

A 1.2Mp indirect-ToF sensor with on-chip ISP for low-power and self-optimization

Seung-Chul Shin, Jiheon Park, Daeyun Kim, Myoungoh Ki, Myunghan Bae, Hoyong Lee, Jonghan Ahn, Myeonggyun Kye, Bumsik Chung, Inho Song, Sunhwa Lee, Il-Pyeong Hwang, Taemin An, Jaecil An, Min-Sun Keel, Young-Gu Jin, Youngchan Kim, Youngsun Oh, Juhyun Ko, JoonSeo Yim

System LSI Division, Samsung Electronics Co., Ltd., Hwaseong, Gyeonggi-do, Korea,
E-mail: sc1225.shin@samsung.com

Abstract — We propose an i-ToF to improve power consumption and motion artifact. AR/VR and Robot devices that require depth information when implementing applications have limitations when using i-ToF due to limited power and moving motion characteristics. In particular, the increase in the amount of AP operation according to the implementation of high resolution and the limitation of latency time due to multi-frame operation occur. The proposed sensor has an embedded depth processor to process depth calculation inside the sensor and enables low-power realization by using a 28nm process. In addition, motion artifact was reduced by operating at 250FPS readout speed. As a result, 30FPS depth output can be output from the sensor itself, and power consumption is reduced by 90%.

I. INTRODUCTION

Recently, 3D information is being used as essential information for machine vision such as 3D map generation and gesture control in AR/VR and robot with RGB image. These devices require low-power operation due to its battery-based operation, and since continuous movement occurs, reduced motion and constant FPS are required. In addition, high-resolution depth information using indirect-ToF (i-ToF) is mainly considered because high-resolution depth information is required due to the characteristics of applications that use surface information. In previous studies, depth performance improvement and distance increase were proposed through demodulation contrast (DC) improvement, modulation frequency increase and tap mismatch calibration [1-4]. However, since all depth processing is software-based in the application processor (AP), there is a limit to the reduction of the overall system power. Also, depending on the performance of the AP, the FPS and depth quality is varying. In addition, since multiple sensors operate simultaneously for tracking and detection, the amount

of data transmitted to the AP and the computational burden further increase. As another issue, since the external light conditions are accumulated during the exposure time, reduced operating distance, and increased noise in outdoor environment compared to the indoor. To solve this problem, a method of converting analog binning and digital binning through external control to implement signal increase and FWC increase effects, respectively, has been proposed. [5] However, there arises a problem that a sensor for determining the external light situation is added or the operation mode needs to be changed through additional calculations in the AP. In this paper, the following three improvements are proposed sensor with the embedded depth processor unit (eDPU) built into the sensor: 1) Reduction of system power, 2) Minimization of motion artifact, 3) Reduction of influence of external conditions.

II. BASIC OPERATION

The i-ToF is to calculate the distance through the phase delay of the modulated transmitter signal reflected back to the object. For this, phase information divided into four per cycle is required, and it is sensed by separating it into each tap inside the one pixel. This feature will reduce noise and improve accuracy. However, each tap has gain and offset errors due to the mismatch during fabrication. To compensate for this, continuous 2-frame information is generally used as tap shuffle is applied in additional frames. In addition, due to the periodicity of the modulation, the maximum operating distance is limited by frequency (fm) as a folding error. To solve this problem, the actual distance is determined through greatest common divisor (GCD) operation by cross-operating the two or more fm. As a result, when calculating i-ToF based processing, 2 to 4 multi frame based calculations are used. As a result, latency increases and the amount of required computations increases.

