An improved distributed compressed video sensing scheme in reconstruction algorithm

Shuai Zheng[1] · Jian Chen[1] · Yonghong Kuo[1]

Abstract Under the new video application scene of resource-constrained coding side such as wireless sensor networks, compressed sensing technique provides the possibility to solve the high-complexity problem of encoder because of its highly efficient compression encoding performance. Distributed compressed video sensing system provides a solution to satisfy the requirements of low encoder complexity and high coding efficiency in the new scene. This paper proposes a new distributed compressed video sensing scheme, which effectively improves the reconstruction quality of non-key frames. An auxiliary iterative termination decision algorithm is proposed to improve the performance of key frames initial reconstruction. An adaptive weights prediction algorithm is put forward to reduce the overall complexity. Besides, this paper proposes a position-based cross reconstruction algorithm to improve the decoded quality of the middle non-key frames in the group of pictures. The simulation results show that the proposed scheme effectively improves the overall performance of the distributed compressed video sensing system especially for high motion sequences.

Keywords Distributed compressed video sensing · Iterative termination decision · Weights prediction · Position-based cross reconstruction

1 Introduction

With the development of wireless sensor networks and low-power video surveillance networks, a new video application scene is gradually evolved. In such a scene, the encoder is

✉ Jian Chen
   jianchen@mail.xidian.edu.cn

   Shuai Zheng
   zhs_xd@163.com

   Yonghong Kuo
   yhkuo@mail.xidian.edu.cn

[1] School of Telecommunications Engineering, Xidian University, Xi'an 710071, China

resource-constrained due to the restricted energy, storage space and computing power. In traditional video codec systems, the motion estimation (ME), motion compensation (MC) and other complex calculations are carried out at encoder [10], which leads to the highly computational complexity of encoder. Therefore, the traditional video codec systems can't give full play to their performance in the new scene. The emergence of distributed video codec (DVC) system successfully reduces the complexity of encoder by moving the ME and MC to the decoder [11]. Compressed sensing (CS) technique is able to sample (or measure) signal at a sampling rate that is much less than the Nyquist sampling rate [1, 2, 7], which greatly simplifies the coding complexity and improves the coding efficiency.

To further reduce the computational complexity, Lu Gan proposes the block-based compressed sensing (BCS) in 2007 [17], which divides the whole frame into non-overlapping small blocks and then samples the frame block by block. BCS greatly reduces the storage pressure and the computational complexity in encoder side. In 2009, Kang LW and Lu CS propose the distributed compressed video sensing (DCVS) codec system [11]. The DCVS system effectively combines the excellent properties of DVC and CS. Later, Mun S and Fowler JE propose a BCS-based smooth projection Landweber reconstruction algorithm (BCS-SPL) [20, 21]. And then, the BCS-SPL based on multihypothesis (MH) reconstruction technique (MH-BCS-SPL) is proposed by Fowler JE et al. [4, 9, 25]. MH-BCS-SPL greatly improves the reconstruction quality of decoder and is widely cited in subsequent studies. Kuo et al. propose a key frame secondary reconstruction algorithm based on MH-BCS-SPL [12], which obviously improves the decoded quality of key frames. In 2015, Chen J et al. propose an Elastic net-based multihypothesis (MH) or single hypothesis (SH) prediction scheme (MS-wEnet) [5]. MS-wEnet effectively improves the accuracy of the side information (SI) and the reconstruction quality of decoder. On this basis, Kuo Y et al. propose a hybrid hypothesis prediction (HHP) system based on hypothesis set optimization technique and adaptive weighted Elastic net (AWEN) technology [13], which further improves the decoding quality. The scheme proposed in [13] is also called Up-Se-AWEN-HHP. However, the effects of different positions in group of pictures (GOP) on the reconstruction quality are not taken into account in the above schemes. Especially for some high motion sequences, the decoding quality of the non-key frames in the middle of GOP is much inferior to the non-key frames in the side of GOP. Moreover, for the reconstruction of non-key frames, the effects of different sampling rates on the MH weights calculation have not been considered in the above systems, which leads to the sharp increase of decoding complexity at high sampling rate in some schemes.

For the above problems, this paper proposes an improved DCVS system based on the key frames secondary reconstruction. The innovation in this paper includes three parts: (1) an auxiliary iterative termination decision algorithm is proposed to improve the reconstructed performance of key frames; (2) for the MH prediction of non-key frames, an adaptive MH weights prediction algorithm based on the number of measurements is proposed to equalize the decoded complexity of the DCVS scheme at different sampling rates; (3) a position-based cross reconstruction algorithm is proposed to improve the decoded quality of the non-key frames in the middle of GOP.

The rest of this paper is organized as follows: Section 2 describes the conventional DCVS systems based on key frames secondary reconstruction and some key techniques in DCVS. The proposed DCVS scheme is illustrated in Section 3. In Section 4, the experiment results are given to verify the performance of the proposed system. Finally, a conclusion for this paper is made in Section 5.

2 Background

2.1 The conventional DCVS systems

In this subsection, we give an overview of the conventional DCVS systems. Figure 1 shows the typical block diagram of the conventional DCVS systems based on key frames secondary reconstruction.

At the encoder side, the video frames are firstly divided into GOP with fixed size. Then the video frames are divided into two different types, the first frame in each GOP is the key (K) frames, others are the non-key (CS) frames. After that, the K frames and CS frames are sampled by BCS at different sampling rates. The sampling process can be described by the following Eq. (1),

$$y = \Phi u = A\theta \tag{1}$$

where $u = \Psi\theta$ is the block to be sampled, Ψ is the sparse basis. y are the measurements obtained in BCS, Φ is a measurement matrix, A denotes $\Phi\Psi$. After encoding, the measurements obtained in BCS are transmitted to the decoder.

At the decoder side, the K frames are firstly decoded, in the initial reconstruction of K frames, some fast reconstruction algorithms are applied, such as BCS-SPL [12], Bregman-SPL etc. Then the decoded K frames, as the reference frames, are utilized to provide hypotheses for the reconstruction of the adjacent CS frames. In the second reconstruction of K frames, the adjacent decoded CS frames are applied to provide the hypotheses, some accurate reconstruction algorithms such as MH-BCS-SPL [12] and MS-BCS-SPL [5] are used in this process. After the processes of hypothesis set acquisition and MH weights prediction, the BCS-SPL residual reconstruction of CS frames is performed. To improve the decoded performance of the CS frames, some improvements are proposed, such as the new MH weights prediction algorithm [5] and the hypothesis set optimization algorithm [13] etc.

