Space-Time Quantization and Motion-Aligned Reconstruction for Block-Based Compressive Video Sensing

Ran Li, Hongbing Liu, Wei He and Xingpo Ma
School of Computer and Information Technology,
Xinyang Normal University
[e-mail: liran358@163.com]
*Corresponding author: Ran Li

Abstract

The Compressive Video Sensing (CVS) is a useful technology for wireless systems requiring simple encoders but handling more complex decoders, and its rate-distortion performance is highly affected by the quantization of measurements and reconstruction of video frame, which motivates us to presents the Space-Time Quantization (ST-Q) and Motion-Aligned Reconstruction (MA-R) in this paper to both improve the performance of CVS system. The ST-Q removes the space-time redundancy in the measurement vector to reduce the amount of bits required to encode the video frame, and it also guarantees a low quantization error due to the fact that the high frequency of small values close to zero in the predictive residuals limits the intensity of quantizing noise. The MA-R constructs the Multi-Hypothesis (MH) matrix by selecting the temporal neighbors along the motion trajectory of current to-be-reconstructed block to improve the accuracy of prediction, and besides it reduces the computational complexity of motion estimation by the extraction of static area and 3-D Recursive Search (3DRS). Extensive experiments validate that the significant improvements is achieved by ST-Q in the rate-distortion as compared with the existing quantization methods, and the MA-R improves both the objective and the subjective quality of the reconstructed video frame. Combined with ST-Q and MA-R, the CVS system obtains a significant rate-distortion performance gain when compared with the existing CS-based video codecs.

Keywords: Compressive video sensing, space-time quantization, motion-aligned reconstruction, median filter based prediction, multi-hypothesis prediction.

1. Introduction

1.1 Challenge and Objective

Compressive Video Sensing (CVS) [1]-[3] refers to a technique that realizes the low-complex video coding by using the Compressive Sensing (CS) [4], which has been attracted a lot of research interest because of its potential application prospect in the resource-constrained environments, e.g., Wireless Multimedia Sensor Networks (WMSN) [5]. So far, the existing two problems in CVS have not been solved effectively yet, and they are quantization of CS measurements and reconstruction of video frame respectively. The rate-distortion performance of CVS is highly affected by them, and therefore we requires to exploit some good features of video signal for improving the accuracy of quantization and reconstruction.

The first objective of this paper is to guarantee a low quantization error with as few bits as possible. Considering that the space-time correlation still remains in the measurement domain, we try to further remove the redundancy between CS measurements by using this property. These low redundant data are easier to be compressed by the entropy coding (e.g., Huffman [6]), and the high frequency of small values close to zero will also limits the intensity of quantizing noise. While reducing the amount of bits, it is still expected to obtain the improvement of reconstruction quality, and therefore the another objective of this paper is to propose an efficient reconstruction algorithm. Each to-be-reconstructed block has always some high-correlated temporal neighborhoods along the motion trajectories between the adjacent reconstructed frames, which provides convenience for guaranteeing the better quality with few measurements, and consequently we will focus on how to extract accurate motion vector field of the to-be-reconstructed video frame on the promise of not introducing excessive computations.

1.2 Related Work

Real-life implementation requires to produce a compressed bitstream from the CS measurements, and therefore recent literature has seen an explosion of interest in the quantization of CS measurements. It is the most straightforward solution to simply apply uniform Scalar Quantization (SQ) to each of the CS measurements, and however its poor rate-distortion performance brings trouble to CS recovery [7],[8]. To make up the performance degradation resulting from the inefficient SQ, some reconstruction algorithms [9],[10] were specially designed to resist quantizing noise, and but the fundamental causes cannot still be addressed only through an optimization of the reconstruction process. A more efficient approach is to make the quantization process better suited to the characteristic of CS measurements, e.g., [11] designs an optimal fixed-rate quantizer to minimize the quantization error depending on the distribution of CS measurements, and the Sigma-Delta quantizer in [12] and [13] even exploits the Restricted Isometry Property (RIP) of measurement matrix to provide robustness to quantizing noise. The one-dimensional (1-D) general signal is regarded as the quantized object in the above-mentioned methods, and but they cannot present the better performance when the quantized object is replaced with the more special 2-D image, which is because the intrinsic spatial correlation existing in natural images is neglected in their quantization process. The exploitation of spatial correlation benefits from the block-based CS (BCS) [14]-[16], i.e., the random measuring of CS is effectuated in blocks. [17] proposed a

framework of quantization via simple uniform SQ coupled with Differential Pulse Code Modulation (DPCM) of the CS measurements, and its surprisingly competitive rate-distortion performance breaks a new path for quantizing CS measurements of image. The main idea of DPCM-plus-SQ is to consider the previous block as a prediction and uniformly quantize the residual between the current block and its prediction of measurements, and however it is not efficient enough to select previous block as the prediction of current one because the non-stationarity of natural images is ignored. To address this problem, [18] proposed the Spatially Directional Predictive Coding (SDPC) to select the more suitable prediction for each block in terms of an optimal directional mode. Although the SDPC achieves the significant improvements in rate-distortion performance as compared with DPCM-plus-SQ, it also brings the two defects: (1) it increases the computational burden of encoder to determine the optimal prediction; (2) missing the flag of best predictive mode in transmission will severely disrupt the inverse quantization at the decoder. Therefore, it remains to be developed that the more economical and safe spatially directional mode. For 3-D video signal, there will be very little difference between consecutive frames in the measurement domain, and but this temporal correlation has not been incorporated into the BCS paradigm of video. Given this, in addition to exploit spatial correlation, we can further take advantage of temporal correlation when quantizing each video frame.

The reconstruction process also plays an important role in the performance improvement of CVS system, and a lot of researches have been carried out for developing the effective joint reconstruction techniques which derived from the decoding strategy of Distributed Video Coding (DVC) [19]. The joint reconstruction consists of Side Information (SI) generation and residual recovery, in which the former provides a prediction of the to-be-reconstructed frame and the latter uses the traditional CS recovery algorithm (e.g., GPSR [20], OMP[21], etc.) to reconstruct the residual between original frame and its SI. The prediction accuracy of SI is especially crucial on the reconstruction performance because the high accurate SI makes the residual more sparse, and consequently the much attention has been devoted to SI generation. Without CS measurements, the SI can be interpolated, e.g., [22] and [23] reconstruct the current video frame with help of the SI interpolated by Frame Rate Up-Conversion (FRUC). Generally speaking, the SI generation with CS measurements has superior performance than that generated by FRUC algorithm, because the form uses the information of current frame to generate SI. Among the SI generation methods with CS measurements, the most classic one is that, the DISCOS framework proposed by [24] uses the sparsity-constrain prediction algorithm based on a dictionary of temporal neighboring blocks. The more prominent work is that Multi-Hypothesis (MH) based SI prediction proposed by [25], which uses a Tikhonov constrain of l_2 norm to provide the better result with a low computational complexity. The exploitation of temporal redundancy is the important feature of MH prediction, and however it still exists some other methods to use temporal correlation, e.g., [26] constructs the structured patch groups by the temporal nonlocal similarity to make a prediction, and [27] trains the PCA dictionary of to-be-reconstructed frame by using the temporal-neighboring blocks. Unfortunately, the MH-based SI generation ignores the motion information between the neighboring reconstructed frames, and its superior performance cannot be effectively exploited for the sequences with moderate or large motions, which will cause performance degradation and meanwhile bring blurring and ghost artifacts in the reconstructed frames. Therefore, we will provide an appropriate motion estimate scheme to compute the motion vector field of the to-be-reconstructed video frame for improving the accuracy of MH prediction.

1.3 Motivation and Contribution

Space-Time Quantization (ST-Q). To improve the rate-distortion performance of quantized CVS, we present a quantization method which exploits the space-time correlation of CS measurements between video frames. Motivated by the two defects of SDPC: (1) the cost of extra computations and bits; (2) the non-stationary of predictive model, the ST-Q is trying to overcome them in basis of SDPC. The proposed method consists of two phases. In the first phase, we compute the difference between CS measurements of consecutive frames for removing the temporal redundancy in the measurement domain. In the second phase, we can further remove the spatial redundancy of the difference by using a spatially predictive strategy based on median filter [28]. The proposed quantization method does not require any supplementary bits, and the computations are also effectively limited by the fixed predictive model.

Motion-Aligned Reconstruction (MA-R). On the basis of the MH-based method, the proposed joint reconstruction algorithm has ability to select highly-correlated temporal neighborhoods of each to-be-reconstructed block in terms of motion information between the adjacent reconstructed frames. The motivation of MA-R is derived from the enforcement of temporal correlation between the to-be-reconstructed block and its MH predictors. The process of motion estimation is implemented in the measurement domain, and the background extraction and 3-D Recursive Search (3DRS) [29] are also used to reduce the computational complexity when computing the motion vector field of to-be-reconstructed video frame.

