研 究 終 了 報 告 書

「バッテリーレス・ワイヤレス動画収集機能を持つ高分散型監視システム」

研究期間： 2017 年 10 月～2021 年 3 月
（新型コロナウイルス感染症の影響を受け 2021 年 6 月まで延長）

研 究 者： 周 金佳

１． 研究のねらい

Driven by the advances in artificial intelligence technologies, smart environments (or "pervasive computing") already penetrate our daily lives. Video surveillance system is one of the most important components in many smart environments. More and more smart homes, smart buildings, and public Utilities like hospitals, airports or universities will be equipped with billions of ubiquitous surveillance devices. However, with the increasing number of wireless video cameras, the surveillance systems including video capturing, compression and transmission, are going to be challenged by the power efficiency. This research aims at realizing an ultra-low-power wireless video surveillance system that achieves 80% power reduction comparing with the current mainstream solutions.

Compared to other undergoing researches that are mostly focus on the power reduction of image sensing and pixel compression separately, this research focuses on the whole system. 1) It is the first time to explore the correlations between the image sensor and subsequent data compression. The feature of the measurement matrix used in the compressive sensing (CS) based sensor will be utilized for the measurement domain compression. 2) It is the first time to apply the measurement domain compression for the CS based sensors. Comparing with pixel domain compression, the power of compression can be significantly reduced with a little loss of compression ratio. 3) It is first time to integrate deep learning methods to decompress the data in measurement domain. 4) Compared to the traditional CS based sensors, this research further reduces the size of measurement, which can be significant for saving the transmission bandwidth. This research will greatly simplify the compression algorithm. The transceiver power will be increased, but the whole system power is expected to be greatly reduced to 40mW. Wind power and solar energy, can substitute for battery.

1) By applying a new compressive sensing (CS) patterns, the power of image sensor is expected to be reduced by more than 80%.
2) A new measurement coding system will be applied to further compress the data from sensor with high compression ratio. Comparing the traditional coding technology (intra-only HEVC, the power of coding is expected to be reduced by more than 80%.
3) Reducing the original pixel data volume by more than 90%. After combining with the CS technology and measurement compression, the compression ratio is expected to be more than 90%.
4) A deep learning based video reconstruction and video enhancement will be applied to improve the quality of the reconstructed video.

2. 研究成果

(1)概要

As show in Figure 1, the proposed system is composed of four main components. In the encoder side, we integrated the compressive sensing (CS) based image sensor and the corresponding measurement domain compression to reduce the power by applying simplified coding algorithm. In the decoder side, decompression and quality enhancement were applied to ensure the quality of the decoded video. We have achieved the following results.

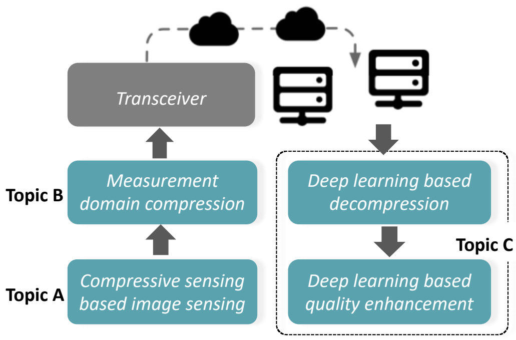

Figure 1. System overview

Topic A. Developing the compressive sensing based image sensing technology. This topic involves designing a new measurement matrix (optical coding patterns) for CS-imager to integrate the pixel domain features into the measurement domain and reduce the power of sensors. We firstly tried some existing sensing patterns, and finally designed new patterns to achieve better results on improving the sensing quality and improving compression ratio of the corresponding measurement coding.

Topic B. Developing Measurement domain compression technology. This topic is to further reduce the output data from CS based sensor. The output data is called measurement. We have implemented intra prediction based measurement coding to reduce the spatial and temporal redundancy. The experimental results show that it can greatly improve visual perception in terms of peak signal-to-noise ratio (PSNR) by 8% and structural similarity (SSIM) by 21%, and reduce 46% in bitrate utilization in term of bit-per-pixel (bpp) compared to state-of-the-art works.

