Acta Automatica Sinica
Vol. 40, No. 10 October, 2014

Residual Distributed Compressive Video Sensing Based on Double Side Information

CHEN Jian[1] SU Kai-Xiong[1] WANG Wei-Xing[1] LAN Cheng-Dong[1]

Abstract Compressed sensing (CS) is a novel technology to acquire and reconstruct sparse signals below the Nyquist rate. It has great potential in image and video acquisition and processing. To effectively improve the sparsity of signal being measured and reconstructing efficiency, an encoding and decoding model of residual distributed compressive video sensing based on double side information (RDCVS-DSI) is proposed in this paper. Exploiting the characteristics of image itself in the frequency domain and the correlation between successive frames, the model regards the video frame in low quality as the first side information in the process of coding, and generates the second side information for the non-key frames using motion estimation and compensation technology at its decoding end. Performance analysis and simulation experiments show that the RDCVS-DSI model can rebuild the video sequence with high fidelity in the consumption of quite low complexity. About $1 \sim 5$ dB gain in the average peak signal-to-noise ratio of the reconstructed frames is observed, and the speed is close to the least complex DCVS, when compared with prior works on compressive video sensing.

Key words Video coding, compressed sensing, distributed compressive video sensing, residual coding

Citation Chen Jian, Su Kai-Xiong, Wang Wei-Xing, Lan Cheng-Dong. Residual distributed compressive video sensing based on double side information. *Acta Automatica Sinica*, 2014, **40**(10): 2316−2323

The compressed sensing (CS) theory[1−2] put forward by Donoho and Baraniuk et al. during $2004 \sim 2006$ shows that the high dimensional signal can be projected to a low dimensional space through an observation matrix incoherent with the transform basis, as long as the signal is sparse in a certain transform domain. Using a few observations, the signal can be reconstructed precisely. Recent years, the researches on reconstruction algorithm and measurement scheme based on CS have made significant progress[3−10]. The application of CS theory about video coding is still in an exploratory stage, but it has showed great development prospects[11].

In 2006, Wakin et al. obtained sampling data through a single pixel camera[12], and reconstructed the frames via the sparsity in the 2-D wavelet domain (referred to as 2D-CS) and a group of frames via the sparsity in the 3-D wavelet domain (referred to as 3D-CS)[13]. In order to reduce the computing burden of image or video compression, Lu put forward block compressed sensing of natural image (Block-CS) in 2007[14]. Then, some scholars applied the Block-CS into video coding[15−16]. It reduced the computational complexity significantly, however, its reconstructing performance was not ideal. To make full use of inter-frame correlation between moving pictures for further improving coding efficiency, some scholars made a CS coding model for the residual video (referred to as RVCS)[17−18]. During 2009, Do et al. proposed a kind of distributed compressed video sensing DISCOS architecture[19], Prades-Nebot et al. suggested the distributed video coding based on CS (DVC-CS)[20], while Kang et al. studied another version of distributed compressive video sensing (DCVS)[21]. After 2010, more and more scholars further researched on video CS based on frame or block, inter-frame residuals, as well as distributed video coding[22−26]. While those methods have improved the quality of video reconstruction to some extent, the complexity is also increasing.

To fully utilize the correlation of intra-frame and inter-frame, a coding algorithm called residual distributed compressive video sensing based on double side information (RDCVS-DSI) is proposed in this paper to rebuild video sequence in high fidelity under the conditions of lower complexity.

1 Compressive video sensing

1.1 Compressed sensing

According to the CS theory[1−3], signal $\boldsymbol{x}$ can be sparsely represented under some basis $\boldsymbol{\psi}^{N\times N}$

$$\boldsymbol{x} = \boldsymbol{\psi\theta}, \quad \boldsymbol{x} \in \mathbf{R}^N \tag{1}$$

where $\boldsymbol{\theta}$ is the transform coefficients of $\boldsymbol{x}$ in $\boldsymbol{\psi}$ domain. When $\boldsymbol{\theta}$ has only S ($S \ll N$) nonzero elements, signal $\boldsymbol{x}$ is S-sparse under the basis of $\boldsymbol{\psi}$. Partial Fourier transform, DCT and DWT are commonly used in sparse transform. In compressed sampling, signal $\boldsymbol{x}$ is projected into a set of measurement vectors of $\boldsymbol{\phi}$ to give the measured value $\boldsymbol{y}$, i.e.,

$$\boldsymbol{y} = \boldsymbol{\phi x}, \quad \boldsymbol{y} \in \mathbf{R}^M \tag{2}$$

where $\boldsymbol{y}$ is an $M \times 1$ measured values matrix, $\boldsymbol{\phi}$ is an $M \times N$ measurement matrix ($M \ll N$) incoherent with $\boldsymbol{\psi}$[27]. Gaussian, Bernoulli, scramble Fourier and scramble block Hadamard ensemble (SBHE) have been shown to be good choices for the measurement matrix $\boldsymbol{\phi}$.

Compressed sampling is a dimension reduction process, which helps reduce the number of collected data from N to M. However, it also makes the recovery of signal $\boldsymbol{x}$ from measurements $\boldsymbol{y}$ an ill-posed problem. The CS theory states that the reconstruction can be formulated as an l_p minimization problem by solving:

$$\min_{\boldsymbol{\theta}} ||\boldsymbol{\theta}||_{l_p} \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{\phi\psi\theta} \tag{3}$$

To solve the above optimization problem, many techniques have been proposed in the literature, e.g., orthogonal matching pursuit (OMP)[28], two-step iterative shrinkage/thresholding (TwIST)[29], gradient projection for sparse reconstruction (GPSR)[30], and sparse reconstru-

Manuscript received September 24, 2013; accepted April 21, 2014
Supported by National Natural Science Foundation of China (61170147), Major Cooperation Project of Production and College in Fujian Province (2012H61010016), and Natural Science Foundation of Fujian Province (2013J01234)
Recommended by Associate Editor HUANG Qing-Ming
1. College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China

ction by separable approximation (SpaRSA)[31]. If signal $\boldsymbol{x}$ is 2-D data, the above theory can be directly applied to image compression combining with existing CS acquisition device.