The remainder of this paper is organized as follows. The overall architecture of the proposed CVS system is firstly presented in Section 2, and the implementation of Space-Time Quantization (ST-Q) is described in detail in Section 3. Then, Section 4 presents how to exploit motion information to realize the Motion-Aligned Reconstruction (MA-R). Experimental results are reported in Section 5 to evaluate the performance of proposed ST-Q and MA-R. Finally, the conclusion is made in Section 6.

2. Framework Overview

The block diagram of CVS system with Space-Time Quantization (ST-Q) and Motion-Aligned Reconstruction (MA-R) is depicted in **Fig. 1**. Due to the feature that the ST-Q quantizes the frame difference, the encoder of proposed CVS system is different from the traditional system (e.g., DISCOS) which directly divides the input video frames into non-key frames and key frames, and but to realize the BCS measuring of non-key frames and key frames by the two-layer structure. At the base layer, each block x_k of size $B \times B$ ($N = B^2$) in the input video frame x is measured by using the measurement matrix Φ_{NK} of size $M_{NK} \times N$ which constructed according to the subrate $S_{NK} = M_{NK}/N$ of non-key frame, and the corresponding measurement vector y_k of each block can be computed as follows,

$$y_k = \Phi_{NK} \cdot x_k \quad k = 1, 2, \cdots, K . \tag{1}$$

At the refinement layer, the key frame x_K is extracted periodically from the video sequence, and then measured again to satisfy the subrate $S_K = M_K/N$ of key frames based on the existing measurement vector obtained at the base layer. According to the number $M_{Sup} = M_K - M_{NK}$ of supplementary measurements, we construct a new measurement matrix Φ_{Sup} of size $M_{Sup} \times N$, and the corresponding measurement vector $y_{Sup,k}$ of each block $x_{K,k}$ can be computed as follows,

$$y_{Sup,k} = \Phi_{Sup} \cdot x_{K,k} \quad k = 1, 2, \cdots, K . \tag{2}$$

After BCS measuring, the measurement vectors of key frames and non-key frames are
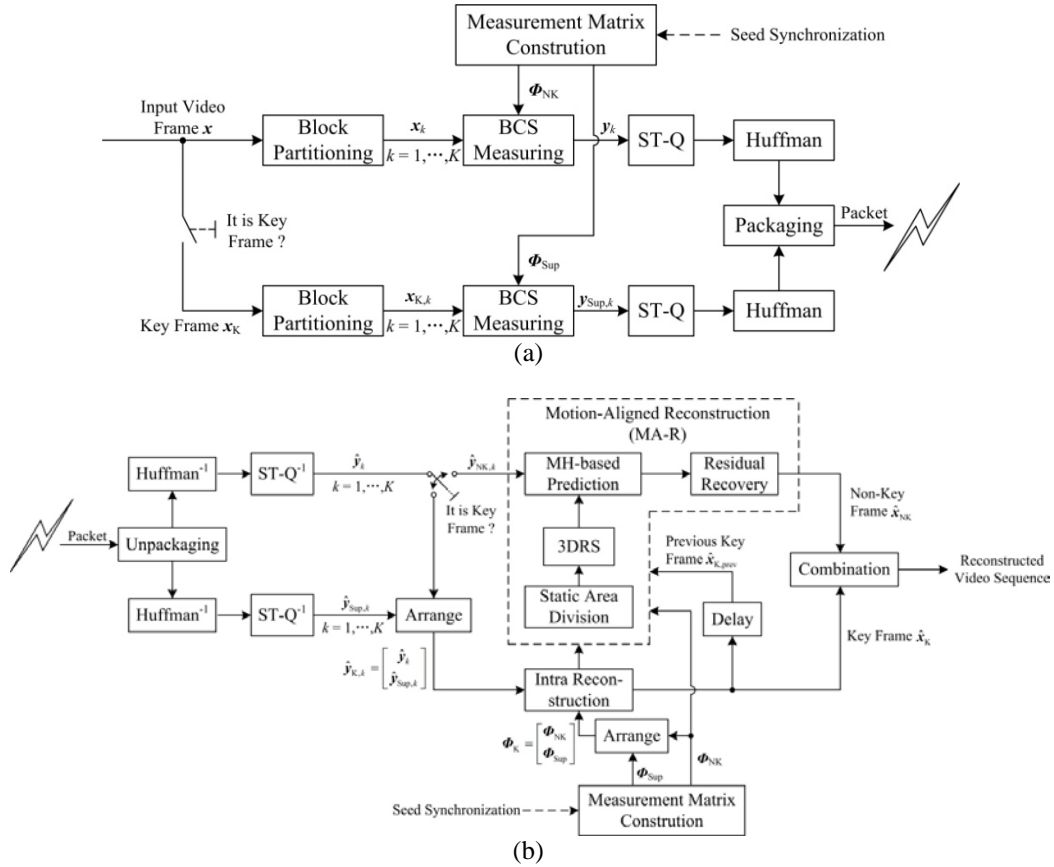
Fig. 1. Block diagram of CVS system with Space-Time Quantization (ST-Q) and Motion-Aligned Reconstruction (MA-R). (a) Encoder. (b) Decoder.

respectively input into the ST-Q, and then compressed to bits by the Huffman coding. Finally, these bits and some important decoding parameters are packed into a packet and sent to the decoder. The implementation of ST-Q will be described in Section 3.

At the decoder, the priority is to generate the measurement vectors $\hat{\mathbf{y}}_{K,k}$ and $\hat{\mathbf{y}}_{NK,k}$ of key frame and non-key frame respectively. From the de-quantized measurement vectors $\hat{\mathbf{y}}_k$ ($k = 1,...,K$) at the base layer, the measurement vector $\hat{\mathbf{y}}_{NK,k}$ of non-key frame can be easily extracted by skipping the measurement vector of key frame periodically. Once the current measurement vector $\hat{\mathbf{y}}_k$ is from the key frame, the $\hat{\mathbf{y}}_{K,k}$ can be generated by arranging the de-quantized supplementary measurement vector $\hat{\mathbf{y}}_{Sup,k}$ in rows behind $\hat{\mathbf{y}}_k$, i.e.,

$$\hat{\mathbf{y}}_{K,k} = \begin{bmatrix} \hat{\mathbf{y}}_k \\ \hat{\mathbf{y}}_{Sup,k} \end{bmatrix}. \tag{3}$$

Similarly, the measurement matrix of each block in key frame can be constructed as follows,

$$\boldsymbol{\Phi}_K = \begin{bmatrix} \boldsymbol{\Phi}_{NK} \\ \boldsymbol{\Phi}_{Sup} \end{bmatrix}. \tag{4}$$

Note that we will make sure to use the same measurement matrix at the both encoder and decoder by synchronously updating the seed (initial state) of pseudorandom generator. After

generating $\hat{\boldsymbol{y}}_{\mathrm{K},k}$ and $\hat{\boldsymbol{y}}_{\mathrm{NK},k}$, the key frame can be reconstructed independently by using the following model,

$$\hat{\boldsymbol{x}}_{\mathrm{K}} = \arg \min_{\boldsymbol{x}} \left\{ \left\| \hat{\boldsymbol{y}}_{\mathrm{K}} - \boldsymbol{\Theta}_{\mathrm{K}} \boldsymbol{E} \cdot \boldsymbol{x} \right\|_{2} + \lambda \left\| \boldsymbol{\Psi} \cdot \boldsymbol{x} \right\|_{1} \right\}, \tag{5}$$

where

$$\hat{\boldsymbol{y}}_{\mathrm{K}} = \begin{bmatrix} \hat{\boldsymbol{y}}_{\mathrm{K},1} \\ \hat{\boldsymbol{y}}_{\mathrm{K},2} \\ \vdots \\ \hat{\boldsymbol{y}}_{\mathrm{K},K} \end{bmatrix}, \ \boldsymbol{\Theta}_{\mathrm{K}} = \begin{bmatrix} \boldsymbol{\Phi}_{\mathrm{K}} & & & 0 \\ & \boldsymbol{\Phi}_{\mathrm{K}} & & \\ & & \ddots & \\ 0 & & & \boldsymbol{\Phi}_{\mathrm{K}} \end{bmatrix}, \ \boldsymbol{E} \cdot \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_{\mathrm{K},1} \\ \boldsymbol{x}_{\mathrm{K},2} \\ \vdots \\ \boldsymbol{x}_{\mathrm{K},K} \end{bmatrix}, \tag{6}$$