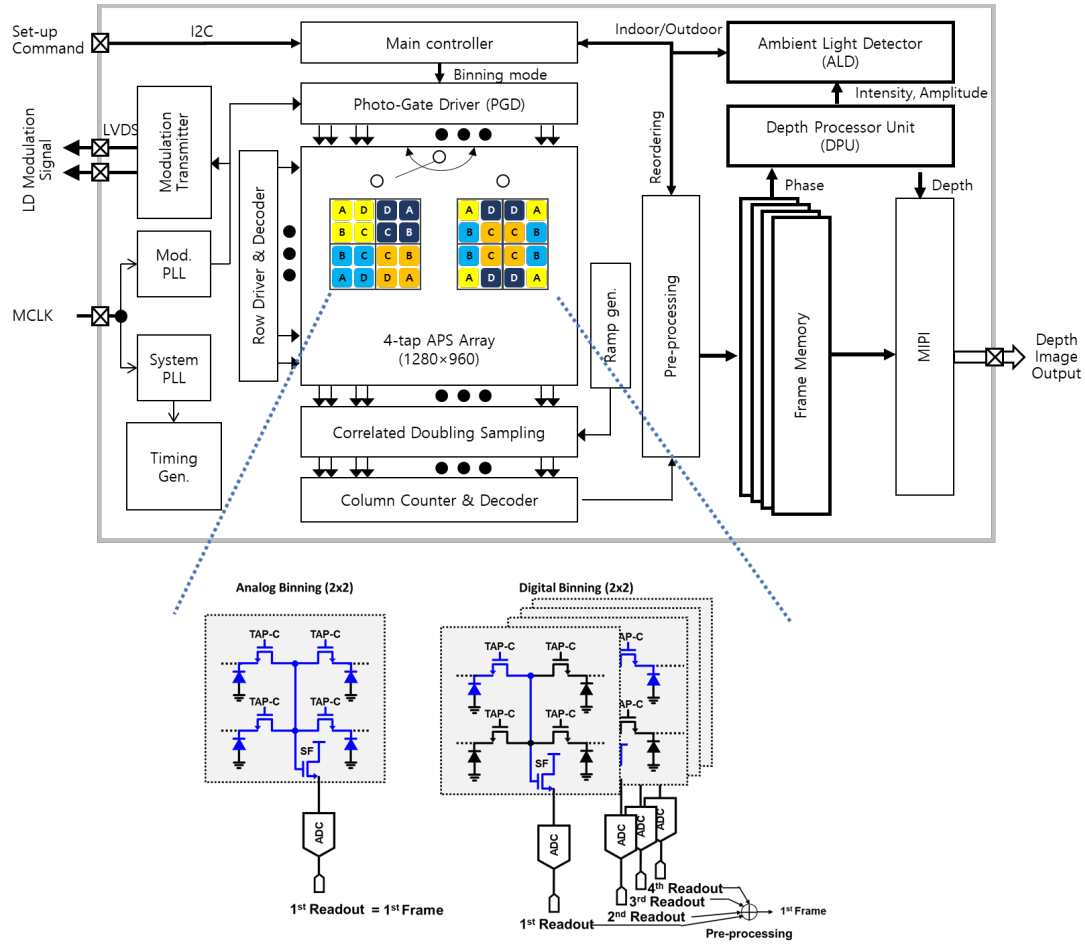


Fig. 1: Block diagram with depth processor and binning mode operation. Pixel structure is referenced [3].

III. PROPOSED ARCHITECTURE

Fig. 1 shows a block diagram of the proposed sensor with depth processor. A unit pixel is composed of 2×2 sub-pixels, each of which is composed of 4-tap (A, B, C, D). The phase information is stored during the exposure integrate time (EIT) section under the control of the photo gate driver signal. After EIT, analog signal is digitized using correlated double sampling (CDS) and counter. These data are subjected to phase reordering through a pre-processing block, and is stored in a frame memory consisting of up to 4-pages of raw data for shuffle and multi frequency. This signal is converted into depth information through the depth processing unit built into the sensor. The eDPU performs phase delay calculation, non-ideality and lens calibration, spatial filtering, and temporal filtering to calculate depth and enhancement. In general, the hard-wired logic has the disadvantage that only pre-defined operations are possible. However, the proposed eDPU implements simple mathematical operations (eg, phase delay sensing) with hard-wired logic, and the parts that

require function change according to the application, such as filtering, are programmable configured. The ambient light detector (ALD) calculates the ambient light environment using the intensity and amplitude calculated from the eDPU and determines the binning mode by sensor itself. This information delivers to the pre-processing block for data reordering. As shown in the bottom of Fig. 1, in case of analog binning since the summation signal of 2×2 sub-pixels is read at once to increase signal. In case of digital binning, 1280×960 signals are captured from all pixels to increase full well capacity (FWC) and converted into 640×480 through average in the pre-processing block.

The timing diagram is shown in Fig.2. The proposed eDPU supports various system scenarios and minimizes motion artifacts by optimizing the raw frame rate based on shuffle, multi frequency, and ISP calculations based on depth frame. Fig.2.(a) shows the shuffle operation under single frequency modulation (s-fm). It calculates depth using two-frame and has the strength of minimizing motion.