1.2 Compressive video sensing

Relative to the CS imaging, compressive video sensing has more stringent requirements on storage resources and real-time processing. Simultaneous temporal and spatial measurement by the 3D-CS is impractical, and thus one opts for frame-by-frame measurement. The most straight compressive video sensing scheme 2D-CS, adopting frame-by-frame CS for image, takes measurement and solution according to formulas (2) and (3).

For alleviating the huge computation and memory burdens, the Block-CS has been introduced into video coding. In the Block-CS, each frame is divided into N_B non-overlapping blocks $\boldsymbol{x}_j$ (sized $B \times B$, subscript j denotes block indicator), and acquired using a suitable $M_B \times B^2$ measurement matrix $\boldsymbol{\phi}_B$, then the corresponding $\boldsymbol{y}_j$ is

$$\boldsymbol{y}_j = \boldsymbol{\phi}_B \boldsymbol{x}_j, \quad \boldsymbol{x}_j \in \mathbf{R}^{B\times B}, \quad \boldsymbol{y}_j \in \mathbf{R}^{M_B} \tag{4}$$

It is straightforward to see that (4) applying block-by-block to an image is equivalent to a whole-image measurement matrix $\boldsymbol{\phi}$ in (2) with a constrained structure that $\boldsymbol{\phi}$ is block diagonal[14, 22],

$$\boldsymbol{\phi} = \begin{bmatrix} \boldsymbol{\phi}_B & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\phi}_B & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\phi}_B \end{bmatrix} \tag{5}$$

When sparsity transform $\boldsymbol{\psi}$ is also a block-based operator, the frame can be reconstructed by block at the decoding end. In general, block-independent reconstruction will produce severe blocking artifacts, thus rebuilding by frame is prior to block. In convenience, we focus on the CS for video by frame. Since a structural measurement matrix[32] in the form of (5) is used in this paper, the following scheme is equally applicable to the compressive video sensing by block.

According to the temporal redundancy of video, the correlation model between successive video frames $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$ can be expressed as:

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{x}_c + \boldsymbol{x}_{t\_u} \\ \boldsymbol{x}_{t+1} = \boldsymbol{x}_c + \boldsymbol{x}_{t+1\_u} \end{cases} \tag{6}$$

where $\boldsymbol{x}_c$ is the common portion between $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$, while $\boldsymbol{x}_{t\_u}$ and $\boldsymbol{x}_{t+1\_u}$ are the specific portions. The RVCS and DCVS are the two typical schemes for compressive video sensing based on the above correlation model.

The basic idea of RVCS[17] comes from the traditional inter-frame coding. By using the same measurement matrix in a group of pictures (GOP), the difference of measurements between adjacent frames is equivalent to the projection of inter-frame residuals, i.e.,

$$\boldsymbol{y}_{t+1} - \boldsymbol{y}_t = \boldsymbol{\phi}\boldsymbol{x}_{t+1} - \boldsymbol{\phi}\boldsymbol{x}_t = \boldsymbol{\phi}(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \tag{7}$$

Therefore, video residuals can be acquired by the single-pixel camera and a subtraction operation. For the scene of slow motion or video surveillance, the neighboring frames are much similar and inter-frame residuals have an intensive sparsity, so it is more conducive to be measured and rebuilt via CS. But for a general video sequence, the reconstructing performance changes with inter-frame residuals.

The DISCOS[19] and DVC-CS[20] introducing the traditional video coding method to key-frame coding demand the traditional camera to sample those data. In view of low cost, we only discuss the DCVS[21] in which all frames can be acquired by CS camera. The DCVS combines the idea of distributed compressed sensing (DCS) and distributed video coding (DVC), and it regards video sequences as the relative sources in the joint sparse model (JSM). Each frame is taken CS measurement individually at the encoding end, and the non-key frames are jointly reconstructed based on the side information at the decoding end. As the coding scheme of DCVS is concise and the decoding algorithm is very flexible, it is especially suitable for many fields such as low-cost digital camera, power-saving and mobile video collecting equipment, distributed sensor network, and so on. However, its reconstruction performance still needs further improving.

2 RDCVS-DSI

2.1 The basic idea and framework for RDCVS-DSI

In general, the coding algorithms of RVCS and DCVS improve the 2D-CS algorithm only in view of inter-frame correlation between video frames, but not increase the coding efficiency through their own characteristics. According to the correlation between successive frames described in (6), finding an appropriate side information ($\boldsymbol{x}_c$), the $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$ can be encoded through compressing the $\boldsymbol{x}_{t\_u}$ and $\boldsymbol{x}_{t+1\_u}$. As the residual sparsity is very strong, it is more advantageous to carry on CS coding.

This study intends to regard the low quality image of the original frame $\boldsymbol{x}_l$ (similar to $\boldsymbol{x}_c$) as the side information of the $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$ for reference. The successive inter-frame model in (6) can be expanded to a related model between the key frame $\boldsymbol{x}_k$ and the multiple non-key frames $\boldsymbol{x}_{nk}$ in a GOP, then

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{x}_{k\_l} + \Delta\boldsymbol{x}_k \\ \boldsymbol{x}_{nk} = \boldsymbol{x}_{nk\_l} + \Delta\boldsymbol{x}_{nk} \\ \boldsymbol{x}_{nk\_l} = f(\boldsymbol{x}_{k\_l}) \end{cases} \tag{8}$$

where $\boldsymbol{x}_{k\_l}$ represents the low quality version of the key frame, $\boldsymbol{x}_{nk\_l}$ represents the low quality version of the non-key frame, $\triangle\boldsymbol{x}_k$ and $\triangle\boldsymbol{x}_{nk}$ represent the residuals between the key/non-key and its low quality version, and $f(\cdot)$ indicates the relationship between the key and non-key frames in low quality version.

In order to guarantee quickly obtaining the reference frame in low quality version both in the encoding and decoding ends, the first side information (SI1) is considered to be generated with a large amount of information and a few data at the encoding end, and it is sent to the decoding end together with the measurement value, so as to be quickly converted to the reference information for decoding. As wavelet transform has a characteristic of time-frequency scalability, and its main energy concentrates in the low frequency, the wavelet coefficients are taken in the lowest layer as SI1.

