Computer Vision, One Photon at a Time

Mohit Gupta

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Joint work with
Sizhuo Ma, Shantanu Gupta, Andreas Velten, Atul Ingle (UW Madison)
Arin Ulku, Paul Mos, Claudio Bruschini and Edoardo Charbon (EPFL)

Support: NSF, ONR, DARPA, WARF, SONY

Images are Interesting

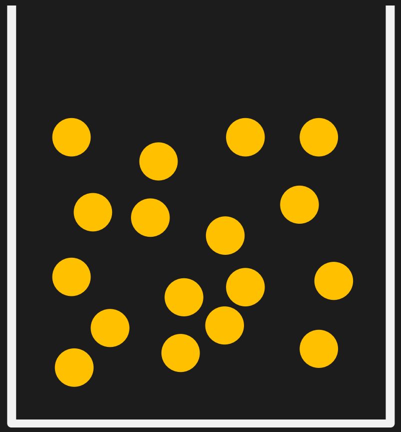
But When you Look Close…

157	159	159	104	104	115	128	131	133	133	132	131	132	130	129	118	132	158	156	153	190	144	117	126	120	81
159	165	153	101	103	113	126	129	130	130	126	124	127	128	127	120	122	158	159	154	160	190	121	118	67	47
162	154	154	98	101	114	124	127	130	132	144	159	155	132	123	119	119	148	154	150	140	185	161	60	48	45
141	132	158	93	98	110	121	125	122	129	143	172	191	188	143	105	117	148	140	145	142	153	105	44	49	71
100	130	157	93	99	110	120	116	116	129	138	163	191	205	211	130	107	153	98	133	147	107	44	47	81	151
87	130	157	92	97	109	124	111	123	134	139	175	194	201	207	205	126	151	74	114	160	57	49	63	141	163
93	131	159	92	98	112	132	108	123	133	162	180	183	192	196	205	184	151	138	199	195	54	47	119	161	156
96	134	164	95	97	113	147	108	125	142	156	171	173	178	184	181	186	191	206	203	161	44	84	158	159	155
95	137	165	95	95	111	168	122	130	137	145	139	144	139	145	179	193	203	194	158	95	49	135	160	157	155
101	139	166	94	96	104	172	130	126	130	108	77	85	80	153	191	188	161	144	113	48	83	161	160	156	153
101	133	167	94	96	100	154	137	123	92	67	57	72	153	182	184	175	101	116	53	48	119	166	163	159	152
99	130	169	97	99	109	131	128	84	55	60	75	149	176	170	194	209	99	79	51	67	150	158	155	154	151
97	129	170	97	98	118	122	94	66	56	56	140	161	114	136	187	163	81	85	52	98	161	159	154	148	137
92	123	173	101	98	129	95	74	74	45	94	174	106	115	126	168	108	60	92	55	128	157	153	148	145	157
81	115	175	104	116	87	78	69	84	56	140	124	158	170	143	173	150	76	90	68	148	153	146	148	186	196
69	108	172	107	103	87	82	54	83	105	93	107	153	166	132	162	153	68	87	97	157	149	141	179	204	206
71	119	172	106	91	78	97	70	99	104	59	116	142	153	141	165	123	55	84	132	154	146	148	199	209	210
61	126	175	112	83	74	92	123	130	53	61	108	137	132	138	154	77	58	82	150	152	143	155	210	211	213
53	128	175	105	71	82	109	127	75	50	57	74	115	139	151	117	47	67	89	154	154	143	159	218	214	199
56	115	173	105	61	76	106	114	70	54	52	60	102	137	160	146	78	67	96	135	130	125	165	215	142	81
117	106	176	101	55	71	81	112	101	57	55	70	117	139	152	188	198	112	87	146	131	112	178	164	81	91
107	121	177	89	50	64	60	103	114	66	56	90	120	140	149	169	201	194	100	148	134	155	208	120	99	99

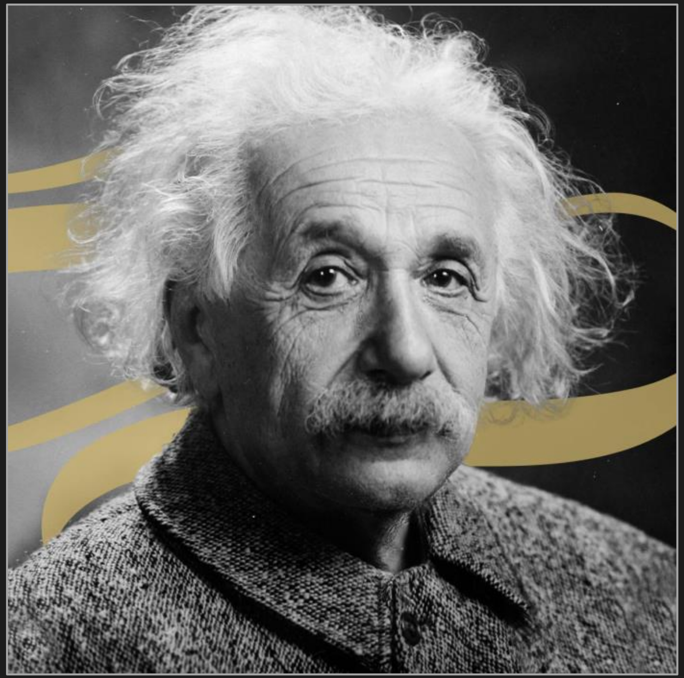
pixel

And When you Look Even Closer...

