 doi:10.20944/preprints201807.0222.v1

Article

A DCVS Reconstruction Algorithm for Mine Video Monitoring Image Based on Block Classification

Xiaohu Zhao [1,2], Xueru Shen [1,2,*], Kuan Wang[1, 2] and Wanmei Li[1, 2]

[1] The National and Local Joint Engineering Laboratory of Internet Application Technology on Mine, Xuzhou 221000, China; TS16060112A3@cumt.edu.cn
[2] China University of Mining and Technology, Xuzhou 221000, China; TS16060247P3@cumt.edu.cn
* Correspondence: TS16060247P3@cumt.edu.cn; Tel.: +86-151-525-01697

Abstract: Aiming at the problems that large amount of video monitoring image data in underground coal mines leads to difficulties in transmission and storage, compressed sensing theory is introduced to encode and decode video images, and a new distributed video coding scheme is proposed. In order to obtain more sparse representation and more general applicability, a block-based adaptive sparse base scheme is proposed. For the acquisition of side information, fixed weight is usually used to synthesize side information and the correlation between different image blocks is neglected, a block-based classification weighted side information generation scheme is proposed. Experimental results show that the block-based classification codec scheme can make full use of inter-frame correlation. Under the appropriate sampling rate, the PSNR value of video reconstruction increases, which effectively improves the quality of video frame reconstruction.

Keywords: compressed sensing; distributed video codec; sparse representation; side information reconstruction

1. Introduction

Coal is an important basic energy source in China and plays a decisive role in the development of economic construction [1]. As a non-renewable resource, the shallow coal mine resources are gradually mined and turned to the exploitation of deep coal seams, and the dangers that come with it are also increasing. Mine smart monitoring, coal seam identification and mineral detection have high requirements on the quality and real-time performance of video images, especially in unattended working areas in coal mines [2]. The amount of data of the monitoring image is large, so the requirement for the device in terms of image transmission and storage is high. The wireless sensor network (WSN) relies on low-power wireless sensor nodes, overcomes the difficulties of wired transmission wiring, incomplete detection information, high deployment costs and poor flexibility, and is widely used in the transmission of information in coal mines. Due to the particularity of underground coal mine working environment, a large number of sensor nodes need to be arranged downhole to monitor various signals such as gas concentration, temperature, humidity, and mine quake signals. Therefore, the amount of data the sensor needs to process is huge. However, the node energy and transmission bandwidth of the sensor are limited, and the energy supply of the mine node is difficult, so it is hard to apply a large amount of data collection. How to reduce the amount of information transmission and energy consumption of nodes is an urgent task to extend the life cycle of wireless network nodes, and data compression is a data processing technology that can effectively reduce the amount of data transmission. Compared with the traditional Nyquist sampling mode, Compressed Sensing [3-4] technology is devoted to searching for sparse solutions of underdetermined linear systems. The signal can be reconstructed

with a sampling rate much lower than the Nyquist sampling rate [5], which is suitable for signal processing.

In the video codec scheme, traditional video coding such as MPEG/H.26X, there will be a large amount of motion estimation and motion compensation [6], which makes the encoding side much more complex than the decoding side, this asymmetric coding scheme is suitable for a situation where multiple encodings are encoded at one time, such as broadcasting. However, more scenarios require low complexity, low power consumption, and low compression ratio at the encoder end. The distributed video coding (DVC) has low coding complexity and is suitable for use in scenarios requiring low complexity. In a distributed video coding scheme, the decoder performs motion estimation and compensation to find the correlation between adjacent frames, which makes the coding side less complex than the decoding side. When using CS theory to reconstruct the signal, the encoding side is less complex than the decoding side. However, using only the CS theory cannot provide us with a sufficiently low sampling rate, so we combine the CS theory and the DVC theory to form a new theory -- distributed compressed video sensing (DCVS) to make up for this deficiency. The DCVS technology allows independent encoding of multiple statistically related signals at the encoder and joint decoding at the decoder, which eliminates the need for complex motion estimation and compensation at the encoder end, thereby reducing the complexity of the encoder.

Combining the compressed sensing theory and the characteristics of distributed video codec, this paper proposes a block-based adaptive sparse representation and weighted side information reconstruction scheme. Firstly, the related theories of compressed sensing and distributed video coding are briefly introduced. Based on the correlation between video frames, a block-based classification sparse representation and classification weighted side information reconstruction scheme is proposed. Experimentally, the reconstructed PSNR value and time complexity are obtained, the experimental results are analyzed to verify the effectiveness of the algorithm.

2. Research Background Theoretical Analysis

2.1.Analysis of Characteristics of Wireless Sensor Network and Video Image in Underground Coal Mine

Under the special working environment of coal mines, if we want to fundamentally solve the problem of safe and efficient production of coal mines, the coal mining industry must shift from labor-intensive to technology-intensive, making it a new industry, new business model, and new model with high-tech features, and take the road of smart, few people (unmanned) and safe mining [7]. Under the transition of technology-intensive, wireless sensor networks are widely used in information transmission in coal mines because of their advantages such as low-power sensor nodes. The communication distance between WSN nodes is limited, and a large number of sensor nodes need to be arranged. However, the energy of the nodes is limited. The more data the nodes send and receive, the greater the energy consumption. But the downhole power supply in coal mines is difficult, so the data transmission needs to be reduced to reduce energy consumption. Large-scale deployment of sensor nodes will cause different nodes in the network to transmit similar data, resulting in a large amount of data redundancy and reducing transmission efficiency. Therefore, it is very necessary to apply compressed sensing theory to deal with video images in coal mines. In the collection of video images in underground coal mines, the collected images are mostly grayscale images composed of black and white. The hue is relatively single, but the correlation between adjacent images is very strong, and there is a large amount of redundant information between video frames. Compressed sensing technology can compress image information better and reduce transmission pressure.

