

VLSI 2026 Review

울산과학기술원 전기및전자공학부 석박통합과정 홍기업

Session 6: Intelligent Vision and Perception Sensors

2026 IEEE/JSAP Symposium on VLSI Technology & Circuits의 Session C6는 Intelligent Vision and Perception Sensors라는 주제로 총 4편의 논문이 발표되었다. Hybrid processing-in-sensor를 이용한 filter-less color vision(C6.1), triple conversion gain과 LOFIC을 결합한 100 dB single-exposure HDR CIS(C6.2), clocked-recharging SPAD와 TTFS encoding을 이용한 neuromorphic image sensor(C6.3), scene-adaptive power saving과 object-of-interest 추출을 결합한 all-digital EVS(C6.4)가 소개되었다. 각 논문은 서로 다른 pixel 및 readout 구조를 사용하지만, sensor 내부에서 데이터 표현을 단순화하고 필요한 정보만 선별하여 전송함으로써 high dynamic range, low power, high throughput 및 edge intelligence를 동시에 확보하려는 공통된 방향성을 보인다.

#6.1 Carnegie Mellon University에서 발표한 논문으로, Computational Mosaic Array(CMA)와 residual-enhanced analog skip connection을 결합한 hybrid processing-in-sensor(PIS)를 제안한다. 기존 processing-in-pixel(PIP)은 높은 병렬성을 제공하지만 pixel area와 weight resolution이 제한되고, processing-in-column(PIC)은 복잡한 연산이 가능하지만 2-D convolution을 위한 row buffer와 memory 부담이 크다. 이를 해결하기 위해 3×3 CMA에서 일부 pixel만 triple-well photodiode와 analog front end로 사용하고 중앙 pixel에 convolution 결과를 저장함으로써 photodiode area를 55% 절감하면서 4-b in-pixel convolution을 구현하였다. Triple-well photodiode는 별도의 color filter 없이 R/G/B 성분을 동시에 획득하며, time-domain AFE가 3-b magnitude와 1-b sign으로 구성된 weight를 current accumulation에 반영한다.

CONV1 출력은 column-wise median 및 max pooling을 통해 2-D feature map에서 21×1 time-domain representation으로 변환되며, 이에 따라 데이터 양을 98.8% 줄이고 row-buffer memory를 제거한다. 이후 3×1 CONV2의 출력은 upscaling된 뒤 CONV1 결과와 analog skip connection으로 합산되고, CONV3와 TDC readout을 거쳐 off-chip classifier로 전달된다. 특히 residual-enhanced network는 analog weight mismatch와 noise가 증가할

때 plain network보다 작은 정확도 저하를 보여 analog PIS의 비이상성에 대한 내성을 향상시켰다. 180-nm CMOS에서 제작된 126×126 sensor array는 40 fps에서 62.08 μW 를 소모하고 97.76 pJ/pixel-frame의 iFoM 및 97.25%의 face-detection accuracy를 달성하였다.

