An elastic net-based hybrid hypothesis method for compressed video sensing

Jian Chen · Yunzheng Chen · Dong Qin · Yonghong Kuo

Abstract Compressed Sensing, an emerging framework for signal processing, can be used in image and video application, especially when available resources at the transmitter side are limited, such as Wireless Multimedia Sensor Networks. For a low-cost and low-power demand, we consider the plain compressive sampling and low sampling rates and propose a Compressed Video Sensing scheme. As a result, most burden of video processing can be shifted to the decoder which employs a hybrid hypothesis prediction method in reconstruction. The Elastic net-based multi-hypothesis mode, one part of the prediction method, combines the multi-hypothesis prediction and the elastic net regression together. And in the process of decoding, either this mode or the single-hypothesis one is implemented based on the threshold which is selected from [1e-11, 1). Both of the prediction modes are carried out in the measurement domain and a residual reconstruction as the final step is executed to accomplish the recovery. According to the performance presented by the simulation results, the proposed multi-hypothesis mode provides a better reconstruction quality than the other multi-hypothesis ones and the proposed scheme outperforms the observed state-of-the-art schemes for compressed-sensing video reconstruction at low sampling rates.

Keywords Compressed sensing · Distributed video coding · Elastic net · Hypothesis prediction · Wireless multimedia sensor networks

1 Introduction

The traditional video coding standards share a common characteristic that the design of the encoders is more complex than the decoders. For this reason, these standards are not suitable for all application backgrounds. For example, in the rapidly developing Wireless Multimedia Sensor Networks (WMSN) [1], the lifetime of the large number of sensors has a great effect on the performance of the network. So the low-complexity transmitter with a good coding

J. Chen (✉) · Y. Chen · Y. Kuo
School of Telecommunications Engineering, Xidian University, Xi'an 710071, China
e-mail: jianchen@mail.xidian.edu.cn

D. Qin
School of Computer Science and Technology, Xidian University, Xi'an 710071, China

efficiency has to be considered to decrease the power consumption. In contrast to traditional video coding, most burden of video processing is shifted to the decoders where abundant resources are available.

At present, one of the techniques that apply the strategy mentioned above is Distributed Video Coding (DVC) [15]. In this technique, the intra-frame encoding is implemented independently without performing motion estimation (ME) and motion compensation (MC). Especially as for Wyner-Ziv (WZ) frames, a kind of channel code is applied and then a portion of the resulting parity bits are transmitted. The decoder employs the received WZ bits and the side information (SI) generated from previous decoded frames to carry out WZ frames reconstruction jointly. Another technique aiming to achieve low complexity for encoders is Compressed Sensing (CS) [5, 9]. CS, a new branch of signal processing, can realize the recovery of sparse or compressible signals by only obtaining a small number of non-adaptive linear measurements. CS breaks the limit of Nyquist sampling theorem and needs no additional compression. Although the video coding technique based on CS, as it is often called Compressed Video Sensing (CVS), has considerable advantages over the ones noted above, a gap still exists in the practical application of CVS.

Existing CVS schemes perform encoding for each frame independently. According to the sampling mode, these schemes can be classified into two categories: the plain compressive sampling [2, 12, 19, 40] and the hybrid sampling [8, 25, 27, 32]. In terms of the plain compressive sampling mode, Kang and Lu [19] proposed the DCVS framework to complete video recovery with the help of SI derived from the reconstructed key frames and the modified GPSR algorithm. Asif et al. [2] considered a motion-adaptive linear dynamical system at the decoder to model the temporal variations and designed the correspondent recovery algorithm which were only suitable for frames with small size. Two CVS schemes: MC-BCS-SPL and MH-BCS-SPL are studied in [12]. In the first one, ME/MC was incorporated into the reconstruction process and three main components were included: multi-hypothesis initialization, residual reconstruction and forward/backward MC. The second one took another form of ME/MC—multi-hypothesis (MH) prediction and proposed a Tikhonov regularization-based MH prediction method which cooperates with the residual reconstruction. Ying et al. [40] exploited how to enhance the sparse representation of each frame block and realized it by means of generating Karhunen-Loève (KL) bases adaptively at the decoder. However, it is hard for this means to meet the real-time processing. This video codec had some differences in sampling form. All frames were treated equally without the partition of key frames and non-key frames. The hybrid sampling mode used a traditional video coding technique for key frames. [27] presented a practical acquisition process that combined the compressive sampling with a conventional one and applied it to non-key frames. The key to this acquisition process was the compressive sampling test that identified which blocks were sparse within key frames. Prades-Nebot et al. [25] proposed a CS-based DVC technique in which three coding modes (SKIP, SINGLE and L1 mode) were utilized to improve the coding efficiency of video. Do et al. [8] designed the DISCOS codec in which both block-based and frame-based measurements were acquired for non-key frames. Their main contributions were the interframe sparsity model, sparsity-constraint block prediction algorithm and sparse recovery with decoder SI method. Tzagkarakis et al. [32] proposed a CVS system for remote imaging applications. This system employed the iterative frame refinement process and super-resolution step to enhance the quality of the reconstructed non-key frames. In brief, compared with the plain compressive sampling without any other form of encoding, the hybrid sampling was more complex and required more computing resources. It was not appropriate for the low-cost, low-power encoding devices applied to WMSN. Based on this fact, we just employ the plain compressive sampling on all frames.

It is worth noting that much attention has been paid to the performance at low sampling rates (which means the fewer amounts of data will be acquired), because it can greatly reduce the power consumption of WMSN. In this paper, we introduce the elastic net regression which is widely utilized in statistical learning to CVS and propose an Elastic net-based hybrid hypothesis prediction technique named MS-wEnet. This technique implements the single-hypothesis (SH) mode or the multi-hypothesis (MH) weighted Elastic net (MH-wEnet) mode selectively. Both modes are carried out in the compressed measurement domain at the decoder to obtain the predictions of non-key frames. After that, the residual reconstruction step is applied to accomplish the final recovery. The simulation results and analyses show our proposed scheme is more superior at low sampling rates when compared with the other CVS schemes which apply MH prediction.

The remainder of this paper is organized as follows: Section 2 describes the background of Compressed Sensing and further reviews its application in video processing. The architecture of the proposed CVS scheme is illustrated in Section 3. The MS-wEnet technique is described in detail in Section 4, and then the performance evaluation is carried out in Section 5. Finally, a conclusion for this paper is made in Section 6.

2 Background

2.1 Compressed sensing

Briefly, CS integrates signal acquisition and compression (dimensionality reduction) together and under certain conditions, a small number of linear measurements obtained contain all the information that is necessary for exact recovery of the signal. Specifically, assume that a real-valued signal of interest $x \in R^N$ has sparse representation in a certain basis Ψ, the transform coefficients vector α can be accurately recovered with high probability from a series of projections acquired through

$$y = \Phi x = A\alpha \qquad (1)$$

where y is a measurement vector, $\Phi \in R^{M \times N}$ is a measurement matrix with $M \ll N$ and A denotes $\Phi\Psi$. It involves searching the sparse solution for the underdetermined linear equations when recovering α from y. Under the condition that A conforms the restricted isometry property [5], this can be realized by some iterative greedy algorithms, such as OMP [31], StOMP [10] and CoSaMP [24]. In addition, with the same condition and a convex relaxation applied, the sparse solution can also be obtained by solving the ℓ_1-minimization problem,

$$\min_{\alpha} \|\alpha\|_1, \quad s.t. \ y = A\alpha \qquad (2)$$

In practice, the acquisition of CS measurements is often corrupted by noise,

$$y = \Phi x + e = A\alpha + e \qquad (3)$$

where e represents the noise vector. In this case, to deal with noisy measurements, (2) should be relaxed to yield the following optimization problem,

$$\min_{\alpha} \frac{1}{2} \|y - A\alpha\|_2^2 + \lambda \|\alpha\|_1 \qquad (4)$$

known as basis pursuit denoising (BPDN) [6] criterion. Here, λ is a non-negative parameter. For the advantage of obtaining the global optimal solution, convex optimization approaches for solving (4) achieve better in reconstruction quality than greedy algorithms, but with more computational complexity.

2.2 Video techniques based on CS

In many cases, natural video frames are compressible in DCT or Wavelet basis and a random matrix is selected as Φ to guarantee the incoherence with Ψ. In the field of reconstruction algorithms, a variety of solvers have been proposed against (4). For example, the FPC_AS [39] algorithm combines the characteristic of greedy algorithms with the improved fixed-point continuation (FPC) to deal with the challenging problems arising in CS, such as high dynamic range; the ADMM [4] solver utilizes the simple but powerful alternating direction method of multipliers to handle the large-scale problems efficiently. Moreover, another recovery algorithm named CGIST [16] has also developed. This algorithm is based on iterative shrinkage/thresholding (IST) and takes advantage of the conjugate gradient partan method to accelerate convergence when the active set remains constant.

