NOVEL IMAGE SENSOR WITH AREA-BASED OPTIMISATION OF SHOOTING CONDITIONS FOR IMMERSIVE CONTENT PRODUCTIONS

K. Kikuchi[1], K. Tomioka[1], T. Usui[1], A. Honji[1], K. Kitamura[1], and S. Kawahito[2]

[1]NHK, Japan and [2]Shizuoka university, Japan

ABSTRACT

We present a novel scene-adaptive imaging technology designed to enhance the image quality of wide-angle immersive videos such as 360-degree videos. It addresses the challenge of balancing resolution, frame rate, and dynamic range due to sensor limitations by dynamically adjusting shooting conditions within a single frame on the basis local subject characteristics. This involves capturing still subjects at high resolution and moving subjects at increased frame rates, adjusting exposure time according to subject brightness while maintaining pixel readout rate. To validate this approach, we developed a block-wise-controlled image sensor prototype with 1.1 million pixels that enables flexible control of shooting conditions individually for 272 separated blocks. Real-time scene analysis and a feedback control system were also developed. Experimental results demonstrate that the proposed method improves subjective image quality compared with conventional imaging that captures the entire frame under a single shooting condition, even at the same data rate.

INTRODUCTION

The global demand for highly immersive video content, such as 360-degree videos and dome screen videos, is escalating. Accompanying this demand is an increasing need for cameras capable of capturing wide viewing angles effectively (e.g. panoramic cameras and omnidirectional multi-cameras). Wide-angle videos typically feature subjects exhibiting diverse textures, movements, and brightness on a single screen, requiring image sensors to meet rigorous performance quality, including not only resolution and frame rates exceeding ultra-high definition television levels (see [1]) but also excelling in dynamic range for incident light. However, developing an image sensor that fulfils all these requirements simultaneously is challenging. Traditional image sensors, such as Complementary Metal Oxide Semiconductor (CMOS) image sensors operating under constant shooting conditions across the entire pixel array, are limited by a trade-off between resolution, frame rate, and the noise performance related to dynamic range (El-Desouki *et al* [2] and Kawahito [3]). Moreover, higher pixel readout rates lead to increased data transfer streams and higher power consumption in image sensors.

On the other hand, from the perspective of improving subjective image quality, uniformity in the sensor's shooting conditions across the screen appears dispensable. For example, areas with still subjects may benefit from high resolution, whereas those with moving subjects may require ensuring temporal resolution rather than spatial resolution to minimise motion blur. In addition, those with high brightness do not necessitate achieving precise dark gradation, and those with low brightness do not require high pixel saturation. Acknowledging this, we propose a new shooting approach, scene-adaptive imaging that facilitates dynamic control of the shooting conditions on the basis of local subject characteristics. To implement this method, a novel image sensor capable of flexibly controlling shooting conditions individually for each separated area on a pixel array is a requisite. This paper presents a Block-Wise-Controlled CMOS Image Sensor (BWC-CIS) prototype with 1.1 million pixels separated into 272 shooting-condition-controllable blocks. We conducted a simulated shooting experience of the developed sensor by using a real-time scene analysis and feedback control system, demonstrating enhancement in subjective image quality achieved through area-based optimisation of shooting conditions.

Related Works

Several methods have been proposed to tackle the challenges involved in enhancing the dynamic range of CMOS image sensors. For instance, several single-shot technologies employ a sensor with a dedicated pixel structure (Miyauchi *et al* [4] or Sakano *et al* [5]), others utilise a pixel-wise digital conversion strategy (Ikeno *et al* [6]) or a block-wise exposure control logic (Hirata *et al* [7]) in a sensor. Although the latter is particularly aligned with our approach, these methods present manufacturing difficulties due to their complex pixel structures or logic circuits, and difficulties in scaling up their configurations. In contrast, our method enables the development of a sensor with relatively simple configurations. Furthermore, it introduces the unique concept of optimising image quality relative to data-rate by flexibly controlling not only dynamic range, but also resolution and frame rate, which provides ease of scalability.

OVERVIEW OF SCENE-ADAPTIVE IMAGING

The concept of scene-adaptive imaging is shown in Figure 1, which briefly describes how different areas of a scene can be captured with various shooting conditions (imaging modes) in a BWC-CIS, as following.