2.2 The key techniques in CS reconstruction

The common method to reconstruct the sparse signal u from the measurements y is based on the l_1 minimization problem [29]. There are many algorithms about CS recovery have been proposed in the existing research, such as BP [24], OMP [26, 27, 30], StOMP [8] et al. However,
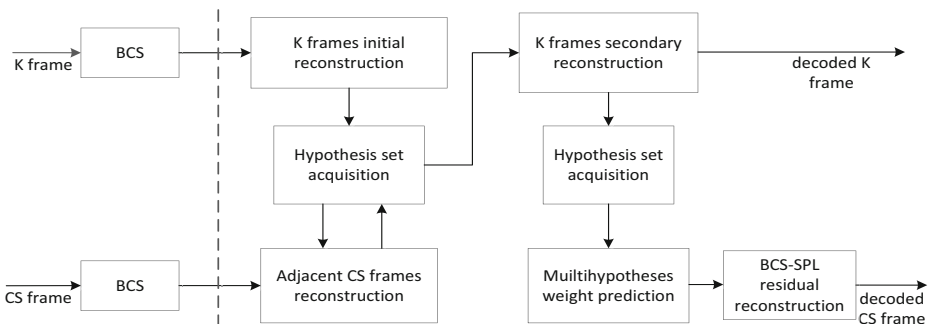
Fig. 1 The conventional DCVS system based on key frames secondary reconstruction

BCS-SPL based on residual reconstruction has been popular in DCVS schemes since it is proposed. Especially the algorithm, MH-BCS-SPL, effectively improves the decoding performance of DCVS. There are many improvements are proposed on the basis of MH-BCS-SPL [5, 12, 13]. In this subsection, the key techniques in MH-BCS-SPL are described in detail, such as the BCS-SPL based on residual reconstruction and the MH weights prediction. Assume that u is the original image or frame to be reconstructed, $\tilde{u}$ denotes the prediction of u, r is the residual between u and $\tilde{u}$. Then u can be expressed as (2).

$$u = \tilde{u} + r \tag{2}$$

The measurements of r under the projection of measurement matrix Φ can be expressed as below,

$$q = \Phi r = b - \Phi\tilde{u} \tag{3}$$

where b denotes the real measurements of u received at decoder. Converting formula (3) to formula (4),

$$u = \tilde{u} + Reconstruct(q, \Phi) \tag{4}$$

where $Reconstruct(\cdot)$ expresses the process of residual reconstruction. The residual is better than the original image in sparsity. Moreover, the MH prediction technique makes the prediction more accurate. Correspondingly, the residual is sparser and the reconstruction performance is better.

The key technique in MH prediction is the calculation of the weight for each hypothesis, i.e. how to acquire more accurate SI. The calculation of the weights can be described by formula (5),

$$w_{i,m}^{mh} = \underset{w}{\arg\min} \left\| y_{i,m} - Q_{i,m} w \right\|_2^2 \tag{5}$$

$w_{i,m}^{mh}$ denotes the weight of each hypothesis, $y_{i,m}$ are the measurements of the original image received at decoder. If $Q_{i,m} = \Phi_B H_{i,m}$ is the projection of hypothesis set $H_{i,m}$ under the measurement matrix Φ_B. The final prediction $\tilde{x}_{i,m}^{mh} = H_{i,m} w_{i,m}^{mh}$. There are some researches about the solving of problem (5) [28]. Literature [19, 25] proposed and validated the excellent performance of the MH prediction based on Tikhonov regularization, the algorithm can be described by the following formula (6),

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

where λ is a non-negative real value parameter. Γ is a Tikhonov regularization matrix, it's form can be described as follows,

$$\Gamma = \begin{pmatrix} \left\| y_{i,m} - q_1 \right\|_2 & & & 0 \\ & \ddots & & \\ & & \left\| y_{i,m} - q_s \right\|_2 & \\ & & & \ddots \\ 0 & & & \left\| y_{i,m} - q_\Sigma \right\|_2 \end{pmatrix} \tag{7}$$

where $\left\| y_{i,m} - q_s \right\|_2$ represents the Euclidean distance between the sth column q_s of matrix $Q_{i,m}$ and the target measurements $y_{i,m}$, the values of the weights decrease

with the increasing of the Euclidean distance. Converting formula (7) to its closed form as below.

$$w_{i,m}^{Tik} = \left(Q_{i,m}^T Q_{i,m} + \lambda^2 \Gamma^T \Gamma \right)^{-1} Q_{i,m}^T y_{i,m} \tag{8}$$

3 Proposed scheme

The block diagram of our proposed DCVS system is presented in Fig. 2. The modules with blue shadow are modified or proposed in this paper.

At the encoder side, our scheme is the same as the system shown in Fig. 1. At the decoder side, for key frames, the total variation (TV) [6, 23] algorithm is used to recover the key frames firstly, then the initial-decoded key frames are used as the reference frames to perform the second reconstruction of key frames by MH-BCS-SPL based on Tikhonov regularization. For CS frames, the decoder acquires hypothesis set from the secondary-recovered key frames firstly, then the MH prediction algorithm based on Elastic net or Tikhonov regularization is selected and performed in the adaptive weights prediction algorithm according to the number of measurements. Then the proposed position-based cross reconstruction algorithm is applied to reconstruct the final CS frames. The hypothesis set optimization technique [13], MH or SH hybrid hypothesis prediction (MS-HHP) [5] and BCS-SPL residual reconstruction are utilized in the reconstructed process.

3.1 A new iterative termination decision algorithm

TV has been popular ever since its introduction by Rudin L et al. [23]. TV minimization model is widely used in image denoising, deconvolution and reconstruction [15, 18]. Then some scholars introduce TV regularization into video decoding algorithm [22]. The TV minimization model can be described by formula (9),

$$\min_u \sum_i \|D_i u\|, \quad \text{s.t. } Au = b \tag{9}$$
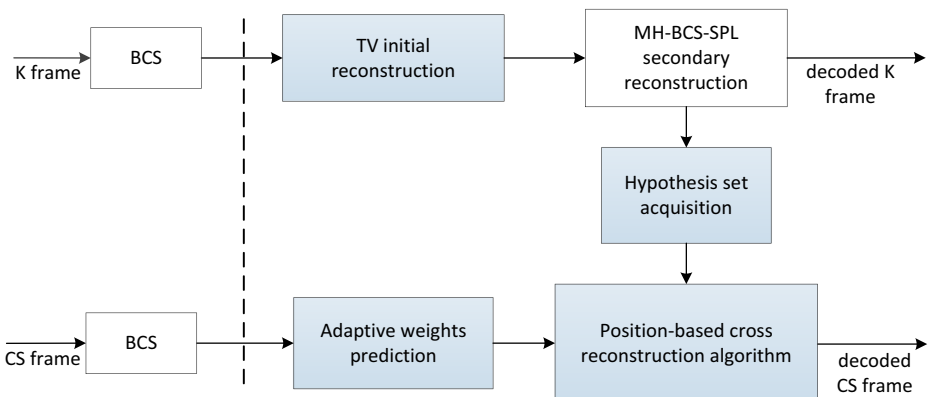

