A Study on a Feature Extractable CMOS Image Sensor for Low-Power Image Classification System

Shunsuke Okura, Ai Otani, Koshiro Itsuki, Yusuke Kitazawa,
Kohei Yamamoto, Yu Osuka, Yudai Morikaku, and Kota Yoshida
Ritsumeikan Univ., Shiga, Japan, sokura@fc.ritsumei.ac.jp

I. Introduction

In development of internet of things (IoT) with trillion sensor universe, the amount of information collected by CMOS image sensors will be drastically increased, and an image recognition based on deep learning (DL) is getting more important to process the big imaging data. However, the data captured by conventional image sensors is redundant for the DL because features of the imaging data is extracted in the deep neural networks (DNNs). For the feature extraction, convolution multiply-accumulate (MAC) operation takes place in the DNNs. Yoneda et al. has proposed an image sensor capable of analog convolution [1], in which convolution MAC operation is conducted in a pixel array. However, crystalline oxide semiconductor FET with an in-pixel capacitor are utilized to suppress leakage current during the convolution, thus resulting in large pixel size. Besides, operational transconductance amplifier (OTA) is utilized for current-domain MAC operation with larger power consumption compared to normal imaging mode. Young et al. has proposed a log-gradient image sensor [2], in which low-complexity feature for machine learning (ML), that is histogram of oriented gradients (HoG), is derived in a column readout circuit with analog four-line-memory. The HoG is aggressively quantized with a 2.75 bit ratio-to-digital converter (RDC), thus conventional RGB color image cannot be readout with the signal chain.

In this paper, a CMOS image sensor which can generate both normal image for human and feature data for the DL is proposed to reduce the power consumption of the image classification system and to save storage space for the big data. In order to keep compatibility with conventional image sensors, the CIS does not employ analog memory for convolution. Simulation results of image classification with horizontal edge as feature data of CMOS image sensor output is shown Sec. II. The CMOS image sensor which can generate the horizontal edge is presented in Sec. III, followed by summary and future work in Sec. IV.

II. Image classification system with a feature extractable CIS

Figure 1 shows a concept overview of our image classification system with the feature extractable CIS. The CIS is switched to an imaging mode according to a trigger generated by a CNN that detects person and/or other objects with the feature data. At the feature mode, the power consumed by the CIS can be reduced with aggressive quantization such as 3 bit. Besides, the power consumption of the CNN will be reduced by omitting redundant layers and filter channels in the feature extractor designed for normal images.

In order to verify image classification accuracy with the feature data, the RGB color INRIA person dataset [3] is converted to a horizontal edge dataset and then input to a 3-layer CNN as shown in Fig. 2. The process to generate the horizontal edge dataset simulates the operation of the proposed CIS. The horizontal edge is difference between vertically adjacent pixels derived with y-derivative but without convolution, thus the 64×128 pixel input image is scaled down to 64×64 pixel. The noise is also added because pixel reset noise is not cancelled in the y-derivative as described in Sec. III. The quantizer represents low-bit analog-to-digital (A/D) conversion. The quantized horizontal edge is resized back to 64×128 pixels in order to use same size CNN for the comparison of original RGB dataset and the horizontal edge dataset. It is expensive to construct a new dataset with our feature-extractable CIS for training CNN models, but the cost can be almost negligible by transforming public datasets according to the behavior of our sensors. Image classification accuracy is summarized in Table I, in which 8 bit RGB color image at the imaging mode and 8 bit horizontal edge at the feature mode are utilized for the training and test of the CNN. The accuracy is 98.3% when the RGB color test dataset are classified with the CNN trained with RGB color image (1), while the accuracy drops to 47.2% with the edge dataset for test (2). The classification accuracy of the edge dataset is improved with the CNN trained with the edge dataset (3). The accuracy further increases to 97.0% when contrast of the edge data is enhanced with histogram equalization (4), which is comparable to the classification accuracy of original RGB color dataset. Figure 3 also shows simulation results of the image classification according to the size of the dataset with quantization. The horizontal edge is robust to the quantization down to 3 bit. Even though the accuracy decreases only by 1.4%, the size of the dataset decreases by 95% with the 3 bit edge dataset compared to the original 8 bit RGB color dataset.

This simulation experiments suggest that the ouput of a CMOS image sensor can be (a) horizontal edge without convolution and (b) low bit-resolution such as 3 bit, for image classification. The feature extractable CIS with the pixel capable of horizontal edge detection and the variable bit-

This work was supported by JSPS KAKENHI Grant Number JP20K04630.

resolution successive approximation register (SAR) analog-to-digital converter (ADC) is described in the following section.