~100-1000 photons



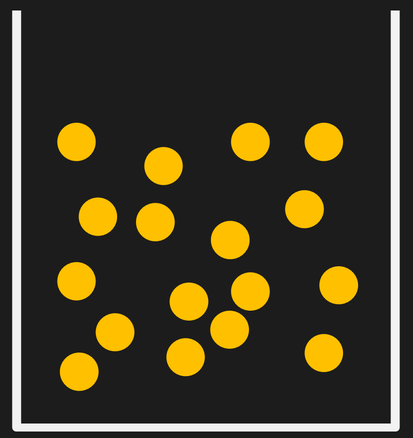
conventional camera pixel



photons as quantas

Single-Photon Cameras

~100-1000 photons



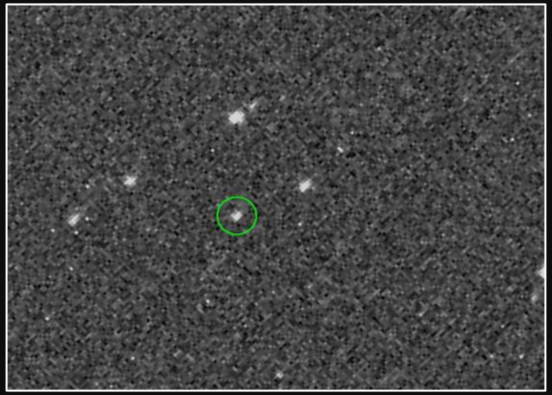
conventional camera pixel

single photon

single-photon camera pixel

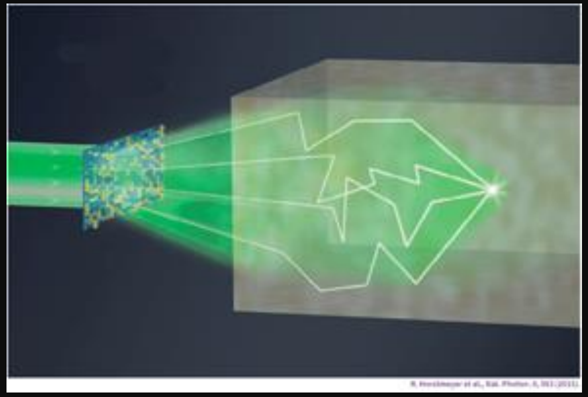
single-photon sensitivity

Single-Photon Cameras: Scientific Imaging

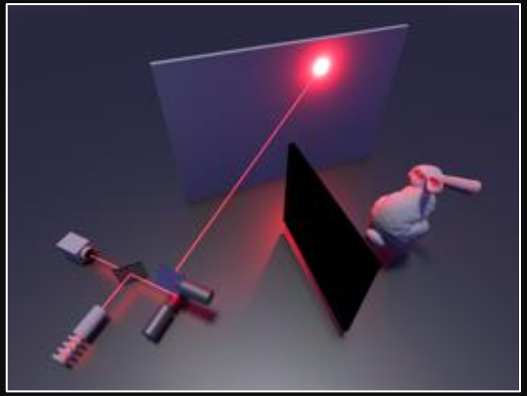


Astronomy

Microscopy



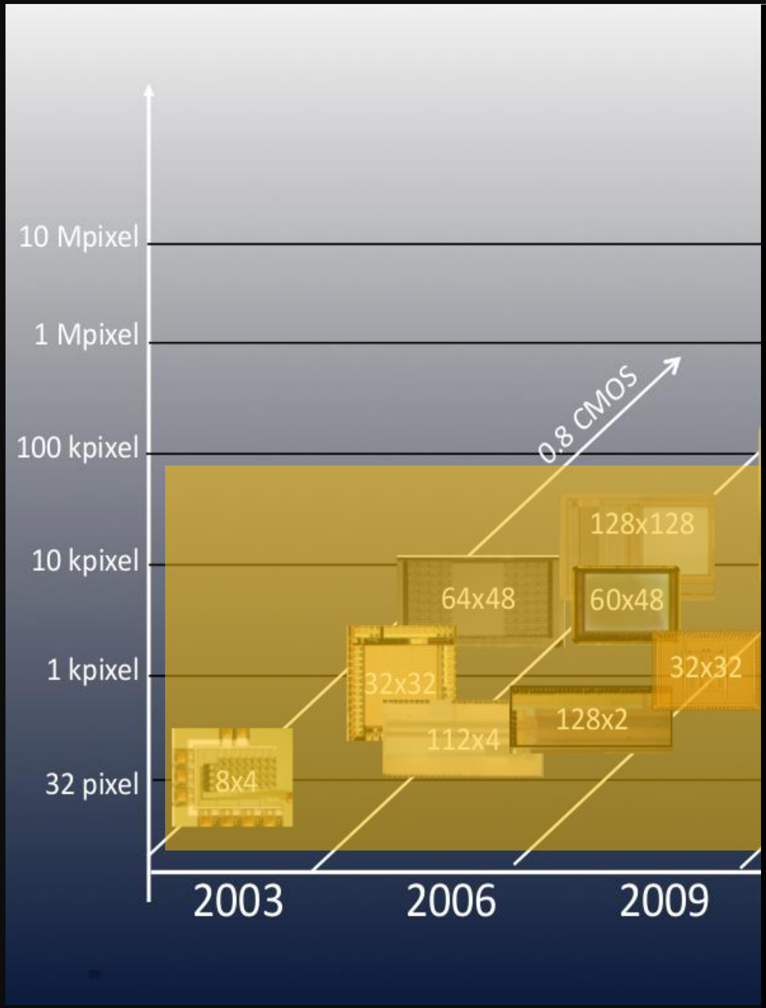
Scattering media



NLOS Imaging

images courtesy: http://www.noao.edu/, http://www.futurahma.it/ , http://www.computationalimaging.org/, http://www.upi.com/ , www.openmv.io

The Single-Photon Revolution

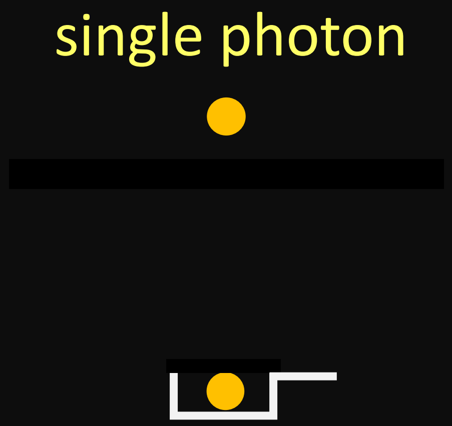


SONY
Canon
Apple
…

Emergence of Large-Format Single-Photon Cameras

A. C. Ulku *et al.*, “A 512 × 512 SPAD Image Sensor With Integrated Gating for Widefield FLIM,” *IEEE J. Select. Topics Quantum Electron.*, vol. 25, no. 1, pp. 1–12, Jan. 2019,
K. Morimoto et al., “Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications,” Optica, vol. 7, no. 4, pp. 346–354, Apr. 2020..