Because there is a strong sparsity in detailed information of the difference between the key/non-key frame and its low quality version in the same GOP, the residuals for a key or non-key frame can be measured respectively. As the SBHE has the advantages of good performance, simple operation, less memory, etc., it is suitable for video measuring

and coding. Thus, the SBHE is taken as the measurement matrix $\boldsymbol{\phi}$ in this paper. The following formulas give two groups of information sources that the coding end needs to sample:

$$\begin{cases} \boldsymbol{\theta}_{L\_k} = \mathrm{LP}_{\mathrm{DWT}}(\boldsymbol{x}_k) \\ \Delta \boldsymbol{y}_k = \boldsymbol{\phi} \Delta \boldsymbol{x}_k \\ \boldsymbol{\phi} = \boldsymbol{qwp} \end{cases} \quad \text{or} \quad \begin{cases} \boldsymbol{\theta}_{L\_nk} = \mathrm{LP}_{\mathrm{DWT}}(\boldsymbol{x}_{nk}) \\ \Delta \boldsymbol{y}_{nk} = \boldsymbol{\phi} \Delta \boldsymbol{x}_{nk} \\ \boldsymbol{\phi} = \boldsymbol{qwp} \end{cases} \tag{9}$$

where $\mathrm{LP}_{\mathrm{DWT}}(\cdot)$ denotes the operation of taking low frequency components of discrete wavelet transform (DWT), while $\boldsymbol{\theta}_{L\_k}$ and $\boldsymbol{\theta}_{L\_nk}$ denote the low frequency coefficients of $\boldsymbol{x}_k$ and $\boldsymbol{x}_{nk}$. The measurement matrix $\boldsymbol{\phi}$ for a residual frame can be decomposed into the product of random selected matrix $\boldsymbol{q}$, the block diagonal Hadamard matrix $\boldsymbol{w}$, and the random permutation matrix $\boldsymbol{p}$. $\triangle \boldsymbol{y}_k$ and $\triangle \boldsymbol{y}_{nk}$ denote the residual measurement values for the key and non-key frames, respectively.

Due to the strong sparsity of $\triangle \boldsymbol{x}_k$ on behalf of the high frequency information of a key frame, it can be directly solved via convex optimization, adding to SI1 and performing inverse discrete wavelet transform (IDWT), and finally the key frame will be reconstructed in high quality. In order to compare with the existing CS video coding schemes, the GPSR has been selected as the basic reconstruction algorithm due to its high efficiency. The decoding for the non-key frame needs to reconstruct $\triangle \boldsymbol{x}_{nk}$ via convex optimization and to combine the second side information (SI2). SI2 is generated completely on the decoding side, relying on motion estimation and compensation (MEMC) by the reconstructed key frames. The initial condition, main iterative formula and auxiliary iterative condition for reconstructing $\triangle \boldsymbol{x}_{nk}$ are shown as follows

$$\begin{cases} \Delta \boldsymbol{x}_{s2} = \mathrm{MEMC}(\boldsymbol{x}_k), \quad \Delta \boldsymbol{\theta}_{s2} = \mathrm{DWT}(\Delta \boldsymbol{x}_{s2}) \\ \min\limits_{\Delta \boldsymbol{\theta}_{nk}} \|\Delta \boldsymbol{y}_{nk} - \boldsymbol{\phi} \Delta \boldsymbol{x}_{nk}\|_{l_2} + \|\Delta \boldsymbol{\theta}_{nk}\|_{l_1} \quad \text{s.t.} \quad \Delta \boldsymbol{\theta}_0 = \Delta \boldsymbol{\theta}_{s2} \\ \max\limits_{\tau} \tau(\Delta \boldsymbol{x}_{nk}^{(i-1)}, \Delta \boldsymbol{x}_{nk}^{(i)}) \quad \text{s.t.} \quad \Delta \boldsymbol{x}_{nk} = \mathrm{IDWT}(\Delta \boldsymbol{\theta}_{nk}) \end{cases} \tag{10}$$

where $\mathrm{MEMC}(\cdot)$ denotes the operation of motion estimation (ME) and motion compensation (MC) using the reconstructed key frames. In the actual decoding structure, to reduce reconstructing complexity of the non-key frame, the two operations are completed respectively by processing the key and non-key frame decoding. $\triangle \boldsymbol{x}_{s2}$ and $\triangle \boldsymbol{\theta}_{s2}$ represent SI2 in a pixel domain and a wavelet domain individually, while $\triangle \boldsymbol{x}_{nk}$ and $\triangle \boldsymbol{\theta}_{nk}$ represent the non-key frame residuals and its coefficients in the wavelet domain. In processing of the convex optimization reconstruction, its initial value is $\triangle \boldsymbol{\theta}_{s2}$, while the main iterative target is minimizing unconstrained function based on l_1-l_2 norm. $\triangle \boldsymbol{x}_{nk}^{(i-1)}$ and $\triangle \boldsymbol{x}_{nk}^{(i)}$ respectively represent the $(i-1)$th and ith iterations of the non-key frame residuals, their correlation can be modeled as a Laplace distribution $\tau(\cdot)$, and the auxiliary iterative target is to maximize τ value.

Based on the above analysis, the basic coding structure for RDCVS-DSI is summarized in Fig. 1, in which the thick line denotes the residual measurement value and SI1 in the frequency domain to be transmitted from the encoding end to the decoding one.

Based on the compressed sensing theory, the sparser is the signal, the lower measurement rate is needed to exactly recover the original signal. In order to protect the reconstruction performance, the key frame, as the reference frame, is usually acquired at a higher sampling rate. Then after deducting the first side information, the residual measured value is sufficient to precisely recover the residual frame. While the reconstructed key frames are more accurate, and the video section is closer to uniform translation, the second side information obtained from motion estimation and motion compensation is more accurate. Surely the reconstruction performance of non-key frames will be improved, and the convergence speed of the convex optimization algorithm be shortened. On the other hand, in order to improve the compression rate, the sampling rate of non-key frame is generally made low, and we must rationally allocate the first side and the residual measurement rates according to the sparsity of residual frame. When the second side of information is more precise, a higher sampling rate can be allocated to residual measurement, which results in fast and efficient reconstruction of the non-key frame; when the second side information is not nice enough, we should increase the sampling rate of the first side information to get a more sparse residual frame, so as to ensure the reconstructing accuracy of non-key frames.

