International Journal of Multimedia and Ubiquitous Engineering
Vol.12, No.1 (2017), pp.389-402
http://dx.doi.org/10.14257/ijmue.2017.12.1.33

WME-KSVD Dictionary Based Distributed Compressive Video Sensing

Chen Rui[1], Wu Minghu[2*], Yang Jie[1] and Rui Xiongli[1]

[1]*School of Communications engineering, Nanjing Institute of Technology, Nanjing 211167, P. R. China*
[2]*Hubei Collaborative Innovation Center for High-efficiency Utilization of Solar Energy, Hubei University of technology, Wuhan 430068, P. R. China*
*chenrui@njit.edu.cn; *wuxx1005@163.com; {yangjie, ruixiongli}@njit.edu.cn*

Abstract

The video reconstruction quality largely depends on the employed sparse representation which approximates the video frame by a sparse linear combination of items from an over-complete dictionary. In this paper, we propose a novel adaptive-weighted side information extraction method is proposed to improve the reconstruction video quality. First, the similarity of the measured values between CS frame and the key frames is calculated. Then the weighted factors are decided according to the calculated similarities. The two key frames which have been made motion estimation multiply the weighted factor to obtain the side information. Then the dictionary is generated by the side information and KSVD algorithm. The simulation results show that the proposed algorithm outperforms the existed non-weighted side information algorithm in terms of peak-signal-to-noise ratio (PSNR) by 0.2~0.5 dB and visual perception.

Keywords: *distributed video compressive sensing; K-SVD algorithm; dictionary learning; video codec; sparse representation*

1. Introduction

The conditional video coding systems based on motion-compensated-prediction, such as H.26x, MPEG, are highly asymmetrical since the computationally intensive motion prediction is performed in the encoder. It satisfies the needs in applications such as video streaming, broadcast system, and digital versatile disk (DVD), where power constrains are less a concern at the encoder. However, new applications with limited access to power, memory, and computational resources at the encoder have difficulties using the conventional video coding systems. Therefore, a simple encoder with low complexity is needed. Distributed coding is one method to achieve low complexity at the encoder. In distributed coding, source statistics are exploited at the decoder and the encoder can be simplified. The theoretical basis for the problem dates back to two theorems in the 1970s. Slepian and Wolf proved a theorem to address lossless compression [1]. Wyner and Ziv extended the results to the lossy compression case [2]. Since these theoretic results were revisited in the late 1990s, several methods have been developed to achieve the results predicted in these two theorems, such as the European DISCOVER codec architecture [3], Bernd Girod's DVC scheme with feedback channel [4], Ramchandran's PRISM (Power-efficient Robust high-compression Syndrome-base Multimedia) scheme [5], and *etc*. These methods are based on the channel coding techniques, for example turbo codes and low-density-parity-check (LDPC) codes. Even though these methods have been proposed, but the encoding efficiency is lower than the traditional inter-frame encoders.

* Corresponding Author

ISSN: 1975-0080 IJMUE

More recently, an emerging signal acquisition technology Compressed Sensing (CS) provides a new way for the signal sampling, signal compression reconstruction based on the sparsity of signal, random measurement matrix and nonlinear optimization algorithm [6]. It broke through the limitations of traditional Nyquist sampling theorem, which has been applicable to directly capture compressed image data efficiently. Combination of distributed video coding and CS, known as distributed compressed video sensing (DCVS) [7], results in more low-complexity and low-cost for video coding. It asserts that a K-sparse signal can be faithfully recovered from less incoherent measurements via linear programming, which exploits linear projection to keep the original structure of image signal, suggesting a significant cost reduction of digital data acquisition. Due to the simplicity of the measurement acquisition at the encoder, the CS framework is a natural fit for distributed applications of limited video coding devices in the wireless network environment. CS is an emerging technology and enables to directly and efficiently capture compressed image data via randomly projecting raw image data to obtain linear and non-adaptive measurements. The image can then be reconstructed at the decoder via solving the convex optimization problem or using some iterative greedy algorithms from the captured data measurement [8].

Prior works have been done in DVCS. In [9], a framework called Distributed Compressed Video Sensing (DISCOS) is introduced. At the encoder, video frames are grouped into group of pictures (GOP) consisting of a key frame and a number of non-key frames. Key frames are encoded using traditional MPEG/H.26x encoding. For non-key frames (*i.e.*, CS frames), both local block-based and global frame-based CS measurements are taken. Side information is generated by using a block-based prediction frame which is created by sparsity-constraint block prediction. In this approach, the block-based measurements of a CS frame are compared with two neighboring decoded key-frames. The measurement vector of the prediction frame is subtracted from that of the input frame to form a new measurement prediction error vector. The reconstructed CS frame is simply the sum of the prediction error and the prediction frame. In this approach, the complex MPEG/H.26x encoding is still required. In [10], the authors' approach is different from [9] in that CS measurement is applied to both key frames and CS frames. Key frames are reconstructed using GPSR at the decoder. For every CS frame, a stopping criteria based on side information generated from the key frames is used during the reconstruction process. Side information is generated by an efficient frame rate up-conversion tool. This work is extended by H. W Chen and *etc.*, [11][12] with the concept of dictionary learning. The dictionary is learned from adjacent video frames. After that, many dictionaries have been applied. In [10], a DVCS framework was proposed with respect to the discrete wavelet transform (DWT) basis, where an efficient initialization and several stopping criteria were proposed to improve and speed up the employed convex optimization algorithm for CS frame reconstruction. The technique presented in [13] incorporated reconstruction from a residual arising from ME/MC that had been widely deployed for several decades in video compression standards and applied discrete cosine transform (DCT) as the sparse representation basis. Besides using the above orthonormal basis (*e.g.*, the DCT basis and wavelet basis), an overcomplete dictionary was proposed for sparse representation of signals in [14]. The overcomplete dictionary contains prototype signal-atoms, and the video signals can be represented by the sparse linear combinations of these atoms. Recent activity in this field concentrated mainly on the study of pursuit algorithms that decompose signals with respect to a given dictionary. Designing dictionaries to better than the above model can be done by *either* selecting one from a pre-specified set of linear transforms, such as multi-scale Gabor function, wavelet and cascaded sine function, or adapting the dictionary