Single-Photon Cameras: Attractive Features



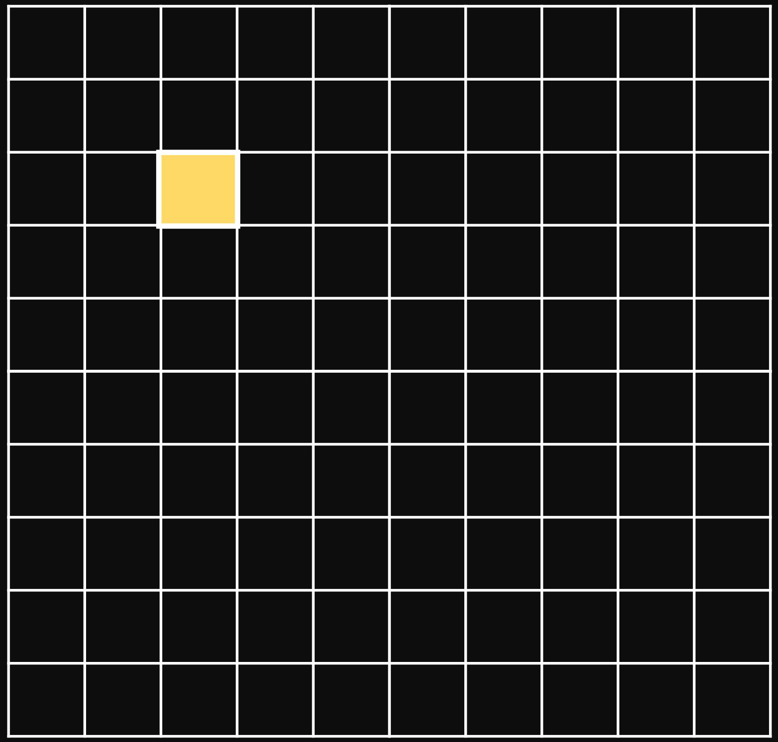
Single-Photon Sensitivity

Room Temperature Operation

CMOS compatible
Low cost, Compact

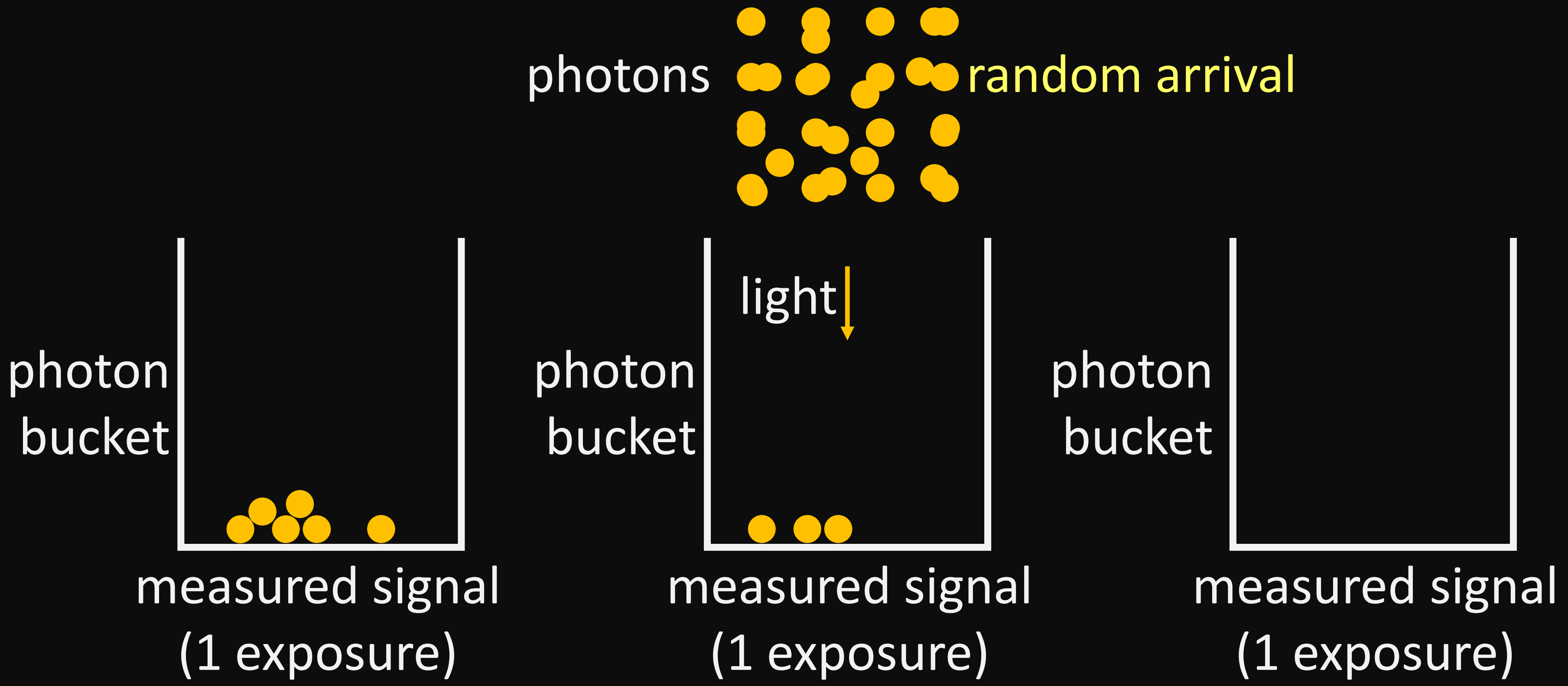
Single-Photon Camera (SPC): Image Formation

Pixel: Photon ‘Tea-Spoon’



Camera Pixel Array

Random Arrival of Photons



arrival of light is a discrete random process (Poisson)

ϕ: mean photon flux (per exposure)

Example: $\phi = 3$ Photons

Random Arrival of Photons: Single-Photon Cameras

SPC pixel measurements are binary random variables

B: recorded value $= 1$

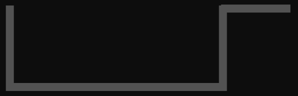
photon teaspoon

B: recorded value $= 1$

photon teaspoon

B: recorded value $= 0$

photon teaspoon



ϕ: mean photon flux (per exposure)

Example: $\phi = 3$ Photons

Random Arrival of Photons: Single-Photon Cameras

SPC pixel measurements are binary random variables

$$P(B = 0) = e^{-\phi}$$

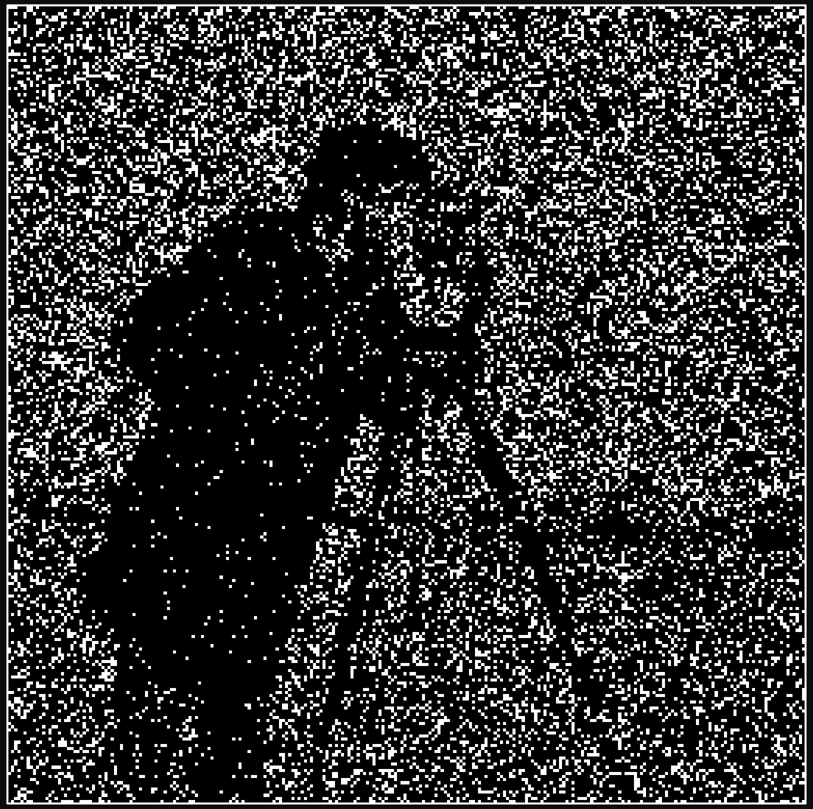
$$P(B = 1) = 1 - e^{-\phi}$$

SPC Image Formation Model

ϕ: mean photon flux (per exposure)