When we employ the reconstruction algorithms noted above to handle the video streaming, the most straightforward approach is to implement global frame-based acquisition and reconstruct the individual frames independently. However, it causes an increasing storage at the encoder and takes much time to recover at the decoder. To solve these problems, we take the strategy of non-overlapping block splitting, which in turn leads to the severe blocking artifacts. Learning from the method of eliminating blocking artifacts in traditional video processing, the BCS-SPL [22] algorithm incorporates a Wiener-filtering step into the reconstruction. In view of its fast implementation and good reconstruction quality, we adopt the BCS-SPL algorithm for the recovery of video frames.

As to the specific dynamic magnetic resonance imaging (MRI), a number of approaches, such as LS-CS [34], KF-CS [33], modified-CS [35], modified-CS-residual [21], and k-t FOCUSS [18] are suggested in CVS community. However, in comparison with natural video sequences, the dynamic MRI ones have fewer amounts of inter-frame motion and smaller temporal variations in video contents. So these approaches are not suitable for natural video scenes, but we can benefit from the idea of "prediction—residual reconstruction". We can get the conclusion from the CVS schemes described in Section 1 that various ME/MC techniques can be embedded into the CS video receiver/decoder efficiently and that the performance of CVS can also be further improved by merging the respective features of CS and DVC. In addition, we can apply the methods in Matrix Completion, such as MATRIX ALPS [20] and SpaRCS [38], to video surveillance environment in which the static background is explained as a low-rank matrix while the moving foreground objects result in a sparse matrix. Also, we may take the form of 3D processing [14, 36, 37]. In this case, a 3D joint reconstruction is implemented. However, the computation and memory issues will be worse with the amount of data increase.

In essence, all these CVS schemes make efforts to take advantage of the redundancy (which exists in both intra-frame and inter-frame forms), especially the temporal redundancy to improve the performance of video reconstruction.

3 Architecture of the proposed CVS scheme

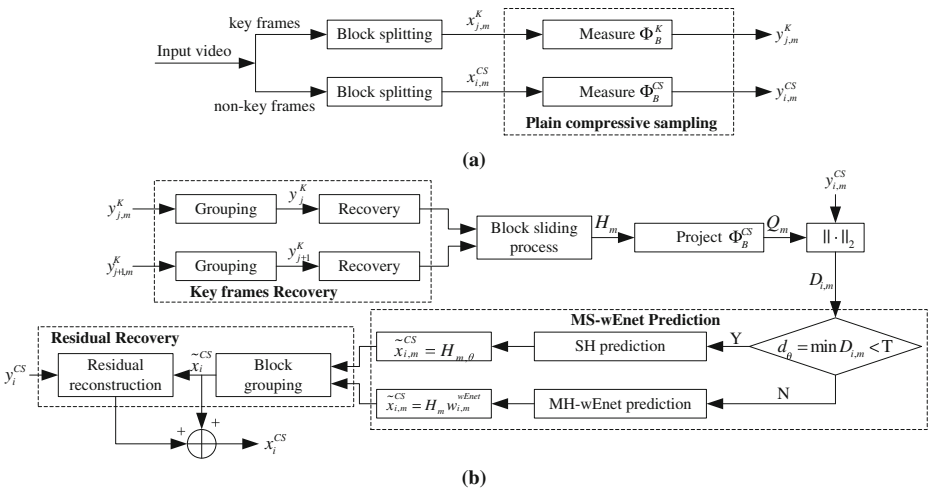The block diagram of our proposed CVS system is presented in Fig. 1.

Fig. 1 Block diagram of the proposed CVS codec: **a** CVS encoder, **b** CVS decoder

Figure 1(a) shows the CVS encoder and Fig. 1(b) illustrates the CVS decoder. The decoder presents the recovery procedure for one non-key frame. It is worth emphasizing that the number of non-key frames plays a decisive role in the amount of data we need to process. In WMSN, with the acceptable visual quality, the more non-key frames, the less power will be consumed on encoding devices. From this block diagram, we can clearly learn that there are four main stages to accomplish the proposed CVS scheme. The first stage, named as plain compressive sampling, carries out at the encoder. At this stage, the block-based measurements will be acquired on key frames and non-key frames separately. Then, at the decoder, key frames will be reconstructed first based on the received and inverse-quantized measurements of them at the second stage. After the block sliding and projection process, the third and the most important stage, MS-wEnet prediction, is implemented to construct the predictions of the blocks in one non-key frame with the aid of the calculated distance sets. The last stage, residual recovery, recovers the difference between the prediction and the target non-key frame. Additionally, the recovery procedure described in Fig. 1(b) can be extended to other non-key frames.

3.1 Design of CVS encoder

As we discussed above, to meet the low-cost and low-power demand, we apply the plain compressive sampling to the video streaming frame-by-frame, block-by-block to obtain measurements. According to the analysis of the distribution of measurements in [2, 12], we divide the video sequence into many Group-of-Pictures (GOPs) which consist of one key frame and several non-key frames. For convenience, we assume $x_{j,m}^K$ is the mth block(vectorized) in the jth key frame and that $x_{i,m}^{CS}$ is the mth block(vectorized) in the ith non-key frame. Moreover, $\Phi_B^K \in R^{M_B^K \times N}$ is defined as the block-based key-frame measurement matrix and $\Phi_B^{CS} \in R^{M_B^{CS} \times N}$ is defined as the block-based non-key-frame measurement matrix. Here, we define M_B^K/N and M_B^{CS}/N as Key-subrate and CS-subrate, respectively. In Fig. 1(a), all of the frames are splitted into blocks with the same size. Each block (vectorized) of key frames is compressively sampled by projecting onto Φ_B^K to get the measurement vector $y_{j,m}^K$, while non-key frames are sampled by Φ_B^{CS} to obtain all measurement vectors $y_{i,m}^{CS}$. Besides, for simple and practical implementation, we adopt the first M_B^{CS} rows of Φ_B^K to form Φ_B^{CS}.

3.2 Design of CVS decoder

The recovery of key frames is completed by using the BCS-SPL algorithm. Same as traditional video coding, to make full use of the temporal redundancy, two neighboring key frames are employed to carry out bi-directional prediction for non-key frames between them. On the reconstructed key frames, the block sliding process is implemented. In this process, a block with the same size of $x_{i,m}^{CS}$ will slide pixel-by-pixel in all directions within the search space associated with $x_{i,m}^{CS}$ and all obtained blocks will be transformed to vectors to generate the hypothesis set H_m. Multiplied by Φ_B^{CS}, H_m is projected to yield the measurement vector set Q_m. The Euclidean distances between every column q_s in Q_m and $y_{i,m}^{CS}$ (the measurement vector corresponds to $x_{i,m}^{CS}$) are calculated to form the distance vector $D_{i,m}$, and then $D_{i,m}$ is used for the implementation of the hybrid hypothesis prediction procedure. Specifically, the hybrid hypothesis prediction is a new method. It selects either the SH or the MH-wEnet prediction mode in the projection domain to produce $\widetilde{x}_{i,m}^{CS}$ as the approximation of $x_{i,m}^{CS}$. These two modes complete the prediction task for non-key frames together. After that, the predicted blocks are arranged via block grouping to acquire the whole predicted $\widetilde{x}_i^{CS}$ followed by the step of residual reconstruction.

As pointed out above, the most essential part in our proposed CVS scheme is the MS-wEnet prediction stage, i.e., the Elastic net-based hybrid hypothesis prediction, and next we will illustrate it in detail.

4 Elastic net-based hybrid hypothesis prediction

In MS-wEnet, MH-wEnet is a new approach to carry out the measurement-domain MH prediction. Therefore, it is necessary to explain the MH prediction employed in CVS first. In the following part, we assume $H_{i,m}$ represents the hypothesis set associated with $x_{i,m}^{CS}$ and that $Q_{i,m}$ denotes the projection set of $H_{i,m}$ by Φ_B^{CS}.

4.1 MH prediction mode in CVS

This mode [8, 30], as well as the SH prediction one, belongs to the ME/MC technique which is more suitable for dealing with complex motion. In CVS, ME/MC only occurs in the CS reconstruction of video. In this process, we can construct the prediction $\widetilde{x}_{i,m}^{CS}$ as the approximation of $x_{i,m}^{CS}$ with the measurement vector $y_{i,m}^{CS}$. Similar to the usage of ME/MC in traditional video coding, in CVS, the more accurate this prediction $\widetilde{x}_{i,m}^{CS}$ is, the more efficiently for the corresponding residual $\widetilde{r}_{i,m}^{CS}$ to be recovered via reconstruction algorithms based on the residual measurements, and then get $x_{i,m}^{CS} = \widetilde{x}_{i,m}^{CS} + \widetilde{r}_{i,m}^{CS}$. Therefore, the key to applying ME/MC to CVS successfully is the prediction methods on the basis of $y_{i,m}^{CS}$.