to a set of training signals, such as MOD, K-SVD, ICA-like and *etc*. [15-17]. A dictionary learning method using K-SVD algorithm was proposed in [12] and [18], based on the training samples extracted from previous reconstructed neighboring frames together with the side information. Furthermore, a "local dictionary" based scheme was proposed in [19-21], for their major core employing the local blocks extracted a set of spatially neighboring blocks of previous decoded neighboring key frames as the dictionary for each block in a CS frame. In our previous work in [22], we adopted PCA algorithm based on the concepts of sparse-land model and nonlocal similarity to construct the global over-complete dictionary, and seek the best representation for video signals.

Another distributed approach to DVCS is reported in [23]. For each image block in CS frame, two different coding modes, SKIP and SINGLE, are used. In the SKIP mode, a block is skipped for decoding of it does not change much from the co-located decoded key frame. This is achieved by increasing the complexity at the encoder. In the SINGLE mode, CS measurements in a dictionary using the MSE criterion. If it is blow some threshold, then the block is marked as a decoded block. A feedback channel is needed to communicate with the encoder that this block has been decoded and no more measurements are required. For blocks that are not encoded by either the SKIP or the SINGLE mode, normal CS reconstruction is performed.

For all these methods have been considered, there is a need for a CS based video codec which does not require a feedback channel with simple side information generation methods to keep the complexity of the encoder low. In this paper, we follow the idea proposed recently in [9] to implement a distributed video code system with an adaptive dictionary. Side information is generated based on the similarity of CS measurements in video frames by an adaptive-weighted extraction algorithm. In our scheme, each source video frame is compressed independently by a number of random sampling operations so as to keep the simplicity at the encoder. On the other hand, all analysis will be conducted at the decoder, leading to a joint and more complicated decoding to deliver a higher performance. Compared with the original work in [9], our contributions in this paper are summarized as follows.

(1) Side information generation. We propose a weighted side information generation method in measurement domain by calculating the similarity factors.

(2) The sparse dictionary in [9] keeps a constant size, which ignores the diversified contents in various blocks within a frame as well as temporal variations among frames. In this paper, we propose to adjust adaptively the block-based sparse dictionary size to improve the coding performance.

The rest of this paper is organized as follows. In Section 2, we introduce the related works including compressive sensing, K-mean Singular Value Decomposition (KSVD), motion estimation, and similarity measurement. Our proposed distributed video compressed sensing framework is presented in Section 3. It is tested using several typical video sequences, and the experimental results and comparing the performance of the proposed algorithm with previously proposed methods are presented in Section 4. Finally, conclusions and future works are discussed in Section 5.

2. Related Works

2.1. The DISCOS Framework

The DISCOS framework in [9] is shown in Figure 1. A source video sequence is divided into several GOPs (group of pictures), where a GOP consists of a key-frame followed by some CS-frames. Each key-frame is intra-coded by a conventional

video coding method (such as MPEG or H.26x). CS-frames are compressively sampled by using two kinds of measurements, block-based (local) and frame-based (global) ones, and all measured data are transmitted to the decoder. The frame-based measurements is similar to the generic CS coding, *i.e.*, each frame F_t (of size $N \times N$, t denoting the time) is first vectorized as x_t (with height N^2) and then compressed via a CS-sampling process as:

$$y_t = \Phi x_t \tag{1}$$

where y_t denotes the output measurement-vector of length M_t, Φ represents the $M_t \times N^2$ measurement (or sampling) matrix generated by the method of structurally random matrices (SRMs) [14]. The measurement rate for x_t is denoted as $R_t = \frac{M_t}{N^2}$.

