MS-EVS: Multispectral event-based vision for deep learning based face detection

Saad Himmi*, Vincent Parret+, Ajad Chhatkuli*, Luc Van Gool*

* Computer Vision Laboratory, ETH Zürich, Switzerland + Stuttgart Laboratory 1, Sony Semiconductor Solutions Europe, Sony Europe B.V.

JAN 3-7 WACV 2024 WAIKOLOA · HAWAII

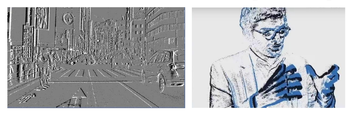
MS-EVS Dataset

What is an event-camera ?

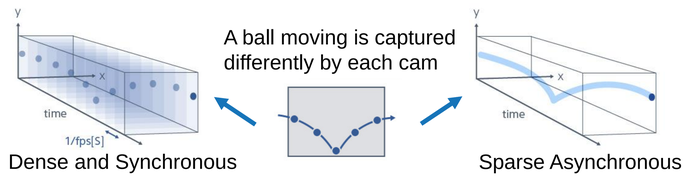
Regular Camera (APS)

- Detailed scene
- ~10 ms latency
- Bad when movement
- Bad at low light

Event Camera (EVS)



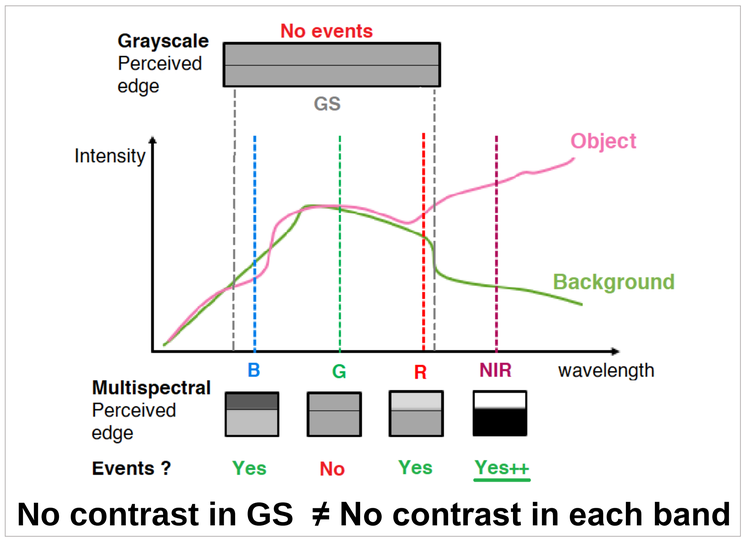
- Sparse (Low power)
- Low (µs) latency
- No motion blur (expos.)
- Better HDR (log)



Motivation

Where is Waldo ?
← Grayscale (GS) vs Color →

Existing EVS are MONOCHROMATIC (Grayscale)
→Any benefit from color or infrared?
No existing dataset to explore if MS-EVS is worth it



No contrast in GS ≠ No contrast in each band

Main contributions

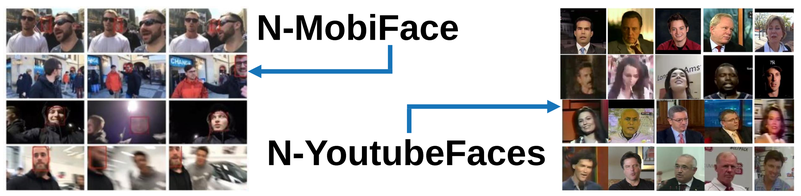
APS+EVS dataset for Face Detection
Use **simulation** to create **large-scale** data
(Sets: 2 large RGB – 1 MS – 1 real events)

First "fair" APS vs EVS comparison
Same model, Same data and simulated EVS
N.B: results are checked for real events too.