$\boldsymbol{\Psi}$ is the sparse representation matrix of video frame $\boldsymbol{x}$. This way is similar to the intra model in traditional video codec, and therefore it is called as intra reconstruction. Some effective intra-reconstruction algorithms can be used to produce the key frame, e.g., [30] proposed the collaborative sparsity based recovery, [31] proposed to reconstruct frame by using the adaptively learned sparsifying basis, and [32] proposed the recovery method to explore the structured Laplacian sparsity of DCT coefficients, etc. In this paper, considering that the above-method algorithms have a high computational complexity because of their requirements to train the sparse representation matrix online, the model (5) uses the the fixed representation matrix based BCS-SPL-DDWT algorithm proposed by [15] to reconstruct each key frame at a high subrate. The proposed MA-R algorithm of non-key frame is realized by residual recovery coupled with SI generation as follows,

$$\hat{\boldsymbol{y}}_{\mathrm{R},k} = \hat{\boldsymbol{y}}_{\mathrm{NK},k} - \boldsymbol{\Phi}_{\mathrm{NK}} \hat{\boldsymbol{x}}_{\mathrm{SI},k} = \boldsymbol{\Phi}_{\mathrm{NK}} \left(\boldsymbol{x}_{\mathrm{NK},k} - \hat{\boldsymbol{x}}_{\mathrm{SI},k} \right), \tag{7}$$

$$\hat{\boldsymbol{r}}_{\mathrm{NK},k} = \arg \min_{\boldsymbol{r}} \left\{ \left\| \hat{\boldsymbol{y}}_{\mathrm{R},k} - \boldsymbol{\Phi}_{\mathrm{NK}} \cdot \boldsymbol{r} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\Psi} \cdot \boldsymbol{r} \right\|_{1} \right\}, \tag{8}$$

$$\hat{\boldsymbol{x}}_{\mathrm{NK},k} = \hat{\boldsymbol{x}}_{\mathrm{SI},k} + \hat{\boldsymbol{r}}_{\mathrm{NK},k}, \tag{9}$$

where λ is a balance factor. The model (8) is still be solved by BCS-SPL-DDWT, and the SI $\hat{\boldsymbol{x}}_{\mathrm{SI},k}$ of current block which can be obtained by using MH-based prediction [16], i.e.,

$$\hat{\boldsymbol{w}}_{\mathrm{NK},k} = \arg \min_{\boldsymbol{w}} \left\{ \left\| \hat{\boldsymbol{y}}_{\mathrm{NK},k} - \boldsymbol{\Phi}_{\mathrm{NK}} \boldsymbol{H}_{\mathrm{K},k} \cdot \boldsymbol{w} \right\|_{2}^{2} + \eta \left\| \boldsymbol{\Gamma} \cdot \boldsymbol{w} \right\|_{2}^{2} \right\}, \tag{10}$$

$$\hat{\boldsymbol{x}}_{\mathrm{SI},k} = \boldsymbol{H}_{\mathrm{K},k} \hat{\boldsymbol{w}}_{\mathrm{NK},k}, \tag{11}$$

where η is a balance factor, $\boldsymbol{H}_{\mathrm{K},k}$ is a matrix of size $N \times L$ whose columns $\boldsymbol{h}_i$ ($i = 1,\ldots,L$) are the rasterizations of temporal neighborhoods in the reconstructed key frames along the motion trajectories of current block, i.e., $\boldsymbol{H}_{\mathrm{K},k} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_L]$, and $\boldsymbol{\Gamma}$ is a diagonal Tikhonov matrix in the form of

$$\boldsymbol{\Gamma} = \begin{bmatrix} \left\| \hat{\boldsymbol{y}}_{\mathrm{NK},k} - \boldsymbol{\Phi}_{\mathrm{NK}} \boldsymbol{h}_1 \right\|_{2} & & 0 \\ & \ddots & \\ 0 & & \left\| \hat{\boldsymbol{y}}_{\mathrm{NK},k} - \boldsymbol{\Phi}_{\mathrm{NK}} \boldsymbol{h}_L \right\|_{2} \end{bmatrix}. \tag{12}$$

The key point of proposed MA-R algorithm is to compute the motion vector field of the to-be-reconstructed frame by using the neighboring reconstructed key frames. To not introduce excessive computations, we design a method to extract the static area from the two consecutive reconstructed key frames, and use the 3DRS strategy to estimate the motion vectors of non-static area, which will be discussed detailedly in Section 4.

3. Space-Time Quantization (ST-Q)

For many types of video (e.g., surveillance video), there will be very little difference between consecutive frames, and besides each block of frame difference has the intrinsic spatial correlation with its neighboring blocks. The Johnson-Lindenstrauss (JL) lemma [1] [33] guarantees that the above space-time characteristic is also be maintained in the measurement domain. It is obvious that the difference after removing space-time redundancy can be represented using far fewer bits than the original CS measurement vector, and meanwhile the Laplacian distribution of difference is also helpful to suppress the quantizing noise of SQ [2]. How to remove the space-time redundancy existing in the each block of CS measurements？ The ST-Q will give an answer below.

The proposed ST-Q quantizer architecture is depicted in **Fig. 2**. We firstly compute the difference vector d_t between the CS measurement vectors of current frame y_t and previous frame $\hat{y}_{t-1}$ to remove the temporal redundancy, and then use the Median Filter based Predictive Quantization (MFPQ) to further reduce the spatial redundancy between blocks in the difference vector d_t. Along the temporal direction, a feedback loop de-quantizes the quantization index i_t to produce the quantized difference vector $\hat{d}_t$ such that $\hat{y}_t = \hat{y}_{t-1} + \hat{d}_t$, and then get the previous quantized measurement vector $\hat{y}_{t-1}$ with a delay buffer. The block diagram of MFPQ is shown in **Fig. 3**, in which it is important how to generate the prediction of current difference vector d_t. As shown in **Fig. 3(c)**, we will save the four quantized difference vectors $\hat{d}_{t,k\cup 1}$, $\hat{d}_{t,k\cup 2}$, $\hat{d}_{t,k\cup 3}$ and $\hat{d}_{t,k\cup 4}$ of spatially neighboring blocks in the register for the difference vector $d_{t,k}$ of current block, and these neighboring difference vectors construct the candidate set CS to generate the prediction of current block.

Different from the SDPC proposed in [18], the Median Filter (MF) is applied to the components $\hat{d}_{t,k\cup i}(m)$ (i = 1, 2, 3, 4) of vectors in candidate set CS, and the output is just the prediction $\hat{d}_{P,k}$ of the difference vector $d_{t,k}$ of current block, i.e.,

$$\hat{d}_{P,k}(m) = median([\hat{d}_{t,k\cup 1}(m), \hat{d}_{t,k\ 2}(m), \hat{d}_{t,k\cup 3}(m), \hat{d}_{t,k\ 4}(m)]). \qquad (13)$$

where *median*($\cdot$) is the MF operation. Some defects of SDPC motivate us to make a prediction by using MF. The SDPC found that the fixed directional model used in DPCM violates the fact that the non-stationary of natural images, and therefore it introduces the four predictive directional models (Vertical, Horizontal, DC and Diagonal) and selects the optimal directional model from them to remove the spatial redundancy as much as possible. Although its superior
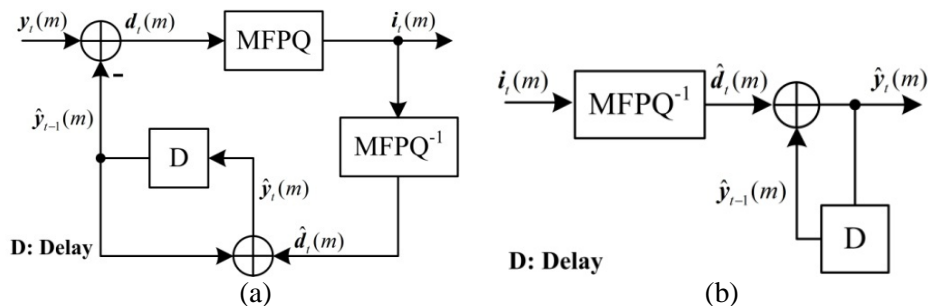

Fig. 2. The proposed ST-Q quantizer architecture. (a) quantization. (b) de-quantization. Note that *m* denotes the component of each vector.