III. Column signal chain of the feature extractable CIS

Figure 4 shows a column signal chain of the feature extractable CIS that consists of pixels and a SAR-ADC. The pixel configuration is same as a conventional DCG pixel composed of 4T pixel and a binning gate BIN. The BIN gate is turned-on during the feature mode to derive the difference between vertically adjacent pixels for the y-derivative. The pixel source follower transistor SF is pre-charged with $\Phi\mathrm{PC}$ [4] inactivating the current source I_d to save power consumption during the feature mode. The SAR-ADC is composed of a comparator CMP and a SAR digital-to-analog converter (DAC). A double clamp circuit [5] and the bias current for the latch in the comparator are inactivated during the feature mode also to save power consumption.

Figure 5 shows a timing diagram of the proposed CIS. At the imaging mode, the operation timing is same as a conventional 4T pixel, and the difference between the pixel reset and signal is converted into 10 bit digital code. At the feature mode, the signal levels of J-th and (J+1)-th row pixels are readout instead of the reset and signal levels of a selected pixel, so that the FD reset noise is not cancelled. However, the reset noise is acceptable because the feature data can be classified with the CNN at low bit-resolution such as 3 bit. The ADC converts the signal difference between J-th and (J+1)-th row pixels, that ranges from negative to positive, into 5 bit digital code with margin. The 9-th bit switch of the SAR-DAC is connected to $\mathrm{V_{RL}}$ and other switches are connected to $\mathrm{V_{RH}}$ during the sampling of J-th row pixel, thus 0.5 V offset voltage is added to the DAC output $\mathrm{V_{DAC}}$ prior to the A/D conversion. Then, 5 MSBs of SAR-DAC are switched during the A/D conversion. According to SPICE simulations, the current dissipation $\mathrm{I_{dis}}$ during the feature mode is only 0.317 [μA] that is reduced by 99.0% compared to that during the imaging mode as summarized in Table II.

The settling error of the pixel SF at the feature mode is a concern for large capacitive load on the pixel output column line due to weak inversion operation. However, the settling error that depends on the input signal level can be divided into gain error and linearity error after the double sampling of J-th and (J+1)-th row pixels, and the gain error affects less to the feature data. SPICE simulation results of settling error is shown in Fig 6, in which signal level of J-th and (J+1)-th row pixels are respectively swept from 0.0 V to 0.5 V thus the input difference of the SF input is swept from -0.5 V to $+0.5$ V. The gain of the pixel SF is given by 0.91, and the maximum linearity error is given by 7.26 mV which corresponds to 7.1 bit resolution and is acceptable for 5 bit A/D conversion.

A test chip of the variable SAR-ADC was implemented with 0.18 μm CMOS process [6]. Figure 7 shows the chip photograph of the 640 column ADC. Figure 8 shows the sample images for a ramp input signal. It is noted that the nonlinearity is caused by capacitance error of the split capacitor $\mathrm{C_{SP}}$ in the SAR-DAC. Except for the split capacitor error, the DNL was 1.51 LSB at the 10 bit imaging mode and 0.12 LSB at the 5 bit feature mode. At the imaging mode, monotonicity and small column FPN were confirmed as shown in Fig. 8(a). At the feature mode, large column FPN was visible as shown in Fig. 8(b).

IV. Summary and Future work

The CMOS image sensor which can generate both normal image and feature data is proposed to reduce the power consumption of the image recognition system and the sensor output data size. Simulation results with 3-layer CNN shows that the recognition accuracy of the feature data is 96.9% and the data size is reduced by 99.0%, in which the original INRIA person dataset was converted to 3 bit horizontal edge dataset. Since the feature data can be aggressively quantized, the pixel source follower and the SAR-ADC of the column signal chain process the difference of vertically adjacent pixel inactivating the bias current and the total current consumption is only 0.371 [μA] for 5 bit feature data.

A test chip of the variable SAR-ADC was implemented with 0.18 μm CMOS process and was evaluated. The effect of the column FPN to the classification with the CNN will be verified in future. A CMOS image sensor with the proposed signal chain will be fabricated so that layer and filter channel structure of the CNN will be studied to reduce the power consumption of the image classification system also in future.

V. Acknowledgments

This work was supported through the activities of VDEC, The University of Tokyo, in collaboration with Cadence Design Systems, with NIHON SYNOPSYS G.K., and with Mentor Graphics.

References

[1] S. Yoneda, Y. Negoro, H. Kobayashi, K. Nei, T. Takeuchi, M. Oota, T. Kawata, T. Ikeda, and S. Yamazaki, "Image sensor capable of analog convolution for real-time image recognition system using crystalline oxide semiconductor fet," in *International Image Sensor Workshop (IISW)*, pp. 322–325, 2019.