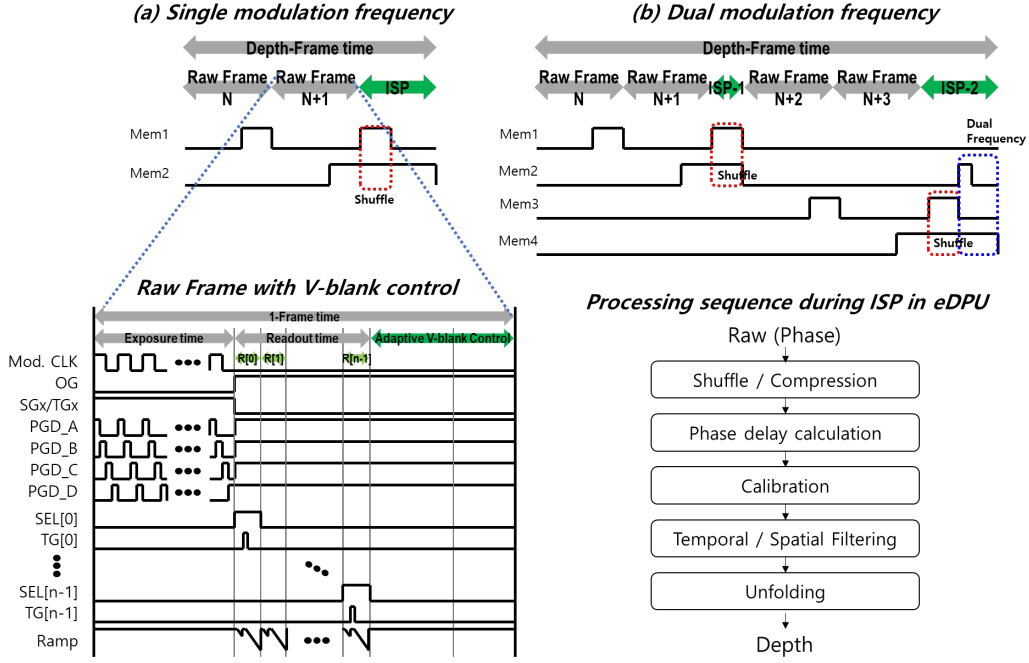


Fig. 2: Timing diagram of frame: depth frame with single and dual modulation frequency, raw frame about demodulation, and eDPU processing sequence

However, since the operating distance is determined only by the modulation frequency, it is suitable for short-distance operation. For memory size optimization, non-shuffle data (n -frame) is directly stored in frame memory, and shuffle frame data ($n+1$ -frame) uses line memory to calculate shuffle, then compressed data stored in frame memory. Next, the depth information is output through phase delay and filtering operations. Fig.2.(b) shows the operation case of dual frequency modulation (d-fm). Although motion lagging time increases because it operates based on four-frame information, it has advantages in securing high-performance depth through distance increase and high-frequency modulation. The data (n , $n+1$ frame) corresponding to modulation 1st frequency (f_1) is shuffled and calculated phase delay then stored in memory, and after reading out ($n+2$, $n+3$ frame) data of 2nd frequency (f_2), shuffle, calibration,

filtering and unfolding operation is performed. This sequence is possible to support motion reduction and distance increase according to the application, respectively.

IV. IMPLEMENTATION AND RESULTS

The proposed sensor is implemented with a 2-stack structure, the top chip is a pixel array with 65nm back-side illumination (BSI), and the bottom chip has been

fabricated with a 28nm logic process for chip-size optimization due to the implementation of eDPU and frame memory. To minimize motion, raw data operates at 250 fps, and Fig. 3 shows the experimental results with rotating at 100 RPM to check the effect of motion artifact. Motion lagging time is defined from 1st raw frame start to EIT of last frame. During dual frequency operation, a lagging time of 25.39ms occurs. Fig. 3.(a) shows the experimental results of operation with dual frequency ($f_1=100\text{MHz}$, $f_2=30\text{MHz}$) in 60fps raw frame condition. The depth information of the moving area is distorted due to the 50.38ms lagging time, and it can be confirmed that some areas are masked due to low SNR degradation. Fig. 3.(b) shows the case of applying shuffle at $f_m=50\text{MHz}$ with 8.73ms lagging time. It shows that the error area is reduced. Additionally, motion free performance can be secured when shuffle is excluded as shown in Fig. 3.(c), however, the noise characteristics are twice deteriorated due to tap mismatch.

