

A Fast Quality Scalable Video Coding Method Based on Compressed Sensing

Min Sun¹, Dong Hu^{2,*} and Jianyu Ding³

¹Education Ministry's Key Lab of Broadband Wireless Communication and Sensor Network Technology

²Education Ministry's Engineering Research Center of Ubiquitous Network and Health Service

³Jiangsu Province's Key Lab of Image Procession and Image Communications, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

*Corresponding author

Abstract—This paper presents a fast quality scalable video coding method based on compressed sensing(CS). The proposed method obtained the coding scheme of the enhancement MJU by using the interlayer and spatial correlation and kept the base layer's coding scheme unchanged. And the part in the enhancement layer which needed to be fine quantified was combined with the compressed sensing theory selectively which based on the sparsity of the signal and the complexity of the reconstruction. In order to satisfy the coding syntax structure of the reference software, the measurement value got by compressed sensing was complemented by 0s and the flag bit was set to distinguish the special sub-blocks coded by CS. Experimental results show that the proposed algorithm can effectively improve the efficiency of scalable video coding and reduce the computational complexity.

Keywords—video coding; compressed sensing; quality scalable video coding

I. INTRODUCTION

Different from the traditional video coding methods, scalable video coding can not only provide video services to different end users, but also can automatically choose different coding methods according to different terminal scenarios and requirements. In the earlier studies, researchers focused on scalable video coding fast algorithms that aimed to reduced its coding complexity. Seon-Tae Kim [1] et al. obtained a weighted model and proposed a fast mode decision algorithm for coarse-grained quality scalability based on correlation between layers. Yu Che [2] et al. proposed a probabilistic model predicting the probability of motion vectors and non-zero coefficients for all scalable video coding. Although researchers have reduced the computational complexity of scalable video coding to a certain extent, they have not really solved the problem of huge resource consumption caused by enhancement layer of scalable video coding.

In recent years, the CS has been a natural tool to reduce the number of samples, then more and more researchers began to take the advantages of CS to solve the problems in scalable video coding. For example, in [3] the authors proposed a scheme for compressed sampling that exploits local sparsity within a frame applied to wireless network environments. The coding framework improves the coding efficiency by abandoning motion estimation and motion compensation, and

only read key frames for reference to reduce the complexity. In [4] Vladimir Stankovic and others proposed a scalable coding framework that improved the base layer coding quality by changing the measurement matrix in compressed sensing, improving the quality of the base layer video image. Compared with previous fast algorithms which decide the coding scheme depending on statistical information, quality scalable video coding based on compressed sensing can improve the performance of scalable video coding more meticulously and effectively. The focus of this paper is to apply sparsity theory to quality scalable video coding and combine compressed sensing with a fast quality scalable coding framework.

The rest of the paper is organized as follows. Section II gives a brief introduction of compressed sensing. Section III describes the architecture of fast quality scalable video coding based on CS. The performance of fast quality scalable video coding based on CS is studied in section IV, followed by concluding remarks in section V.

II. BACKGROUND

Compressed sensing or sampling [5][6] was proposed as a new acquisition framework which can sample and compress sparsity or compressible signals in a single operation. Suppose that a signal $x \in R^n$ can be transformed to a coefficient vector θ with some basis Ψ , $x = \Psi\theta$. Ψ can be any representing basis such as DCT or wavelet [7]. The measurements of compressed sensing, $y = R^m$, are obtained by multiplying signal x with a measurement matrix $\Phi \in R^{m \times n}$, i.e., $y = \Phi x$. Since $m < n$, (1) is an underdetermined system with infinite solutions. Using the reverse operation in (1) to recover x is infeasible. Finally, complete content and organizational editing before formatting [8]. Please take note of the following items when proofreading spelling and grammar:

$$y = \Phi x \quad (1)$$

To make recovery stable and accurate, sensing matrix Φ must satisfy the restricted isometry property (RIP). Reference [6] showed the methods of generating sensing matrix holding

RIP. One of them is to randomly select m rows from the Fourier matrix or Gaussian random matrix.

