Intelligent Imager with Processing-in-Sensor Techniques

Chih-Cheng Hsieh
Department of Electrical Engineering, National Tsing Hua University
101, Section 2, Kuang-Fu Road, Hsinchu 300044, Taiwan R.O.C.
E-mail: cchsieh@ee.nthu.edu.tw

Abstract Imaging systems with signal processing have found widespread use in DSP-based and AI-aided applications. Processing-in-sensor (PIS) techniques take advantage of reducing power consumption and data transfer latency by enabling data processing at the sensing node. In smart edge applications, intelligent imagers utilizing PIS techniques with in/near-sensor feature extraction present a promising solution. This talk will explore the existing literature and ongoing research that leverage PIS techniques while also addressing the associated challenges and future potential.

Keywords: processing-in-sensor, smart sensor, feature extraction, edge device

1. Introduction

The demand forecast for the CMOS image sensor market is still growing and optimistic contributed by the AI-aided sensing usage nowadays. The AI-aided smart imager is an integration of image sensing and AI computing capabilities. It can exceed human eye's capability by adding intelligence in it to extract meaningful information beyond the image itself, such as feature extraction for machine vision. Processing-in-sensor (PIS) technique further enhances the system efficiency of the smart imager across diverse applications, including smart surveillance, automotive, and robotics. In the intelligent imager using PIS techniques, the processing-in-sensor circuit is expected to be implemented in the CMOS image sensor between the pixel array and ADC. The PIS circuit will perform the pre-processing task before data digitization. By doing so, the ADC's spec requirement including resolution and bandwidth can be relieved. Furthermore, the required data transfer is feature or ROI only. This effectively reduces the demand for power/latency in interconnections and the required memory on the processor. According to various applications and conditions, the PIS circuit can be implemented in various approaches and roughly classified into two categories, including spatial domain and temporal domain feature extractions [1].

2. Spatial domain feature extraction

The concept of spatial domain feature extraction is to implement static texture filtering to get spatial information such as texture, coarseness, contrast, and more to remove redundant raw data. Several well-known processing engines for spatial information extraction using the PIS technique have been reported, such as LBP, HOG, and NN. The concept of LBP is to extract spatial information by encoding the relationship between the intensity of a central pixel and its surrounding pixels [2]. It can be applied to texture classification and recognition, offering advantages such as computational efficiency and immunity of illumination and rotation. However, it comes with limitations, including restricted global feature description and sensitivity to noise. The HOG method extracts the spatial information by calculating the distribution of gradient orientations in a specific image region. It offers the advantage of immunity to illumination as well but suffers from sensitivity to rotation [3]. On the other hand, the use of Convolutional Neural Networks (CNN) in AI has become increasingly powerful and dominant for spatial information extraction. Through multiple layers of operations, CNN can extract various spatial content from an image, including color, texture, shape, and more. The interesting thing is that the functions in the CNN model can be easily realized in the analog domain [4-6]. The proposed imager in [6] is even equipped with a customized tiny CNN and accomplishes the task of face detection using mixed-mode PIS circuits.

3. Temporal domain feature extraction

Temporal domain feature extraction is to detect the temporal change in each pixel, and then report the level-difference image or locations of triggered pixels. It is useful for motion detection of consecutive images by eliminating the static information (like background) to remove the redundant data. It can be applied to various applications, including motion detection, direction detection, saliency detection, dynamic depth sensing, temporal derivative, and more. There are two main methodologies employed in this process. One is event-based reporting, such as the dynamic vision sensor, and the other is frame-based reporting such as the frame differencing sensor. The idea of an event-based reporting (ER) operation is to identify and report the location of events by thresholding the temporal changes per pixel. It is commonly implemented with a sensor featuring real-time logarithmic I_{ph}-V conversion and asynchronous x-y location reporting readout. Unlike the conventional frame-based operation in standard cameras, it achieves continuous high-speed and high dynamic range temporal feature extraction [7-10]. On the contrary, the concept of frame-differencing (FD) operation is to report temporal level difference or thresholding event between two consecutive frames [11-13]. This is typically implemented with a sensor using linear integrating I_{ph}-V conversion and synchronous frame reporting readout. Unlike the ER sensor, the FD sensor requires no in-pixel amplifier thanks to the inherent I-V conversion gain of integrating operation, featuring a smaller pixel size and low power consumption, but with a smaller dynamic range.

4. Conclusion

The evolution from traditional IoT to the cutting-edge era of cognitive AIoT is in progress. This transformation involves migrating essential information from centralized cloud to edge devices and transforming raw data into meaningful insights with commendable energy efficiency for specific tasks. We believe that the processing-in-sensor technique is a promising solution for application-driven intelligent vision systems.

