

Research Article

Distributed Compressed Video Sensing in Camera Sensor Networks

Yu Liu,¹ Xuqi Zhu,¹ Lin Zhang,¹ and Sung Ho Cho²

¹Key Lab of Universal Wireless Communications, Ministry of Education of PRC, Beijing University of Posts and Telecommunications, Beijing 100876, China

²Department of Electronics and Computer Engineering, Hanyang University, Seoul 133791, Republic of Korea

Correspondence should be addressed to Yu Liu, liuy@bupt.edu.cn

Received 5 June 2012; Revised 8 December 2012; Accepted 9 December 2012

Academic Editor: Sartaj K. Sahni

Copyright © 2012 Yu Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the booming of video devices ranging from low-power visual sensors to mobile phones, the video sequences captured by these simple devices must be compressed easily and reconstructed by relatively more powerful servers. In such scenarios, distributed compressed video sensing (DCVS), combining distributed video coding (DVC) and compressed sensing (CS), is developed as a novel and powerful signal-sensing and compression algorithm for video signals. In DCVS, video frames can be compressed to a few measurements in a separate manner, while the interframe correlation is explored by the joint recovery algorithm. In this paper, a new DCVS joint recovery scheme using side-information-based belief propagation (SI-BP) is proposed to exploit both the intraframe and interframe correlations, which is particularly efficient over error-prone channels. The DCVS scheme using SI-BP is designed over two frame signal models, the mixture Gaussian (MG) model and the wavelet hidden Markov tree (WHMT) model. Simulation results evaluated on two video sequences illustrate that the SI-BP-based DCVS scheme is error resilient when the measurements are transmitted through the noisy wireless channels.

1. Introduction

Current video coding paradigms, such as MPEG and the ITU-T H.26x, are traditionally designed for the applications followed the so-called “broadcast” model, as shown in the left part of Figure 1. The video sequence is complicatedly encoded at the powerful server only once, and then the compressed video stream is distributed and decoded frequently on many cheap and simple user devices. So MPEG and H.26x standards both have complicated encoder and light decoder.

However, with the booming of video devices ranging from low-power visual sensors to camera mobile phones, visual applications now have already developed beyond this broadcast model. The video processing paradigm in camera sensor networks, which are composed of spatially distributed smart camera devices capable of processing images or videos of a scene from a variety of viewpoints, is more like a “multiple-access” model, as shown in the right part of Figure 1. In this scenario, these video devices with limited battery power and storage memory need to send their captured video

streams to the monitor server. Meanwhile, high compression efficiency is also required considering both the limitations of wireless bandwidth and transmission power. The requirements of the video processing paradigms here are diametrically opposed to MPEG and H.26x. Thus, lightweight and efficient encoding technologies are developed to satisfy such requirements of these multiaccess applications.

The first step to satisfy the multiaccess model has been taken by exploiting the interframe statistics at the decoder only, known as distributed video coding (DVC) [1], and the information-theoretic basis of DVC is distributed source coding (DSC) [2]. It states that the correlated sources can be separately encoded and jointly decoded using the correlation between them; furthermore, this separate encoding will achieve the same coding efficiency as the joint encoding. The concept of DSC has achieved a lot of attentions with the booming of wireless sensor networks where the correlated sources captured by sensors have to be encoded without communications with each other while decoded jointly at the sink node.

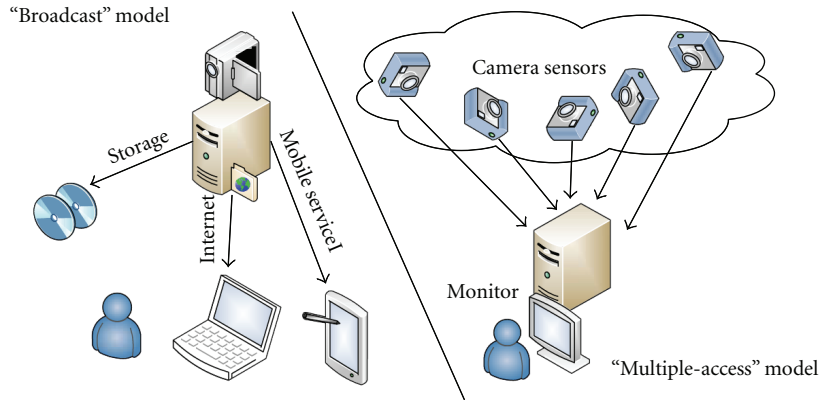


FIGURE 1: The comparison of the "Broadcast" model and the "Multiple-access" model.

Slepian-Wolf theorem [2] for lossless coding and Wyner-Ziv theorem [3] for lossy coding are the most important theoretical foundations of DSC. The first practical strategy of DSC is proposed in [4] to exploit the potential of the Slepian-Wolf theorem by introducing channel codes. The statistical dependence between the two sources is modeled as a virtual correlation channel. One source is used as the side information to help the decoding of the other source.

DVC combines DSC methods with traditional intraframe video coding systems so as to shift the complicated motion search operations from the encoder to the decoder, and perfectly suits the "multiple-access" model. A classical framework called power-efficient, robust, high-compression, syndrome-based multimedia coding (PRISM) is proposed in [5]. They encoded the P frames following the same procedures as that of I frames but at lower rates, while the motion search is used at the decoder to estimate the side information from the neighboring recovered I frames. Based on the side information, the P frames can be successfully reconstructed using DSC decoding algorithms.

Girod et al. [6] investigated DVC in the similar format but using turbo codes instead of the trellis codes in [5]. The scheme in [7] tackled the multiview correlations using DVC methods. DVC not only provides light video encoder but also contributes to the low-cost protection of traditional video coding. The layered Wyner-Ziv video coding system [8] achieved robust video transmission by adding Wyner-Ziv bitstream layers as the enhanced layers.