2.2 Performance analysis for RDCVS-DSI

As the RDCVS-DSI measures the residual signal, under

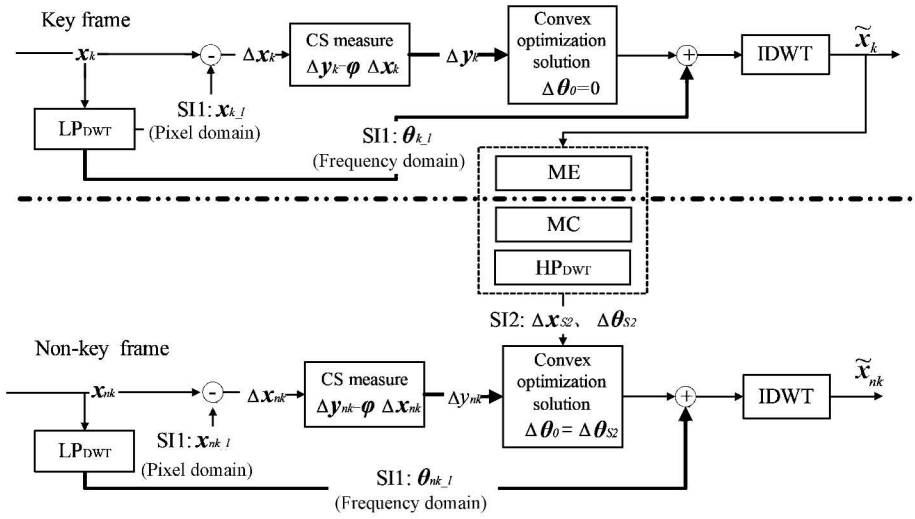


Fig. 1 Basic coding structure for RDCVS-DSI

the precondition of using the same measure number and the reconstructing algorithm, its reconstruction quality mainly depends on the residual sparsity (rs), the accuracy of initial iterative value (iv) in reconstructing and the suitability of stopping rule (sr). Then the reconstruction quality for the key and non-key frames (Q-k and Q-nk) in a GOP can be expressed as

$$Q\text{-}k = \alpha \times rs\_if_k + \beta \times iv_k + \gamma \times sr_k \quad (11)$$

$$Q\text{-}nk = \alpha \times rs\_if_{nk} + \beta \times iv_{nk} + \gamma \times sr\_rdcvs_{nk} \quad (12)$$

where rs_if_k, rs_if_{nk} and α respectively denote the average residual sparsity for the key and non-key frames, as well as their weight factors in reconstruction quality. iv_k, iv_{nk} and β respectively denote the accuracy of initial value for the key and non-key frames, and their weight factors in the residual reconstructing. sr_k, sr_rdcvs_{nk} and γ respectively denote the suitability of stop criterion for the key and non-key frames, and their weight factors.

The encoding for the key frame mainly includes low-pass filtering (lp), subtraction (sub) and residual measurement (mea), while the decoding for the key frame mainly includes convex optimization solving (sol), inverse discrete wavelet transform ($idwt$), motion estimation (me) and addition (add) operations. The encoding steps for the non-key frames are similar to the key ones, while its decoding includes convex optimization solving with initial value and multiple stopping criterions (sol_rdcvs), motion compensation (mc), high-pass filtering (hp), inverse discrete wavelet transform ($idwt$), and addition (add) operations. The computational complexity of $idwt$ is equal to dwt, while lp and hp have the similar computational complexity, and their multiplications are not more than a half of dwt. Compared to dwt, the computational complexity for add and sub operations should be negligible. Therefore, the coding complexity for the key and non-key frames (C-k and C-nk) can be approximately expressed as

$$C\text{-}k = lp + mea + sol + idwt + me \approx mea + 1.5dwt + sol + me \quad (13)$$

$$C\text{-}nk = lp + mea + sol\_rdcvs + mc + hp + idwt \approx mea + 2dwt + sol\_rdcvs + mc \quad (14)$$

For checking the coding performance of RDCVS-DSI algorithm, the reconstruction quality and computational complexity are illustrated in Tables 1 and 2, together with 2D-CS, RVCS and DCVS algorithms.

Regardless of the key or non-key frame, the sparsity of a frame (fs_k or fs_{nk}) is less than that of the residual frame, and the sparsity of intra-frame residuals (rs_if_k or rs_if_{nk}) reflecting the high frequency components of a frame is usually stronger than that of inter-frame residuals (rs_df_{nk}) reflecting the inter-frame motion. That is,

$$rs\_if_k > fs_k, \quad rs\_if_{nk} > rs\_df_{nk} > fs_{nk} \quad (15)$$

The convex optimization solving for a key frame via four algorithms named 2D-CS, RVCS, DCVS and RDCVS-DSI, as well as for a non-key frame via the first two algorithms, only has main iterative target, so sr_k and sr_{nk} can be considered the same. As the convex optimization solving for a non-key frame by the last two algorithms has introduced iterative initial value and auxiliary iterative target, the relevant experiment[21] has showed that the joint effect of iv_{nk} and sr_dcvs_{nk} is better than sr_{nk}. In addition, sr_rdcvs_{nk} for a residual frame is much finer than sr_dcvs_{nk} for a frame, thus three algorithms are the same, and DCVS are intermediate.

In the computational complexity problem, the multiplication computation by fast wavelet transform and weighted motion compensation for an image containing N pixels are both O($4N$), and the measurement computation on it using block Hadamard matrix sized M is about O(MN). It is difficult to precisely calculate the computational cost of the convex optimization solving via GPSR, however it contains a moderate amount of inner product, measurement, vector-scalar multiplication and vector addition in each iteration step. As the complexity of each iteration exceeds the total complexity of lp, dwt, mea, add and sub, sol without initial condition is the most complicated step in decoding. Due to the introduction of reference side information and auxiliary iterative condition, the iterative number for sol_dcvs (or sol_rdcvs) decreases, meanwhile, the operation for generating side information (me + mc) adds. The relevant experiment[21] has showed the joint complexity of the later two is below sol. Therefore, the coding complexity via DCVS is the lowest, RDCVS-DSI takes second place. As the key frame coding by RVCS needs sol twice, its complexity is supreme.