Topic C: Developing the deep learning based decompression technologies. This topic is to Integrate deep learning technologies to decompress the measurement data and increase the quality of decoded video. We integrated deep learning based three technologies including video decompression, image enhancement, and super resolution are integrated together to achieve better results. For the reconstruction, a PSNR of 22dB can be achieved at 64x compression ratio, which is about 4% to 9% better than state-of-the-art works.

Topic D: Integration and evaluation. We integrated the encoding and decoding components together to show video capturing, encoding and decoding. Moreover, we have implemented the hardware of measurement coding and simulated the power of the encoder.

During the extended 3 months, we have implemented the whole measurement encoding system into the FPGA platform with a camera. And then, the system was verified and evaluated. Moreover, the video reconstruction algorithm has been further improved by applying in-loop and out-loop filters.

(2)詳細

研究ターマ A:Developing the compressive sensing based image sensing technology

(5. 主な研究成果リスト.(3).[1.2]).

We have realized the target of designing sensing patterns to embed the pixel information. The pixel information was used for the further compression (Task B). Firstly, we applied partly-structural and partly-random matrix for the optical coding. The sensing power can be reduced to 25% while keeping the video quality. Moreover, we proposed a continuously ordered Walsh-Hadamard matrix. The sensing power can be reduced to 1/64 with PSNR higher than 30dB.

研究ターマ B:Developing the measurement domain compression technologies.

(5. 主な研究成果リスト.(3).[1.2], (2).[2])

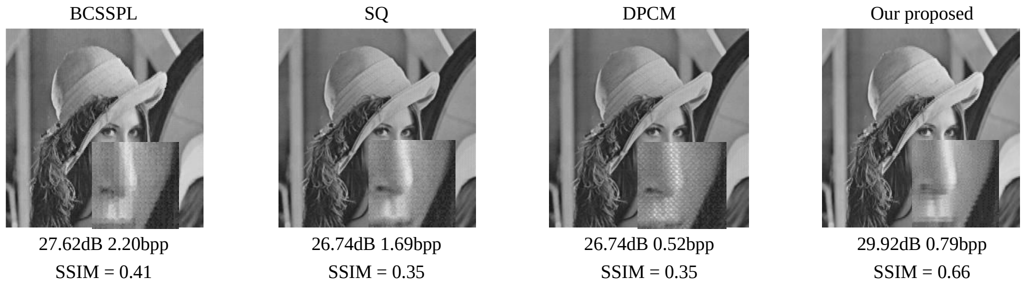

Figure 2. Comparison of visual qualities and compression artifacts of four test images: Lenna, Goldhill, Mandrill, and Pentagon among our proposal, BCSSPL, SQ, and DPCM with $B = 4$,

This topic is to further reduce the output data from CS based sensor. The output data is called measurement. By embedding the pixel information to the measurement domain (Topic A), an intra-prediction based coding system is proposed to reduce the spatial redundant in measurement domain by using the pixel domain features. Furthermore, we employed image quantization and Huffman coding to further compress the data. The experimental results show that the proposed system can greatly improve the coding efficiency, which increases 1.94dB

– 2.3dB in PSNR and reduces 42% – 65% bitrate in terms of bit-per-pixel as shown in Fig 2. Comparing the pixel compression (HEVC-intra), the complexity of measurement can be reduced more than 90%. The target of topic B is realized.

研究ターマ C: Developing the deep learning based decompression technologies.
(5.主な研究成果リスト.(3).[3,5])
We integrated deep learning based three technologies including video decompression, image enhancement, and super resolution are integrated together to achieve better results. A deep neural network with interpretable motion compensation called CS-MCNet is proposed to realize high-quality and real-time decoding of video compressive sensing. By using the information from the neighboring frames, the video quality can be improved. PSNR of 22dB can be achieved at 64x compression ratio, which is about 4% to 9% better than state-of-the-art methods. In addition, due to the feed-forward architecture, the reconstruction can be processed by our network in real time and up to three orders of magnitude faster than traditional iterative methods.