In the CS reconstruction of video, both MH and SH prediction modes can be implemented in the pixel domain or in the measurement domain. However, from [3, 7, 12, 17], $\widetilde{x}_{i,m}^{CS}$ made in the measurement domain is probably more close to $x_{i,m}^{CS}$ as compared to that made in the pixel domain according to the Johnson-Lindenstrauss (JL) lemma. Therefore, the idea of "prediction—residual reconstruction" is more suitable for the measurement domain and this has been confirmed in [12]. Besides, by selecting multiple candidate hypotheses (columns) h_s from $H_{i,m}$ and then executing an optimal linear combination, the MH mode produces a prediction superior to any other ones constructed by the SH mode.

Based on these facts, we conclude that the predictions obtained by the MH mode in the measurement domain achieve better accuracy than the other ones.

When this mode is carried out, the MH coefficients vector $w_{i,m}^{mh}$ is acquired in the projection domain first and then applied to all hypotheses to obtain the linear prediction $\widetilde{x}_{i,m}^{mh}$ $= H_{i,m} w_{i,m}^{mh}$. Because of the dimensionality reduction caused by Φ_B^{CS}, the optimization problem of acquiring $w_{i,m}^{mh}$ should be added a certain penalty term to obtain a preferable solution. At present, two methods have been proposed for MH prediction in the projection domain.

(a) ℓ_1 regularization-based MH prediction. In [8], the proposed sparsity-constraint block prediction scheme supposes $x_{i,m}^{CS}$ can be represented as a linear combination of a few h_s. In other words, $w_{i,m}^{mh}$ is assumed to be sparse and the support used for combination is a subset of that of $H_{i,m}$. As a result, the ℓ_1 regularization-based MH prediction involves solving

$$w_{i,m}^{\ell_1} = \underset{w}{\operatorname{argmin}} \left|\left| y_{i,m}^{CS} - Q_{i,m} w \right|\right|_2^2 + \lambda \left|\left| w \right|\right|_1 \tag{5}$$

where $w_{i,m}^{\ell_1}$ is a form of $w_{i,m}^{mh}$.

(b) Tikhonov regularization-based MH prediction. This method, as the main part of MH-BCS-SPL [12, 30] which is called MH-Tikhonov in our work assumes the support used for combination is just that of $H_{i,m}$. All h_s make contributions to the prediction. Based on the Euclidean distances between each column q_s in $Q_{i,m}$ and $y_{i,m}^{CS}$, the contributions can be further adjusted. With these distances as prior knowledge, the Tikhonov regularization-based MH prediction can be written as

$$w_{i,m}^{Tik} = \underset{w}{\operatorname{argmin}} \left|\left| y_{i,m}^{CS} - Q_{i,m} w \right|\right|_2^2 + \left|\left| \lambda \Gamma w \right|\right|_2^2 \tag{6}$$

Here, Γ is a Tikhonov matrix which contains the prior knowledge stated, and $w_{i,m}^{Tik}$ is another form of $w_{i,m}^{mh}$. Compared with (5) which is solved by iteration, a closed-form solution is available for (6).

4.2 Elastic net

A conclusion can be drawn from the analysis noted above that the performance of prediction is dominated by the selected support and the associated coefficients. In fact, both ℓ_1 -based and Tikhonov-based prediction methods in the measurement domain can be viewed as two limiting cases in the MH mode for their supports applied in the linear combination. However, to achieve the best prediction performance, h_s which is highly correlated with $x_{i,m}^{CS}$ should be selected as much as possible and the others should be dropped in order to avoid the corruption. With this principle, we introduce and modify Elastic net which is widely used in statistical learning to execute MH prediction in the projection domain for video data.

Associated an ℓ_1 penalty with an ℓ_2 one together, the optimization problem, as it is often called Elastic net [42], can be written as

$$w_{i,m}^{Enet} = \left(1 + \frac{\lambda_2}{n}\right) \underset{w}{\operatorname{argmin}} \left|\left| y_{i,m}^{CS} - Q_{i,m} w \right|\right|_2^2 + \lambda_1 \left|\left| w \right|\right|_1 + \lambda_2 \left|\left| w \right|\right|_2^2 \tag{7}$$

with both λ_1 and λ_2 be non-negative values. $w_{i,m}^{Enet}$ represents the acquired MH coefficients vector. From (7), it is obvious to find out the connection between Elastic net and ℓ_1 regularization or Tikhonov regularization. Here, the introduction of the factor $\left(1 + \frac{\lambda_2}{n}\right)$ is to eliminate the double shrinkage caused by these two regularization terms. It is well known that the elastic net regression combines the merits of the LASSO [28] and ridge regression. Specifically, the ℓ_1 part in the elastic net penalty promotes the sparsity in $w_{i,m}^{Enet}$ and the ℓ_2 part not only overcomes the limitation on the number of columns in $Q_{i,m}$ to be selected but also stabilizes the solution path when $p \gg n$ ($Q_{i,m} \in R^{n \times p}$). More importantly, a grouping effect on $w_{i,m}^{Enet}$ can be encouraged by the ℓ_2 part. This makes Elastic net be able to handle many highly correlated columns in $Q_{i,m}$ perfectly. Unlike the LASSO which tends to select one column and ignore the others, all of these columns can be selected or dropped. From the construction of $H_{i,m}$, we can learn that the correlations between some columns (hypotheses) are so high that they can be treated as a group in $H_{i,m}$. According to the JL lemma, the projections in $Q_{i,m}$ which correspond to the group are also highly correlated, thus it is reasonable for Elastic net to deal with these columns.

To make the statements above more understandable, we take the first frame of *Foreman* sequence (CIF format, 352×288) as the reference frame. On it, the block-based compressive sampling with a block size of B=16 and Key-subrate=0.7 is executed. Then, the BCS-SPL algorithm is used for the recovery of this frame. On the reconstructed reference frame, we extract the search space related to the 25th block of non-key frames with a spatial window size W (15 pixels) to form $H_{i,25}$ and $Q_{i,25}$. Figure 2 presents the acquisition of the search space and the CS-subrate we set for $Q_{i,25}$ is 0.2. The experimental data in Tables 1 and 2 present the correlations between columns 40~43 in $H_{i,25}$ and $Q_{i,25}$, separately.

From Tables 1 and 2, we can conclude that the columns 40~43 are highly correlated with all the $\rho_{a,b}$ above 0.8, especially in $Q_{i,25}$. In this case, Elastic net can be utilized to process these columns effectively. All of them will be abandoned and the corresponding coefficients are set zero, or be selected and allocated similar non-zero coefficients.
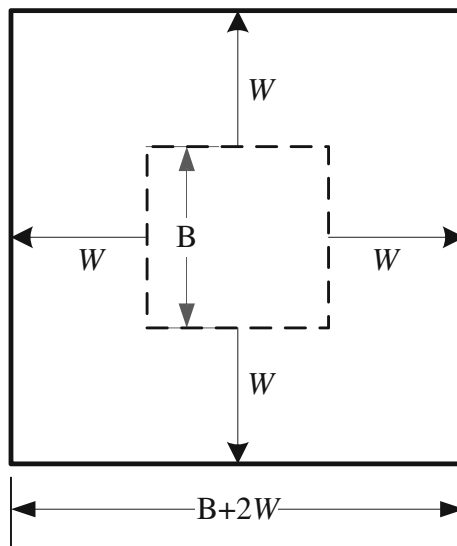

Fig. 2 A block with B×B and the search space with (B+2W)×(B+2W); W is the spatial window size

Table 1 The correlations between columns 40~43 in $H_{i,25}$

$\rho_{a,b}^{H}$	40	41	42	43
40	1	–	–	–
41	0.9314	1	–	–
42	0.8537	0.9327	1	–
43	0.8023	0.853	0.931	1

LARS-EN, an efficient algorithm, is also developed in [42] to obtain the entire Elastic net solution path by the updating and downdating of Cholesky decomposition and [26] makes a further description of this algorithm with the available Matlab software: SpaSM. In the situation where $p \gg n$, we adopt the early stopping [42] which means the optimal solution is achieved before the algorithm runs to end. Besides, LARS-EN requires $O(k^3+pk^2)$ operations after k iterations, so the early stopping can be used to ease the computational burden to acquire the solution.