[2] C. Young, A. Omid-Zohoor, P. Lajevardi, and B. Murmann, "A data-compressive 1.5/2.75-bit log-gradient qvga image sensor with multi-scale readout for always-on object detection," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, pp. 2932–2946, 2019.

[3] "Inria person." https://paperswithcode.com/dataset/inria-person. Accessed: 6/4/2023.

[4] M. Guy, B. Jan, W. Xinyang, and G. Vanhorebeek, "Backside illuminated global shutter cmos image sensors," in *International Image Sensor Workshop (IISW)*, no. R51, 2011.

[5] T. Sugiki, S. Ohsawa, H. Miura, M. Sasaki, N. Nakamura, I. Inoue, M. Hoshino, Y. Tomizawa, and T. Arakawa, "A 60 mw 10 b cmos image sensor with column-to-column fpn reduction," in *2000 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, pp. 108–109, 2000.

[6] K. Itsuki, A. Otani, H. Ogawa, and S. Okura, "A variable-resolution sar adc with 10-bit image capturing mode and 5-bit feature extraction mode," in *5th International Workshop on Image Sensors and Imaging Systems (IWISS2022)*, 2022.

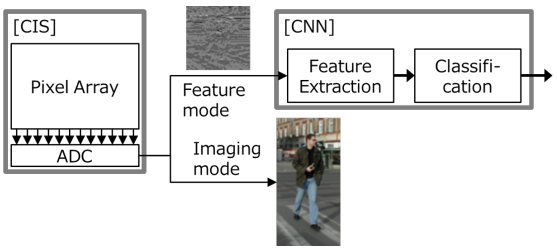


Fig. 1. Overview of an image classification system with feature extractable CIS

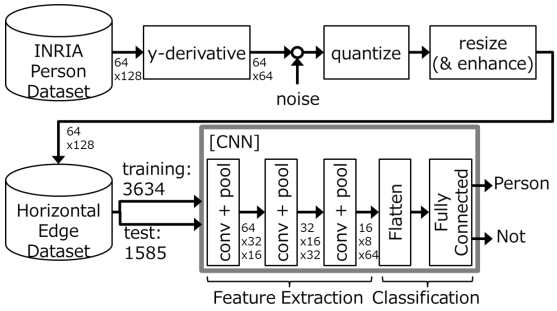


Fig. 2. Block diagram of feature dataset conversion and a 3-layer CNN

	(1)	(2)	(3)	(4)
training	RGB	RGB	Edge-1	Edge-2
test	RGB	Edge-1	Edge-1	Edge-2
accuracy	98.3%	47.2%	95.7%	97.0%

TABLE I
SIMULATION RESULTS OF IMAGE CLASSIFICATION.
EDGE-1: 8 BIT HORIZONTAL W/O CONTRAST ENHANCEMENT.
EDGE-2: 8 BIT HORIZONTAL WITH CONTRAST ENHANCEMENT.