Single-Photon Camera Image: Binary

Quanta Image

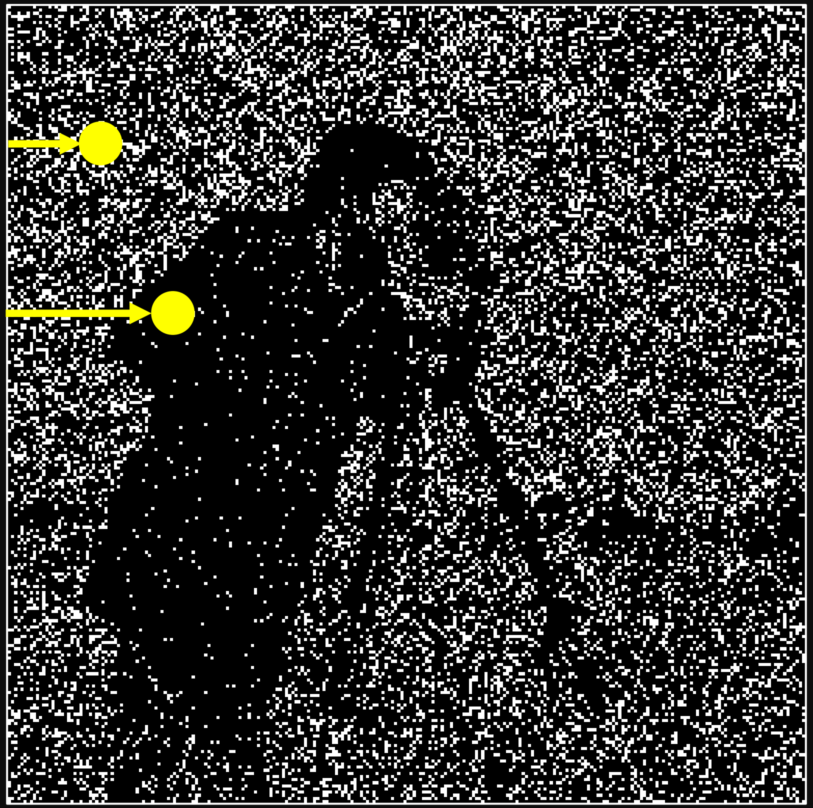


Noise: randomness of photon arrivals (shot noise)
Negligible read noise

Single-Photon Camera Image: Random

Bright pixel: High frequency of 1s

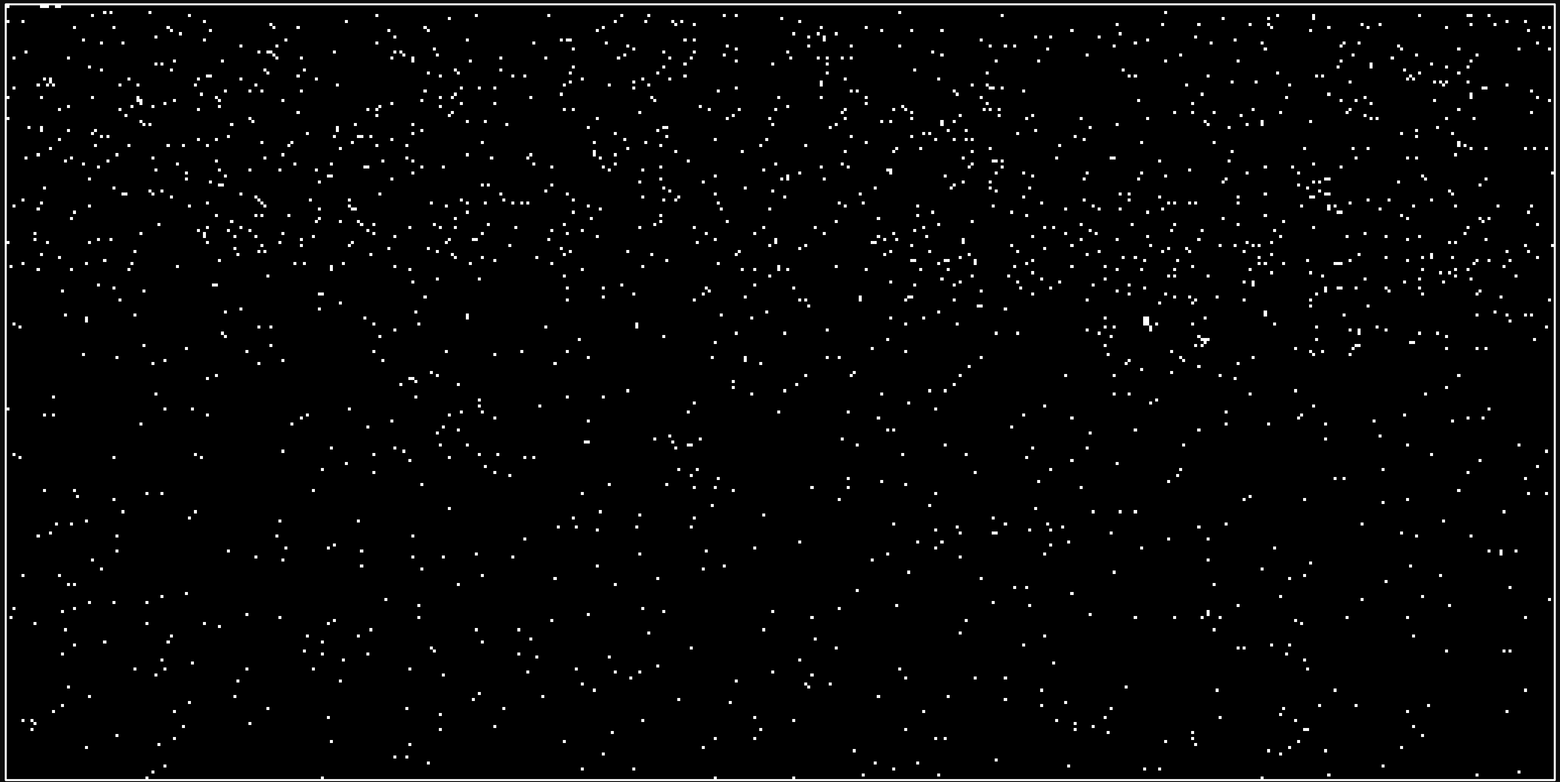
Dark pixel: Low frequency of 1s



High-Speed Binary Capture (~100K FPS)

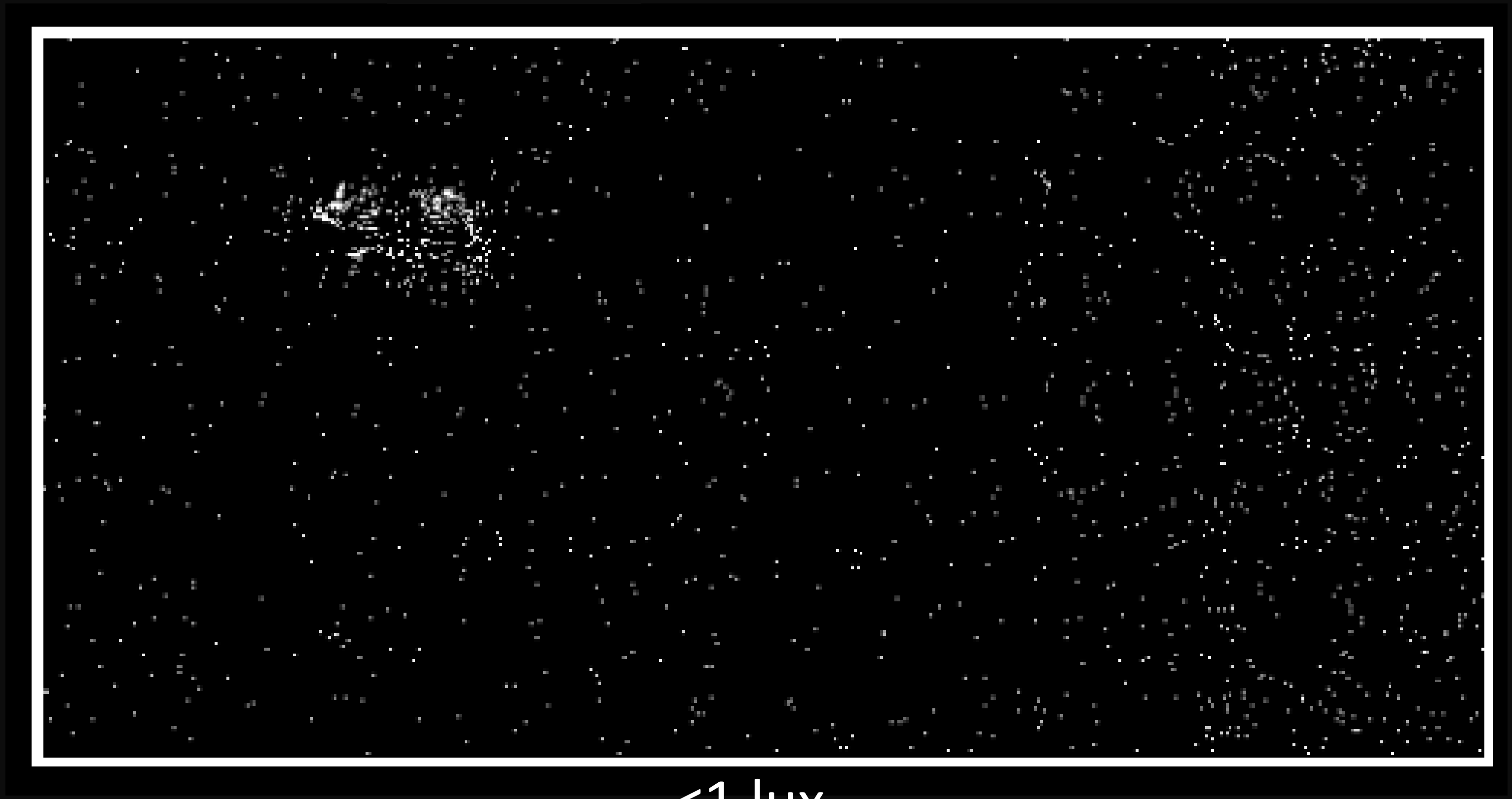
Single-Photon Cameras Capture Photon Streams

pixel measurements are Bernoulli (binary) random variables



Can you tell what the scene looks like?