2.2. Compressed Sensing Theory

The compressed sensing theory proposed by Donoho et al. is a novel sampling theory. From the perspective of analog to digital conversion, the compressed sensing theory provides a method of sampling at a rate lower than the Nyquist sampling rate without distortion recovery [8]. Different

from the traditional sampling theorem, the theory points out that the precondition for the signal sparse representation is that the signal is sparse. The signal x is thus converted to other domain space, represented by the sparse signal y. Consider an image signal x of length N whose transform coefficient on some orthogonal base Ψ is expressed as:

$$x = \Psi\theta \tag{1}$$

Where θ is the coefficient vector of x on the orthogonal basis Ψ, the nonzero number of θ is K and K<<N, The signal x is called K-sparse signal. The linear prediction of the signal x is performed using an observation matrix $\Phi_{M\times N}$ (M<<N) that is incoherent with the orthogonal basis Ψ:

$$y = \Phi x = \Phi\Psi\theta \tag{2}$$

Where y is the observation value and Φ is the observation matrix. In this way, M linear predictions which are much smaller than N are obtained, and by solving an optimization problem, the original signals x can be reconstructed with high probability from these few projections. It can be proved that such a projection contains enough information to reconstruct the signal. In this theoretical framework, the sampling rate is not determined by the bandwidth of the signal, but depends on the structure and content of the information in the signal.

2.3 Distributed Video Coding Theory

Distributed video coding (DVC) is a specific application of distributed source coding (DSC) in video processing. DSC is the codec problem of distributed sources that are related to each other in time or space. The information theory basis of DVC is Slepian-Wolf and Wyner-Ziv theorem [9]. Slepian-Wolf theory describes the conditions that need to be satisfied for lossless coding of related sources. It is proposed that the related source "independent coding, joint decoding" can be the same compression efficiency as "joint coding, and joint decoding". The Wyner-Ziv theory is supplement and development based on the Slepian-Wolf theory and discusses the description of the rate-distortion function $R_{X/Y}^{WZ}(D)$ of the relevant source coding in the case of lossy coding. The Wyner-Ziv rate-distortion function shows the lower limit of the code rate for distributed coding under a distortion constraint. In the case of lossy coding, the rate distortion function obtained by using the side information only at the decoding end and using at the codec side is consistent [6]. Generally, Wyner-Ziv encoding can be equivalent to quantifying sources and then performing Slepian-Wolf encoding.

3. Video Image Sparsity Analysis and Compression Coding Processing

3.1.Sparse Representation of the Signal

3.1.1. Block-prediction and DCT Mixture Sparse Basis

From Fourier transform to wavelet transform to later multi-scale geometric analysis, the purpose of the scientists' research is to study how to provide a more concise and direct analysis method for signals in different functional spaces. All of these transformations are designed to exploit the characteristics of the signal and sparsely represent it, or to improve the nonlinear function approximation of the signal, and further studying the degree of sparsity of the signal or the degree of

 doi:10.20944/preprints201807.0222.v1

energy concentration of the decomposition coefficients using a set of bases in a certain space. From the perspective of signal analysis, the Fourier transform is the basis for signal and digital image processing. Wavelet analysis brings signal and digital image processing to a whole new field. Multi-scale geometric analysis is a new generation of signal analysis tool following wavelet analysis. It has many excellent features such as multi-resolution, localization and multi-directionality. It is more suitable for processing multi-dimensional signals such as images. All these studies laid the foundation for the theory of compressed sensing.

In order to get a better sparse representation, many methods have been studied. In the traditional DCVS framework, linear or orthogonal transform bases such as discrete cosine transform (DCT) or discrete wavelet transform (DWT) are widely used due to their simple and efficient features. Traditional video compression or distributed video coding (DVC) technology based on motion estimation and motion compensation all rely on a high-cost mechanism, that is, perception or sampling and compression are not performed continuously, which leads to waste of resources. In other words, most of the collected raw video data may be discarded in the complex compression process [10]. However, learning-based dictionaries can provide a more sparse representation of image signals than these pre-specified base classes. A dictionary learned from a set of blocks globally extracted from the previous reconstructed neighboring frames is used as sparse basis in [11]. Nevertheless, this dictionary is based on K-SVD algorithm [12] and can be only applied in some specific applications. Dictionaries reported in [13] and [14] are generated by a linear combination of neighboring blocks in preceding and following key frames. However, with limited reference information, the learned dictionary may not be sufficiently redundant. In [15], a simple side information generation technique based on correlation analysis of CS measurements is effective in complexity reduction. But the rate-distortion performance is unsatisfying compared with other schemes.

In the traditional video codec, a hybrid coding method combining prediction and DCT can quickly reduce spatiotemporal correlation and obtain a more sparse representation of the video frame. In order to obtain better rate-distortion performance, we propose a new block prediction and hybrid sparse basis for non-key frames with reference to this hybrid coding method. As shown in equation (3), the new sparse basis consists of two parts, the initial DCT matrix and the block-prediction basis. In this way, initial DCT basis is acquired by linear transform in DCT domain while block-prediction basis is based on the SI generated from the adjacent decoded key frames.

$$\left[\Psi_{DCT}; \Psi_{inter}\right] = \Psi_i \tag{3}$$

Where $\Psi_{DCT} \in R^{N\times N}$ is an initial DCT basis, $\Psi_{inter} \in R^{N\times 1}$ is a block-prediction basis and $\Psi_i \in R^{N\times(N+1)}$ is a newly built hybrid DCT basis.

3.1.2. Classification Judgment Standard

In order to obtain more sparse representations and more general applicability, this paper proposes a new hybrid sparse scheme that combines linear DCT bases and block prediction bases to solve sparse representation problems. Adaptive block-based prediction is used to generate side information and to learn block prediction basis. Since the video is composed of consecutive frames

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

and the time redundancy is particularly large, that is, the inter-frame correlation is very strong. By utilizing the correlation between adjacent frames, the scheme can achieve more sparse representation while reducing the complexity.