Fig. 2 Block diagram of the proposed DCVS codec scheme

where u is the image to be recovered, $D_i u$ denotes the discrete gradient of u at pixel i, $D = [D_V, D_h]$, D_V and D_h denote the vertical and horizontal finite difference operators [33], respectively. The detailed mathematical expansion of $\|D_i u\|$ is given in [3], furthermore, more knowledge about discrete gradient can be found in [16]. A is the measurement matrix, b are the measurements received at decoder. Usually, the common TV reconstruction model [14] can be described as below,

$$\min_u \sum_{i=1}^{n^2} \|D_i u\|_2 + \frac{\mu}{2} \|Au - b\|_2^2 \tag{10}$$

$\|D_i u\|_2$ is the Euclidean norm of $D_i u$ in pixel domain. The most remarkable property of TV is that it can effectively preserve the edges of image due to the linear penalty on differences between adjacent pixels [31]. However, TV reconstruction may lead to the serious distortion for the images with complex texture. Later, a multiple regularization constraints TV model is proposed in [6] by combining nonlocal regularization [32, 33] and TV regularization as the following formula (11),

$$\min_u c_1 \|Du\|_2 + c_2 \|\Psi u\|_1 + c_3 \|u - Wu\|_2^2 \quad s.t. \ Au = b \tag{11}$$

Ψ denotes wavelet transform basis, c_1, c_2, c_3 denote the corresponding weights of different regularization, W is the weight matrix of the nonlocal mean. This model solved by the alternating iteration method [31] is used in this paper to reconstruct the key frames for the first time.

However, there is huge difference in texture, edge and geometric features for different video sequences, which makes it difficult for the decoding algorithm to judge the optimal iteration termination point. As a result, the alternating iteration method may lead to the over iteration for some images with complex texture. In such a situation, the final reconstructed image is not optimal. In the original scheme, the iteration termination is decided by the following formula (12),

$$\Delta U = \frac{norm(U_i - U_{i-1})}{norm(U_{i-1})} \tag{12}$$

where the U_i and U_{i-1} denote the ith and $(i-1)^{th}$ iteration results, respectively. $norm(\cdot)$ denotes the Frobenius norm. When $\Delta U < T$ (T is the threshold), iteration over. In practice, for some texture-complex images, the iteration result has achieved the optimum solution even ΔU may do not reach the threshold, then the iteration algorithm will go on, which leads to the final iteration result is no longer optimal. To solve this problem, the best solution is comparing the similarity between the iteration result and the real original image. However, the original image is unknown at decoder except its measurements. To obtain the similarity information, we propose an auxiliary iterative termination decision method based on measurement domain as shown in formula (13),

$$\Delta F = norm(AU_i - b) - norm(AU_{i-1} - b) \tag{13}$$

where A is the measurement matrix, $norm(AU_i - b)$ denotes the similarity (or the distance) between the ith iteration result U_i and the original frame in measurement domain. In the later stage of the iteration algorithm, if $\Delta F < 0$, the similarity is reduced, the iteration stops. Conversely, the iteration goes on. The detailed decision flow diagram is shown in Fig. 3.

Fig. 3 Block diagram of the proposed TV iteration termination decision scheme
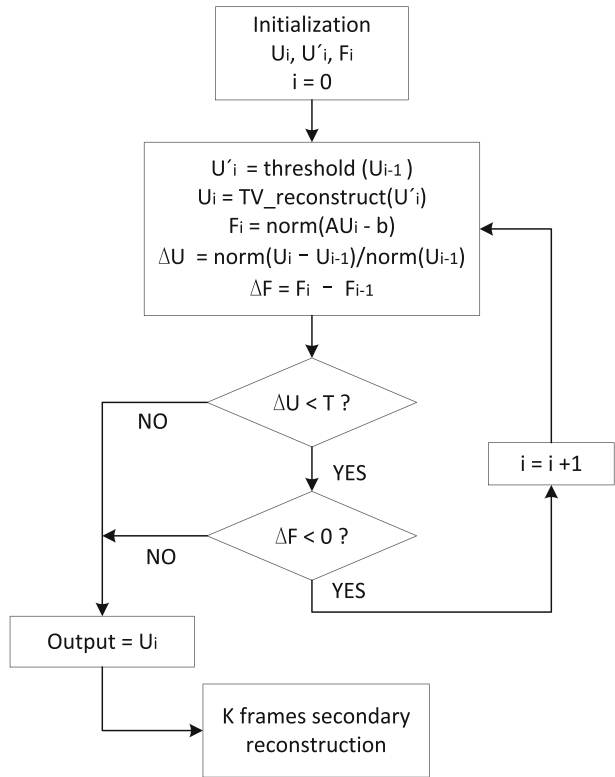
As shown in Fig. 3, to further improve the quality of the TV reconstruction, the thresholding process applied in BCS-SPL reconstruction is introduced in each iteration. The sparsity of the signal to be reconstructed is enhanced in each iteration, which improves the decoding performance. Moreover, in traditional DCVS schemes based on K frames secondary reconstruction, the second reconstruction of K frames has to be performed after the decoding of the adjacent CS frames [12, 13]. It increases the whole decoding delay of DCVS systems. In our proposal, the decoding of K frames is only related to the K frames without the participation of CS frames. Furthermore, the quality of reference frames is better than the traditional DCVS systems in the second reconstruction of key frames.

