Reconstruction for Distributed Video Coding: A Context-Adaptive Markov Random Field Approach

Yongsheng Zhang, *Member, IEEE,* Hongkai Xiong, *Senior Member, IEEE,* Zhihai He, *Senior Member, IEEE,* Songyu Yu, and Chang Wen Chen, *Fellow, IEEE*

***Abstract*—Within the existing reconstruction process of distributed video coding (DVC), there are two major approaches: the maximum probability reconstruction and the minimum mean square error (MMSE) reconstruction. Both of them assume that each node, a pixel in pixel domain DVC or a coefficient in transform domain DVC, is i.i.d., and reconstruct the value of each node independently by only exploiting statistical correlation between source and side-information. These kinds of models produce considerable amount of artifacts in decoded Wyner–Ziv (WZ) frames and degrade the objective performance. In this paper, we propose a context-adaptive Markov random field (MRF) reconstruction algorithm which exploits both the statistical correlation and the spatio-temporal consistency by modeling the corresponding MRF of a generic DVC architecture, and solve the inference by finding its MRF-based maximum a posteriori (MAP) estimate. The *energy function* of the MRF model consists of two terms: a *data term* measuring the statistical correlation, and a *geometric regularity term* enforcing local spatio-temporal structure consistency which is modeled by optical flow estimation with regard to the critical parameters under a wide variety of DVC scenarios. In case the unreliability of the derived local structure, a confidence parameter is introduced to prevent inappropriate penalizing. To find the reconstructed patch assignment with the largest expected probability in the context-adaptive MRF, the energy minimization for the MRF-based MAP estimate of the WZ frames is solved by global optimization and greedy strategies. Compared to the existing maximum probability and MMSE reconstruction with i.i.d. model, a better subjective and objective performance is validated by extensive experiments.**

***Index Terms*—Distributed video coding (DVC), Markov random fields (MRFs), maximum *a posteriori*, reconstruction, Wyner–Ziv coding.**

I. Introduction

DISTRIBUTED video coding (DVC), also known as Wyner–Ziv (WZ) video coding, acts as a new video coding paradigm motivated by the requirements of recent emerging applications, e.g., mobile camera phone and wireless visual sensor networks, which desire low encoding complexity due to the battery life constraint [1], [2]. It avoids the computationally intensive temporal prediction loop at the encoder by shifting the exploitation of the temporal redundancy to the decoder. Although Wyner and Ziv have theoretically proved that WZ coding suffers no performance loss for Gaussian memoryless sources under mean-squared error (MSE) distortion metric [3], practical DVC schemes often have a considerable performance loss compared with traditional predictive coding engines, e.g., H.264/AVC. To improve the rate-distortion performance, researchers have extensively developed a variety of approaches which are mainly composed of constructing more accurate side information [4], adopting decorrelation transform to exploit spatial correlation [5], using advanced channel codes, e.g., Turbo [6] and low-density parity-check (LDPC) codes [7], and designing advanced reconstruction algorithms [8], [9]. In this paper, we will devote to the DVC reconstruction problem by formulating it as a Markov random field (MRF) model to exploit the intrinsic geometric regularity (GR) constraint of WZ frames.

So far, there are two major reconstruction algorithms in DVC applications: the maximum probability [10] and the minimum mean-square error (MMSE) [8], [11], [12]. In maximum probability reconstruction algorithm, the reconstruction value is chosen in the quantization bin with the maximum conditional probability for the given side-information. The MMSE algorithm chooses the value so as to minimize the MSE of the reconstructed frame. Recently, *Roca et al.* proposed an adaptive reconstruction algorithm [9] which adopts a spatial non-stationary correlation noise model instead of devising a new reconstruction algorithm. Overall, all these algorithms adopt i.i.d. models for both the WZ frame and the correlation noise. For the i.i.d. model, the MMSE algorithm is optimal in the sense of MSE distortion metric.

However, a natural image has high spatial correlation between neighboring pixels [13]. Although decorrelation transforms, e.g., DCT, are usually introduced to eliminate spatial redundancy [5], GR still exists between neighboring coefficients in low frequency coefficient bands. Coefficient bands are formed by grouping together the DCT coefficients of the entire Wyner–Ziv frame, according to the position occupied by each DCT coefficient. In this sense, the i.i.d. model cannot characterize the property of natural images well

Manuscript received September 1, 2010; revised January 2, 2011; accepted February 15, 2011. Date of publication March 28, 2011; date of current version August 3, 2011. This work was supported in part by the National Natural Science Foundation of China, under Grants 60772099, 60928003, 60736043, and 60632040, and by the Program for New Century Excellent Talents in University, under Grant NCET-09-0554. This paper was recommended by Associate Editor M. Hannuksela.

Y. Zhang, H. Xiong, and S. Yu are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yongsheng.zhang.cn@gmail.com; xionghongkai@sjtu.edu.cn; syyu@sjtu.edu.cn).

Z. He is with the Department of Electrical and Computer Engineering, University of Missouri, St. Louis, MO 65211 USA (e-mail: hezhi@missouri.edu).

C. W. Chen is with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: chencw@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2133830

and degrades the objective performance. Furthermore, it also induces annoying artifacts: contouring and blocking artifacts for pixel domain and transform domain DVC, respectively. Martins *et al.* [14] also noticed the blocking artifacts and proposed an adaptive deblocking filter to smooth the artifacts according to the boundary strength of blocks. To improve both the subjective and objective performances of reconstruction, a new model is desired to exploit not only the statistical correlation between the WZ frame and the side-information but also the GR constraint between neighboring pixels and coefficients of the WZ frame [15].

In this paper, the WZ frame is modeled as an MRF which reflects interactions among neighboring nodes on a 2-D lattice for natural video, and its maximum *a posteriori* (MAP) estimate is solved for the desired reconstruction to achieve better subjective and objective performance than the i.i.d. model of the existing maximum probability and the MMSE algorithms. The major contributions of this paper are: 1) to formulate the reconstruction process of a generalized DVC scheme as an MRF to exploit the spatio-temporal consistency of video sequences along with the statistical correlation between the WZ-frame and side information, and 2) to propose an *energy function* with context-adaptive spatio-temporal constraint to exploit local structures of video sequences. All parameters in the *energy function* are derived by analyzing the temporal consistency property of successive frames. With the *energy function*, the energy minimization for the MRF-based MAP estimate of the WZ frames is solved by global optimization, the sequential tree reweighted message passing (TRW-S) algorithm [16], and greedy strategies, the iterated conditional modes (ICM) algorithm [17]. Compared to the existing maximum probability and the MMSE reconstruction with i.i.d. model, a better subjective and objective performance is validated by extensive experiments.

The rest of this paper is organized as follows. The existing DVC schemes are briefly summarized in Section II. In Section III, we analyze the existing reconstruction algorithms and formulate the reconstruction problem by an MRF model. Implementation details are presented in Section IV. Extensive experimental results are shown in Section V with both objective and visual metrics. Section VI concludes this paper.

II. Overview of DVC Architectures

Existing practical DVC architectures are composed of pixel-domain DVC (PD-DVC) [18]–[20],[1] transform-domain DVC (TD-DVC) [5], and DVC-related applications without feedback channel [21]–[23]. PD-DVC is the simplest DVC scheme with less encoding complexity and worse RD performance than TD-DVC since no decorrelation transform is used to eliminate spatial redundancy.

[1]To increase performance and preserve the low complexity property of PD-DVC, MRF-based algorithms have been introduced to exploit spatial correlations in the Slepian–Wolf decoding process [18]–[20]. These approaches attempt to reduce the required bit-rate encoding a bit-plane, while the proposed algorithm in this paper is dedicated to increasing the reconstruction quality of Wyner–Ziv frames after the bit-planes have been decoded.

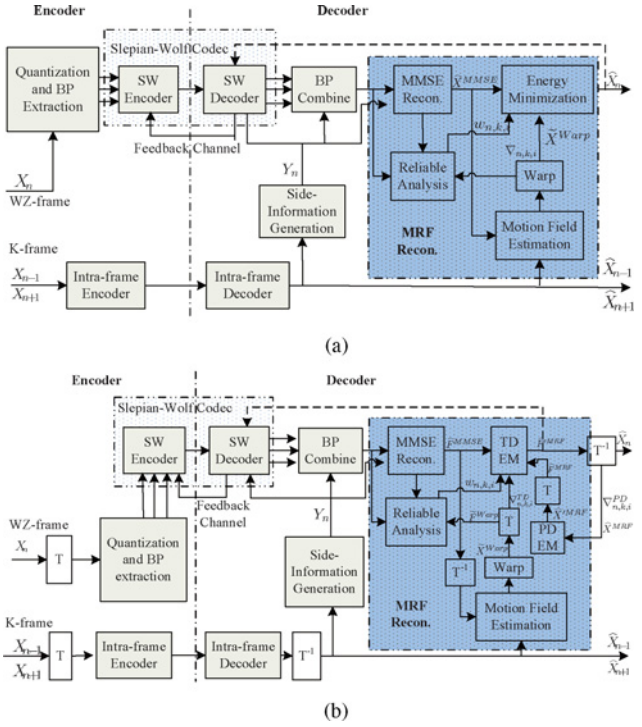


Fig. 1. DVC codec architecture with the proposed context-adaptive MRF-based reconstruction algorithm. (a) PD-DVC. (b) TD-DVC.

A. Pixel-Domain DVC (PD-DVC)

In the PD-DVC scheme, a subset of frames, termed as "Key frames," are encoded and decoded using conventional intra-frame encoding schemes, e.g., H.264/AVC Intra coding. The rest frames, called "WZ frames," are intra-frame encoded but inter-frame decoded. Fig. 1(a) presents the corresponding PD-DVC codec architecture with the proposed MRF-based reconstruction algorithm.

For a WZ frame X_n, each pixel $x_{n,i}$ is uniformly quantized into M bits. The bits of all pixels in a WZ frame would form M bit-planes which are fed into the Slepian–Wolf (SW) encoder to generate the WZ bit-stream. The low-density parity-check accumulate (LDPCA) approach [24], [25] with degree distribution in (14), is used as SW codec.

At the decoder side, side-information Y_n is generated by bi-directional motion compensation of previously decoded Key frames [26]. Specifically, forward motion estimation is adopted to get the candidate motion vectors for each non-overlapped block in the Wyner–Ziv frame. From the available candidate vectors, the motion vector that intercepts the Wyner–Ziv frame closer to the center of the block is taken as the final motion vector, and bidirectional motion compensation is performed to produce the side-information.