Different regions in the video sequence have different inter-frame correlations. We propose an adaptive block prediction scheme for sparse representation. Perform the non-overlapping block processing of the reconstruction results of the two adjacent key frames before and after the current non-key frame. The difference between the *i*-th block x_{t+1}^{i} of the following frame and the *i*-th block x_{t-1}^{i} of the previous frame at the corresponding frame may reflect the correlation between the inter-frames in the video sequence. Consider the residual energy values of the two corresponding frame sub-blocks as the classification criteria, and define the residual energy value as shown in equation (4):

$$E(x_{t-1}^{i}, x_{t+1}^{i}) = \left\| x_{t-1}^{i} - x_{t+1}^{i} \right\|_2^2 \tag{4}$$

However, for video frames where video scenes change rapidly, the residual energy value of the corresponding position may be large. Therefore, simply using the residual energy value as the classification criterion will cause the threshold value to be too dependent on the sequence itself, making the algorithm less versatile. Based on this, it is thought that the ratio of the residual energy to the energy block of the previous key frame is used as the classification criterion, as defined by equation (5):

$$e(x_{t-1}^{i}, x_{t+1}^{i}) = \frac{E(x_{t-1}^{i}, x_{t+1}^{i})}{E(x_{t-1}^{i})} \tag{5}$$

The video frame sub-blocks are classified using thresholds T_1, T_2 ($T_1 < T_2$) for video sequences. Calculate the value of the judgment function of the *i*-th block according to equation (5).If $e(x_{t-1}^{i}, x_{t+1}^{i}) < T_1$,then the block is defined as block with slight movement; if $T_1 < e(x_{t-1}^{i}, x_{t+1}^{i}) < T_2$,then the block is defined as block with big movement; and if $e(x_{t-1}^{i}, x_{t+1}^{i}) > T_2$, then the block is defined as block with severe movement. In the experiment, the video frame is subjected to block classification processing. The classification judgment thresholds T_1 and T_2 take 0.003 and 0.015 respectively. After block classification of non-key frames, the classification results and corresponding sparse base strategies are shown in Figure 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

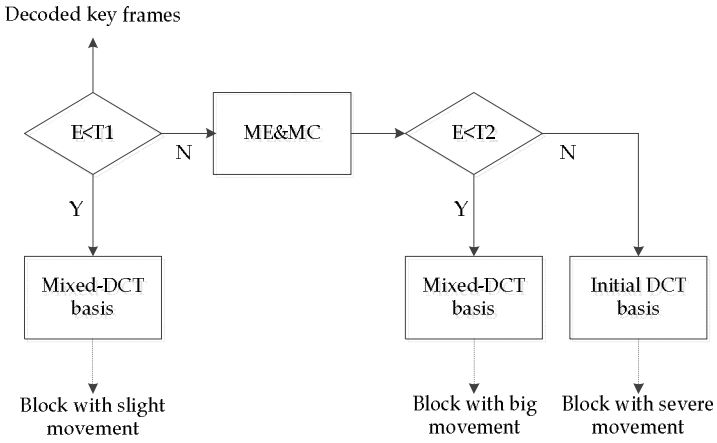

Figure 1. Block classification of non-Key frames

After the non-key frames are blocked according to the defined classification criteria, different sparse-based strategies are proposed for different types of blocks. Algorithm 1 describes the process of adaptive hybrid DCT-based learning. For the block with slight movement, predictions are generated by generating the side information by performing forward and backward motion estimation on adjacent decoded key frames. The block prediction basis is learned and combined with the DCT basis to construct a hybrid DCT basis. For the block with big movement, by selecting different weights and combining forward and backward motion estimation to obtain side information, the construction method of the hybrid DCT base is similar to the block with slight movement. For the block with severe movement, the inter-frame correlation between adjacent frames is small, and motion estimation and compensation cannot accurately predict these fast-changing blocks. Therefore, for this type of block, only the initial DCT basis can be used. The description of algorithm 1 is as follows.

Algorithm 1: The hybrid DCT sparse basis algorithm based on block prediction.

Input: x_{t-1}^i, x_{t+1}^i: Neighboring decoded key frames as reference for the *i*-th non-key frames.

1. Calculate the residual energy value of the corresponding frame sub-block: $E(x_{t-1}^i, x_{t+1}^i) = \left\| x_{t-1}^i - x_{t+1}^i \right\|_2^2$;
2. Calculate the ratio $e(x_{t-1}^i, x_{t+1}^i) = \dfrac{E(x_{t-1}^i, x_{t+1}^i)}{E(x_{t-1}^i)}$ of the residual energy to the energy block of the previous key frame as a classification criterion;
3. Calculate side information $SI^i = \alpha x_{t-1}^i + (1-\alpha) x_{t+1}^i$, Where α selects different weights according to the classification;
4. If $e(x_{t-1}^i, x_{t+1}^i) < T_1$, $\alpha = 0.8$, $\Psi_{inter} = x_t^i / 255$, $\Psi_t(i) = \left[\Psi_{DCT}; \Psi_{inter}\right]$;
5. If $T_1 < e(x_{t-1}^i, x_{t+1}^i) < T_2$, $\alpha = 0.5$, $\Psi_{inter} = x_t^i / 255$, $\Psi_t(i) = \left[\Psi_{DCT}; \Psi_{inter}\right]$; otherwise,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

$\alpha = 0.3$, $\Psi_t(i) = \Psi_{DCT}$.

Output: $\Psi_t(i)$: Hybrid DCT sparse basis for the *t*-th block in the *i*-th non-key frame.

3.2. Design of Observation Matrix

After the sparse representation, how to design a stable $M \times N$ dimensional observation matrix Φ that is not related to the transform basis Ψ, and ensure that the important information of the sparse vector θ when descending from the N dimension to the M dimension is not destroyed, this is the second problem to be solved.