4.3 Weighted Elastic net

Applying Elastic net simply to video processing does not utilize all the underlying available information. In [43], the described adaptive Elastic net combines the adaptive LASSO [41] and Elastic net together. This adaptive Elastic net, with the pre-computed Elastic net estimator as the penalty factor embedded into the ℓ_1 part, allocates different weight to the corresponding coefficient. As to the multi-hypothesis setting in the projection domain for video data, the distance-weighted rule is employed to form the Tikhonov regularization-based MH prediction in [30]. The greater the distance is, the heavier penalty to the correspondent coefficient will be. We can adopt this rule to Elastic net or replace the pre-computed Elastic net estimator in the adaptive Elastic net by the distance vector $D_{i,m}$ defined in Section 3.2. Compared with the adaptive Elastic net, the resulting weighted Elastic net (wEnet), as we call, is also data-dependent for $D_{i,m}$ varies with the block number m in the frame number i but more efficient in computing without acquisition of the Elastic net solution. Additionally, the wEnet also inherits the characteristics of sparse representation and grouping effect in Elastic net mentioned above. In summary, the proposed Multi-Hypothesis weighted Elastic net (MH-wEnet) which is used in the measurement domain to make the prediction $\widetilde{x}_{i,m}^{wEnet}$ has the following form:

$$w_{i,m}^{wEnet} = \left(1+\frac{\lambda_2}{n}\right)\arg\min_{w}\left\|y_{i,m}^{CS}-Q_{i,m}w\right\|_2^2 + \lambda_1\left\|\mathcal{D}_{i,m}w\right\|_1 + \lambda_2\left\|w\right\|_2^2 \qquad (8)$$

where

$$\widetilde{x}_{i,m}^{wEnet} = H_{i,m}w_{i,m}^{wEnet} \qquad (9)$$

$$\mathcal{D}_{i,m} = \begin{pmatrix} d_1 & & & & 0 \\ & \ddots & & & \\ & & d_s & & \\ & & & \ddots & \\ 0 & & & & d_\Sigma \end{pmatrix}$$

Table 2 The correlations between columns 40~43 in $Q_{i,25}$

$\rho_{a,b}^{Q}$	40	41	42	43
40	1	–	–	–
41	0.997	1	–	–
42	0.9931	0.9971	1	–
43	0.9904	0.992	0.9958	1

$w_{i,m}^{wEnet}$ is the MH coefficients vector in MH-wEnet and $d_s = \|y_{i,m}^{CS} - \Phi_B^{CS} h_s\|_2$ is also the element in the distance vector $D_{i,m}$. Σ represents the total number of columns in $Q_{i,m}$.

For the computation, the LARS-EN algorithm can still be employed to obtain the solution path efficiently based on the adjusted $Q_{i,m}$ by $D_{i,m}$ and the early stopping is also adopted.

4.4 MH/SH wEnet

We have designed the measurement-domain MH-wEnet mode to handle the problems of support selection and coefficients adjustment. But from the construction of $D_{i,m}$, we can see that the value of d_s will be extremely close to zero when $y_{i,m}^{CS}$ is approximately equal to q_s and that in the special case where $y_{i,m}^{CS} = q_s$, d_s will precisely be zero. When one of these two cases happens, the LARS-EN algorithm used to solve the wEnet will become unstable because of the implementation of Cholesky decomposition. To avoid this and improve the robustness of the prediction procedure, we take the approach of SH prediction and combine it with MH-wEnet to complete the prediction task for non-key frames.

In CVS situation, the SH mode has two options for implementation. In the pixel domain [2, 19, 23, 32], it is similar to the traditional ME/MC framework and an ME operation is made between an initial predicted frame and a reference frame to form the associated field of motion vectors. This field is applied to produce the more precise prediction. In the measurement domain [25], a single and best-matching hypothesis h_s is chosen as $\tilde{x}_{i,m}^{CS}$ according to some distortion standards, such as the minimum mean square error (MMSE), carried out between each measurement vector q_s and the target $y_{i,m}^{CS}$. As noted in Section 4.1, the prediction calculated in the measurement domain is more precise than that in the pixel domain. Moreover, because MH-wEnet has its implementation in the measurement domain, the SH mode should also keep the same manner to guarantee the efficiency of prediction.

Specifically, as Fig. 1(b) shows, after the calculation of $D_{i,m}$, the minimum d_θ in it will be found and then compared with a suitable threshold T which is determined by experience and requirements. According to the results of judgment, either the SH or the MH-wEnet prediction mode in the projection domain will be selected to execute. That is the reason why we call MS-wEnet a hybrid prediction method. Once the SH mode is chosen to carry out, the θth column in H_m, i.e., $H_{m,\theta}$ will be viewed as the prediction of the vectorized $x_{i,m}^{CS}$ directly and then added to predictions of the other blocks to constitute the whole prediction of x_i^{CS} via block grouping.

Here, there may exist a problem that since SH prediction is less attractive than MH prediction in quality, the participation of SH prediction may lead to a reduction in the final reconstruction quality. It should be emphasized that the number of blocks which go through the process of SH prediction is much less than that through MH prediction. And if the threshold T is small enough, $x_{i,m}^{CS}$ can be accurately approximated by $H_{m,\theta}$ according to the JL lemma. As a result, the quality achieved by the hybrid hypothesis prediction method (MH/SH wEnet, and MS-wEnet for simplicity) will not be deteriorated compared with that by the solely MH mode. Moreover, the prediction procedure will also be speeded up because

the implementation of SH prediction is much faster than MH one. This will be presented in the simulation part.

With the discussion above, we have described the proposed CVS system in detail, especially the vital component—MS-wEnet at the decoder. The performance of this scheme will be shown and analyzed in the next section.

5 Experiments and results

In order to make the performance evaluation, in all our experiments, we take the standard test video sequences with CIF format available at http://trace.eas.asu.edu/yuv/ and use the luminance component. A block size of B=16 and an orthonormalized Gaussian Φ_B are applied to the block-based plain compressive sampling. The Key-subrate we use is 0.7 and CS-subrate varies from 0.1 to 0.5. The dual-tree discrete wavelet transform (DDWT) with four levels of decomposition is used as the sparsity-inducing basis for reconstruction and the residual reconstruction is only executed once by using the BCS-SPL algorithm. Moreover, in those schemes implementing MH/MS prediction, the spatial window size W, as shown in Fig. 2, is set to be 15 pixels. All of the experiments are carried out under the configuration of Windows 7 SP1, Intel Core i3-2330 M, CPU 2.20GHz, 2GB RAM. The version of Matlab is R2010(b).

Before the performance of the proposed CVS scheme is demonstrated, there are still several problems that need to be handled.

5.1 Parameters in MS-wEnet

In the MH-wEnet mode, the Matlab software: SpaSM is adopted to solve the wEnet presented in (8) and (9). There are two parameters λ_1 and λ_2 that need to be predetermined in the wEnet. In other words, we should define the values of *delta* and *stop* (*delta* is equal to λ_2 and $-stop$ which is related to λ_1 denotes the number of hypotheses to be selected) first for SpaSM. We can learn from the JL lemma that $\|y_{i,m}^{CS}-Q_{i,m}w\|_2$ will be close to $\|x_{i,m}^{CS}-H_{i,m}w\|_2$ with a higher probability and meanwhile the correlations between each q_s in $Q_{i,m}$ and $y_{i,m}^{CS}$ will be distinguished more efficiently as CS-subrate increases. So it allows us to select more
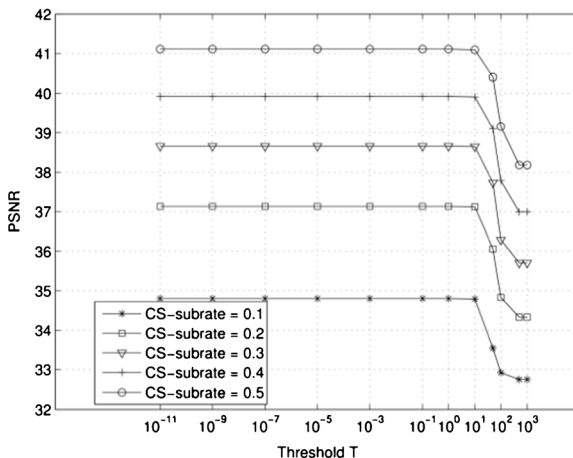

Fig. 3 Reconstruction quality of frame 5 of *Foreman* sequence with different T at various CS-subrates
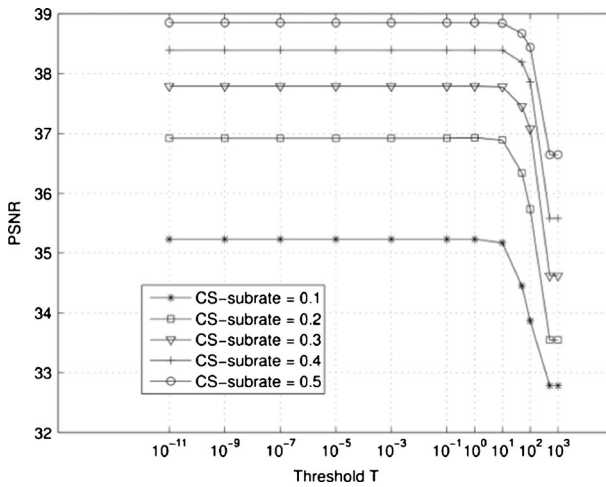
Fig. 4 Reconstruction quality of frame 5 of *News* sequence with different T at various CS-subrates

h_s in $H_{i,m}$ to make contributions to the prediction with higher CS-subrate. Based on the analysis and experimental observations over a large set of different frames, we find $delta \in [0.01, 0.2]$ and $stop = -1000 \times CS$−subrate provide the better results. As a result, we use $delta = 0.1$ and $stop = -1000 \times CS$−subrate from this point on.