---

[1] In brief, the JL lemma holds that the L points in $\mathbf{R}^N$ can be projected into a K-dimensional subspace while approximately maintaining pairwise distances as long as $K \geq O(\log L)$.
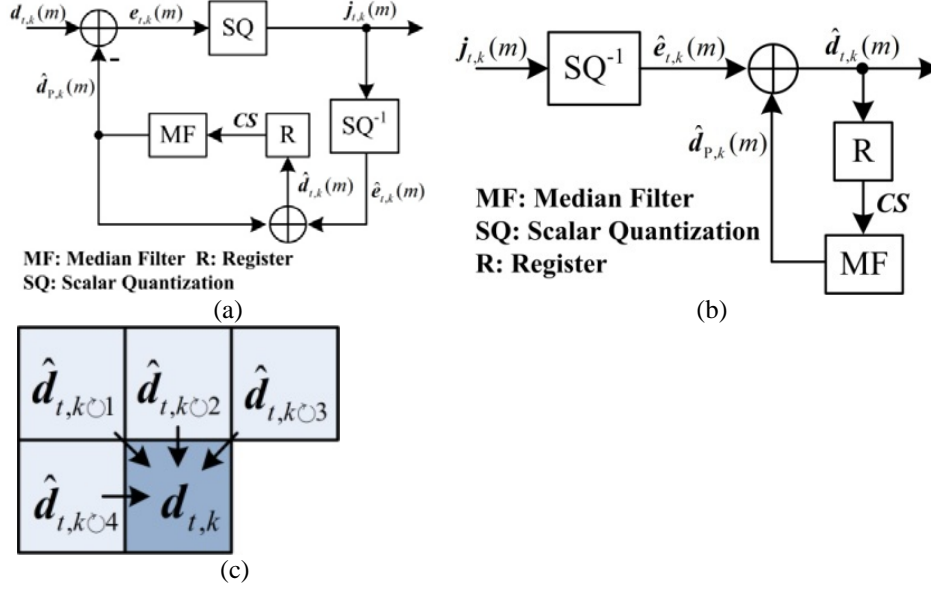
Fig. 3. Block diagram of median filter based predictive quantization (MFPQ). (a) quantization. (b) de-quantization. (c) position of candidates saved in register. Note that m denotes the component of each vector.

performance than DPCM, this improvement is traded at the cost of extra computations and bits. Importantly, it results in the complete failure of de-quantization to lose the bits representing the flag of the best predictive model in the process of transmission, and consequently the practicability of SDPC has yet to be investigated. The use of MF cannot produce the extra bits so that it can avoid the risk of losing the key flags, and besides the MF has also a low computational complexity than the selection of optimal directional model. A fundamental property of MF states the prediction $\hat{d}_{P,k}(m)$ statisfies

$$\sum_{i=1}^{4}\left\|\hat{d}_{P,k}(m)-\hat{d}_{t,k爇j}(m)\right\|_{p} \leq \sum_{i=1}^{4}\left\|\hat{d}_{t,k_{j}}(m)-\hat{d}_{t,k_{i}}(m)\right\|_{p} \quad j=1,2,3,4 , \qquad (14)$$

which guarantees that a larger variability of spatial correlation cannot happen suddenly. Therefore, for the predictive accuracy, the MFPQ can still provide a comparable result with SDPC.

To demonstrate the better performance of proposed MFPQ, we make a further quantitative comparison with SDPC for correlation and computations respectively. We use the correlation coefficient [17] to measure the correlation of measurement vector y_k of current block and its prediction $y_{P,k}$, defined as

$$\rho_k = \frac{y_k^{\mathrm{T}} y_{P,k}}{\left\|y_k\right\|_2 \left\|y_{P,k}\right\|_2}. \qquad (15)$$

For each block, the DPCM uses directly the measurement vector y_{k-1} of previous block as the prediction, the SDPC selects the optimal prediction from four directional models by using the least square error criterion, and the proposed MFPQ computes the median value of components in measurement vectors of neighboring blocks as the prediction. **Table 1** presents the average correlation coefficients for DPCM, SDPC and MFPQ, which are computed by Eq. (15) over all 16×16 blocks for three 512×512 test images at the different subrates. There is no doubt that the SDPC obtains the largest correlation coefficient among all methods depending

Table 1. Average correlation coefficients in measurement domain for various quantization methods

Subrate	0.10				0.25			
Image	*Lenna*	*Peppers*	*Clown*	Avg.	*Lenna*	*Peppers*	*Clown*	Avg.
DPCM	0.9610	0.9396	0.8446	0.9151	0.9576	0.9338	0.8291	0.9068
SDPC	0.9710	0.9616	0.8900	0.9409	0.9674	0.9558	0.8775	0.9336
MFPQ	0.9653	0.9495	0.8516	0.9221	0.9608	0.9425	0.8343	0.9125
Subrate	0.35				0.50			
Image	*Lenna*	*Peppers*	*Clown*	Avg.	*Lenna*	*Peppers*	*Clown*	Avg.
DPCM	0.9630	0.9401	0.8423	0.9151	0.9653	0.9432	0.8488	0.9191
SDPC	0.9713	0.9608	0.8859	0.9393	0.9731	0.9621	0.8902	0.9418
MFPQ	0.9666	0.9498	0.8510	0.9225	0.9688	0.9526	0.8584	0.9266

Table 2. Comparisons of computational complexity for prediction scheme when quantizing each block

Quantization Method	Multiplications	Additions	Comparisons
DPCM	0	0	0
SDPC	$9M$-4	M	3
MFPQ	0	0	$6M$

*M denotes the length of measurement vector of each block

on the more computations and bits, and the DPCM shows the worst performance due to the fixed predictive direction. However, the average correlation coefficient of MFPQ obtains the moderate values between those of DPCM and SDPC at any subrate, e.g., the MFPQ has 0.0074 and -0.0171 gains on average when compared with DPCM and SDPC respectively at the subrate of 0.35. Although the average correlation coefficients of MFPQ are inferior than those of SDPC, it can be seen from **Table 2** that the proposed MFPQ only requires $6M$ comparison operators when making a prediction for the measurement vector of length M, and its computations are far below those of SDPC including the more complex floating-point addition and multiplication. Therefore, the proposed MFPQ makes a good tradeoff between predictive performance and computational complexity.

4. Motion-Aligned Reconstruction (MA-R)

It is helpful for improving the accuracy of SI in the reconstruction of non-key frame to exploit motion information between the two consecutive key frames because a high correlation exists between the current block and its motion-aligned block, which motivates us to design the MA-R algorithm that selects all L ($= W^2$) temporal neighborhoods in the $W \times W$ windows centered on the motion-aligned block $x_{c,k}$ of current block $x_{NK,k}$ to construct the MH matrix $H_{K,k}$ in Eq. (11), and the MA predictive model is shown in **Fig. 4**. The key of MA-R algorithm is how to compute the motion vector field of each non-key frame, which is described as follows.

Firstly, we divide a large static area in several successive non-key frames from their adjacent reconstructed key frames to avoid the unnecessary search points for static blocks. The similarity between their adjacent reconstructed key frames is measured block-by-block as follows,

$$\gamma_{\mathrm{NK},k} = \left(1-\alpha\right)\left\|\hat{\boldsymbol{x}}_{\mathrm{K},k}^{\mathrm{prev}} - \hat{\boldsymbol{x}}_{\mathrm{K},k}^{\mathrm{next}}\right\|_1 + \alpha\left\|\mathrm{D}\left(\hat{\boldsymbol{x}}_{\mathrm{K},k}^{\mathrm{prev}}\right) - \mathrm{D}\left(\hat{\boldsymbol{x}}_{\mathrm{K},k}^{\mathrm{next}}\right)\right\|_1, \tag{16}$$
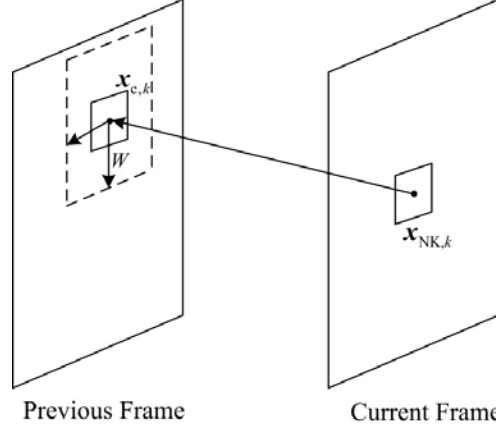
Fig. 4. MA predictive model illustration. $\boldsymbol{x}_{c,k}$ denotes the motion-aligned block.

