領域別に適正露光制御可能な高ダイナミックレンジ1型 17 Mpixel 1000 fps 積層CMOSイメージセンサ

平田友希，村田寛信，有井　卓，松田英明，米持　元，手塚洋二郎，綱井史郎

A 1-inch 17 Mpixel 1000 fps Block-Controlled Coded-Exposure Back-Illuminated Stacked CMOS Image Sensor for Computational Imaging and Adaptive Dynamic Range Control[†]

Tomoki HIRATA, Hironobu MURATA, Taku ARII, Hideaki MATSUDA, Hajime YONEMOCHI, Yojiro TEZUKA and Shiro TSUNAI

人間の知覚を超えた，広い明暗差に対応かつ，高速高解像で画像取得可能なビジョンシステムを提案する．イメージセンサは観賞用カメラだけではなく，ロボットビジョンなど幅広い分野に使われている．例えば自動運転用途では，暗いトンネル内や真夏の日差しの中で，動きに追従して高精細に，かつ高速に物体を認識する必要がある．

しかしながら，1つのカメラで明暗差のある移動被写体を撮る場合，高精細とフレームレート，画素サイズとダイナミックレンジ（DR）にトレードオフの関係があるため，すべての性能を満たすことは困難であった．我々は，4 K×4 Kの解像度で 1000 fps の読み出し動作が可能で，2.7 μm の微細画素でありながら 110 dB の DR を実現する高速ビジョンシステムを開発した．具体的には，高 DR 手法として画素アレイを複数のブロックに分割して個別に露光時間を制御する機能を採用し，ブロック毎に AD 変換器を並列配置することで高速読み出しを実現し，3 次元ウェハ積層技術により回路と画素を高密度に集積することでこれらの課題を克服した．本システムは暗い箇所と明るい箇所が画面内に混在するような高 DR が必要な用途に加えて，符号化露光（Coded Exposure）を用いたコンピューテーショナルイメージングへの応用が期待される．

This study introduces a vision system that can acquire images at high speeds and high resolutions. It can handle wide differences in brightness beyond human perception. Image sensors are not only used in digital still cameras but also in a wide range of fields, such as robot vision. Object recognition and movement tracking at high speeds and high definitions are essential in automated driving systems, especially in dark tunnels or in the mid-summer sunshine. However, capturing a moving subject with a high contrast and achieving a high performance concerning all aspects is challenging because of the tradeoff between high definition and high frame rate, and pixel size and dynamic range (DR). We developed a high-speed vision system capable of 1000-fps readout operation at a resolution of 4K x 4K and achieving a DR of 110 dB and fine pixels of 2.7 μm. These characteristics were achieved simultaneously using coded exposure (CE), which divides the image plane into smaller blocks and controls the exposure time of each block individually. A high-speed readout was achieved by arranging analog to digital converters in parallel for each block. A high resolution was realized by integrating both circuits and pixels at a high density using three-dimensional wafer stacking technology. This system is expected to be applied to computational imaging using CE, in addition to applications that require high DR in which dark and bright areas coexist in a scene.

Key words ビジョンシステム，ブロック並列，積層型 CMOS イメージセンサ，高ダイナミックレンジ，高速撮影，符号化露光
vision system, block parallel, stacked CMOS image sensor, high dynamic range, high-speed imaging, coded exposure

1 Introduction

Image sensors are not only used for taking photos but also increasingly expected to serve as intelligent systems with surrounding configurations. Coded exposure (CE) [1], [2] is a method applied in intelligent system approaches, and