3 Simulation

In order to validate the performance of RDCVS-DSI, we have taken simulations by Matlab tool, using peak signal-to-noise ratio (PSNR) to reflect the reconstruction quality,

Table 1 Performance comparison for the key frame among 4 algorithms

	Q-k	C-k
2D-CS		$mea + dwt + sol$
RVCS	$\alpha \times fs_k + \gamma \times sr_k$	$mea + 2dwt + 2sol$
DCVS		$mea + dwt + sol$
RDCVS-DSI	$\alpha \times rs\_if_k + \beta \times iv_k + \gamma \times sr_k$	$mea + 1.5dwt + sol + me$

Table 2 Performance comparison for the non-key frame among 4 algorithms

	Q-nk	C-nk
2D-CS	$\alpha \times fs_{nk} + \gamma \times sr_{nk}$	$mea + dwt + sol$
RVCS	$\alpha \times rs_{nk} + \gamma \times sr_{nk}$	$mea + dwt + sol$
DCVS	$\alpha \times fs_{nk} + \beta \times iv_{nk} + \gamma \times dcvs_{nk}$	$mea + dwt + sol\_dvcs + me + mc$
RDCVS-DSI	$\alpha \times rs\_if_k + \beta \times iv_{nk} + \gamma \times sr_{nk}$	$mea + 2dwt + sol\_rdcvs + mc$

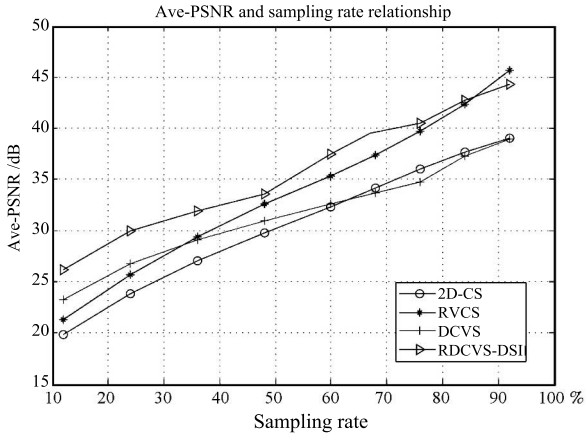
encoding and decoding time (enc-dec-time) to reflect the computational complexity. We employ 9/7 wavelet as the sparse operator, SBHE as measurement matrix, and adopt GPSR as reconstructing algorithm. The initial material is taken from the first 100 frames in two groups of standard QCIF sequences with different video features — foreman (smoothing image and medium-speed moving scenes) and coastguard (various texture and slowly moving scenes). The experimental platform is a desktop computer, configured as AMD Athlon (tm) 64 × 2 Dual Core Processor 4800+, dominant frequency 2.53 GHz and memory 2 G.

For convenient comparison, we also have taken the related modules of three algorithms — 2D-CS[13], RVCS[17] and DCVS[21] as references. To be fair, the amount of SI1 requiring transferring to the decoder via RDCVS-DSI algorithm should be converted into the measurement number and recorded in the sampling rate together with the residual measurement number. That is, for three reference algorithms, the average sampling rate = the measurement data quantity/the original data quantity, while for RDCVS-DSI, the average sampling rate = (the measurement data quantity + the side information data quantity)/the original data quantity.

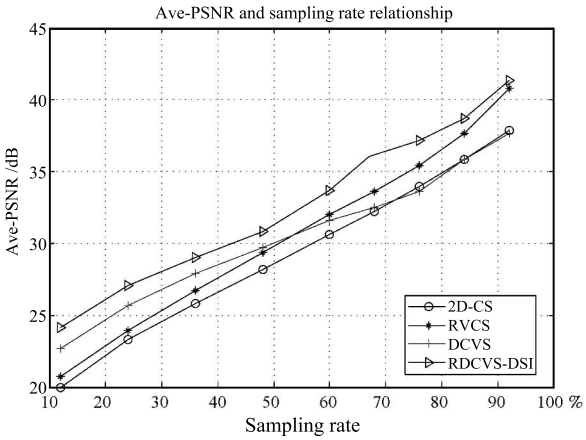
Figs. 2 and 3 show the graphical charts about the average PSNR (Ave-PSNR) and average enc-dec-time (Ave-enc-dec-time) for recovering video sequence by the four algorithms as the sampling rate slopes up, when the size of GOP (GOPSize) is 5. Figs. 4 and 5 show these two graphical charts as the GOPSize grows, when fixing the sampling rate of the key frame to be 70 %, and the non-key frame 30 %. The subgraphs (a) and (b) in Figs. 2 ∼ 5 show the simulation results for the two groups of the given test sequences separately.

Being seen from Fig. 2, in the process of the sampling rate changing, the RDCVS-DSI considering the correlations both inside frame and between frames has achieved better PSNR than the other three algorithms; the RVCS and DCVS merely considering the inter-frame correlation take the second place, while 2D-CS considering neither the intra-frame nor inter-frame correlation gets the worst PSNR.

Fig. 3 shows the computational complexity in view of the average encoding and decoding time. Although the coding complexity of RDCVS-DSI increases slightly compared with DCVS at low sampling rate, its reconstruction quality ascends. With the growth of the sampling rate, the total coding time of RDCVS-DSI is close to, even less than that of DCVS, and the 2D-CS is longer, RVCS the longest.