where $\gamma_{NK,k}$ is the similarity measure of the k-th block in the non-key frame, $\hat{\boldsymbol{x}}_{K,k}^{prev}$ and $\hat{\boldsymbol{x}}_{K,k}^{next}$ denotes the reconstructed blocks in the previous and next key frames respectively, $\|\cdot\|_1$ is the l_1 norm, and α is a weighting factor. The first item of Eq. (16) is to directly compute the sum of the absolute differences between pixel values, and but it cannot reveal the similarity between blocks completely. Considering the human eyes' sensitivity to the edge of object, the block-matching error of edge region is more easy to reflect the similarly between blocks, and therefore we construct the operator $D(\cdot)$ to extract the medium frequency DCT coefficient including edge information, and form the second item of Eq. (16) to be the auxiliary of measuring the similarity between blocks. At $\alpha = 0$, only the luminance differences are considered in the similarity measure. As α gets larger, the DCT coefficient differences play a more important role to match the edge information. However, too large an α increases the effects of noise components, degrading the matching performance. In this work, α is fixed to 0.3 experimentally. By implementing the operator $D(\cdot)$, in terms of the spatial position of 2D-DCT coeffient matrix, the range from the $0.382 \times B^2$-th DCT coefficient to $0.618 \times B^2$-th DCT coefficient in the zig-zag order can be extracted as the medium frequency coefficients of each block with size of $B \times B$. To improve the block-matching accuracy, we use the overlapped block with size of $1.5B \times 1.5B$ to compute the matching error due to the fact that the block with a large size contains the more features. After getting the similarity measure of each block, the block classification map of non-key frame is generated by a hard-thresholding decision as follows,

$$\boldsymbol{map}_{NK,k}(\boldsymbol{s}) = \begin{cases} 1, & \gamma_{NK,k} \leq T \\ 0, & \gamma_{NK,k} > T \end{cases}, \tag{17}$$

where s denotes the spatial position of the k-th block, and T is the threshold. If $\boldsymbol{map}_{NK,k}(s)$ is labeled as 1, the k-th block is a static block, and otherwise the k-th block is a non-static block. The blocks existing in the boundary region rarely contain movements (as shown in **Fig. 5**), and therefore they can be used to determine the value of threshold T. Suppose n_b is the total number of boundary blocks, we sort their similarity measures in ascending order, and then the threshold T is set to the average value of remaining similarity measures after removing the first and last $0.15n_b$ values respectively.

Secondly, we design a smoothing operator to remove the isolated blocks in the block classification map. The occurrence of isolated blocks derives from the fact that the variational

luminance and texture details on the surface of object leads to the problem of mismatch. As shown in **Fig. 6(a)**, the isolated blocks cause the discontinuity in static area or non-static area, and they can be identified according to any one of the following conditions when the $\textit{map}_{\mathrm{NK},k}(s)$ is labeled as 0,

<div align="center">The Previous Key Frame   The Next Key Frame</div>

Fig. 5. The comparison of blocks existing in the boundary region for the previous and next key frames. The blocks surrounded by red line lie in the boundary region of video frame.

- The 8 neighbors of $\textit{map}_{\mathrm{NK},k}(s)$ are all labeled as 1(the first row of **Fig. 6(b)**);
- The only one in the 4 neighbors of $\textit{map}_{\mathrm{NK},k}(s)$ is labeled as 0, and the four diagonal neighbors of $\textit{map}_{\mathrm{NK},k}(s)$ are all labeled as 1 (the second row of **Fig. 6(b)**);
- The 4 neighbors of $\textit{map}_{\mathrm{NK},k}(s)$ are all labeled as 1, and there are two at most to be labeled as 0 in the four diagonal neighbors of $\textit{map}_{\mathrm{NK},k}(s)$ (the third row of **Fig. 6(b)**).

If the $\textit{map}_{\mathrm{NK},k}(s)$ is labeled as 1, all logic arrays in **Fig. 6(b)** are required to be inverted. After determining the positions of isolated blocks, the $\textit{map}_{\mathrm{NK},k}(s)$ is smoothed by using a 3×3 mask as follows,

$$\textit{map}_{\mathrm{NK},k}(s) = \begin{cases} 1, & R \geq 10 \\ 0, & R < 10 \end{cases}, \tag{18}$$

$$R = \sum_{p=-1}^{1} \sum_{q=-1}^{1} mask(p,q) \cdot \textit{map}_{\mathrm{NK},k}(s_x + p, s_y + q), \tag{19}$$

$$\textbf{mask} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{20}$$

where $s = (s_x, s_y)$ denotes the position of isolated block. As shown in **Fig. 6(c)**, it can be seen that the isolated blocks are completely eliminated after smoothing block classification map, and the static blocks and non-static blocks are respectively segmented into a set of connected regions.

After determining the static blocks in the several successive non-key frames, we finally compute the motion vectors of the non-static blocks. As a consequence of JL lemma, the full-search based block matching algorithm can be realized in the measurement domain [25], that is,

$$\textbf{x}_{\mathrm{c},k} = \arg\min_{\textbf{x} \in V_k} \left\| \hat{\textbf{y}}_{\mathrm{NK},k} - \boldsymbol{\Phi}_{\mathrm{NK}} \textbf{x} \right\|_2, \tag{21}$$

where $\hat{\textbf{y}}_{\mathrm{NK},k}$ is the de-quantized measurement vector of current non-static block, V_k denotes

the candidate set including all blocks in the search area, x denotes the matching block candidate, and $x_{c,k}$ denotes the best matching block of current non-static block in the previous key frame. However, the full search not only introduces the excessive computational complexity because it requires to traverse all possible candidates within the search area, but
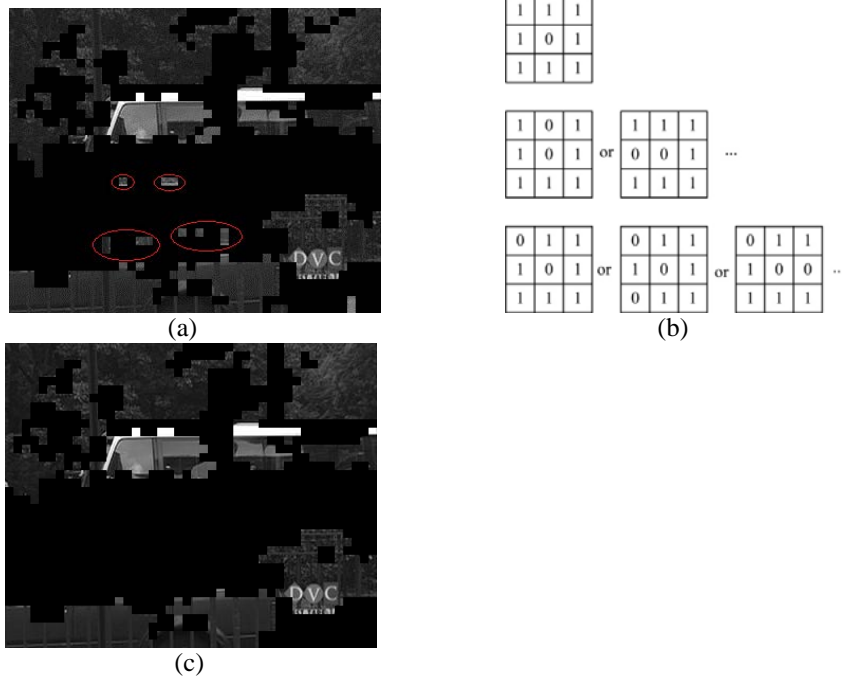

(a)

(b)

(c)

Fig. 6. Static area illustration before and after smoothing block classification map. The isolated blocks are highlighted in the red circle. (a) static area before smoothing. (b) logic arrays used to identify isolated block. (c) static area after smoothing.

also it results in some inaccurate motion vectors without the smoothness constraint of motion vector field, To overcome the defects of full search, the 3DRS is used to construct the candidate block set V_k, in which each matching block candidate is extracted by using the candidate motion vectors of current non-static block. As shown in **Fig. 7**, the candidate motion vector set C of current non-static block is composed by the seven candidate motion vectors (the coordinate of current non-static block is defined as B): zero vector 0; the motion vectors of spatio-neighboring block locations S_a (upper) and S_b (left); the motion vectors of temporal-neighboring block locations T_a (lower) and T_b (right); the update motion vectors of spatio-neighboring block locations U_a (upper-left) and U_b (upper-right), that is,

$$C = \{0, MVF(S_a), MVF(S_b), MVF^K(T_a), MVF^K(T_b), MVF(U_a) \\ + R_a, MVF(U_b) + R_b\} \ R_a, R_b \in US \tag{22}$$

$$US = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \end{pmatrix} \right\}, \tag{23}$$

where $MVF(\cdot)$ represents the motion vector field of the current non-key frame, and $MVF^K(\cdot)$ represents the motion vector field of the previous key frame. This approach imposes the smoothness constraint implicitly through predictive search, and therefore it guarantees the accuracy of motion vector with a low complexity.