References

[1] T. -H. Hsu et al., "AI Edge Devices Using Computing-In-Memory and Processing-In-Sensor: From System to Device," in IEEE IEDM, 2019, pp. 22.5.1-22.5.4.

[2] X. Zhong et al., "A 2PJ/Pixel/Direction MIMO Processing Based CMOS Image Sensor for Omnidirectional Local Binary Pattern Extraction and Edge Detection," in IEEE Symposium on VLSI Circuits, 2018, pp. 247-248.

[3] C. Young et al., "A Data-Compressive 1.5/2.75-bit Log-Gradient QVGA Image Sensor With Multi-Scale Readout for Always-On Object Detection," in IEEE JSSC, vol. 54, no. 11, pp. 2932-2946, Nov. 2019.

[4] M. Lefebvre et al., "A 0.2-to-3.6TOPS/W Programmable Convolutional Imager SoC with In-Sensor Current-Domain Ternary-Weighted MAC Operations for Feature Extraction and Region-of-Interest Detection," in IEEE ISSCC, 2021, pp. 118-120.

[5] T. -H. Hsu et al., "A 0.5V Real-Time Computational CMOS Image Sensor with Programmable Kernel for Feature Extraction," in IEEE JSSC, vol. 56, no. 5, pp. 1588-1596, May 2021.

[6] T. -H. Hsu et al., "A 0.8 V Intelligent Vision Sensor With Tiny Convolutional Neural Network and Programmable Weights Using Mixed-Mode Processing-in-Sensor Technique for Image Classification," in IEEE JSSC, vol. 58, no. 11, pp. 3266-3274, Nov. 2023.

[7] P. Lichtsteiner et al., "A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," in *IEEE JSSC*, vol. 43, no. 2, pp. 566-576, Feb. 2008.

[8] K. Kodama et al., "1.22μm 35.6Mpixel RGB Hybrid Event-Based Vision Sensor with 4.88μm-Pitch Event Pixels and up to 10K Event Frame Rate by Adaptive Control on Event Sparsity," in *IEEE ISSCC*, 2023, pp. 92-94.

[9] M. Guo et al., "A 3-Wafer-Stacked Hybrid 15MPixel CIS + 1 MPixel EVS with 4.6GEvent/s Readout, In-Pixel TDC and On-Chip ISP and ESP Function," in *IEEE ISSCC*, 2023, pp. 90-92.

[10] A. Niwa et al., "A 2.97μm-Pitch Event-Based Vision Sensor with Shared Pixel Front-End Circuitry and Low-Noise Intensity Readout Mode," in *IEEE ISSCC*, 2023, pp. 4-6.

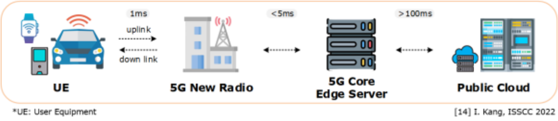
[11] X. Zhong et al., "A Fully Dynamic Multi-Mode CMOS Vision Sensor With Mixed-Signal Cooperative Motion Sensing and Object Segmentation for Adaptive Edge Computing," in IEEE JSSC, vol. 55, no. 6, pp. 1684-1697, June 2020.

[12] 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," in IEEE JSSC, vol. 58, no. 10, pp. 2767-2777, Oct. 2023.

[13] T. -H. Hsu et al., "A 0.8 V Multimode Vision Sensor for Motion and Saliency Detection With Ping-Pong PWM Pixel," in IEEE JSSC, vol. 56, no. 8, pp. 2516-2524, Aug. 2021.

[14] I. Kang, "The Art of Scaling: Distributed and Connected to Sustain the Golden Age of Computation," in IEEE ISSCC, 2022, pp. 25-31.

From Centralized Cloud to Edge Devices



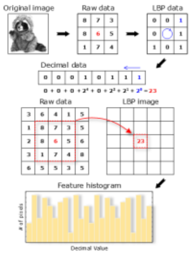
*UE: User Equipment

[14] I. Kang, ISSCC 2022

- Centralized cloud
  - High-level processing/complex task
  - Transmission cost (power & latency)
  - User privacy concerns
  - Network requirement
- Edge devices
  - Local/real-time decision-making
  - Low/Mid-level processing
  - Preserve data privacy
  - Power-constrained environments

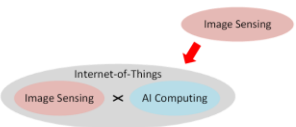
Local Binary Pattern (LBP)

- Concept
  - Encode the relationship between the central pixel and surrounding pixels
- Applications
  - Texture classification
  - Object recognition
- Pros
  - Computational efficiency
  - Immune to illumination and rotation
- Cons
  - Limited global feature description
  - Noise sensitive