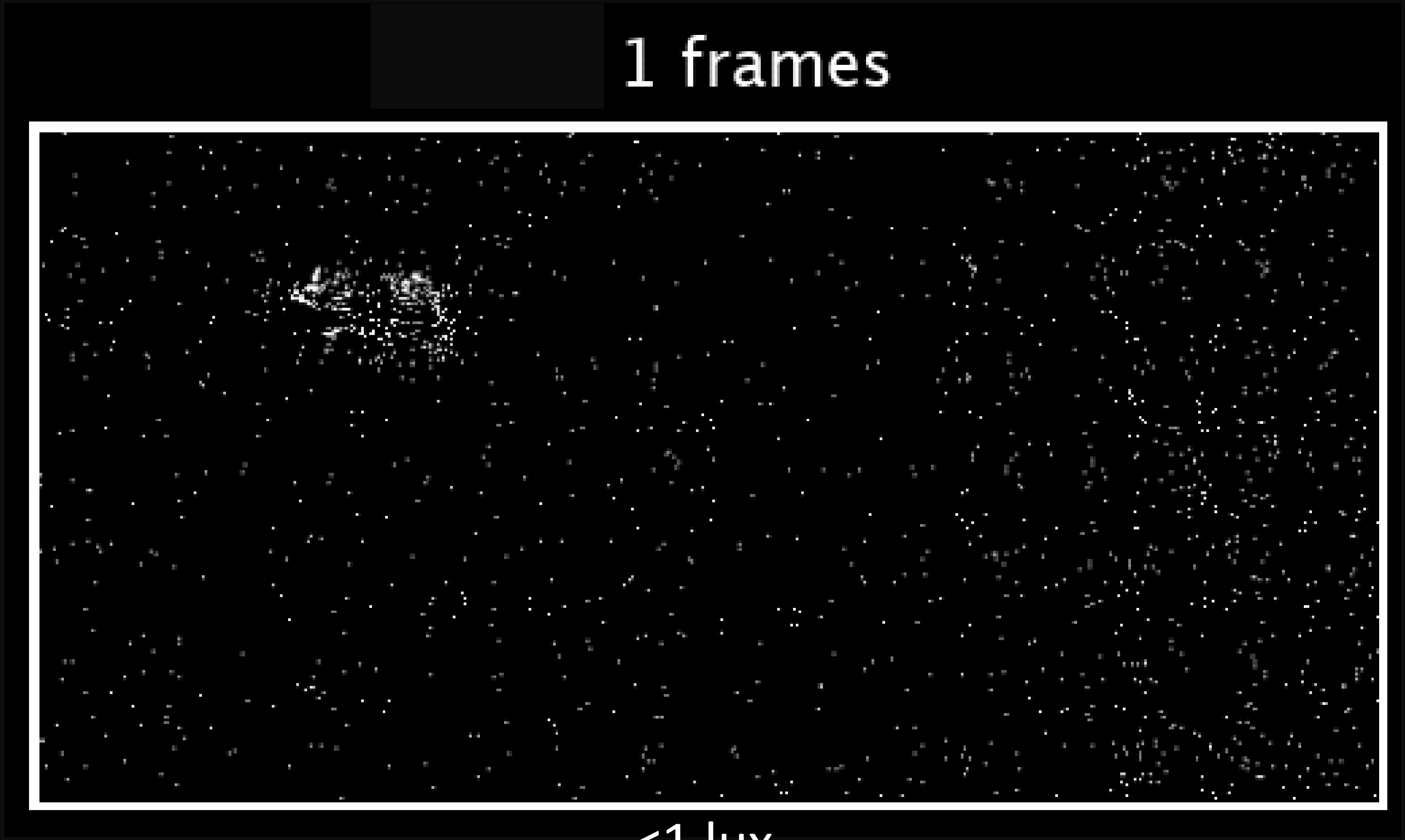
Object Detection: Single Quanta Frame



<1 lux

Extremely Dark Scene

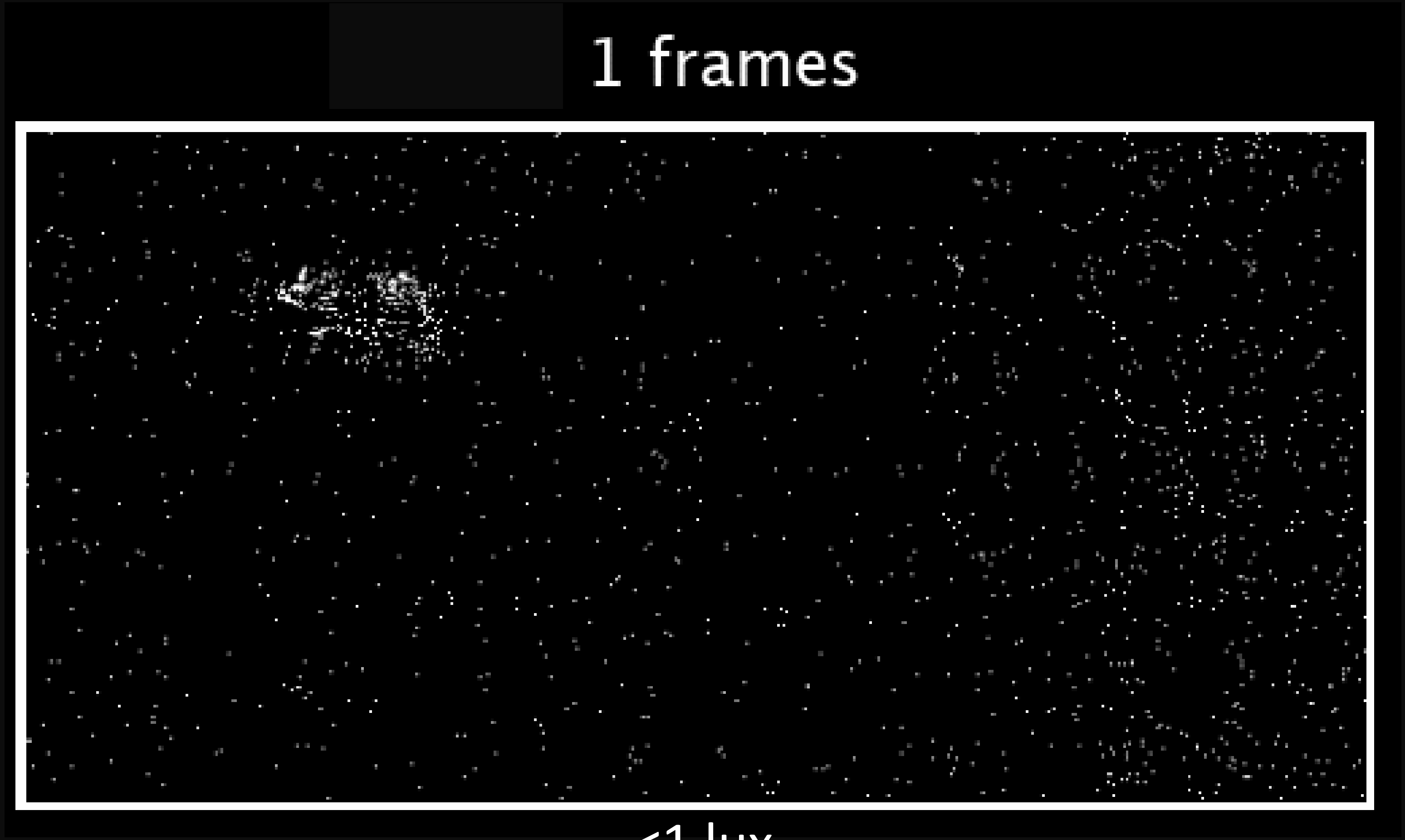
Object Detection: Naïve Approach



<1 lux

High Blur or High Noise: Detection Fails in All Cases

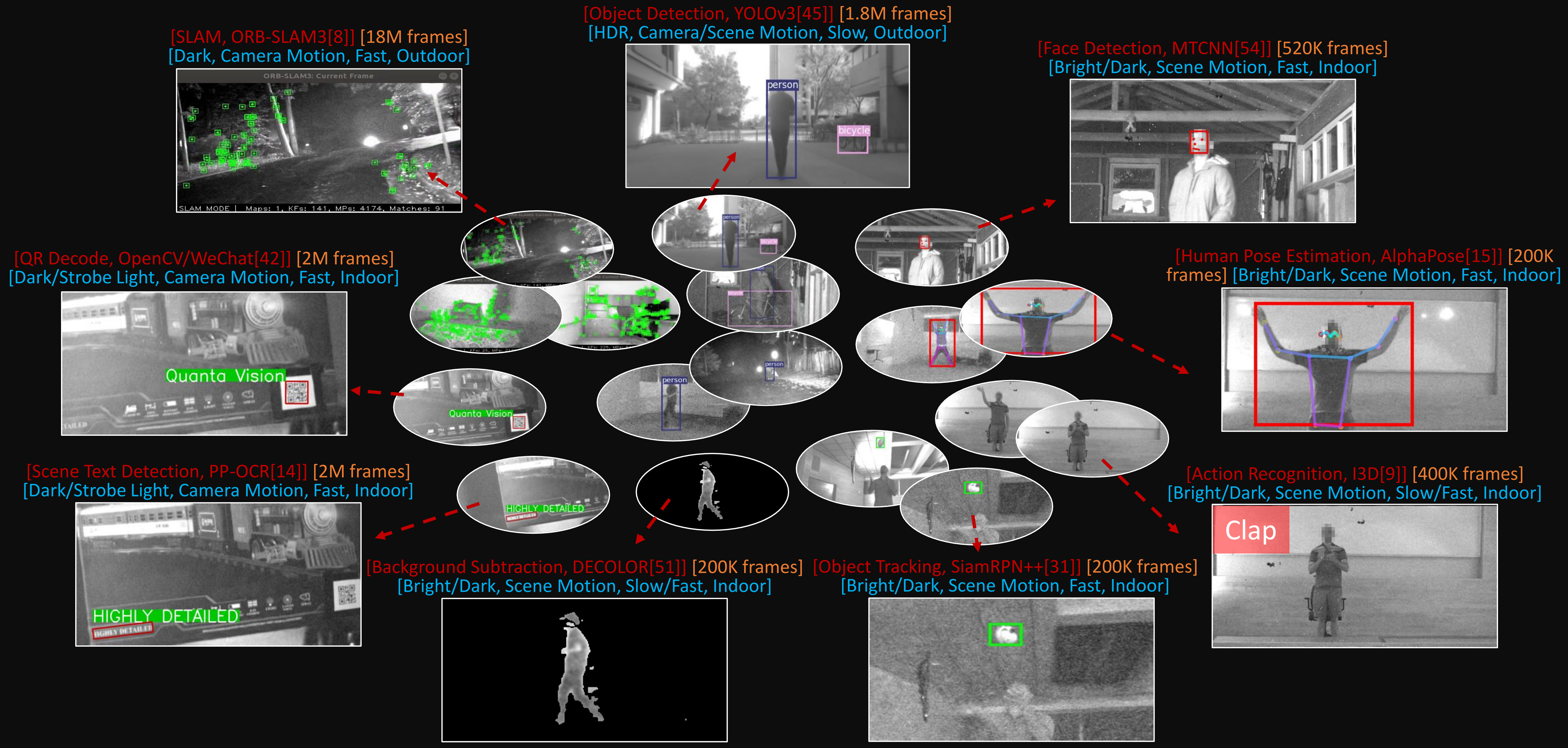
Object Detection: Quanta Vision Algorithm



<1 lux

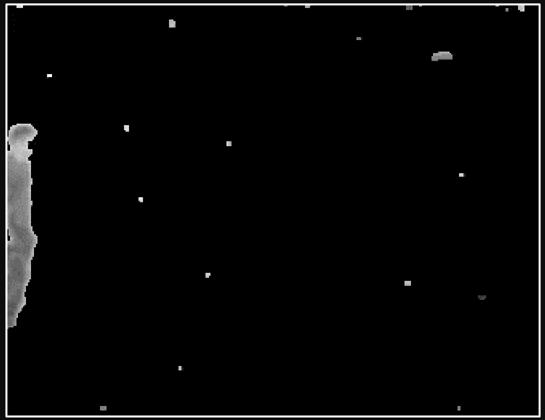