Another recent direction to achieve light encoder is to exploit the intraframe signal's sparsity property, known as compressed sensing (CS) [9–14]. In traditional intraframe coding, a large number of pixel values are firstly transformed, and then only the important low-frequency coefficients are entropy coded while other coefficients are discarded. In CS, we denote an N -dimensional vector as $X = (x_1, \dots, x_N) \in \mathbb{R}^N$ that can be represented as $X = \Psi\theta$, where Ψ is an orthogonal N -by- N matrix and $\theta \in \mathbb{R}^N$ is the projection of X on basis Ψ . If θ has only K nonzero entries, $K \ll N$, we say that X is K -sparse with respect to representing matrix Ψ . If θ is approximately sparse with K larger entries, X is called compressible.

As long as an M -by- N sensing matrix Φ can be found incoherent with the representing matrix Ψ , the signal X can be sampled as $Y = \Phi X$, where $Y = \{y_1, \dots, y_M\} \in \mathbb{R}^M$, $M < N$, is the measurement vector. With the measurements, the signal can be recovered by solving the ℓ_p -norm minimization problem:

$$\min_{\tilde{\theta} \in \mathbb{R}^N} \|\tilde{\theta}\|_p \quad \text{s.t. } Y = \Phi\Psi\tilde{\theta}. \quad (1)$$

The solution to the optimization program is $\hat{\theta}$, and then the estimated signal is $\hat{X} = \Psi\hat{\theta}$. According to the compressive sensing theory [9–14], the recovery can be achieved with probability close to one, if M satisfies $M \sim 2K \log(N/K)$ [14].

Sparsity [11] is one of the essential factors of CS which guarantees the signal can be compressively sampled, and the other is the incoherence [12] which provides the sensing matrices. Usually, random matrices are largely incoherent with any fixed representing matrix. In the CS methodology, the image or video signals are usually sparse on the basis of discrete cosine transform (DCT) matrix and discrete wavelet transform (DWT) matrix [15], where the frame signal can be directly compressed to a small number of measurements. Thus, CS may greatly improve the efficiency and decrease the complexity of intraframe compression. CS has been applied in image coding by many researchers [16, 17], and [18, 19] discussed the modified transforming, blocking and quantizing methods for video coding according to the feature of CS measurements.

Combining DVC and CS is a main research direction for video compressed sensing, so called distributed video compressed sensing (DCVS). In [20], the authors reconstructed the difference between frame signals firstly using ordinary gradient projection for sparse reconstruction (GPSR) [21] algorithm, and then recovered each signal. GPSR is essentially a gradient projection algorithm applied to a quadratic programming formulation of (1), in which the search path from each iteration is obtained by projecting the negative-gradient direction onto the feasible set. [22] tried to explore the correlation between random measurements of signals using Wyner-Ziv method, but such cascaded design needs two sets of encoders and decoders, which will consequently

increase the complexity. Taking the side information into account, the scheme in [23] modified the initializations and various stopping criteria within GPSR recovery. The idea is similar to ours, but the recovery algorithm using basis pursuit is different from ours. It also has to be noted that all these schemes did not consider the transmission noise of the measurements.

In this paper, a novel DCVS scheme using side-information-based belief propagation (SI-BP) to utilize both the interframe and intraframe correlations of video sequence is proposed. Each frame is compressed separately using structured sensing matrices. However, the P frame has lower compression ratio because it can be jointly decoded using SI-BP where the reconstructed adjacent I frames can be utilized to generate the side information. The algorithm SI-BP is derived from the Bayesian inference, so the frame signal model is crucial for the performance of the DCVS scheme. Two signal models are introduced, the mixture Gaussian (MG) and the wavelet hidden Markov tree (WHMT). Due to the recovery method based on Bayesian inference, our scheme is resilient to the noisy transmission channel which is inevitable in wireless networks. Although the recovery method based on the belief propagation (BP) has been introduced into CS in [24], it is only designed for single signal not for image or video signals.

The proposed DCVS scheme has three advantages. First, the system structure of DVC is introduced so that the motion search is moved from encoder to decoder to alleviate the complexity of the encoder. Second, a coding-theory-like sensing and recovery scheme is proposed based on Bayesian inference where the SI-BP algorithm is used, which is quite different from the optimization recovery schemes in prior work. Third, the MG and WHMT models of wavelet coefficients are exploited to recover signals not only from the point of sparsity but also the point of statistical distribution. The sufficient utilization of the video frame signal's properties makes our scheme robust to noise-prone transmission channels.

The rest of the paper is organized as follows. Section 2 introduces the two MG and WHMT frame signal models. In Section 3, we present the details of the proposed DCVS scheme consisting of separate compression and joint recovery algorithms. In Section 4, simulation results are illustrated and discussed. And finally, Section 5 gives some concluding thoughts with directions for future works.

2. Frame Signal Models

In our SI-BP based DCVS scheme, the correlations between the frames are used to initialize the decoding iteration. So the correlation models deeply affect the performance of the decoder. Generally, the frame signal of videos is compressible in the DCT basis or DWT basis, and the transform coefficients constitute a compressible signal with special construction model. In this section, we will discuss the effectiveness of the MG model and the WHMT model.