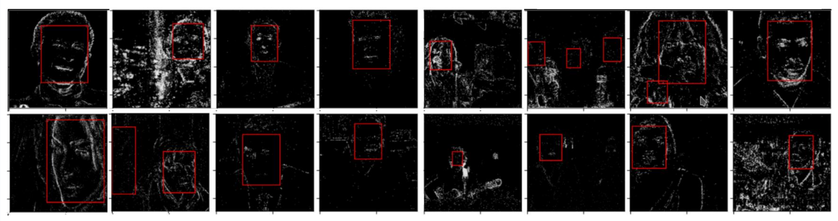
Introduce Multispectral events (MS-EVS)
Capture event-based data over multiple bands and **explore benefits of MS-EVS over GS**

Introducing MS-EVS Dataset

Large RGB datasets for pre-training



Contains both APS and simulated events for 3 bands

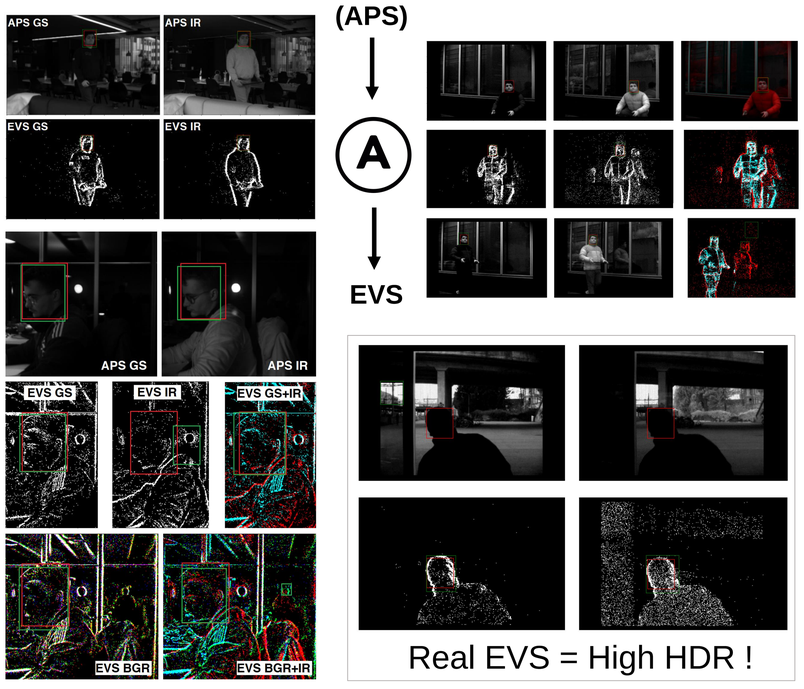


TOTAL : ~3.5k videos at 24 fps (APS) and their events

HOW ? Reuse Open-Source Datasets (A)

First Multispectral EVS dataset

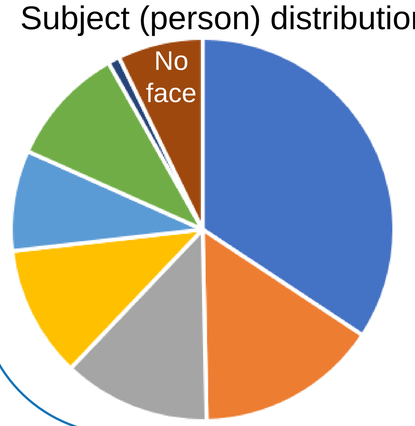
N-SpectralFace		Real-SpectralFace
Multispectral cam	Cam A	EVS + IR cut (=GS)
GS APS + IR pass	Cam B (APS)	EVS + IR pass (=IR)



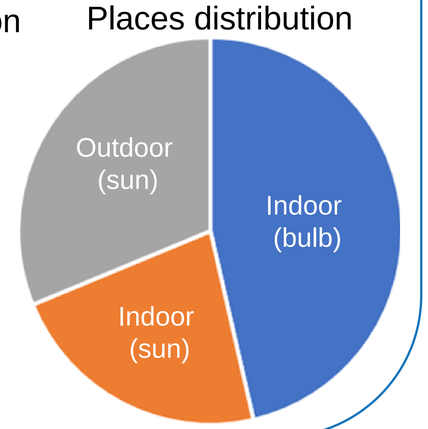
Real EVS = High HDR !