Low Blur and Noise: Successful Detection with Few Photons

Quanta Vision in the Real World

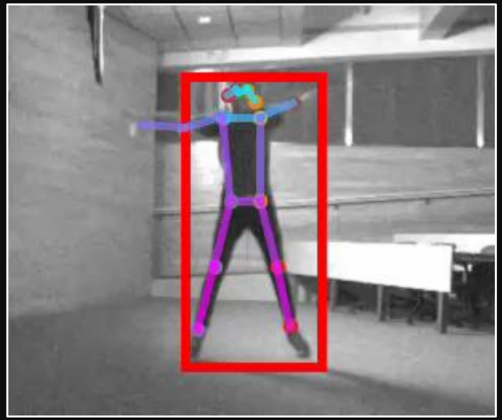


Recovering Semantics for a Variety of Tasks

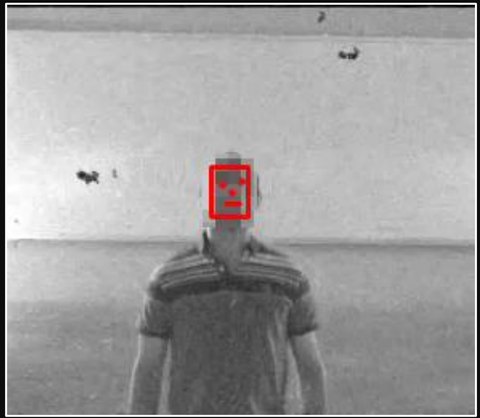
Background Subtraction [1]



Human Pose Estimation [2]



Face Detection [3]

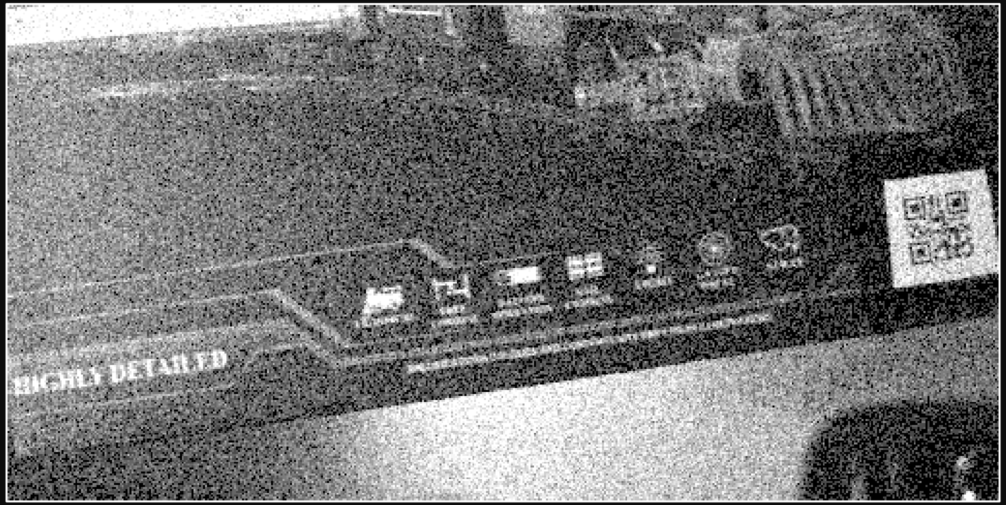


Action Recognition [4]

