A 128×128 Dual-Resolution CMOS Image Sensor with In-Sensor Temporal and Spatial Gradient Calculations for Optical Flow Pre-Processing

Liang-Yu Huang, John Carl Joel Salao Marquez, Guan-Cheng Chen, Chung-Chuan Lo, Ren-Shuo Liu, Min Sun, Meng-Fan Chang, Kea-Tiong Tang, and Chih-Cheng Hsieh

National Tsing Hua University, Hsinchu, Taiwan

2025 IEEE Asian Solid-State Circuits Conference (A-SSCC) | 979-8-3315-8631-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/A-SSCC67472.2025.11349365

Optical flow (OF) describes object motion relative to the observer and is widely used in computer vision applications, requiring real-time processing, compact design, and low power consumption. Gradient-based OF relies on the constraint equation $I_xV_x + I_yV_y + I_t = 0$, where I_x and I_y are the spatial gradients, and I_t is the temporal gradient. Conventional methods using high-frame-rate sensors and processors are power-intensive due to complex computations and frequent frame buffer access. Recent CIS designs [1-4] compute 1-D spatial and temporal gradients but are limited in circuit complexity and energy efficiency. Compared to time-stamped approaches, gradient-based methods offer better angle information and higher speed limits. The proposed smart CIS employs a processing-in-sensor (PIS) approach [5] to compute 2-D spatial gradients and temporal gradient (I_x, I_y, I_t) in-sensor while outputting dual-resolution images for simple and pyramid OF [6]. The signal pre-processing architecture can reduce backend computational workload and memory access significantly and thereby achieving a better system power efficiency.

Figure 1 illustrates the architecture of the prototyped chip, which features a 128x128 pulse-width-modulation (PWM) pixel array with in-pixel frame-differencing (FD) operation for I_t, in-column 2-D spatial gradient calculation circuits for I_x and I_y, in-column binning circuits for dual-resolution output to support pyramid OF, and the peripheral supporting blocks. According to the OF scenario, the maximum detectable motion speed in the scene, which is fundamentally constrained by the extent of projected displacement that can be captured within a single pixel pitch, is significantly improved through the integration of the binning operation in the sensor. In this case, the maximum detectable motion speed is enhanced by three times, enabling more robust performance in the high-speed dynamic scenes.

Figure 2 shows the pixel operation and timing. Within a single frame time, this design supports the simultaneous output of multimodal information, including the raw image (I_{raw}), the pre-processed gradients (I_x, I_y, I_t), and the binning data, ensuring the required synchronicity for post processing. A 6T1C PWM-based pixel [7] is employed for in-pixel FD with a 9-bit output. Unlike previous work [7] that supports only I_t or I_{raw} in a single frame, this design enables simultaneous FD (I_t) and raw image (I_{raw}) outputs using data resampling operation.

Figure 3 shows the implemented in-column circuits for spatial gradient calculations and binning operation. Leveraging the time-domain output of PWM pixel, the spatial gradient calculations (I_x and I_y) are efficiently achieved using low-voltage combinational logic in the column level circuit, eliminating the need for OPAMP-based analog subtraction or adder-based digital subtractors. This low-voltage logic implementation is power- and area-efficient, robust, and well-suited to advanced technology nodes. To compute the vertical spatial gradient (I_y) of a certain pixel, the PWM outputs from three consecutive rows are selected simultaneously via three vertical buses within a column. Signals from the up ($V_{pw,U}$) and down ($V_{pw,D}$) neighboring pixels are multiplexed and subtracted using column-wised XOR gates. The sign bit, $I_{y,sign}$ (= 1 when $V_{pw,U} > V_{pw,D}$, and = 0 when $V_{pw,U} < V_{pw,D}$), is determined via a simple dynamic NAND operation between I_y and the lower-side signal $V_{pw,D}$. Since the rising edge of I_y is triggered by the falling edge of $V_{pw,D}$, the gated result of sign bit is guaranteed to be glitch-free. Similarly, the horizontal spatial gradient (I_x) and sign bit ($I_{x,sign}$) calculations are realized using an identical circuit as well, which efficiently processing the signals from the right ($V_{pw,R}$) and left ($V_{pw,L}$) neighboring pixels.

To support the computation of pyramid OF, the column-parallel 3-by-3 binning operations for the raw image and temporal gradient I_t are implemented using time-to-voltage-to-time (TVT) operation. In the time-to-voltage (T2V: integration) phase with Φ_{INT} = "1", the nine PWM signals from 3-by-3 subarray are used to control the turn-on duration of the integration current I_{int}, which is then integrated on the capacitor C_{bin} (=0.57pF) to sum up the signals from the nine pixels. Then, in the voltage-to-time (V2T: conversion) phase with Φ_{INT} = "0", the voltage V_{bin} on C_{bin} is discharged using the mirrored current I_{dis} to obtain the pulse width $V_{pw.bin}$ related to the summed signal in the T2V phase, and converted to 9-b digital code using a simple counter. By employing a locally matched current ratio I_{int}/I_{dis} and a single capacitor C_{bin} for both T2V and V2T conversions, it achieves an accurate 3-by-3 binning operation and is insensitive to process variations.