3.2 Adaptive MH weights prediction

In 2015, a hybrid hypothesis prediction scheme that combines MH and SH prediction algorithms is proposed in [5]. In this scheme, an Elastic net-based weights prediction algorithm (wEnet), which combines the ℓ_1 penalty term and ℓ_2 penalty term, is proposed to estimate the MH weights as shown in formula (14),

$$w_{i,m}^{\text{Enet}} = \left(1 + \frac{\lambda_2}{n}\right) \underset{w}{\arg\min} \left\|y_{i,m} - Q_{i,m}w\right\|_2^2 + \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2 \qquad (14)$$

where λ_1 and λ_2 are non-negative real parameters. Benefiting from the perfect performance of the wEnet-based MH prediction algorithm, the prediction accuracy of SI is greatly improved

especially at low sampling rate. On this basis, a hypothesis set update algorithm is proposed in [13], which effectively improves the quality of the hypothesis set. However, with the increase of sampling rate, the complexity of the Elastic net-based hybrid hypothesis prediction algorithm increases greatly, and a slowdown in quality growth occurs compare with the scheme based on Tikhonov-based MH prediction algorithm. To compare the performance of the Tikhonov-based MS prediction algorithm (MS-Tik) and Elastic net-based MS prediction algorithm (MS-wEnet) under the same conditions, the decoding performance of Coastguard and Mother-daughter, measured in [13], is given as example in this paper. The comparison data, the peak signal-to-noise ratio (PSNR) and the average run-time tested in [13], is shown in Tables 1 and 2.

It is observed that the performance of MS-wEnet is greatly related to sampling rate. At low sampling rate, it is better than MS-Tik in PSNR and run-time. However, when sampling rate is greater than 0.3, the complexity of MS-wEnet increases dramatically with the increase of sampling rate, the decoding quality is also inferior to MS-Tik. The reason is that the accuracy of SI increases with the increase of sampling rate, while the wEnet becomes unstable when SI is close enough to the real image [5]. Therefore, the performance of MS-wEnet is inferior to MS-Tik at high sampling rate. To improve the universal applicability of DCVS systems for different sampling rates, an adaptive MH weights prediction method based on the number of measurements in each block is proposed in this paper as shown in Fig. 4.

Where Y_cs are the measurements received at decoder, Num_y denotes the number of the measurements of each block and T_m is the threshold which decided by the following formula (15),

$$T_m = \alpha \times (block\_size^{\wedge}2) \tag{15}$$

where α is a positive parameter related to the sampling rate. If $Num\_y > T_m$, the Tikhonov-based MH prediction algorithm is adopted to estimate the weight of each hypothesis; otherwise the wEnet-based weights prediction algorithm is adopted.

3.3 The position-based cross reconstruction algorithm of CS frames

In conventional DCVS schemes, to reduce the complexity of encoder, the video sequences are usually divided into GOP with fixed size. However, for high motion sequences, the decoding quality of CS frames in the middle of GOP is much inferior to the CS frames in the side of GOP because of the following reasons: 1) the CS frames in the middle of GOP are far away from the decoded K frames, some high-quality hypotheses in K frames are in the outside of the searching window due to its complex motion; 2) the similarity between CS frames in the middle of GOP and the reference frames is reduced, the quality of hypothesis set becomes

Table 1 The comparison results of average PSNR (dB)

Sequences	Method	Sampling rate				
		0.1	0.2	0.3	0.4	0.5
Coastguard	MS-wEnet	30.95	33.05	34.81	36.47	38.15
	MS-Tik	30.40	32.74	34.66	36.54	38.49
Mother-daughter	MS-wEnet	42.31	43.74	44.94	46.01	47.04
	MS-Tik	41.44	43.33	44.75	46.03	47.21

Table 2 The comparison results of average reconstruction time (s)

Sequences	Method	Sampling rate				
		0.1	0.2	0.3	0.4	0.5
Coastguard	MS-wEnet	98.9	168.4	326.7	591.4	1038.2
	MS-Tik	238.3	245.8	254.5	258.9	260.6
Mother-daughter	MS-wEnet	98.0	168.6	324.2	607.9	1053.8
	MS-Tik	240.7	246.5	259.9	262.4	262.5

worse. To display the difference of decoding quality caused by the different positions in GOP, the decoded average PSNR of Soccer sequence in the scheme [13] is tested as an example. In Table 3, the 1st means the first CS frame in GOP, the size of GOP is 8.

As shown in Table 3, the average decoding quality of CS frames at different positions in GOP presents the symmetrical distribution. The quality of CS frames in the middle of GOP is inferior to the CS frames at both sides of the GOP, especially for the 3rd, 4th and 5th CS frames, their PSNR is decreased by 2.5 dB (on average) relative to the 1st and 7th CS frames. The aforementioned analyses indicate that the position of CS frame has great impact on the decoding quality. To solve this problem, a position-based cross reconstruction (PBCR) algorithm is proposed in this paper. The block diagram of the proposed algorithm is shown in Fig. 5.

As shown in Fig. 5, the PBCR mainly consists of two parts: the hypothesis set acquisition algorithm based on multiple reference frames (HSA-MRF) and the searching window expansion algorithm (SWEA) based on fixed hypothesis set size. As mentioned above, it's the reduced similarity and complex motion that leads to the poor decoding quality of the middle CS frames in GOP. To improve the similarity between the middle CS frames and the SI predicted by MH, the HSA-MRF is proposed to make full use of the decoded K frames and CS frames to obtain better hypotheses. Here, as an example, the size of GOP is set to 8 and Fig. 6 describes the HSA-MRF in detail.

As shown in Fig. 6, K1 and K2 are the decoded adjacent K frames, CSi denotes the ith CS frame in each GOP. Rec$i$ denotes the decoding order of the CS frames. In HSA-MRF, K1 and K2 are the necessary reference frames for each CS frame. However, with the increase of the distance between the CS frame and the decoded K frames, their similarity is reduced and the quality of hypotheses obtained from the decoded K frames becomes worse. In this case, comparing with the K frames, the former decoded CS frames are more similar to the next CS frame to be reconstructed. Therefore, the decoded CS frames are also adopted as the reference frames for the next CS frames to obtain more and better hypotheses. As described in Fig. 6, in order to effectively use the decoded CS frames, the decoding order of CS frames is designed to be crossed, that is, starting from the both sides of GOP and then gradually shifting to the inside of GOP (CS1, CS7, CS2, CS6…CS4). For the CS frames in the middle of GOP (CS2-CS6), the former decoded two CS frames are applied to provide more hypotheses. Then, the
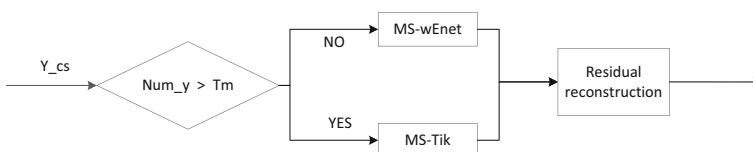
Fig. 4 The block diagram of the adaptive MH weights prediction method

Table 3 The average reconstruction PSNR (dB) of different positions in GOP (only CS frames)

| Position		1st	2nd	3rd	4th	5th	6th	7th
Sampling rate	0.1	29.5732	27.5424	26.7354	26.9651	26.8824	27.5320	29.5722
	0.2	32.2422	30.3522	29.4390	29.7749	29.5613	30.2642	32.2271
	0.3	34.4475	32.7058	31.8690	32.2226	31.9298	32.5561	34.5445
	0.4	36.7228	35.2213	34.2766	34.7522	34.4183	34.9138	36.8810
	0.5	39.1746	37.8369	37.0633	37.3729	37.0679	37.6844	39.2635

hypothesis set optimization technique [13] is utilized to eliminate the poor hypothesis according to the Euclidean distance in measurement domain. After optimization, the size of hypothesis set is the same as the original hypothesis set without HSA-MRF, but the hypotheses have been updated with better hypotheses obtained from the decoded CS frames. The quality of hypothesis set is improved effectively, the reconstructed quality has higher stability due to the existence of the hypothesis set optimization technique.