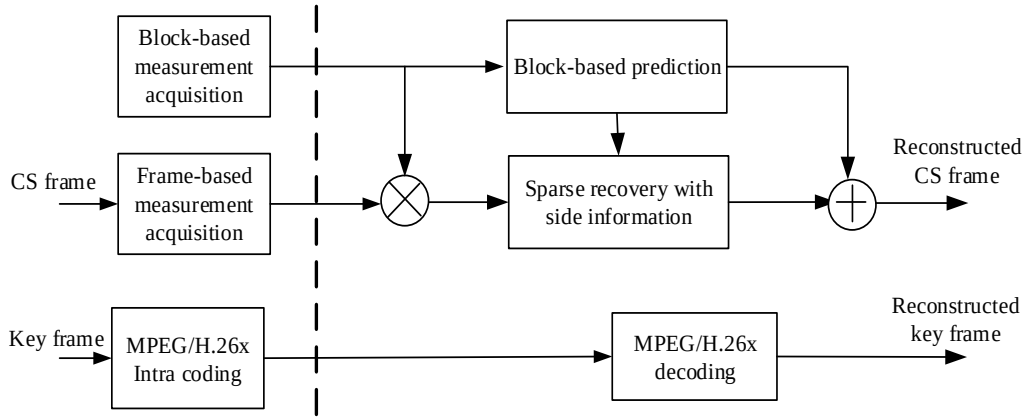

Figure 1. The DISCOS Framework

The block-based measurements are also exploited to preserve local information that helps the decoder construct more accurate side information (SI) in DISCOS. Each CS-frame is first partitioned into non-overlapped blocks of size $B \times B$. Then, each vectorized block x_i (where i stand for the block's index) is sampled with the same CS operator as:

$$y_i = \Phi_B x_i \tag{2}$$

where Φ_B is the measurement matrix. The equivalent sampling operator Φ appeared in (1) for the whole frame is a block-wise diagonal matrix composed by Φ_B. At the decoder side, an independent reconstruction by using the necessary video decoding method is first carried out for all key-frames, while the reconstruction for CS-frames are much more complicated. As shown in Figure 1, each block in a frame is reconstructed via solving an l_1 minimization problem as:

$$\hat{\alpha}_i = \arg\min \|\alpha_i\|_1 \quad s.t. \quad y_i = \Phi_B \Psi_i \alpha_i \tag{3}$$

where y_i is obtained from (2), Ψ_i is a sparse basis matrix which can provide a sparse representation for x_i, i.e., $x_i = \Psi_i \cdot \alpha_i$. Instead of using a fixed linear transform (*e.g.*, the block DCT), DISCOS uses a dictionary formed from a set of spatially neighboring blocks of previously decoded neighboring key-frames as the sparsifying matrix Ψ_i. Block-based prediction uses the sparsity adaptive matching pursuit (SAMP) [19] reconstruction algorithm to solve the l_1-minimization. Then, DISCOS employs a sparse recovery with SI from its global measurements and its local block-based prediction to jointly reconstruct a CS-frame: to subtract the measurement vector of an original CS-frame from that of a block-based prediction frame to form a new measurement vector of the prediction error. Finally, the CS-

frame is recovered by adding the prediction error to the prediction frame, and the gradient projection for sparse reconstruction (GPSR) algorithm [20] is used.

2.2. K-means Singular Value Decomposition (K-SVD)

K-SVD is a dictionary learning algorithm for creating a dictionary for sparse representations, via a singular value decomposition approach [24]. K-SVD is a generalization of the *K*-means clustering method, and it works by iteratively alternating between sparse coding the input data based on the current dictionary, and updating the atoms in the dictionary to better fit the data. Let Y be a set of n-dimensional N input signals, *i.e.*, $Y=\left[y_1...y_N\right]\in R^{n\times N}$. Learning a reconstructive dictionary with K items for sparse representation of Y can be accomplished by solving the following problem

$$\min_{D,X}\{\left\|Y-DX\right\|_F^2\} \qquad s.t. \quad \forall i, \quad \left\|x_i\right\|_0 \le T_0 \tag{4}$$

where D is the learned dictionary, *i.e.*, $D=\left[d_1...d_K\right]\in R^{n\times K}$ ($K>n$, making the dictionary over-complete), $X=\left[x_1...x_N\right]\in R^{K\times N}$ are the sparse codes of input signals Y, and T_0 is a sparsity constraint factor (each signal has fewer than T_0 items in its decomposition). The term $\left\|Y-DX\right\|_F^2$ denotes the reconstruction error.

The construction of D is achieved by minimizing the reconstruction error and satisfying the sparsity constraints. The K-SVD algorithm is an iterative approach to minimize the energy in Eq. (4) and learns a reconstructive dictionary for sparse representations of signals. It is highly efficient and works well in image compression. Given D, sparse coding computes the sparse representation X of Y by solving:

$$\min_{D,X}\sum_i \| x_i \|_0 \qquad s.t. \ \ \forall i \ \| Y-DX \|_F^2 \le \varepsilon \tag{5}$$

where ε is the reconstruction error.

The implementation of K-SVD algorithm includes two steps: sparse coding and dictionary updating, described as:

Step1. Sparse coding. For a given initial dictionary D and the objective function, adapt a purchasing algorithm to approach the optimal approximate sparse representation matrix X of the sampling points. The reconstruction algorithm used in this paper is OMP algorithm.

Step2. Dictionary updating. Update each dictionary column and the corresponding values of sparse matrix X orderly, according to the SVD (singular value decomposition). Suppose that the column to be updated is K-column, then the objective function can be written in the following form:

$$\begin{aligned} \| Y-DX \|_F^2 &= \| Y-\sum_{j=1}^{k} d_j X_T^2 \|_F^2 = \| (Y-\sum_{j\neq k} d_j X_T^2)-d_k X_T^k \|_F^2 \\ &= \| E_k - d_k X_T^k \|_F^2 \end{aligned} \tag{6}$$

In order to remove the zero samples, we define ω_k as the set of K values which satisfied $x_T^k(i)\neq 0$, *i.e.*, $\omega_k=\left\{i \mid 1\le i\le k, x_T^k(i)\neq 0\right\}$. Then we define matrix Ω_k, which size is $N\times\Omega_k$. The values of matrix Ω_k is 1 at $(\omega_k(i),i)$, and the others is 0. Let $x_R^k=x_T^k\Omega_k$, $E_k^R=E_k\Omega_k$, then (6) can be rewritten as

$$\| E_k \Omega_k - d_k X_T^k \Omega_k \|_F^2 = \| E_k^R - d_k x_R^k \| \tag{7}$$

The SVD (singular value decomposition) of matrix E_k^R is $E_k^R = U\Delta V^T$. The atoms d_k of the initial dictionary D is replaced by the first column of matrix U. The first column of matrix V multiplies $\Delta(1,1)$, the result is used to update the sparse matrix x_R^k. The above two steps will be repeated until the convergence condition is satisfied.