[1] Xiaowei Zhou, Can Yang, and Weichuan Yu, “Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 597–610, Mar. 2013.
[2] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, “RMPE: Regional Multi-person Pose Estimation,” in IEEE International Conference on Computer Vision (ICCV), Venice, Oct. 2017, pp. 2353–2362.
[3] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, Oct. 2016.
[4] J. Carreira and A. Zisserman, “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6299–6308.

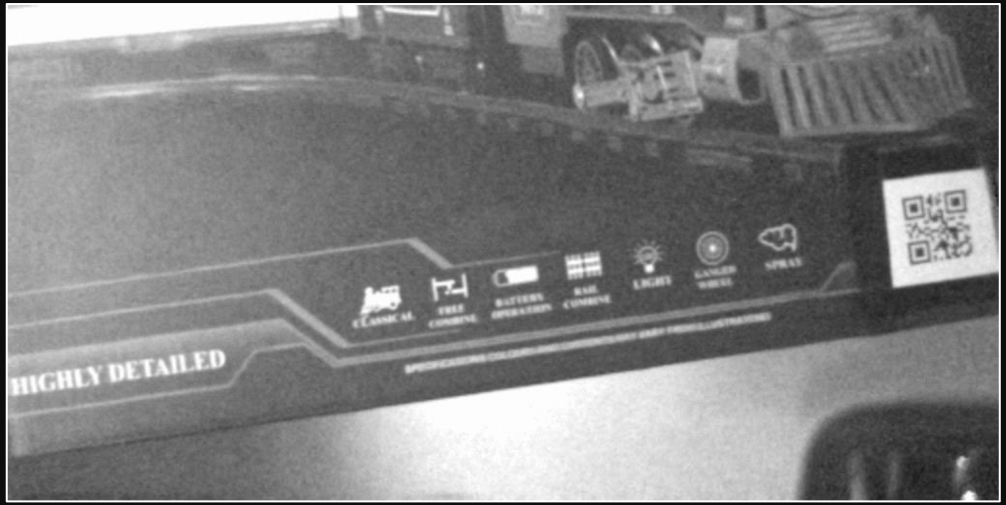
Recovering High-Frequency Spatial Details

Naive Average



Noisy & Blurry, Detection Fails

Burst Vision



Code Decoded Successfully

QR Decoding

WeChat QR code detector for detecting and parsing QR code, https://github.com/opencv/opencv_contrib/

Recovering High-Frequency Spatial Details

Naive Average

Noisy & Blurry, Detection Fails

Burst Vision

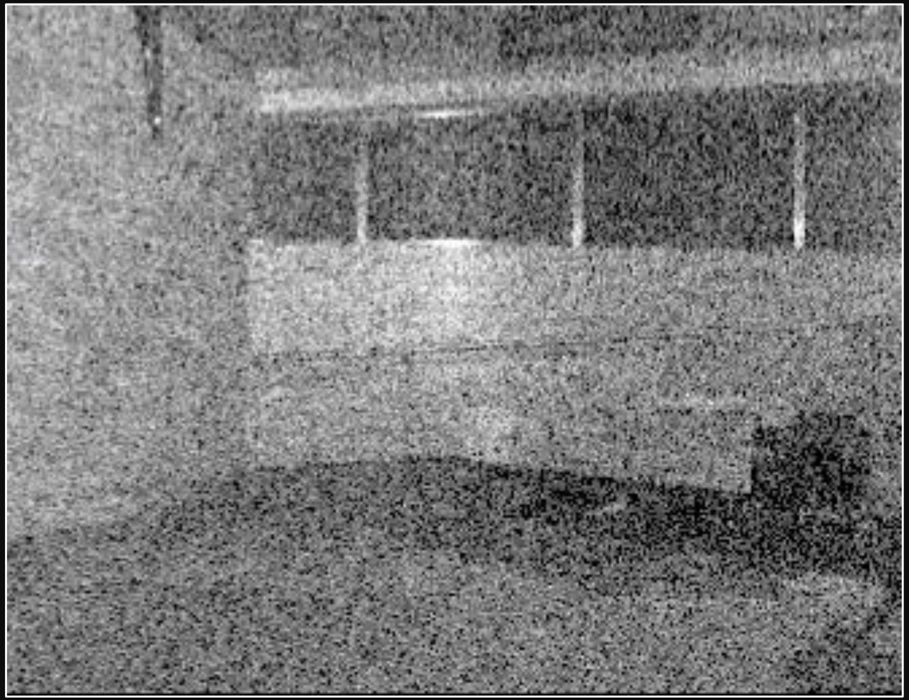
Text Recognized Successfully

Text Detection and Recognition

Y. Du et al., “PP-OCR: A Practical Ultra Lightweight OCR System,” arXiv:2009.09941 [cs], Oct. 2020, Accessed: Nov. 10, 2021. [Online]. Available: http://arxiv.org/abs/2009.09941

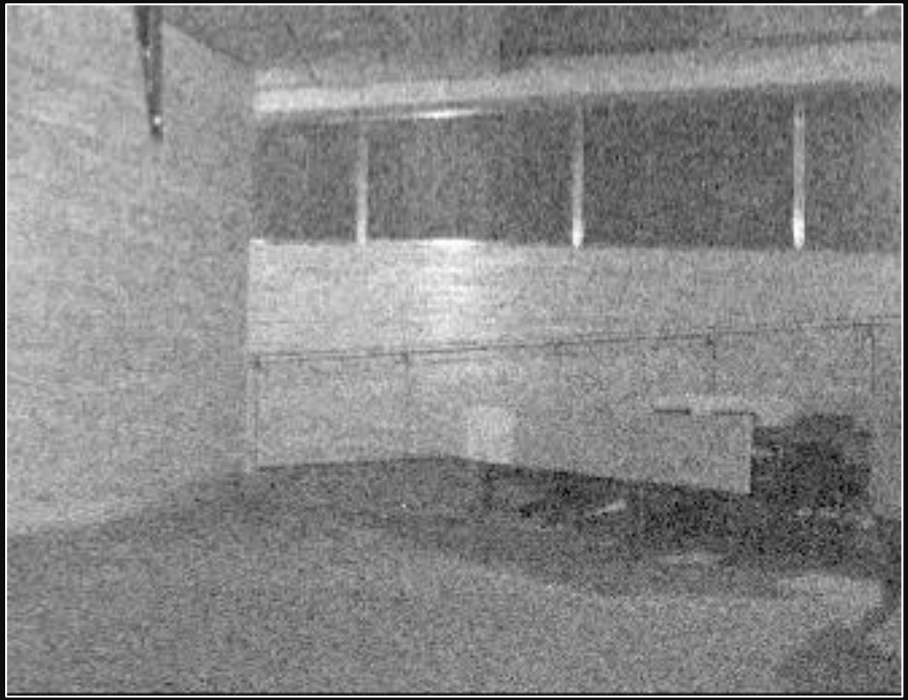
Existing SPAD Cameras: Naive Average

Naive Average (Short Exposure)