A 1V/1.1V 128x128 smart CMOS image sensor using the proposed PIS techniques for spatial and temporal gradient calculation and 3-by-3 binning operation for pre-processing of pyramid optical flow computations was fabricated in TSMC 0.18-μm CMOS process.

Figure 4 presents captured images of moving objects, including 126x126 9-bit raw data (I_{raw}), I_x, I_y, I_t, 42x42 binning results, and the OF results (V_x, V_y) computed off-chip. In the doll scenario, edge information (I_x, I_y) and motion contrast (I_t) are clearly depicted. For the moving car, the temporal gradient effectively filters out the static background, enhancing data sparsity and reducing the energy required for data transfer and processing. Compared to the ideal results from Matlab using captured raw image I_{raw}, the achieved root-mean-square-errors (RMSEs) of the spatial gradients (I_x, I_y) and binning ($I_{raw.bin}$) operations are around 0.4 and 1.6 LSB (8-bit), respectively. The OF results achieve an end-point-error (EPE) of 0.375 pixel.

Figure 5 demonstrates the OF results using measured gradients (I_x, I_y, I_t) with and without the pyramid method. The implemented 3x3 binning operation extends the detectable speed range by approximately 3x. In addition, the hand gestures experiments further validate the sensor's capability to capture motion information for hand pose detection.

Figure 6 presents the measured performance and a comparison table. This work supports comprehensive PIS capabilities, including 9-bit outputs for raw image, gradients (I_x, I_y, I_t), and 3x3 binning operations for gradient-based OF pre-processing. With the 1V/1.1V supplies, it consumes 198.44uw at 30fps. For a fair comparison across different operational modes and data bit depths, a modified iFoM is defined by normalizing to the summation of feature levels. The proposed sensor demonstrates competitive iFoM results of 0.02~0.197 pJ/(pix*fps* level) among multimodal outputs (Mode1~3), offering an energy-efficient solution for wearable edge applications involving real-time motion and depth sensing.

References:

[1] M. -J. Park *et al.*, "A Real-Time Edge-Detection CMOS Image Sensor for Machine Vision Applications," Sensors Journal 2023.

[2] H. -J. Kim *et al.*, "A Dual-Imaging Speed-Enhanced CMOS Image Sensor for Real-Time Edge Image Extraction," JSSC 2017.

[3] K. Lee *et al.*, "A 272.49 pJ/pixel CMOS image sensor with embedded object detection and bio-inspired 2D optic flow generation for nano-air-vehicle navigation," Symposium on VLSI Circuits 2017.

[4] S. Park *et al.*, "7.2 243.3pJ/pixel bio-inspired time-stamp-based 2D optic flow sensor for artificial compound eyes," ISSCC 2014.

[5] T. -H. Hsu *et al.*, "A 0.5-V Real-Time Computational CMOS Image Sensor With Programmable Kernel for Feature Extraction," JSSC 2021.

[6] J.-Y. Bouguet *et al.*, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," Intel corporation, vol. 5, no. 1-10, p. 4, 2001.

[7] M.-Y. Chiu *et al.*, "A multimode vision sensor with temporal contrast pixel and column-parallel local binary pattern extraction for dynamic depth sensing using stereo vision," JSSC 2023.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on July 01,2026 at 06:12:39 UTC from IEEE Xplore. Restrictions apply.

979-8-3315-8631-7/25/$31.00 ©2025 IEEE

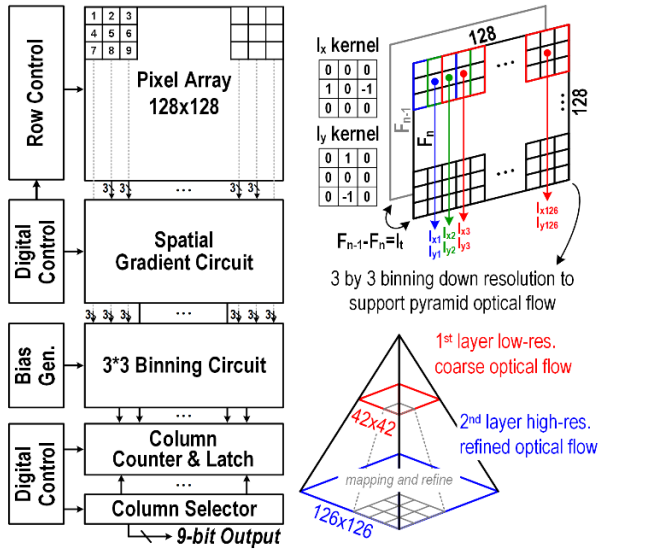


Fig. 1. Chip architecture and implemented features, including raw data output, temporal and spatial gradients, and binning operations.