The statistical correlation between the generated side-information and current WZ frame is modeled at the decoder to compute the log-likelihood of each bit. Together with received WZ bit-stream, the log-likelihood is fed into the SW decoder to recover the original bit-plane. If it fails, the decoder would request additional bits from the encoder's buffer through feedback channel (FBC). The correctly decoded bit-planes are combined to determine the quantization bin $[B_L, B_U]$ of each

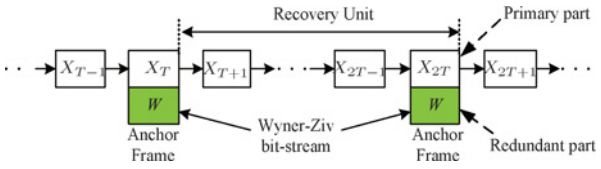


Fig. 2. Bit-stream structure of the error-resilient video coding scheme for TD-DVC without feedback channel [31].

pixel $x_{n,i}$ in the WZ frame. Finally, reconstruction process would derive a best estimate of the original pixel value $x_{n,i}$ as the output of the DVC decoder. If a refinement decoding process is activated, the best estimate will in turn be used to assist the SW decoding of less significant bit-planes [27].

In the proposed reconstruction algorithm, MMSE reconstruction $\widehat{X}^{\text{MMSE}}$ is firstly derived as an initial estimate. A good initial estimate can accelerate the following energy minimization process and prevent falling into a local minimum. With the MMSE estimate, the warping-based motion compensation $\widetilde{X}^{\text{Warp}}$ is adopted to estimate the local GR constraints $\nabla_{n,k,i}$ through an optical flow based method [28] between the WZ frame and the adjacent Key frames. Considering some regions are unreliable because of occlusion or abrupt motion, a weighting parameter $w_{n,k,i}$ is derived based on the reliability of the warping results. With the local GR constraint $\nabla_{n,k,i}$, weighting parameter $w_{n,k,i}$, and the quantization bin $[B_L, B_U]$, energy minimization is solved to find an optimized estimate of the WZ frame $\widehat{X}^{MRF}$.

B. Transform-Domain DVC (TD-DVC)

Decorrelation transform, e.g., DCT, is widely used in DVC applications to eliminate the spatial redundancy. A corresponding TD-DVC scheme with the proposed reconstruction algorithm is shown in Fig. 1(b). At the encoder side, a Wyner–Ziv frame is partitioned into non-overlap 4×4 blocks, and each block is transformed into DCT domain with 4×4 DCT. The generated coefficients at the same position are grouped to form a coefficient band, e.g., DC band is composed of the DC coefficients of all the 4×4 blocks. The DC band is quantized with uniform scalar quantizer and AC bands are quantized with scalar quantizer with a deadzone around zero [5]. The quantization step Q_s is derived from a quantization parameter for WZ coding (QPW). The QPW range is 4–51, and QPW value 4 corresponds to Q_s value 1. Like the standard H.264/AVC quantization process [29], Q_s doubles in size for every increment of 6 in QPW, and increases by 12.5% for each increment of 1 in QPW. After a DCT band is quantized with quantization step Q_s, the quantized coefficients in the band are extracted to $\lceil \log_2(V_{\max}) \rceil$ bit-planes for DC band, or $\lceil \log_2(V_{\max}) \rceil + 1$ bit-planes for AC band where the extra bit-plane is the sign, $V_{\max}$ is the maximum absolute value of quantized coefficients in the band, and $\lceil \cdot \rceil$ is the ceiling function. In this configuration, all bit-planes extracted from quantized coefficients are fed into SW encoder, and induce no distortion. At the decoder side, the WZ-frame is decoded on the basis of the derived side-information and statistical correlation model. With the correctly decoded bit-planes and side information, the reconstruction module is activated to produce an optimized estimate of the original coefficients. Finally, 4×4 IDCT operation is implemented to generate the decoded WZ frame.

In the proposed MRF-based reconstruction, MMSE reconstruction is likewise implemented to acquire an initial estimate $\widehat{F}^{\text{MMSE}}$. In turn, $\widehat{F}^{\text{MMSE}}$ is transformed into pixel domain with IDCT to derive the optical flow field between the WZ frame X_n and the adjacent Key frames X_{n-1} or X_{n+1}, and then produce the warping image $\widetilde{X}^{\text{Warp}}$ to estimate the *GR constraint* $\nabla_{n,k,i}$. The weighting parameter $w_{n,k,i}$ in transform domain is derived based on the coefficients reliability of the warping image, and is fed into the "transform-domain energy minimization" (TDEM) model with the quantization bin $[B_L, B_U]$. The output $\widehat{F}'^{MRF}$ is transformed into the pixel domain to eliminate the blocking artifact along with the "pixel-domain energy minimization" (PDEM) module, and the output $\widehat{X}'^{MRF}$ would be transformed into DCT domain for the next iteration. The reconstruction process terminates till the iteration limit is reached or the energy minimization converges.

C. TD-DVC With Side-Information Refinement (SIR)

In a standard DVC, the side-information is generated only once before SW decoding and unchanged in the entire decoding process. To generate a better side-information and improve the R-D performance, side-information refinement (SIR) approaches have been proposed [12], [27] which iteratively refine the side-information along the decoding of DCT bands or bit-planes instead of keeping it unchanged as the standard DVC.

In this paper, we also test the proposed reconstruction algorithms for TD-DVC with the SIR algorithm [12], where the side-information is gradually improved as the decoding proceeds: The DC coefficient band is primarily decoded, reconstructed and transformed into pixel domain with IDCT; with the current reconstructed frame from the already decoded coefficient bands and the current side-information, the blocks for SIR would be determined on basis of their difference; these selected blocks will search for their better side information candidates within a given window; the refined side information will be generated by weighted averaging over the candidate side information blocks, and be used for the decoding of the next coefficient band. The proposed MRF-based reconstruction for TD-DVC with SIR is similar to the traditional TD-DVC with the feedback channel, except that the side-information is gradually enhanced along the WZ decoding process. It is worth mentioning that the proposed algorithm itself is not an SIR scheme. The involved optical flow warping is devoted to estimating the local gradient of the WZ frame instead of generating a better estimate of the WZ frame.

D. TD-DVC Without Feedback Channel (FBC)

To allocate the proper number of bits, most DVC schemes resort to the feedback channel. In most practical scenes, a feedback channel does not exist or the end-to-end delay is too large to realize real-time rate allocation [21]–[23]. To deal with the SW bit-plane decoding failure without feedback channel, there exist two kinds of approaches: 1) use only the correctly decoded bit-planes for reconstruction and discard all

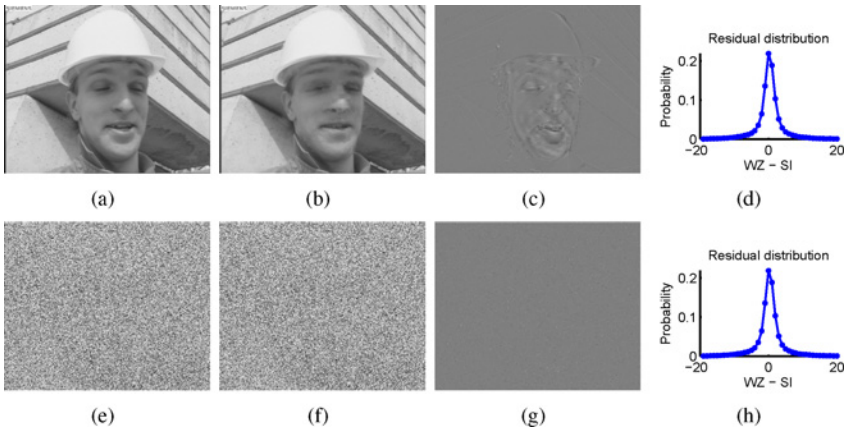


Fig. 3. Illustrative example of a WZ frame and the derived side-information. (a) Original WZ frame of the *Foreman* sequence. (b) Derived side-information. (c) Difference between (a) and (b). (d) Probability distribution of (c). (e)–(g) Random permutation of (a)–(c). (h) Probability distribution of (g).

the subsequent bit-planes of the coefficient band if one bit-plane fails to decode [12], and 2) use all bit-planes in the reconstruction process [30]. In fact, the reconstruction process is still similar to a conventional DVC reconstruction except the quantization bin $[B_L, B_U]$ varies.

In this paper, we adopt an error resilient video coding scheme to represent the TD-DVC architecture without the feedback channel [31]. As shown in Fig. 2, a set of frames, called anchor frames, are protected by a WZ bit-stream to stop temporal error propagation, and the state-of-the-art H.264/AVC is used to generate the primary bit-stream. The WZ redundant bit-stream will be discarded if the anchor frame is error free; otherwise, it will be decoded to correct errors and terminate error propagation. To allocate appropriate bit-rate for each bit-plane, the conditional entropy of each bit-plane is firstly estimated given previous decoded bit-planes and the transmission distortion, and the encoding rate is then determined together with the error correction capacity of the LDPCA codec and the predefined bit-plane decoding failure probability [32]. The proposed reconstruction algorithm will be validated under bit-plane decoding failure conditions.

III. Problem Formulation

The target of the reconstruction algorithm in a generic DVC system is to minimize the distortion based on the decoded bit-planes and the priors of video sequence. The motivation of the proposed algorithms is addressed, and we formulate the DVC reconstruction process as an MRF problem.

A. The Maximum Probability and MMSE Approaches

At the decoder side, the decoded bit-planes restrict the reconstructed value $\hat{x}_i$ in quantization bin $[B_L, B_U]$, where $B_U = B_L + Q_s - 1$ and Q_s is the quantization step. The maximum probability reconstruction of x_i is calculated with the mode of conditional probabilistic distribution $f_{X|Y}(x_i|y_i)$ in quantization bin $[B_L, B_U]$ [10]

$$\hat{x}_i^{MP} = \arg \max_{x_i \in [B_L, B_U]} f_{X|Y}(x_i|y_i) \tag{1}$$

where $f_{X|Y}(x_i|y_i)$ is the conditional probability distribution of x_i given the side-information y_i.

In comparison, the MMSE reconstruction aims to minimize the MSE of the output by computing the conditional expectation of variable x_i given the quantization bin $x_i \in [B_L, B_U]$ and side-information y_i [8]

$$\begin{aligned} \hat{x}_i^{\text{MMSE}} &= E[x_i | x_i \in [B_L, B_U], y_i] \\ &= \frac{\sum_{x_i=B_L}^{B_U} x_i f_{X|Y}(x_i|y_i)}{\sum_{x_i=B_L}^{B_U} f_{X|Y}(x_i|y_i)}. \end{aligned} \tag{2}$$

B. Motivation

As shown in (1) and (2), both maximum probability and MMSE reconstruction algorithms are based on an i.i.d. model which does not reflect interactions between neighboring nodes, e.g., pixels for PD-DVC or coefficients for TD-DVC. Fig. 3 describes an illustrative example, where Fig. 3(e)–(g) displays a random permutation of the corresponding images in Fig. 3(a)–(c), respectively. Despite being visually different, they have identical statistical distribution of the correlation noise and the source. In this sense, the i.i.d. model cannot benefit from the geometric smoothness property of natural images. In addition, the MMSE reconstruction results also suffer annoying artifacts, as depicted in Fig. 4, where contouring artifacts for the PD-DVC schemes and blocking artifacts for the TD-DVC schemes appear in high motion regions.

To improve the subjective and objective quality of reconstruction algorithms, we model a WZ frame as an MRF to exploit the intrinsic spatio-temporal consistency [33]–[35], in addition to the statistical correlation for reconstruction.