→ **Data for Finetuning and Experiments**

Subject (person) distribution



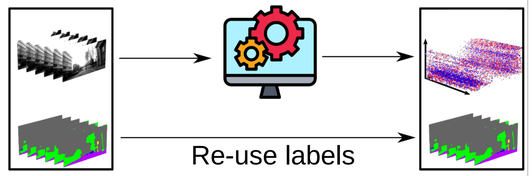
Places distribution



Method

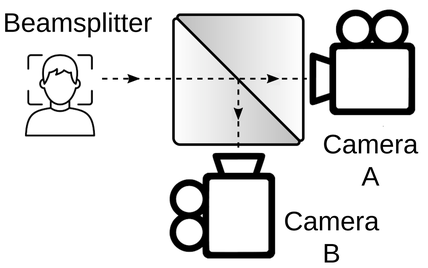
A

- Use existing Data
- "Anticipate" MS-EVS sensor



B

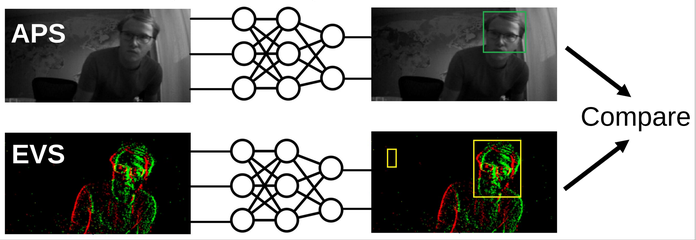
- Capture same scene from 2 cameras = 2 spectrum
- Combined with labelling then (A)= MS-EVS



C

Same spectral band **BUT** Different sensor

Same model (RetinaFace) – Different weights



Results

Single-band (1 color) comparison C

Face Detection mAP after finetuning on N-SpectralFace

Channel	Blue	460nm	500nm	Green	570nm	610nm	Red	700nm	Gray	NIR
APS	**0.672**	**0.670**	**0.666**	**0.667**	**0.673**	**0.670**	**0.663**	**0.672**	**0.668**	0.582
EVS	**0.589**	**0.590**	**0.593**	0.579	0.577	0.576	**0.589**	0.559	**0.592**	0.521

→ **APS > EVS** (more FP in events)
→ **All bands perform the same** (except IR, worse)

Multispectral comparison + Real events

C Now, **the models are trained on multiple bands**:

Channel	N-SpectralFace		Real-SpectralFace	
	APS	EVS	APS	EVS
Grayscale (GS)	0.668	0.592	0.413	0.578
Infrared (IR)	0.582	0.521	**0.548**	0.535
GS+IR	**0.693**	0.626	0.473	**0.607**
Blue+Red (BR)	0.670	0.615	-	-
BR+IR	0.679	0.639	-	-
Color (BGR)	0.665	0.628	-	-
BGR+IR	**0.684**	**0.647**	-	-
BR+GS	0.672	0.626	-	-
BR+GS+IR	**0.688**	0.645	-	-
Hyperspectral (8ch)	0.664	0.645	-	-
8ch+IR	0.681	**0.656**	-	-
8ch+GS (9ch)	0.677	0.645	-	-
9ch+IR	0.671	**0.657**	-	-

Background color is mAP relative to best single band
<0%
0 to 2%
2 to 5%
> 5%

→ **APS - EVS gap reduces** (MS-EVS have less FP)
→ **IR improves mAP for both APS/EVS** (1 to 3%)
→ **MS-EVS benefits MORE from MS than APS**

Lior Wolf, et al. « Face recognition in unconstrained videos with matched background similarity», CVPR 2011.
Jiankang Deng, et al. « Retinaface: Single-stage denseface localisation in the wild», CoRR 2019.
Yiming Lin, et al. « Mobiface: A novel dataset for mobile face tracking in the Wild», CoRR 2019.
Gehrig, Daniel, et al. «Video to events: Recycling video datasets for event cameras», CVPR 2020.