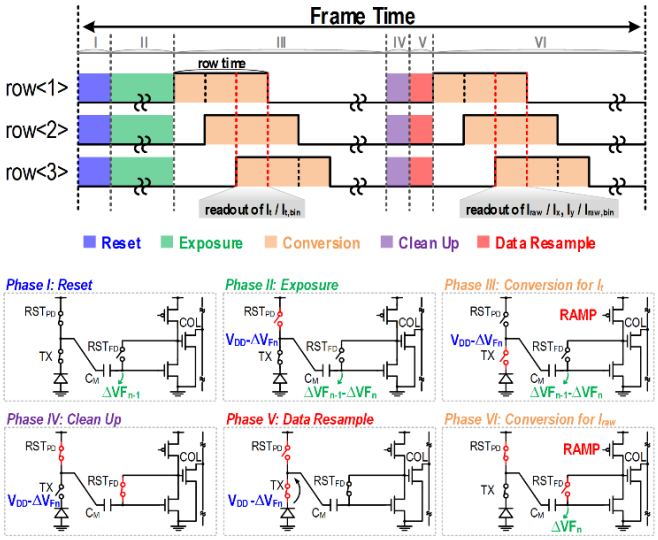


Fig. 2. Pixel operations and timing for simultaneous output, supporting I_t, $I_{t.bin}$, I_{raw}, I_x, I_y, and $I_{raw.bin}$ using data resampling.

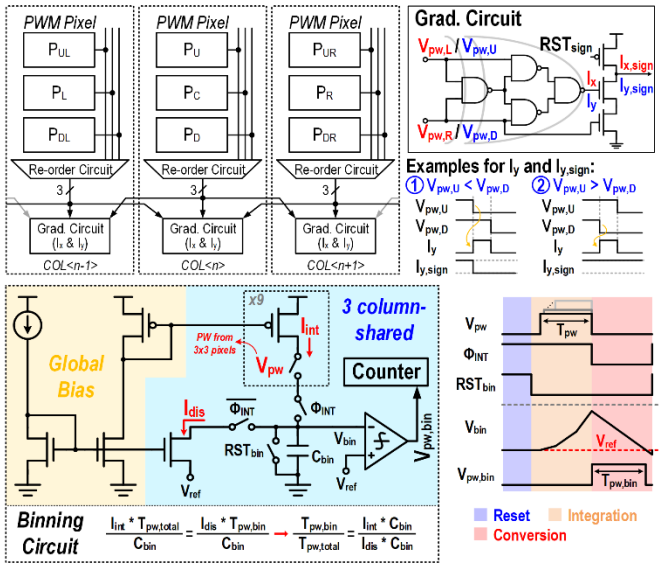
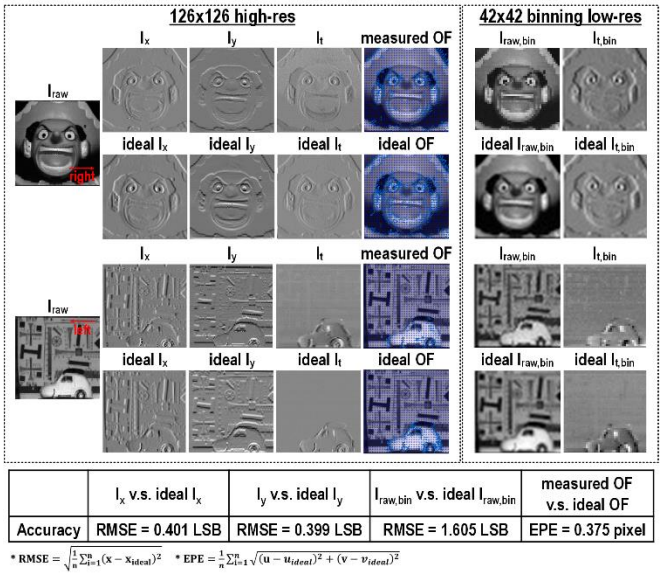


Fig. 3. In-column circuits for spatial gradient calculations and binning operation.