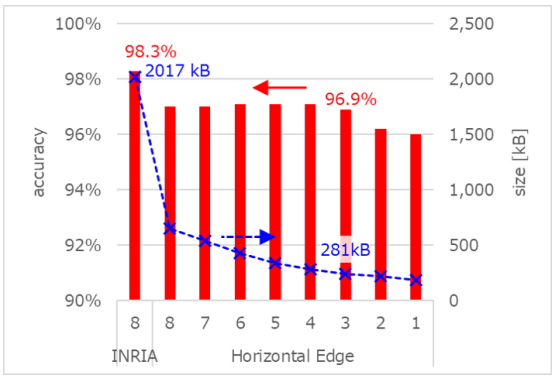


Fig. 3. Simulation results of image classification accuracy and data size

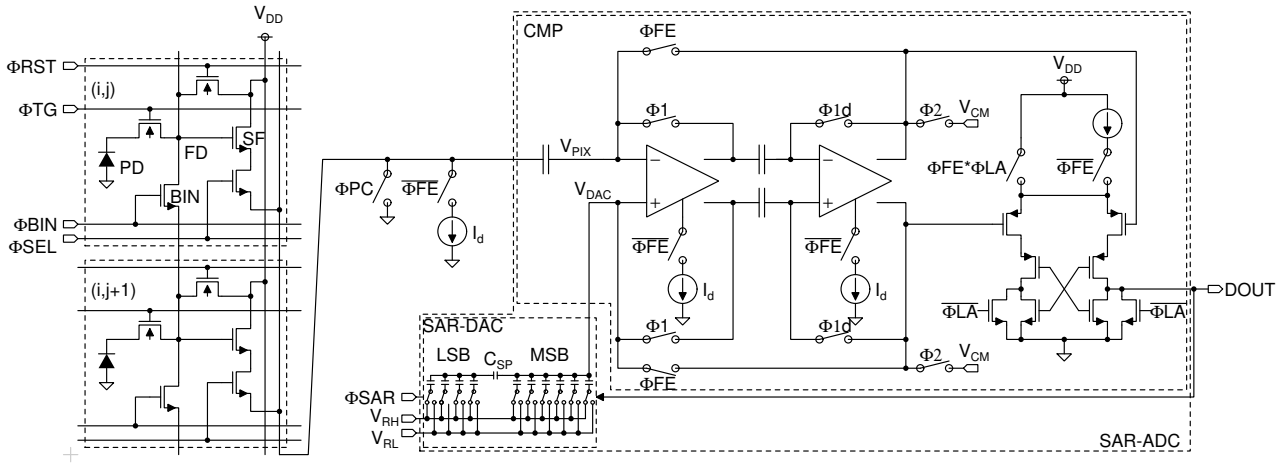
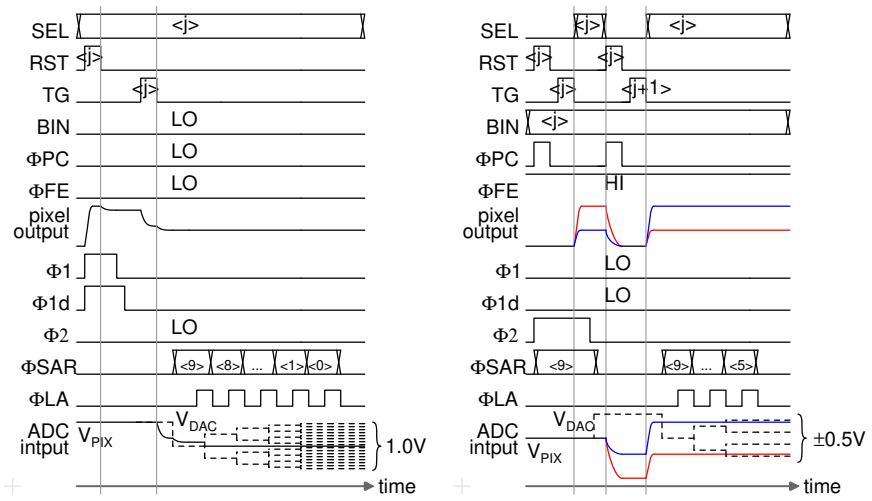


Fig. 4. Schematic diagram of a feature extractable pixel and a variable SAR-ADC



(a) Imaging mode (b) Feature mode

Fig. 5. Timing diagram

I_{dis} [μA]	Imaging mode	Feature mode
Pixel SF	9.2	0.26
ADC	24.1	0.057
total	33.3	0.317

TABLE II
SIMULATION RESULTS OF THE CURRENT CONSUMPTION

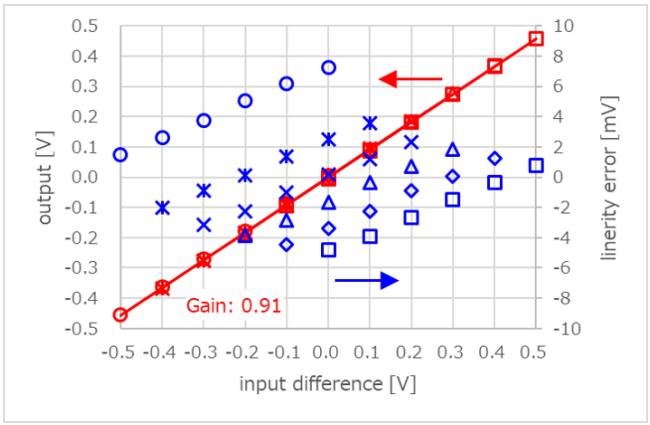


Fig. 6. Simulation result of the pixel SF at the feature mode

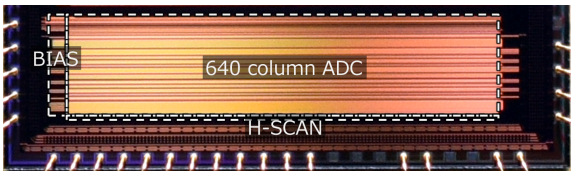
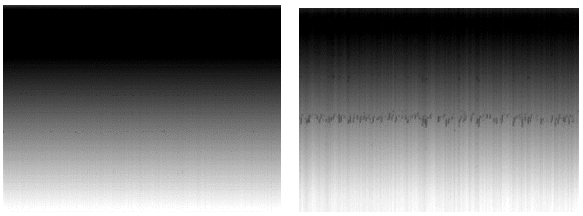


Fig. 7. Chip photograph of the proposed variable SAR-ADC



(a) 10 bit imaging mode (b) 5 bit feature mode

Fig. 8. Sample image for a ramp input signal