研究ターマ D: Integration and evaluation
We implemented the measurement coding in hardware level using field-programmable gate array (FPGA). The detailed implementation results are shown in section 2*.

３． 今後の展開

This project has finished evaluating the algorithm of the proposed system including image acquisition, compression, transmission, and decompression. The hardware architecture and function were evaluated on FPGA. The power consumption is simulated by the synthesis tools. In the future, we will integrate the whole system to an ASIC chip. Based on the ASIC, we will analyze the function and measure the power consumption. The proposed new coding framework can not only be utilized in the image sensing system, but also can be used in various research field such as video/image understanding, interactive visual learning, and so on.

４． 自己評価

研究目的の達成状況

1) By applying a new compressive sensing (CS) patterns, the power of image sensor is targeted to be reduced by more than 80%. Finally, a better result was achieved. The proposed work reduced the power of image sensor by more than 90% with the sampling rate of 1/16. At the same time, video quality in terms of PSNR was maintained to higher than 30dB.
2) A new measurement coding system will be applied to further compress the data from sensor with high compression ratio. Comparing the traditional coding technology (intra-only HEVC,

the power of coding is targeted to be reduced by more than 80%. Finally, a better result was achieved. For the resolution of 3820×2140 with the sampling rate of 1/16, the required power is 28mW. The traditional HEVC-intra requires more than 120mW for the resolution of 352x288 with 30fps. The normalize energy consumption of this work is reduced by 92%.

3) After combining with the CS technology and measurement compression, the compression ratio is expected to be more than 90%. Finally, a better result was achieved. A controller was added to control the video quality and compression ratio. When the video quality in terms of PSAR is higher than 30dB, the compression ratio is from 75% to 98%.
4) The target is to apply a deep learning based video enhancement to improve the quality of the reconstructed video. Finally, not only a deep learning based video enhancement but also a deep learning based video reconstruction were applied to achieve high quality of the reconstructed video. The PSNR was improved more than 5dB.

研究の進め方

This research requires a lot of work to collect and analyze huge data. It is great that JST Presto can provide a long-term support to hire research assistants. Moreover, JST also supports international collaboration. Through the collaboration, we obtained a video data set with the resolution of 1 billion pixels to do the test, and our collaborated work achieved a good result.

研究成果の科学技術及び社会・経済への波及効果

For this research, compressive sensing technology, measurement domain compression technology, and deep learning technology were integrated into a system to capture high-quality video with low power consumption. Comparing with the conventional solution, the power of the whole system was reduced by more 80%. This technology can be widely used in surveillance system to capture video with batteryless and powerless. The proposed ultra-low-power surveillance system plays a vital role in building smart environments for the next-generation information society. Besides the surveillance system, the applied technologies also can be used in various research field such as video/image understanding, interactive visual learning, and so on.

５． 主な研究成果リスト

（１） 代表的な論文（原著論文）発表

研究期間累積件数：8件（うち1件はコロナ延長時の成果）

1. Man M. Ho, Jinjia Zhou, Gang He, "RR-DnCNN v2.0: Enhanced Restoration-Reconstruction Deep Neural Network for Down-Sampling Based Video Coding", IEEE Transactions on Image Processing (TIP). DOI:10.1109/TIP.2020.3046872. Jan. 2021.
This work proposes a restoration-reconstruction deep neural network (RR-DnCNN) using the degradation-aware techniques to integrate super resolution technologies to video compression. Moreover, to prevent the loss of essential features in the very deep network from

restoration to super-resolution, we leverage up-sampling skip connections to compensate for the lost information from restoration layers. It is called restoration-reconstruction u-shaped deep neural network (RR-DnCNN v2.0). As a result, our RR-DnCNN v2.0 can achieve 17.02% BD-rate reduction on UHD resolution compared to the standard H.265/HEVC.