Because of the application background of WMSN, there are abundant computing resources available at the receiver. So compared with the reconstruction time, the reconstruction quality is more important. Next, we will analyze the effect that the threshold T makes on the reconstruction quality in detail.

In fact, the lower limit of the threshold T is the minimum value of d_θ that the LARS-EN algorithm allows when solving (8) stably. It is found that when d_θ is lower than 1e-11, the adjusted $Q_{i,m}$ (each adjusted column $q'_s = q_s./d_s$) is badly conditioned and this leads to the instability of LARS-EN. For this reason, the lower limit of T we set is 1e-11. Based on this
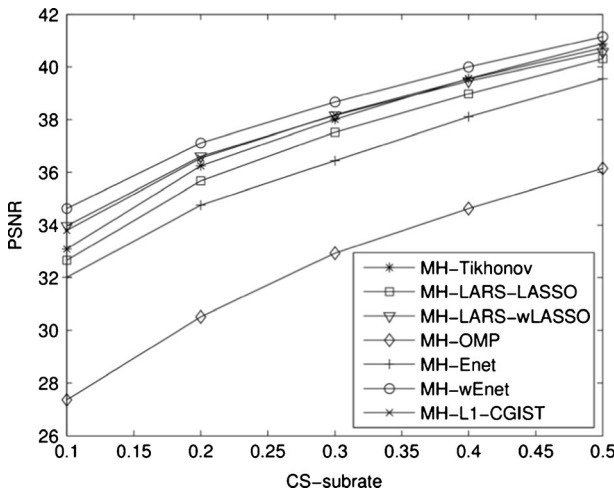

Fig. 5 Reconstruction quality of frame 5 of *Foreman* sequence for various MH prediction methods

Table 3 PSNR(dB) and reconstruction time(s) of frame 5 of *Foreman* sequence for various MH methods

Method		CS-subrate				
		0.1	0.2	0.3	0.4	0.5
MH-OMP	PSNR	27.36	30.51	32.93	34.63	36.15
	Time	10.7	16.0	25.3	41.1	72.8
MH-LARS-LASSO	PSNR	32.66	35.68	37.52	38.98	40.31
	Time	42.9	78.0	119.0	173.9	229.2
MH-LARS-wLASSO	PSNR	33.97	36.61	38.16	39.47	40.57
	Time	30.3	61.0	97.2	138.8	200.8
MH-Enet	PSNR	32.01	34.76	36.44	38.11	39.56
	Time	81.3	170.2	286.7	479.5	766.1
MH-wEnet	PSNR	34.63	37.12	38.67	40	41.15
	Time	47.4	101.0	174.2	311.6	520.7
MH-Tikhonov	PSNR	33.09	36.25	38.01	39.57	40.88
	Time	173.3	168.2	170.6	173.7	179.2
MH-L1-CGIST	PSNR	33.80	36.54	38.18	39.55	40.72
	Time(h)	1.41	2.05	2.78	3.55	4.50

lower limit and without loss of generality, we utilize the frames 1,5,9 of *Foreman* and *News* sequences to determine the scope of T in which the reconstruction quality of MS-wEnet achieves the optimum. The frames 1,9 are treated as key frames and frame 5 as non-key frame to be recovered. Figures 3 and 4 show the relationship between the threshold T and the reconstruction quality of frame 5 at different CS-subrates. The reconstruction quality is measured in peak signal-to-noise ratio (PSNR).

It can be observed that the trend of all curves is consistent. When $T \in [\ 1e{-}11,\ 1\ )$, the reconstruction quality remains unchanged. The reason is that few d_θ are in $[\ 1e{-}11,\ 1\ )$ and the
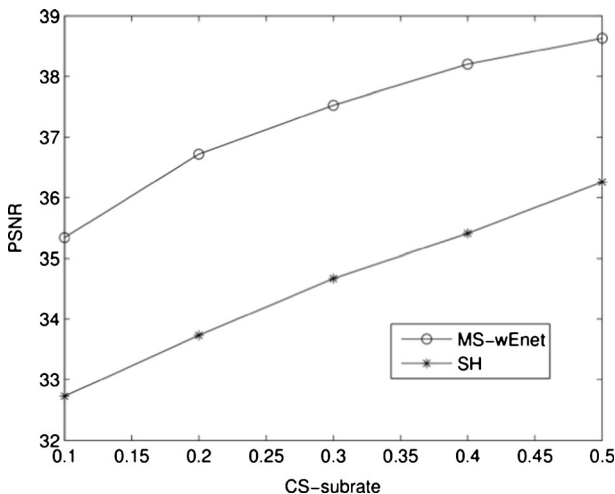

Fig. 6 Reconstruction quality for frame 5 of *News* sequence for MS-wEnet/SH

Table 4 PSNR(dB), reconstruction time(s) and SH proportion for frame 5 of *News* sequence for MS-wEnet/SH

Method	CS-subrate				
		0.1	0.2	0.3	0.4	0.5
MS-wEnet	PSNR	35.34	36.72	37.53	38.2	38.63
	Time	38.16	74.99	137.12	241.64	385.98
	SH proportion	0.313	0.313	0.313	0.313	0.313
SH	PSNR	32.72	33.73	34.67	35.41	36.26
	Time	13.78	14.47	14.31	14.93	16.1
	SH proportion	1	1	1	1	1

block corresponding to d_θ which is in [1e−11, 1) can be approximated by SH prediction. When $T \in$ [1, 500], the reconstruction quality reduces rapidly with the increasing T which indicates most d_θ are in [1, 500]. When T>500, for the reason that all blocks employ the reconstruction method based on SH prediction, the reconstruction quality does not change any more. With this analysis, in order to achieve the best reconstruction quality stably by applying MS-wEnet, we should select T from [1e−11, 1). Therefore, T=1e-7 is used from this point on.

5.2 Comparison of various MH prediction methods

In our decoding procedure, the acquired $H_{i,m}$ and $Q_{i,m}$ can be viewed as two special dictionaries in the spatial domain and measurement domain, respectively. It exists many kinds of techniques that can be employed to settle the MH prediction problem based on $Q_{i,m}$. Here, we take five typical algorithms: OMP, CGIST, LARS [26], LARS-EN and Tikhonov regularization [30]. The OMP is an iterative greedy algorithm, while CGIST is an ℓ_1 - regularization solver by convex optimization. The LARS and LARS-EN are two fast algorithms based on Least Angle Regression [11] for line regression: the LASSO and Elastic net. At last, in Tikhonov regularization, an ℓ_2 -penalty term that contains the prior knowledge
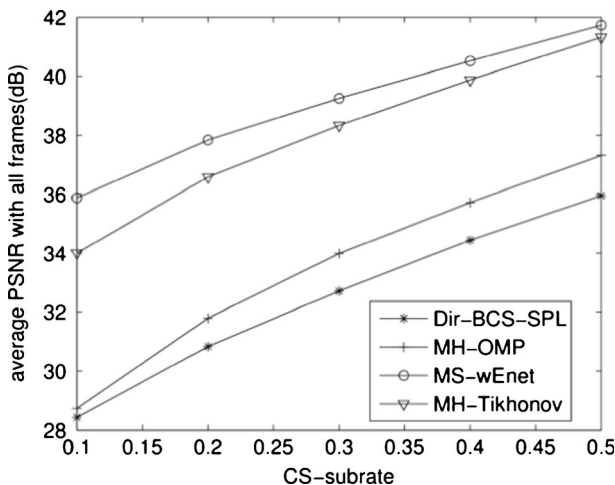

