

추론과 센싱의 통합을 통한 효율적인 AI 기반 비전 센서 시스템 설계

□ 고흥환 / 성균관대학교

요약

온디바이스 환경에서의 실시간 비전 센서 시스템은 제한된 전력과 대역폭 아래에서 빠르게 변하는 시각 조건에 대응하여 효율적이고 강건한 추론 태스크를 수행할 수 있어야 한다. 그러나 센서에서 획득된 데이터가 프로세서로 전달되어 처리되는 전통적인 단방향 pipeline은, 뒷단의 추론 태스크에 필요하지 않은 정보까지 프로세서로 전달하는 과정에서 큰 전력소모와 지연시간을 유발한다. 이를 해결하려는 흐름은 두 방향으로 나타나고 있는데, 하나는 추론을 센서에 가깝게 또는 센서 안에 통합하여 데이터 이동을 줄이는 near-sensor 및 in-sensor 컴퓨팅의 방향이고, 다른 하나는 직전 프레임의 추론 결과를 다음 프레임의 센싱 및 연산에 피드백하여 태스크를 더 정확하고 효율적으로 수행하는 적응적 피드백 방향이다. 본 기고문에서는 이러한 관점에서 개발된 두 가지 비전 센서 시스템 사례를 통합적으로 조망한다. 첫 번째는 동적 피드백(dynamic feedback) ISP(Image signal processor)와 손 자세 추정(hand pose estimation, HPE) 가속기를 FPGA에 통합한 near-sensor 시스템이고, 두 번째는 이를 센서 칩 내부로 발전시켜 추론 결과로 센싱 파라미터를 실시간 조절하는 in-sensor 이벤트 카메라이다. 두 사례는 직전 프레임의 추론 결과로 센싱-연산을 조절하는 접근이 near-sensor에서 in-sensor로 발전하면서 동적 환경의 강건성과 저전력 실시간 처리를 동시에 달성할 수 있음을 보여준다.

I. 서론

온디바이스 AI 시장의 빠른 성장으로, AR/VR, 스마트폰, IoT 기기 등에서의 저전력·실시간 비전 처리에 대한 요구가 빠르게 커지고 있다. 이러한 기기는 주로 배터리로 구동되기 때문에, 단순히 인식 정확도를 높이는 것을 넘어

제한된 전력 예산 안에서 실시간성과 강건성을 함께 확보하는 것이 핵심 과제이다.

전통적인 비전 센서 시스템은 센서가 장면이나 태스크의 상태와 무관하게 고정된 방식으로 데이터를 획득하고, 이를 별도의 처리·추론 장치로 전달하는 단방향 구조로 동작한다. 이 구조에는 두 가지 근본적 비효율이 있는데 우

선 센서가 데이터를 취득할 때 태스크를 고려하지 않는다는 점이다. 센서가 추론에 실제로 필요한 정보와 무관하게 고정된 해상도와 파라미터로 모든 화소를 균일하게 획득할 경우 시간적·공간적으로 불필요한 데이터를 획득하므로 이에 따른 에너지 낭비가 발생하며 때로는 태스크 추론에 방해가 되기도 한다. 또한, 추론을 위한 프로세서가 센서와 분리되어 있는 경우 대량의 센서 데이터를 연산 장치로 옮기는 데이터 이동이 큰 전력과 지연을 유발한다.

이러한 비효율을 줄이기 위해, 추론을 센서에 가깝게 배치하거나(near-sensor computing) 센서 내부에서 직접 수행하는(in-sensor computing) 흐름이 등장하였다 [1, 2]. 더 나아가, 직전 프레임의 추론 결과나 태스크 특성을 센싱 과정에 반영하여 다음 프레임의 추론이 더 정확하고 효율적으로 이루어지도록 센서가 동작을 스스로 조절하는 태스크 인지형 적응 센싱(task-aware adaptive sensing)이 제안되고 있다.

본 기고문은 이러한 발전 과정을 두 가지 실제 하드웨어 사례로 구체화한다. 먼저 동적 피드백 ISP와 추론 가속기를 near-sensor FPGA에 통합한 ToF 손 자세 추정 시스템(III장)을 살펴보고, 이어서 유사한 개념을 센서 칩 내부의 적응형 closed-loop로 발전시킨 in-sensor 이벤트 카메라(IV장)를 소개한다. II장에서는 두 사례를 관통하는 공통 목표와 구조, 설계 원리를 정리하고, V장에서 향후 전망을 제시한다.