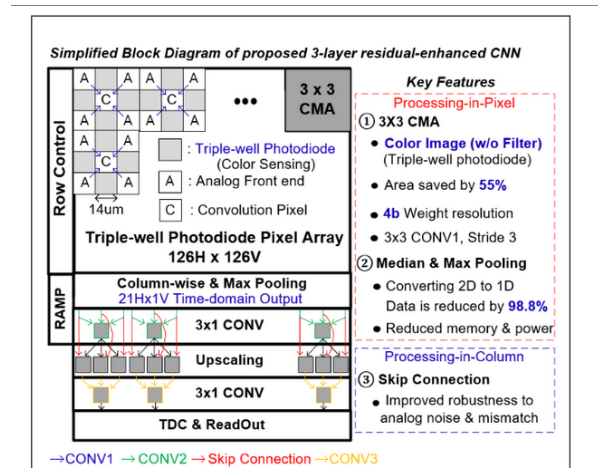


그림. 제안된 3-layer residual-enhanced PIS 전체구조

#6.2 Sony Semiconductor Solutions와 Sony Semiconductor Manufacturing에서 발표한 논문으로, 1.22 μm pixel pitch에서 triple conversion gain과 transfer-gate(TRG)-type LOFIC을 결합한 50-Mpixel single-exposure HDR image sensor를 제안한다. 기존 sub-pixel HDR 구조는 하나의 pixel에 두 개의 photodiode를 배치하므로 2 μm 이하 pitch에서 면적 확보가 어렵다. 제안 pixel은 4개의 photodiode가 FD1을 공유하며, FDG를 통해 metal capacitance가 추가된 FD2를 형성하고 FCG를 통해 대용량 3-D MIM LOFIC capacitor에 연결된다. 조도가 증가하면 charge가 PD에서 FD1, FD2, LOFIC capacitor 순서로 overflow 되므로 1st gain, 2nd gain, LOFIC gain의 세 구간에서 넓은 입력 범위를 읽을 수 있다.

TRG 아래에 overflow path를 형성하면 PD의 charge-storage capacity가 감소할 수 있으므로, 본 설계는 PD maximum potential과 FD reset potential을 함께 높였다. 이때 RST_VDD와 source-follower의 SF_VDD를 분리하여 reset headroom은 확보하면서 source-follower의 전력 증가는 방지하였다. 또한 exposure 동안 MIM_VDD를 3.3 V에서 2.4 V로 낮춰 LOFIC storage node의 electric field를 완화하고 FD dark current를 5,000 e^-/s 에서 3,700 e^-/s 로 감소시켰다. 90-nm CIS와 22-nm ASIC으로 구현된 sensor는 최대 analog gain에서 0.73 e^-_{rms} read noise, LOFIC gain에서 250 ke^- full-well capacity, 그리고 100 dB 이상의 single-exposure dynamic range를 달성하였다.

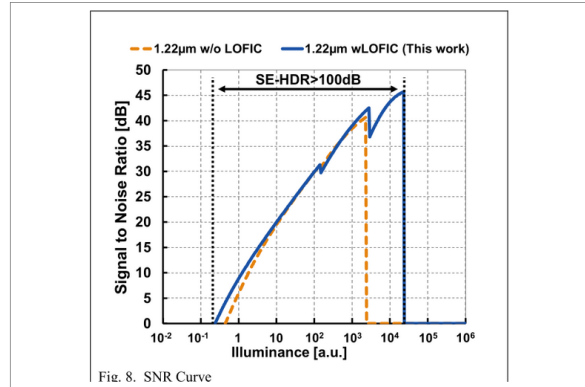
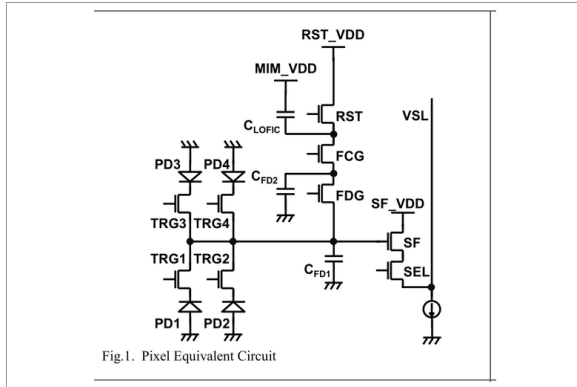


그림. Triple conversion gain과 TRG-type LOFIC을 구현한 pixel equivalent circuit(좌)과 기존 DCG sensor 대비 100 dB 이상의 single-exposure dynamic range를 보이는 SNR curve(우)