2.1. Mixture Gaussian Model. The MG model has been proved to be a simple yet effective model of real sparse signals

in image processing and inference problems. The wavelet coefficients of the video frames can be regarded as a K -sparse signal $X = (x_1, \dots, x_i, \dots, x_N) \in \mathbb{R}^N$. According to the magnitude of these coefficients, there are divided into K large elements and $N - K$ small elements. Thus, it is modeled as a two-state MG, where the large elements ("large" state) and small elements ("small" state) are picked from Gaussian distributions $N(0, \sigma_1^2)$ and $N(0, \sigma_0^2)$, respectively, where $\sigma_1^2 > \sigma_0^2$. So the probability density function of the element in the frame signal is given as

$$p(x_i) \sim \frac{K}{N}N(0, \sigma_1^2) + \frac{N-K}{N}N(0, \sigma_0^2). \quad (2)$$

The investigation of the information-theoretic bounds for the performance of CS is always a focal point [25–27] based on the particular sparse representations. For such two-state MG sparse signal, [25] has derived the rate-distortion bounds using the mean squared error (MSE) distortion measure. The simple upper bound on $D(R)$ is obtained using an adaptive two-step code. Firstly, the appearance of the "large" and "small" states obeys the Bernoulli distribution. Then, the Gaussian distributions of the two states are encoded respectively. Therefore, the distortion rate function of the DWT coefficients is upper bounded by

$$D(R) \leq c(t)\sigma^2 e^{-2R},$$

$$c(t) = \exp\left(2H_b(\mu(t)) + (1 - \mu(t)) \log \frac{\sigma_1^2(t)}{\sigma^2} + \mu(t) \log \frac{\sigma_0^2}{\sigma^2}\right). \quad (3)$$

The lower bound on $D(R)$ is found by considering the Markov chain of the $S \rightarrow X \rightarrow \hat{X}$, where S is a hidden discrete variance of the sparse pattern. The lower bound is

$$D(R) \geq \min_{\{R_s\}} \sum \pi_s \sigma_s^2 2^{-2R_s} \quad \text{s.t.} \quad \sum \pi_s R_s = R, \quad R_s \geq 0. \quad (4)$$

These results are for a single DWT vector, while for video coding, the correlation between I frame and P frame should be exploited. Thus, we discuss the rate-distortion bound for one sparse signal X with the side information Y . Based on the correlation model $X = Y + Z$, where Y and Z are independent sparse signals, we have the lower bound according to the conclusions in Wyner-Ziv coding [25]:

$$R_{X|Y}(D) \doteq H(Z) - \frac{1}{2} \log 2\pi e D, \quad (5)$$

where \doteq denotes asymptotic equality for $D \rightarrow 0$.

These results on rate-distortion bounds of DWT coefficients are the theoretical foundations of our proposed scheme. It can be noticed that the MG model is not unique for video frame signal but a universal model for any sparse signal. In other words, the particular features of video frame signals are not demonstrated fully through this model. So we adopt another model, WHMT model, which will be introduced in the next sub-section.

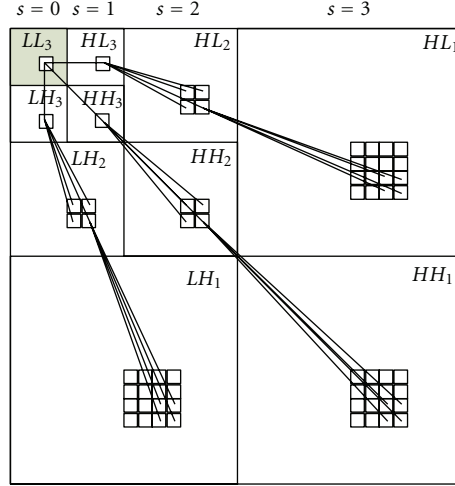


FIGURE 2: The quad tree structure of wavelet transform coefficients.

2.2. Wavelet Hidden Markov Tree Model. The DWT coefficients of an image or a video frame have the quad tree structure [15] which has been studied well before the advent of CS. And recent wavelet-based CS [28, 29] used this prior information into the CS reconstruction. We also introduce it into our joint recovery scheme.

Figure 2 shows the tree structure of the 3-level DWT coefficients of an image. The coefficients were decomposed with high (H) and low (L) pass filters at each level. HL_l , LH_l , HH_l represent the sub-band directions at level $l = 1, 2, 3$, respectively, and LL_3 is the basic representation of the image. Due to the analysis of DWT coefficient values, they tend to persist through scales for each sub-band. Therefore, we can construct a quad tree, where the coefficients at the highest level $l = L$ (L is the number of levels) is called the “root” with scale $s = 1$, and coefficients at the lowest level $l = 1$ is the “leaf” with scale $s = L$. The coefficients at LL_L is denoted as scale $s = 0$.

The DWT coefficients are modeled as a WHMT which is the general version of the two-state MG model. The two hidden states of this WHMT are also the “large” state and “small” state. The coefficient values of each state are drawn from a Gaussian distribution $N(0, \sigma_1^2)$ or $N(0, \sigma_0^2)$. However, the transition probability of a coefficient (not a root coefficient) between the two states is conditioned on the state of its parent coefficient. If the parent coefficient is in the “small” state, its children coefficients are in the “small” state with probability close to one. If the parent coefficient is in the “large” state, the “large” state and the “small” state are both possible for its children coefficients. And the root coefficient at scale $s = 1$ has high probability to be in the “large” state, while the coefficient at scale $s = 0$ must be in the “large” state. Thus, the prior probability of the i th coefficient at scale s is given as

$$P(\theta_{(s,i)}) \sim \pi_{(s,i)} N(0, (\sigma_{s1})^2) + (1 - \pi_{(s,i)}) N(0, (\sigma_{s0})^2), \quad (6)$$

where $\pi_{(s,i)}$ is the transition probability to the “large” state of coefficient $\theta_{(s,i)}$, and σ_{s1} and σ_{s0} are the standard deviations

in the “large” state, and “small” state at scale s respectively. Particularly, the setting of $\pi_{(s,i)}$ have some hints from the tree structure that

$$\pi_{(s,i)} = \begin{cases} 1, & s = 0 \\ \pi_1 \rightarrow 1, & s = 1 \\ \pi_{s1}, & s = 2, \dots, S, \omega_{p(s,i)} \text{ in “large” state} \\ \pi_{s0} \rightarrow 0, & s = 2, \dots, S, \omega_{p(s,i)} \text{ in “small” state,} \end{cases} \quad (7)$$

where $\omega_{p(s,i)}$ denotes the parent coefficient of $\theta_{(s,i)}$, and π_{s1} and π_{s0} mean the transition probability when the parent coefficient is in “large” state and “small” state, respectively.