Fig. 7 Performance comparison with the first 88 frames of *Foreman* sequence
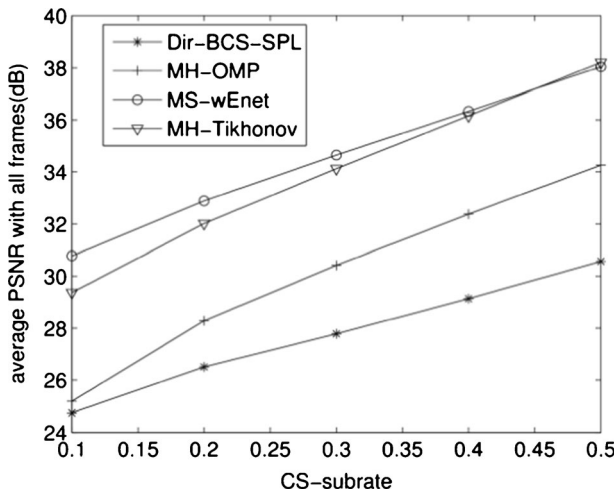
Fig. 8 Performance comparison with the first 88 frames of *Coastguard* sequence

is imposed on the norm of $w_{i,m}^{mh}$. All of the algorithms except Tikhonov regularization, will provide the sparse solution $w_{i,m}^{mh}$. Additionally, we also extend the distance-weighted rule to the LASSO, yielding another MH prediction method, MH-LARS-wLASSO. So there are seven methods which are demonstrated in Fig. 5 in total to compare.

As a representative, we take the frames 1,5,9 of *Foreman* sequence as an example. The frames 1,9 are treated as key frames and frame 5 as non-key frame to be recovered. In OMP, the maximum number of iterations K equals 500 and in CGIST, the value of the parameter *mu* we set is 100. To solve Enet (Elastic net) and wEnet, we take *delta*=0.1 and *stop*=−1000×CS−subrate. As for the LASSO and wLASSO, SpaSM is also utilized to obtain the solution with *delta*=0 and *stop*=−1000×CS−subrate. Note that the number of the selected hypotheses will be limited by n when applying LASSO or wLASSO. In MH-Tikhonov,
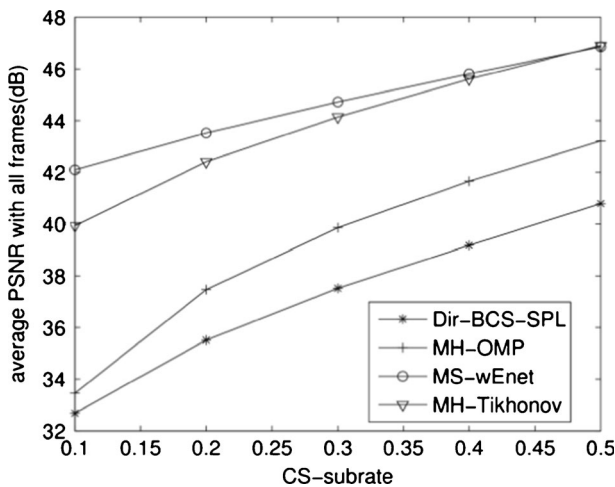

Fig. 9 Performance comparison with the first 88 frames of *Mother and Daughter* sequence
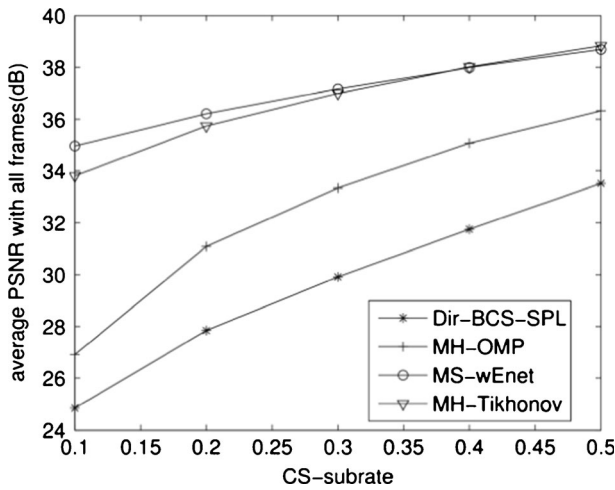
Fig. 10 Performance comparison with the first 88 frames of *Hall Monitor* sequence

lambda=0.25, the same setting with [12, 30]. The comparison of reconstruction quality for various methods is shown in Figure 5. And Table 3 presents the associated experiment data.

Here, it should be emphasized that when using MH-wEnet or MH-LARS-wLASSO on the three frames we select, there is no unstable situation which is described in Section 4.4. In other words, the SH mode has no contribution to the reconstruction.

As Fig. 5 shows, MH-wEnet almost outperforms the other methods in reconstruction quality but MH-OMP is the worst. The suboptimal ones are MH-L1-CGIST and MH-LARS-wLASSO. The corresponding curves of these methods are very close to each other. MH-Tikhonov has a slight advantage over the suboptimal ones at high CS-subrates. From the quality formed by MH-LARS-LASSO and MH-Enet, we can see that the improvement of PSNR can not always benefit from the simply more selected h_s from the hypothesis set $H_{i,m}$. It is essential to apply the prior knowledge contained in $D_{i,m}$ to further distinguish the contribution made by each h_s to the linear combination.
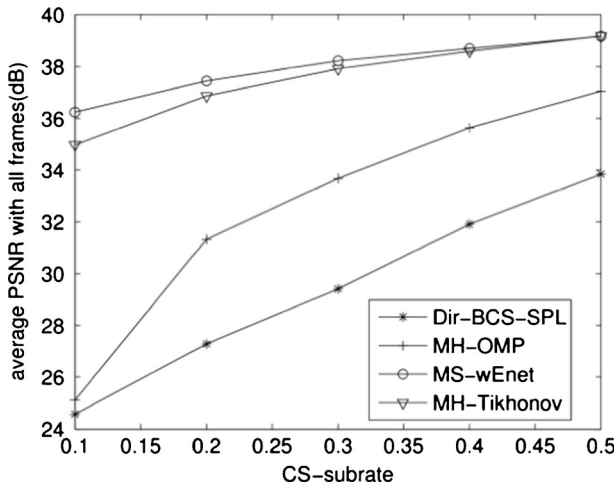

Fig. 11 Performance comparison with the first 88 frames of *News* sequence

According to the reconstruction time (CPU time) for various MH methods provided in Table 3, clearly, MH-OMP is the fastest method, while MH-L1-CGIST is the slowest one. So it is not suitable for the real-time processing. It takes nearly the same amount of time to implement the recovery process for MH-Tikhonov at each CS-subrate. The remaining four methods need more and more time for reconstruction with the increasing CS-subrate. Additionally, it can be found that the application of the distance-weighted rule makes MH-LARS-wLASSO and MH-wEnet much faster than that without it.

With the analysis above, MH-wEnet, MH-Tikhonov, and MH-LARS-wLASSO are the better ones. However, for the application background of WMSN, much attention is paid to the performance at low sampling rates, especially the reconstruction quality because of the abundant computing resources available at the receiver. From Fig. 5 and Table 3, we can see that with MH-Tikhonov to be the baseline, MH-wEnet is more competitive than MH-LARS-wLASSO in terms of reconstruction quality at low CS-subrates, though some longer time is required. In addition, MH-LARS-wLASSO does not only ignore the grouping effect but also inherit the instability of the LASSO solution path when q_s are highly correlated. Therefore, between the two proposed methods, MH-wEnet and MH-LARS-wLASSO, we select the former to compare with MH-Tikhonov which is one of the best methods in reconstruction quality as we know at present. But for the reason that MH-wEnet is not suitable for all video sequences, we adopt MS-wEnet instead of MH-wEnet.