2.3. Motion Estimation

At DCVS decoder, for a CS frame x_t, its side information I_t can be generated from the motion-compensated interpolation of its previous and next reconstructed key frames, respectively, denoted by x_{t-j} and x_{t+j}. Motion estimation is the process of determining motion vectors that describe the transformation from adjacent reconstructed key frames.

Block matching is the most popular and efficient motion estimation technique [25]. The key steps of the block-matching method are reviewed as follows:

(1) partition frame F_{i-1} (*e.g.*, previous frame) into $P \times P$ (pixels) blocks;

(2) pre-define a window size $M \times M$ (pixels);

(3) search all the $P \times P$ blocks in the $M \times M$ windows in frame F_i (*e.g.*, current frame) around the selected block in frame F_{i-1};

(4) find the best-matching block in the window according to some metric (*e.g.*, mean squared error), and use this to compute the block motion. The block size used in this paper is 16×16.

The most well known matching criteria for block matching motion estimation are Mean Absolute Error (MAE), Mean Squared Error (MSE), Cross Correlation Function (CCF), and *etc*. Because the Sum of Absolute Difference (SAD) criterion gives good performance in terms of the operation speed and accuracy, so we use SAD criterion which can be calculated by

$$SAD(h,v) = \sum_{i=1}^{M}\sum_{j=1}^{N} | x_k(i,j) - x_{k-1}(i+h, j+v) | \tag{8}$$

where $x_k(i,j)$ is the pixel value of current frame, $x_{k-1}(i+h, j+v)$ is the pixel value of the decoded key frame, $M \times N$ is the block size, (h, v) is the relative displacement.

2.4. Similarity Measurement

Measurement of the similarity between key frames and CS frames is very important for compression. At the encoder, we only know the CS measurement, the reconstructed key frame and the measuring matrix, so the similarity between the key frame and the CS frame is valued by the similarity between the measurement values of the key frame and the CS frame.

Lemma 2.1 [Johnson-Lindenstrauss (JL) lemma] [26]

Let $\varepsilon \in (0,1)$ be given. For every set Q of $\#Q$ points in R^N, if n is a positive integer such that $n > n_0 = O(\ln(\#(Q))/\varepsilon^2)$, there exists a Lipschitz mapping $f : R^N \to R^n$ such that

$$(1-\varepsilon)\|u-v\|^2 \le \|f(u)-f(v)\|^2 \le (1+\varepsilon)\|u-v\|^2 \tag{9}$$

for all $u, v \in Q$.

JL lemma states that a small set of points in a high-dimensional space can be embedded into a space of much lower dimension by a random linear projection, and the similarity between the high dimensional space mapping and low dimensional space remains within a controllable range of similarity. Key frames and CS frames

after measuring the equivalent transformation to a low dimensional space, according to the similarity between the measured values can be inferred before measuring the similarity.

The most common similarity measures are mean squared error (MSE) and peak signal-to-noise ratio (PSNR). In this paper, we use MSE to measure the similarity between the measured values of key frames and CS frames:

$$u(x_{t-1}, x_t) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_{t-1,k}, x_{t,k})^2} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_{t-1}, x_t)^T(x_{t-1}, x_t)} = \sqrt{\frac{1}{n}}\left\|x_{t-1}, x_t\right\|^2 \quad (10)$$

where x_{t-1} and x_t are the measured values of key frame and CS-frame, respectively. Greater values of u (x_{t-1}, x_t) indicate lower similarity.

3. Proposed WME-KSVD dictionary based DCVS

The distributed compressive video sensing (DCVS) is a solution for distributed video coding based on the compressed sensing theory. The DCVS framework compressively samples each video frame independently at the encoder and recovers video frame jointly at the decoder by exploiting an interframe sparsity model by performing sparse recovery with side information. In this paper, we propose a new DCVS based on adaptive weighted side information and KSVD algorithm, abbreviated as WME-KSVD algorithm (WME: Weighted Motion Estimation).