However, at low sampling rate, the quality of the decoded CS frames is bad relative to the K frames. Correspondingly, the hypotheses, obtained from the former reconstructed CS frames, are not good enough to effectively improve the overall quality of the final hypothesis set, which limits the performance of the HSA-MRF algorithm. Furthermore, as mentioned above, for high motion video sequences, some high-quality hypotheses in K frames are likely to be in the outside of the searching window when the distance between the CS frame and the decoded K frame is relatively far. Therefore, how to make full use of the decoded K frames becomes a problem to be solved. To solve the aforementioned problems, when sampling rate is less than 0.3, the SWEA is applied in PBCR for the middle CS frames in GOP i.e. the CS frames with green shadow shown in Fig. 6.

Figure 7 shows the process of the searching window expansion. W is the size of the original searching window. The block size is B. ΔW is the increased size. If the CS frame is in the center of GOP and the sampling rate is less than 0.3, the size of searching window in the reference K frames is expanded to $W + \Delta W$. After the process of SWEA, the hypothesis set optimization technique is applied to select the high-quality hypotheses from the hypothesis set obtained in SWEA. The size of the finial hypothesis set is the same as the original set without SWEA.

4 Experimental results

To validate the performance of the position-based cross reconstruction DCVS (PBCR-DCVS) scheme, a series of experimental results are given in this section. In our experiments, the first 88 frames of the standard test video with CIF format are tested. MS-wEnet [5] and Up-Se-AWEN-HHP [13] are introduced as comparison. To be fair, the experimental parameters are consistent with [13]. At the encoder, we set the block size as $B = 16$ and the size of GOP is set to 8. The initial searching window size $W = 15$ pixels, in SWEA, $\Delta W = 8$, and the SWEA is
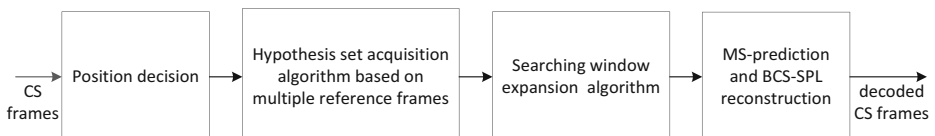
Fig. 5 The position-based cross reconstruction algorithm for CS frames
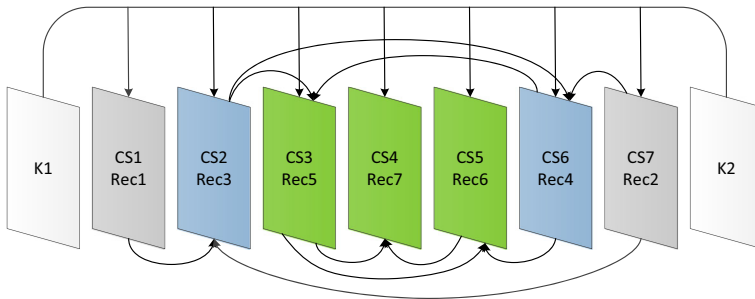
**Fig. 6** The acquisition process of multiple reference frames in HSA-MRF

only performed at the sampling rate less than 0.3. The sampling rate of key frames is 0.7, while the sampling rates of non-key frames vary from 0.1 to 0.5. The iterative termination threshold $T = 0.001$ in the original TV reconstruction scheme. $\alpha = 0.3$ in Eq. (15). The orthonormal matrix is selected as Φ. The reconstructed quality of the video is measured by PSNR, the computational complexity is examined by the reconstructed speed, which is measured by the CPU run-time. The configuration of the simulation is 64-bit Windows 7 SP1, Inter (R) Core (TM) i7–4790 CPU, 3.60GHz, 16G RAM. The version of Matlab is R2015a.

4.1 The performance of the new iterative termination decision scheme

In this subsection, we compare our new iterative termination decision scheme with the original iterative termination decision method. The average PSNR of the K frames TV initial reconstruction is tested. Furthermore, to directly reflect the performance of the new iterative termination decision scheme, the average relative distance ($D = \text{norm}(AU_i - b)/\text{norm}(b)$) between the iteration result and the original image in measurement domain is also adopted to measure the quality of the TV initial reconstruction. The smaller relative distance indicates that the iteration result is closer to the original image, and the reconstruction quality is better.

As shown in Table 4, comparing with the original iterative termination decision scheme, for Coastguard, the PSNR is increased by 0.44 dB and the relative distance is reduced by 0.12% in our proposed scheme, mainly because the over iteration phenomenon is avoided by the proposed auxiliary iterative termination decision algorithm in measurement domain. For Foreman, Mother-daughter and Soccer sequences, the PSNR and relative distance are only marginally improved. This can be explained by the following reasons: 1) comparing with the complex water ripple information in Coastguard, the above three sequences are relatively simple in texture details, the original TV iterative termination decision scheme can accurately determine the optimal results, especially for Foreman and Mother-daughter, there is no over

Fig. 7 The process of searching window expansion
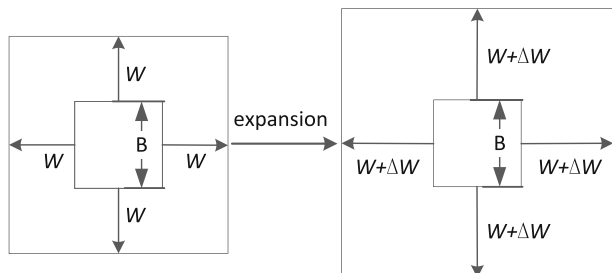
Table 4 The comparison results of the TV initial reconstruction for K frames (on average)

Sequences	Method	Relative distance (%)	PSNR (dB)
Coastguard	original TV	2.4472	38.2957
	proposed	2.3257	38.7326
Foreman	original TV	1.0592	43.1022
	proposed	1.0579	43.1090
Mother-daughter	original TV	0.9222	47.1246
	proposed	0.9190	47.1513
Soccer	original TV	1.4480	42.9033
	proposed	1.4431	42.9341

iteration problem, and there are only three key frames subject to the problem of over iteration for Soccer; 2) in our new iterative termination decision scheme, the thresholding process is introduced in each iteration, the sparsity is enhanced in each iteration, which improves the decoding quality.