thus various functions can be realized by selecting an integration variable in the plenoptic function. High dynamic range (HDR) can be realized when the integration variable is time. Various methods have been proposed to achieve HDR. A method, such as a lateral overflow integration capacitor (LOFIC), was introduced that could provide the plurality of detection capacitors [3]. In addition, another method was presented to prevent the photodiode saturation by adding low-sensitivity pixels [4]. However, various proposed methods require an enlarged pixel size. Alternatively, a high-speed readout, such as an array-parallel analog-to-digital converter (ADC) structure [5], is useful for integrating multiple frames [6] to realize an HDR. However, it increases the noise level and requires a faster readout to reduce motion artifacts. To mitigate these adverse effects, a method was proposed in which a pixel array is divided into multiple blocks, and the signal integration time of each block is individually controlled [7]. In another method, CE was demonstrated using the pixel-level control of the exposure time [8]. However, in these methods, the readout path and control circuitry should be arranged within the same plane because these are unstacked sensors; thus, the pixel size is relatively large and high resolution is difficult to realize. Therefore, we designed a sensor that could simultaneously achieve a 4 K × 4 K resolution and a high-speed readout of 1000 fps. Using a stacked structure, we demonstrated the CE capability by individually controlling the exposure time for each block of the pixels.

2 Sensor Architecture

2.1. Block Diagram

Fig. 1 shows a conceptual diagram of the image sensor. This 1-inch image sensor has two layers: the top chip comprises BSI pixels with a 65-nm process and the bottom chip is used for sig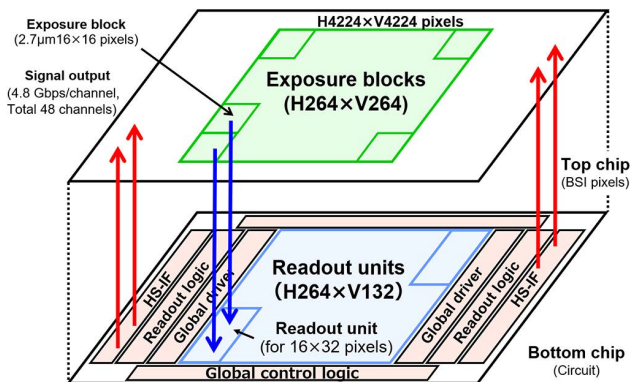nal processing; the layers are bonded to each other. The highest-density area consists of two contacts per pixel. The top chip has a pixel array (H4224 × V4224, 2.7-μm pitch), which is divided into H264 × V264 exposure control blocks with a basic unit of H16 × V16 pixels. The bottom chip has ADC circuits, logic circuits, high-speed interfaces (HS-IF), and H264 × V132 readout circuits (readout units), which are arranged directly below the pixels. Each readout unit corresponds to H16 × V32 pixels as a basic unit. Pixel signals are converted into parallel digital data using ADC circuits located in each readout unit. Binary conversion and correlated double sampling (CDS) operations are performed using the readout logic located in the bottom chip. Subsequently, signals are outputted through 48 channels of the HS-IF, which operates at 4.8 Gbps per channel.

2.2. Block Array Structure

Fig. 2 shows a diagram of the readout unit and two exposure blocks. Each pixel comprises a photodiode (PD) and five transistors. The reset transistor (RST) and select transistor (SEL) are controlled by a global pixel driver placed in the periphery of the bottom chip. Both the photoelectron transfer transistor (TX1) and PD reset transistor (TX2) are controlled for each block by the local pixel driver in the readout unit. The readout unit is composed of 16 column ADCs, a data transfer circuit, an integration time controller, and a local pixel driver. Moreover, ADC is a 12-bit single slope type, and the ramp signal and counter are supplied from the peripheral circuitry of the bottom chip. Each ADC is con-
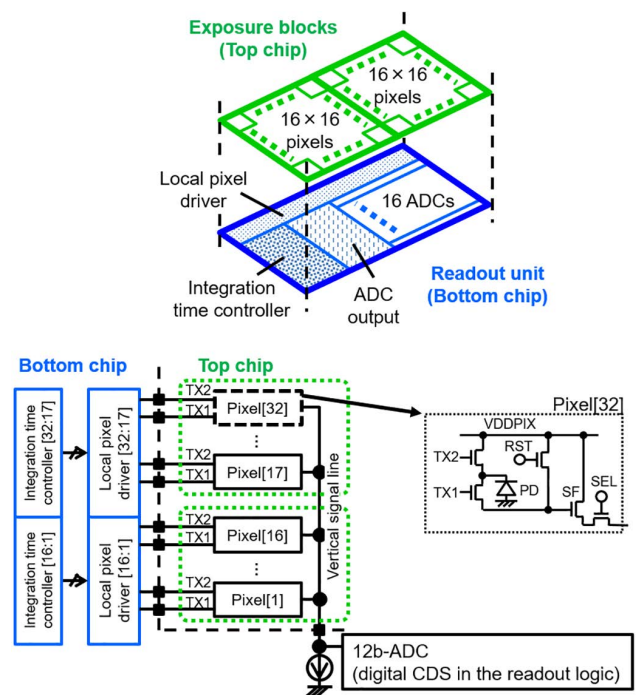

