On-Chip Data Reduction and Object Detection for a Feature Extractable CMOS Image Sensor

Yudai MORIKAKU, Ryuichi UJIIE, Daisuke MORIKAWA, Hideki SHIMA, Kota YOSHIDA, and Shunsuke OKURA

***Abstract*—Our research group has proposed an image sensor that can extract features suitable for AI. In this paper, we present object recognition accuracy of the feature data with a deep neural network trained. Furthermore, to reduce the power consumption of the image recognition system, we propose a data reduction scheme in the image sensor chip. According to simulation results, object recognition accuracy is 56.6% even though data amount is reduced by 97.7% compared to the conventional RGB images.**

***Index Terms*—image sensors, feature extraction, object recognition**

I. Introduction

For the coming IoT age, our research group has proposed an image sensor that can extract lightweight features suitable for AI in order to save a amount of data and energy consumption [1]. In this paper, we propose further data reduction scheme of the features extracted in a CMOS image sensor and present simulation results of object recognition of the feature data using COCO dataset and YOLOX. Since horizontal edge extraction, pseudo RGB to grayscale conversion, horizontal binning, and low-bit quantization are performed in a CMOS image sensor, the data amount of the feature is aggressively compressed compared to that of RGB image while the feature is still recognizable with the YOLOX.

II. Proposed CMOS Image Sensor

Figure 1 shows an overview of a proposed CMOS image sensor which can output both RGB color image and feature data. The feature data is extracted as follows. With the 4-shared pixel structure, RGB color pixel signal is pseudo converted to grayscale given by $R+2G+B$ when photoelectrons integrated in a shared pixel are transferred to a shared charge detection node simultaneously, reducing the number of pixels to $1/4$. The horizontal edge of the image is then derived by subtracting vertical adjacent pixel cells instead of correlated double sampling of a single pixel, reducing the number of pixels to $1/2$. The analog feature output from the pixel array is quantized by setting the column parallel ADC to 3-bit resolution mode, reducing the bit-resolution to $3/8$. The horizontally adjacent pixel signal is then binned reducing the number of pixels to $1/2$. Therefore, the output data size of the feature data is reduced to only $2.3\%(=3/128)$ compared to that of the RGB color image.

Y. Morikaku, K. Yoshida. and S. Okura. are with Ritsumeikan Univ.. R. Ujiie., D. Morikawa. and H. Shima. are with Nisshinbo Micro Devices Inc..
This work was supported by JSPS KAKENHI Grant Number JP20K04630.

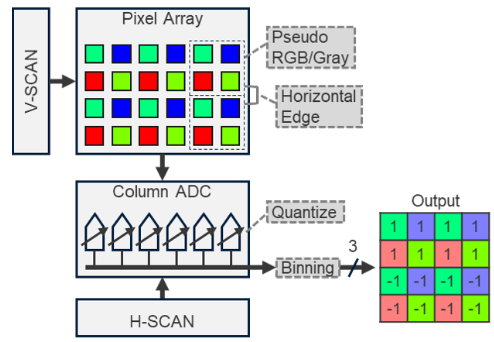
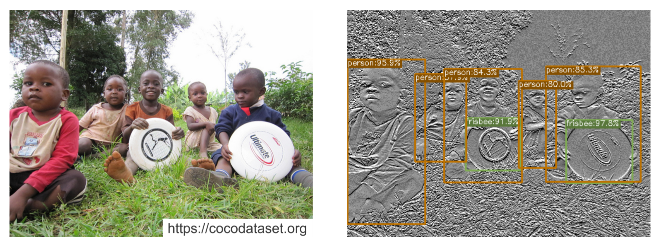


Fig. 1: Overview of a proposed CMOS image sensor.



(a) RGB color image (b) feature data

Fig. 2: Simulation result of a feature extraction

III. Simulation Results

The RGB color images of the COCO dataset are converted to the feature data to simulate the image recognition accuracy with the proposed feature extractable CMOS image sensor, as an example is shown in Fig. 2. The YOLOX pretrained with the feature data achieved the precision AP50 of 56.6%, while the YOLO with RGB color image achieved 69.6%. Since the AP50 was 56.5% when twice noise of the theoretical pixel noise is imposed to the feature data, robustness to random noise is also confirmed.

IV. Conclusion and Future Work

The proposed CMOS image sensor output data is reduced by 97.7% even though the object recognition accuracy is reduced only by 13.0%. The image sensor will be fabricated to save a amount of data and energy consumption of image recognition system in the IoT age.

References

[1] S. OKURA and *et al.*,, in Proc. IISW, May 21-25, 2023.