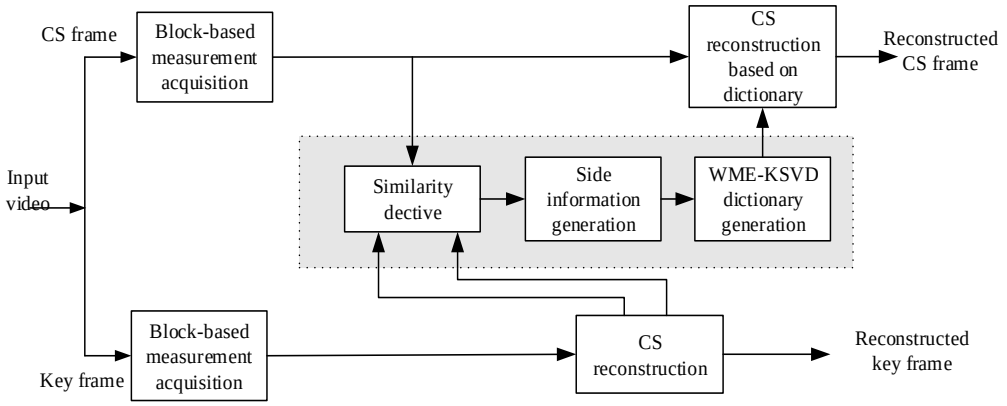

Figure 2. Block Diagram of WME-KSVD Dictionary based Distributed Compressive Video Sensing

3.1. WME-KSVD Dictionary based DCVS Framework

The proposed WME-KSVD dictionary based DCVS framework is depicted in Figure 2. At the encoder, video frames are divided into key frames (also called K-frames) and non-key frames (also called CS-frames). In order to make the coding simple, both K-frames and CS-frames adopt consistent block-based random measurements. The difference lies in that: the sampling rate of K-frame is higher and fixed, while the sampling rate of CS-frame is lower and variable. At the decoder, K-frames are directly reconstructed using CS reconstruction algorithm, and CS-frames are jointly reconstructed by the side information *SI*, which is obtained by the previous reconstructed K-frame to produce the prediction of the CS-frame. It is a typical "separate encoding, joint decoding" distributed video coding framework.

3.2. WME-KSVD dictionary generation

To achieve a well-performed sparse basis for CS frames, blocked-based prediction technique is used to generate side information by exploiting temporal

correlation between adjacent decoded key frames and the current CS frame. Note that different regions in a video sequence have different inter-frame correlation, we propose an adaptive weighted motion estimation scheme to generate side information I_w, and the dictionary D is trained by KSVD algorithm At the decoder, for a CS-frame X_t, its side information I_t is generated from the motion-compensated interpolation of its previous and next reconstructed K-frames, respectively, denoted by X_{t-1} and X_{t+1}. Then we use X_{t-1}, X_{t+1} and an adaptive weighted side information I_w to train the dictionary for this CS frame X_t as follows.

(1) as mentioned in section 2.3, $\overset{\circ}{X}_{t-1}$ is generated from the forward motion estimation of key frame X_{t-1}, and $\overset{\circ}{X}_{t+1}$ is generated from backward motion estimation of key frame X_{t+1}.

(2) the measurements of $\overset{\circ}{X}_{t+1}$ and $\overset{\circ}{X}_{t+1}$ (denoted as x_{t-1} and x_{t+1}, respectively) can be calculated as

$$\begin{cases} x_{t-1} = Phi \times \overset{\circ}{X}_{t-1} \\ x_{t+1} = Phi \times \overset{\circ}{X}_{t+1} \end{cases} \tag{11}$$

Then the similarity $u(x_{t-1}, x_t)$ is calculated by (11), and the weighted factor α is chosen according to the calculated result of $u(x_{t-1}, x_t)$. The great $u(x_{t-1},x_t)$ means a great weighted factor α is chosen. In our experiments, we set 9 weighted factors: α_i={0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1}. According to $u(x_{t-1}, x_t)$ and $u(x_{t+1}, x_t)$, the weighted factor α_i is adaptively selected. Then the side information I_t is calculated by

$$I_t = \alpha_i \overset{\circ}{X}_{t-1} + (1-\alpha_i)\overset{\circ}{X}_{t+1} \tag{12}$$

(3) Finally, constitutes the WME-KSVD dictionary by using KSVD algorithm and side information I_t in Step2.

In summary, the description of WME-KSVD algorithm is described as follows.

Table 1. WME-KSVD Algorithm

Algorithm 1 WME-KSVD algorithm
Input: X_{t-1} and X_{t+1} are the two neighboring decoded key frames as reference for the *i*-th CS-frame.
Output:
1: motion estimation to search the optimal blocks in adjacent decoded fames: $\overset{\circ}{X}_{t-1}$, $\overset{\circ}{X}_{t+1}$
2: calculates $x_{t-1} = Phi \times \overset{\circ}{X}_{t-1}$ and $x_{t+1} = Phi \times \overset{\circ}{X}_{t+1}$
3:calculates $u(x_{t-1}, x_t)$ and $u(x_{t+1}, x_t)$, and selects α_i according to $u(x_{t-1}, x_t)$ and $u(x_{t+1}, x_t)$
4: generates the side information *SI*: $SI = \alpha_i \overset{\circ}{X}_{t-1} + (1-\alpha_i)\overset{\circ}{X}_{t+1}$
5: initializes the dictionary *D* using *SI* in step4, then constitutes the dictionary *D* during KSVD algorithm
6: returns *D*

3.3. CS Reconstruction with WME-KSVD Dictionary

In our DVCS, applying WME-KSVD dictionary D for CS-frame reconstruction, (2) can be rewritten as

$$\alpha_{\text{opt}} = \arg\min_{\alpha}\{\| y - \Phi D\alpha \|_2 + \lambda \| \alpha \|_1\} \quad (13)$$

where $\lambda(\lambda > 0)$ is the factor that trades off between the reconstruction error and sparsity, D is WME-KSVD dictionary, and α is the representation coefficient matrix of CS-frame $\mathbf{x}(t_i)$ over D. (14) is a reweighted l_1-minimization problem, which can be effectively solved by the GPSR (Gradient Projection for Sparse Reconstruction algorithm) [27] or the iteration shrinkage algorithm [28].