In compressed sensing theory, after the sparse coefficient vector $\theta = \Psi^T X$ of the signal is obtained by transformation, the observation part of the compressed sample is designed. The purpose of the observation matrix design is how to sample and obtain M observations, and it can be ensured that a signal X of length N or a sparse coefficient vector Ψ of the base Ψ can be reconstructed from it. The importance of the observation matrix design is that if the signal X is destroyed during the observation process, it is impossible to reconstruct signal successfully. The observation process actually uses the M row vectors $\{\varphi_j\}_{j=1}^{M}$ of the $M \times N$ dimensional observation matrix Φ to project the sparse coefficient vector. That is, to calculate the inner product between θ and each observation vector $\{\varphi_j\}_{j=1}^{M}$, and get M observations $y_j = \langle \theta, \varphi_j \rangle (j = 1,2,...,M)$, for the observation vector $Y = (y_1, y_2, ..., y_M)$, that is

$$Y = \Phi\theta = \Phi\Psi^T X = A^{CS} X \tag{6}$$

The sampling process here is adaptive, that is, θ does not have to change according to the change of X, and the observed data is no longer a point sampling, but a more general K linear functional of the signal.

Given a vector Y, finding θ from equation (6) is a linear programming problem. However, since M<<N, that is, the number of equations is less than the number of unknowns, this is an underdetermined problem and the deterministic solution cannot be calculated. However, if θ has K-sparseness (K<<M), it is expected to find a certain solution just try to determine the appropriate position of the K non-zero coefficients θ_i in θ. Since the observation vector Y is a linear combination of K non-zero coefficient column vectors, a linear equation of $M \times N$ is formed to solve the specific values of these non-zero terms. For whether there are necessary and sufficient conditions for determining the solution, Restricted Isometry Property (RIP)[16] gives the definition, The necessary and sufficient conditions for the definition are consistent with the geometric properties proposed by Candes and Tao et al that the sparse signals must maintain under the observation matrix. In the reconstruction process of the signal, in order to completely reconstruct the signal, it must be ensured that the observation matrix does not map two different K-term coefficient signals into the same sampling set, which requires that the matrix formed by the M column vectors extracted from the observation matrix is non-singular. So the key to solving such problems is how to determine the position of the non-zero coefficients to form a linear system of equations.

How to judge the RIP property of a given matrix A^{CS} is a combination complexity problem. Finding an alternative method that can easily implement RIP characteristics is the key to successfully constructing the observation matrix. If the observation matrix θ and the sparse basis Ψ are guaranteed to be uncorrelated, A^{CS} can largely satisfy the RIP property. Irrelevant means that the vector $\{\varphi_j\}$ cannot be sparsely represented by $\{\Psi_j\}$. The stronger the irrelevance, the more coefficients are needed to represent each other; on the contrary, the stronger the relevance.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

Candes E J et al. proved that when Φ is a Gaussian random matrix, the incoherence and RIP conditions can be satisfied with a large probability. Therefore, the observation matrix Φ in this paper adopts a random Gauss matrix. A random Gaussian matrix has a useful property: for a $M \times N$ random Gaussian matrix Φ, it can be shown that when $M \geq cK \log(N/K)$, $\Phi\Psi^T = A^{CS}$ satisfies the RIP property with a large probability (where c is a small constant). Therefore, a K-term sparse signal length of N can be reconstructed with high probability from M observations. The random Gaussian matrix is not related to the matrix formed by most fixed orthogonal bases. This property determines the choice that we choose it as an observation matrix. When other orthogonal bases are used as the sparse transform base, the RIP can be satisfied.

3.3. Compressed Codec Processing Research

Compressed sensing is a new way of signal acquisition that allows us to design very simple video encodings that can be implemented on mobile devices with limited resources. However, the CS-based video codec proposed by the predecessors either requires a conventional video encoder or a feedback channel, which increases the complexity of the codec. The distributed compressed video sensing (DCVS) codec scheme proposed in this paper uses CS only at the encoding end, and the decoder uses a correlation between CS measurements of adjacent frames to form a new side information generating scheme. The new DCVS block diagram we proposed is shown in Figure 2. This codec is completely CS-based and does not involve traditional video encoders. Both key and non-key frames are encoded as CS measurements and no feedback channel is required.

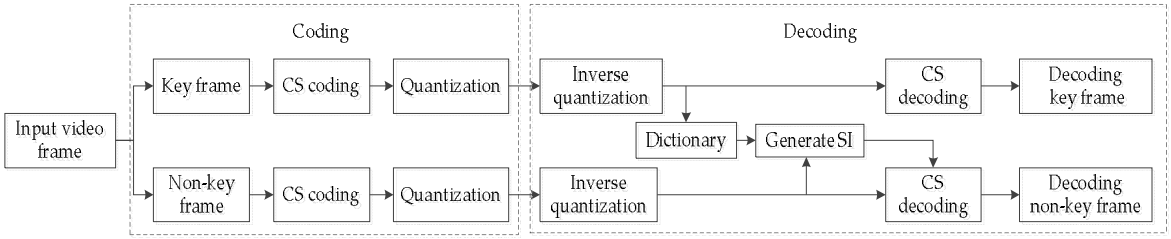

Figure 2. DCVS video codec scheme

At the encoding end, the video sequence is divided into a number of Group of Pictures (GOPs). Video frames are divided into key frames and non-key frames, non-key frames also called Wyner-Ziv (WZ) frames, which form a group of pictures. Each GOP contains one key frame and several non-key frames. Both key frames and non-key frames are coded with similar CS theory.

At the encoding end, each key frame is reconstructed by the SAMP algorithm, which redefines the l_1 minimization problem, as shown in equation (7).

$$\min_{\alpha_k} \frac{1}{2}\left\|y_k - A\alpha_k\right\|_2^2 + \tau\left\|\alpha_k\right\|_1 \quad (7)$$

Where y_k is the CS measurement of the key frame obtained at the decoding end, $A = \Phi\Psi$ has been described in the previous, $\alpha_k \in R^N \times 1$ is the sparse coefficient vector of the solution. Key frame $\hat{x}_k$ is obtained by $\hat{x}_k = \Psi\hat{\alpha}_k$, where $\hat{\alpha}_k$ is the optimal solution of α_k in equation (7).