II. 추론과 센싱의 통합: 목표 및 설계 원리

1. 단방향 pipeline의 한계

단방향 pipeline의 한계는 센싱과 추론이 분리되어 독립적으로 동작한다는 데서 비롯된다. 고정된 센싱 파라미터는 조도나 움직임이 급변할 때 정보량이 낮은 데이터를 생성하여 추론 정확도를 떨어뜨리고, 태스크와 무관한 배경까지 빠짐없이 획득하여 처리하는 구조는 불필요한 전력을 낭비한다. 또한 센서와 프로세서 사이의 데이터 이동이 시스템 성능과 전력의 병목이 된다. 결과적으로 강건성과 효율이 동시에 저하되며, 단순히 더 빠른 가속기를 붙이는 것만으로는 해결되지 않는다.

2. 통합의 목표와 구조

추론과 센싱의 통합이 지향하는 공통 목표는, 직전 프레임의 추론 결과를 바탕으로 다음 프레임의 추론이 더 정확하고 효율적으로 이루어지도록 시스템 파라미터를 조절하는 것이다. 즉, 추론과 센싱이 매 프레임 서로의 결과를 참고하며 점진적으로 최적의 동작점을 찾아가는 구조이다. 이러한 구조는 세 가지 구성요소로 일반화할 수 있는데 첫째는 대상 센서, 둘째는 그 위에서 수행하는 추론 태스크, 셋째는 추론 결과나 태스크 특성에 따라 조절하는 파라미터이다. 본 기고문의 두 사례는 이 세 요소를 서

<표 1> 두 사례의 구조 기반 비교: 대상 센서·추론 작업·조절 파라미터

구분	사례 1: ToF 손 자세 추정	사례 2: 이벤트 카메라
대상 센서	iToF 3D 센서	digital-EVS (이벤트 센서)
추론 task	손 자세 추정 (HandFoldingNet)	객체 검출
조절 파라미터	처리 영역(valid area)·해상도·연산 자원	노출 시간(T_{exp})·이벤트 임계값(E_{thr})·RoI readout
통합 위치	near-sensor (FPGA)	in-sensor (센서 칩 내부)
적용 근거	직전 프레임의 추론 유효 영역	직전 프레임의 추론 결과(신뢰도·마스크)

로 다르게 채우면서, 추론과 센싱의 통합이 near-sensor에서 in-sensor로 깊어지는 과정을 보여준다(〈표 1〉). 사례 1은 추론과 ISP를 near-sensor에 통합하고 직전 프레임에서 응용이 사용한 유효 영역을 피드백으로 받아 ISP의 처리 영역과 해상도를 조절하여 효율과 정확도를 함께 높였고, 사례 2는 추론을 센서 칩 안으로 가져와 직전 프레임의 검출 결과로 노출 시간과 이벤트 임계값을 직접 조절하였다.

3. 공통 설계 원리

통합의 지점은 다르지만 두 사례는 다음과 같은 공통된 설계 원리를 따르고 있다. 첫째는 태스크 인지형 적용으로, 직전 프레임의 추론 결과를 센싱·연산을 조절하는 피드백 신호로 사용하여 추론 태스크의 탐지 성능을 향상하고자 하였다. 둘째는 센싱·연산의 효율화로, 태스크에 필요한 영역만 선택적으로 센싱·처리·출력하고 불필요한 부분에 대해서는 수행하지 않음으로써 데이터의 획득, 이동, 처리, 출력에 필요한 전력을 함께 줄일 수 있다는 점이다.

