this paper, we have investigated correlation estimation for distributed image and video applications under rate and complexity constraints. Focusing on the situations when sampling techniques are employed to estimate the correlation, we first analyzed how the number of samples relates to the p.d.f. of the rate penalty when compressing a binary input source. We then extended the analysis to the cases when multiple binary input sources are to be compressed and proposed a strategy to allocate samples to the sources such that the overall rate penalty can be minimized. Furthermore, we proposed a model-based estimation for the particular but important situations when bit-planes are extracted from a continuous-valued input source, and discussed extensions to the cases when bit-planes are extracted based on the significance of the data for wavelet-based DSC applications. Experimental results including real image compression demonstrated model-based approach can achieve accurate estimation.

APPENDIX I
DERIVATION OF OPTIMAL SAMPLE ALLOCATION

Here we give the detail of the derivation of the optimal sample allocation given in (17). We consider

$$\min_{\{n_l:\sum_{l=0}^{L-1} n_l=n_T;n_l\leq K_l\}} E[\triangle H], \tag{36}$$

where $E[\triangle H]=\frac{1}{K_T}\sum_{l=0}^{L-1} K_l\alpha_l n_l^{-1/2}$ and $\alpha_l=ln(\frac{1}{p_l}-1)z_{\omega/2}\sqrt{p_l(1-p_l)}$. Using Lagrange multipliers, we construct

$$J(\mathbf{n})=\frac{1}{K_T}\sum_{l=0}^{L-1} K_l\alpha_l n_l^{-1/2}+\lambda\sum_{l=0}^{L-1} n_l. \tag{37}$$

Differentiate (37) and set to zero,

$$\frac{\partial J}{\partial n_l}=\frac{-1}{2K_T}K_l\alpha_l n_l^{-3/2}+\lambda=0, \tag{38}$$

or

$$n_l^*=(K_l\alpha_l)^{2/3}(\frac{1}{2\lambda K_T})^{2/3}=(K_l\alpha_l)^{2/3}\gamma. \tag{39}$$

This is true as long as $n_l^*<K_l$. Otherwise, we need to use the Kuhn-Tucker conditions to find n_l^*:

$$\frac{\partial J}{\partial n_l}\begin{cases}=0, & \text{if } n_l^*<K_l,\\ \leq 0, & \text{if } n_l^*=K_l.\end{cases} \tag{40}$$

Solving (40) we obtain the results in (17).

ACKNOWLEDGMENT

The work was supported in part by NASA-JPL. The authors would like to thank Sam Dolinar, Aaron Kiely and Hua Xie of NASA-JPL for helpful discussions and for providing data for comparison. The authors would also like to thank the anonymous reviewers for their comments which helped improve the paper significantly.

REFERENCES

[1] N.-M. Cheung, H. Wang, and A. Ortega, "Correlation estimation for distributed source coding under information exchange constraints," in *Proc. Int'l Conf. Image Processing (ICIP)*, 2005.

[2] N.-M. Cheung and A. Ortega, "A model-based approach to correlation estimation in wavelet-based distributed source coding with application to hyperspectral imagery," in *Proc. Int'l Conf. Image Processing (ICIP)*, 2006.

[3] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, vol. 19, pp. 471–480, July 1973.

[4] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. 22, pp. 1–10, Jan. 1976.

[5] S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, pp. 51–60, Mar. 2002.

[6] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, pp. 80–94, Sept. 2004.

[7] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[8] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *Proc. Data Compression Conference (DCC)*, 1999.

[9] R. Puri and K. Ramchandran, "PRISM: a new robust video coding architecture based on distributed compression principles," in *Proc. Allerton Conf. Communications, Control, and Computing*, Oct. 2002.

[10] 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, Oct. 2002.

[11] N.-M. Cheung, C. Tang, A. Ortega, and C. S. Raghavendra, "Efficient wavelet-based predictive Slepian-Wolf coding for hyperspectral imagery," *EURASIP Journal on Signal Processing - Special Issue on Distributed Source Coding*, Nov. 2006.

[12] V. Thirumalai, I. Tosic, and P. Frossard, "Distributed coding of multiresolution omnidirectional images," in *Proc. Int'l Conf. Image Processing (ICIP)*, 2007.

[13] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. Data Compression Conference (DCC)*, 2002.

[14] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver," in *Proc. Int'l Conf. Image Processing (ICIP)*, 2004.

[15] X. Zhu, A. Aaron, and B. Girod, "Distributed compression for large camera arrays," in *Proc. Workshop on Statistical Signal Processing*, 2003.

[16] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences*. Prentice-Hall, 1995.

[17] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec using set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, pp. 243–250, June 1996.

[18] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Nov. 2002.

[19] J. Wang, A. Majumdar, K. Ramchandran, and H. Garudadri, "Robust video transmission over a lossy network using a distributed source coded auxiliary channel," in *Proc. Picture Coding Symposium (PCS)*, 2004.

[20] C. Tang, N.-M. Cheung, A. Ortega, and C. S. Raghavendra, "Efficient inter-band prediction and wavelet based compression for hyperspectral imagery: A distributed source coding approach," in *Proc. Data Compression Conference (DCC)*, 2005.

[21] Z. He, L. Cao, and H. Cheng, "Correlation estimation and performance optimization for distributed image compression," in *Proc. Visual Communications and Image Processing (VCIP)*, 2006.

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[23] P. Lassila, J. Karvo, and J. Virtamo, "Efficient importance sampling for monte carlo simulation of multicast networks," in *Proc. Conf. Computer Communications (INFOCOM)*, 2001.

[24] S. R. Smoot and L. A. Rowe, "Study of DCT coefficient distributions," in *Proc. SPIE*, Jan. 1996.

[25] R. P. Westerlaken, S. Borchert, R. K. Gunnewiek, and R. L. Lagendijk, "Analyzing symbol and bit plane-based LDPC in distributed video coding," in *Proc. Int'l Conf. Image Processing (ICIP)*, 2007.

[26] N.-M. Cheung and A. Ortega, "An efficient and highly parallel hyperspectral imagery compression scheme based on distributed source coding," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2006.

[27] D. Taubman and M. Marcellin, *JPEG2000: Image compression fundamentals, standards and practice*. Kluwer, 2002.

[28] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proc. Data Compression Conference (DCC)*, 1996.

[29] A. Kiely, M. Klimesh, H. Xie, and N. Aranki, "ICER-3D: A progressive wavelet-based compressor for hyperspectral images," NASA-JPL IPN Progress Report, Tech. Rep., 2006, http://ipnpr.jpl.nasa.gov/progress_report/42-164/164A.pdf.