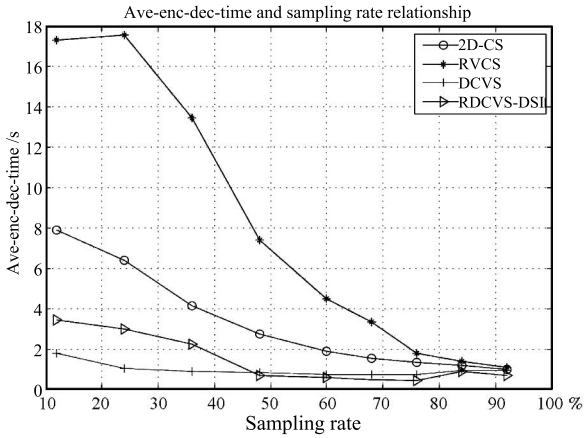


(a) Foreman

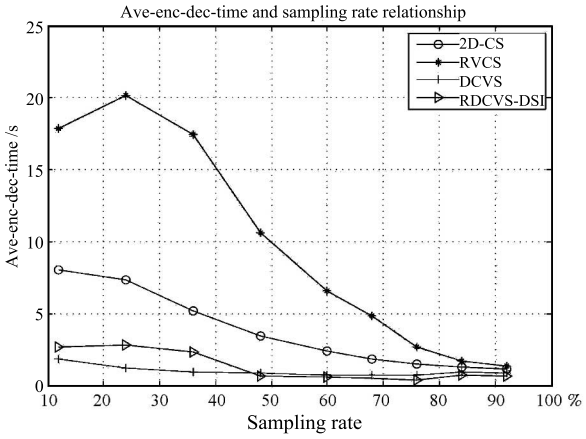


(b) Coastguard

Fig. 2 Ave-PSNR and sampling rate relationship in QCIF format



(a) Foreman



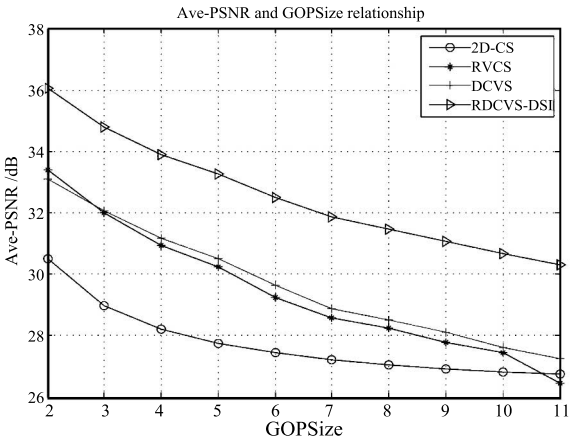
(b) Coastguard

Fig. 3 Ave-enc-dec-time and sampling rate relationship in QCIF format

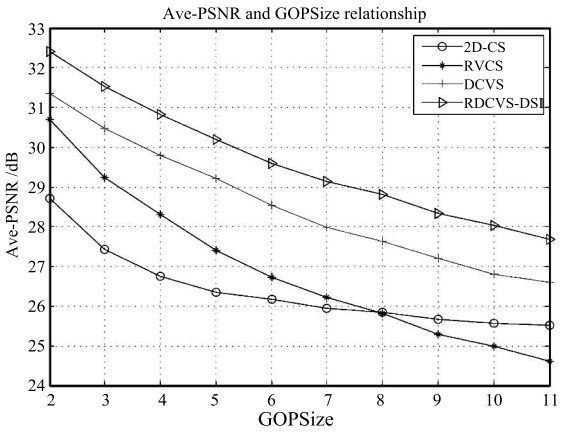
From Figs. 4 and 5, as the GOPSize increases the correlation between frames abates, the PSNR for RDCVS-DSI algorithm is always above other algorithms, and its coding time gets close to DCVS with the lowest cost.

Further, we get the second material from the first 50 frames in the two groups of standard CIF sequences with

different video features — mobile (intermediate-speed moving) and bus (fast moving). The relationship of ave-enc-dec-time is similar to the result in QCIF. So we only show the PSNR as the sampling rate grows, as that in Fig. 6. For saving space, we ignore the PSNR with GOPSize, since the tendency is easy to estimate from Figs. 4 and 6.