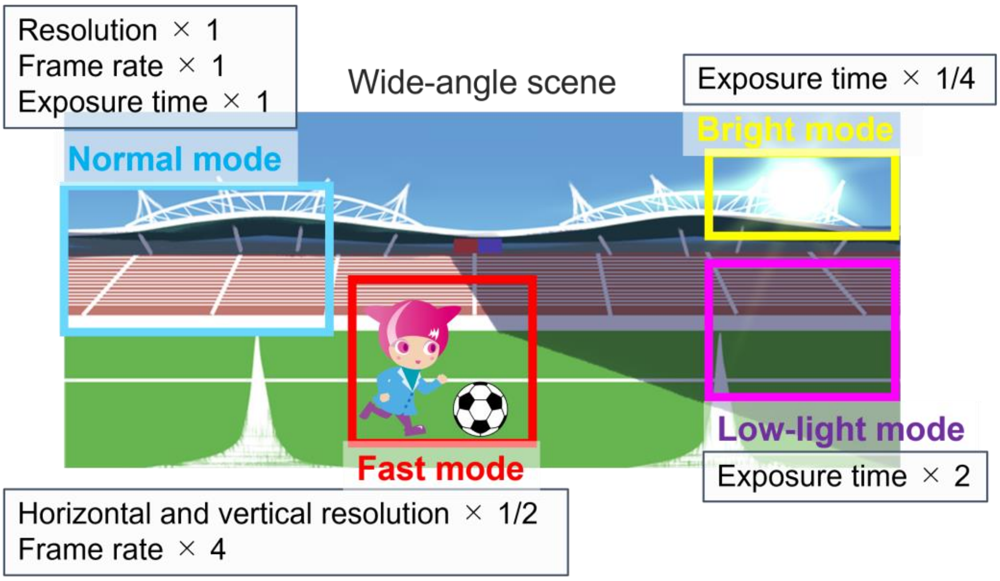

Figure 1 – Overview of scene-adaptive imaging

- **Normal mode** captures still and detailed subjects with standard resolution (maximum resolution based on the sensor's pixel array) and standard frame rate.
- **Fast mode** captures fast moving subjects with horizontally and vertically half-reduced resolution (reduced 1/4 in total) and four times faster frame rate.
- **Bright mode** captures highly bright subjects with one-fourth shorter exposure time.
- **Low-light mode** captures dark subjects with two times longer exposure time with two times lower frame rate.

The exposure time in Bright and Low-light modes exemplifies our current implementation, where both shorter and longer exposure times are adjustable. Note that the pixel readout rate remains consistent in Normal and Fast modes. Consequently, the proposed method addresses the sensor's limitation challenge by optimising local shooting conditions of resolution, frame rate, and dynamic range, which improves image quality without escalating data rates. To fully harness the potential of the BWC-CIS, the concurrent utilisation of a scene analysis and feedback control system dedicated for the sensor is needed. As shown in Figure 2, this system determines the optimal imaging modes of the sensor by analysing the subject's brightness distribution and movement with low latency and subsequently feeds back the information of determined imaging modes to the sensor. In the following sections, we provide details of the developed experimental imaging systems.

BLOCK-WISE-CONTROLLED IMAGE SENSOR

To evaluate the proposed method, we developed a monochrome BWC-CIS prototype grounded in CMOS image sensor technology (Tomioka et al (8)). As shown in Figure 3, our sensor comprises 1,024 × 1,088 pixels with a pixel pitch of 2.6 µm, a mode control circuit, pixel drive circuit, Analogue-to-Digital Converter (ADC) circuit, and output circuit. The right part of Figure 3 shows the pixel structure of our sensor, where each element circuit shares a pixel amplifier for four pixels (A, B, C, and D). Although this structure closely resembles typical CMOS image sensors, our sensor distinguishes itself by incorporating the mode control circuit and selection switches. These additional components can manage controlling pixel signal charge, reset, and readout of four pixels by the control signal regulated individually for each block of 64 × 64 pixels (control block). This configuration results in 16 × 17 control blocks across the pixel array, facilitating flexible control of imaging modes in 272 areas of the sensor, where the effective counts of pixels and control blocks are 960 × 960 pixels and 15 × 15 blocks, respectively, due to the presence of light shielding (optical black) pixels.