When the original signal x is measured to form the measurement value y , it is necessary to reconstruct original signal in the decode side. OMP reconstruction algorithm is adopted in this paper [9]. The algorithm first performs Schmidt orthogonalization for selected atoms, and then the signal is projected on the space of the atoms that are processed by orthogonalization, and finally the projection components of the signal in the selected atomic space are calculated. The advantage of this method is that it can guarantee that the residuals used in each iteration after schmidt's orthogonalization process are orthogonal to the selected atoms, so it can effectively speed up the convergence speed of the algorithm and improve the efficiency of the algorithm [10].

III. PROPOSED VIDEO CODING ALGORITHM

In this section, we describe the proposed fast quality scalable video coding scheme based on the compressed sensing. Our work is divided into two parts. The first is to analyze the correlation between the base layer and the enhancement layer, then use the sub-blocks that have been coded in the base layer to obtain sub-blocks of the enhancement layer quickly. The second part combines sub-blocks with compressed sensing theory to achieve the effect of reducing the coding rate of the encoder.

A. Correlation Analysis

1) Spatial correlation analysis:

Spatial correlation means that neighboring macroblocks have a high degree of similarity in their motion characteristics and structural characteristics, which leads to a very high correlation between their coding modes. When the spatial correlation is used for prediction, the correlation degree between the current macroblock and adjacent macroblocks, that is probabilistic meaning between them, must be obtained primarily. The spatial correlation of the enhancement layer can then be used to judge the spatial correlation of the enhancement layer. When the correlation between the neighboring blocks of the base layer is greater, the probability that the neighboring blocks of the enhancement layer at the same position use the same pattern is greater. These correlations are not only manifested in the type of macroblocks, but also in the spatial position of macroblocks. Spatial correlation can provide an effective basis for the next fast mode decision.

2) Inter-layer correlation:

In quality scalable video coding, the base layer and the enhancement layer have the same resolution and frame rate, only the quantization steps are different, so the base layer and enhancement layer are similar in coded macroblocks, namely inter-layer correlation. To illustrate the correlation between the base layer and the enhancement layer clearly, the experiment code different sizes test sequences and calculate the degree of correlation between the base layer and enhancement layer in case of 16×16 , 16×8 , 8×16 , 8×8 , and the SKIP mode respectively. This experiment is performed in the H.264[11] standard code reference software JSVM 9.18. The relevant

parameters are configured as follows: the quantization parameters of the base layer and the enhancement layer are 24 and 32 respectively; the GOP size is 4, the video sequences size are 1280×720 , the experiment results are shown in the Table I.

Table I shows results of the Inter-layer correlation between base layer and enhancement layer in 1280×720 video sequence. We can see that the best coding mode of enhancement layer has a certain correlation with the base layer coding mode.

TABLE I. CORRELATION BETWEEN BASE LAYER AND ENHANCEMENT LAYER IN 1280×720 VIDEO SEQUENCE (%)

Video Sequence	EL BL	Conditional Probability Distribution				
		SKIP	16×16	16×8	8×16	8×8
PartyScene	SKIP	0.701	0.124	0.033	0.042	0.089
	16×16	0.262	0.576	0.045	0.051	0.056
	16×8	0.102	0.193	0.388	0.038	0.279
	8×16	0.120	0.206	0.075	0.310	0.288
	8×8	0.123	0.151	0.170	0.168	0.388
ParkRunter	SKIP	0.679	0.157	0.025	0.023	0.116
	16×16	0.271	0.512	0.069	0.063	0.054
	16×8	0.121	0.135	0.377	0.113	0.254
	8×16	0.148	0.178	0.062	0.352	0.260
	8×8	0.159	0.167	0.183	0.126	0.365
ParkJoy	SKIP	0.628	0.166	0.039	0.060	0.107
	16×16	0.259	0.561	0.064	0.057	0.059
	16×8	0.142	0.158	0.363	0.060	0.287
	8×16	0.114	0.194	0.100	0.346	0.276
	8×8	0.151	0.177	0.167	0.152	0.358