4.2 The performance of the PBCR algorithm

In this subsection, we compare our scheme with the system without PBCR under the same K frames reconstruction scheme proposed in this paper. The average PSNR of the decoded CS frames in each position of GOP is measured. To reflect the performance of the PBCR algorithm at different sampling rates, we give the comparison results of the CS frames reconstruction quality at sampling rates equal to 0.2 and 0.4, respectively.

As shown in Figs. 8 and 9, comparing with the scheme without PBCR, the reconstruction quality of the CS frames in the middle of GOP is effectively improved for all tested video
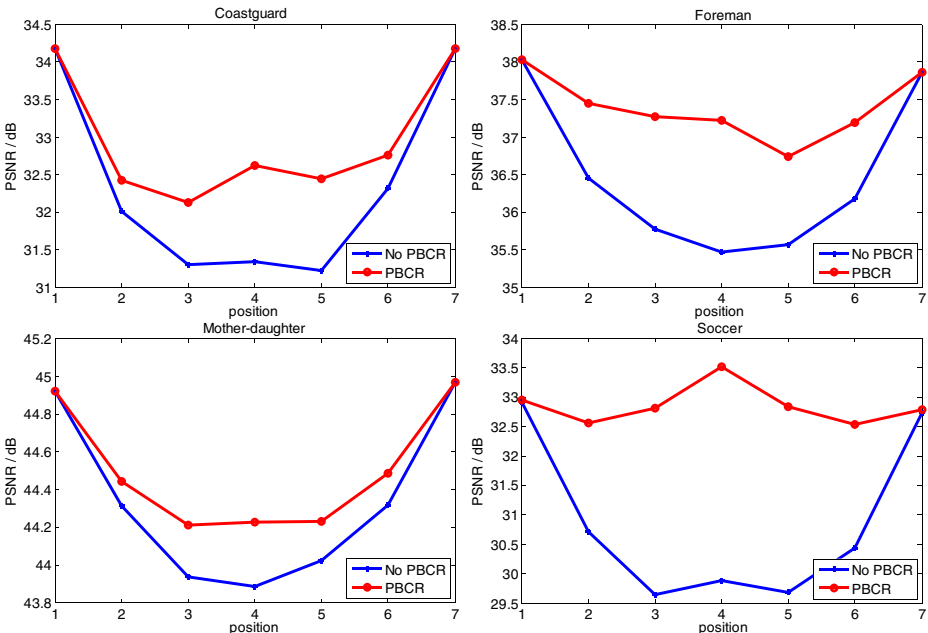
Fig. 8 The average PSNR of CS frames in each position of GOP (sampling rate equals to 0.2)
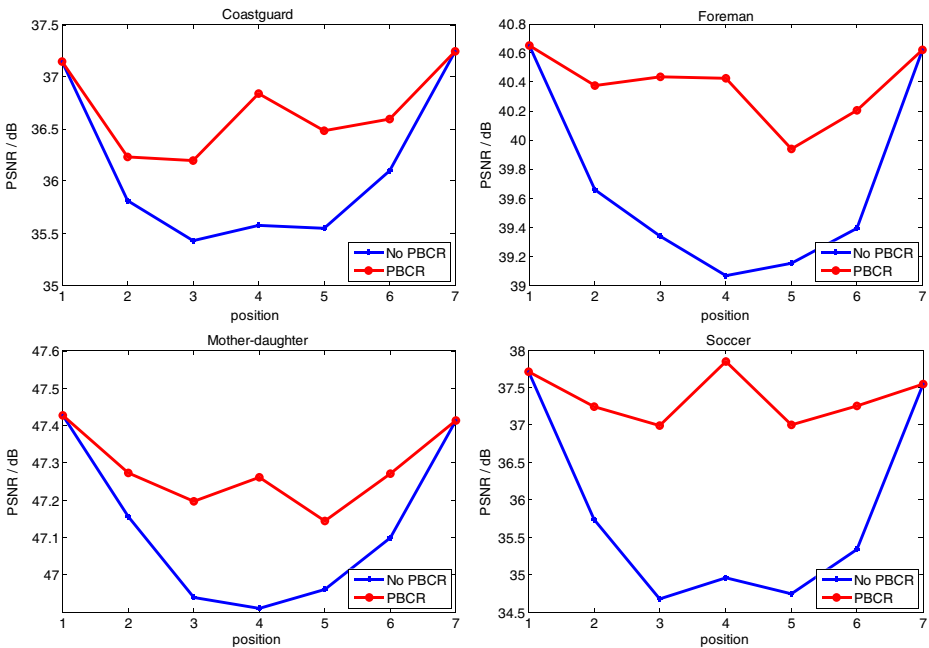
Fig. 9 The average PSNR of CS frames in each position of GOP (sampling rate equals to 0.4)

sequences in our proposed system. Especially for high motion sequences such as Foreman and Soccer, at sampling rate equals to 0.2, the average PSNR of the middle CS frames (CS2-CS6) is increased by 1.29 dB and 2.80 dB, respectively. For low motion sequences such as Mother-daughter, the average PSNR of the middle CS frames is increased by 0.23 dB. These can be explained by the following reasons: 1) the HSA-MRF algorithm in PBCR makes full use of the similarity among non-key frames, more high-quality hypotheses are obtained from the former reconstructed CS frames, the quality of the hypothesis set is improved effectively; 2) the SWEA in PBCR further deepens the using of key frames by expanding the size of searching window, which further improves the reconstruction quality of the medial CS frames (CS3-CS5) at low sampling rate; 3) for low motion sequences, the similarity between key frame and non-key frame is high, the PSNR of non-key frames at different positions in GOP is balanced relative to the high motion sequences. Therefore, the improvement bring by PBCR is not as obvious as the high motion sequences at low sampling rate. Moreover, since the decoded quality of CS frames increases with the increase of the sampling rate, the hypotheses obtained from the former reconstructed CS frames are also improved. Therefore, it is observed that the decoding quality of the CS frames in different positions is more balanced with the increase of the sampling rate.