By the above process, we can get the motion vector field of each non-key frame. In terms of

the motion vector field, we find the motion-aligned block of to-be-reconstructed block, and then the corresponding MH matrix can be constructed whether non-static block or static block. Implementing Eq. (10) and (11), we predict the SI of each non-key frame, and reconstruc the residual frame between the SI and its original by Eq. (7) and (8). Finally, we combine the residual frame into SI to generate the reconstructed frame by using Eq. (9).
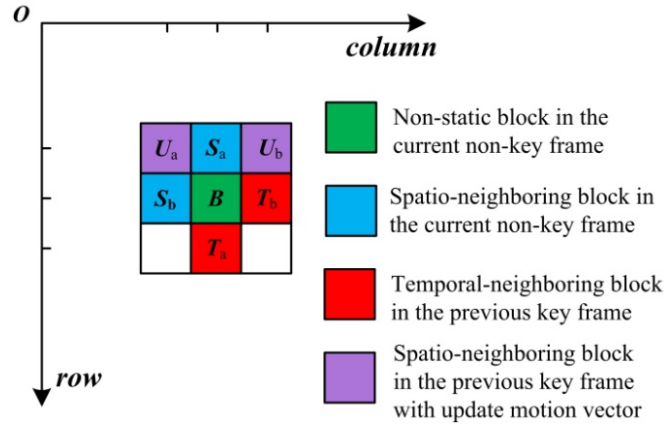

Fig. 7. Relative positions of candidate motion vectors used in 3DRS.

Table 3. Average MSE and bitrate (in kbps) comparison of various quantization methods for all test sequences at the different subrates.

Subrate	Quantization Method of CS Measurements							
	Uniform SQ		DPCM-plus-SQ [17]		SDPC [18]		ST-Q	
	MSE	Bitrate	MSE	Bitrate	MSE	Bitrate	MSE	Bitrate
0.1	1.1483	3473.05	1.1483	2610.30	1.1441	2738.43	0.5756	2932.10
0.2	1.3308	5515.93	1.3308	4005.37	1.3265	4284.59	0.5896	4118.84
0.3	1.3316	7290.15	1.3316	5292.36	1.3274	5593.58	0.5942	5198.71
0.4	1.3315	9301.04	1.3315	6756.31	1.3277	7097.66	0.5989	6448.61
0.5	1.3407	11006.62	1.3407	7989.64	1.3376	8380.76	0.5976	7497.47
Avg.	1.2966	7317.36	1.2966	5330.80	1.2927	5619.00	0.5912	5239.15

5. Experimental Results

In this section, various experiments are conducted to evaluate the performance of proposed CVS system[2]. First, we compare the performances of proposed ST-Q with those of the uniform SQ, the DPCM-plus-SQ proposed by [17] and the SDPC proposed by [18]. Second, the performance of proposed MA-R algorithm is evaluated by using Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) [34], and the comparisons with the recovery algorithm of DISCOS system [24] and MH-based Predictive-Residual (MHPR) reconstruction [25] are also presented. Finally, the rate-distortion performance of proposed CVS system is compared with those of CS-KLT video codec (one of state-of-the-art CVS system) proposed in [35] and CS-based video codec in our previous work [36].

---

[2] The MATLAB source code of the proposed CVS system including ST-Q and MA-R can be downloaded from the website: https://sites.google.com/site/draner358.
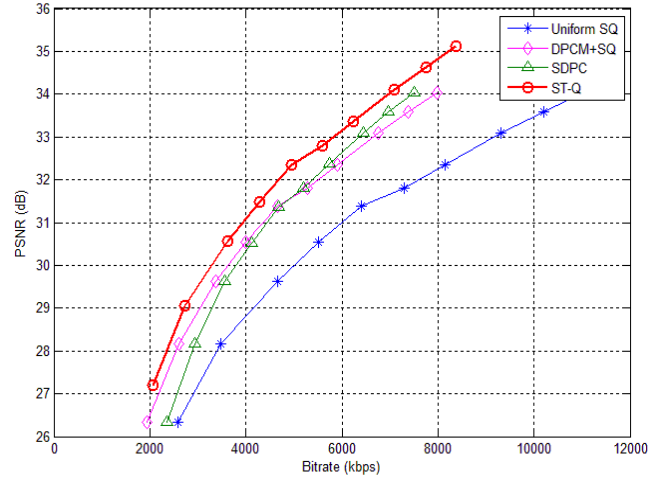
Fig. 8. Rate-distortion curve for various quantization methods according to the average PSNR values and bitrates on all test sequences.

The eight test sequences with CIF resolution of 352×288 pixels and frame rate of 30 fps are used in the experiments, and these are *Foreman*, *Mobile*, *Football*, *Highway*, *Container*, *Bus*, *Soccer* and *Stefan*. At the encoder side of proposed CVS system, the block size $B×B$ of each frame is 8×8, the bitdepth b of each quantization is set to 8 in order to guarantee a fair comparison among all quantization methods, the subrate S_K of key frame is set to 0.7, the subrate S_{NK} of non-key frame varies from 0.1 to 0.5, and the measurement matrices Φ_K and Φ_{NK} both use the structurally random Hadamard matrix proposed by [37]. The key frame frequency is set to one key frame for every 10 frames, and the first 100 frames of each test sequence are input into the proposed CVS system. For the construction of MH matrix $H_{K,k}$, the radius W of temporal neighboring window is set to $3B/8$, and the balance factor η is set to 0.0625. The compared algorithm uses their original parameter settings so that they provides the best results respectively. All experiments are implemented on a PC with an Intel Core i7 CPU at 3.6 GHz and 8GB RAM.

5.1 Quantization Performances

The uniform SQ, DPCM-plus-SQ, SDPC and proposed ST-Q are respectively used in the proposed CVS system to encode the first 100 frames of each test sequence, and the average Mean Square Error (MSE) of quantization and bitrate in kilo-bits per second (kbps) on all test sequence are presented in **Table 3**. It can be seen that the MSE of proposed ST-Q is more smaller than others at any subrate, and it decreases 53.49% on average over the SDPC, which benefits from the fact that the spatial-temporal prediction of ST-Q reduces the number of nonzero values in measurement residuals so as to suppresses the accumulation of quantization errors on the large values. Meanwhile, we can also observe from **Table 3** that the ST-Q requires few bits to encode measurements when compared with the other three methods, which indicates that the ST-Q effectively removes the spatial-temporal redundancy of measurements in video sequences. **Fig. 8** presents the rate-distortion performance of the proposed CVS system combined with the different quantization methods, in which the average PSNR values and bitrates on all test sequences are computed for the various quantization methods, and it sufficiently demonstrates that the superiority of ST-Q over the other compared methods.

5.2 Reconstruction Comparison

Table 4. Average PSNR (in dB), SSIM and Processing Time (s/frame) comparison of various reconstruction methods for all test sequences at the different subrates

Subrate	Reconstruction Method								
	DISCOS [24]			MHPR [25]			MA-R		
	PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time
0.1	26.39	0.7391	11.25	27.84	0.7984	1.97	28.24	0.8111	3.14
0.2	28.31	0.8093	8.39	30.16	0.8740	2.03	30.68	0.8814	3.70
0.3	28.98	0.8325	7.89	31.50	0.9047	2.07	31.99	0.9082	3.88
0.4	29.86	0.8583	8.44	32.90	0.9292	2.11	33.32	0.9300	3.97
0.5	30.61	0.8791	8.82	33.94	0.9451	2.16	34.31	0.9446	3.90
Avg.	28.83	0.8237	8.96	31.27	0.8903	2.07	31.71	0.8951	3.72

Fig. 9. Visual comparison of the reconstructed 2-th frame of *Foreman* by different methods ($S_{NK} = 0.1$). From left to right: DISCOS [24], MHPR [25], and the proposed MA-R.

Fig. 10. Visual comparison of the reconstructed 2-th frame of *Football* by different methods ($S_{NK} = 0.1$). From left to right: DISCOS [24], MHPR [25], and the proposed MA-R.

