# Distributed Compressed Video Sensing based on Convolutional Sparse Representation

1st Yosuke Higuchi, 2nd Yoshimitsu Kuroki
*National Institute of Technology, Kurume College*
Fukuoka, Japan
a3922yh, kuroki@kurume.kosen-ac.jp

***Abstract*—Conventional Convolutional Sparse Representation (CSR) has difficulty in learning a set of filters, called a dictionary, for many images due to memory requirements. In this study, we have solved it by using consensus-framework CSR and have shown experimentally that decoding accuracy can be improved by using a large number of keyframes for learning.**

***Index Terms*—vision / image processing**

## I. Introduction

Convolutional Sparse Representation (CSR) [1], [2] approximates images by the convolution of a dictionary of image features and sparse coefficient maps. Therefore, Distributed Compressed Video Sensing (DCVS) with CSR is robust to image shifting, while conventional DCVS with block-wise method causes block noise due to block-by-block processing. In addition, CSR learn a dictionary that extracts feature common to each image by using a large number of training images. Using this property, this study aims to improve decoding accuracy by using a large number of key frames for training.

## II. Proposed Method

In this study, to improve decoding accuracy, we use a large number of key frames to learn the dictionary. In general CSR, the memory requirement depends on the number of training images, so it is not possible to use a large number of images for training at one time. Therefore, we adopt the Consensus Framework whose memory requirement does not depend on the number of training images.

## III. Experimental Results

We experimented with a foreman (300 frames of 128×128 pixels) video data sets. The video was divided into 10 Group of Pictures (GOPs), each GOPs consisting of one key frame located at the beginning and 29 non-key frames. In learning process, the proposed method uses all key frames to learn the single dictionary for decoding. In contrast, the conventional method learns dictionaries for each GOPs using the key frames located at the beginning of each GOPs. In decoding process, coefficients are optimized for the non-key frames belonging to each GOPs by the proximal gradient method. Then, the original image is estimated by the convolution of the learned dictionary and the optimized coefficients. Fig. 1 shows the Peak signal-to-noise ratio (PSNR) values of the decoded non-key framses when the compression ratio is set to 0.5. Fig. 2 shows the average difference in PSNR values between the proposed and the conventional method for the decoded non-key frames at each GOPs.

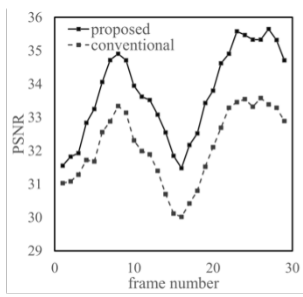


Figure 1. PSNR of proposed and conventional

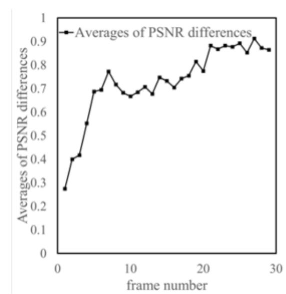


Figure 2. Average of PSNR differences

Experimental results show that the proposed method improves the decoding accuracy over the conventional method.

## IV. Conclusion

In this study, we performed DCVS based on CSC using a large number of key frames to improve the decoding accuracy. The experimental results show that the proposed method were able to improve the decoding accuracy by learning the dictionary using a large number of key frames with CSC based on Consensus Framework. Future work includes the use of an L1-norm error term when learning with a large number of key frames.

## Aknowlegdement

This work was supported by JSPS KAKENHI Grant Numbers JP20K11878 and JP23K11159.

## References

[1] (Cristina and Wohlberg, 2018) G. Cristina and B. Wohlberg, Convolutional Dictionary Learning: "A Comparative Review and New Algorithms", IEEE, pp.366-381, 2018.

[2] (Mizokami and Kuroki, 2019) Tomohito Mizokami and Yoshimitsu Kuroki, "Distributed compressed video sensing based on convolutional sparse coding", International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Dec. 2019. DOI: 10.1109/ISPACS48206.2019.8986358