The reconstructed algorithm of CS frames is described as follows.

Table 2. The Reconstructed Algorithm of CS Frames

The reconstructed algorithm
Input: y, Φ and λ output: 1. initialization: set initial $Z^{(0)} \in \Omega$, $\alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$, $0 < \alpha_{\min} < \alpha_{\max} \le 1$, counter k=0 2. projection: if $Z^{(k)}$ meet the conditions, then stop; otherwise, the search direction is $d^{(k)} = P_{\Omega}(Z^{(k)} - \alpha_k(GZ^{(k)} + q)) - Z^{(k)}$ Where $GZ^{(k)} + q$ is the gradient of the objective function, $P_{\Omega}(g)$ is orthogonal projection on Ω. 3: line searching: calculate $Z^{(k+1)} = Z^{(k)} + \lambda_k d^{(k)}$, $\lambda_k \in (0,1]$ 4: iteration: calculate $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$, k ++ 5: repeats step 2 6: returns α_{opt}

4. Experimental Results

To testify the proposed WME-KSVD Dictionary based DCVS, we implement it and assess the performance on four QCIF video sequences: *Foreman*, *Coastguard*, *News* and *Akiyo*. We perform experiments including ME-KSVD dictionary [19] and the proposed WME-KSVD dictionary. The dictionary size is 256×256, the atom size is 16×16.

In DCVS, the block size is 16×16, the sampling rate of key frame is 0.5, the sampling rate of CS frame varies from 0.1 to 0.5, and both key frame and CS frame are adopted GPSR algorithm for reconstruction. The other parameters are setting as: λ=0.8; block size=16×16.

The average reconstructed qualities are shown in Figure 3 and Table 3. As we can see, under the same condition, the reconstructed quality of the proposed WME-KSVD dictionary outperforms the ME-KSVD dictionary above 0.2~0.5dB in PSNR.

In our experiments, we find that the times costs of GPSR reconstruction algorithm is a little higher than training the dictionary. The use of specialized decoding hardware can reduce the reconstruction time to achieve real-time decoding. Note that in our experiments, the video frames are divided into blocks for measuring and reconstruction, so there are block effects in the reconstructed frames, which can be eliminated by filtering and post-processing.

(a) (b) (c)

Figure 3. Reconstructed CS Frame of *Akiyo* at Sampling Rate of 0.3: (a) Original; (b) ME-KSVD Dictionary; (c) Proposed WME-KSVD Dictionary

We have also carried out experiments on 4 video sequences with different CS sampling rates. Table 1 lists the average PSNR values of while the CS frame sampling rate varies from 0.1 to 0.5. As we can see from Table 3, the proposed WME-KSVD algorithm is better than ME-KSVD algorithm, and the reconstructed video quality is improved in terms of PSNR by 0.2~0.5dB. Because of the intensity of video sequences, the results of these two methods are different. When the sampling rate $r \geq 0.3$, improving sampling rate cannot greatly improve the reconstruction quality. So, the CS sampling rate is set to 0.3. The performance comparison between ME-KSVD and the proposed WME-KSVD for CS frames is shown in Figure 4, while the sampling rate is 0.3.

Table 3. The Average Reconstructed Quality of CS Frames in PSNR

	Average PSNR(dB)				
	rate				
Foreman	0.1	0.2	0.3	0.4	0.5
ME-KSVD	26.71	28.18	28.96	29.45	29.74
WME-KSVD	27.24	29.02	29.22	29.79	29.98
Coastguard					
ME-KSVD	23.68	25.23	25.99	26.44	26.89
WME-KSVD	24.27	26.09	26.41	26.94	27.11
News					
ME-KSVD	24.05	26.54	27.81	28.59	29.17
WME-KSVD	25.58	27.94	28.16	29.17	29.54
Akiyo					
ME-KSVD	27.85	29.45	30.41	31.34	31.93
WME-KSVD	28.54	30.56	30.86	31.64	32.13