3. Implementation of DCVS Using SI-BP

In this section, we will focus on the implementation of DCVS with SI-BP, including the separate compression and joint recovery algorithms. The framework of DCVS is shown in Figure 3. A frame group is assumed to consist of one I frame and two P frames. The I frames X_I , X_{I+1} and P frames X_P , X_{P+1} are all transformed into their DWT coefficients denoted by θ_I , θ_{I+1} and θ_P , θ_{P+1} respectively.

The coefficients of I frames and P frames are sampled by sensing matrices $\Phi_I^{M_I \times N}$ and $\Phi_P^{M_P \times N}$, respectively, where $M_P < M_I < N$. The measurements Y_I , Y_{I+1} , Y_P , and Y_{P+1} are transmitted through an AWGN channel with noise $W \sim N(0, \sigma_W^2)$, and the received versions are represented by Y'_I , Y'_{I+1} , Y'_P , and Y'_{P+1} . At the more powerful decoder, firstly the coefficients of two I frames are reconstructed as $\hat{\theta}_I$ and $\hat{\theta}_{I+1}$ so that the side information $\hat{\theta}_{SI}$ can be constructed via interpolation. Then the P frames are reconstructed as $\hat{\theta}_P$ and $\hat{\theta}_{P+1}$ with the help of the side information $\hat{\theta}_{SI}$ via the SI-BP algorithm. Finally, these frames are inversely transformed as \hat{X}_I , \hat{X}_{I+1} , \hat{X}_P , and \hat{X}_{P+1} to recover the original video.

3.1. Separate Compression. The BP and SI-BP algorithms for recovery are processed on the bipartite graph, so the sensing

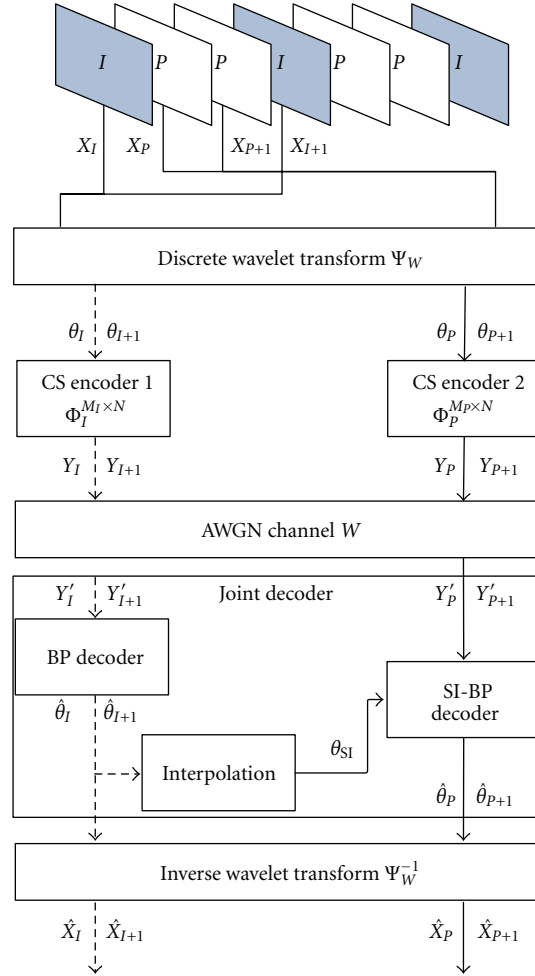


FIGURE 3: The system model of the proposed DCVS.

matrices have to be designed as low-density sensing matrices (LDSMs).

For the MG model, the LDSMs $\Phi_I^{M_I \times N}$ and $\Phi_P^{M_P \times N}$, are defined as $\Phi_I^{M_I \times N} \sim (d_I^v, d_I^c)$ and $\Phi_P^{M_P \times N} \sim (d_P^v, d_P^c)$ similarly to the regular LDSM that has d_I^v (d_P^v) non-zero elements in each column (variable node degree) and d_I^c (d_P^c) non-zero elements in each row (check node degree). The non-zero elements in this matrix are uniformly picked from the set $\{1, -1\}$. Intuitively, the larger d_*^v is, the more information of signal samples is preserved in measurements.

For the WHMT model, according to the unequal importance of different part of the tree, the LDSM is redefined as

$$\Phi^{M \times N} \sim (\{d^{v,1}, d^{v,2}, \dots, d^{v,l}, \dots, d^{v,L}\}, d^c), \quad (8)$$

where $d^{v,l}$ is the number of non-zero elements of the columns corresponding to the coefficients of layer l , $l = 1, 2, \dots, L$.