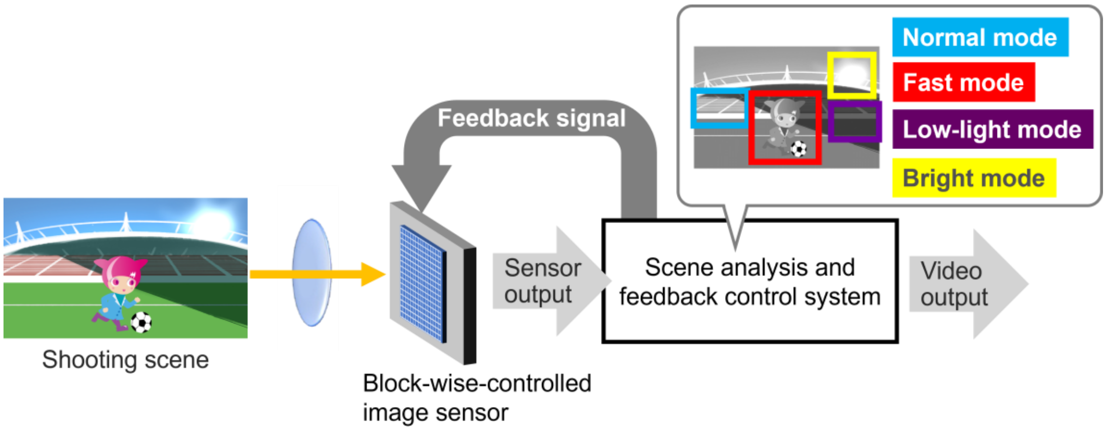

Figure 2 – System configuration of scene-adaptive imaging

Figure 4 presents two types of pixel scanning methods employed in our sensor: (a) sub-pixel readout and (b) pixel binning readout. The former sequentially applies readout pulses to charge transfer switches for each A, B, C, and D pixel within a control block at 240 frames per second (fps) and reads out the pixel signal through a pixel amplifier, resulting in full resolution at 60 fps (64 × 64 pixels at 60 fps). Conversely, the latter reads out the signal after combining signals from all four pixels (2 × 2 pixel binning) in a single scan. Although the exposure time per pixel scan is one-fourth compared with sub-pixel readout, the four-pixel binning compensates for the reduction in signal value. With this offsetting outcome, the frame rate can be quadrupled while the horizontal and vertical resolution is halved (32 × 32 pixels at 240 fps).