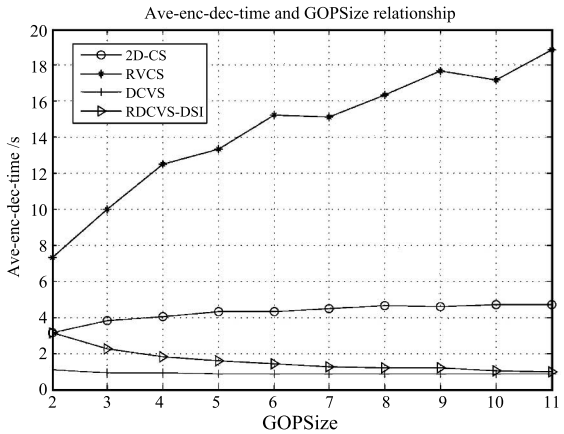


(a) Foreman

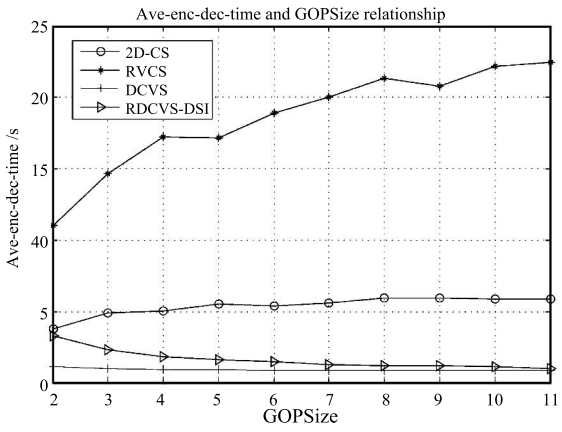


(b) Coastguard

Fig. 4 Ave-PSNR and GOPSize relationship in QCIF format

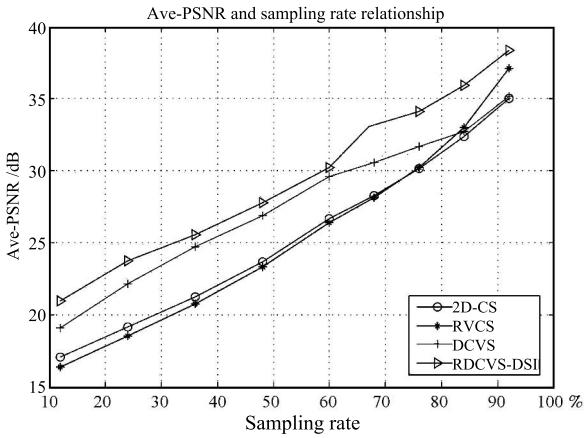


(a) Foreman

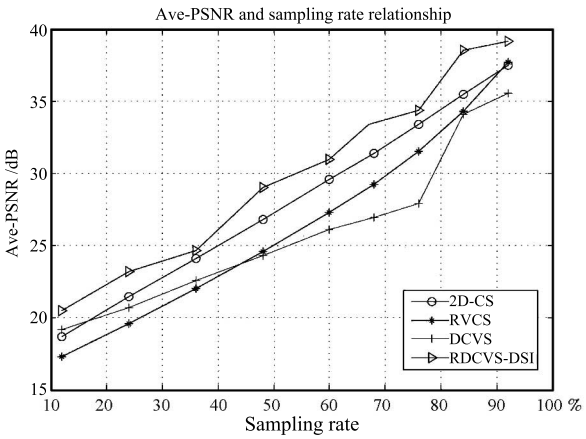


(b) Coastguard

Fig. 5 Ave-enc-dec-time and GOPSize relationship in QCIF format



(a) Mobile



(b) Bus

Fig. 6 Ave-PSNR and sampling rate relationship in CIF format

We can see that when the object is making translation as in mobile sequence, the DCVS obtains a higher PSNR than the 2D-CS and RVCS, because the relevant model coincides with its moving track. When the target motion is too intense as in bus sequence, the inter-frame advantage of

the RVCS and DCVS might become invalid. Fortunately, our RDCVS-DSI is optimal in both situations.

Finally, we test the four algorithms with a real AVI stream — racing car (512 × 384, varied movement). We can see from Fig. 7 that the performances of the former three algorithms vary in the real situation, as compared with the simulation sequence. In any case, the RDCVS-DSI still keeps ahead by reasonably adjusting the sampling rate between the first side information and residual measurement.

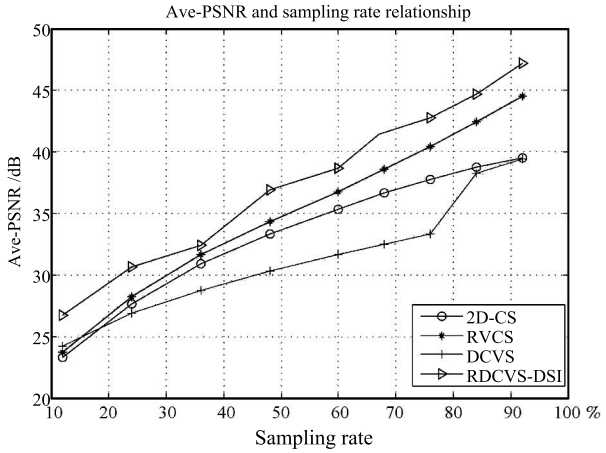


Fig. 7 Ave-PSNR and sampling rate relationship in AVI format

4 Conclusion

For taking full advantage of the video spatial and temporal redundancy so as to improve the existing performance of compressive video sensing, a structure of residual distributed compressive video sensing based on double side information (RDCVS-DSI) is proposed in this paper. In the RDCVS-DSI, in the pre-filtering process is carried on before measuring, and the transform coefficients through the low-pass filtering are regarded as SI1. Then the high-frequency components are measured after filtering, and the residual measured values together with SI1 are acquired and transmitted to the decoding end. In the corresponding decoding side, the residuals in sparse domain are rebuilt via convex optimization, and then the key frame is reconstructed by adding SI1 and performing inverse transform from the sparse domain. The residuals of the non-key frames are jointly reconstructed by SI2, which is obtained by motion estimation, motion compensation and the high-pass filtering.

The innovations lie in: 1) enhance the information content of side information only for the decoding of the non-key frame in DCVS, and decompose it into SI1 (the low-frequency components for both video frames) in both encoding and decoding sides and SI2 (the high-frequency components for the non-key frame) in the decoding side, so as to improve compression efficiency and reconstruction performance; 2) introduce the joint reconstruction method to residual rebuilding of the non-key frame, and generate the side information through motion estimation by the key frame, motion compensation and high-pass filtering by the non-key frame individually, so as to improve the reconstruction speed to some extent. Performance analysis and simulation are conducted compared with the existing video CS algorithms. Test results show that the coding algorithm of the RDCVS-DSI, according to the changing conditions of video sequences with different characteristics, sampling rate and image group size, can obtain good reconstruction quality with quite low complexity. About $1 \sim 5$ dB gain in PSNR of the reconstructed frames is observed, and the speed is close to the least complex DCVS, when compared with prior works on compressive video sensing. The next step will be focused on intra-frame and inter-frame characteristics of video, level selection of side information and sampling rate allocation to further improve the overall performance of the RDCVS-DSI.

References

1 Donoho D L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, **52**(4): 1289−1306

2 Baraniuk R G, Candes E J, Nowak R, Vetterli M. Compressive sampling. *IEEE Signal Processing Magazine*, 2008, **25**(2): 12−13

3 Dai Qiong-Hai, Fu Chang-Jun, Ji Xiang-Yang. Research on compressed sensing. *Chinese Journal of Computers*, 2011, **34**(3): 425−434 (in Chinese)

4 Jiao Li-Cheng, Yang Shu-Yuan, Liu Fang, Hou Biao. Development and prospect of compressive sensing. *Acta Electronica Sinica*, 2011, **20**(7): 1651−1662 (in Chinese)

5 Davenport M A, Duarte M F, Eldar Y C, Kutyniok G. Introduction to compressed sensing. *Compressed Sensing: Theory and Applications*. Cambridge: Cambridge University Press, 2012.