In other words, the LDSM is written in a layered format

$$\Phi^{M \times N} = [\Phi_1^{M \times N_1} \Phi_2^{M \times N_2} \dots \Phi_l^{M \times N_l} \dots \Phi_L^{M \times N_L}], \quad (9)$$

where N_l is the number of coefficients of layer l , and $\sum_{l=1}^L N_l = N$. $\Phi_l^{M \times N_l} \sim (d^{v,l}, d^c)$ is the sub-LDSM for layer l .

So, for both the MG and the WHMT model, the measurements of frames are generated as

$$Y_I = \Phi_I^{M_I \times N} \theta_I = \Phi_I^{M_I \times N} \Psi_W X_I, \quad (10)$$

$$Y_P = \Phi_P^{M_P \times N} \theta_P = \Phi_P^{M_P \times N} \Psi_W X_P.$$

The compression ratios are calculated as

$$R_I = \frac{M_I}{N} = \frac{d_I^v}{d_I^c}, \quad R_P = \frac{M_P}{N} = \frac{d_P^v}{d_P^c}. \quad (11)$$

3.2. Joint Recovery. Let us firstly recall the principles of BP algorithm in traditional channel decoding. BP algorithm approximately calculates the marginal distribution of all the variable nodes from the global distribution by passing messages iteratively along the edges of the bipartite graph. At the beginning of iterations, the variable nodes are initialized by the received codeword, but the check nodes have no external information. In contrast, the BP algorithm in CS [24] has no prior information for variable nodes, while it has external information for check nodes.

The proposed SI-BP algorithm is different from the aforementioned situations because both the variable nodes

and check nodes have prior information in SI-BP. The correlation between the side information frame and the current P frame to be decoded can be modeled as a virtual channel. So the side information θ_{SI} is used to initialize the variable nodes. And the measurements Y'_p are used to correct the errors caused by the correlation noise and the transmission noise.

For clarifying the iterative BP decoding algorithms, we give the following assumptions firstly.

- (i) Denote the variable nodes by $\theta_F^1, \theta_F^2, \dots, \theta_F^j, \dots, \theta_F^N$ and the check nodes by $y_F^1, y_F^2, \dots, y_F^k, \dots, y_F^{M_F}$, where $F \in \{I, P\}$. The side information θ_{SI} is denoted as $s_1, s_2, \dots, s_j, \dots, s_N$, and the measurements Y'_F is denoted as $m_F^1, m_F^2, \dots, m_F^k, \dots, m_F^{M_F}$.
- (ii) The messages sent on the edges of the bipartite graph are probability density functions (pdfs). As the pdf is real function over the interval $(-\infty, +\infty)$, so we represent the pdf with Q samples.
- (iii) $q_{j,k}(x)$ represents the message sent from variable node θ_F^j to check node y_F^k . $r_{k,j}(x)$ represents the message sent from check node y_F^k to variable node θ_F^j .
- (iv) $C(j)$ is the set of check nodes connected with the variable node θ_F^j . $R(k)$ is the set of variable nodes connected with the check node y_F^k .

The decoding algorithms for MG model and WHMT model are similar but different at the initialization. However, the WHMT model can be degenerated to MG model by setting all $\pi_{(s,i)}$ equal to K/N , where K is the sparsity of the frame. So we just discuss the WHMT model in what follows.

Each variable node of I frames is initialized as

$$q_j^{(0)}(\theta) = \pi_{(s,i)} \frac{1}{\sqrt{2\pi(\sigma_{s1})^2}} \exp\left(-\frac{\theta^2}{2(\sigma_{s1})^2}\right) + (1 - \pi_{(s,i)}) \frac{1}{\sqrt{2\pi(\sigma_{s0})^2}} \exp\left(-\frac{\theta^2}{2(\sigma_{s0})^2}\right), \quad (12)$$

which is send to the corresponding check nodes. While for the P frame, the variable node is initialized as

$$q_j^{(0)}(\theta) = \pi_{(s,i)} \frac{1}{\sqrt{2\pi(\sigma_{s1})^2}} \exp\left(-\frac{(\theta - s_j)^2}{2(\sigma_{s1})^2}\right) + (1 - \pi_{(s,i)}) \frac{1}{\sqrt{2\pi(\sigma_{s0})^2}} \exp\left(-\frac{(\theta - s_j)^2}{2(\sigma_{s0})^2}\right). \quad (13)$$

It can be found that the side information s_j is used as the prior mean of the variable node of P frame. This is determined by the maximum *a posteriori* probability (MAP) estimate. Then the other steps in recovery of I frames and P frames are the same, and we listed them below.

The check node y_F^k receives the messages sent from variable nodes in the previous half-iteration, and then calculates

the message to be transmitted back to variable nodes in following steps:

$$\begin{aligned} \tilde{r}_{k,k(l)}(\theta) &= w(\theta) * q_{k(1),k}(h_{k(1)}\theta) * q_{k(2),k}(h_{k(2)}\theta) * \dots \\ &\quad * q_{k(l-1),k}(h_{k(l-1)}\theta) * q_{k(l+1),k}(h_{k(l+1)}\theta) * \dots \\ &\quad * q_{k(d_F^c),k}(h_{k(d_F^c)}\theta) \\ r_{k,k(l)}(\theta) &= \tilde{r}_{k,k(l)}(-h_{k(l)}\theta - \Delta(m_F^k)), \end{aligned} \quad (14)$$

where $\theta_{k(l)} \in R(k)$ and $h_{k(l)}$ is the non-zero element in row k , $l = 1, 2, \dots, d_F^c$. $w(\theta)$ is the pdf of the channel noise. And $\Delta(m_F^k)$ represents a function of measurement m_F^k to adjust the message. The function $r_{k,k(l)}(\theta)$ is the final message sent from check node y_F^k to variable node $\theta_{k(l)}$.