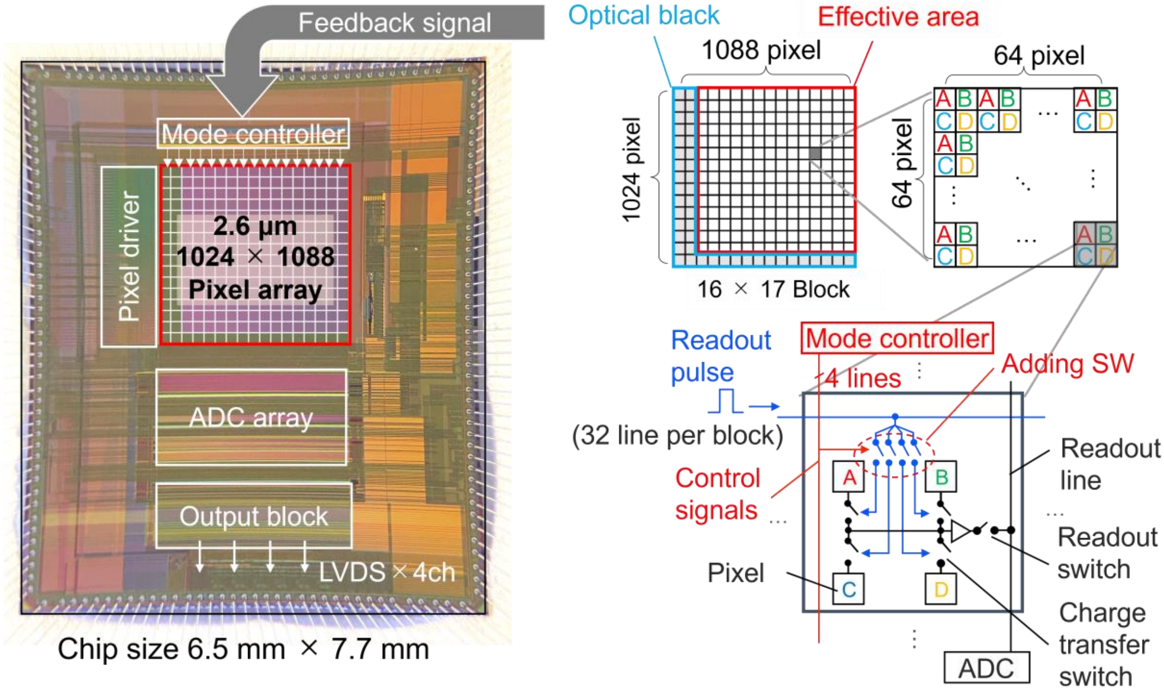

Figure 3 – Die photograph (left) and pixel architecture (right) of our sensor

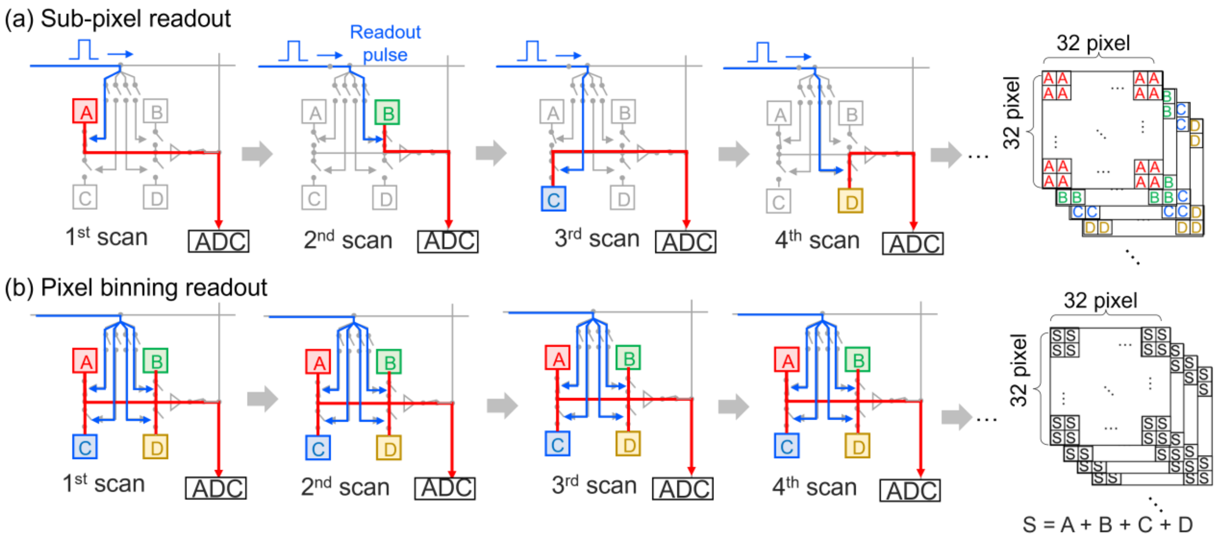

Figure 4 – Readout methods of our sensor

Table 1 and Figure 5 display the four imaging modes assignable in our sensor and their corresponding operational flow within a horizontal block line, respectively. Normal mode operation is achieved by using the sub-pixel readout, whereas Fast mode employs the comparable pixel binning readout. The sub-pixel readout also facilitates in Bright mode, albeit with a sequential reset of A, B, C, and D pixel charges per scan, constraining each pixel's exposure time to 1/240 second to avoid pixel saturation. Low-light mode alternates sub-frame readout with a four-scan pause to extend exposure to 1/30 seconds (30 fps) to enhance pixel sensitivity. The scene-adaptive imaging is achieved by switching these operations for the entire control blocks in response to feedback signals.

Mode	Resolution (per control block)	Frame rate	Exposure time
Normal	64 x 64 pixels	60 fps	1/60 s
Fast	32 x 32 pixels	120 fps	1/240 s
Bright	64 x 64 pixels	60 fps	1/240 s
Low-light	64 x 64 pixels	30 fps	1/30 s