The decoding of non-key frames are assisted by the side information generated by the dictionary and the side information is generated by the inverse quantized CS measurement of the key frame. The effect of side information is only apparent when there is sufficient correlation between the measured values of the key frame and the WZ frame. In a video sequence, adjacent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

frames in the same scene are highly correlated. Therefore, we assume that even if the CS measurement process is very different from a linear transformation (such as DCT), the CS measurement of such adjacent frames is highly correlated. On the one hand, the DCT coefficients follow a Laplacian distribution. On the other hand, CS measurements follow a more or less normal (Gaussian) distribution.

4. DCVS Framework Based on Block Classification Weighted Side Information

4.1. Description of the Framework

In the DCVS framework, in order to improve the reconstruction quality of non-key frames at the decoding end, side information is introduced when reconstructing non-key frames. Therefore, the acquisition of side information plays a very important role in the DCVS framework. Especially for the reconstruction of non-key frames, if the acquisition of side information is not accurate, the performance of its reconstruction will be greatly affected. The process of obtaining the side information may be generated by using a K-SVD training dictionary or performing motion estimation on the decoded key frame, interpolating in the time domain to generate side information. Most of the traditional methods of obtaining the side information are to estimate the forward and backward motion of the reconstructed values of the two key frames before and after the current frame, then according to a fixed weight (usually 1/2), adding to obtain the side information to assist in the reconstruction of non-key frames [17]. However, the inter-frame correlations of different regions in the video sequence are different, and the varying scenes and motion levels of the video are different. If the inter-frame correlation between the forward and backward motion estimation is weak, still using a fixed weight does not predict the current frame very well, so the accuracy of the generated side information is not high, which in turn affects the reconstruction of non-key frames. Therefore, the traditional method of synthesizing side information by using fixed weight 1/2 does not make good use of inter-frame correlation. According to the correlation between different blocks of video frames, a DCVS framework based on different blocks is proposed. The framework is shown in Figure 3.

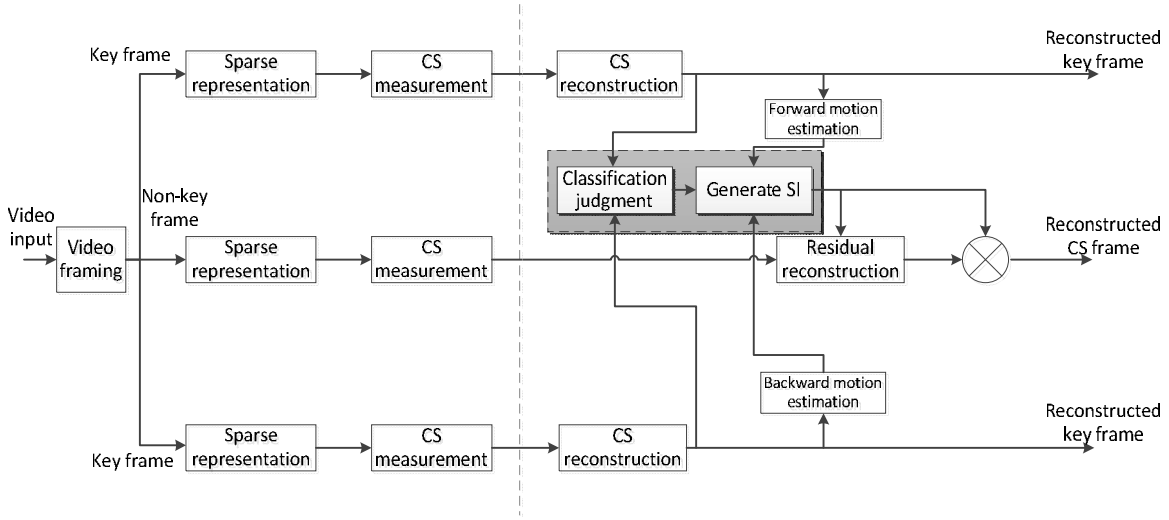

Figure 3. Block-based classification weighted side information DCVS framework

At the encoding end, the current frame and the two key frames are subjected to block sparse representation and CS sampling measurement. The key frames are sampled by the traditional DWT algorithm, the non-key frames are sampled by the hybrid sparse algorithm, and the measurement matrix still selects the random Gauss matrix. At the decoding end, the key frame is first reconstructed by SAMP algorithm, and then the classification judgment method introduced in the previous section is used to classify non-key(CS) frames, performing forward and backward motion

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

estimation on reconstructed values of adjacent key frames, different weighting schemes are used to generate side information according to different block categories. And the reconstruction of non-key frames requires the use of side information in conjunction with residual decoding of non-key frames.

4.2. Non-key Frame Reconstruction

The reconstruction of non-key frames uses the motion estimation of the classification weights to generate the side information, and then uses the side information and the measured values of the current frame to reconstruct the residual of the current frame. The reconstruction result of the non-key frame is the combination of the side information and the residual reconstruction.

Assuming that the current frame of a video is x and the predicted value of the current frame is $\tilde{x}$, the residual between the actual result and the predicted value of the current frame is $r = x - \tilde{x}$, the predicted value $\tilde{x}$ of the current frame is generated by predicting the current frame from the pre- and post-decoding key frames. Since the decoding end has not yet obtained the reconstructed data of the decoded key frames before and after the current frame, so we convert the residual to the measurement domain:

$$d = \Phi r = \Phi x - \Phi \tilde{x} = y - \Phi \tilde{x} \quad (8)$$

If the difference between the actual value and the predicted value of the current frame is smaller, the smaller the residual of the two is, the sparser it is, the smaller the sampling under the measurement matrix is, and the better the reconstruction effect of the residual is. The reconstruction result of the current frame is:

$$x_{rec} = \tilde{x} + r_{rec} \quad (9)$$

Where x_{rec} is the reconstruction result of the current frame, $\tilde{x}$ is the predicted value of the current frame, and r_{rec} is the result of the residual reconstruction.