The variable node θ_F^j receives the messages sent from connected check nodes, and calculates the message to be transmitted back to check node as

$$\begin{aligned} q_{j,j(t)}(\theta) &= q_j^{(0)}(\theta) \cdot r_{j(1),j}(\theta) \\ &\quad \cdot r_{j(2),j}(\theta) \cdot \dots \cdot r_{j(t-1),j}(\theta) \\ &\quad \cdot r_{j(t+1),j}(\theta) \cdot \dots \cdot r_{j(d_F^v),j}(\theta), \end{aligned} \quad (15)$$

where $y_F^{j(t)} \in C(j)$, $t = 1, 2, \dots, d_F^v$. The function $q_{j,j(t)}(\theta)$ is the message sent from θ_F^j to check node $y_F^{j(t)}$.

The iteration is repeated for the desired number of iterations. Finally, we get the pdf of each variable node as

$$p_j(\theta) = q_j^{(0)} \prod_{k \in C(j)} r_{k,j}(\theta). \quad (16)$$

The MAP estimate is used to determine the value of each variable node.

In the SI-BP algorithm, the side information θ_{SI} is used to initialize the pdfs of variable nodes, and affects the pdfs in each iteration. This initialization is more reasonable than using zeros uniformly when there is no side information. And the θ_{SI} 's deviation from θ_p is gradually corrected by the measurements Y'_p during the iterations.

4. Simulation Results and Analysis

For the simulations, we select the commonly used YUV 4:2:0 video sequences "Coastguard" and "Foreman" to test the performance of the DCVS algorithm, where the first one is in QCIF format and the last one is in CIF format. Here only the Y frames are processed. For the CIF sequence, a frame is divided into 32×32 blocks, and they are encoded individually. While for the QCIF sequence, the block's size is 16×16 . These settings are used to make good trade-off between SI-BP efficiency and complexity. The group of pictures (GOP) consists of one I frame and two P frames. The forward and backward I frames are reconstructed firstly to generate the side information by interpolation for decoding the P frames.

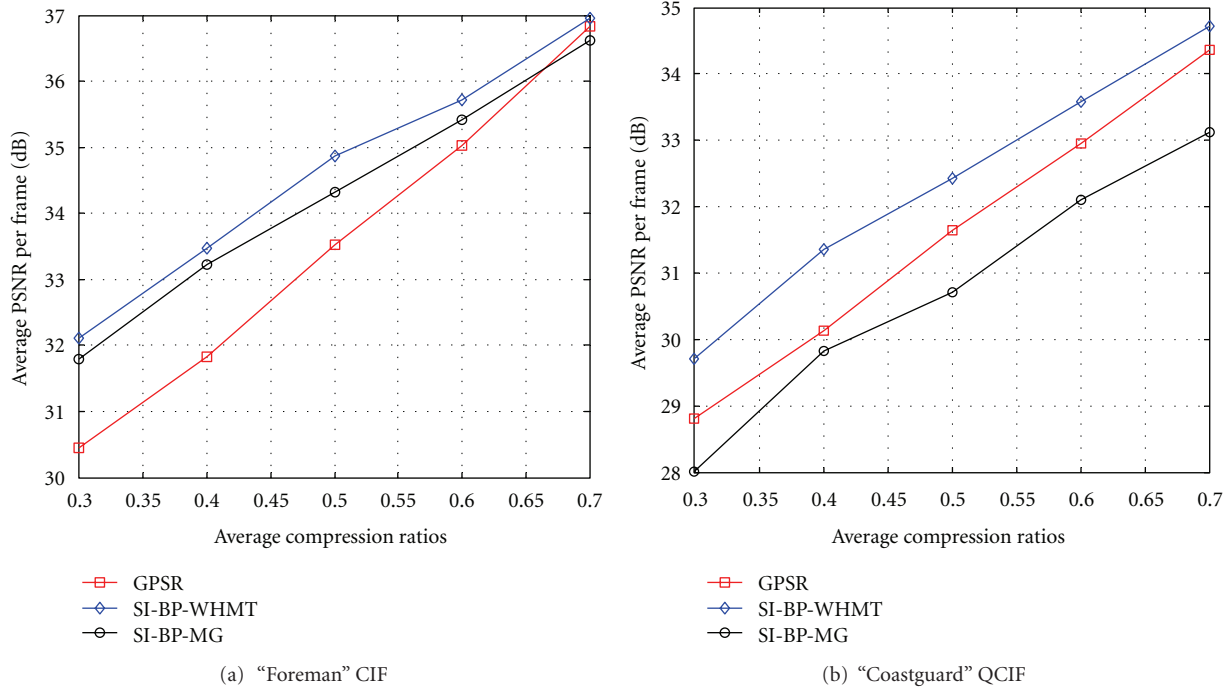


FIGURE 4: The PSNR performances for the: (a) "Foreman" CIF and (b) "Coastguard" QCIF sequences with the noiseless channel.

In order to evaluate the performance of SI-BP based recovery scheme compared with traditional convex optimization recovery schemes, we use the DCVS algorithm [23] using GPSR in [21] for comparisons with our scheme using SI-BP. The two-state MG model and WHMT model are both simulated in our scheme. For the MG model, the sparsity K , the standard deviation σ_1 and σ_0 are empirically decided. And the LDSM is designed with $d^c = d_l^c = d_p^c = 20$, and the variable node degrees are determined by the compression ratios. The WHMT model is trained before coding using the software in [30]. The average peak signal to noise ratios (PSNRs) at different average compression ratios R_{avg} is demonstrated to justify the performance, where $R_{\text{avg}} = (R_l + R_p \times 2)/3$. For example, $R_{\text{avg}} = 0.3$ is achieved by setting $R_l = 0.5$ and $R_p = 0.2$. When R_l and R_p both increase 0.1, the R_{avg} will increase 0.1 too.