Table 1 – Imaging modes assignable in our sensor

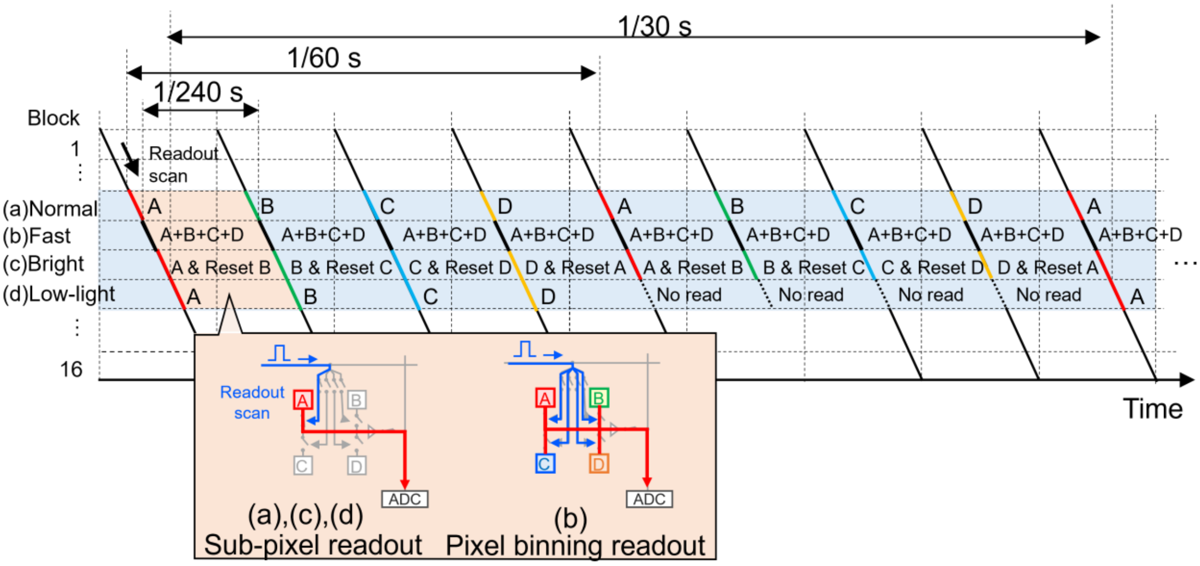

Figure 5 – Operation flow corresponding imaging modes in our sensor

SCENE ANALYSES SYSTEM

By integrating the fundamental imaging functions of the BWC-CIS prototype into Field Programmable Gate Array (FPGA) devices for real-time processing, we developed a scene-adaptive imaging experimental system. Figure 6 shows the comprehensive set-up of the system on the upper side and its photographs on the lower side, comprising the developed sensor installed on a sensor driving board, optics including a beam splitter, a sub-sensor (the Event Vision Sensor, described below), and two FPGA boards designed for scene analysis and feedback control and for video signal processing. This system was devised to analyse the brightness distribution and motion detection for subjects independently, and then merge the results to provide feedback signal to the BWC-CIS. Figure 7 shows the processing pipeline for the scene analysis processing. The followings detail specific functionalities of each component.

Brightness Distribution Analysis

The brightness distribution analysis is conducted by directly examining the BWC-CIS's output. The approach is outlined in the following, maximizing the exploitation of the sensor's scanning method to achieve analysis with minimal latency.

1. Subdividing the control block signals from 3rd scan (sub-pixel C or S in Figure 4, 32 × 32 pixels) into smaller segments (8 × 8 pixels, totalling 16 segments).
2. Averaging and applying thresholding to each segment, resulting a brightness distribution map that categorises the segments into three labels (high, middle, and low).
3. Determining the optimal mode regarding exposure time (i.e. Normal, Bright, or Low-light mode) for each control block through a mode filter applied to the map (e.g., Bright mode is assigned to the block exhibiting the highest frequency occurrence of "high").

These series of processes are designed to be completed within 1/240 seconds in the FPGA after acquiring the signal from the 3rd scan.

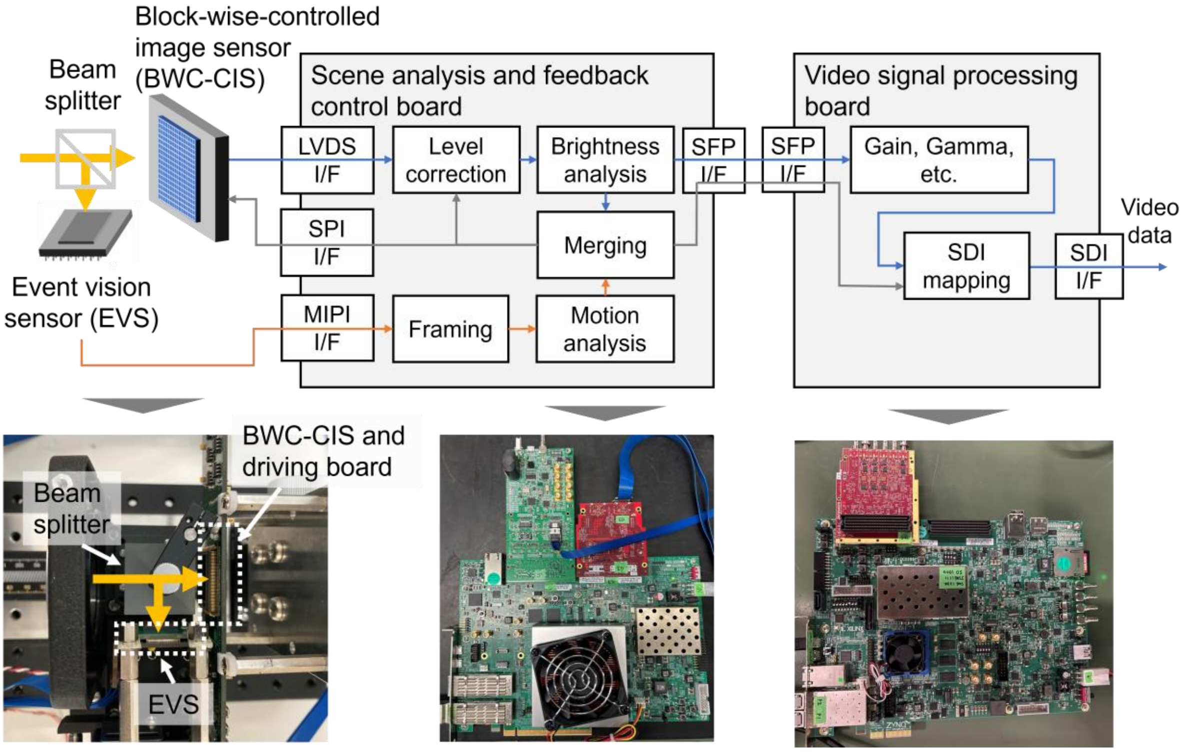

Figure 6 – Block diagram (upper) and photographs (lower) of the experimental system

