Cube-based Video Coding Framework for Compressive Imaging

Jirayu Peetakul, Yibo Fan, Jinjia Zhou

Hosei University
Graduate School of Science and Engineering
Tokyo, Japan
jirayu.peetakul.6u@stu.hosei.ac.jp
zhou@hosei.ac.jp

Fudan University
State Key Laboratory of ASIC and System
Shanghai, China
fanyibo@fudan.edu.cn

JST, PRESTO
Japan Science and Technology Agency,
Tokyo, Japan

Introduction

This work takes a fresh look at raw data by viewing it as cube made up of multiple downsampled images rather than a vector. As a result, we can view each data point as a pixel. Following that, we propose a tailored video coding framework for cube structure that includes directional 9 modes intra and inter prediction with block-matching motion estimation, transformation, quantization, and entropy coding. When compared to state-of-the-art works, this proposal provides a significant improvement in coding performance and flexibility. We evaluated coding performance using various 4K datasets, resulting in 60-65% lower bit-per-pixels while maintaining visual quality compared to state-of-the-art works.

Cube-based data structure

$$\underbrace{\begin{bmatrix} y_1 = (x_1\Phi_{1,1}) + (x_2\Phi_{1,2}) + \cdots + (x_n\Phi_{1,n}) \\ y_2 = (x_1\Phi_{2,1}) + (x_2\Phi_{2,2}) + \cdots + (x_n\Phi_{2,n}) \\ \vdots \\ y_m = (x_1\Phi_{m,1}) + (x_2\Phi_{m,2}) + \cdots + (x_n\Phi_{m,n}) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} \Phi_{1,1} & \Phi_{1,2} & \cdots & \Phi_{1,n} \\ \Phi_{2,1} & \Phi_{2,2} & \cdots & \Phi_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{m,1} & \Phi_{m,2} & \cdots & \Phi_{m,n} \end{bmatrix}}_{\Phi} \times \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{x}$$

Figure 1: An illustrate of undetermined system.

Fig 1. represents the traditional perspective we utilize compressed sensing.
However, look closely at the block-based CS procedure [1]; an entire frame is partitioned into several blocks with window size of b×b. An elements in the block will be vectorized into x and iteratively linearized into element of y via each 1D kernel order of Φ. It expects to obtain several downsampled images corresponding to m as output by grouping the same element order from all blocks as shown in Fig 2.

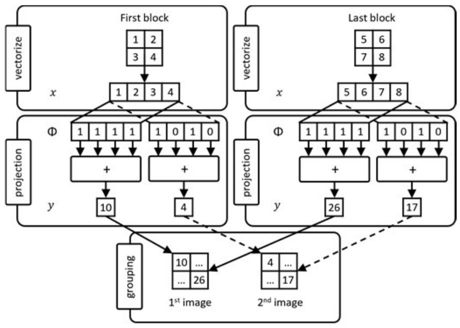

Figure 2: An example of block-based CS procedure, where the block size is set to be 2×2, $n = 4$, $m = 2$, which generate 2 downsampled image.

Cube-based video coding framework #1

In terms of cube-based data structures, we can encode sampled data using 9 modes of intra-prediction, inter-prediction, DCT transform, and quantization table. By coding sub-images in cube, sampled data can be packed more efficiently and transmitted to the receiver. It uses less bandwidth than uncoded and cutting-edge works that proposed vector-based coding algorithms. The explanation of each coding modes are shown in Fig 4. for intra-prediction and Fig 5. for inter-prediction.

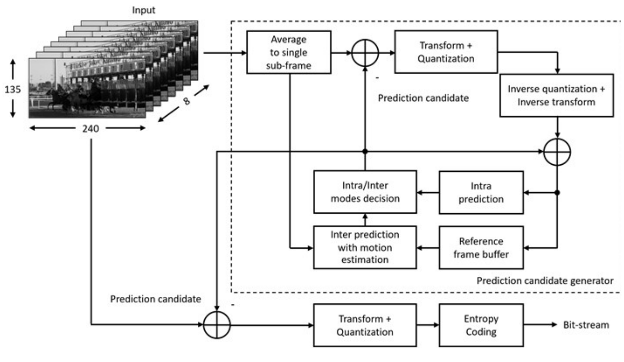

Figure 3. The proposed framework architecture tailored to a data cube for compressive imaging.

[1] Lu Gan, "Block Compressed Sensing of Natural Images," *2007 15th International Conference on Digital Signal Processing*, 2007, pp. 403-406, doi: 10.1109/ICDSP.2007.4288604.
[2] A. Mercat, M. Viitanen, and J. Vanne, "UVG dataset: 50/120fps 4K sequences for video codec analysis and development," Accepted to ACM Multimedia Syst. Conf., Istanbul, Turkey, June 2020.
[3] S. Mun and J. E. Fowler, "DPCM for quantized block-based compressed sensing of images," 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), 2012, pp. 1424-1428.
[4] Zhang, J., Zhao, D. \& Jiang, F. Spatially directional predictive coding for block-based compressive sensing of natural images. {\em 2013 IEEE International Conference On Image Processing}. pp. 1021-1025 (2013)
[5] J. Zhou, D. Zhou, L. Guo, Y. Takeshi and S. Goto, "Measurement-domain intra prediction framework for compressively sensed images," 2017 IEEE International Symposium on Circuits and Systems (ISCAS), 2017, pp. 1-4, doi: 10.1109/ISCAS.2017.8050262.
[6] J. Peetakul, Y. Fan and J. Zhou, "Intra Prediction-Based Measurement Coding Algorithm for Block-Based Compressive Sensing Images," in IEEE Access, vol. 9, pp. 56031-56040, 2021, doi: 10.1109/ACCESS.2021.3068579.
[7] T. T. T. Tran, J. Peetakul, C. D. K. Pham and J. Zhou, "Bi-directional intra prediction based measurement coding for compressive sensing images," 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP), 2020, pp. 1-6, doi: 10.1109/MMSP48831.2020.9287074.