Table 5 Average PSNR(dB) for the first 88 frames of several video sequences

Sequence	Method	CS-subrate				
		0.1	**0.2**	**0.3**	0.4	0.5
Foreman	Dir-BCS-SPL	28.42	30.83	32.72	34.44	35.95
	MH-OMP	28.73	31.79	33.99	35.71	37.32
	MS-wEnet	**35.87**	**37.84**	**39.26**	**40.51**	**41.73**
	MH-Tikhonov	34.0	36.58	38.33	39.88	41.32
Coastguard	Dir-BCS-SPL	24.75	26.51	27.78	29.14	30.56
	MH-OMP	25.18	28.28	30.41	32.37	34.26
	MS-wEnet	**30.78**	**32.88**	**34.65**	**36.33**	38.04
	MH-Tikhonov	29.37	32.02	34.13	36.16	**38.20**
Mother and daughter	Dir-BCS-SPL	32.68	35.51	37.52	39.18	40.79
	MH-OMP	33.45	37.47	39.86	41.66	43.22
	MS-wEnet	**42.10**	**43.52**	**44.72**	**45.81**	46.84
	MH-Tikhonov	39.93	42.41	44.14	45.61	**46.90**
Hall monitor	Dir-BCS-SPL	24.84	27.84	29.91	31.75	33.52
	MH-OMP	26.91	31.09	33.34	35.07	36.33
	MS-wEnet	**34.97**	**36.21**	**37.17**	**37.99**	**38.68**
	MH-Tikhonov	33.81	35.74	37.0	38.02	38.83
News	Dir-BCS-SPL	24.57	27.28	29.41	31.91	33.84
	MH-OMP	25.11	31.33	33.68	35.64	37.05
	MS-wEnet	**36.24**	**37.44**	**38.22**	**38.69**	39.16
	MH-Tikhonov	34.98	36.87	37.91	38.59	**39.18**

The bold PSNR values represent the highest PSNR for each video sequence at different CS-subrates

5.3 Contribution comparison of MH-wEnet and SH to MS-wEnet

Because MS-wEnet involves the mode selection, it is necessary to figure out the contributions two modes make to the reconstruction performance gain of MS-wEnet. In this time, to make SH mode play a role in MS-wEnet under the threshold T=1e-7, we choose frames 1,5,9 of *News* sequence and the treatment is the same as that in Section 5.2. Figure 6 shows the reconstruction quality of the 5th frame by applying MS-wEnet and solely SH method, separately. Note that MMSE distortion standard is used in the solely SH method. Table 4 lists the PSNR values, the reconstruction time and the proportion of SH mode in either method.

From these results, we can see the reconstruction quality achieved by MS-wEnet is significantly better than that by solely SH method. So it can be inferred that in terms of prediction quality, the MH-wEnet mode is much better than the SH mode. From Table 4, we can know the reconstruction time for solely SH method is much less than that for MS-wEnet. So it can be concluded that in terms of prediction speed, the SH mode is much faster than the MH-wEnet mode. It should be emphasized that in the reconstruction method applying MS-wEnet, the proportion of blocks using the SH mode is 0.313 only. Therefore, in comparison with the SH mode, the contribution made by the MH-wEnet mode is dominant in the reconstruction performance gain of MS-wEnet.

Table 6 Average reconstruction time(s) per frame for all methods at each CS-subrate

Sequence	Method	CS-subrate				
		0.1	**0.2**	**0.3**	0.4	0.5
Foreman	Dir-BCS-SPL	10.8	7.2	5.7	4.4	3.9
	MH-OMP	65.9	70.7	79.2	93.3	118.7
	MS-wEnet	96.3	137.1	202.0	313.0	503.5
	MH-Tikhonov	187.3	191.7	195.6	200.5	206.5
Coastguard	Dir-BCS-SPL	9.7	6.0	5.1	4.2	3.9
	MH-OMP	66.0	70.2	78.4	91.5	116.3
	MS-wEnet	94.9	137.4	202.8	314.0	502.3
	MH-Tikhonov	187.5	190.3	194.0	197.1	200.9
Mother and daughter	Dir-BCS-SPL	11.1	5.9	5.2	4.2	3.8
	MH-OMP	66.7	73.0	78.7	93.5	117.3
	MS-wEnet	96.1	139.3	201.8	315.3	502.8
	MH-Tikhonov	190.9	200.7	194.7	202.0	203.9
Hall monitor	Dir-BCS-SPL	12.8	6.8	5.9	4.9	4.3
	MH-OMP	72.0	76.1	86.9	100.9	127.6
	MS-wEnet	105.1	150.3	222.1	338.5	538.8
	MH-Tikhonov	204.8	206.9	213.1	215.7	221.7
News	Dir-BCS-SPL	9.6	7.9	6.0	4.8	4.4
	MH-OMP	69.1	74.2	79.1	89.3	106.7
	MS-wEnet	90.0	120.3	162.8	239.2	367.2
	MH-Tikhonov	199.8	208.3	206.4	211.5	215.6

Sampling rates: 0.1, 0.2, and 0.3 in bold represent the low sampling rates

5.4 Performance of the proposed CVS scheme for video

We are more interested in the performance of the proposed CVS system in entire video sequences. So a comprehensive comparison of the proposed CVS scheme for video sequences with the other ones is made in this section. Note that all CVS schemes take the same plain compressive sampling, though different strategies are utilized for recovery. As a result of Sections 5.2 and 5.3, MS-wEnet and MH-Tikhonov are the main participants in the final comparison. Besides these, MH-OMP and Dir-BCS-SPL which directly applies BCS-SPL to reconstruction are added. The MH-OMP method can be considered as a benchmark for the methods using MH prediction and by comparison with Dir-BCS-SPL, we can see how efficient MH prediction is to improve the reconstruction quality.

In the experiments, we employ the first 88 frames of the commonly used video test sequences, *Foreman*, *Coastguard*, *Mother and Daughter*, *Hall Monitor*, and *News*. At the CVS encoder, we define a GOP size of $P=8$. At the CVS decoder, in the schemes that take MH/MS prediction strategy, for fair comparison, the same reconstructed key frames must be used for the recovery of non-key frames. Moreover, the better accuracy key frames are reconstructed, the more margins will be obtained in quality for non-key frames. So the enhancement process [29] in which the temporally neighboring non-key frames are utilized to re-recover and boost key frames should be applied to MH-OMP, MH-Tikhonov, and MS-

(a)

(b)

(c)

(d)

Fig. 12 Reconstruction of frame 15 of *Foreman* sequence with CS-subrate = 0.1: **a** Dir-BCS-SPL,PSNR = 26.83 dB **b** MH-OMP,PSNR = 26.93 dB **c** MH-Tikhonov,PSNR = 33.31 dB **d** MS-wEnet,PSNR = 35.86 dB

wEnet. Additionally, the parameter setting is the same with that in Sections 5.1 and 5.2 and the Matlab code for MH-Tikhonov can be available at http://www.ece.msstate.edu/~ewt16/publications/.

Figures 7, 8, 9, 10 and 11 illustrate the performance of the four CVS schemes for varying CS-subrates. The PSNR is averaged over all 88 frames of each sequence. Table 5 shows the same experiment data as well and Table 6 presents the average reconstruction time per frame for various methods at each CS-subrate. Visual results of the reconstructed frame 15 of *Foreman* sequence with CS-subrate=0.1 are depicted in Fig. 12.

In terms of reconstruction quality, we can observe intuitively that the CVS schemes based on MS-wEnet and MH-Tikhonov have a great improvement compared with the remaining ones. Furthermore, at low sampling rates (CS-subrate<0.3), MS-wEnet always outperforms MH-Tikhonov and at high sampling rates, such as CS-subrate=0.5 for all video sequences except *Foreman*, MH-Tikhonov has a very slight advantage over MS-wEnet. The comparison between MH-OMP and Dir-BCS-SPL shows the usage of MH prediction can enhance the reconstruction quality greatly.

Besides the reconstruction quality, we also take the average reconstruction time per frame measured by the CPU time into consideration, though none of these schemes has been specially optimized for execution speed. As Table 6 presents, it takes the shortest time for Dir-BCS-SPL to reconstruct each frame and a longer time for MH-OMP. As for MH-Tikhonov, the average time spent on recovery has small changes with CS-subrates for each video sequence. At last, we can find the corresponding time for MS-wEnet has a characteristic that more time is needed with the increasing CS-subrate because of the linear relation between the parameter *stop* and CS-subrate. However, by comparison with MH-Tikhonov, MS-wEnet is faster at low sampling rates. Moreover, in the special case of *News* sequence, we can observe the reconstruction time is reduced further for the reason that more blocks need to be processed by the SH mode. As an extension, many measures can be taken to speed up the CVS scheme based on MS-wEnet. For example, we can replace LARS-EN with the glmnet [13] algorithm to solve the wEnet for its faster implementation; in the enhancement process for key frames, MS-wEnet can be utilized to produce the temporally neighboring non-key frames at low sampling rates; the value of threshold T in MS-wEnet can also be increased greatly to reduce the number of blocks that need to be dealt with by MH-wEnet, though a little loss will be made in reconstruction quality.

In conclusion, because of the application background of WMSN, the low sampling rates draw much attention and at such rates, the proposed CVS scheme based on MS-wEnet is superior to that based on MH-Tikhonov not only in the reconstruction quality but also in the recovery time.