In the acquisition of side information, it is necessary to utilize the reconstruction result of adjacent decoding key frames [18]. First, forward and backward motion estimation is performed on the reconstruction results of the *i*-th blocks x_{t-1}^{i} and x_{t+1}^{i} corresponding to the adjacent decoding key frames to obtain $\hat{x}_{t-1}^{i}$ and $\hat{x}_{t+1}^{i}$ respectively, then find the side information of the *i*-th block according to equation (10):

$$SI^{i} = \alpha \hat{x}_{t-1}^{i} + (1-\alpha)\hat{x}_{t+1}^{i} \quad (10)$$

Where α is the weight coefficient.

In the previous section, the non-key frame blocks have been classified and judged. After the blocks of different regions are classified, different weights α are adopted to obtain the side information, thereby further reconstructing the non-key frames. For the block with slight movement, the ratio of the residual energy to the energy block of the previous key frame is small, and the inter-frame correlation of the preceding and succeeding frames is large. According to equation (5), the forward motion estimation can predict the current frame very well, and then take α to a larger value. Conversely, for the block with severe movement, the inter-frame correlation is small, and the backward motion estimation can better predict the current frame, and then take α to a smaller value. For the block with big movement, the inter-frame correlation is moderate, and the current frame can be predicted by weighted averaging of forward and backward motion estimation. Based on the above analysis, the weights of the block with slight, big and severe movement are respectively set to $\alpha = \{0.7, 0.5, 0.3\}$.

In summary, for the reconstruction process of non-key frames summarized as follows: Firstly, the adjacent key frames are reconstructed, and the reconstruction results are classified and judged, and the forward and backward motion estimation is performed on adjacent key frames. Combined with the classification criteria and forward and backward motion estimation, the side information of the non-key frame is obtained according to the equation (10). According to the motion estimation, the side information and the current frame are obtained, and the residual reconstruction of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

non-key frame residual is performed. The reconstruction result of the non-key frame is the sum of the side information obtained by the motion estimation and the non-key frame residual reconstruction.

Algorithm 2: The description of the non-key frame reconstruction algorithm.

Input: $y, \Phi_s, x_{t-1}, x_{t+1}$

1. Perform forward and backward motion estimation for the pre- and post-decoding key frames x_{t-1} and x_{t+1} respectively to obtain $\hat{x}_{t-1}$ and $\hat{x}_{t+1}$;
2. According to the classification judgment result of x_{t-1} and x_{t+1}, the side information is obtained by the formula $SI = \alpha x_{t-1} + (1-\alpha)x_{t+1}$, Where α selects different weights depending on the classification;
3. Calculate the residual of the measured value y and the SI in the measurement domain: $r = y - cs\_Encoder(SI, \Phi_s)$;
4. Reconstruct the residual r to $\tilde{x}_r$;
5. Reconstruction results for non-key frames: $\tilde{x}_t = SI + \tilde{x}_r$.

Output: $\tilde{x}_t$: Reconstruction of the *t*-th block non-key frame.

5. Simulation Experiment Results and Analysis

The experiment selects two sets of video sequences collected in the underground coal mine to test the performance of the classified weighted side information method under hybrid sparse basis. In the experiment, the classification weighted side information generation method and the fixed weight side information generation method under the hybrid sparse basis are compared with the original DCVS algorithm. The reconstruction algorithm in the classification weighted side information selects the SAMP algorithm. The measurement and reconstruction of key frames and non-key frames in the video sequence are block-based. We define the first frame as the key frame. Since the key frame has a great influence on the generation and reconstruction of the side information, so we choose a key frame with a sampling rate of 0.9. The non-key frame sampling rate is selected by comparing the PSNR values reconstructed by the classified weighted side information and the fixed weighted side information at different sampling rates. Select two different video scenes, and the comparison results of PSNR values at different sampling rates are shown in Figure 4. The classification thresholds T_1 and T_2 are: T_1=0.003，T_2=0.015。In order to verify the effect of the algorithm, by comparing the time complexity of the algorithm under different sampling rates, the peak signal-to-noise ratio PSNR value is used to objectively evaluate the reconstruction effect of the algorithm. To calculate the PSNR value, the value of the mean square error MSE is first calculated. The MSE is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\|I(i,j) - K(i,j)\right\|^2 \tag{11}$$

The peak signal to noise ratio PSNR is defined as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{(2^n - 1)^2}{MSE}\right) \tag{12}$$

It is not difficult to understand that it can be obtained from equation (12) that the smaller the value of the MSE is, the closer the reconstructed video frame effect is to the original video frame image, and the larger the PSNR value is.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

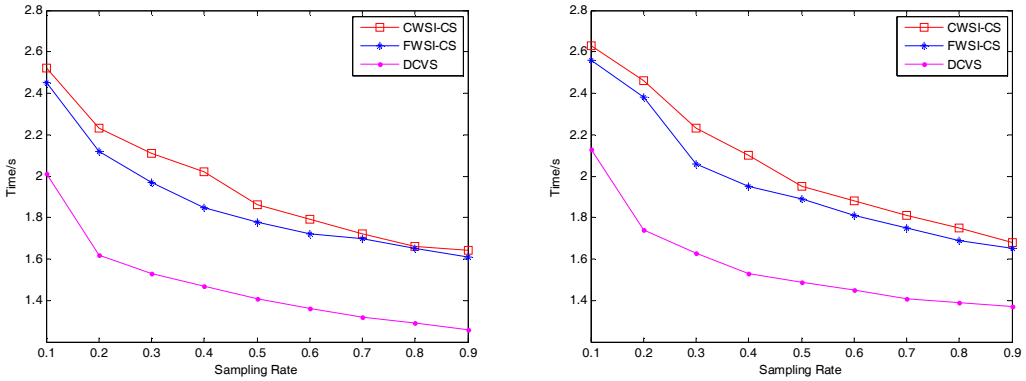

Figure 4. Relationship between sampling rate and time complexity of video sequences