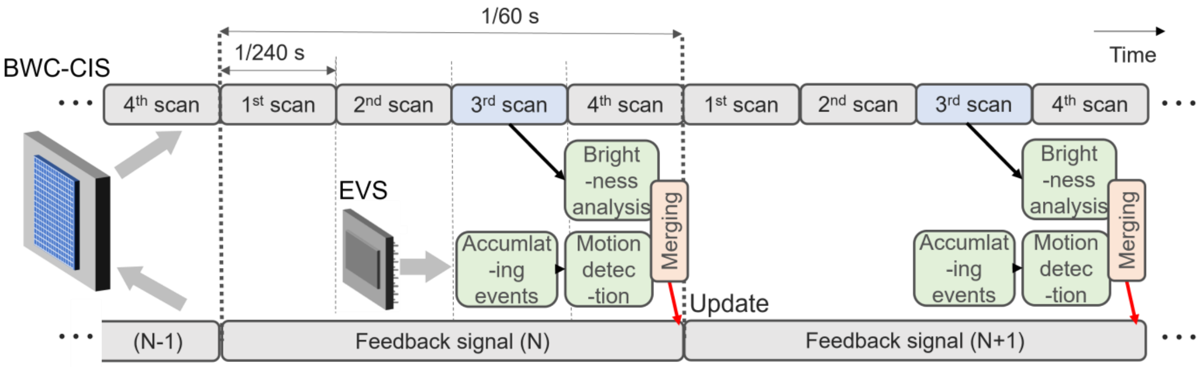

Figure 7 – Processing pipeline of scene analysis and feedback control system

Motion Detection Analysis

Detecting the subject’s motion domain with minimal delay poses a significant challenge. The widely-used motion detection methods, such as background separation (Piccardi [9]), require buffering multiple images for processing, resulting in delayed detection output. Given the system's need to consistently determine the subject’s motion domain during shooting, minimising processing delays is crucial. Therefore, as shown in the optics of Figure 6, we aimed to detect motion with the aid of the Event Vision Sensor (EVS, Finateu *et al* [10]) that is incorporated into the same light path of the BWC-CIS by using the beam splitter. An EVS detects pixel-level changes as events and transmits data in ultra-low latency, enabling us to promptly obtain cues for the subject’s motion. The processing flow with the acquired event follows the steps following:

1. Accumulating events from the EVS for a 3rd scan period and generating binary images (1: with event, 0: without event).
2. Filtering out isolated events (eliminating false events arising from shot noise, etc.).
3. Calculating the total number of events for each segment corresponding to the control block of the BWC-CIS.
4. Applying thresholding to the control blocks and classifying them as a motion area (preferred area for Fast mode) or not.

Similar to the brightness distribution analysis, this procedure is executed within 1/240 seconds after accumulating events on an FPGA.

Finally, combining the preceding two analyses establishes imaging modes for the subsequent scans of A, B, C, and D pixels in the entire sensor, where the priority of selecting modes regarding brightness (Normal, Bright, and Low-light modes) or motion (Fast mode) can be set arbitrarily. This mode information is promptly transmitted to the sensor via feedback signals at the appropriate timing. The comprehensive operation enables continuous real-time updates of the sensor's shooting conditions every 1/60 seconds and enables it to respond to transitions in the local brightness or movement of the scene with a minimal delay. Additionally, our system performs a signal level correction of data from the BWC-CIS, compensating for sensitivity differences due to variations in exposure times between Normal, Bright, and Low-light modes by multiplying the pre-calculated correction coefficients.

EXPERIMENTAL RESULTS

Shooting experiments were conducted using the developed system to assess its performance. The experiment involved still subjects with varying local luminance and rotating radial charts. Images were acquired using a lens with an f-number of 4.0 and a focal length of 17 mm, and were recorded using an uncompressed serial digital interface (SDI) recorder. Since the spatial-temporal mixed sample structures of the raw data from our sensor made it difficult to handle as a conventional raster image, it was rendered into a final output format of 1K × 1K pixels at 240 fps video, wherein areas captured in Fast mode had their horizontal and vertical resolution doubled, while those in other modes were interpolated four times finer in temporal resolution. For comparison, the experiment also included images captured when all control blocks were set in Normal mode to simulate the conventional imaging method.