Fig. 4 shows the depth and eDPU performance of the proposed sensor. For evaluation, lens is used F/1.3, 78° FoV, and transmitter was used with 940nm 3.4W peak power based vertical-cavity surface-emitting laser (VCSEL). The maximum EIT was set to 0.4ms, and saturation was prevented by controlling the exposure time in the main controller. For performance comparison with SW-ISP, an experiment is used with an

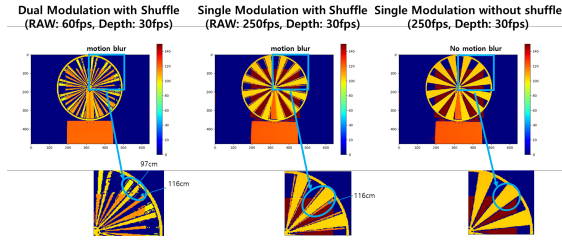


Fig. 3: Comparison of motion artifacts according to modulation frequency and shuffle application

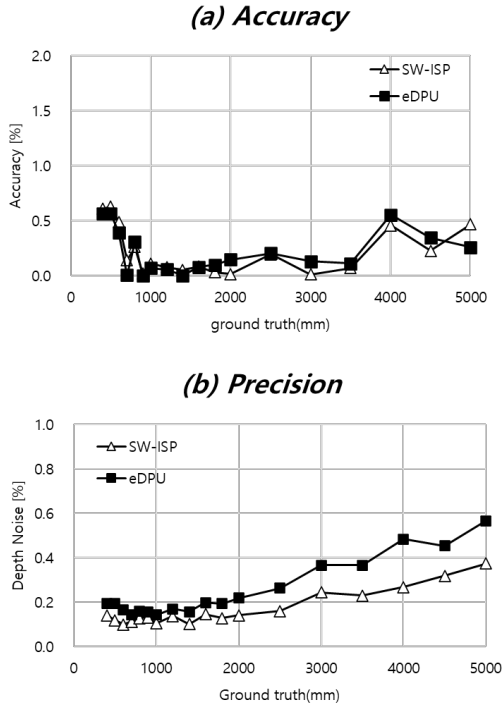


Fig. 4: Depth performance comparison of depth processing performance of AP-based SW and eDPU: Accuracy, Noise, Latency, and Power

AP Processor (HDK8350, 64bit float) by implementing the same function of 640x480 resolution. As a result of comparison, accuracy was less than 0.6% within the working range (0.4 ~ 5m). Depth noise was achieved to be less than 0.57%, but it was degraded by 0.2% compared to SW-ISP. The reason is that the quantization noise increases during the compression process for optimizing the frame memory capacity. As a result of comparison of latency time, it was to 13.59ms, and the amount of output data was also reduced to 6.25% ~ 12.5%. Processing power consumption is only 0.12W, which is reduced by 90% than SW-ISP as shown in Fig. 5. Finally, the sensor + depth operation power was reduced from 1.4W to 0.29W. Fig. 5 shows the depth performance according to the ambient condition.

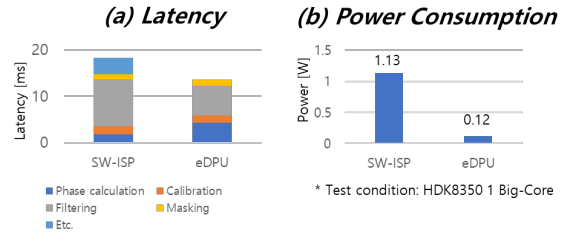


Fig. 5: Latency and power consumption comparison of depth processing performance of AP-based SW and eDPU

V. CONCLUSION

The proposed i-ToF system enables low-power design of miniaturized devices such as AR/VR with the embedded ISP. In addition, the same performance and high-speed FPS compared to SW were confirmed. Through this, it is possible to use depth data more easily and efficiently when implementing applications and systems.

REFERENCES

- [1] Bamji, Cyrus S., et al. "1Mpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3 μ m global shutter pixels and analog binning." 2018 IEEE International Solid-State Circuits Conference-(ISSCC). IEEE, 2018.
- [2] Keel, Min-Sun, et al. "A VGA Indirect Time-of-Flight CMOS Image Sensor With 4-Tap 7- μ m Global-Shutter Pixel and Fixed-Pattern Phase Noise Self-Compensation." IEEE Journal of Solid-State Circuits 55.4 (2019): 889-897.
- [3] Keel, Min-Sun, et al. "7.1 A 4-tap 3.5 μ m 1.2 Mpixel Indirect Time-of-Flight CMOS Image Sensor with Peak Current Mitigation and Multi-User Interference Cancellation." 2021 IEEE International Solid-State Circuits Conference (ISSCC). Vol. 64. IEEE, 2021.
- [4] Kang, Jubin, et al. "A 640 \times 480 Indirect Time-of-Flight Image Sensor with Tetra Pixel Architecture for Tap Mismatch Calibration and Motion Artifact Suppression." 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits). IEEE, 2022.
- [5] Shin, Seung-Chul, et al. "Indirect-ToF system optimization for sensing range enhancement with patterned light source and adaptive binning", Proc. IISW, 2021, [online] Available: <https://imagesensors.org/Past%20Workshops/2021%20Workshop/2021%20Papers/P12.pdf>