4.3 The overall reconstruction performance of the proposed DCVS

The overall performance comparison of the proposed PBCR-DCVS scheme with other DVCS schemes is made in this subsection. MS-wEnet [5] and Up-Se-AWEN-HHP [13] as the state-of-the-art DCVS schemes are exploited as comparison systems. The comparison results of the average reconstruction PSNR in different DCVS schemes are given in Fig. 10. Moreover, to compare the complexity of our proposed system with other schemes, the average CPU run-time of different systems is measured in experiment as shown in Table 5.
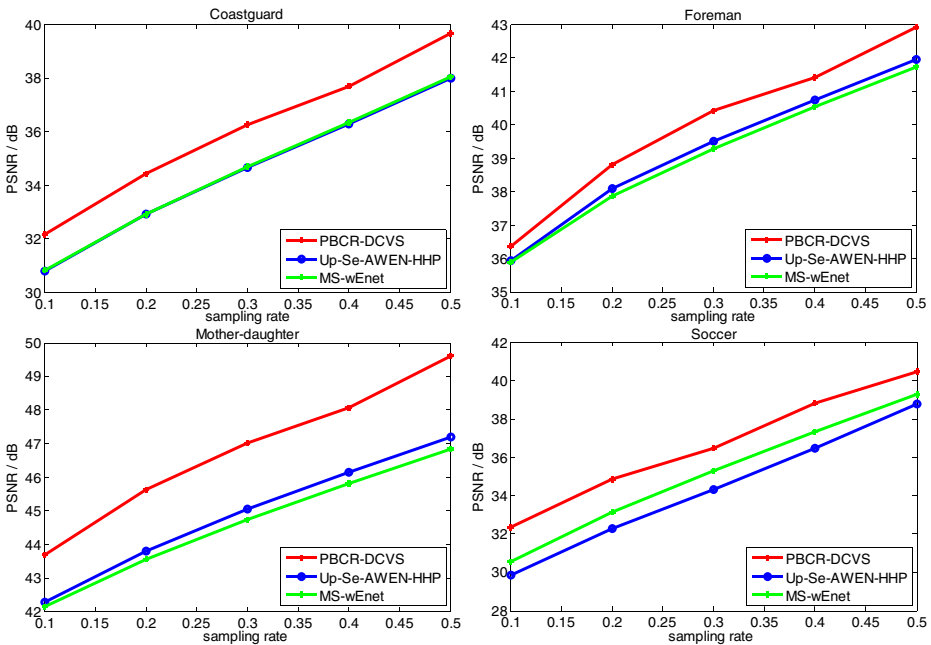
Fig. 10 The reconstruction performance of different DCVS schemes

As shown in Fig. 10, compare with MS-wEnet [5] and Up-Se-AWEN-HHP [13] schemes, the reconstruction performance is significantly improved in our PBCR-DCVS scheme. Benefitting from the improvement of the hypothesis set's quality bring by PBCR, the reconstruction quality of the middle non-key frames in GOP is improved greatly, which improves the overall performance of the proposed scheme. For Coastguard and Foreman, the average PSNR is increased by 1.5 dB and 0.75 dB, respectively. Moreover, in our scheme, key frames are first decoded by TV algorithm, then the BCS-SPL is applied to recover the key frames for the second time, the decoded quality of key frames is improved. Correspondingly, there is a general promotion for the reconstruction quality of non-key frames in each position

Table 5 Time (s) consumed by single frame for different schemes

Sequences	Method	Sampling rate				
		0.1	0.2	0.3	0.4	0.5
Coastguard	PBCR-DCVS	45.45	73.57	85.07	52.52	53.88
	MS-wEnet	38.19	58.75	92.47	150.01	255.82
	Up-Se-AWEN-HHP	36.14	65.26	102.46	166.87	330.00
Foreman	PBCR-DCVS	47.29	76.46	88.27	50.41	51.67
	MS-wEnet	40.79	60.63	94.76	152.18	256.21
	Up-Se-AWEN-HHP	37.11	65.46	120.39	192.43	299.66
Mother-daughter	PBCR-DCVS	45.47	73.17	85.46	54.45	55.49
	MS-wEnet	39.41	58.83	92.32	148.50	252.28
	Up-Se-AWEN-HHP	36.03	64.48	96.85	160.80	295.63
Soccer	PBCR-DCVS	47.82	76.04	87.55	48.16	49.22
	MS-wEnet	40.48	60.43	94.16	151.14	254.45
	Up-Se-AWEN-HHP	35.89	67.84	117.32	200.01	315.18

of GOP. Especially for Soccer and Mother-daughter, the average PSNR is increased by more than 1.9 dB relative to Up-Se-AWEN-HHP [13]. Compare with MS-wEnet [5], the average PSNR is increased by 1.47 dB for Soccer.

Table 5 shows the comparison results of the CPU run-time for different DCVS schemes. First of all, compare with MS-wEnet [5] and Up-Se-AWEN-HHP [13], the time consumed in key frames reconstruction is effectively reduced, due to the following reasons: 1) the non-key frames are not need to provide side information for the key frames secondary reconstruction, the proposed key frames secondary reconstruction scheme acquires the hypotheses from the initial-recovery key frames, which do not have to wait for the reconstruction of non-key frames; 2) the multiple regularization constraints TV algorithm has high convergence rate, the optimal solution can be obtained after a small amount of threshold decision processes. For the recovery of non-key frames, the proposed adaptive MH weights prediction algorithm adaptively selects the Tikhonov-based weights prediction algorithm rather than the wEnet when sampling rate is more than 0.3. It avoids the problem that the complexity of the wEnet-based prediction algorithm increases drastically with the increase of sampling rate. Moreover, the SWEA is only performed at the sampling rate less than 0.3. Therefore, it is observed that the time consumed in single frame is the least in our scheme when the sampling rate is more than 0.2. However, when sampling rate is less than 0.3, with the increase of the reference frames and the size of the searching window, the number of hypotheses increase dramatically. Correspondingly, the algorithms of HSA-MRF and SWEA in PBCR have to consume more time to select the better hypotheses from the large number of hypotheses. Therefore, the computational complexity of our scheme increases slightly compared with MS-wEnet [5] and Up-Se-AWEN-HHP [13] at low sampling rate.