Figure 8 shows the imaging results. With the conventional method, a reduction in spatial resolution due to motion blur was observed in area (a), where the rotating subject was captured. Additionally, clipped whites and crushed shadows were observed in areas (b) and (c), where bright specular reflection subjects or subjects hidden in the shadows were located, respectively. On the other hand, the results of proposed method demonstrated that shooting conditions were suitably adjusted for each subject, as depicted in Figure 8. This resulted in improvements in motion blur in region (a'), even though the spatial resolution is reduced in Fast mode, as well as improvements in overexposure and underexposure in regions (b') and (c'), respectively. However, it can be observed that the proposed method may result in blocking artifacts at the imaging mode boundaries, as shown in region (d). To address this issue, potential solutions include developing CIS with finer control blocks or implementing post-processing techniques to smooth the boundaries.

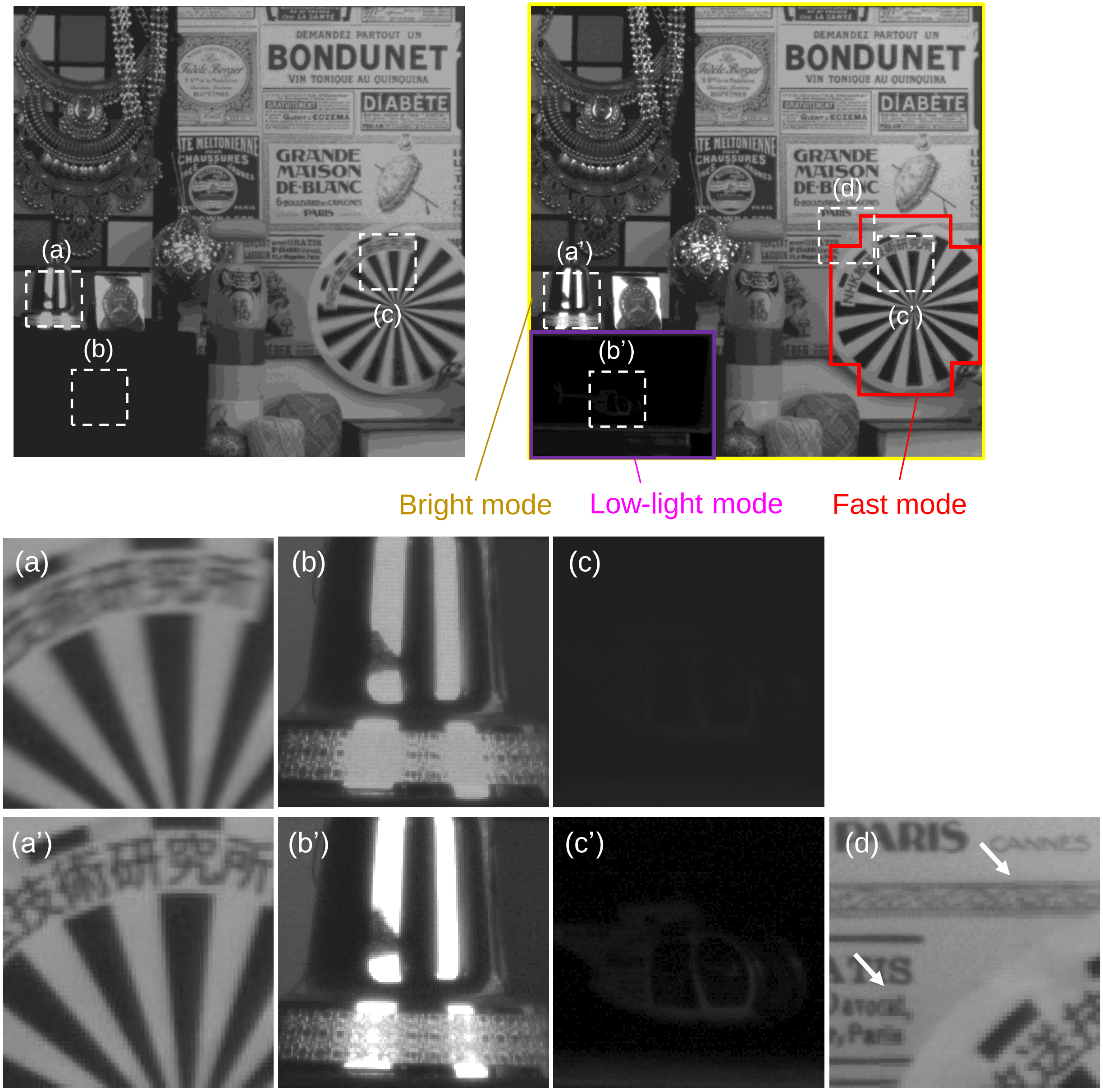

Figure 8 – Comparison of images acquired by the proposed and conventional methods and their mode settings (upper) and comparison of enlarged images (lower).