B. Fast Mode Selection Model

To accelerate the selection of enhancement layer coding modes and reduce coding complexity effectively, we proposed early termination conditions for spatial and inter-layer correlation. Since the macroblocks content are approximately similar in the same position of base layer and enhancement layer, it also means that the different of quantization steps does not play an important role in the mode selection. Therefore, the early termination condition between layers can be estimated by:

$$z_e - z_b = k_1 \quad (2)$$

where z_e and z_b are the quantization coefficients of the base layer and the enhancement layer, respectively, and k_1 is threshold from experiment. Formula (2) can be written as:

$$r_e \leq Q_e r_b / Q_b + k_1 Q_e \quad (3)$$

where Q_b , Q_e are the quantization steps of base layer and enhancement layer respectively, r_b and r_e are DCT coefficients of base layer and enhancement layer respectively. Since the DCT coefficients calculation can be described as:

$$r = \sum_{u=0}^3 \sum_{v=0}^3 d_{iu} x_{uv} d_{jv} \quad (4)$$

where d_{iu} is the value of (i, u) in the integer DCT transform, x_{uv} is the residual signal value, since d_{jv} is less than $\sqrt{\frac{2}{5}}$, so

$$|r| \leq \frac{2}{5} \sum_{u=0}^3 \sum_{v=0}^3 |x_{uv}| \quad (5)$$

combine (2) and (5), that can describe as:

$$\sum_{u=0}^3 \sum_{v=0}^3 |x_{uv}^e| \leq Q_{estep} \sum_{u=0}^3 \sum_{v=0}^3 |x_{uv}^b| / Q_{bstep} + \frac{5}{2} k_1 Q_{estep} \quad (6)$$

When condition

$$RD_e \leq Q_e * RD_b / Q_b + 40k_1 Q_e \quad (7)$$

is satisfied, the selection of the inter-layer decision mode ends, where RD is the rate-distortion cost, which is the constant obtained from experiments, and the best is 2.43. According to the spatial correlation of DCT quantized coefficients, similar condition of early termination of spatial correlation is proposed as:

$$RD_1 - RD_2 \leq Q_{estep} * (RD_3 - RD_4) / Q_{bstep} + \frac{5}{2} k_2 Q_{estep} \quad (8)$$

When spatial correlation meets formula (8), mode selection ends.

C. Fast Quality Scalable Coding Algorithm Based on CS

When we combined inter-frame coding with CS, the signal meets the sparsity requirements so we can compress the signal in a high rate. In scalable video coding, signal of the enhancement layer, especially the enhancement layer of the inter-frame coding, is a residual signal with a high degree of sparsity, so it naturally meets the sparsity demand. Therefore, the sub-block selected by the quick mode combined with CS can improve the efficiency of the encoding.

When compressed sensing is combined with inter-frame coding, it has a high compression rate because it can achieve the sparsity requirement of the signal. In scalable video coding, the enhancement layer especially enhancement inter-frame coding layer, the residual signals of the signal can satisfy the requirement of the signal sparsity. Therefore, it can reduce the efficiency of coding by combining the compressed sensing with the sub-block selected by fast mode. When the enhancement layer is sparse, it does bring down the reduction of the coding end code of the scalable video encoding, which improves the coding efficiency greatly.