Fig. 1 Device structure

Fig. 2 Exposure block and readout unit

nected to its corresponding pixels (1–32) via a vertical signal line and source followers (SF). The digital signal, converted in ADC, is transferred to the readout logic via the transfer path for each set of 16 columns.

2.3. Exposure Control

The exposure control has two operation modes. One is "No Skip Mode" for short exposure time less than one frame, and the other is "Skip Mode" for long exposure time more than one frame. Fig. 3 shows a simplified block diagram and timing chart of exposure control. The integration time controller comprises a row counter, address decoder, and four flip-flops that hold an exposure time register setting. The register has four bits per block, of which three are used to encode eight different exposure times within one frame. The remaining one bit is a mode select signal, used as a mask signal that skips the reading of photoelectrons. The register can be updated for each frame. TX2 is controlled using the AND logic applied to the eight exposure time settings and timing signal linked with the register. TX1 is controlled using the AND logic of the TX1SEL signal from the global control logic and TX1MASK signal. The local pixel driver, composed of a level shift circuit and driver circuit, scans pixels in a predetermined row based on the decode signal of the integration time controller and SEL signal. Each readout unit is associated with 32 rows of pixels, corresponding to two exposure blocks per frame. In the No Skip Mode, TX1 is sequentially controlled to cross two exposure blocks, and TX2 is independently controlled for each block according to the integration time, as presented by blocks (1,1) and (1,2) in Fig. 3. Short exposure times of one horizontal period or less can be achieved because the controls for TX1 and TX2 are independent. Skip Mode can

be realized by skipping the TX1 and TX2 operations using the mask signal, as presented by block (264,263) in Fig. 3. With the above configuration, the exposure time of each block is individually set by the integration time controller, and the rolling shutter reading of 16 × 32 pixels in each unit is performed simultaneously. Each block has an exposure time controller that includes the selection of two operations. The exposure table is changed for each frame. Thus, various exposure patterns can be created in a two-dimensional manner, resulting in HDR imaging.

3 Configuration of the Experimental System

Fig. 4 shows the system configuration. This system comprises a camera head, camera control unit, and host computer (PC). The camera head has a lens unit and image sensor board. The camera control unit is composed of two field programmable gate arrays (FPGA) and a power supply board and functions to control the image sensor, supply power, and receive image data. The PC controls the entire system, processes, and then saves the images.

The image data from the sensor are inputted to each FPGA in the camera control unit via 24 channels at a time. The FPGA adjusts the data rate and data width and transfers data to the PC through the optical fiber.