Cube-based video coding framework #2

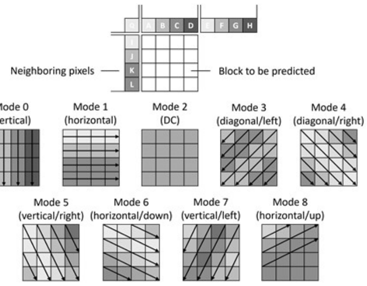

Figure 4: Standard 4 ×4 block directional 9 modes intra-prediction.

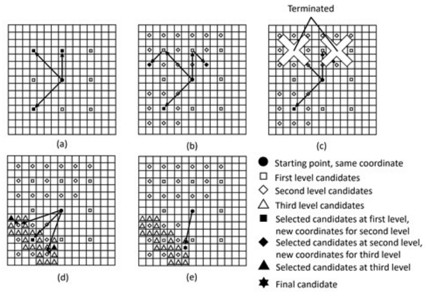

Figure 5: Fast three-level hierarchical search motion estimation capable of stopping motion vectors with v equal to 3, where (a) the top-level of hierarchical motion estimation, (b) the middle-level of hierarchical motion estimation, (c) two motion vectors in the middle-level of hierarchical motion estimation are stopped, (d) the lowest-level of hierarchical motion estimation received computational resources to active motion vector, and (e) the matched candidate is returned.

Simulation results

We evaluate the proposed algorithm on 4K resolution datasets such as Beauty, ReadySetGo, and Bosphorous [2]. In this paper, we use a binary random sensing matrix with the following configuration: n = b ×b = 16 ×16 = 256 and m = 64. We use the L1-minimization method via a primal-dual interior-point method to convert CS data to image. We compare coding performance using a variety of quantitative matrices, including PSNR and SSIM and bit-per-pixel (bpp). To compare with existing works, we run an experiment in two parts: intra-prediction only in Table 1 and intra-prediction with inter-prediction in Table 2.

Table 1. Intra-prediction performance comparison with the state-of-the-art on various datasets, where b×b = n = 16 ×16, m = 64, (a) binary random matrix with DPCM-SQ [3], (b) binary random matrix with SDPC-SQ [4], (c) binary random matrix with modified-binary random matrix with 3 modes intra prediction-SQ [5], (d) hadamard with 4 modes intra prediction-SQ [6], (e) sequency-ordered walsh-hadamard with 7 modes intra prediction-SQ [7], (f) this work only intra-prediction with Q_p = 3, (g) this work only intra-prediction with Q_p = 6, and (h) this work only intra-prediction with Q_p = 9.

Coding methods	Beauty dataset	ReadySetGo dataset	Bosphorus dataset
	PSNR / SSIM / bpp	PSNR / SSIM / bpp	PSNR / SSIM / bpp
(a)	34.77 / 0.79 / 2.46	34.35 / 0.85 / 2.84	35.24 / 0.93 / 2.73
(b)	34.53 / 0.79 / 2.23	34.06 / 0.82 / 2.52	34.37 / 0.90 / 2.52
(c)	34.77 / 0.79 / 1.98	36.04 / 0.92 / 2.21	35.42 / 0.93 / 2.28
(d)	35.85 / 0.84 / 1.92	36.04 / 0.92 / 2.14	39.13 / 0.97 / 2.29
(e)	37.13 / 0.86 / 1.65	35.79 / 0.90 / 1.91	40.37 / 0.94 / 1.94
(f)	47.92 / 0.98 / 0.83	39.11 / 0.97 / 0.99	43.22 / 0.98 / 0.95
(g)	47.65 / 0.98 / 0.64	39.07 / 0.97 / 0.84	43.13 / 0.98 / 0.78
(h)	47.24 / 0.98 / 0.53	39.01 / 0.97 / 0.75	42.98 / 0.97 / 0.68

Table 2: Intra-prediction with 2 modes inter-prediction performance on 100 frames of various datasets, where b×b = n = 16×16, m = 64. The results in (a), (b), (c), represent intra-prediction with slow mode inter-prediction; (d), (e), (f), represent intra-prediction with fast mode inter-prediction, where Q_p = 3, 6, and 9, respectively.

Coding methods	Beauty dataset	ReadySetGo dataset	Bosphorus dataset
	PSNR / SSIM / bpp	PSNR / SSIM / bpp	PSNR / SSIM / bpp
(a)	48.79 / 0.98 / 0.4496	39.27 / 0.97 / 0.7151	43.17 / 0.98 / 0.6237
(b)	48.66 / 0.98 / 0.3078	39.26 / 0.97 / 0.5581	43.68 / 0.98 / 0.4632
(c)	48.50 / 0.98 / 0.2387	39.24 / 0.97 / 0.4719	43.62 / 0.98 / 0.3794
(d)	48.79 / 0.98 / 0.4495	39.27 / 0.97 / 0.7151	43.71 / 0.98 / 0.6245
(e)	48.67 / 0.98 / 0.3084	39.26 / 0.97 / 0.5581	43.68 / 0.98 / 0.4634
(f)	48.05 / 0.98 / 0.2387	39.24 / 0.97 / 0.4719	43.64 / 0.98 / 0.3794