However, due to the sparsity representation of all enhancement layer sub-block modes, the computational complexity of the reconstruction algorithm will be very high. The next step is to do a detailed analysis of computational complexity for the different macroblocks with sparsity representation. Since the 4×4 block have not met the requirements of sparsity, there is no prerequisite for measuring the sparsity. Therefore, only 8×8 blocks and 16×16 blocks are used for reference comparison. The experimental performance reference indexes are $\Delta Bitrate$, $\Delta PSNR$, $\Delta Time$, the calculation formula is:

$$\Delta Bitrate = \frac{Bitrate_8 - Bitrate_{16}}{Bitrate_8} \times 100\% \quad (9)$$

$$\Delta PSNR = PSNR_8 - PSNR_{16} \quad (10)$$

$$\Delta Time = \frac{Time_8 - Time_{16}}{Time_8} \times 100\% \quad (11)$$

The experimental comparison results are shown in the Table II.

TABLE II. PERFORMANCE COMPARISON OF 8×8 MACROBLOCK AND 16×16 MACROBLOCK

Video Sequence	QP (BL, EL)	ΔBitrate (%)	ΔPSNR (dB)	ΔTime (%)
PartyScene	(30,22)	9.33	0.049	-279
	(30,24)	7.05	0.047	-242
	(30,26)	8.52	0.037	-255
ParkRunter	(30,22)	6.41	0.077	-303
	(30,24)	7.56	0.078	-281
	(30,26)	6.42	0.082	-289
ParkJoy	(30,22)	9.21	0.056	-271
	(30,24)	8.15	0.078	-254
	(30,26)	9.26	0.091	-317
Stockholmter	(30,22)	5.88	0.084	-286
	(30,24)	6.97	0.048	-249
	(30,26)	7.18	0.062	-236
Intotree	(30,22)	5.92	0.051	-265
	(30,24)	7.10	0.072	-325
	(30,26)	8.04	0.090	-216

From the Table II we can see that when 16×16 macroblocks are combined with compressed sensing, the complexity of its reconstruction algorithm is about 3 times of 8×8 blocks, and its code rate is not significantly improved. Therefore, in this paper, only 8×8 blocks are sparsely represented in the quality scalable fast coding method, it can not only provide reliable guarantee for the accuracy of encoding, but also alleviate the sharp increase of coding algorithm complexity.

Since the sparsity and measurement matrices used in the compressed sensing process need to satisfy the principle of RIP, DCT or DWT is often used as a sparse matrix and Gaussian random matrices or Bernoulli matrices are used as measurement matrices in actual experiments. Integer DCT transform has been used in the reference software JSVM of the H.264 standard video coding. Therefore, this algorithm chooses integer DCT as the sparse matrix and uses Gaussian random matrix as the measurement matrix. The specific steps of the algorithm are as follows:

Step 1: Initialize the parameters: (1) Generate a 64×64 Gaussian random matrix using Gaussian random function. (2) Set the scalable video coding layer to 2, one is base layer and the other is enhancement layer.

Step 2: Judge whether the current frame code is base layer code, and if so, keep the original code mode; if not, obtain the block mode of the enhancement layer by the fast mode selecting condition.

Step 3: Judge whether the current enhancement layer sub-block contains the flag transform_size_8×8_flag. If not, perform fine-quantization coding according to the original mode.

Step 4: After the 8×8 sub-blocks are quantized and then sparsely coded as follows:

(1) Sparse the 8×8 sub-blocks using an integer DCT, then transform the 8×8 sparse block to a 64×1 sparse vector N .

(2) Select a Gaussian random matrix of size $m \times N$ as the measurement matrix for the CS according to the sparsity k of the vector N . Multiply the sparse vector N with a Gaussian random matrix ϕ to obtain a measurement M of size $m \times 1$.

(3) In order to satisfy the coding syntax structure of the reference software, the measurement value M is complemented by 0s and the flag bit F_m is set, and then entropy coding is performed.

Step 5: Judge whether the block to be decoded has a flag bit at the decoder side. If not, perform the normal decoding step; otherwise, use the Φ and m to calculate M and ϕ , then reconstruct the original signal using the Orthogonal Matching Pursuit (OMP) algorithm.