To calculate the exposure time of each block, we prototyped two types of systems according to the application. The first system is to calculate the exposure time in FPGAs in the camera control unit, assuming that the subject moved at a high speed. The exposure value can be updated in every 8 frames by calculating the exposure value using the pipeline operation without a frame memory. The second system is to
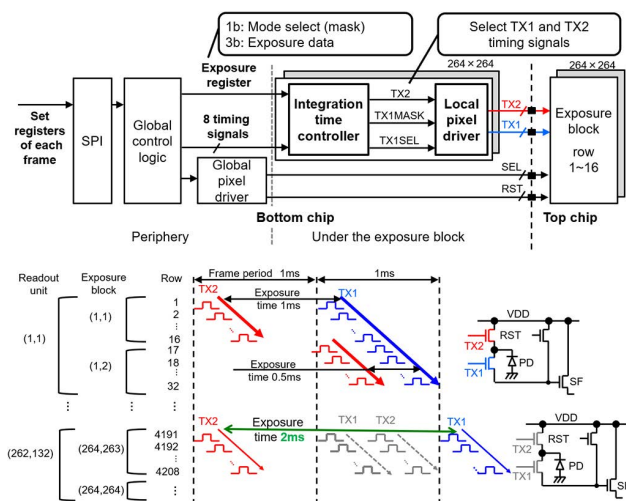

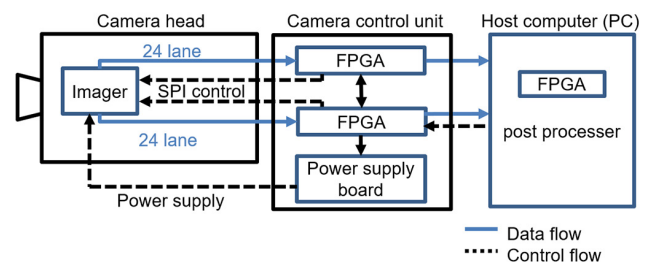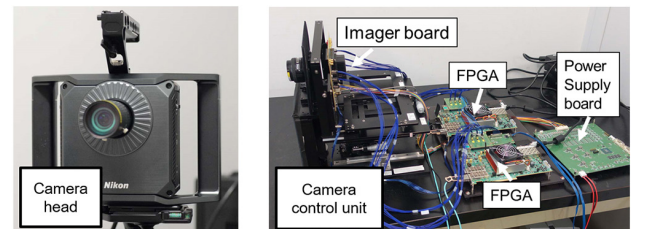Fig. 3   Block exposure control structure and timing chart

Fig. 4   Camera system architecture

calculate it in PC. More advanced and precise exposure control is possible in case of considering the surrounding blocks or multiple frames.

4 Results and Discussion

4.1. Coded Exposure

Fig. 5 presents the experimental results of CE. To generate an exposure table, the original image was first divided into blocks of 264 × 264, and the average value of each block was converted into 3-bit gradation data. Subsequently, the exposure time was set from 1/128 to 1 ms corresponding to the 3-bit data. The photography was performed on a light box without patterns under a uniform illumination using the exposure table. The block exposure method can produce a coded image in a single shot. Furthermore, setting the exposure time frame-by-frame or across frames yields various coded patterns.

4.2. HDR Imaging

Fig. 6 shows the experimental results for an HDR. The image shown on the left side was obtained by exposure bracketing, which is a conventional technique of taking multiple shots with different exposure times. Moreover, dynamic

range (DR) of each image was relatively small where blackouts and whiteouts occurred. Therefore, many images were required, which required a long time to capture. The image on the right side is a HDR image acquired with CE using this sensor and calculated in PC. The sensor was driven at 1000 fps, and the post-processing system and integration
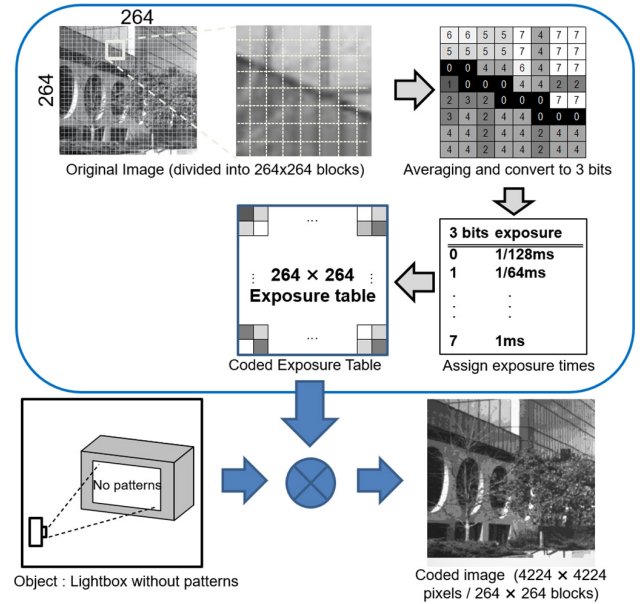