4.1. The Noiseless Channel Case. Firstly, the ideal transmission channel is considered. The results are shown in Figure 4 and give the performance comparison between the proposed SI-BP schemes and the GPSR scheme of these two video sequences "Foreman" and "Coastguard", respectively. It can be found that the SI-BP scheme based on WHMT model performs better than the SI-BP scheme based on MG model and the GPSR scheme. This is because the WHMT model is specific for the video frame signal, and it just relies on the statistical properties of the signal which can be obtained by training. Besides, the unregular density distribution of LDSM protects the important part of the low-layer DWT coefficients, so it performs better than the MG model. For the MG model, its PSNR performance on Coastguard sequence is worse than that obtained from GPSR scheme, while for

the Foreman sequence, it is better than that of GPSR. This is due to that the MG model also considers both of the sparse and statistical properties. And the SI-BP recovery algorithm depends heavily on the accuracy of the signal model. We can infer that the sparse property plays a more important role in the noiseless channel case.

4.2. The Noisy Channel Case. And then the PSNR performances of these three schemes in the error-prone channel case are simulated, where the noise standard deviation σ_w of the AWGN channel is set to 15. As shown in Figure 5. For the "Foreman" CIF sequence, the SI-BP schemes on WHMT and MG models both outperform the GPSR scheme. The SI-BP scheme based on WHMT and MG model is superior to GPSR scheme by at least 6 dB. For the "Coastguard" QCIF sequence, the SI-BP scheme on MG still outperforms the GPSR scheme up to 2.5 dB. It can be found that the GPSR scheme almost keeps the same PSNR through all the compression ratios. The performance is not improved with increased measurements because the GPSR recovery algorithm has little ability of resisting the errors of measurements. In contrast, the noise does not degrade the performance of our proposed SI-BP scheme. And because of the unequal protection provided by the un-regular LDSM matrix, the performance of SI-BP based on WHMT is about 0.5 dB higher than that of the MG model.

Figure 6 gives additional demonstration of the excellent resiliency of the noisy measurements of the SI-BP scheme. The PSNR performances of the SI-BP scheme based on WHMT and MG models and the GPSR scheme are compared at the average compression ratio $R_{\text{avg}} = 0.4$ with the changing standard deviation of the channel noise. When the noise

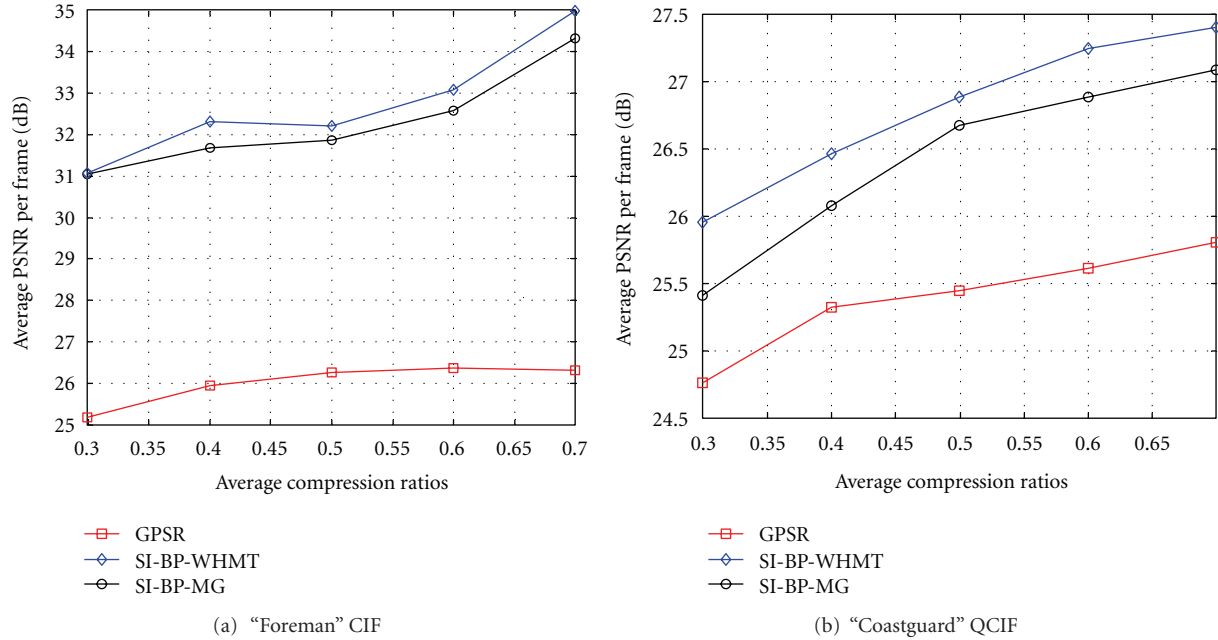


FIGURE 5: The PSNR performances for the: (a) "Foreman" CIF and (b) "Coastguard" QCIF sequences, where the transmission channel is AWGN channel with the noise standard deviation $\sigma_W = 15$.