IV. EXPERIMENT

In this section, we carry out experiments on fast quality scalable fast coding architecture based on CS follow the above steps in section III. The test sequences are ParkScene, ParkRunter, ParkJoy, Stockholmter and Intotree. The experiments are performed on the standard reference software JSVM9.18 of H.264. The relevant configuration is as follows:

- 1) Number of coded frames: 200;
- 2) GOP size: 4;
- 3) Video resolution: base layer: 1280×720, enhancement layer: 1280×720;
- 4) Quantization step: (30, 22) (30, 24) (30, 26) (30, 28) and (38, 30) (38, 32) (38, 34) (38, 36).

The Quantization Parameter(QP) of the first set of experiments are (30,22) (30,24) (30,26) (30,28). Table III shows the results of the first set of experiments. The QP of the second set of experiments are (38,30) (38,32) (38,34) (38,36). Table IV shows the results of the second set of experiments. In addition, Figure I shows the comparison between coding and decoding image of the video sequences PartyScene and ParkRunter.

The following indexes are used to evaluate the algorithm experiments results presented in this paper:

$\Delta PSNR$ indicates the increase in Peak Signal to Noise Ratio(PSNR). The formula is as follows:

$$\Delta PSNR = PSNR_{JSVM} - PSNR_{proposed} \quad (12)$$

where in, the $PSNR_{JSVM}$ obtained from the standard reference software JSVM9.18 of H.264, $PSNR_{proposed}$

TABLE III. EXPERIMENTAL RESULTS OF FAST QUALITY SCALABLE ALGORITHM BASED ON CS (GROUP 1)

Video Sequence	QP(BL,E L)	$\Delta Bitrate$ /%	$\Delta PSNR$ /dB
PartyScene	(30,22)	29.87	0.018
	(30,24)	31.04	0.055
	(30,26)	26.39	0.049
	(30,28)	27.15	0.031
ParkRunter	(30,22)	25.83	0.064
	(30,24)	27.40	0.053
	(30,26)	32.05	0.039
	(30,28)	27.26	0.076
ParkJoy	(30,22)	31.58	0.051
	(30,24)	28.36	0.087
	(30,26)	29.28	0.036
	(30,28)	25.10	0.047
Stockholmter	(30,22)	27.02	0.086
	(30,24)	26.15	0.040
	(30,26)	31.36	0.068
	(30,28)	33.47	0.070
Intotree	(30,22)	25.21	0.063
	(30,24)	27.57	0.073
	(30,26)	28.88	0.057
	(30,28)	34.64	0.063

TABLE IV. EXPERIMENTAL RESULTS OF FAST QUALITY SCALABLE ALGORITHM BASED ON CS (GROUP 2)

Video Sequence	QP(BL,EL)	$\Delta Bitrate$ /%	$\Delta PSNR$ /dB
PartyScene	(38,30)	25.42	0.027
	(38,32)	25.16	0.071
	(38,34)	26.07	0.034
	(38,36)	25.95	0.049
ParkRunter	(38,30)	27.92	0.963
	(38,32)	33.13	0.080
	(38,34)	30.05	0.017
	(38,36)	34.21	0.024
ParkJoy	(38,30)	27.02	0.050
	(38,32)	23.38	0.066
	(38,34)	29.28	0.053
	(38,36)	30.61	0.058
Stockholmter	(38,30)	22.26	0.067
	(38,32)	28.27	0.029
	(38,34)	31.79	0.095
	(38,36)	30.04	0.072
Intotree	(38,30)	22.63	0.081
	(38,32)	27.57	0.052
	(38,34)	28.88	0.079
	(38,36)	28.64	0.069

represents the PSNR obtained by this algorithm.