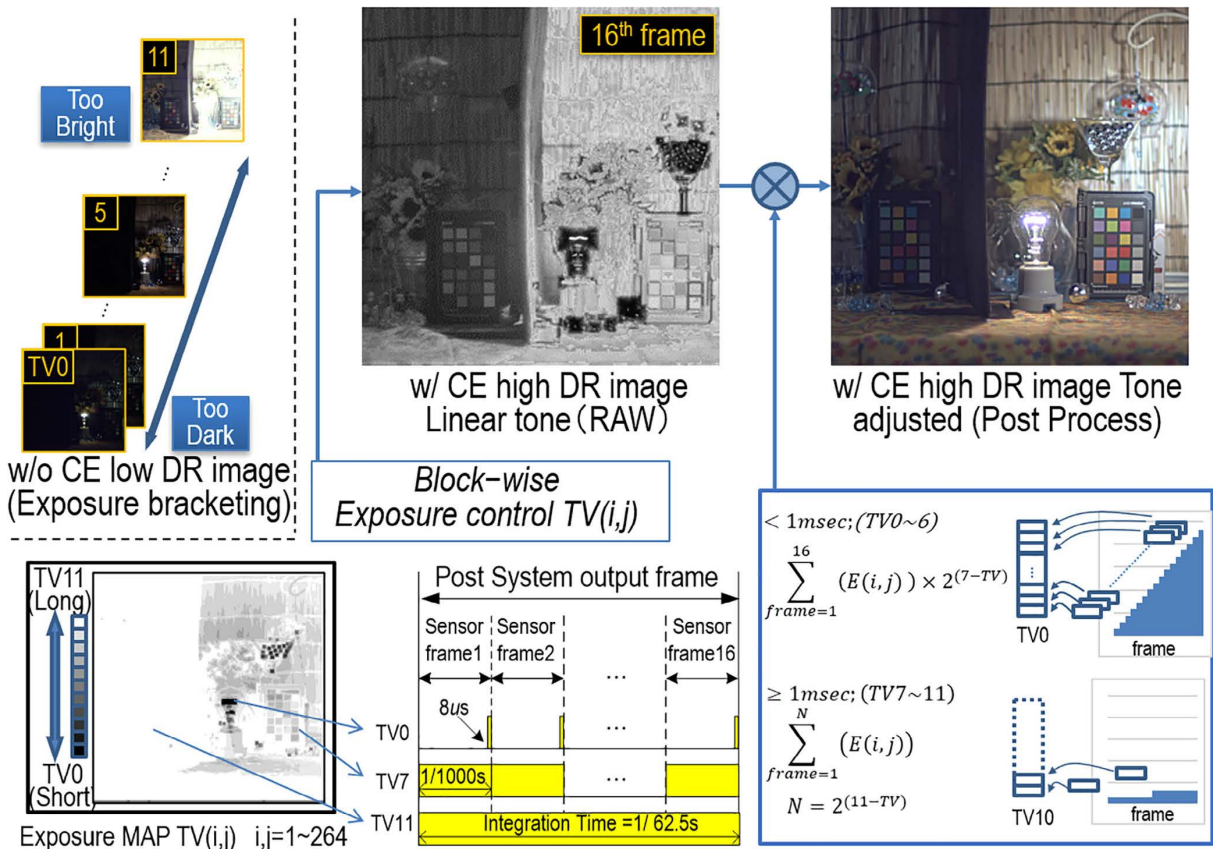Fig. 5　Experimental results of CE

Fig. 6　Experimental results of HDR

time operated in every 16 frames. Image acquisition was performed by changing the exposure time for each block based on the exposure table, indicated by time value (TV), which logarithmically defines the exposure time. The calculation performed multiplication and addition processing of the acquired image according to the TV. Consequently, images with a total of 11 exposure stops can be acquired at a high speed by setting the exposure time for 7 stops within one frame and 4 stops for over a period of 16 frames.

Considering the block steps, if the exposure times of the adjacent blocks are different, the signal-to-noise ratio (SNR) steps occur at the boundary of the blocks. To mitigate the block steps, smoothing the number of stops of the exposure time between adjacent blocks is preferable. In addition, the SNR can be improved by adding multiple frames to the blocks with a short exposure time.

4.3. HDR Tracking for Moving Object