III. 사례 1: 동적 피드백 ISP 기반 ToF 손 자세 추정 시스템

1. 배경: ToF 센싱과 손 자세 추정의 과제

손 자세 추정(HPE)은 모바일·AR/VR 기기에서 사용자의 의도를 이해하는 핵심 상호작용 수단이다[3]. 3차원 손 자세를 얻으려면 깊이 정보가 필요하며, iToF 센서는 변조광의 위상 차이로 거리를 측정한다[4, 5]. 그러나 센서가 출력하는 위상 정보는 추론 모델이 곧바로 사용할 수 없어 깊이·point cloud로의 변환이 필요하고, 더 정밀한 분석을 위해 고해상도 영상을 사용하면 연산량과 전력이 급격히 늘어난다. 딥러닝 기반 HPE 역시 많은 연산과 자원

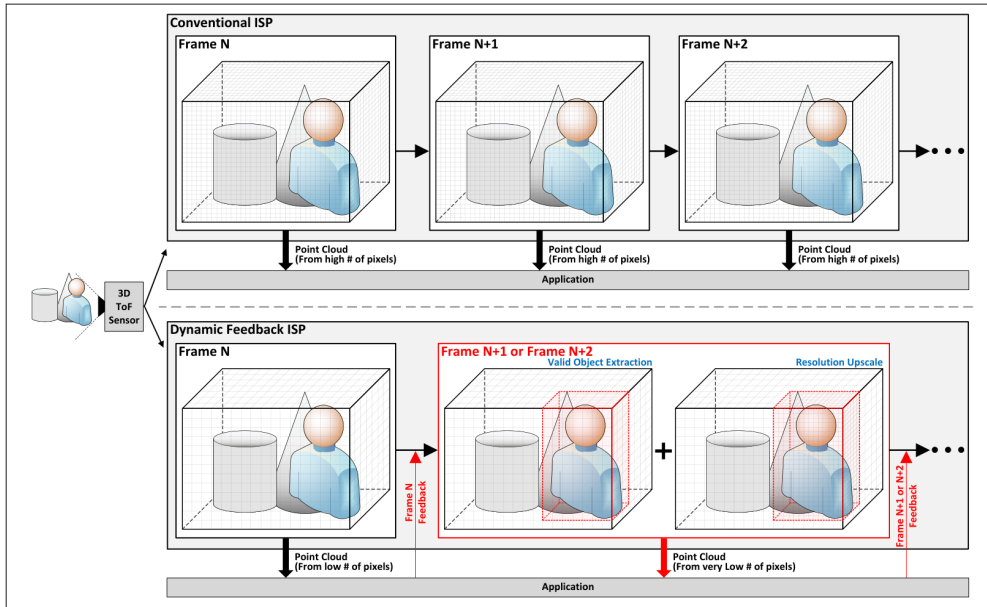
을 요구하여 소형·저전력 기기에서 실시간으로 구동하기 어렵다. 따라서 정확도·효율·속도 사이의 균형이 핵심 과제가 된다.

2. 구조와 동작

이 사례는 iToF 센서를 위한 ISP 및 HPE 가속기를 FPGA에 통합하여 추론에 이르는 전 과정을 near-sensor에서 처리하도록 하였으며, 추론 모델로는 정확도가 높으면서 파라미터 수가 적은 HandFoldingNet[6]을 사용하였다. 이때 ISP의 기본적인 역할은 센서의 위상 정보를 depth와 point cloud로 변환 및 전처리하여 단순히 가속기에 전달하는 역할이나, 본 사례에서는 ISP가 추론 태스크의 결과를 이용하여 다음 프레임 변환과 전처리를 최적으로 수행할 수 있도록 동적 피드백(dynamic feedback) 구조로 설계하였다는 점이다.

〈그림 1〉과 같이, 동적 피드백 ISP[7]는 두 가지 적응적 작업을 수행한다. 첫째, 동적 영역 추출(dynamic area extraction)로 매 프레임 응용이 실제로 사용하는 유효 영역(valid area)의 화소에 대해서만 연산과 후처리를 수행한다. 둘째, 동적 해상도(dynamic resolution)로 프레임마다 적절한 해상도를 판단해 적용한다. 즉, 직전 프레임에서 손 영역이 차지한 위치와 크기를 피드백으로 받아 다음 프레임에서 처리할 영역과 해상도를 조절함으로써, 센서 데이터 전체를 처리하지 않고도 정확도를 유지하거나 오히려 높일 수 있다.