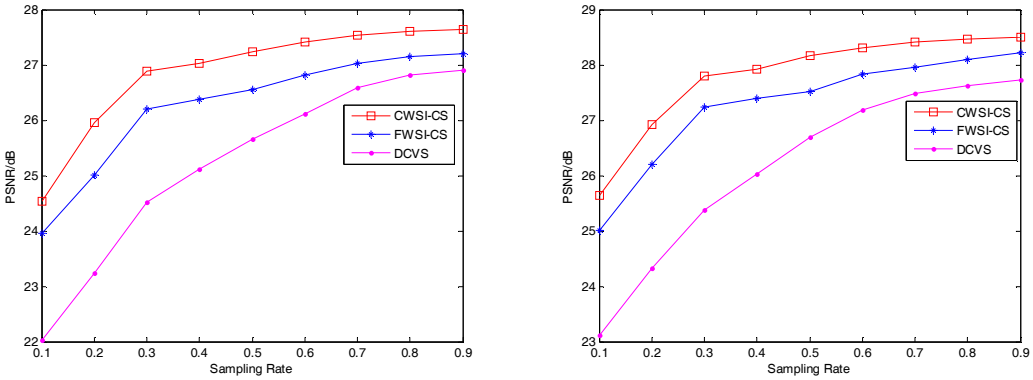

Figure 5. PSNR value of video reconstruction quality at different sampling rates

As can be seen from Figure 4, the complexity of the proposed algorithm is similar to the fixed-weighted side information algorithm at different sampling rates. The original DCVS algorithm has the shortest running time, that is, the complexity of the algorithm is lower than the former two. However, as can be seen from Figure 5, as the sampling rate increases, the PSNR value increases correspondingly. The reconstruction effect is relatively good under the condition of high sampling rate. Obviously, the proposed algorithm has the highest reconstruction quality. Although the complexity of its reconstruction is relatively high, its complexity is within an acceptable range relative to the improvement of reconstruction quality. Figure 5 shows that when the sampling rate is greater than 0.3, the increase of the PSNR value is not large as the sampling rate increases. Considering that, in the case of meeting the general needs, the low sampling rate requires less transmission data. Therefore, the sampling rate of non-key frames is selected to be 0.3, and the lesser transmission data required at low sampling rates saves energy while ensuring reconstruction quality.

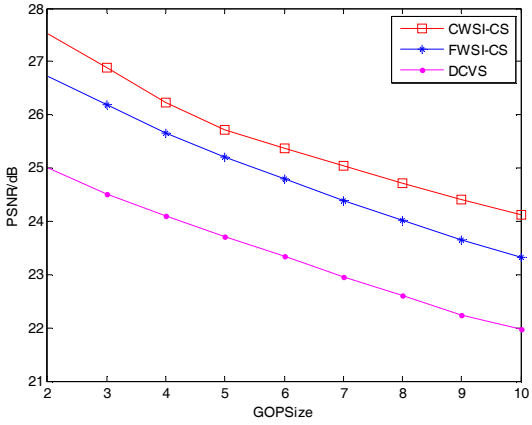

Figure 6. PSNR values of video reconstruction quality under different GOP sizes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 July 2018 doi:10.20944/preprints201807.0222.v1

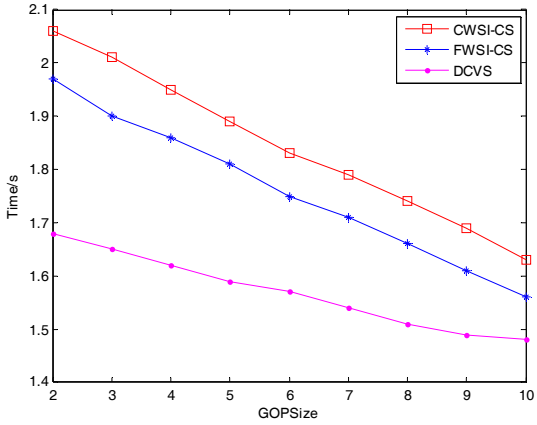

Figure 7. Time complexity under different GOP sizes

As can be seen from Figure 6 and Figure 7, as the number of frames in each GOP increases, the inter-frame correlation becomes smaller, resulting in a decrease in the reconstructed PSNR value. However, the reconstruction effect of the proposed algorithm has always been above the fixed weight side information and the original DCVS algorithm, and its reconstruction complexity is close to the original DCVS algorithm, which saves energy consumption while ensuring the reconstruction quality.

Experiment to select the 30 GOPs of the video sequence, there are 3 frames in each GOP, and a total of 90 frames are respectively simulated. The two frames before and after the GOP are taken as key frames, and the intermediate frames are non-key frames. Through sparse representation, side information generation, frame reconstruction, the quality of the two side information generation method is compared between the classification weighted side information and the fixed weight side information, and is objectively measured by the PSNR value. The curve of the reconstructed quality PSNR value of 30 frames of non-key frames in the video sequence GOP is shown in Figure 8.

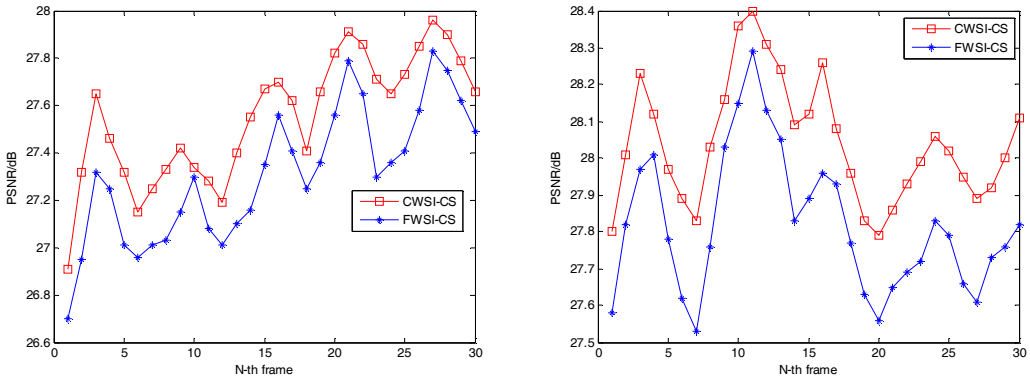

Figure 8. PSNR value of non-key frame reconstruction (sampling rate 0.3)