Fig. 7 shows the results of the responsiveness of exposure control to a moving object. The sensor was operated at 1000 fps and an object moved horizontally on the screen. The image on the left was captured without an exposure control. When the object dashes from the dark to bright area of frame #345 with respect to the initial state of frame #1, the characters on the body are oversaturated. The results of dynamic exposure control in every 8 frames are demonstrated on the right side. Updating the exposure table in every 8 frames suppresses the oversaturation of characters on the body and realizes an HDR image.
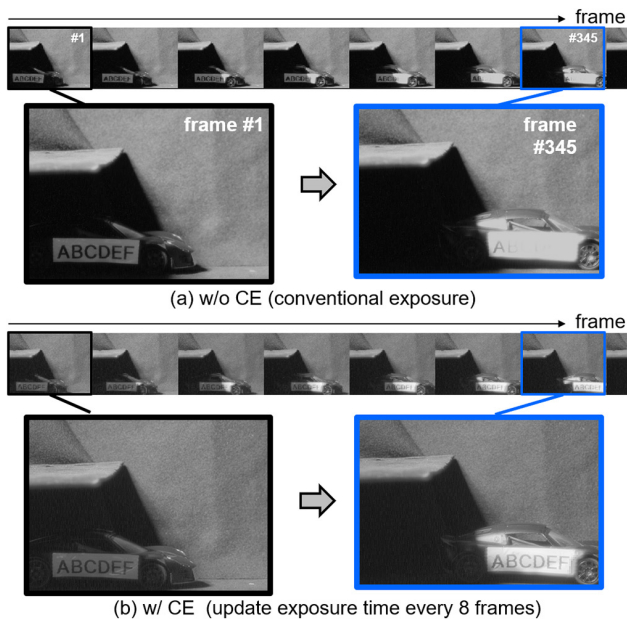
4.4. Sensor Specifications

Table 1 presents a comparison between the performance of our sensor with that of other sensors presented in the existing studies. At first, our developed sensor achieves high speed readout with a resolution of 17 Mpixel. In addition, this paper demonstrates HDR with small pixel of 2.7-um pitch. The total power of the sensor was 7.4 W at 1000 fps. The chip micrograph is shown in Fig. 8. Both chips were fabricated using a 65-nm process.