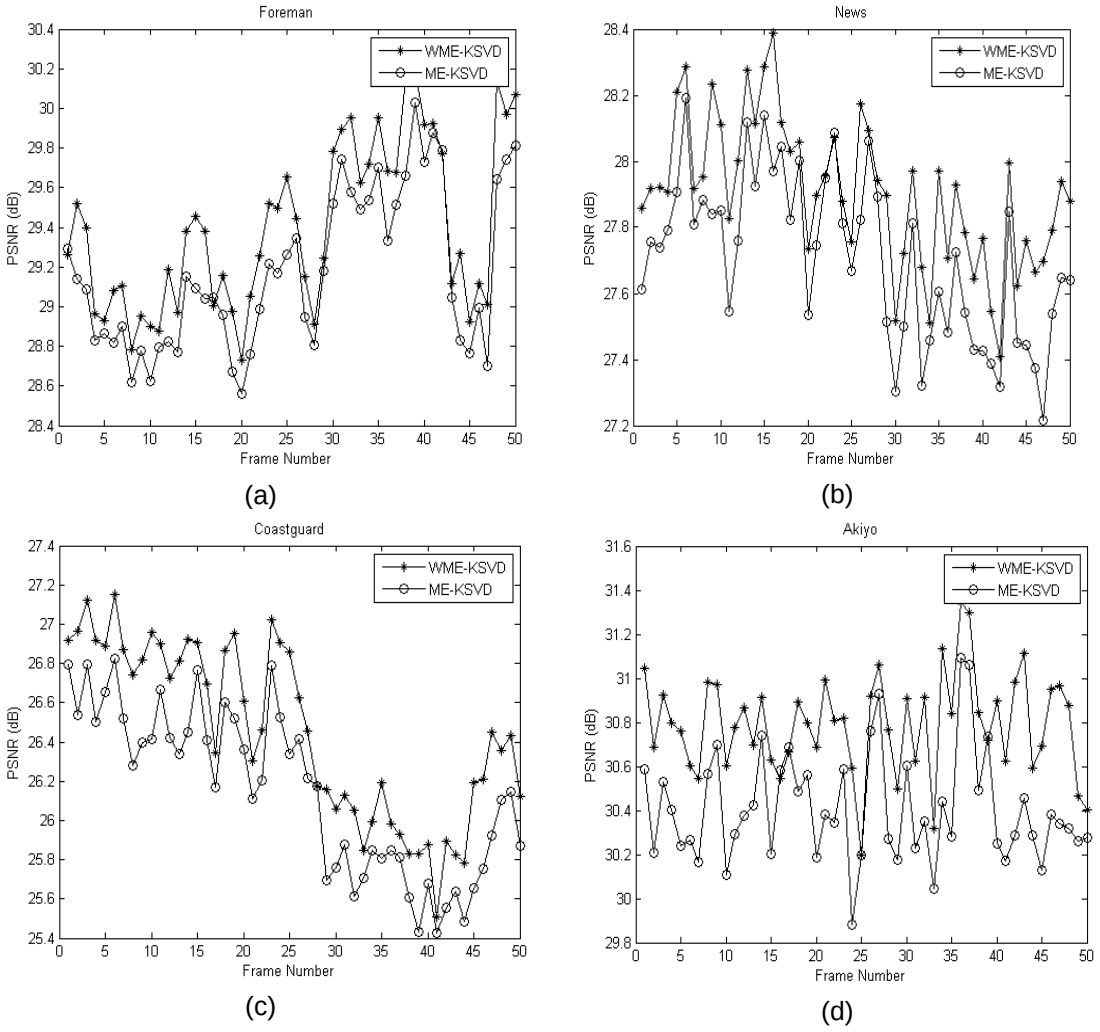

Figure 4. Performance Comparison between ME-KSVD and the Proposed WME-KSVD(CS Frames): (a) *Foreman* Sequence; (b) *News* Sequence; (c) *Coastguard* Sequence; (d) *Akiyo* Sequence

5. Conclusions

In this paper, a distributed compressive video sensing (DCVS) framework via adaptive weighted motion estimation and KSVD dictionary training method is proposed to improve the reconstructed quality. The weighted factors have been selected according to the calculated similarity between the key frame and the CS frame. The simulation results have shown that our algorithm outperforms ME-KSVD dictionary method in [15] in both PSNR and visual quality. For future works, several important issues need to investigate in depth for achieving a complete CS-based video coding system are described as follows. (1) More efficient sparse representation for video signals to exploit the underlying video in DCVS. (2) Adaptive measurement matrix learning: If a measurement matrix can be adaptively learned based on the characteristics of current signal to be captured, the number of captured measurements should be reduced while preserving a certain performance. (3) Fast dictionary training at the decoder and more accurate side information generation. If more accurate side information for a CS frame can be generated, the trained dictionary can provide much sparser representation for this frame, resulting in better compression performance.

Acknowledgments

This research was supported by National Natural Science Foundation of China (No.61471162), Program of International science and technology cooperation (2015DFA10940）, NSF of Jiangsu Province (BK20141389), Technology Research Program of Hubei Provincial Department of Education (D20141406), NSF of Hubei Province (2014CFB589), and Innovation Project of Nanjing Institute of Technology (ZKJ201305, QKJA201304).