동적 피드백 ISP를 HPE 가속기와 함께 단일 FPGA에 통합한 시스템[8]에서는 추론 가속기 또한 저전력으로 설계하였다. 처리 요소(PE) 단위의 동적 on/off 제어를 통해 사용하지 않는 연산기 영역을 끄고, half-size shift buffer로 systolic array의 buffer 면적을 줄였으며, 센싱·ISP·추론을 병렬·pipeline으로 결합하여 순차 처리 대비 약 2배의 성능을 얻었다. 또한 DSP 대신 LUT와 FF만 사용하여 전력과 면적을 추가로 낮추었고, 유연한 sensor connector로 다양한 iToF 센서를 하드웨어 재설계 없이



<그림 1> 기존 ISP와 동적 피드백 ISP의 동작 비교 개요. 응용의 피드백으로 유효 영역만 추출하고 해상도를 조절해 처리 화소 수를 줄임으로써 연산과 전력을 절감한다.

연결하였다.

3. 결과

동적 피드백 ISP는 최대 178 fps의 처리 속도와 123.15

fps/W의 에너지 효율을 달성하였으며, HPE 가속기와 연결했을 때 <표 2>와 같이 손 자세 추정의 평균 오차를 최대 20.25mm에서 10.03mm로 낮추어 유효 영역 추출과 해상도 조절이 추정 정확도를 함께 높임을 확인하였다. 동적 피드백 ISP와 가속기를 통합한 전체 시스템

<표 2> 동적 피드백 ISP 적용 시(Proposed)와 기존 ISP(Baseline)의 손 자세 추정 결과 비교(HandFoldingNet, NYU 데이터셋)

Original	Baseline	Proposed	Original	Baseline	Proposed
	 Error = 7.09 mm	 Error = 4.39 mm		 Error = 8.89 mm	 Error = 7.44 mm
	 Error = 39.07 mm	 Error = 12.3 mm		 Error = 14.15 mm	 Error = 7.86 mm
	 Error = 22.16 mm	 Error = 10.06 mm		 Error = 22.10 mm	 Error = 4.39 mm

은 KU115 FPGA 보드에서 검증되었으며, 평균 오차 7.78mm로 소프트웨어 모델과 유사한 정확도를 유지하면서 초당 52프레임(fps)의 실시간 처리와 최대 61.74 GOPs/W의 에너지 효율을 달성하여, GPU가 적합하지 않은 저전력·소형 기기 환경에서의 실시간 HPE 가능성을 입증하였다.

IV. 사례 2: in-sensor neural 피드백 이벤트 카메라

사례 1이 추론과 ISP를 near-sensor FPGA에 통합하고 응용의 유효 영역을 피드백으로 활용한 구조였다면, 다음 단계의 목표는 이 통합을 센서 칩 내부로 가져와 센싱 동작 자체를 실시간으로 제어하는 것이다. 이 사례에서는 추론을 위한 프로세서 부분을 센서와 함께 동일 칩 안에 두고, 직전 프레임의 추론 결과로 센서의 동작 파라미터와 출력을 직접 조절하는 in-sensor 적응형 closed-loop를 구현하였다.