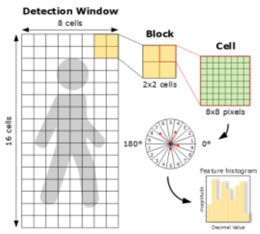
Smart Imager – Exceed Human Eyes' Capabilities

- What is really meaningful?
  - What? / Where? / Who? / When?
- More than the image itself
  - Feature extraction
  - Machine vision

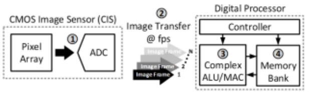


Histogram of Oriented Gradients (HOG)

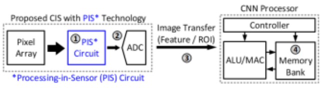
- Concept
  - Compute the distribution of gradient orientations
- Applications
  - Pedestrian detection
  - Human pose estimation
- Pros
  - Immune to illumination
- Cons
  - Sensitivity to rotation



Advantages of Processing-in-Sensor (PIS)

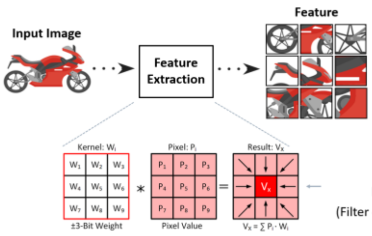


- Conventional imager
  - ① High-resolution ADC
  - ② Data transfer bottleneck
  - ③ High-complexity computation
  - ④ High-capacity memory



- Intelligent imager using PIS
  - ① PIS circuit for feature extraction
  - ② Relieve ADC resolution
  - ③ Transfer feature / ROI only
  - ④ Relieve memory capacity

Convolutional Neural Network (CNN)

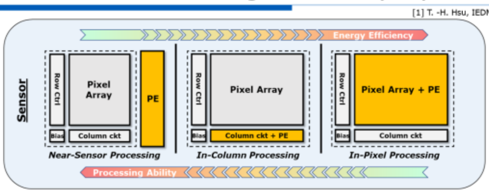


- Content of Image
  - Color
  - Texture
  - Shape
  - Position

Convolution Neural Network (Filter descriptor → Feature)

Architectures of Processing-in-Sensor (PIS)

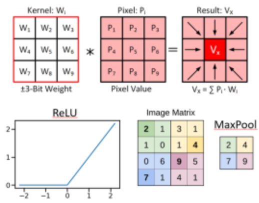
[1] T. -H. Hsu, IEDM 2019



- Near-sensor processing
  - Capability & flexibility
  - Suffered latency/efficiency
  - e.g., classification
- In-column processing
  - In-between solution
  - Medium performance
  - e.g., attention
- In-pixel processing
  - High parallelism
  - Suffered FF
  - e.g., pre-processing

Spatial Domain Image Processing

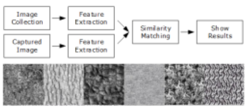
- Convolution: multiply-accumulate (MAC) operation
  - Multiplication: I=V*(1/R), Q=I*T
  - Accumulation: Wired sum (I)
    Integration (Q)
- Rectifier linear activation (ReLU)
  - Polarity judgement: Comparator
- Maximum Pooling (MaxPool)
  - Max finding: Winner-take-all



MAC

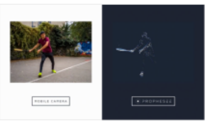
Feature Extraction – Spatial Domain Information

- Concept
  - Static texture filtering
  - Coarseness, contrast, directionality, regularity
- Purpose
  - Extract key characteristics
  - Remove redundant data
- Methodologies
  - Local binary pattern (LBP)
  - Histogram of Oriented Gradients (HOG)
  - Neural network (NN)
- Applications
  - Fingerprint, retina, and face for forgery detection
  - Biomedical diagnosis
  - Environment recognition, autonomous vehicle
  - Defect detection, smart agriculture



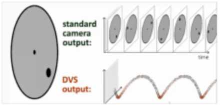
Feature Extraction – Temporal Domain Information

- Concept
  - Detect temporal changes in each pixel
  - Report the level-difference image or locations of the triggered pixels
- Purpose
  - Motion detection between consecutive images
  - Remove redundant data
- Methodologies
  - Event-based reporting: dynamic vision sensor (DVS)
  - Frame-based reporting: frame differencing sensor (FDS)
- Applications
  - Motion detection
  - Motion direction
  - Saliency detection
  - Dynamic depth sensing
  - Temporal derivative



Event Reporting (ER)

- Concept
  - Report event location by thresholding temporal changes per pixel
- Architecture
  - Sensor: real-time logarithmic Iph-V conversion
  - Readout: asynchronous x-y location report
- Applications
  - High-speed / high-dynamic-range
  - Image deblurring
  - Eye-tracking
  - Obstacle avoidance