To intuitively illustrate the proposed MRF-based reconstruction algorithm, we show the reconstruction results in Fig. 5 with and without inter-node constraint for part of one row in a frame. In this example, the difference between side-information and source is approximated by Laplacian distribution, and the target gradient between neighboring nodes is available. According to the maximum probability algorithm, the reconstruction result is the one in the

(a) (b)

Fig. 4. Artifacts in the MMSE reconstruction (the 14th frame of the *Foreman* sequence). (a) Contouring artifacts for PD-DVC. (b) Blocking artifacts for TD-DVC.

quantization bin $[B_L, B_U]$ with the maximum conditional probability $f_{X|Y}(x_i|y_i)$ given y_i, while the MMSE reconstruction minimizes the expected MSE distortion for the given $f_{X|Y}(x_i|y_i)$ and y_i. In fact, the MMSE reconstruction is the centroid of the conditional probability $f_{X|Y}(x_i|y_i)$ and is closer to the center of quantization bin than the maximum probability reconstruction. Without exploiting the inter-node constraint, both the maximum probability algorithm and the MMSE algorithm suffer from the i.i.d. model and result in artifacts. For example, nodes 7 and 8 have identical reconstruction value with either the maximum probability or the MMSE reconstruction algorithm since both of them have the same conditional probability distribution in the quantization bin. In this way, the difference between nodes 8 and 9 increases and results in artifacts. However, the proposed methods with inter-node constraint would penalize the inter-node difference, and thus improve both the subjective and objective performance.

C. The MRF Model for DVC Reconstruction

If an image is modeled as an MRF, the overall probability of each assignment $\mathbf{x}$ can be written as the product of potential functions $\psi_C(\mathbf{x}_C)$ over cliques of the graph [36]

$$p(\mathbf{x}) = \frac{1}{Z}\prod_C \psi_C(\mathbf{x}_C) \tag{3}$$

where $Z = \sum_{\mathbf{x}}\prod_C \psi_C(\mathbf{x}_C)$ is a normalization constant, and $\mathbf{x}_C$ is a set of variables in clique C. Assuming that the potential functions $\psi_C(\mathbf{x}_C)$ are strictly positive, it is convenient to express them in exponential form. Accordingly, (3) turns out to be a Gibbs random field

$$p(\mathbf{x}) = \frac{1}{Z}\prod_C e^{-E_C(\mathbf{x}_C)} = \frac{1}{Z}e^{-\sum_C E_C(\mathbf{x}_C)} \tag{4}$$

where $E_C(\mathbf{x}_C)$ is called an *energy function* for clique C. Therefore, finding an MAP estimate of the MRF is equivalent to minimize the *energy function* $E(\mathbf{x}) = \sum_C E_C(\mathbf{x}_C)$, that is

$$\hat{\mathbf{x}}^{MRF} = \arg\max_{\mathbf{x}} \; p(\mathbf{x}) = \arg\min_{\mathbf{x}} \sum_C E_C(\mathbf{x}_C). \tag{5}$$

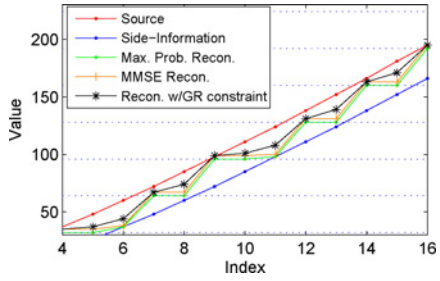


Fig. 5. Demonstrative example to show how the inter-node constraint can improve the reconstruction performance.

1) *Constraints in DVC Reconstruction:* To formulate the *energy function* $E(\mathbf{x})$, we need to analyze the prior constraints that can be used in the reconstruction process at the DVC decoder.

a) The bit-planes of the WZ frame are sequentially decoded, as shown in Fig. 6(a). The decoded bit-planes define the quantization bin $[B_L, B_U]$, e.g., the possible range of each $x_{n,i}$ in the WZ frame, as shown in Fig. 6(b).
b) $f_{X|Y}(x_{n,i}|y_{n,i})$, the statistical correlation between the WZ frame and the side-information.
c) The spatio-temporal consistency constraint, as shown in Fig. 6(b) and (c).

These constraints are graphically depicted in Fig. 7, where $x_{n,i}$ is the desired node value, $y_{n,i}$ is the value of the side-information, $x_{n,k,i}$ is a neighbor of $x_{n,i}$ in direction k, and $\nabla_{n,k,i}$ is the gradient between $x_{n,i}$ and its neighbor in direction k. In this paper, we regard a frame as an 8-connected regular lattice, and ignore second and higher order cliques for simplicity. The underlying nodes consist of two types of cliques: clique $\{x_{n,i}, y_{n,i}\}$ and cliques $\{x_{n,i}, x_{n,j}\}$. The clique $\{x_{n,i}, y_{n,i}\}$, in Fig. 7(b), is denoted the *data term*, which uses the *energy function* $E_i(\hat{x}_{n,i})$ to measure the cost of assigning label $\hat{x}_{n,i}$ to pixel $x_{n,i}$ given side-information $y_{n,i}$, the quantization bin $[B_L, B_U]$ and statistical correlation $f_{X|Y}(x_{n,i}|y_{n,i})$. The cliques $\{x_{n,i}, x_{n,j}\}$ in Fig. 7(c) are called the *GR term*, which measures the spatio-temporal consistency constraint with *energy function* $E_{i,j}(\hat{x}_{n,i}, \hat{x}_{n,j})$ to penalize the cost of assigning labels $\hat{x}_{n,i}$ and $\hat{x}_{n,j}$ to a pair of neighboring nodes, where $j \in \mathcal{N}(i)$ indicates the index of neighbors of node i. In consequence, the *energy function* $E(\hat{\mathbf{x}})$ can be written as

$$E(\hat{\mathbf{x}}) = \sum_i E_i(\hat{x}_{n,i}) + \eta \sum_{i,j\in\mathcal{N}} E_{i,j}(\hat{x}_{n,i}, \hat{x}_{n,j}) \tag{6}$$

where parameter η regularizes the relative ratio between the *data term* and the *GR term*. Obviously, the MMSE reconstruction acts as a special independent case of the proposed reconstruction model when $E_{i,j}(\hat{x}_{n,i}, \hat{x}_{n,j})$ is set to 0 and $E_i(\hat{x}_{n,i})$ adopts MSE metric.

2) *Data Term:* The *data term* is measured with the expected MSE distortion of reconstruction $\hat{x}_{n,i}$

$$\begin{aligned} E_i(\hat{x}_{n,i}) &= E[(\hat{x}_{n,i} - x_{n,i})^2] \\ &= \sum_{x_{n,i}} f_{X|Y}(x_{n,i}|y_{n,i}) \cdot (\hat{x}_{n,i} - x_{n,i})^2 \end{aligned} \tag{7}$$

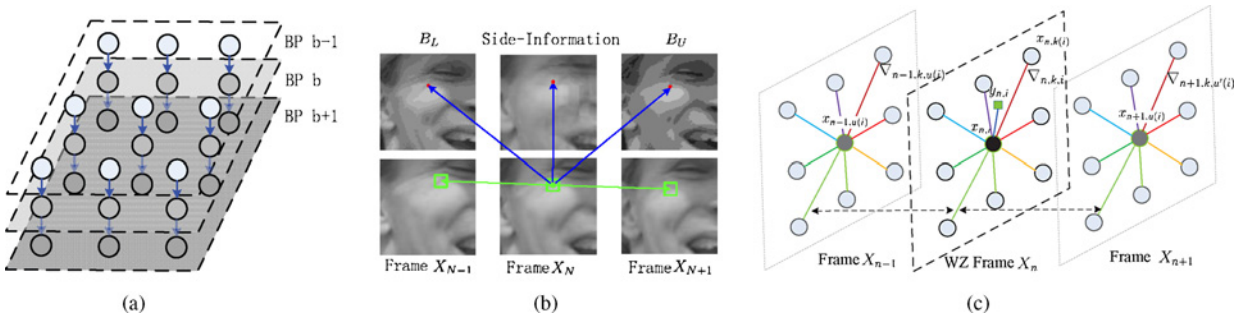


Fig. 6. Constraint information that can be used in the DVC reconstruction process, and the graphical representation of the spatio-temporal constraint. (a) Bit-plane constraint. (b) Spatio-temporal consistency and statistical correlation. (c) Graphical representation of the spatio-temporal correlation.

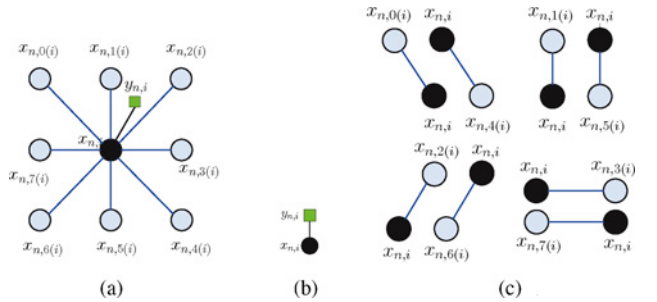


Fig. 7. Graphical representation of the MRF model for DVC reconstruction. (a) MRF model. (b) Clique for *data term*. (c) Cliques for *GR term*.

where $f_{X|Y}(x_{n,i}|y_{n,i})$ is the conditional probability of $x_{n,i}$ given side-information $y_{n,i}$, and $\hat{x}_{n,i}$ is the reconstruction value.

In the TD-DVC scheme, the quantization bin $[B_L, B_U]$ is only available in DCT domain and we define the corresponding *data term* for pixel-domain optimization as

$$E_i(\hat{x}_{n,i}) = (\hat{x}_{n,i} - x'_{n,i})^2 \tag{8}$$

where $x'_{n,i}$ is the initial value for the pixel domain optimization.

3) *GR Term:* The choice of the GR term is critical, and there are different functions to meet the requirements of spatial constrained priors [36]. In this paper, a weighted quadratic function is adopted to impose the spatio-temporal consistency constraint

$$\begin{aligned} E_{i,j}(\hat{x}_{n,i}, \hat{x}_{n,k(i)}) &= w_{n,k,i} V(\hat{x}_{n,i}, \hat{x}_{n,k(i)}) \\ &= \sum_{s=\pm 1} w_{n+s,k,i} (|\hat{x}_{n,i} - \hat{x}_{n,k(i)}| - |\nabla_{n+s,k,i}|)^2 \end{aligned} \tag{9}$$

where $k(i)$ is the index of the neighbor of node $x_{n,i}$ in direction k, and $\nabla_{n,k,i}$ is the target gradient of $x_{n,i}$ in direction k which measures the spatio-temporal consistency of video sequence. Since the actual value of $\nabla_{n,k,i}$ is unavailable, it is approximated by $\nabla_{n-1,k,u(i)}$ (and $\nabla_{n+1,k,u(i)}$ if available) which is the gradient of pixel $x_{n-1,u(i)}$ in the reference frame X_{n-1} in direction k. $w_{n\pm1,k,i} \geq 0$, and $\sum_{s=\pm1} w_{n+s,k,i} = 1$, is the weighting factor, which indicates the reliability of derived target gradient $\nabla_{n,k,i}$. Specifically, if both the warping results $\tilde{x}_{n-1,u(i)}$ and $\tilde{x}_{n-1,k(u(i))}$ are in the quantization bin $[B_L, B_U]$ of $x_{n,i}$ and $x_{n,k(i)}$, the $w_{n-1,k,i}$ is set 1, and otherwise set to 0. If both $\nabla_{n-1,k,u(i)}$ and $\nabla_{n+1,k,u(i)}$ are available, the $w_{n-1,k,i}$ and $w_{n+1,k,i}$ are set to 0.5.