	I_x v.s. ideal I_x	I_y v.s. ideal I_y	$I_{raw,bin}$ v.s. ideal $I_{raw,bin}$	measured OF v.s. ideal OF
Accuracy	RMSE = 0.401 LSB	RMSE = 0.399 LSB	RMSE = 1.605 LSB	EPE = 0.375 pixel

$^*\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x - x_{ideal})^2}$ $^*\text{EPE} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(u - u_{ideal})^2 + (v - v_{ideal})^2}$

Fig. 4. Measured accuracy and capture images of 126x126 I_{raw}, I_x, I_y, I_t, and 42x42 binning output of $I_{raw.bin}$ and $I_{t.bin}$.

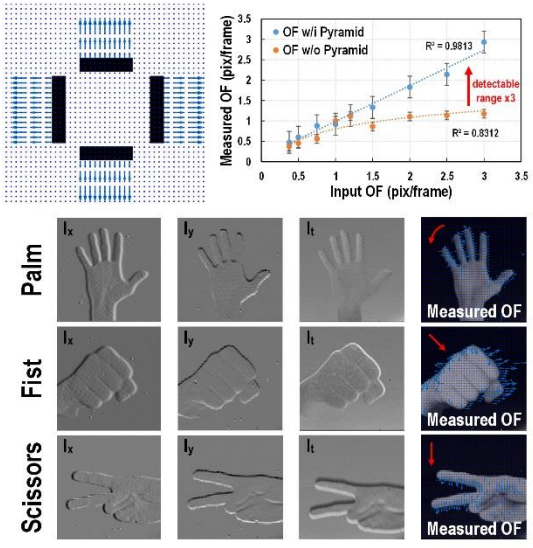


Fig. 5. Measured optical flow wi/wo pyramid method and hand pose verification.

	This work	[1] Park, Sensors Journal'23	[2] Kim, JSSC'17	[3] Lee, VLSI'17	[4] Park, ISSCC'14
Process	**180nm**	180nm	180nm	180nm	180nm
Supply	**1 (analog) / 1.1 (digital)**	2.8 (analog) / 1.8 (circuit)	2.8 (pixel) / 1.8 (circuit)	3.3 (pixel) / 1.8 (analog)	3.3 (pixel) / 0.9 & 1.8 (analog) / 1.8 (digital)
Pixel pitch	**12um**	8um	4.9um	31um	28.8um
Array size	**128 x 128**	320x320	160x120	256x256 64x64	64x64
Fill factor	**44%**	N.A.	53%	19%	18.32%
Raw image resolution	**9-bit 126x126 & 42x42**	10-bit 320x320	10-bit 160x120	X	X
Spatial gradient	**2D, 9-bit**	1D, 1-bit	1D, 5-level	1D, 2-bit	X
Temporal gradient	**9-bit 126x126 & 42x42**	X	X	1-bit	1-bit
Support algorithm	**Gradient-based optical flow**	X	X	Time-stamp-based optical flow	Time-stamp-based optical flow
Frame rate	**30 fps (Mode1) 89 fps (Mode2) 556 fps (Mode3)**	240 fps	3200 fps	30 fps	120 fps
Memory	**No**	No	No	Yes (digital)	Yes (digital)
Power	**198.44uW**	17.72mW	4.3mW	Spatial: 2.18mW Temporal: 29.94uW	109.38uW
iFoM*	**0.1971 (Mode1) 0.0885 (Mode2) 0.0212 (Mode3)**	0.7027	0.0680	277.2 121.825	111.265

*iFoM = pJ/(pix*fps*total level), total level = summation of feature level, level = 2^n, n = bit
*Mode1 = Raw image+I_t(126x126, 42x42)+I_x+I_y(126x126) *Mode2 = Raw image+I_t+I_x+I_y(126x126)
*Mode3 = Raw image+I_x+I_y(126x126)

Fig. 6. Comparison table.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on July 01,2026 at 06:12:39 UTC from IEEE Xplore. Restrictions apply.

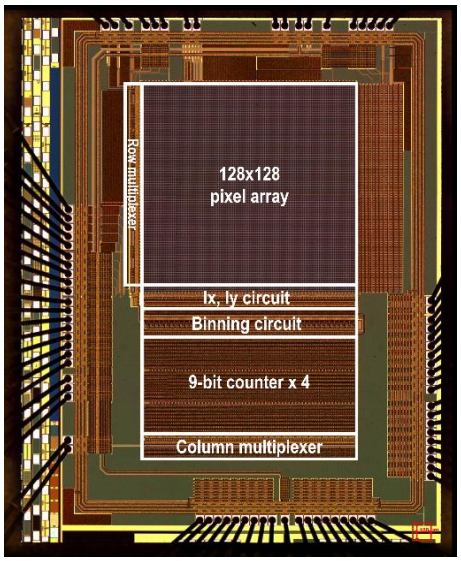


Fig. 7. Chip micrograph.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on July 01,2026 at 06:12:39 UTC from IEEE Xplore. Restrictions apply.