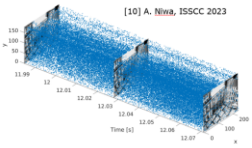


[UZH Robotics and Perception Group]
[7] P. Lichtsteiner , JSSC 2008

Readout Speed

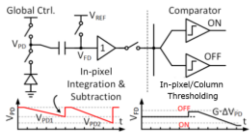
- Event-based reporting
  - Realtime I-V conversion and Asynchronous readout
  - Event data are available once triggered
  - Event rate (events per second, eps)
  - Need accumulation for the following frame-based processing
- Frame difference
  - Integrating I-V conversion and Synchronous readout
  - Generate event frame at a fixed time interval
  - Frame rate (frame per second, fps)
  - Limited by the exposure time

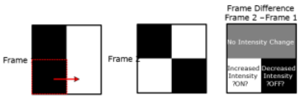
[10] A. Niwa, ISSCC 2023



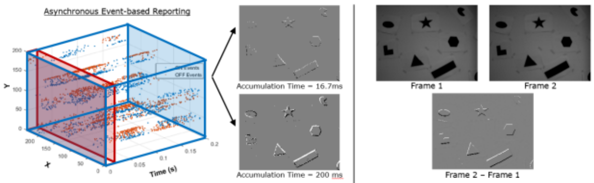
Frame Differencing (FD)

- Concept
  - Report temporal level difference or thresholding event of two consecutive frames
- Architecture
  - Sensor: integrating Iph-V conversion
  - Readout: Synchronous frame report
- Applications
  - Saliency detection
  - Motion detection
  - Motion direction detection
  - Dynamic depth sensing





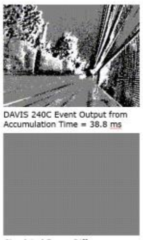
Compatibility with Processors: Event Frame



- Asynchronous event-based readout: event accumulation
  - Similar concept with exposure time
  - Requires additional post-processing
- Synchronous frame difference
  - No additional post-processing

Event Reporting (ER) vs. Frame Differencing (FD)

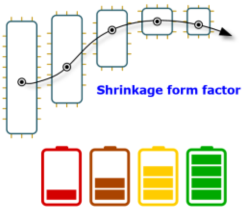
- Sensing
  - ER: real-time logarithmic Iph-V conversion with non-linear HDR response, need in-pixel amplification
  - FD: integrating Iph-V conversion with a linear response, inherent gain from exposure.
- Readout
  - ER: Asynchronous x-y event location reporting of thresholding temporal difference
  - FD: Synchronous frame reporting of temporal level difference or thresholding event of two consecutive frames
- Specifications to be considered:
  - Event sensitivity, Dynamic range, Speed, Compatibility with processors


DAVIS 240C Event Output from Accumulation Time = 38.8 ms

Simulated Frame Difference

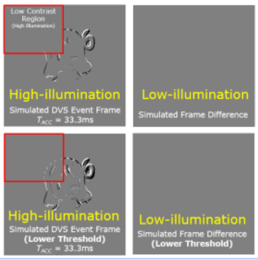
Challenges of PIS

- Analog / mixed-mode approaches
  - Low / mid-level processing
  - Accuracy / flexibility
  - Application-specific task
- High parallelism / low latency
  - Pixel / Column level processing
  - Pitch limitation
  - Real-time sensing
- High energy efficiency
  - Reduced data transmission
  - Limited power source
  - Always-on / portable / AIoT application



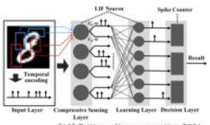
Event Sensitivity and DR Improvement

- Sensitivity improvement:
  - ON and OFF event **threshold reduction**
  - May result in **increased noise-triggered events**
- Dynamic range improvement
  - Dual conversion gain
  - Multiple exposure
- Possible solutions
  - Adaptive gain
  - Adaptive thresholding
  - Adaptive exposure
  - **Tradeoff of event sensitivity, noise, and dynamic range**

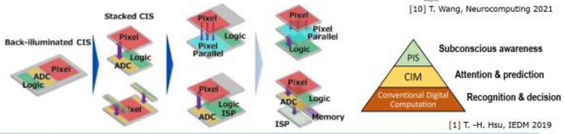


Future Prospects of PIS

- Customized network + error tolerance
  - PIS+Tiny-CNN, PIS+SNN, Retraining
- Pitch limitation relief
  - 3-D stacking technique
- Energy harvesting



[10] T. Wang, Neurocomputing 2021



[1] T. -H. Hsu, IEDM 2019