D. Energy Minimization Algorithms

With the derived *energy function*, energy minimization can be used to find the MAP solution on the joint probability of the observed variables (the known pixels) and the hidden variables (the possible assignment with a probability). It aims to find the marginal distribution for all the variable nodes in the MRF in order to achieve the maximum of the joint distribution, i.e., the most likely assignment for each node in the MRF. The well-known energy minimization algorithms are graph cuts [37] and message passing algorithm [38]. In view of the fact that the *GR term* in (9) is not regular [39], we adopt the TRW-S algorithm to approach the global minimum [16]. To reduce the decoder's complexity, the ICM [17] algorithm is also evaluated.

1) *Sequential Tree-Reweighted Message Passing (TRW-S):* TRW-S is a message-passing algorithm similar to loopy belief propagation [40], but it computes a lower bound on the energy. TRW-S can achieve similar energy with the graph cuts and BP algorithms, and can be applied to any function of the form as (6) [16].

2) *Iterated Conditional Mode (ICM):* The ICM algorithm uses a deterministic greedy strategy to find a local minimum [17]. It starts with an initial estimate of the desired node value, and chooses the value which could give the largest decrease of the *energy function* for each node. This process is repeated till convergence or the iteration limit is reached.

Although ICM has low computational complexity, it is very sensitive to the initial estimate and is easily been trapped in a local minimum. Since the MMSE estimator can play as a good initial estimate of the WZ frame [8], it helps to find a solution close to the global minimum while preserving the low complexity advantage.

3) *Accelerate the Optimization Process:* To accelerate the convergence speed and reduce the computational complexity, the structured sparsity of natural images can be used. Specifically, with the decoded bit-planes, we can analyze the reliability of side-information in different regions of the WZ-frame by verifying the distance between side-information and corresponding quantization bin. With this reliability information, we can change the message passing order for TRW-S or the searching order for ICM to use reliable message first.

IV. Implementation Details

In the defined *energy function*, there are three parameters that need to be determined: target gradient $\nabla_{n,k,i}$, weighting parameter $w_{n,k,i}$, and regularization factor η. The target gradient $\nabla_{n,k,i}$ reflects the spatial correlations between neighboring nodes. However, it is unavailable at the DVC decoder side since the original WZ frame is inaccessible. Therefore, we resort to the temporal consistency property of video sequences by compensating the relative motion between successive frames. Considering that the derived motion field maybe not very accurate and/or occlusion may exist, the reliability of the derived $\nabla_{n,k,i}$ is measured by the weighting parameter $w_{n,k,i}$. The relative ratio η between the *data term* and the *GR term* should be adjusted for different quantization steps to prevent over-smoothing. In what follows, we will discuss the spatio-temporal consistency for different DVC scenarios, and analyze the impact of parameter η on the reconstruction performance.

A. Spatio-Temporal Consistency: Target Gradient $\nabla_{n,k,i}$

1) *Temporal Consistency Analysis:* To find the target gradient $\nabla_{n,k,i}$, the relative displacement between successive video frames should be derived and compensated. During the estimation of the motion field, the following constraints are taken into account.

a) *Gray value constancy.* In optical flow estimation, it is assumed that the gray value of pixel is consistent with small variation along the displacement, i.e., $x_{n,i} \cong x_{n-1,u(i)}$, where $u(i)$ is the displacement vector between frame X_n and frame X_{n-1}.

b) *Gradient constancy.* The gray value constancy assumption is sensitive to slight changes in brightness, and it is useful to determine the displacement vector by the gradient constancy assumption [41], $\nabla_{n,k,i} = \nabla_{n-1,k,u(i)}$.

c) *Motion smoothness.* Considering that the motion of a homogeneous contour is locally ambiguous (*aperture problem*), a smoothness assumption of the flow field is introduced.

d) *Multiscale.* To prevent being trapped in a local minimum, it is useful in a multiscale sense: from a coarse, downsampled version to the refined version of the problem.

Accordingly, it is reasonable to define a cost function that penalizes deviations from these conditions. The global deviations from both the *gray value constancy* and the *gradient constancy* conditions are measured by the cost

$$D_D(u) = \int_{i\in\mathcal{B}} \Psi(|x_{n,i} - x_{n-1,u(i)}|^2 + \gamma|\nabla_{n,k,i} - \nabla_{n-1,k,u(i)}|^2)dx \tag{10}$$

where γ is a weight between the two conditions, $\Psi(s^2) = \sqrt{s^2+\epsilon^2}$ is an increasing concave function which would result in an L^1 norm minimization and is used to limit the influence of outliers under quadratic distortion metric, ϵ is a fixed small positive constant to offer advantages in the minimization process, and is set to 0.001 in experiments [41].

In addition, a smoothness term is used to penalize the *motion smoothness* condition. It penalizes the total variation of the flow field

$$D_S(u) = \int_{i\in\mathcal{B}} \Psi(|\nabla_{s,h}|^2 + |\nabla_{s,v}|^2)dx \tag{11}$$

where spatial gradient $\nabla_{s,v}$ and $\nabla_{s,h}$ indicate the spatial gradient in vertical and horizontal directions.

The total cost is the weighted sum between $D_D(u)$ and $D_S(u)$

$$D(u) = D_D(u) + \alpha D_S(u) \tag{12}$$

where $\alpha > 0$ is a regularization parameter.

The detailed numerical approximation algorithm for (12) could refer to [28] and [41].

2) *Target Gradient $\nabla_{n,k,i}$:*

a) *PD-DVC with feedback channel.* With the derived optical flow field between the MMSE reconstruction of the WZ frame and the reference frame, parameter $\nabla_{n,k,i}$ of pixel $x_{n,i}$ in direction k can be approximated by the gradient of corresponding pixel $x_{n-1,u(i)}$ in direction k in the reference frame X_{n-1}, that is

$$\nabla_{n,k,i} = \nabla_{n-1,k,u(i)} \tag{13}$$

where $u(i)$ is the index of the matching pixel of $x_{n,i}$ in frame X_{n-1}. The examples of the motion compensation based warping results for the PD-DVC scheme are shown in Fig. 8.

Fig. 8. Examples of the warping results for PD-DVC (the second frame of *Foreman* sequence with $Q_s = 64$). (a) Original WZ frame. (b) MMSE reconstruction. (c) Warping result from the previous frame. (d) Warping result from the next frame.

b) *TD-DVC with feedback channel.* As for the TD-DVC scheme, the pixel domain target gradient $\nabla^{PD}_{n,k,i}$ and optical flow field are initially generated. For the transform domain optimization, a node is a coefficient in the corresponding coefficient band, and the neighboring nodes are neighboring coefficients in that band. For example, neighboring nodes in the DC coefficient band are DC coefficients of adjacent 4×4 blocks in a specific direction. The target gradient $\nabla^{TD}_{n,b,k,i}$ for the transform domain optimization is predicted by the gradient of corresponding coefficient band of warping results. Fig. 9 shows the warping results for TD-DVC with $Q_s = 64$ (QPW = 40) and $Q_s = 32$ (QPW = 34). It can be seen that the warping results are acceptable even with a coarse MMSE estimate as the target frame, i.e., QPW = 40.

c) *TD-DVC with side-information refinement.* In TD-DVC scheme with SIR, the reconstruction process is implemented after each bit-plane has been decoded. The quantization bin $[B_L, B_U]$ is progressively halved along with the bit-plane decoding. With the specific quantization bin $[B_L, B_U]$ for a bit-plane, the target gradient $\nabla^{PD}_{n,k,i}$ and

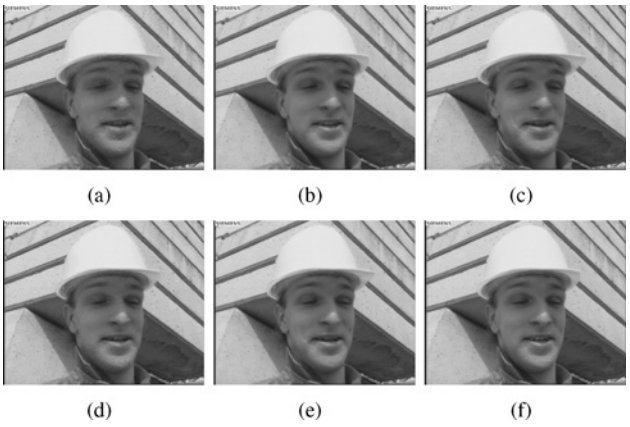


Fig. 9. Examples of the warping results for TD-DVC (the second frame of *Foreman* sequence). (a) MMSE reconstruction ($Q_s = 64$). (b) Warping result from the previous frame ($Q_s = 64$). (c) Warping result from the next frame ($Q_s = 64$). (d) MMSE reconstruction ($Q_s = 32$). (e) Warping result from the previous frame ($Q_s = 32$). (f) Warping result from the next frame ($Q_s = 32$).

$\nabla^{TD}_{n,b,k,i}$ can be calculated as that in TD-DVC scheme with feedback channel.

d) *TD-DVC without feedback channel.* In DVC approaches without feedback channel, bit-plane decoding failure may occur [21], [23]. Once bit-planes decoding failure exists in the WZ frame, there are two kinds of treatment.

i) Discard all subsequent less significant bit-planes of the coefficient band if one bit-plane decoding failure is detected, and only correctly decoded bit-planes are used for reconstruction. The decoding failure can be detected by verifying the parity-check constraint for LDPCA approaches. Cyclic redundancy check (CRC) can be used to detect decoding failure if Turbo codes are adopted as the SW codec. In this case, the target gradient $\nabla_{n,k,i}$ can be computed as that in TD-DVC with feedback channel.

ii) Use all bit-planes in the reconstruction process. In this case, bit errors in significant bit-planes will induce outlier blocks, as shown in Fig. 10. Therefore, outlier blocks might be detected with a spatial variant at the boundary of blocks, and replaced by the side-information. The modified reconstruction frame is used to acquire the optical flow field as in the conventional TD-DVC scheme with feedback channel.

B. Reliability of Target Gradient: $w_{n,k,i}$

Since the derived optical flow field is imperfect, the derived target can be unreliable in some regions. We introduce a parameter $w_{n,k,i}$ to measure the reliability of target gradient $\nabla_{n,k,i}$ to prevent inappropriate penalizing.