6 Conclusions

In this paper, we have studied how to make efficient use of the temporal redundancy to improve the performance of Compressed Video Sensing in WMSN. The proposed CVS scheme based on MS-wEnet performs the hybrid hypothesis prediction in the measurement domain for each non-key frame. It means either the SH or the MH-wEnet prediction mode is selected to carry out and that both modes are implemented in the projection domain constructed by the measurement matrix. With the prediction process completed, a residual reconstruction step utilizing the measurement-domain residual is followed to accomplish the recovery of non-key frames. By comparison with various MH prediction methods, we can see that MH-wEnet provides a better reconstruction quality. Next, by the comparison made

between MS-wEnet and MH-Tikhonov applied to the video sequences, we can see the proposed CVS scheme outperforms that based on MH-Tikhonov, especially at low sampling rates. In terms of future work, to make the proposed CVS scheme more practical in WMSN, we will work on the issue of quantization for CS measurements which is largely underdeveloped in CS community, and the adaptive strategy for compressive sampling at the encoder will also be considered to reduce the amount of data more.

References

1. Akyildiz IF, Melodia T, Chowdhury KR (2007) A survey on wireless multimedia sensor networks. Comput Netw 51(4):921–960
2. Asif MS, Fernandes F, Romberg J (2011) Low-complexity video compression and compressive sensing. In: Duke workshop http://users.ece.gatech.edu/~sasif/
3. Baraniuk R, Davenport M, DeVore R, Wakin M (2008) A simple proof of the restricted isometry property for random matrices. Constr Approx 28(3):253–263
4. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends® Mach Learn 3(1):1–122
5. Candès EJ, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory 52(2):489–509
6. Chen SS, Donoho DL, Saunders MA (2001) Atomic decomposition by basis pursuit. SIAM Rev 43(1):129–159
7. Chen C, Tramel EW, Fowler JE (2011) Compressed-sensing recovery of images and video using multihypothesis predictions. In: Proceedings of the asilomar conference on signals, systems and computers, pp 1193–1198. Pacific Grove, CA
8. Do TT, Chen Y, Nguyen DT, Nguyen N, Gan L, Tran TD (2009) Distributed compressed video sensing. In: Proceedings of the international conference on image processing, pp 1393–1396. Cairo, Egypt
9. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52(4):1289–1306
10. Donoho DL, Tsaig Y, Drori I, Starck J-L (2012) Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. IEEE Trans Inf Theory 58(2):1094–1121
11. Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least angle regression. Ann Stat 32(2):407–499
12. Fowler JE, Mun S, Tramel EW (2012) Block-based compressed sensing of images and video. Found Trends® Signal Process 4(4):297–416
13. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33(1):1
14. Gamper U, Boesiger P, Kozerke S (2008) Compressed sensing in dynamic MRI. Magn Reson Med 59(2):365–373
15. Girod B, Aaron AM, Rane S, Rebollo-Monedero D (2005) Distributed video coding. Proc IEEE 93(1):71–83
16. Goldstein T, Setzer S (2010) High-order methods for basis pursuit. UCLA CAM Report, 10–41
17. Johnson WB, Lindenstrauss J (1984) Extensions of Lipschitz mappings into a Hilbert space. Contemp Math 26(189–206):1
18. Jung H, Sung K, Nayak KS, Kim EY, Ye JC (2009) k–t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. Magn Reson Med 61(1):103–116
19. Kang L-W, Lu C-S (2009) Distributed compressive video sensing. In: Proceedings of the international conference on acoustics, speech, and signal processing, pp 1169–1172. Taipei, Taiwan
20. Kyrillidis A, Cevher V (2012) Matrix alps: Accelerated low rank and sparse matrix reconstruction. In: IEEE statistical signal processing workshop, pp 185–188
21. Lu W, Li T, Atkinson IC, Vaswani N (2011) Modified-cs-residual for recursive reconstruction of highly undersampled functional mri sequences. In: Proceedings of the international conference on image processing, pp 2689–2692. Brussels, Belgium
22. Mun S, Fowler JE (2009) Block compressed sensing of images using directional transforms. In: Proceedings of the international conference on image processing, pp 3021–3024. Cairo, Egypt

23. Mun S, Fowler JE (2011) Residual reconstruction for block-based compressed sensing of video. In: IEEE data compression conference. Snowbird, UT, pp 183–192
24. Needell D, Tropp JA (2009) CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Appl Comput Harmon Anal 26(3):301–321
25. Prades-Nebot J, Ma Y, Huang T (2009) Distributed video coding using compressive sampling. In: Proceedings of the picture coding symposium, pp 1–4. Chicago, IL
26. Sjöstrand K, Ersbøll B (2012) SpaSM—a Matlab toolbox for sparse statistical modeling. Software available at http://www2.imm.dtu.dk/projects/spasm/
27. Stankovic V, Stankovic L, Cheng S (2008) Compressive video sampling. In: Proceedings of the European signal processing conference, pp 2–6. Lausanne, Switzerland
28. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Roy Stat Soc Ser B 58(1):267–288
29. Tramel EW (2012) Distance-weighted regularization for compressed-sensing video recovery and supervised hyperspectral classification. Ph.D. thesis, Mississippi State University
30. Tramel EW, Fowler JE (2011) Video compressed sensing with multihypothesis. In: IEEE data compression conference, pp 193–202. Snowbird, UT
31. Tropp JA, Gilbert AC (2007) Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inf Theory 53(12):4655–4666
32. Tzgakarakis G, Woiselle A, Tsakalides P, Starck J-L (2012) Design of a compressive remote imaging system compensating a highly lightweight encoding with a refined decoding scheme. In: Proceedings of the international conference on computer vision theory and applications, pp 24–26. Rome, Italy
33. Vaswani N (2008) Kalman filtered compressed sensing. In: Proceedings of the international conference on image processing, pp 893–896. San Diego, CA
34. Vaswani N (2010) LS-CS-residual (LS-CS): compressive sensing on least squares residual. IEEE Trans Signal Process 58(8):4108–4120
35. Vaswani N, Lu W (2010) Modified-CS: modifying compressive sensing for problems with partially known support. IEEE Trans Signal Process 58(9):4595–4607
36. Wakin M, Laska J, Duarte M, Baron D, Sarvotham S, Takhar D, Kelly K, Baraniuk RG (2006) Compressive imaging for video representation and coding. In: Proceedings of the picture coding symposium. Beijing, China
37. Wakin MB, Laska JN, Duarte MF, Baron D, Sarvotham S, Takhar D, Kelly KF, Baraniuk RG (2006) An architecture for compressive imaging. In: Proceedings of the international conference on image processing, pp 1273–1276. Atlanta, GA
38. Waters AE, Sankaranarayanan AC, Baraniuk RG (2011) SpaRCS: Recovering low-rank and sparse rmatrices from compressive measurements. In: Neural Information Processing Systems, pp 1089–1097
39. Wen Z, Yin W, Goldfarb D, Zhang Y (2010) A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation. SIAM J Sci Comput 32(4):1832–1857
40. Ying L, Ming L, Pados DA (2013) Motion-aware decoding of compressed-sensed video. IEEE Trans Circ Syst Video Technol 23(3):438–444
41. Zou H (2006) The adaptive lasso and its oracle properties. J Am Stat Assoc 101(476):1418–1429
42. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J Roy Stat Soc B 67(2):301–320
43. Zou H, Zhang HH (2009) On the adaptive elastic-net with a diverging number of parameters. Ann Stat 37(4):1733

Jian Chen was born in Jiangsu, P.R. China. He received the B.S. degree from Xian Jiaotong University, China in 1989, the M.S. degree from Xian Institute of Optics and Precision Mechanics of Chinese Academy of Sciences in 1992, the Ph.D. in teletcommunication engineering in Xidian University, China in 2005. He is a professor in the school of telecommunications engineering in Xidian University. He was a visitor scholar in the University of Manchester from 2007 to 2008. His research interests include cognitive radio, OFDM and wireless multimedia sensor networks.

Yunzheng Chen was born in Hebei, P.R. China. He received his Bachelor's degree in Electronic Information Engineering from Agricultural University of Hebei, P.R. China in 2011. Now he is pursuing his Master's degree of Communication and Information Systems in Xidian University. His current research interests involve compressed sensing, image and video processing, and distributed video coding.

Dong Qin was born in Shaanxi, P.R. China. He received his Bachelor's degree in Xidian University, P.R. China in 2011. Now he is pursuing his Master's degree of Computer Architecture in Xidian University. His current research interests involve network security, sparse coding, and network information processing.

Yonghong Kuo was born in Henan, P.R. China. She received the B.S. and M.S. degrees from Xian Jiaotong University in 1989 and 1992 respectively, the Ph.D. degree from Peking Union Medical University in 1998. She is a professor in the school of telecommunications engineering in Xidian University. Her research interests include OFDM systems, cognitive radio and wireless multimedia sensor networks.