It can be seen from the figure that under the same conditions, the PSNR value of the block-based classification adaptive sparse basis and the classification weighted side information is improved by 0.2-0.8dB compared with the fixed weight side information reconstruction. For image blocks with too high movement intensity, the correlation between adjacent frames is weak, and the reconstruction effect is relatively weak, but the reconstruction effect is still improved compared with the fixed weight side information. Therefore, the proposed algorithm can improve the quality of video reconstruction in different video scenarios.

6. Conclusions

In order to solve the problem that the correlation between the different video sequence image blocks is different in the process of applying the compressed sensing method to the video decoding side to acquire the side information for assisting non-key frame reconstruction, a block-based

classification weighting method is proposed for non-key frame reconstruction. In terms of the sparse representation of the encoding end, the video sequence is divided into different blocks according to the inter-frame correlation, and different sparse basis strategies are selected for different blocks; In the decoding end of the non-key frame, in the algorithm for generating the side information by motion estimation, the generation of the SI selects different weights according to the difference of the inter-frame correlation. The experimental results show that the scheme can adaptively select different sparse basis and side information generation schemes to assist non-key frame reconstruction according to different video scenes, making full use of inter-frame correlation and improving reconstruction quality.

Author Contributions: Conceptualization, K.W. and M.W.; Methodology, R.X.; Software, R.X.; Validation, R.X., K.W. and M.W.; Formal Analysis, R.X.; Resources, K.W.; Writing-Original Draft Preparation, R.X.; Writing-Review & Editing, H.X.; Visualization, R.X.; Supervision, H.X.; Project Administration, H.X.; Funding Acquisition, H.X.

Funding: This work is supported by the National Key Research and Development Project of China (No.2017YFC0804404).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhao, X.H.; Deng, Y.F.; Mu, D.C. Applied study on compressed sensing technology to mine Internet of things. *Coal Science and Technology*, **2016**, 44(7), 69-72[CrossRef].
2. Zhao X.H.; Liu S.S.; Shen X.R.; et al. Micro-seismic data compression and reconstruction based on distributed compressed sensing. *Journal of China University of Mining & Technology*, **2018**, 1, 172-182[CrossRef].
3. Zhang F.; Yan X.X.; Li Y.J. A novel image reconstruction method of mine intelligent surveillance based on adaptive sparse representation. *Journal of China Coal Society*, **2017**, 42(5), 1346-1354[CrossRef].
4. Zhang F.; Yan X.X. The block compressed sensing of mine monitoring images based on DFT basis. *Journal of Transduction Technology*, **2017**, 30(1), 94-100[CrossRef].
5. Wang G.; Zhao Z.K.; Ning Y.J. Design of Compressed Sensing Algorithm for Coal Mine IoT Moving Measurement Data Based on a Multi-Hop Network and Total Variation. *SENSORS*, **2018**, 18(6),1732[CrossRef]
6. Lin B.L.; Zheng B.Y.; Qian C. The reconstruction methods of frames classification of distributed video compression coding. *Signal Processing*, **2015**, 2, 201-207[CrossRef].
7. Zhao X.H.; Liu S.S.; Shen X.R.; et al. Research on processing algorithm of image in underground coal mine based on CS framework. *Coal Science and Technology*, **2018**, 46(2), 219-224[CrossRef].
8. Zhang R. A preliminary study of image reconstruction and denoising based on compressed sensing. Southwest Jiaotong University, **2010**[CrossRef].
9. Wu M.H.; Zhu X.C. Dynamic measurement rate allocation for distributed compressive video sensing. *Journal of Nanjing University of Posts and Telecommunications*, **2013**, 33(1), 62-67[CrossRef].
10. Dong H.; Zhuang B.; Su F.; et al. A novel distributed compressive video sensing based on hybrid sparse basis. // Visual Communications and Image Processing Conference. IEEE, **2014**, 320-323[CrossRef].
11. Chen H.W.; Kang L.W.; Lu C.S. Dictionary learning-based distributed compressive video sensing.// *Picture Coding Symposium*. **2011**, 210-213[CrossRef].
12. Aharon M.; Elad M.; Bruckstein A. rmK-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, **2006**, 54(11), 4311-4322[CrossRef].
13. Prades N.J.; Ma Y.; Huang T. Distributed video coding using compressive sampling.// *Picture Coding Symposium. IEEE*, **2009**, 1-4[CrossRef].
14. Liu L.; Wang A.; Li Z.; et al. An Improved Distributed Compressive Video Sensing Based on Adaptive Sparse Basis. // *International Conference on Robot. IEEE*, **2011**, 137-140[CrossRef].
15. Baig Y.; Lai E.M.K.; Punchihewa A. Distributed Video Coding Based on Compressed Sensing.//*IEEE International Conference on Multimedia and Expo Workshops. IEEE*, **2012**, 131(5), 325-330[CrossRef].
16. Baraniuk R.; Davenport M.; Devore R.; et al. A Simple Proof of the Restricted Isometry Property for Random Matrices.// *CONSTR APPROX*, **2008**, 253-263[CrossRef].

17. Dai Y.Y.; Cao X.Q.; Chen R.; et al. Reconstruction algorithm with classified weighted side Information for distributed video compressive sensing. *Computer Technology and Development*, **2017**, 27(5), 87-91[CrossRef].
18. Jian C.; Su K.X.; Wang W.X.; et al. Residual Distributed Compressive Video Sensing Based on Double Side Information. *Acta Automatica Sinica*, **2014**, 40(10), 2316-2323[CrossRef].