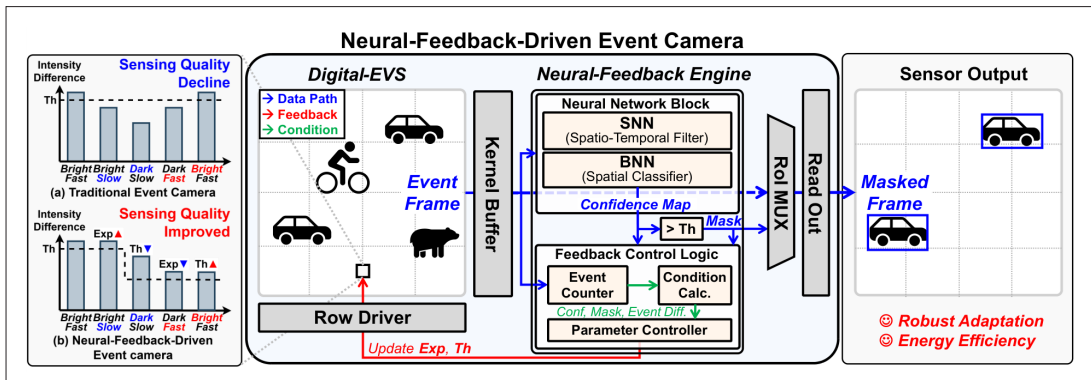
벤트를 발생시켜, 마이크로초 수준의 시간 분해능과 높은 동적 범위, 최소한 데이터 출력을 제공한다[9, 10]. 이벤트 발생률과 민감도는 이벤트 임계값(E_{thr})으로 결정되는데, 조도·움직임이 급변하는 open-loop 동작에서는 event flooding이나 event starvation에 빠져 후단 태스크의 탐지 성능이 저하된다[11]. 기존 아날로그 bias 기반 E_{thr} 제어가 공정·전압·온도(PVT) 변동에 취약하고 정착 거동이 비결정론적이라는[12, 13] 단점을 보완하여, 최근의 digital-EVS는 아날로그 비교기를 화소별 디지털 counter로 대체하여 노출 시간(T_{exp})과 E_{thr} 를 register로 결정론적으로 제어할 수 있게 하였다[14, 15]. 그러나 이들도 대부분 파라미터를 고정하거나 출력 안정화를 타겟으로만 동작한다. 전역 이벤트 통계로 bias를 조정하거나[16] 검출기 신뢰도로 bias를 수정하는[17] 적응형 센싱, 신경 제어기로 시스템 수준 거동을 조정하는 기법[18] 등도 제안되었으나, 태스크와의 직접 연결이 약하거나 off-chip 처리에 의존하여 프레임 단위의 결정론적 제어를 제공하지 못했다.

2. 구조와 동작

1. 배경: 이벤트 비전 센서와 open-loop 한계

이벤트 비전 센서(event vision sensor, EVS)는 각 화소가 밝기 변화량이 임계값을 넘을 때만 비동기적으로 이

〈그림 2〉와 같이, 이 사례는 digital-EVS 전단부와 경량화된 neural network 엔진을 하나의 센서 시스템 안에 두어 실시간 closed-loop를 형성하였다[19, 20]. 경



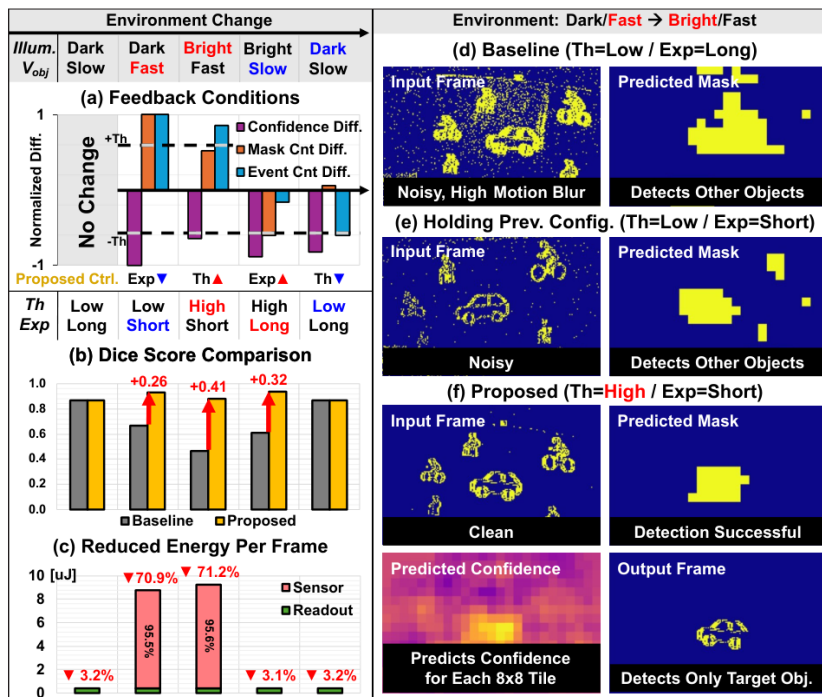
〈그림 2〉 in-sensor 이벤트 카메라의 적응형 closed-loop 구조. 직전 프레임 객체 검출 결과가 노출·임계값 제어와 Roi 기반 선택적 readout에 사용된다.

