Temporal Redundancy Reduction in Compressive Video Sensing by using Moving Detection and Inter-Coding

Jirayu Peetakul*, Jinjia Zhou*+

*Graduate School of Science and Engineering, Hosei University

+JST, PRESTO Tokyo, Japan

INTRODUCTION

Compressive sensing (CS) is a simultaneously signal acquisition and compression technique for efficiently acquiring and reconstructing a signal from a small number of measurements, which can be obtained by linear projections onto sparse signal [1]. In order to further compress the measurements, many works applied intra prediction-based measurement coding. In this paper, we proposed temporal redundancy reduction in compressive video sensing by using moving detection and inter-coding. The experimental results show that our proposed can greatly reduce bandwidth usage in terms of BPP by 63.15%, improve in PSNR by 1.56dB, and SSIM by 14.81% on average when compared to the state-of-the-art.

PROPOSED ARCHITECTURE

We divided the type of measurement into two portions: static measurement as non-moving part and dynamic measurement as moving part in pixel-domain. In general, an information of consecutive frames are similar, resulting in temporal redundancy. We used moving detection and inter-coding to further reduce temporal redundancy and bandwidth usage in compressive video sensing as shown in Fig. 1.

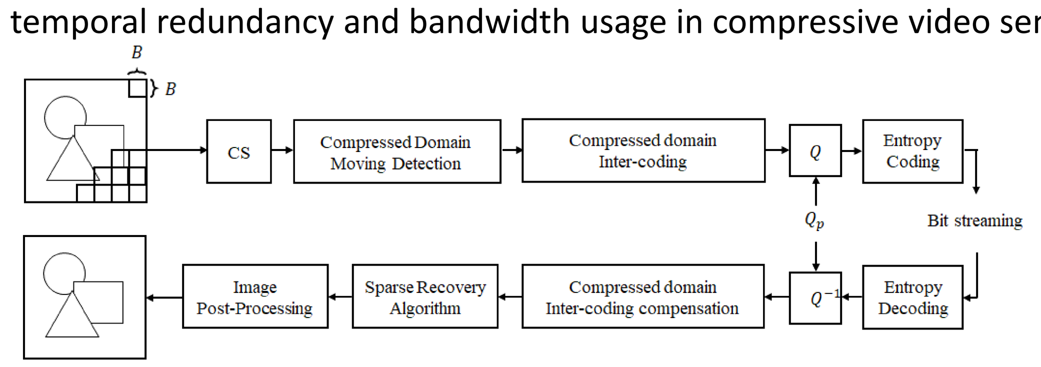

Figure 1. Proposed measurement coding architecture of temporal redundancy reduction in compressive video sensing by using moving detection and inter-coding.

To implement motion detection, we estimated the difference between current block of y_t and reference block y_{ref} in co-location via mean squared errors (MSE). In order to account for variations in measurement, local adaptive threshold has become a primary method to classify the distinction of measurement.

We used local adaptive threshold to classify the measurement with an association of error distinction, where threshold levels are chosen automatically and independently. It is allowed framework to operate without difficulty to non-uniform of illumination. We assumed that y_t is static measurement, which is necessary to be omitted. Otherwise, the change of y_t is represented to illumination change or moving objects in pixel-domain, which is necessary to further compress.

False positive detection can be occurred randomly during the process that caused BPP increasing uncertainty and unpredictable. Therefore, we proposed straightforward algorithm that adjusts the quantization parameter (Q_p) to improve quality in subjective important area and eliminate false positive detection automatically. The parameter can be varying depend on how frequently the area become detected. If the specific area of interest is frequently detected, we assumed that the area should be paid more attention, where adaptive quantization parameter by $q_p \in \{2, 4, 6, 8\}$ and parameter step size denoted by q_{ss} = 2.

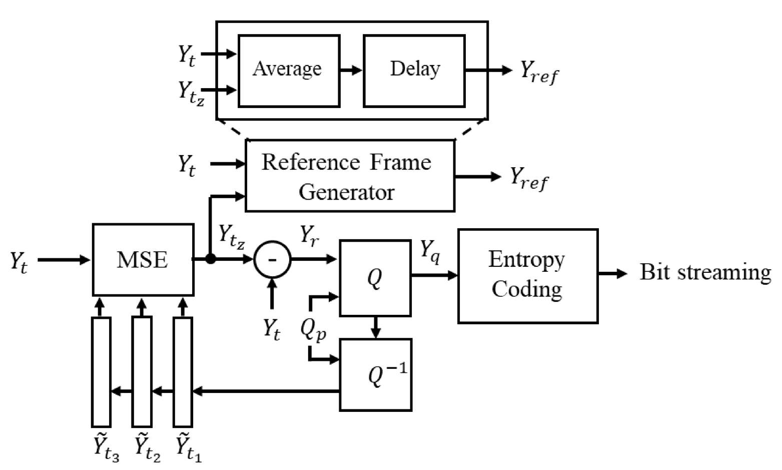

Figure 2. Proposed of compressed-domain inter-coding architecture.

By directly transfer an entire of detected block can cause block artifact due to smoothness constrained between the neighboring blocks when the frame is fully reconstructed. Therefore, we implemented compressed domain inter coding by storing three neighboring frames in first-in-first-out (FIFO) memory and finding the best matching blocks of y_t through the candidates in FIFO denoted by $\hat{y}_{t_1}$, $\hat{y}_{t_2}$, and $\hat{y}_{t_3}$ by using MSE.

The selected candidate denoted by $\hat{y}_{t_z}$, where z is selected candidate position in FIFO. Hence, it is required to transfer to the decoder. We further compressed and increased smoothness constrained between the neighboring blocks by subtracting y_t with $\hat{y}_{t_z}$, resulting in residual of motion denoted by y_r.