Next, the performance of proposed MA-R algorithm is evaluated from objective and subjective views by comparing it with the recovery algorithm of DISCOS [24] and MHPR reconstruction algorithm [25]. The average PSNR results reconstructed by various methods on all test sequences are provided in **Table 4**, and it can be seen that the proposed MA-R algorithm obtains the higher PSNR value than other two compared methods at any subrate, in which its average PSNR gains are 2.88 and 0.44 dB when compared with DISCOS and MHPR respectively. The similar results can also be achieved from the viewpoint of the SSIM, which can be observed in **Table 4**, and the MA-R algorithm obtains the 0.0714 and 0.0048 SSIM gains on average when compared with DISCOS and MHPR respectively. The inferior objective evaluation of DISCOS is attributed to its way of sparsity-constrain prediction which is not suitable to express the relationship between the current reconstructed blocks and its

temporal neighbors, and however the MH prediction used in MHPR and MA-R can better reveal the temporal correlation between the neighboring video frames. Due to the exploitation of motion information between frames, the reconstruction performance of MA-R algorithm acquires the effective improvement when compared with the MHPR algorithm without the motion estimation. The average processing time to reconstruct a video frame for each method is also provided in **Table 4**. It can be seen that the proposed MA-R algorithm has a moderate computational complexity, and its reconstruction time is slightly over that of MHPR, which is derived from the computational burden of estimating the motion vector field. Because we simplify the procedure of motion estimation, and the MA-R algorithm has only 1.65 s time
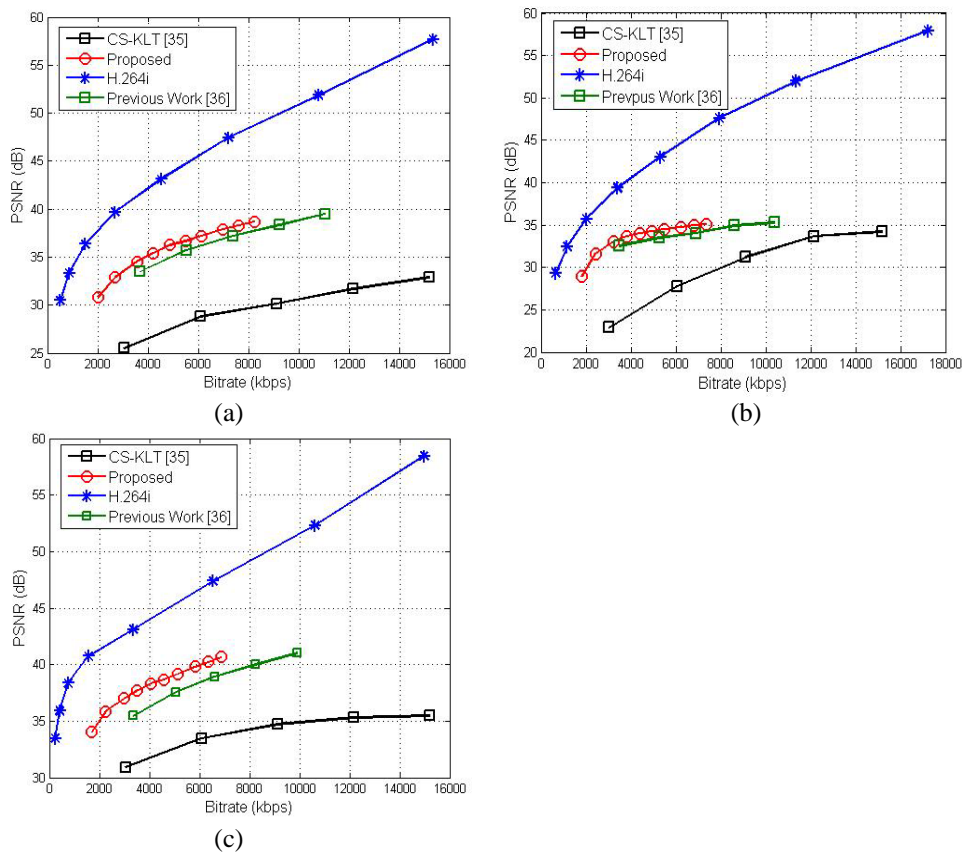

(a)   (b)

(c)

Fig. 11. Rate-distortion curve for H.264/AVC-Intra (H.264i), CS-KLT codec [35], our previous work [36], and the proposed CVS system: (a) *Foreman*. (b) *Highway*. and (c) *Container*.

gains on average when compared with MHPR algorithm. We also visually assesses some video frames constructed by different methods. **Figs. 9-10** show the reconstructed frames of *Foreman* and *Football* at the subrate $S_{NK} = 0.1$ respectively. It can be observed that the proposed MA-R method provides the pleasant results for each test sequence, e.g., for *Foreman* sequences as shown in **Fig. 9**, the face recovered by these competing methods contains many annoying artifacts, and but the proposed method clearly perceives the face.

5.3 Rate-Distortion Performances

Fig. 11 compares the rate-distortion performances, averaged over the first 100 frames of *Foreman*, *Highway* and *Container* sequences respectively, of the Intra coded results by the H.264/AVC JM9.5 software (H.264i), the CS-KLT codec proposed by [35], the CS-based

video codec in our previous work [36] and the proposed CVS system. The CS-KLT codec implements the motion estimation and motion compensated interpolation at the decoder by the sparsity-aware reconstruction using interframe Karhunen-Loève Transform basis, and it exhibits the excellent performance among the existing CS-based video codecs. Note that the results of CS-KLT codec are directly taken from the order-10 decoding in [35]. The CS-based video codec in our previous work [36] is different from the proposed CVS system: (1) it adopts a DPCM-based nonuniform quantizer; (2) its reconstruction is realized by the autoregressive prediction and adaptive PCA residual recovery. From **Fig. 11**, it is observed that the proposed codec is superior over the whole range of bitrates compared to the CS-KLT codec, e.g., the highest PSNR gain can be up to about 8 dB for *Container* sequence. Compared with [36], the proposed CVS system has a better rate-distortion performance for *Foreman* and *Container* sequences, and a similar PSNR curve is also obtained for *Highway* sequences with large static scenes. Besides, the CS-KLT codec requires lots of computations, e.g., its order-2 decoding time is about 332.81 seconds per frame on average when the subrate is 0.125, the codec in [36] requires about 10.14 seconds on average to decode a frame, and but the decoding time of our codec is only about 3.72 seconds on average per frame. Our success is mainly attributed to the three points that 1) the ST-Q quantization acquires both a low MSE and bits amount than the uniform and non-uniform SQs used respectively in the CS-KLT codec and our previous work [36], 2) the MA-R algorithm uses the better MH prediction with motion estimation and but CS-KLT codec still uses the traditional sparisity-constrain prediction, and 3) the MA-R guarantees a low computational complexity by replacing the KLT basis (or PCA basis) computed by multiple loop iterations in CS-KLT codec and video codec in [36] with the MH matrix constructed directly by these temporal neighbors. However, it can be seen that there is still a considerable performance gap between all CS-based video codecs and H.264i, which is attributed to the difficulty of solving inverse problem in CS reconstruction.

6. Conclusion

In this paper, we presented a practical CVS system combined with the ST-Q and MA-R. At the encoder, the ST-Q is used to quantize the CS measurements of video sequence. To remove the space-time redundancy between the measurement vectors of neighboring video frames, the ST-Q firstly computes the measurement vector of frame difference to remove the temporal redundancy, and then uses MFPQ to further reduce the spatial redundancy between blocks in the frame difference. After removing space-time redundancy, the predictive residual vector of each frame can be represented using far fewer bits than its original CS measurement vector, and meanwhile the quantization errors are also effectively reduced due to suppressing the accumulation of quantization noises on the lager values. At the decoder, the MA-R algorithm exploits the motion information between the two consecutive key frames to improve the MH prediction, and the MH matrix is constructed by selecting temporal neighborhoods in the windows centered on the motion-aligned block of current block. To not introduce the excessive computations, the MA-R algorithm firstly extracts the static area in several successive non-key frames to avoid the unnecessary search points for static blocks, and then computes the motion vectors of non-static blocks by using 3DRS to further reduce the search points while guaranteeing the accuracy of motion vectors with the smoothness constraint of motion vector field. Various experiments are performed to evaluate the performance of proposed CVS system from some perspectives, and their results demonstrated that the proposed ST-Q used in our codec reduces effectively the quantization errors of measurements so as to guarantee a better rate-distortion performance when compared with the uniform SQ,

DPCM-plus-SQ and SDPC, and the proposed MA-R algorithm has the superior performance than the existing reconstruction methods DISCOS and MHPR in both PSNR and visual quality. Combined with the ST-Q and MA-R, the proposed CVS system obtains a significant rate-distortion performance gain when compared with the existing CS-based video codecs. However, our codec still has the inferior performance when compared with the H.264/AVC. Therefore, we still require some future works to further improve the rate-distortion performance of CS-based video codec. Besides, the buffer sizes of SDPC and proposed MFPQ quantization are proportional to the number of CS measurements, which could be an issue in the resource constrained sensing environment. In the future, we will study how to reduce the memory size of predictive quantization.