References

[1] J Slepian, J Wolf, Noiseless Coding of Correlated Information Sources, IEEE Transactions on Information Theory, 4, 19, pp. 471- 480 **(1974)**
[2] A. D. Wyner, J. Ziv, The Rate-distortion Function for Source Coding with Side Information at the Decoder, IEEE Transactions on Information Theory, 1, 22, pp. 1-10 **(1976)**
[3] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, The DISCOVER Codec: Architecture, Techniques and Evaluation, Proceedings of the Picture Coding Symposium, **(2007)** Apirl 24-26; Beijing, China
[4] Dong W, Yang X, Shi G., Compressive Sensing via Reweighted TV and Nonlocal Sparsity Regularisation, Electronics Letters, 3, 49, pp. 184-186 **(2013)**
[5] R. Puri, A. Majumdar, K. Ramchandran, PRISM: A Video Coding Paradigm with Motion Estimation at the Decoder, IEEE Transactions on Image Processing, 10, 16, pp. 2436-2448 **(2007)**
[6] David L, Donoho, Compressed Sensing, IEEE Transactions on Information Theory, 4, 52, pp. 1289-1306 **(2006)**
[7] Vladimir Stankovic, Lina Stankovic, Samuel Cheng, Compressive Image Sampling with Side Information, Proceedings of IEEE International Conference on Image Processing, **(2009)** ; Cairo, Egypt
[8] Girod B, Aaron A, Rane S, Distributed video coding, Proceedings of the IEEE, 1, 93, pp. 71-83. **(2005)**
[9] Do T T, Chen Y, Gan, Distributed Compressed Video Sensing, Proceedings of IEEE International Conference on Image Processing, **(2009)** November 7-12; Cairo, Egypt
[10] Kang L W, Lu C S, Distributed Compressive Video Sensing, Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, **(2009)** April 19-24; Taipei, China
[11] H. W. Chen, L. W. Kang, and C. S. Lu, Dictionary Learning-based Distributed Compressive Video Sensing, Proceedings of SPIE Visual Communications and Image Processing, **(2010)** July 7-10; Huangshan, China
[12] H. W. Chen, L. W. Kang, and C. S. Lu, Dynamic Measurement Rate Allocation for Distributed Compressive Video Sensing, Proceedings of SPIE on Visual Communications and Image Processing, **(2010)** July 7-10; Huangshan, China
[13] Mun S, Fowler J E, Residual Reconstruction for Block-Based Compressed Sensing of Video, Proceeding of the Data Compression Conference, **(2011)** March 29-31; Snowbird, USA
[14] Elad M, Sparse and Redundant Representation from Theory to Applications in Signal and Image Processing, Springer Publishers, **(2010)**
[15] Michal A, Elad M, Alfred B, K-SVD: An algorithm for Designing Over-complete Dictionaries for Sparse Representation, IEEE Transactions on Signal Processing, 11, 54, pp. 4311-4322 **(2006)**
[16] Engan K, Aase S O, Hakon H J, Method of Optimal Directions for Frame Design Acoustics, Proceedings of IEEE International Conference on Acoustic Speech and Signal Processing, **(1999)** March 15-19; Washington DC, USA
[17] Duarte Carvajalino J M, Sapiro G, Learning to Sense Sparse Signals: Simultaneous Sensing Matrix and Sparsifying Dictionary Optimization, IEEE Transactions on Image Processing, 7, 18, pp. 1395-1408 **(2009)**
[18] Wu M H, Gan Z L, Zhu X C, Adaptive Dictionary Learning for Distributed Compressive Video Sensing, International Journal of Digital Content Technology and its Applications, 4, 6, pp. 141-149 **(2012)**
[19] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, Sparsity Adaptive Matching Pursuit for Practical Compressed Sensing, Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers, **(2008)** October 26-29; Pacific Grove, California, USA
[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, 4, 1, pp. 586-597 **(2007)**
[21] Liu H X, Song B, Qin H, A Dictionary Generation Scheme for Block-based Compressed Video Sensing, Proceedings of IEEE International Conference on Signal Processing, Communications and Computing, **(2011)** Sept. 14-16; Xian, China
[22] Wu Minghu, Chen Rui, Li Ran, and Zhou Shangli, Dynamic Global-principal Component Analysis Sparse Representation for Distributed Compressive Video Sampling, China Communications, 5, 10, pp. 20-29 **(2013)**

[23] J. Peades-Nebot, M. YI, and T. Huang, Distributed Video Coding Using Compressive Sampling, Proceedings of Picture Coding Symposium, **(2009)** May 6-8; Chicago, USA
[24] Rubinstein R, Faktor T, and Elad M, K-SVD Dictionary-learning for the Analysis Sparse Model. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, **(2012)** March 25-30; Kyoto Japan, pp. 5405-5408
[25] Pandian S, Bala G J, George B A., A Study on Block Matching Algorithms for Motion Estimation, International Journal on Computer Science & Engineering, 1, 3, pp.54-58 **(2011)**
[26] Kane D M, Nelson J, Sparser Johnson-Lindenstrauss Transforms, Journal of the ACM, 1, 61, pp. 1-24 **(2014)**
[27] Figueiredo M A T, Nowak R D, Wright S J, Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems, IEEE Journal of Selected Topics in Signal Processing, 12, pp. 586-597 **(2007)**
[28] Needell D, Tropp J A, CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples, Applied and Computational Harmonic Analysis. 3, 26, pp. 301–321 **(2009)**

Authors

Chen Rui, received the B.E degree, the M.E. degree from Southeast University, Nanjing, China, in 1991 and 1996, respectively. She received PhD degree from Nanjing University of Post and Telecommunications in 2013. Her major research interests include distributed video coding and wireless multimedia communication.

Wu Minghu (Corresponding author), received the B.S. degree from Communication University of China, Beijing, China and the M.S. degree from Huazhong University of Science and Technology in 1998 and 2002, respectively. He received PhD degree from Nanjing University of Post and Telecommunications in 2014. His major research interests include signal processing, video coding and compressive sensing.

Yang Jie, received the B.E. degree, the M.E. degree from Lanzhou University of Technology, Lanzhou, China. She received PhD degree from Nanjing University of Posts and Telecommunications in 2015. Her major research interests include cooperative communication and signal processing.

Rui Xiongli, received the B.E. degree, the M.E. degree from Hohai University, Nanjing, China. She is currently a Ph.D. candidate at Nanjing University of Posts and Telecommunications, Nanjing, China. Her major research interests in multimedia signal processing.