1) *Pixel Domain DVC (PD-DVC):* Within the PD-DVC scheme, the pixel value $x_{n-1,u(i)}$ in the warping result is checked by the quantization bin $[B_L, B_U]$: $w_{n,k,i}$ is set to 1 if $x_{n-1,u(i)} \in [B_L, B_U]$ and $x_{n-1,u(k(i))} \in [B_L, B_U]$; otherwise it is set to 0. Finally, $w_{n,k,i}$ is normalized for pixel $x_{n,i}$ among all directions k, that is, to impose an even weight on each edge of the MRF. Fig. 11 presents an example of the derived reliability mask of *Foreman* sequence, where black dots denote unreliable pixels in the corresponding warping results. It can be seen that, the unreliable pixels mainly be located at either the object edges where the occlusion could exist or the regions where significant artifacts appear from the MMSE reconstruction, e.g., the chin of *Foreman*.

Fig. 10. MMSE reconstruction results with bit-error in TD-DVC approaches without feedback channel. (a) 8th frame and (b) 48th frame of *Foreman* sequence.

Fig. 11. Reliability mask for the second frame of *Foreman* sequence. The black dots denote unreliable pixels. (a) Warping result from the previous frame ($Q_s = 64$). (b) Warping result from the next frame ($Q_s = 64$). (c) Warping result from the previous frame ($Q_s = 32$). (d) Warping result from the next frame ($Q_s = 32$).

2) *Transform Domain DVC (TD-DVC):* Because the MRF-based reconstruction for TD-DVC would be alternately implemented in the pixel domain and the transform domain, we calculate the weighting parameter $w_{n,k,i}$ in the pixel domain and the transform domain, respectively. It is worth mentioning that the quantization bin $[B_L, B_U]$ of TD-DVC is only available in the transform domain. For neighboring node pair where an unreliable coefficient is involved, the weighting parameter $w^{TD}_{n,b,k,i}$ is set to 0 to prevent inappropriate penalizing, and set to 1 if both of the neighboring coefficients are reliable. Finally, $w^{TD}_{n,b,k,i}$ is normalized for the coefficient $f_{n,b,i}$ in direction k. Fig. 12 shows examples of the derived unreliable mask. Since $[B_L, B_U]$ is undefined in the pixel domain, we always set weighting parameter $w^{PD}_{n,k,i}$ for the pixel domain to 1 and normalize them over all directions of a pixel.

3) *TD-DVC With Side-Information Refinement (SIR):* In TD-DVC scheme with SIR, the reconstruction process is implemented after each bit-plane has been decoded. The quantization bin $[B_L, B_U]$ is progressively halved along with the bit-plane decoding. With the specific quantization bin $[B_L, B_U]$ for a bit-plane, the target gradient $\nabla^{PD}_{n,k,i}$ and $\nabla^{TD}_{n,b,k,i}$ can be calculated as in the TD-DVC scheme with feedback channel.

4) *TD-DVC Without Feedback Channel:* In DVC approaches without feedback channel, the size of quantization bin $[B_L, B_U]$ varies for different coefficient bands because

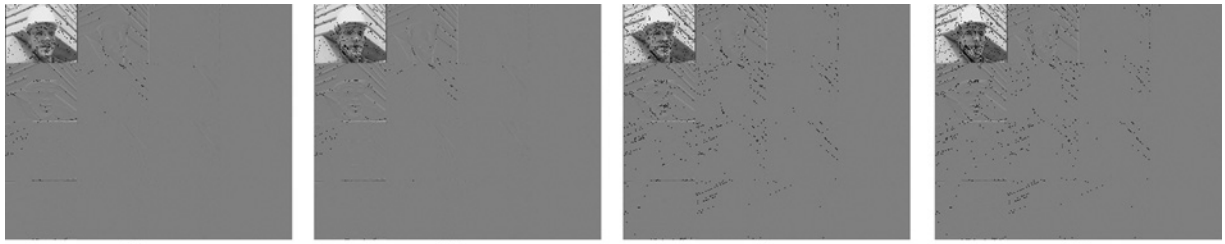


Fig. 12. Unreliable mask of the second frame of *Foreman* sequence for TD-DVC. (a) Mask for the warping results from the previous frame (Q_s = 64). (b) Mask for the warping results from the next frame (Q_s = 64). (c) Mask for the warping results from the previous frame (Q_s = 32). (d) Mask for the warping results from the next frame (Q_s = 32).

bit-plane decoding failure may appear [21], [23]. Because LDPC codes can detect decoding failure and CRC bits can be used for Turbo codes to detect the decoding failure [12], the bit-plane decoding failure can be detected at the decoder side. To verify the reliability of derived target gradient $\nabla_{n,b,k,i}$, the quantization bin $[B_L, B_U]$ is counted with only correctly decoded significant bit-planes. If one bit-plane fails to be decoded, all subsequent bit-planes for this coefficient band will be discarded. The weighting parameter $w^{TD}_{n,b,k,i}$ is set to 1 if the warping results $f_{n,b,i} \in [B_L, B_U]$, otherwise it is set to 0. Likewise, for optimization, the weighting parameter $w^{PD}_{n,k,i}$ in the pixel domain is always set to 1. Finally, the weighting parameters $w^{PD}_{n,k,i}$ and $w^{TD}_{n,b,k,i}$ are normalized over all directions, i.e., $\sum_k w^{PD}_{n,k,i} = 1$ and $\sum_k w^{TD}_{n,b,k,i} = 1$.

C. Analysis of Regularization Parameter η

The parameter η is introduced to adjust the relative ratio between the *data term* and the *GR term*. Fig. 13 presents the performance gain of the proposed "MRF-ICM" algorithm over the MMSE reconstruction along with the increase of η for the PD-DVC and TD-DVC schemes, where the first 100 frames of each sequence are tested. It can be seen that: in case of large quantization step, the performance gain of the proposed algorithm would be enhanced with the increase of η; when quantization step is small, large η may degrade the final reconstruction performance because of over-smoothing. To prevent generating the over-penalized effect, η should be reduced in accordance with the decrease of the quantization step, and could be set to 0 for small quantization step size, e.g., the number of bit-planes (BPN) > 5 for PD-DVC or $QPW < 28$ for TD-DVC.

D. Summary of the Proposed Algorithm

In the PD-DVC scheme, the proposed reconstruction algorithm initially derives an initial estimate of the WZ frame by the MMSE algorithm, which could accelerate the energy minimization process and achieve a satisfying output even with a deterministic greedy strategy, e.g., the ICM algorithm. As for the TD-DVC scheme, the GR constraint would be enabled in both the DCT domain and the pixel domain. The pixel domain optimization aims to eliminate the blocking artifacts, while the DCT domain optimization imposes the quantization bin constraint on the output to prevent inappropriate penalizing.

The major steps of the proposed MRF-based reconstruction algorithm are summarized as follows.

Input: Quantization bin $[B_L, B_U]$,
side-information Y_n,
previously decoded key frames X_{n-1} and X_{n+1},
statistical correlation between Y_n and WZ frame X_n.
Output: reconstruction result $\widehat{X}_n$
Important intermediate variables:
$\widetilde{X}^{\text{Warp}}$: warping result,

Algorithm begin
/ Initiation */*

$$\widehat{x}_i^{\text{MMSE}} = E[x_i | x_i \in [B_L, B_U], y_i] = \frac{\sum_{x=B_L}^{B_U} x_i\, p(x_i|y_i)}{\sum_{x_i=B_L}^{B_U} p(x_i|y_i)};$$

/ Estimate target gradient ∇ */*
$\mathbf{u} = \arg\min \mathbf{D_m}$ (12)
*/*Warping*/*
$\widetilde{x}_i^{\text{Warp},n-1} = x_{n-1,u(i)}$
$\widetilde{x}_i^{\text{Warp},n+1} = x_{n+1,u'(i)}$
if PD-DVC
$\nabla_{n-1,k,i} = \widetilde{x}_i^{\text{Warp},n-1} - \widetilde{x}_{k(i)}^{Warp,n-1}$,
$\nabla_{n+1,k,i} = \widetilde{x}_i^{\text{Warp},n+1} - \widetilde{x}_{k(i)}^{\text{Warp},n+1}$
else
$\nabla^{PD}_{n-1,k,i} = \widetilde{x}_i^{\text{Warp},n-1} - \widetilde{x}_{k(i)}^{\text{Warp},n-1}$,
$\nabla^{PD}_{n+1,k,i} = \widetilde{x}_i^{\text{Warp},n+1} - \widetilde{x}_{k(i)}^{\text{Warp},n+1}$;
$\widetilde{F}_{\text{Warp}} = DCT(\widetilde{X}_{\text{Warp}})$,
$\nabla^{TD}_{n-1,b,k,i} = f_{n-1,b,u(i)} - f_{n-1,b,k(u(i))}$,
$\nabla^{TD}_{n+1,b,k,i} = f_{n+1,b,u(i)} - f_{n+1,b,k(u'(i))}$;
end if
*/*Analyze reliability of the warping results $\widetilde{X}^{\text{Warp}}$ */*
$w_{n,k,i} = 1$ **if** $\widetilde{x}_i \in [B_L, B_U]$ **and** $\widetilde{x}_{k,i} \in [B_L, B_U]$
Normalize $w_{n,k,i} = \frac{w_{n,k,i}}{\sum_k w_{n,k,i}}$
Implement energy minimization
Algorithm end

Energy minimization for PD-DVC
Input: quantization bin $[B_L, B_U]$,
statistical correlation between Y_n and WZ frame X_n,
target gradient $\nabla_{n-1,k,i}$ and $\nabla_{n+1,k,i}$
weighting parameter $w_{n-1,k,i}$ and $w_{n+1,k,i}$.
Output: reconstruction result $\widehat{X}^{MRF}$

Algorithm begin
/ Initialization */*
$\widehat{X}^{MRF} = \widehat{X}^{\text{MMSE}}$
*/*Perform optimization*/*
repeat
Calculate cost of energy function, $E(\widehat{X}^{MRF})$
Optimization
Calculate cost of energy function, $E(\widehat{X}'^{MRF})$
$\widehat{X}^{MRF} = \widehat{X}'^{MRF}$
until $|E(\widehat{X}^{MRF}) - E(\widehat{X}'^{MRF})| < \sigma_{threshold}$ or $Iteration > MaxIteration$
Algorithm end