량 하이브리드 SNN-BNN 검출기가 각 이벤트 프레임에서 tile 단위의 신뢰도 점수 맵과 관심영역(RoI) 마스크를 생성한 후 이를 다음의 두 가지의 작업에 활용하도록 설계하였다. 우선 검출된 객체 영역만 센서 외부로 전송하도록 readout을 제어하여 데이터 이동과 에너지를 줄이고, 직전 프레임의 신뢰도와 마스크 움직임 단서를 입력으로 다음 프레임의 노출(T_{exp})과 임계값(E_{thr})을 조절하도록 하였다. 이러한 제어를 위해 신뢰도 변화, 마스크 변위, 이벤트 개수의 세 단서를 기반으로 물체 움직임에 의한 교란과 조도 변화에 따른 교란을 구분하여 노출과 임계값을 조정하고, 신뢰도가 안정적인 때는 노출을 최소화하여 에너지를 절감할 수 있다. 센서 readout과 neural network 엔진은 row-synchronous pipeline으로 공동 설계되어, 전체 프레임 획득을 기다리지 않고 현재 프레임 안에서 센싱 파라

미터와 출력 영역을 함께 조정하는 프레임 내 적응을 가능하도록 하였다.

3. 검증 및 결과

구현된 HW는 programmable LED panel과 모터 구동 선형 actuator를 통해 다양한 조도·속도의 장면 전환 sequence를 구성하여 성능과 효율성을 측정하였으며, 고정 파라미터로 동작하는 정적 EVS를 baseline으로 비교하였다. <그림 3>과 같이, 제안 구조는 정적 기준선 대비 객체 검출 품질(Dice 점수)을 최대 0.41만큼 회복하였고, 고속 움직임 조건에서 최대 71.2%의 에너지를 절감하여, 직전 프레임 추론 결과로 센서를 조절하는 적응형 closed-loop가 동적 환경에서 강건성과 효율을 동시에 달성함을 보였다.



<그림 3> 사례 2의 평가 결과. 다양한 조도·움직임 전환에서 (b) 검출 품질(Dice) 회복과 (c) 프레임당 에너지 절감을 보인다. 예시-(d) 제어가 없는 베이스라인, (e) 직전 프레임의 파라미터 사용, (f) 제안 기법

V. 결론 및 전망

본 기고문에서는 직전 프레임의 추론 결과를 바탕으로 다음 프레임의 센싱·연산을 조절하여 추론을 더 정확하고 효율적으로 수행한다는 공통 목표 아래, 대상 센서·추론 태스크·조절 파라미터의 세 요소가 서로 다른 두 가지 하드웨어 사례를 살펴보았다. ToF 손 자세 추정 시스템은 응용의 유효 영역을 피드백으로 받아 ISP의 처리 영역과 해상도를 조절하는 동적 피드백 ISP를 추론 가속기와 함께 near-sensor FPGA에 통합한 구조로 저전력 실시간 처리

를 달성하였고, 이벤트 카메라는 동일 개념을 센서 칩 내부로 발전시켜 직전 검출 결과로 센싱 파라미터를 실시간으로 조절하는 in-sensor 적응형 closed-loop로 동적 환경의 강건성과 효율을 함께 확보하였다.

향후에는 두 접근의 장점을 결합하여, 센서 내부의 적응형 센싱과 응용 인지형 처리·연산을 하나의 시스템으로 통합하는 방향을 생각해 볼 수 있다. 나아가 다양한 task와 센서 modality로의 확장, 추론과 센싱 사이의 표준화된 제어 interface 등이 이러한 지능형 센서를 엣지 비전의 일반적인 설계 기반으로 확립하는 데 중요한 과제가 될 것이다.