To update the new value to FIFO, it is necessary to perform quantization and de-quantization to y_t for ensuring the similarity of information between encoder and decoder. An average of y_t and $\hat{y}_{t_z}$ will turn into reference for the next frame comparison denoted by y_{ref}. The local adaptive threshold levels can be updated automatically by average of summation between y_t and y_{ref} denoted by T_h. To further reduce bandwidth usage and packet length, we performed 4-bit quantization to perform left and right shift on bit patterns to y_r, where q_p is adaptively selected. The quantization results denoted by y_q. Consequence, we reduced bits length for data streaming needed to stream and store a packet by using Huffman coding as entropy coding before transferring to the communication channel as shown in Fig. 2.

SIMULATION RESULTS

Figure 3. Simulation results of proposed temporal redundancy reduction with crowded of people by using VIRAT [6], AVL-Town Centre [7], and WILDTRACK [8] sequences, respectively. The first row is the ground truth. The second row is residual results of moving detection in pixel-domain. The third row is residual motion in pixel-domain. The fourth row is fully reconstruction results. The fifth row is crop results to show the remaining details, where B = 16×16, $q_p \in \{2, 4, 6, 8\}$, q_{ss} = 2, and SR = 1⁄2.

Figure 4. The visual comparison among three methods with our proposed using WILDTRACK [8] sequence. The top row is the original scene. The second row is residual motion in pixel-domain. The third row is fully reconstruction results in pixel-domain. The fourth row is cropped and zoomed results for comparing the remaining details, where B = 16×16, q_p =4, and SR = 1⁄2

SIMULATION RESULTS

Table 1. The average PSNR (dB), SSIM, and BPP comparisons of 100 frames, where B = 16×16, $q_p \in \{2, 4, 6, 8\}$, q_{ss} = 2, and SR = 1⁄2, 1⁄4, and 1⁄8.

Sequences	PSNR			SSIM			BPP		
	1⁄2	1⁄4	1⁄8	1⁄2	1⁄4	1⁄8	1⁄2	1⁄4	1⁄8
VIRAT	38.21	36.11	34.29	0.92	0.86	0.76	0.10	0.04	0.01
AVL	41.07	38.17	35.80	0.96	0.91	0.84	0.24	0.10	0.03
WILDTRACK	36.94	34.94	33.22	0.93	0.85	0.74	0.98	0.40	0.16
Average	38.74	36.40	34.43	0.93	0.87	0.78	0.44	0.18	0.06

Table 2. An average PSNR (dB), SSIM, and BPP comparison of 100 frames of WILDTRACK [8] sequence.

	[2]	**[3]**	**[4]**	**This work**
Sub-block size	16×16	16×16	16×16	16×16
Sampling ratio	1⁄2	1⁄2	1⁄2	1⁄2
Measurement matrix	Checker-board	Modified random	Walsh-Hadamard	Walsh-Hadamard
Coding Method	Intra	Intra	Intra	MD + Inter
Q_p	4	4	4	variable
Recovery algorithm	ℓ_1-min. + IDCT	ℓ_1-min. + IDCT	ℓ_1-min. + IFWHT	ℓ_1-min. + IFWHT
Post-processing	De-blocking filter	Median filter	Horizontal kernel filter	Horizontal kernel filter
Avg. PSNR (dB)	33.74	34.64	37.76	36.94
Avg. SSIM	0.67	0.83	0.93	0.93
Avg. BPP	3.17	2.84	1.99	0.98

CONCLUSION

We demonstrated that proposed work can further reduce temporal redundancy by using moving detection and inter-coding for compressive video sensing. Our proposed is fast in restoration along with good visual qualities and significantly reduce in BPP. According to the experimental results, the encoder can perform moving detection and show that the test scenes were recovered accurately. The moving detection performance were not disturbed by the initial frame and environmental noise. Moreover, our further compress in detected area can greatly reduce in bandwidth usage. The coding efficiency and performances are measured in terms of PSNR, BPP, and SSIM for a perceptual metric that quantifies image quality degradation. This proposed method can reduce sampling costs and alleviate communication and storage burdens while obtaining comparable estimation performance, which results as a straightforward approach against bandwidth usages.

REFERENCES

[1] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," in *IEEE Signal Processing Magazine,* vol. 25, no. 2, pp. 21-30, March 2008.

[2] Yun Song, Wei Cao, Yanfei Shen, Gaobo Yang, “Compressed sensing image reconstruction using intra

[3] prediction”, Neurocomputing, Volume 151, Part 3, 2015, pp 1171-1179,

[4] J. Zhou, D. Zhou, L. Guo, T. Takeshi and S. Goto, "Measurement-domain intra prediction framework for compressively sensed images," *2017 IEEE International Symposium on Circuits and Systems (ISCAS),* Baltimore, MD, 2017, pp. 1-4.

[5] Peetakul, J. Zhou and K. Wada, "A Measurement Coding System for Block-Based Compressive Sensing Images by Using Pixel-Domain Features," *2019 Data Compression Conference (DCC)*, Snowbird, UT, USA, 2019, pp. 599-599.

[6] VIRAT dataset via S. Oh et al., "A large-scale benchmark dataset for event recognition in surveillance video," CVPR 2011, Colorado Springs, CO, USA, 2011, pp. 3153-3160.

[7] Town Centre Dataset by Active Vision Laboratory via http://tiny.cc/xq0uez

[8] T. Chavdarova et al., "WILDTRACK: A Multi-camera HD Dataset for Dense Unscripted Pedestrian Detection," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 5030-5039.

法政大学 HOSEI University

+This work is supported by JST, PRESTO Grant Number JPMJPR1757 Japan.

Data Compression Conference (DCC), March 23 – 28, 2020