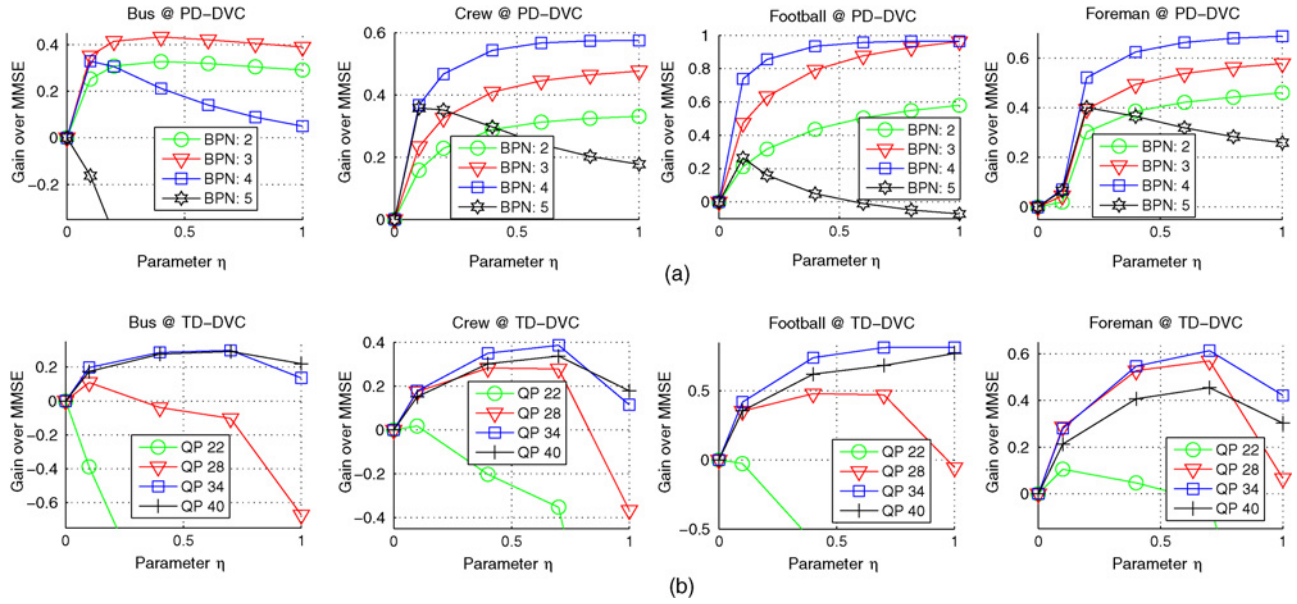


Fig. 13. Reconstruction performance gain over the MMSE reconstruction along with the increase of η. The y-axis indicates the performance gain of the proposed "MRF-ICM" algorithm over the MMSE reconstruction algorithm, and the x-axis denotes the regularization parameter η. (a) Results for PD-DVC with $2 \sim 5$ bit-planes (BPN). (b) Results for TD-DVC from $QPW = 22$ to $QPW = 40$.

Energy minimization for TD-DVC

Input: quantization bin $[B_L, B_U]$,
statistical correlation between Y_n and WZ frame X_n,
pixel domain target gradient $\nabla^{PD}_{n-1,k,i}$ and $\nabla^{PD}_{n+1,k,i}$
transform domain target gradient $\nabla^{TD}_{n-1,b,k,i}$ and $\nabla^{TD}_{n+1,b,k,i}$
weighting parameter $w^{TD}_{n-1,k,i}$ and $w^{TD}_{n+1,k,i}$.

Output: reconstruction result $\widehat{X}^{MRF}$

Important intermediate variables:
X': intermediate output of pixel domain optimization
F': intermediate output of transform domain optimization

Algorithm begin
/ Initialize */*
$\widehat{F}^{MRF} = \widehat{F}^{\text{MMSE}}$
$\widehat{X}^{MRF} = \widehat{X}^{\text{Warp}}$
*/*Optimization*/*
repeat
Calculate cost of energy function, $E(\widehat{X}^{MRF})$
Pixel domain optimization
Calculate cost of energy function, $E(\widehat{X}'^{MRF})$
$\widehat{F}^{MRF} = DCT(\widehat{X}'^{MRF})$
DCT domain optimization
$\widehat{X}^{MRF} = IDCT(\widehat{F}'^{MRF})$
until $|E(\widehat{X}^{MRF}) - E(\widehat{X}'^{MRF})| < \sigma_{threshold}$ or $Iteration > MaxIteration$
Algorithm end

V. Experimental Results

In this section, we conduct experiments to evaluate the performance of the proposed reconstruction algorithm for DVC applications. In experiments, four different reconstruction approaches are evaluated: maximum probability (Max. Prob.) reconstruction, MMSE reconstruction, and the proposed spatio-temporal MRF-based reconstruction with the TRW-S and the ICM algorithms. Four CIF (352×288 at 15 Hz) video sequences (*Bus*, *Crew*, *Football*, and *Foreman*) and two 720×480 video sequences (*Driving* and *WhaleShow*) are tested in the encoding structure "I-WZ-I-WZ-I-." The bidirectional motion compensation with search window size 8×8 and the search range 16 are adopted to generate the side-information [26]. The SW codec adopts the LDPCA approach with a block length of 6336 bits and the degree distribution [24], [25] is shown in (14). It can produce a set of coding rates $\frac{2}{66}, \frac{3}{66}, \frac{4}{66}, \ldots, \frac{66}{66}$

$$\lambda(x) = 0.316x^1 + 0.415x^2 + 0.128x^6 + 0.069x^7 + 0.020x^{18} + 0.052x^{20}. \quad (14)$$

The LDPCA decoding stops when the parity-check node constraints are satisfied or the maximum iteration (200 iterations in experiments) is reached. If the check-node constraints (i.e., the syndrome) are still not satisfied after the LDPCA decoding stopped for TD-DVC without feedback channel, this bit-plane would be regarded as a decoding failure. Although the decoded bit-planes may still contain errors even the check-node constraint is satisfied, it only appears with a very small probability and induces slightly performance decrease. In experiments, these undetected errors are regarded as correct bits for reconstruction. In energy minimization algorithms for reconstruction, iteration operation would terminate when the energy of two successive iterations is identical or the iteration limit (herein the maximum iteration number is set to 5) is reached. The experiments are implemented on Intel Xeon CPU E5520@2.27 GHz with 6.0 GB RAM.

A. Rate-Distortion Performance

Figs. 14 and 15, respectively, show the RD performance of the different reconstruction algorithms for PD-DVC and TD-DVC. It can be seen that the proposed MRF-based reconstruction algorithm either with the TRW-S or the ICM method would significantly outperform the MMSE and the maximum probability reconstruction algorithm. In addition, it is also observed: the "MRF-TRW-S" and the "MRF-ICM" algorithms achieve similar RD performance because the MMSE

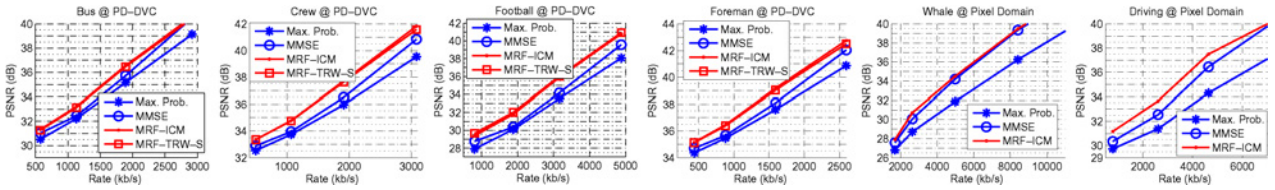


Fig. 14. Rate-distortion performance for the PD-DVC scheme.

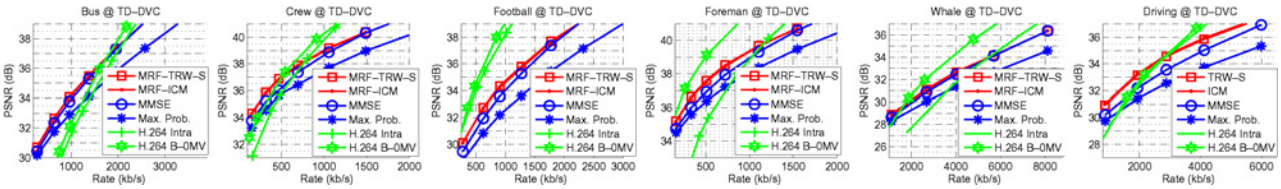


Fig. 15. Rate-distortion performance for the TD-DVC scheme.

reconstruction provides a good initial estimate for the ICM strategy.

To evaluate the proposed reconstruction algorithm more precisely, the Bjøntegaard (BD) delta values are also evaluated based on rate-distortion curve fitting [42]. Table I shows the BD-PSNR and BD-bitrate comparison values of the proposed "MRF-ICM" algorithm against the MMSE algorithm. The BD-PSNR indicates the average PSNR difference in dB over the whole range of bit-rate, which is calculated as follows:

$$\Delta PSNR = \left(\int_{r_L}^{r_H} (D_2(r) - D_1(r))\,\mathrm{d}r\right)/(r_H - r_L) \qquad (15)$$

where $D_1(r)$ and $D_2(r)$ are two R-D curves which, respectively, represent the reconstructed distortions in PSNR of the MMSE and the "MRF-ICM" reconstruction algorithm approximated by a third order polynomial as

$$D_{\mathrm{PSNR}} = D(r) = a_0 r^3 + a_1 r^2 + a_2 r + a_3 \qquad (16)$$

and $r = \log R$ represents the logarithm of the bitrate, so that r_H and r_L in (15) are the logarithmic forms of R_H and R_L which bound the bitrate range of coding results of two fitted R-D curves, $a_0, \cdots, a_3$ are the parameters of fitting polynomial R-D curves. An inverse process which fits the interpolation to find bitrate as a function of PSNR as $D_{\mathrm{bitrate}} = D(PSNR) = a_0 PSNR^3 + a_1 PSNR^2 + a_2 PSNR + a_3$ can be employed to find BD-bitrate. BD-bitrate indicates the average bitrate difference in % over the whole range of PSNR. The results in Table I show that the proposed algorithm achieves up to 1.309 dB PSNR improvement and 17.489% bit-rate saving for PD-DVC scheme, and up to 0.954 dB PSNR improvement and 20.017% bit-rate saving for TD-DVC scheme.

For the TD-DVC scheme without feedback channel in Section II-D, Fig. 16 shows the RD performance of the "MRF-ICM" and the MMSE reconstruction algorithms under "bit-plane (BP) error" and "bit error" conditions, where the packet loss ratio is set to 8% and the target bit-plane decoding failure probability is 0.1. The quantization bin $[B_L, B_U]$ is determined by only correctly decoded bit-planes, and in turn varies for different coefficient bands. Under the "BP error" condition, all subsequent bit-planes are discarded once a bit-plane decoding failure is detected. Under the "bit error" condition, all the decoded bit-planes are used for reconstruction. Compared to TD-DVC (with feedback channel condition) shown in Fig. 15, the performance gain of the proposed MRF-based reconstruction over the MMSE reconstruction is slightly better because only correct regions in the previous frames are used to generate the local gradient to prevent inappropriate penalizing.