참 고 문 헌

- [1] Y. Baek et al., "Edge intelligence through in-sensor and near-sensor computing for the artificial intelligence of things," npj Unconventional Computing, vol. 2, no. 1, p. 25, 2025.
- [2] Y. Liu, R. Fan, J. Guo, H. Ni, and M. U. M. Bhutta, "In-sensor visual perception and inference," Intelligent Computing, vol. 2, p. 0043, 2023.
- [3] W. Chen et al., "A survey on hand pose estimation with wearable sensors and computer-vision-based methods," Sensors, vol. 20, no. 4, p. 1074, 2020.
- [4] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," IEEE Sensors J., vol. 11, no. 9, pp. 1917-1926, 2011.
- [5] C. Bamji et al., "A review of indirect time-of-flight technologies," IEEE Trans. Electron Devices, vol. 69, no. 6, pp. 2779-2793, 2022.
- [6] W. Cheng, J. H. Park, and J. H. Ko, "HandFoldingNet: A 3-D hand pose estimation network using multiscale-feature guided folding of a 2-D hand skeleton," in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2021, pp. 11260-11269.
- [7] Y. Kim, J. So, C. Hwang, W. Cheng, and J. H. Ko, "An energy-efficient dynamic feedback image signal processor for three-dimensional time-of-flight sensors," Sensors, vol. 24, no. 21, art. 6918, 2024.
- [8] Y. Kim, J. So, C. Hwang, W. Cheng, J. Choi, and J. H. Ko, "An FPGA-based energy-efficient real-time hand pose estimation system with an integrated image signal processor for indirect 3-D time-of-flight sensors," IEEE Internet of Things J., vol. 12, no. 2, pp. 1817-1830, 2025.
- [9] G. Gallego et al., "Event-based vision: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 1, pp. 154-180, 2020.
- [10] B. Chakravarthi, A. A. Verma, K. Daniilidis, C. Fermuller, and Y. Yang, "Recent event camera innovations: A survey," in Proc. European Conf. Computer Vision (ECCV). Springer, 2024, pp. 342-376.
- [11] W. Shariff et al., "Event cameras in automotive sensing: A review," IEEE Access, vol. 12, pp. 51275-51306, 2024.
- [12] T. Serrano-Gotarredona and B. Linares-Barranco, "A 128×128 1.5% contrast sensitivity 0.9% FPN 3 μs latency 4 mW asynchronous frame-free dynamic vision sensor using transimpedance preamplifiers," IEEE J. Solid-State Circuits, vol. 48, no. 3, pp. 827-838, 2013.
- [13] B. Son et al., "A 640×480 dynamic vision sensor with a 9 μm pixel and 300 Meps address-event representation," in Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC), 2017, pp. 66-67.
- [14] J. Han et al., "A digital dynamic vision sensor with SPAD pixels and multi-event generation for motion/vibration-adaptive detection," in Proc. IEEE Symp. VLSI Technology and Circuits, 2024, pp. 1-2.
- [15] H. Lee et al., "All-digital event-based vision sensor with multi-event generation for motion/vibration-adaptive detection," IEEE J. Solid-State Circuits, 2025.

참 고 문 헌

- [16] T. Delbruck, R. Graca, and M. Paluch, "Feedback control of event cameras," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 1324-1332.
- [17] M. S. Dilmaghani, W. Shariff, C. Ryan, J. Lemley, and P. Corcoran, "Autobiasing event cameras," arXiv:2411.00729, 2024.
- [18] B. A. Mudassar et al., "CAMEL: An adaptive camera with embedded machine learning-based sensor parameter control," IEEE J. Emerg. Sel. Topics Circuits Syst., vol. 9, no. 3, pp. 498-508, 2019.
- [19] J. So, H. Lee, C. Hwang, W. Chung, J. Choi, and J. H. Ko, "A neural-feedback-driven event camera for robust and efficient vision processing," in Proc. IEEE/ACM Int. Symp. Low Power Electronics, 2026.
- [20] H. Lee, J. So, C. Hwang, W. Chung, K. Kim, J. Lee, J. Ko, and J. Choi, "All-digital event-based vision sensor with scene adaptive power-saving pixels and three-layer neural network for object detection," in Proc. IEEE Symp. VLSI Technology and Circuits, 2026.

저 자 소 개



고 종 환

- 2004년 : 서울대학교 기계항공공학부, 컴퓨터공학과 학사
- 2006년 : 서울대학교 전기컴퓨터공학과 석사
- 2006년 ~ 2013년 : 국방과학연구소 선임연구원
- 2018년 : Georgia Institute of Technology ECE 박사
- 2019년 ~ 현재 : 성균관대학교 전자전기공학부 교수
- 주관심분야 : 딥러닝 모델 경량화 및 가속, In-memory / in-sensor computing