6 Li Ran, Gan Zong-Liang, Zhu Xiu-Chang. A fast compressed-sensing image reconstruction algorithm based on best linear estimate. *Journal of Electronics and Information Technology*, 2012, **34**(12): 3006−3012 (in Chinese)

7 Li Zhi-Lin, Chen Hou-Jin, Yao Chang, Li Ju-Peng. Compressed sensing reconstruction algorithm based on spectral projected gradient pursuit. *Acta Automatica Sinica*, 2012, **38**(4): 1218−1223 (in Chinese)

8 Wang Rong-Fang, Jiao Li-Cheng, Liu Fang, Yang Shu-Yuan. Block-based adaptive compressed sensing of image using texture information. *Acta Electronica Sinica*, 2013, **41**(8): 1506−1514 (in Chinese)

9 Song Xiao-Xia, Shi Guang-Ming. Fewer Bernoulli measurements satisfying the constraint of reconstruction probability. *Acta Automatica Sinica*, 2013, **39**(1): 53−56 (in Chinese)

10 Liu Fang, Wu Jiao, Yang Shu-Yuan, Jiao Li-Cheng. Research advances on structured compressive sensing. *Acta Automatica Sinica*, 2013, **39**(12): 1980−1995 (in Chinese)

11 Edwards J. Focus on compressive sensing (special reports). *IEEE Signal Processing Magazine*, 2011, **28**(2): 11−13

12 Duarte M F, Davenport M A, Takhar D, Laska J N, Ting S, Kelly K F, Baraniuk R G. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 2008, **25**(2): 83−91

13 Wakin M B, Laska J N, Duarte M F, Baron D, Sarvotham S, Takhar D, Kelly K F, Baraniuk R G. Compressive imaging for video representation and coding. In: Proceedings of the 2006 Picture Coding Symposium (PCS). Beijing, China, 2006. 711−716

14 Lu G. Block compressed sensing of natural images. In: Proceedings of the 2007 International Conference, Digital Signal Processing (DSP). Cardiff, UK: IEEE, 2007. 403−406

15 Stankovic V, Stankovic L, Cheng S. Compressive video sampling. In: Proceedings of the 2008 Signal Processing Conference (EUSIPCO). Lausanne, Switzerland: Elsevier, 2008. 1−5

16 Wahidah I, Suksmono A B, Mengko T L R. A comparative study on video coding techniques with compressive sensing. In: Proceedings of the 2011 Electrical Engineering and Informatics (ICEEI). Bandung, Indonesia: IEEE. 2011: 1−5

17 Zheng J, Jacobs E L. Video compressive sensing using spatial domain sparsity. *Optical Engineering*, 2009, **48**(8): 1−10

18 Yantian H, Feng L. A low-complexity video coding scheme based on compressive sensing. In: Proceedings of the 2011 International Symposium, Computational Intelligence and Design (ISCID). Hangzhou, China: IEEE, 2011. 326−329

19 Do T T, Chen Y, Nguyen D T, Nguyen N, Gan L, Tran T D. Distributed compressed video sensing. In: Proceedings of the 43rd Annual Conference on Information Sciences and Systems. Baltimore, MD: IEEE, 2009. 1−2

20 Prades-Nebot J, Yi Ma, Huang T. Distributed video coding using compressive sampling. In: Proceedings of the 2009 Picture Coding Symposium (PCS). Chicago, USA: 2009. 1−4

21 Kang L W, Lu C S. Distributed compressive video sensing. In: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Taipei, China: IEEE, 2009. 1169−1172

22 Fowler J E, Mun S, Tramel E W. Block-based compressed sensing of images and video. *Foundations and Trends in Signal Processing*, 2012, **4**(4): 297−416

23 Wang A, Zhao M, Deng S, Zhang X. An efficient residual-based distributed compressive video sensing. *Journal of Computational Information Systems*, 2011, **7**(16): 5732−5737

24 Liu Z R, Elezzabi A Y, Zhao H V. Maximum frame rate video acquisition using adaptive compressed sensing. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, **21**(11): 1704−1718

25 Wu M H, Zhu X C, Gan Z L, Li X. Adaptive dictionary learning for distributed compressive video sensing. *International Journal of Digital Content Technology and Its Applications*, 2012, **6**(4): 141−149

26 Li X X, Wei Z H. Compressed sensing video images recursive reconstruction algorithm based on local autoregressive model. *Acta Electronica Sinica*, 2012, **40**(9): 1795−1800

27 Hegde C, Baraniuk R G. Signal recovery on incoherent manifolds. *IEEE Transactions on Information Theory*, 2012, **58**(12): 7204−7214

28 Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2007, **53**(12): 4655−4666

29 Bioucas-Dias J M, Figueiredo M A T. A new twIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 2007, **16**(12): 2992−3004

30 Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 2007, **1**(4): 586−597

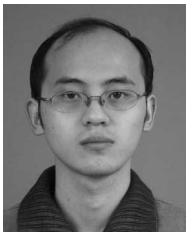
31 Wright S J, Nowak R D, Figueiredo M A T. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 2009, **57**(7): 2479−2493

32 Do T T, Gan L, Nguyen N H, Tran T D. Fast and efficient compressive sensing using structurally random matrices. *IEEE Transactions on Signal Processing*, 2012, **60**(1): 139−154

CHEN Jian Lecturer at the College of Physics and Information Engineering, Fuzhou University. She received her master degree in communication and information system from Fuzhou University in 2007. Her research interest covers image processing, video coding, and compressed sensing. E-mail: chenjian-fzu@163.com

SU Kai-Xiong Professor at the College of Physics and Information Engineering, Fuzhou University. He received his master degree from University of Science and Technology of China in 1988. His research interest covers wireless communications, video broadcasting, and embedded systems. Corresponding author of this paper. E-mail: skx@fzu.edu.cn

WANG Wei-Xing Professor at the College of Physics and Information Engineering, Fuzhou University. He received his Ph. D. degree from Royal Institute of Technology of Switzerland in 1997. His research interest covers image processing, pattern recognition, and machine vision.
E-mail: wxwwx@fzu.edu.cn



LAN Cheng-Dong Lecturer at the College of Physics and Information Engineering, Fuzhou University. He received his Ph. D. degree from Wuhan University in 2011. His research interest covers video coding, image processing, and multimedia communication. E-mail: lcd217@263.net