TABLE I

AVERAGE BD DELTAS OF THE PROPOSED ''MRF-ICM'' RECONSTRUCTION ALGORITHM AGAINST THE MMSE RECONSTRUCTION ALGORITHM FOR PD-DVC AND TD-DVC SCHEMES

	PD-DVC		TD-DVC	
Sequences	ΔPSNR (dB)	ΔRate (%)	ΔPSNR (dB)	ΔRate (%)
Bus	0.636	−8.001	0.271	−5.481
Crew	0.880	−17.135	0.588	−15.804
Football	1.309	−17.489	0.645	−10.359
Foreman	0.816	−16.006	0.530	−16.483
Driving	0.640	0.249	0.954	−20.017
Whale	0.344	−2.175	0.215	−2.886

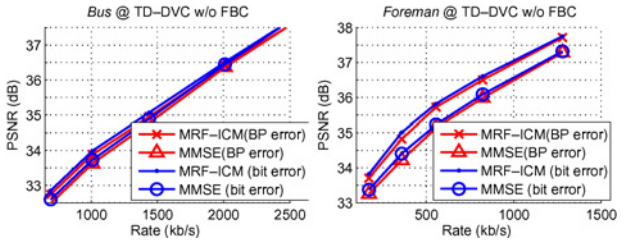


Fig. 16. RD performance comparison with different reconstruction algorithms for the TD-DVC scheme without feedback channel (FBC), where he packet loss ratio is 8% and the bit-plane decoding failure probability is 0.05.

For the DVC scheme with SIR [12] in Section II-C, Fig. 17 shows the RD performance of WZ frames with different reconstruction algorithms. It can be seen that the proposed MRF-based approach achieves superior performance than the existing MMSE reconstruction results.

B. PSNR of Each WZ Frame

Figs. 18 and 19, respectively, present the PSNR of the reconstructed WZ frames among all reconstruction algorithms

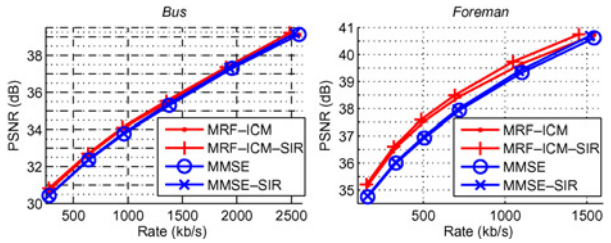


Fig. 17. RD performance comparison with different reconstruction algorithms for the TD-DVC scheme with side-information refinement (SIR).

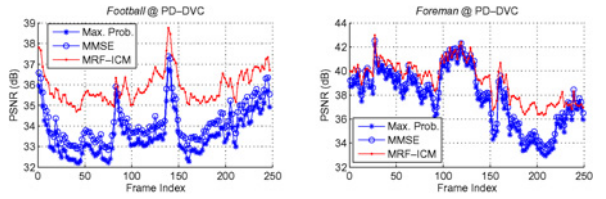


Fig. 18. PSNR comparison of the reconstructed WZ frames for the PD-DVC scheme.

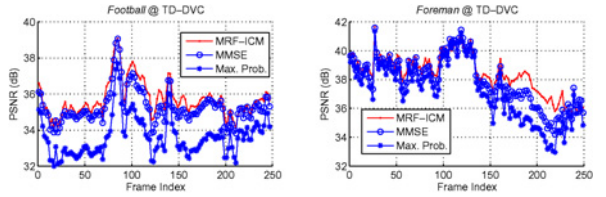


Fig. 19. PSNR comparison of the reconstructed WZ frames for the TD-DVC scheme.

for the PD-DVC and TD-DVC schemes. It can be seen that the proposed reconstruction would achieve a significant performance gain over the existing MMSE and maximum probability reconstruction, and especially in favor of the frames with relative low side-information quality. Based on the good initial MMSE estimator, the proposed "MRF-TRW-S" and the "MRF-ICM" approaches are observed to achieve very close objective performance.

C. Subjective Quality Comparison

Fig. 20 presents the subjective quality comparison on the sampled frames from different reconstruction algorithms for the TD-DVC scheme. Since the "MRF-TRW-S" and "MRF-ICM" algorithms have similar performance, only results with "MRF-ICM" reconstruction algorithm are presented and compared to that of the "MMSE" reconstruction algorithm. It can be seen that the proposed MRF-based algorithm would achieve better subjective performance without the blocking artifact which obviously appears in the MMSE reconstruction.

D. Computational Complexity

In this section, we will give the computational complexity of the proposed reconstruction algorithm and compare it to that of the MMSE and the maximum probability reconstruction algorithm, and also compare it to that of other modules in the decoding process to demonstrate the ratio that the proposed reconstruction algorithm accounts for in the whole WZ decoding process.

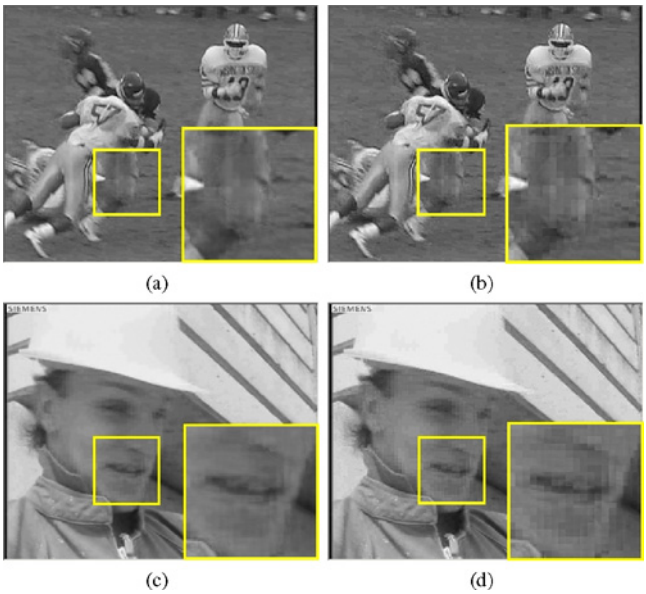


Fig. 20. Subjective performance comparison of different reconstruction algorithms for the TD-DVC scheme. (a), (c) MMSE algorithm. (b), (d) Proposed "MRF-ICM" algorithm.

TABLE II
AVERAGE COMPLEXITY FOR ONE WYNER-ZIV DECODING FOR THE PD-DVC SCHEME

Reconstruction Methods	Running Time of Different Sequences (S)			
	Bus	*Crew*	*Football*	*Foreman*
Maximum probability	0.0014	0.0013	0.0011	0.0013
MMSE	0.230	0.231	0.245	0.251
MRF-ICM	12.475	13.704	13.598	13.776
MRF-TRW-S	1287.331	1117.353	1321.326	1290.408
Optical flow analysis	5.671	5.581	5.641	5.356
LDPCA decoding	292.130	225.833	310.578	311.673

Table II presents the computational complexity (the average time for one WZ frame) of different reconstruction algorithms for the PD-DVC scheme, where four bit-planes are transmitted. It shows that the MRF-ICM algorithm is much faster than the MRF-TRW-S algorithm and only accounts for a small amount of the overall decoding time, typically about 6% of the Slepian–Wolf decoding.

Likewise, Table III presents the computational complexity (the average time for one WZ frame) of different reconstruction algorithms for the TD-DVC scheme, where the quantization parameter for the WZ coding is set to $QPW = 28$. The MRF-ICM algorithm only accounts for a small amount of the overall decoding time, typically about 20% $\sim$ 30% of the SW decoding.

E. Convergence Analysis

In Figs. 21 and 22, we present the convergence property of the proposed "MRF-ICM" reconstruction algorithm for the PD-DVC and TD-DVC schemes. It shows that the blocking artifact is gradually alleviated with the number of iterations. It is worth mentioning that the over-smoothing in the pixel domain optimization of the TD-DVC scheme would be made a tradeoff with the iterations, because it is not strictly restricted

TABLE III
AVERAGE COMPLEXITY FOR ONE WYNER-ZIV DECODING FOR THE TD-DVC SCHEME

Reconstruction Methods	Running Time of Different Sequences (S)			
	Bus	*Crew*	*Football*	*Foreman*
Maximum probability	0.0905	0.0944	0.0910	0.0923
MMSE	0.332	0.338	0.319	0.318
MRF-ICM	25.069	26.356	24.344	25.836
MRF-TRW-S	1171.484	1203.500	1236.355	1231.663
Optical flow analysis	5.671	5.581	5.641	5.356
LDPCA decoding	131.707	80.656	102.080	110.241

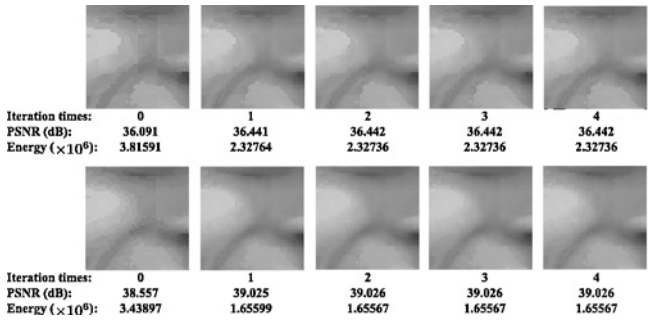


Fig. 21. Convergence analysis of the proposed "MRF-ICM" reconstruction algorithm for the PD-DVC scheme. Top: 3 WZ bit-planes. Bottom: 4 WZ bit-planes.

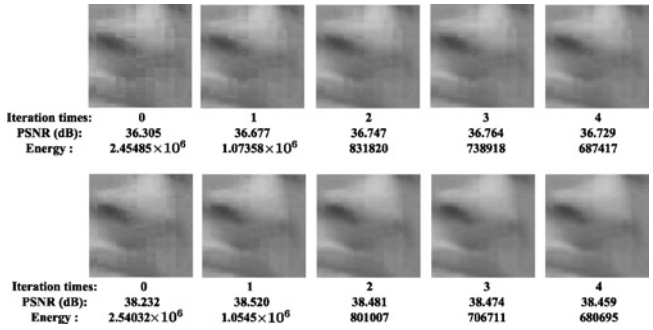


Fig. 22. Convergence analysis of the proposed "MRF-ICM" reconstruction algorithm for the TD-DVC scheme. Top: $QPW = 34$. Bottom: $QPW = 28$.

by the quantization bin like the PD-DVC scheme. Certainly, the blocking artifacts are steadily alleviated with the iterations.

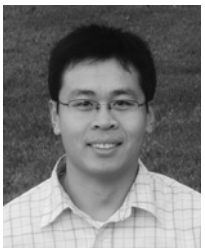
VI. CONCLUSION

In this paper, we proposed an MRF-based reconstruction algorithm for a generic DVC architecture which exploits the spatio-temporal consistency of video sequences as well as the statistical correlation between the WZ frame and the corresponding side-information to achieve an optimized reconstruction performance. The WZ frame is modeled as an MRF which reflects interactions among neighboring nodes on a 2-D lattice for natural video, and its MAP estimate is solved for the desired reconstruction to achieve better subjective and objective performance than the i.i.d. model of the existing maximum probability and the MMSE algorithms. The *energy function* consists of two terms: a *data term* measuring the statistical correlation, and a *GR term* enforcing local spatio-temporal consistency. The *data term* is measured by the MSE metric in accord with the PSNR objective quality measurement, and the *GR term* is calculated from a weighted summation of gradient deformation which is measured by the difference between the gradient of the desired output and that of the reference frames. With the derived *energy function*, the energy minimization for the MRF-based MAP estimate of the WZ frames is solved by global optimization (the TRW-S algorithm) and greedy strategies (the ICM algorithm) with the MMSE reconstruction as the initial estimate. The implementation of the proposed reconstruction algorithm is discussed with regard to the critical parameters under a wide variety of DVC scenarios.