$\Delta Bitrate$ indicates the reduction of the enhancement layer bitrate. The formula is as follows:

$$\Delta Bitrate = \frac{Bitrate_{JSMV} - Bitrate_{proposed}}{Bitrate_{JSMV}} \times 100\% \quad (13)$$

where $Bitrate_{JSMV}$ represents the encoding bit rate of the standard reference software JSVM9.18 of H.264, and $Bitrate_{proposed}$ represents the encoded bit rate of the proposed algorithm. Our presented results reduce the rate bit than reference JSVM9.18.

The effects on PSNR and bit rate are given in Table III and Table IV, which show the increasement of PSNR and reduction of Bitrate. And our algorithm can keep the reduction of PSNR in an acceptable range and improve the bit rate obviously. Figure I is the coding images and decoding images of two video sequences, we can see that our algorithm keeps the image quality unchanged.



(A) PARTYSCENE CODING IMAGE



(B) PARTYSCENE DECODING IMAGE



(C) PARKRUNTER CODING IMAGE



(D) PARKRUNTER DECODING IMAGE

FIGURE I. FAST QUALITY SCALABLE VIDEO CODING BASED ON CS CODING AND DECODING IMAGE COMPARISON

V. CONCLUSION

In this paper, we proposed a low-complexity, fast quality scalable video coding architecture based on compressed

sensing. As the proposed algorithm significantly achieves good scalability and reduces the coding rate of enhancement layer, it is more effective and efficient than the traditional solution of video coding. While fully exploiting the inter-layer and spatial correlation and residual sparsity, it also guarantees the quality of video coding. At the same time, the operation of zero-filling for measured value also makes the method satisfy the standard coding syntax so it is practical in current video coding standard reference software.

REFERENCES

- [1] Seon-Tae Kim, Krishna Reddy Konda, ChunSu Park, ChangSik Cho, SungJea Ko, Fast Mode Decision Algorithm for Inter-Layer Coding in Scalable Video Coding, *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, 1572-1580, 2009.
- [2] Yu Che Wey, MeiJuan Chen, ChiaHung Yen, ChiaYen Chen, Fast Mode Decision Algorithm for Scalable Video Coding Based on Probabilistic Models[J], *Information and Engineering*, vol. 32, no. 5, 931-945, 2016
- [3] Siyuan Xiang, Lin Cai, Scalable Video Coding with Compressive Sensing for Wireless Videocast, *International Conference on Communication(ICC)*, IEEE, 1-5, 2011
- [4] Vladimir Stankovic, Lina Stankovic, Scalable compressive video, In *Image Processing(ICIP)*, IEEE, 921-924, 2011
- [5] Donoho D L. Compressed sensing[J]. *IEEE Transactions on Information Theory*, 2006, 52(4):1289-1306.
- [6] Candès E J. Compressive Sampling[J]. *Marta Sanz Solé*, 2006, 17(2):. 1433-1452.
- [7] J. Chen, J. Boyce, Y. Yan and M. M. Hannuksela, G. J. Sullivan, Y-K. Wang. Scalable High Efficiency Video Coding Draft 7, JCTVC-R1008, 18th JCTVC Meeting, Sapporo, JP, 2014.
- [8] S. N. Karishma, B. K. N. Srinivasarao, Indrajit Chakrabarti, Compressive Sensing based Scalable Video Coding for Space Applications, In *National Conference on Communication(NCC)*, IEEE, 1-6, 2016..
- [9] Hong Jiang, Chengbo Li, Raziel Haimeir-Cohen, Paul A. Wilford, Yin Zhang, Scalable Video Coding Using Compressive Sensing[J], *Bell Labs Technical Journal*, vol.16, no. 4, 149-170, 2012..
- [10] Zhang Y, Mei S, Chen Q, et al. A novel image/video coding method based on Compressed Sensing theory[C]. *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008:1361-1364..
- [11] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra. Overview of the H.264/AVC video coding standard, *IEEE Trans. Circuits Syst.Video Technol*, 2003, 13(7): 560-576.