5 Conclusions

In this paper, we propose a position-based cross reconstruction DCVS scheme. An auxiliary iterative termination decision algorithm is proposed to improve the performance of key frames TV initial reconstruction. Then, an adaptive MH weights prediction algorithm is proposed to reduce the complexity of the wEnet-based DCVS schemes at high sampling rate. To improve the decoded quality of the middle non-key frames in GOP, we propose a PBCR algorithm, which mainly consists of two components: the HSA-MRF and the SWEA. The HSA-MRF makes full use of the similarity among the non-key frames, while the SWEA deepens the exploiting of the key frames by expanding the searching window. Numerical results verify that our PBCR-DCVS scheme effectively improves the overall performance of the DCVS system compare with other schemes with the slightly increase of complexity at low sampling rate.

References

1. Candès EJ (2006) Compressive sampling. Marta Sanz Solé 25(2):1433–1452
2. Candes EJ, Romberg J, Tao T (2006) Roubust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory 52(2):489–509

3. Chang K, Qin T, Xu W et al (2014) Reconstruction of multi-view compressed imaging using weighted total variation. Multimedia Systems 20(4):363–378

4. 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, Pacific Grove, CA, p 1193–1198

5. Chen J, Chen Y, Qin D, Kuo YH (2015) An Elastic net-based hybrid hypothesis method for compressed video sensing. Multimedia Tools and Applications 74(6):2085–2108

6. Chen J, Gao Y, Ma C et al (2016) Compressive sensing image reconstruction based on multiple regulation constraints. In: Circuits Systems & Signal Processing, pp 1–18

7. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52(4):1289–1306

8. Donoho DL, Tsaig Y, Drori I, Starck JL (2012) Sparse solution of under determined systems of linear equations by stagewise orthogonal matching pursuit. IEEE Trans Inf Theory 58(2):1094–1121

9. Fowler JE, Mun S, Tramel EW (2012) Block-based compressed sensing of images and video. Foundations and Trends in Signal Processing 4(4):297–416

10. ITU-T Rec. H.264 and ISO/IEC 4496-10 (MPEG-4/AVC) (2003) Advanced video coding for generic audio visual services, Standard version 7, ITU-T and ISO/IEC JTC 1. https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.264-200305-S!!PDF-E&type=items

11. Kang LW, Lu CS (2009) Distributed compressive video sensing. In: Proceedings of the international conference on acoustics, speech and signal processing, Taipei, p 1169–1172

12. Kuo Y, Wang S, Qin D, Chen J (2013) High-quality decoding method based on resampling and re-reconstruction. Electron Lett 49(16):991–992

13. Kuo Y, Wu K, Chen J (2015) A scheme for distributed compressed video sensing based on hypothesis set optimization techniques. In: Multidimensional Systems and Signal Processing, pp 1–20

14. Li C, Yin W, Zhang Y (2009) TVAL3: TV minimization by augmented Lagrangian and alternating direction algorithm. http://www.caam.rice.edu/~optimization/L1/TVAL3/. Accessed May 2016

15. Li C, Yin W, Jiang H et al (2013) An efficient augmented Lagrangian method with applications to total variation minimization. Comput Optim Appl 56(3):507–530

16. Liu Z, Dong FF et al (2009) Discrete total variation model with gradient fidelity term for image restoration. In: Proceedings of the 2nd International congress on Image and Signal Processing, p 1–5

17. Lu G (2007) Block compressed sensing of natural images. In: Proceedings of the international conference on digital signal processing, Cardiff, p 403–406

18. Ma S, Yin W, Zhang Y et al (2008) An efficient algorithm for compressed MR imaging using total variation and wavelets. In: IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, p 1–8

19. Mallat S, Zhang Z (1993) Matching pursuit with time-frequency dictionaries. IEEE Trans Signal Process 41(12):3397–3415

20. Mun S, Fowler JE (2009) Block compressed sensing of images using directional transforms. In: Proceedings of the IEEE international conference on Image processing, Cairo, p 3021–3024

21. Mun S, Fowler JE (2011) Residual reconstruction for block-based compressed sensing of video. Data Compression Conference 4:183–192

22. Roohi S, Noorhosseini M, Zamani J et al (2013) Low complexity distributed video coding using compressed sensing. In: Proceedings of the 8th Iranian Conference on Machine Vision and Image Processing, p 53–57

23. Rudin L, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Phys D 60:259–268

24. Scott SC, Donoho DL, Saunders MA (2001) Atomic decomposition by basis pursuit. SIAM Rev 58(1):33–61

25. Tramel EW, Fowler JE (2011) Video compressed sensing with multihypothesis. In: Proceeding: Data Compression Conference, Snowbird, UT, p 193–202

26. Tropp JA (2006) Algorithms for simultaneous sparse approximation. Part ii: convex relaxation. Signal Process 86(3):589–602

27. Tropp JA, Gilbert A (2007) Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inf Theory 53(12):4655–4666

28. Wen J, Chang XW (2017) Success probability of the babai estimators for box-constrained integer linear models. IEEE Trans Inf Theory 63(1):631–648

29. Wen J, Li D, Zhu F (2015) Stable recovery of sparse signals via lp-minimization. Appl Comput Harmon Anal 38(1):161–176

30. Wen J, Zhou Z, Wang J et al (2017) A sharp condition for exact support recovery of sparse signals with orthogonal matching pursuit. IEEE Trans Signal Process 65(6):1370–1382

31. Xiao Y, Yang J, Yuan X (2012) Alternating algorithms for total variation image reconstruction from random projections. Inverse Problems and Imaging 6(3):547–563
32. Zhang X, Tony FC (2010) Wavelet inpainting by nonlocal total variation. Inverse problems and Imaging 4(1):191–210
33. Zhang J, Liu S, Xiong R, et al (2013) Improved total variation based image compressive sensing recovery by nonlocal regularization. In: Proceedings of the IEEE International Symposium on Circuits and Systems, Beijing, p 2836–2839

Shuai Zheng was born in Henan, P.R. China. He received his Bachelor's degree in Electronic Information Engineering from Technology University of Henan, P.R. China in 2015. Now he is pursuing his Ph.D. degree of Military Communication in Xidian University. His current research interests involve compressed sensing and distributed video coding.

Jian Chen received the B.S. degree from Xi'an Jiaotong University, China, in 1989, the M.S. degree from Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences in 1992, the Ph.D. degree in telecommunications engineering in Xidian University, China, in 2005. He is a professor in the School of Telecommunications Engineering, 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 sensor networks.

Yonghong Kuo received the B.S. and M.S. degrees from Xi'an Jiaotong University in 1989 and 1992 respectively, and the Ph.D. degree from Peking Union Medical University in 1998. She is a professor in the School of Telecommunications Engineering, Xidian University. Her research interests include signal processing, cognitive radio and wireless sensor networks.