REFERENCES

[1] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.
[2] F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi, and S. Klomp, "Distributed video coding: Selecting the most promising application scenarios," *Signal Process. Image Commun.*, vol. 23, no. 5, pp. 339–352, Jun. 2008.
[3] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
[4] M. Maitre, C. Guillemot, and L. Morin, "3-D model-based frame interpolation for distributed video coding of static scenes," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1246–1257, May 2007.
[5] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner–Ziv video coding performance," in *Proc. IEEE ICASSP*, vol. 2. May 2006, pp. 525–528.
[6] Y. Zhang and J. Garcia-Frias, "Data compression of correlated nonbinary sources using punctured turbo codes," in *Proc. IEEE Data Compression Conf.*, Apr. 2002, pp. 242–251.
[7] A. D. Liveris, Z. X. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, Oct. 2002.
[8] D. Kubasov, J. Nayak, and C. Guillemot, "Optimal reconstruction in Wyner–Ziv video coding with multiple side information," in *Proc. 9th IEEE Workshop MMSP*, Oct. 2007, pp. 183–186.
[9] A. Roca, J. PradesNebot, and E. J. Delp, "Adaptive reconstruction for Wyner–Ziv video coders," *Proc. SPIE Visual Commun. Image Process.*, vol. 7257, pp. 725708(1–9), Jan. 2009.
[10] A. Aaron, R. Zhang, and B. Girod, "Wyner–Ziv coding of motion video," in *Proc. Asilomar Conf. Signals Syst.*, Nov. 2002, pp. 240–244.
[11] E. Peixoto, R. L. de Queiroz, and D. Mukherjee, "A Wyner–Ziv video transcoder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 2, pp. 189–200, Feb. 2010.
[12] R. Martins, C. Brites, J. Ascenso, and F. Pereira, "Refining side information for improved transform domain Wyner–Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 9, pp. 1327–1341, Sep. 2009.
[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.
[14] R. Martins, C. Brites, J. Ascenso, and F. Pereira, "Adaptive deblocking filter for transform domain Wyner–Ziv video coding," *IET Image Process. Special Issue Distrib. Video Coding*, vol. 3, pp. 315–328, Jul. 2009.
[15] Y. Zhang, H. Xiong, Z. He, and S. Yu, "Reconstruction for distributed video coding: A Markov random field approach with context-adaptive smoothness prior," in *Proc. VCIP*, vol. 7744. Jul. 2010, pp. 774417 (1–10).
[16] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.
[17] J. Besag, "On the statistical analysis of dirty pictures (with discussion)," *J. R. Statist. Soc. Series B (Methodol.)*, vol. 48, no. 3, pp. 259–302, 1986.
[18] A. Avudainayagam, J. M. Shea, and D. Wu, "Hyper-trellis decoding of pixel-domain Wyner–Ziv video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 557–568, May 2008.
[19] J. Zhang, H. Li, and C. W. Chen, "Distributed image coding based on integrated Markov random field modeling and LDPC decoding," in *Proc. IEEE 10th Workshop MMSP*, Oct. 2008, pp. 265–270.

[20] D. Varodayan, A. Aaron, and B. Girod, "Exploiting spatial correlation in pixel-domain distributed image compression," in *Proc. Picture Coding Symp.*, Apr. 2006, pp. 1–4.
[21] R. Bernardini, M. Naccari, R. Rinaldo, M. Tagliasacchi, S. Tubaro, and P. Zontone, "Rate allocation for robust video streaming based on distributed video coding," *Signal Process. Image Commun.*, vol. 23, pp. 391–403, Jun. 2008.
[22] M. Morbee, J. Prades-Nebot, A. Pizurica, and W. Philips, "Rate allocation algorithm for pixel-domain distributed video coding without feedback channel," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2007, pp. 521–524.
[23] T. Sheng, G. Hua, H. Guo, J. Zhou, and C. W. Chen, "Rate allocation for transform domain Wyner–Ziv video coding without feedback," in *Proc. 16th ACM Int. Conf. Multimedia*, Oct. 2008, pp. 701–704.
[24] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Process.*, vol. 86, pp. 3123–3130, Nov. 2006.
[25] D. Varodayan. *Rate-Adaptive LDPC Accumulate Codes for Distributed Source Coding* [Online]. Available: http://www.stanford.edu/%7Edivad/ldpca.html
[26] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," in *Proc. 5th EURASIP Conf. Speech Image Process. Multimedia Commun. Services*, Jul. 2005, pp. 176–181.
[27] J. Ascenso, C. Brites, and F. Pereira, "Motion compensated refinement for low complexity pixel based distributed video coding," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance*, Sep. 2005, pp. 593–598.
[28] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Comput. Sci. Artif. Intell. Lab., Massachusetts Instit. Technol., Cambridge, May 2009.
[29] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[30] W. A. Weerakkody, W. A. C. Fernando, and A. M. Kondoz, "An enhanced reconstruction algorithm for unidirectional distributed video coding," in *Proc. IEEE Int. Symp. Consumer Electron.*, Apr. 2008, pp. 1–4.
[31] Y. Zhang, H. Xiong, L. Song, and S. Yu, "Spatial non-stationary correlation noise modeling for Wyner–Ziv error resilience video coding," in *Proc. IEEE ICIP*, Nov. 2009, pp. 2929–2932.
[32] Y. Zhang, H. Xiong, Z. He, S. Yu, and C.-W. Chen, "An error resilient video coding scheme using embedded Wyner–Ziv description with decoder side non-stationary distortion modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 498–512, Apr. 2011.
[33] J. Chen and C. K. Tang, "Spatio-temporal Markov random field for video denoising," in *Proc. IEEE Conf. CVPR*, Jun. 2007, pp. 2232–2239.
[34] R. Huang, V. Pavlovic, and D. N. Metaxas, "A new spatio-temporal MRF framework for video-based object segmentation," in *Proc. MLVMA Conjunction ECCV*, Oct. 2008, pp. 1–12.
[35] S. Dai, S. Baker, and S. B. Kang, "An MRF-based deinterlacing algorithm with exemplar-based refinement," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 956–968, May 2009.
[36] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed. London, U.K.: Springer, 2009.
[37] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[38] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, vol. 1. Jul. 2004, pp. 261–268.
[39] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.
[40] J. Yedidia, W. Freeman, and Y. Weiss, "Generalized belief propagation," in *Proc. Adv. Neural Inform. Process. Syst.*, 2000, pp. 689–695.
[41] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. ECCV*, May 2004, pp. 25–36.
[42] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves (VCEG-M33)*, VCEG Meeting (ITU-T SG16 Q.6), Austin, TX, Apr. 2001.



Yongsheng Zhang (M'11) received the B.S. and M.S. degrees in electronic engineering from Shandong University, Jinan, Shandong, China, in 2002 and 2005, respectively. He is currently working toward the Ph.D. degree in electrical engineering from the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China.

His current research interests include distributed video coding and image processing.

Hongkai Xiong (M'01–SM'10) received the Ph.D. degree in communication and information systems from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2003.

Since 2003, he has been with the Department of Electronic Engineering, SJTU, where he is currently an Associate Professor. From December 2007 to December 2008, he was with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, as a Research Scholar. He has published over 90 international journal/conference papers. In SJTU, he directs the Intelligent Video Modeling Laboratory and multimedia communication area in the Key Laboratory of the Ministry of Education of China—Intelligent Computing and Intelligent System which is also co-granted by Microsoft Research, Beijing, China. His current research interests include source coding/network information theory, signal processing, computer vision and graphics, and statistical machine learning.

Dr. Xiong was the recipient of the New Century Excellent Talents in University Award in 2009. In 2008, he received the Young Scholar Award of Shanghai Jiao Tong University. He has served for various IEEE conferences as a technical program committee member. He acts as a member of the Technical Committee on Signal Processing of the Shanghai Institute of Electronics.

Zhihai He (S'98–M'01–SM'06) received the B.S. degree in mathematics from Beijing Normal University, Beijing, China, in 1994, the M.S. degree in mathematics from the Institute of Computational Mathematics, Chinese Academy of Sciences, Beijing, in 1997, and the Ph.D. degree in electrical engineering from the University of California, Santa Barbara, in 2001.

In 2001, he joined Sarnoff Corporation, Princeton, NJ, as a Technical Staff Member. In 2003, he joined the Department of Electrical and Computer Engineering, University of Missouri, St. Louis, as an Assistant Professor. His current research interests include image/video processing and compression, network transmission, wireless communication, computer vision analysis, sensor networks, and embedded system design.

Dr. He received the 2002 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award, and the SPIE VCIP Young Investigator Award in 2004. Currently, he is an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the *Journal of Visual Communication and Image Representation*. He is a Guest Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Special Issue on Video Surveillance. He is a member of the Visual Signal Processing and Communication Technical Committee of the IEEE Circuits and Systems Society, and serves as a technical program committee member or session chair for a number of international conferences.

Songyu Yu graduated from Shanghai Jiao Tong University, Shanghai, China, in 1963.

Since 1992, he has been a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. He has published more than 50 research papers, and serves as the Secretary-General of the Shanghai Images and Graphics Association. His current research interests include image processing, video coding, and digital TV.

Chang Wen Chen (F'04) received the B.S. degree from the University of Science and Technology of China, Hefei, Anhui, China, in 1983, the M.S.E.E. degree from the University of Southern California, Los Angeles, in 1986, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, Urbana, in 1992.

He has been a Professor of computer science and engineering with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, since 2008. Previously, he was an Allen S. Henry Distinguished Professor with the Department of Electrical and Computer Engineering, Florida Institute of Technology, Melbourne, from 2003 to 2007. He was with the Faculty of Electrical and Computer Engineering, University of Missouri-Columbia, Columbia, from 1996 to 2003, and with the University of Rochester, Rochester, NY, from 1992 to 1996. From 2000 to 2002, he was the Head of the Interactive Media Group, David Sarnoff Research Laboratories, Princeton, NJ. He has consulted with Kodak Research Laboratories, Rochester, Microsoft Research, Beijing, China, Mitsubishi Electric Research Laboratory, Cambridge, MA, NASA Goddard Space Flight Center, Greenbelt, MD, and the U.S. Air Force Rome Laboratories, Rome, NY.

Dr. Chen was the Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from January 2006 to December 2009. He has served as an editor for the PROCEEDINGS OF THE IEEE, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE MULTIMEDIA, the *Journal of Wireless Communication and Mobile Computing*, the *EUROSIP Journal of Signal Processing: Image Communications*, and the *Journal of Visual Communication and Image Representation*. He has chaired and served on numerous technical program committees for the IEEE and other international conferences. He was elected a fellow of the IEEE for his contributions to digital image and video processing, analysis, and communications, and a fellow of the SPIE for his contributions to electronic imaging and visual communications.