#6.3 UNIST와 Sogang University에서 발표한 논문으로, clocked-recharging(CR) SPAD pixel이 광세기를 time-to-first-spike(TTFS)로 직접 encoding하는 특성을 이용한 neuromorphic image sensor를 제안한다. 기존 CNN 기반 PIS는 dense multi-bit data와 반복적인 accumulation으로 인해 데이터 이동과 연산 에너지의 부담이 남아 있는 반면, TTFS 기반 SNN은 하나의 spike 발생 시점만으로 입력 크기를 표현하므로 높은 sparsity와 낮은 latency를 제공한다. 96×96 pixel array의 각 pixel은 CR SPAD, 4-b TTFS memory, 8-b exponential counter, 5-b SRAM 및 pulse generator를 포함하며, 3×3 sub-pixel-array(SPA) 단위로 comparator와 output memory를 공유하여 array-parallel convolution을 수행한다. Intensity mode에서는 recharging cycle 내 첫 spike를 TS0–TS3의 one-hot code로 저장하고, spike timing에 따라 8-b counter에 +8, +4, +2, +1을 누적한다. 16회의 averaging으로 dark-count 영향을 줄인 뒤, convolution mode에서는 counter의 상위 4-b로 dominant spike timing을 decode하고 SRAM의 5-b weight를 pulse 수와 polarity로 변환한다. Chopping analog counter의 4-b CDAC은 TS0–TS3에 각각 8C, 4C, 2C, 1C의 temporal weight를 부여하며, threshold를 negative weight로 함께 주입해 단일 comparator로 integrate-and-fire를 구현한다. Input-stationary dataflow를 통해 동일한 spike input을 재사용하므로 filter마다 다시 exposure하거나 encoding할 필요가 없다. 110-nm BSI 공정으로 제작된 prototype은 150 μs frame time에서 6.67 kfps로 동작하며, convolution mode에서 5.12 mW를 소모한다. 측정된 compute power는 0.5 mW이고, 2.6 pJ/pixel·fps·filter의 FoM, 3.86 TOPS/W의 energy efficiency 및 91.5% face-detection accuracy를 달성하였다.

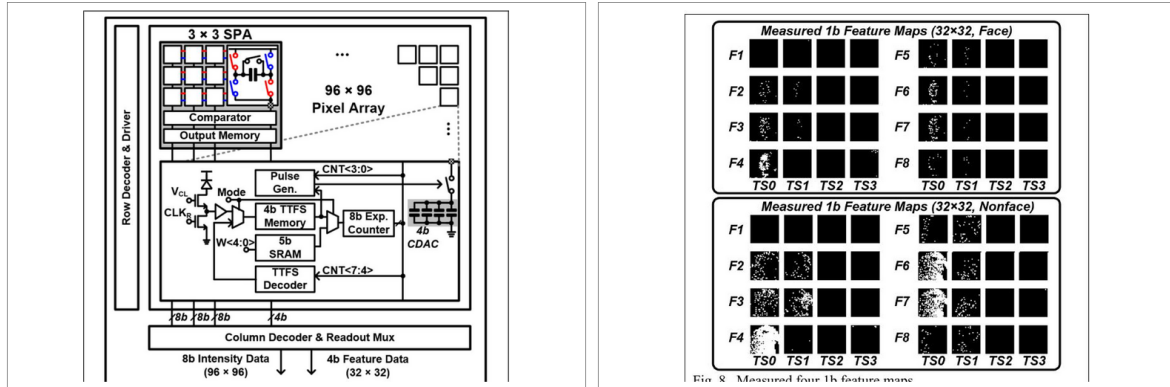


그림. CR SPAD, TTFS memory, exponential counter 및 SPA-level IF 회로로 구성된 overall sensor architecture(좌)와 face/non-face 입력에 대해 측정된 1-b feature maps(우)