It is essential to emphasise that the sensor's output rates are the same for both the proposed and conventional methods. Consequently, although there is room for improvement in resolution and the granularity of control block size in our sensors, the experimental results indicate its advantages in enhancing subjective image quality. This also indicates that the performance of CMOS image sensors can be drastically extended in a relatively straightforward manner to meet the growing demand for stringent specification requirements in immersive video content production.

CONCLUSIONS

We proposed a scene-adaptive imaging method designed to overcome the limitations of conventional image sensor technology by adjusting shooting conditions on the basis of local subject features, with the aim of achieving high resolution, high frame rate, and high dynamic range simultaneously. To validate the principle of our approach, a block-wise-controlled image sensor was developed, capable of selecting different resolutions, frame rates, and exposure times in units of 64 × 64 pixels over a 1K × 1K pixel array. Through imaging experiments combining this sensor with a real-time scene analysis system, we confirmed the enhancement in subjective image quality achieved by locally optimising shooting conditions on the basis of subjects. Our next objectives include developing a practical sensor with higher resolution and more flexible imaging adjustability, as well as constructing a wide-field camera system for high-quality immersive video production.

REFERENCES

1. ITU-R, 2019. Video parameter values for advanced immersive audio-visual systems for production and international programme exchange in broadcasting. Recommendation ITU-R BT.2123-0.

2. El-Desouki M., Deen M. J., Qiyin F., Liu L., Tse F., and Armstrong D., 2009. CMOS Image Sensors for High Speed Applications. Sensors 2009. 9(1). pp. 430 to 444.

3. Kawahito S., 2018. Column-Parallel ADCs for CMOS Image Sensors and Their FoM-Based Evaluations. IEICE Trans. Electron., vol. E101-C, no. 7. pp. 444 to 456.

4. Miyauchi K., Okura S., Takayanagi I., Nakamura J., and Sugawa S., 2019. A High Optical Performance 2.8μm BSI LOFIC Pixel with 120ke-FWC and 160 μV/e^- Conversion Gain. Int. Image Sensor Workshop, 2019. R30.

5. Sakano Y., Toyoshima T., Nakamura R., Asatsuma T., Hattori Y., Yamanaka T., Yoshikawa R., Kawazu N., Matsuura T., Iinuma T., Takahiro T., Watanabe T., Suzuki A., Motohashi Y., Asami J., Tateshita Y., and Haruta T., 2020. A 132dB Single-Exposure-Dynamic-Range CMOS Image Sensor with High Temperature Tolerance. IEEE Int. Solid-State Circuits Conf., 2020. pp. 106 to 108.

6. Ikeno R., Mori K., Uno M., Miyauchi K., Isozaki T., Takayanagi I., Nakamura J., Wuu S., Bainbridge L., Berkovich A., Chen S., Chilukuri R., Gao W., Tsai T., and Kiu C., 2022. A 4.6-μm, 127-dB Dynamic Range, Ultra-Low Power Stacked Digital Pixel Sensor With Overlapped Triple Quantization. IEEE Trans. Electron Devices, vol. 69 no. 6. pp. 2943 to 2950.

7. Hirata T., Murata H., Matsuda H., Tezuka Y., and Tsunai S., 2021. A 1-inch 17Mpixel 1000fps Block-Controlled Coded-Exposure Back-Illuminated Stacked CMOS Image Sensor for Computational Imaging and Adaptive Dynamic Range Control. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Vol. 64. pp.120 to 122.

8. Tomioka K., Kikuchi K., Usui T., Kitamura K., and Kawahito S., 2023. Feedback Control of a Block-Wise-Controlled Image Sensor Based on Brightness Distribution Analysis. Int. Image Sensor Workshop, 2023. p.25.
9. Piccardi M., 2004. Background subtraction techniques: a review. Proc. IEEE Int. Conf. Systems, Man and Cybernetics, vol. 4. pp. 3099 to 3104.
10. Finateu T., Niwa A., Matolin D., Tsuchimoto K., Mascheroni A., Reynaud E., Mostafalu. P., Brady F., Chotard L., LeGoff F., Takahashi H., Wakabayashi H., Oike Y., and Posch C., 2020. A 1280×720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86µm Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline. IEEE Int. Solid-State Circuits Conf., 2020. pp. 112 to 114.