4.5. More Applications Using Coded Exposure

Fig. 9 shows an application that overlays the information

Table 1　Sensor specifications

	This work	IISW2019 [3]	ISSCC2020 [4]	VLSI2017 [5]	IISW2015 [7]	ISSCC2019 [8]
Process	Stacked BSI Top: 65nm Bottom: 65nm	BSI 65nm 1P4M	Stacked BSI Top: 90nm/65nm 1P4Cu Bottom: 40nm 1P6Cu1AL	Stacked BSI Top: 90nm 1P4M Bottom: 55nm 1P7M	FSI 0.18μm 1P4M	0.11μm
HDR technology	Integration time controllable for each block	LOFIC	Sub-pixel	-	Integration time controllable for each block	Integration time controllable for each pixel
Number of pixels [Mpix]	17.8	0.6	5.7	4.1	0.4	0.05
Pixel pitch [μm]	2.7	2.8	3.0	4.8	5.0	11.2
Bit depth [bit]	12	-	12	-	-	-
Frame rate [fps]	1000	-	30	630	-	25
Conversion gain [μV/e-]	161	10 (Low) 160 (High)	6.7 (Low) 197 (High)	65	-	-
Random noise [e-]	2.9	-	0.6	4.2	-	-
Sensitivity [ke-/lux·s]	20.7	-	38.0	28.4	-	-
FWC[ke-]	7.4	120.0	165.8	-	-	-
Dynamic-range [dB]	110 @ 1 frame 134 @ 16frames	-	132	-	120	-

Fig. 8　Chip micrograph

(a) w/o CE (conventional exposure)

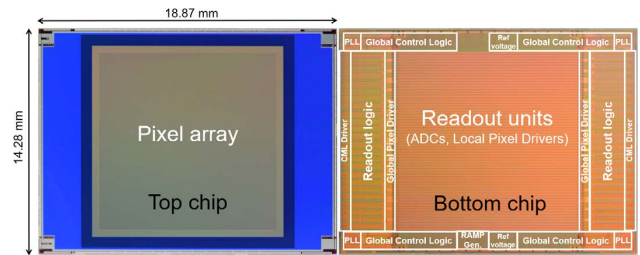(b) w/ CE (update exposure time every 8 frames)

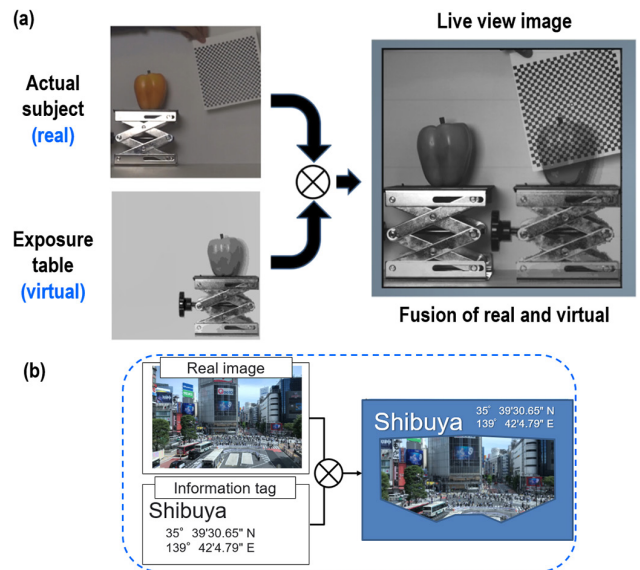Fig. 7　Experimental results of exposure tracking

Fig. 9　(a) Experimental results of the overlay information on the real image and (b) an application example of CE

on a real image. Encoding the exposure time for each region implies that an arbitrary virtual pattern can be generated in the image. The experimental result is shown in Fig. 9(a). The photograph on the upper left shows the actual subject. The lower left figure presents an image of the exposure table, where a pseudo image is formed by expressing eight different exposure settings as gradations. The dark area indicates a short exposure, whereas the bright area indicates a long exposure. The results obtained using this exposure map are illustrated on the right side. It can be observed that a virtual image is embedded in the subject in the real space. In addition, this exposure table is created based on an actual subject; that is, it can also be used as an image-recording function.

Fig. 9(b) shows an application, where it is possible to embed tag information or a head-mounted display for virtual or augmented reality [9], [10]. Overlaying them on the image sensor reduces the processing power of the subsequent system and improves latency.