#6.4 Sungkyunkwan University와 SolidVue에서 발표한 논문으로, scene-adaptive pixel power saving과 object-of-interest(OOI) selective readout을 결합한 all-digital synchronous event-based vision sensor(EVS)를 제안한다. 200×128 SPAD array는 saturation-based(SAT-PS), temporal(T-PS), spatial(S-PS)의 세 가지 power-saving 기법을 사용한다. SAT-PS는 pixel output이 saturation에 도달하면 이후의 불필요한 counting을 중단하고, T-PS는 짧은 sub-window에서 급격한 intensity change를 검출한 뒤 pixel을 비활성화한다. S-PS는 3개의 primitive pixel과 1개의 reference pixel로 구성된 2×2 macro-pixel에서 temporal intensity trend를 비교하고, 인접 pixel이 유사할 경우 primitive pixel을 끈 뒤 reference event로 overwrite하여 공간적으로 중복된 동작을 줄인다.

Event decision 및 interpolation logic에서 생성된 event는 48-row kernel buffer에 저장되는 동시에 3-layer hybrid SNN-BNN으로 전달된다. 첫 번째 SNN layer가 spatial-temporal filtering을 수행하고, 두 개의 BNN layer가 1-b weight와 activation을 이용해 8×8 tile 단위의 OOI confidence map을 생성한다. 이 OOI mask로 readout을 gating함으로써 관심 영역의 event만 외부로 전송하고, 선택된 event 수를 이용해 다음 frame의 threshold도 scene-adaptive하게 갱신한다. 90-nm 공정의 prototype은 200 lux에서 9.85 mW를 소모하며 pixel 및 core power를 각각 47.6%와 21.6% 절감하였다. S-PS 단독 조건에서는 최대 53.77%, 세 기법이 동시에 동작하는 composite scene에서는 51.4%의 pixel power saving을 보였다. 측정 throughput은 DAQ에 의해 제한된 5.58 kfps 및 714 Meps이다. On-chip OOI block은 IR-drop으로 목표 100 MHz에서 안정적으로 동작하지 않아 동일 network를 FPGA로 구현해 검증하였으며, sensor-captured event image에서 0.7452의 Dice score를 달성하였다.

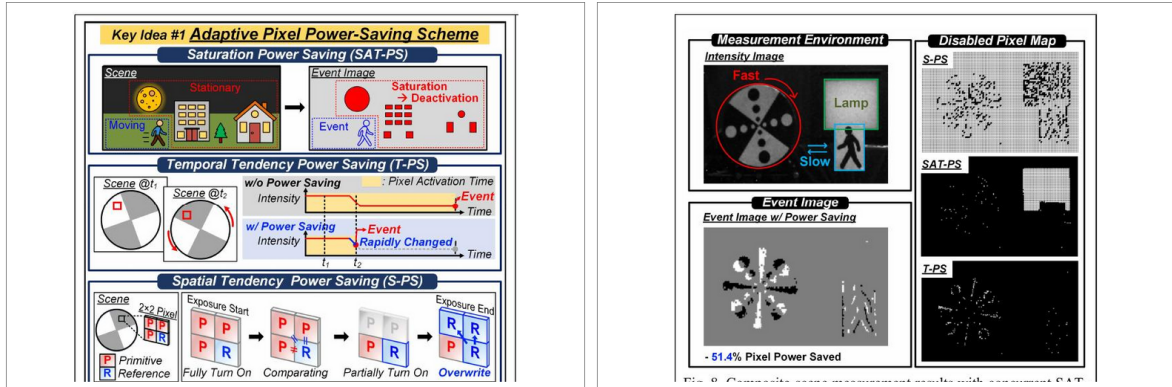


그림. SAT-PS, T-PS 및 S-PS로 구성된 scene-adaptive pixel power-saving 개념(좌)과 세 기법을 동시에 적용한 composite-scene 측정 결과(우)

저자정보



홍기업 석박사통합과정 대학원생

- 소속 : 울산과학기술원
- 연구분야 : 센서디자인, 혼성회로설계
- 이메일 : slsnsep357@unist.ac.kr
- 홈페이지 : <https://sites.google.com/view/bias-sogang/home>