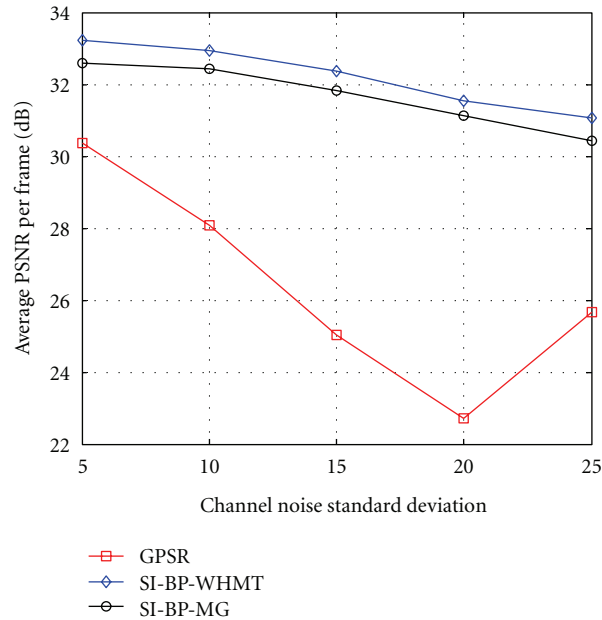


FIGURE 6: The PSNR performances for the "Foreman" CIF sequences with the average compression ratio $R_{avg} = 0.4$.

is very small, the GPSR scheme is considerable with our schemes, just as that in the noiseless case, but degrades rapidly with increased noise variance. While the proposed DCVS scheme using SI-BP, achieve the stable PSNR performance, whether the frame model is WHMT or MG. So we can come to the conclusion that the SI-BP based DCVS is more suitable for the practical applications as the channel noise is inevitable in wireless networks.

5. Conclusions and Future Works

In this paper, a novel DCVS scheme using side-information based belief propagation (SI-BP) is proposed to deal with the multiaccess model of video processing. Each video frame signal is compressed to its measurements separately, and the intra- and inter-correlations are utilized at the joint decoder. The SI-BP recovery algorithm is proposed based on Bayesian

inference, which is quite different from the optimization recovery schemes in prior CS work and is error-resilient when the measurements are transmitted through the noise-prone channels.

The proposed DCVS scheme shifts the complexity of video coding to the decoder and guarantees a light and efficient encoder for the constrained camera sensors. The decoding algorithm based on SI-BP is more suitable in practical applications where the transmission noise is inevitable. In the future, we will further improve the performance of SI-BP based DCVS in noiseless scenarios and expand the achievements of this paper in other video and image analysis tasks, such as motion tracking and so on.

Acknowledgments

This work is supported by National Science Foundation of China (no. 61201149), the 111 Project (no. B08004), and the Fundamental Research Funds for the Central Universities. This work is also supported (in part) by Korea Evaluation Institute of Industrial Technology (KEIT), under the R&D support program of Ministry of Knowledge Economy, Korea. The authors would like to thank all of the reviewers and editors for their detailed comments that have certainly improved the quality of their paper.

References

- [1] S. Takamura, "Distributed video coding: trends and future," *IPSI SIG Technical Reports*, vol. 2006, no. 102, pp. 71–76.
- [2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [4] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 2–10, 1974.
- [5] R. Puri and K. Ramchandran, "PRISM: a new robust video coding architecture based on distributed compression principles," Proceedings of the 40th Allerton Conference on Communication, Control, and Computing, Allerton, Ill, USA, October 2002.
- [6] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.
- [7] M. Panahpour Tehrani, T. Fujii, and M. Tanimoto, "The adaptive distributed source coding of multi-view images in camera sensor networks," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 10, pp. 2835–2843, 2005.
- [8] Q. Xu and Z. Xion, "Layered Wyner-Ziv video coding," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3791–3803, 2006.
- [9] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.
- [10] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [11] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.
- [12] D. L. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 1–53, 2009.
- [13] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [14] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [15] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic, San Diego, Calif, USA, 2nd edition, 1999.
- [16] M. B. Wakin, J. N. Laska, M. F. Duarte et al., "An architecture for compressive imaging," in *Proceedings of IEEE International Conference on Image Processing (ICIP '06)*, pp. 1273–1276, October 2006.
- [17] M. Akakaya, J. Park, and V. Tarokh, "A coding theory approach to noisy compressive sensing using low density frames," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5369–5379, 2011.
- [18] Y. Zhang, S. Mei, Q. Chen, and Z. Chen, "A novel image/video coding method based on compressed sensing theory," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 1361–1364, April 2008.
- [19] D. Venkatraman and A. Makur, "A compressive sensing approach to object-based surveillance video coding," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 3513–3516, April 2009.
- [20] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proceedings of IEEE International Conference on Image Processing (ICIP '09)*, pp. 1393–1396, November 2009.
- [21] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.
- [22] S. Xie, S. Rahardja, and Z. Li, "Wyner-Ziv image coding from random projections," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '07)*, pp. 136–139, July 2007.
- [23] L. W. Kang and C. S. Lu, "Distributed compressive video sensing," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 1169–1172, April 2009.
- [24] D. Baron, Sh. Sarvotham, and R. Baraniuk, "Bayesian compressed sensing via belief propagation," Technical Report TREE 0601, Rice ECE Department, 2006.
- [25] C. Weidmann and M. Vetterli, "Rate distortion behavior of sparse sources," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 4969–4992, 2012.
- [26] F. Wu, J. Fu, Z. Lin, and B. Zeng, "Analysis on rate-distortion performance of compressive sensing for binary sparse source," in *Proceedings of Data Compression Conference (DCC '09)*, pp. 113–122, March 2009.
- [27] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 492–504, 2010.

- [28] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Wavelet-domain compressive signal reconstruction using a hidden Markov tree model," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 5137–5140, April 2008.
- [29] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3488–3497, 2009.
- [30] H. Choi and J. Romberg, "Wavelet-domain hidden Markov models," DSP Group, Rice University, 2009, <http://dsp.rice.edu/software/wavelet-domain-hidden-markov-models>.