2. Do Kim Chi Pham, Jinjia Zhou, "Deep Learning-based Luma and Chroma Fractional Interpolation in Video Coding", IEEE Access, Vol. 7, pp. 112535-112543, Aug. 2019. DOI: 10.1109/ACCESS.2019.2935378

This work proposes CNN-based fractional interpolation for Luminance (Luma) and Chrominance (Chroma) components in motion compensated prediction to improve the coding efficiency. As a result, our proposal gains 2.9%, 0.3%, 0.6% Y, U, V BD-rate reduction, respectively.

3. Muchen Li, Jinjia Zhou, Satoshi Goto, "A configurable fixed-complexity IME-FME Cost ratio based Inter mode filtering method in HEVC encoding", IIEEJ Transactions on Image Electronics and Visual Computing, Vol. 8, No. 1, page58-70 (2020.6)

This works proposes a configurable fixed-complexity IME-FME Cost ratio based Inter mode filtering method. The main idea can not only be used in traditional video coding standard such as HEVC, but also in the compressive sensing based measurement domain coding.

4. Pham Do Kim Chi, Jian Yang, Jinjia Zhou, "CSIE-M: Compressive sensing image enhancement using multiple reconstructed signals for IoT surveillance systems", IEEE Transactions on Industrial Informatics, DOI: 10.1109/TII.2021.3082498. May. 2021 （※コロナ延長時の成果）

Different from the other image compression standards, compressive sensing (CS) can get various reconstructed images by applying different reconstruction algorithms on coded data. Using this property, it is the first time to propose a deep learning-based compressive sensing image enhancement framework using multiple reconstructed signals (CSIE-M). Firstly, images are reconstructed by different CS reconstruction algorithms. Secondly, reconstructed images are assessed and sorted by a No-reference quality assessment module before being inputted to the quality enhancement module by order of quality scores. Finally, a multiple-input recurrent dense residual network is designed for exploiting and enriching the useful information from the reconstructed images. Experimental results show that CSIE-M obtains 1.88-8.07dB PSNR improvement while the state-of-the-art works achieve a 1.69-6.69 dB PSNR improvement under sampling rates from 0.125 to 0.75. On the other hand, using multiple reconstructed versions of the signal can improve 0.19-0.23 dB PSNR, and only 4% reconstructing time is increasing compared to using a reconstructed signal.

（2） 特許出願

研究期間累積件数:1 件(特許公開前のものも含む)

[1] 周金佳, ピエタク ジラユー, ″イメージング装置及び測定符号化装置″, JP Patent, 出願番号: 特願 2021-008179, 出願日:令和 3 年(2021 年)1 月 21 日

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

[1] Jirayu Peetakul, Jinjia Zhou, Koishi Wada, ″A Measurement Coding System for Block-based Compressive Sensing Images by Using Pixel-Domain Features″, IEEE Data Compression Conference (DCC 2019), Snowbird, USA, March 2019. DOI:10.1109/DCC.2019.00111

[2] Jirayu Peetakul and Jinjia Zhou, ″Temporal Redundancy Reduction in Compressive Video Sensing by using Moving Detection and Inter-Coding,″ 2020 Data Compression Conference (DCC), Snowbird, UT, USA, 2020, pp. 387-387, doi: 10.1109/DCC47342.2020.00052

[3] Man M. Ho, Gang He, Zheng Wang, and Jinjia Zhou*, ″Down-Sampling Based Video Coding with Degradation-aware Restoration-Reconstruction Deep Neural Network″, The 26th International Conference on Multimedia Modeling (MMM), Daejeon, Korea, Jan. 2020.DOI: 10.1007/978-3-030-37731-1_9 (Best Paper Runner-Up Award)

[4] Jinjia Zhou, “Compressive sensing based coding system,” Video Live Stack Conference, Shanghai, China, April. 2019, invited talk

[5] Bowen Huang, Jinjia Zhou, Xiao Yan, Ming'e Jing, Rentao Wan, Yibo Fan, ″CS-MCNet: A Video Compressive Sensing Reconstruction Network with Interpretable Motion Compensation″, Asian Conference on Computer Vision 2020 (ACCV), Virtual Tokyo. Nov. 2020.