5 Conclusions

In this study, we have newly developed a CMOS image sensor with a stack structure that operates at 1000 fps while having a high resolution of 17 Mpixel with small pixel of 2.7-um pitch. By using the block-wise coded exposure function, 110 dB DR is achieved at 1000 fps in a single frame, and 134 dB DR in 16 frames. This vision system can be applied to various computational imaging using the coded exposure function, in addition to high dynamic range imaging in scenes where dark and bright areas of the subject are mixed in the frame.

References

[1] R. Raskar, et al., "Coded exposure photography: Motion deblurring using fluttered shutter," *ACM SIGGRAPH* 2006 Papers, SIGGRAPH '06, 2006, pp. 795–804.

[2] Y. Hitomi, et al., "Video from a single coded exposure photograph using a learned over-complete dictionary," *IEEE ICCV*, 2011, pp. 287–294.

[3] K. Miyauchi, et al., "A high optical performance $2.8\,\mu m$ BSI LOFIC pixel with 120ke FWC and 160 µV/e conversion gain," in *Proc. Int. Image Sensor Workshop*, 2019.

[4] Y. Sakano, et al., "A 132dB single-exposure-dynamic-range CMOS image sensor with high temperature tolerance," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2020, pp. 106–108.

[5] T. Takahashi, et al., "A 4.1Mpix 280fps stacked CMOS image sensor with array-parallel ADC architecture for region control," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2017, pp. 244–245.

[6] S. Velichko, et al., "140 dB dynamic range sub-electron noise floor image sensor," in *Proc. Int. Image Sensor Workshop*, 2017.

[7] A. Peizerat, et al., "A 120dB DR and 5µm pixel pitch imager based on local integration time adaptation," in *Proc. Int. Image Sensor Workshop*, 2015.

[8] N. Sarhangnejad, et al., "Dual-tap pipelined-code-memory coded-exposure-pixel CMOS image sensor for multi-exposure single-frame computational imaging," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2019, pp. 102–104.

[9] J. Wang, et al., "Augmented reality navigation with automatic marker-free image registration using 3-d image overlay for dental surgery," *IEEE Transaction on Biomedical engineering*, 2014, vol. 61, no. 4, pp. 1295–1304.

[10] H. Hile and G. Borriello, "Positioning and orientation in indoor environments using camera phones," *IEEE Computer Graphics and Applications*, 2008, vol. 28. no. 4, pp. 32–39.

[11] T. Hirata, et al., "A 1-inch 17Mpixel 1000fps block-controlled coded-exposure back-illuminated stacked CMOS image sensor for computational imaging and adaptive dynamic range control," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 120–121.

平田友希　Tomoki HIRATA
先進技術開発本部　光デバイス開発センター
Opto Device Development Center
Advanced Technology Research & Development Division

村田寛信　Hironobu MURATA
先進技術開発本部　光デバイス開発センター
Opto Device Development Center
Advanced Technology Research & Development Division

有井　卓　Taku ARII
先進技術開発本部　光デバイス開発センター
Opto Device Development Center
Advanced Technology Research & Development Division

松田英明　Hideaki MATSUDA
先進技術開発本部　光デバイス開発センター
Opto Device Development Center
Advanced Technology Research & Development Division

米持　元　Hajime YONEMOCHI
先進技術開発本部　光デバイス開発センター
Opto Device Development Center
Advanced Technology Research & Development Division

手塚洋二郎　Yojiro TEZUKA
先進技術開発本部　光デバイス開発センター
Opto Device Development Center
Advanced Technology Research & Development Division

綱井史郎　Shiro TSUNAI
先進技術開発本部　光デバイス開発センター
Opto Device Development Center
Advanced Technology Research & Development Division

平田友希
Tomoki HIRATA

村田寛信
Hironobu MURATA

有井　卓
Taku ARII

松田英明
Hideaki MATSUDA

米持　元
Hajime YONEMOCHI

手塚洋二郎
Yojiro TEZUKA

綱井史郎
Shiro TSUNAI