References

[1]   M. S. Hosseini, and K. N. Plataniotis, "High-accuracy total variation with application to compressed video sensing," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3869-3884, Sept. 2014. Article (CrossRef Link)

[2]   S. Pudlewski, and T. Melodia, "Compressive video streaming: design and rate-energy-distortion analysis," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 2072-2086, Dec. 2013. Article (CrossRef Link)

[3]   H. Liu, B. Song, F. Tian, and H. Qin, "Joint sampling rate and bit-depth optimization in compressive video sensing," *IEEE Transactions on Multimedia*, vol. 16, no. 6, pp. 1549-1562, Oct. 2014. Article (CrossRef Link)

[4]   R. G. Baraniuk, "Compressive sensing," *Signal Processing Magzine*, vol. 24, no. 4, pp. 118-124, Jul. 2007. Article (CrossRef Link)

[5]   I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921-960, Oct. 2007. Article (CrossRef Link)

[6]   M. Nelson and J. L. Gailly, *The Data Compression Book*, 2nd ed. New York: M & T Books, 1995. Article (CrossRef Link)

[7]   W. Dai, H. V. Pham, and O. Milenkovic, "A comparative study of quantized compressive sensing schemes," in *Proc. of the 2009 IEEE International conference on Symposium on Information Theory*, pp. 11-15, Jun. 2009. Article (CrossRef Link)

[8]   W. Dai, H. V. Pham, and O. Milenkovic, "Quantized compressive sensing," *arXiv preprint arXiv*:0901.0749, pp. 1-18, 2009. Article (CrossRef Link)

[9]   A. Zymnis, S. Boyd, E. J. Candes, "Compressed sensing with quantized measurements," *IEEE Signal Processing Letter*, vol. 17, no. 2, pp. 149-152, Feb. 2010. Article (CrossRef Link)

[10] U. S. Kamilov, V. K. Goyal, and S. Rangan, "Message-passing de-quantization with applications to compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6270-6281, Dec. 2012. Article (CrossRef Link)

[11] J. Z. Sun, and V. K. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. of the 2009 IEEE International conference on Symposium on Information Theory*, pp. 6-10, Jun. 2009. Article (CrossRef Link)

[12] C. S. Gunturk, M. Lammers, A. Powell, R. Saab, and O. Yilmaz, "Sigma delta quantization for compressed sensing," in *Proc. of the 44th Annual Conference on Information Sciences and Systems*, pp. 1-6, Mar. 2010. Article (CrossRef Link)

[13] Feng Joe-Mei, and F. Krahmer, "An RIP-based approach to $\Sigma\Delta$ quantization for compressed sensing," *IEEE Signal Processing Letter*, vol. 21, no. 11, pp. 1351-1355, Nov. 2014. Article (CrossRef Link)

[14] L. Gan, "Block compressed sensing of natural images," in *Proc. of the 15th International Conference on Digital Signal Processing*, pp. 403-406, Jul. 2007. Article (CrossRef Link)

[15] S. Mun, and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *Proc. of the International Conference on Image Processing*, pp. 1103-1120, Nov. 2009. Article (CrossRef Link)

[16] C. Chen, E. W. Tramel, and J. E. Fowler, "Compressed-sensing recovery of images and video using multihypothesis predictions," in *Proc. of the 45th IEEE Asilomar Conference on Signals, Systems and Computers*, pp.1193-1198, 2011. Article (CrossRef Link)

[17] S. Mun, and J. E. Fowler, "DPCM for quantized block-based compressed sensing of images," in *Proc. of European Conference on Signal Processing*, pp. 1424-1428, Aug. 2012. Article (CrossRef Link)

[18] J. Zhang, D. Zhao, and F. Jiang, "Spatially directional predictive coding for block-based compressive sensing of natural images," in *Proc. of the International Conference on Image Processing*, pp. 1021-1025, Sept. 2013. Article (CrossRef Link)

[19] B. Girod, A. M. Aaron, S. Rane, and R. David, "Distributed video coding," in *Proc. of IEEE*, vol. 93, no. 1, pp. 71-83, Jan. 2005. Article (CrossRef Link)

[20] 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 of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586-597, Apr. 2007. Article (CrossRef Link)

[21] J. A. Tropp, and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007. Article (CrossRef Link)

[22] J. Prades-Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proc. of the Picture Coding Symposium*, pp. 1-4, May 2009. Article (CrossRef Link)

[23] L. W. Kang, and C. S. Lu, "Distributed compressive video sensing," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1169-1172, Apr. 2009. Article (CrossRef Link)

[24] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan and T. D. Tran, "Distributed compressed video sensing," in *Proc. of the International Conference on Image Processing*, pp. 1393-1396, Nov. 2009. Article (CrossRef Link)

[25] E. W. Tramel, and J. E. Fowler, "Video compressed sensing with multihypothesis," in *Proc. of the Data Compression Conference*, pp. 193-202, Mar. 2011. Article (CrossRef Link)

[26] C. Zhao, S. W. Ma, and W. Gao, "Video compressive sensing via structured Laplacian modelling." in *Proc. of the IEEE International Conference on Visual Communications and Image Processing*, pp. 402-405, Dec. 2014. Article (CrossRef Link)

[27] M. H. Wu, and X. C. Zhu, "Distributed video compressive sensing reconstruction by adaptive PCA sparse basis and nonlocal similarity." *KSII Transactions on Internet and Information Systems*, vol. 8, no. 8, pp. 2851-2865, Aug. 2014. Article (CrossRef Link)

[28] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. USA: Prentice Hall, 2007. Article (CrossRef Link)

[29] G. D. Haan, P. W. A. C. Biezen, H. Huijgen, and O. A. Ojo, "True motion estimation with 3-D recursive search block matching," *IEEE Transactions on Circuits System and Video Technology*, vol. 3, no. 5, pp. 368–379, Oct. 1993. Article (CrossRef Link)

[30] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no .8, pp. 3336-3351, Aug. 2014. Article (CrossRef Link)

[31] J. Zhang, C. Zhao, D. Zhao, and W. Gao, "Image compressive sensing recovery using adaptively learned sparsifying basis via L0 minimization," *Signal Processing*, vol. 103, pp. 114-126, 2014. Article (CrossRef Link)

[32] C. Zhao, S. Ma, and W. Gao. "Image compressive-sensing recovery using structured laplacian sparsity in DCT domain and multi-hypothesis prediction," in *Proc. of the IEEE International Conference on Multimedia and Expo.*, pp. 1-6, Jul. 2014. Article (CrossRef Link)

[33] S. Dasgupta, and A. Gupta, "An elementary proof of a theorem of Johnson and Lin denstrauss," *Random Structures and Algorithms*, vol. 22, no. 1, pp. 60–65, Jan. 2003. Article (CrossRef Link)

[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment : From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004. Article (CrossRef Link)

[35] Y. Liu, M. Li and D. A. Pados, "Motion-aware decoding of compressed-sensed video," *IEEE Transactions on Circuits System and Video Technology*, vol. 23, no. 3, pp. 438-444, Mar. 2013. Article (CrossRef Link)

[36] Ran Li, Hongbing Liu, Rui Xue, and Yanling Li, "Compressive-sensing based video codec by autoregressive Prediction and adaptive residual recovery," *International Journal of Distributed Sensor Networks*, vol. 2015, pp. 1-19, Sept. 2015. Article (CrossRef Link)

[37] T. T. Do, L. Gan, N. Nguyen and T. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 139-154, Jan. 2012. Article (CrossRef Link)
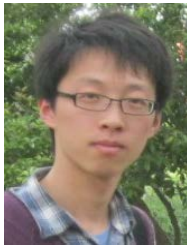
Ran Li received the B.Sc. degree in communication engineering from Huanghe College of Science and Technology, Zhengzhou, China, in 2006, and the Ph.D. degree in signal and information processing from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2014. He is currently a Lecturer in the School of Computer and Information Technology, Xinyang Normal University, Xinyang, China. His current research interests include video processing, compressed sensing, distributed video coding, and image super-resolution reconstruction.

Hongbing Liu received the B.Sc. degree in computer science from Henan University, Kaifeng, China, in 1993, the M.Sc. degree in computer application from Central China Normal University, Wuhan, China, in 2008, and the Ph.D. degree in computer science from Wuhan University of Technology, Wuhan, China, in 2011. He is currently an Associate Professor in the School of Computer and Information Technology, Xinyang Normal University, Xinyang, China. His research interests include intelligent information processing, data mining, rough set and image processing.

Wei He received the B.Sc. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2006, and the Ph.D. degree in computer science from Zhengjiang University, Hangzhou, China, in 2014. She is currently a Lecturer in the School of Computer and Information Technology, Xinyang Normal University, Xinyang, China. Her current research interests include compressed sensing and biological imaging.

Xingpo Ma received the B.Sc. degree, the M.Sc. degree and the Ph.D. degree in computer science from Central South University, Changsha, China, in 2005, 2009 and 2013, respectively. He is currently a Lecturer in the School of Computer and Information Technology, Xinyang Normal University, Xinyang, China. His research interests include wieless sensor networks, wireless media sensor networks and RFID networks.