Noisy

Naive Average (Long Exposure)



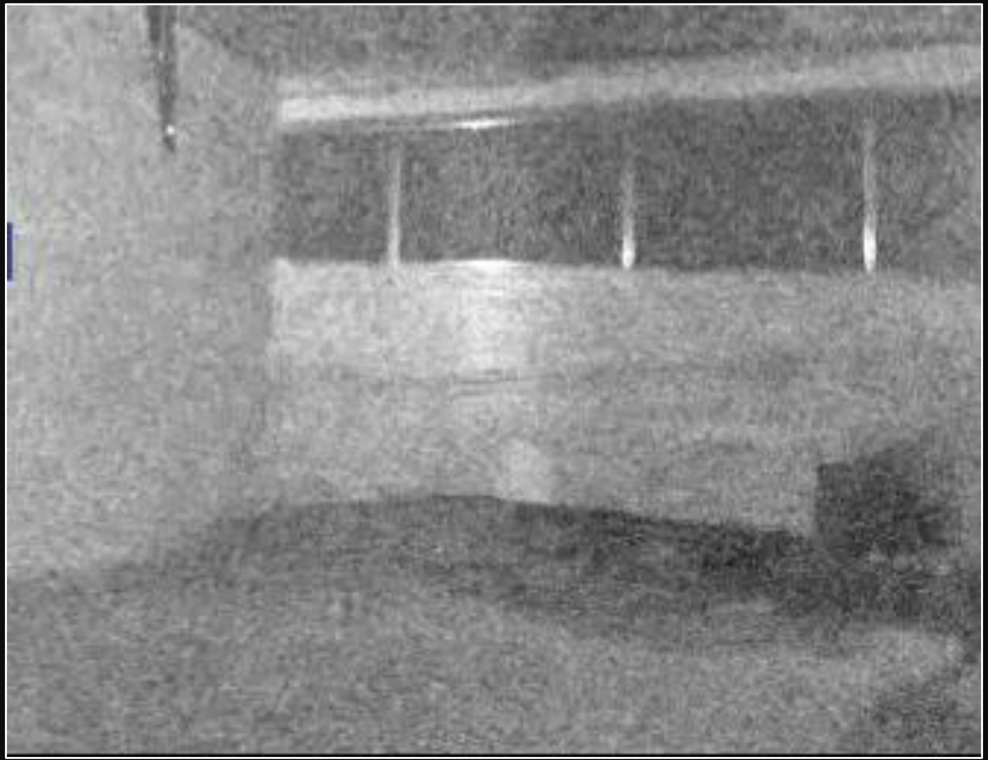
Blurred

Object Detection Fails

J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” *arXiv:1804.02767 [cs]*, Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767

Quanta Burst Vision

Burst Reconstruction [1]



Much higher image quality and detection

[1] S. Ma, S. Gupta, A. C. Ulku, C. Bruschini, E. Charbon, and M. Gupta, "Quanta burst photography," *ACM Trans. Graph.*, vol. 39, no. 4, Jul. 2020.

Comparison to Other Low-Light Sensors?

Night Vision Camera

Thermal Camera

Single-Photon Cameras

Tracking Failure

Tracking Maintained

Object Tracking

B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SiamRPN++: Evolution of Siamese Visual Tracking With Very Deep Networks," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 4277–4286.

Quanta Vision in the Wild

SPAD with Burst Vision

Successful Detection

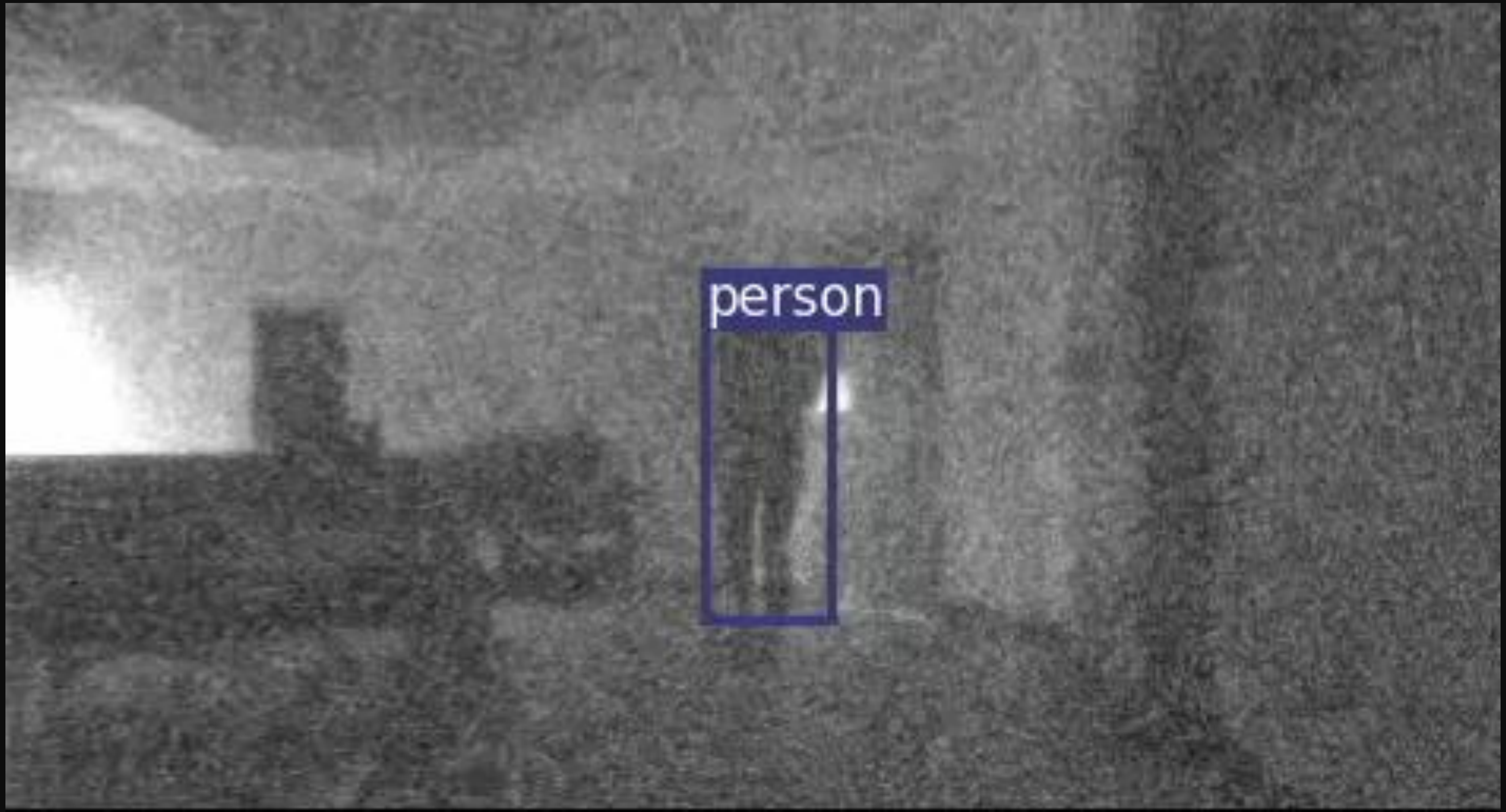
SPAD with Naive Averaging (Noise)

No Detection

Night Vision (Noise, Blur, Low Contrast)

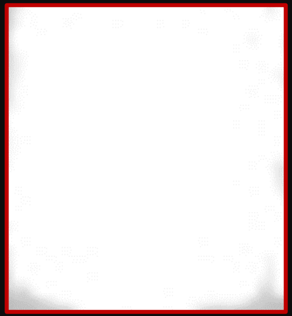
No Detection

Quanta Vision in Dark and Bright

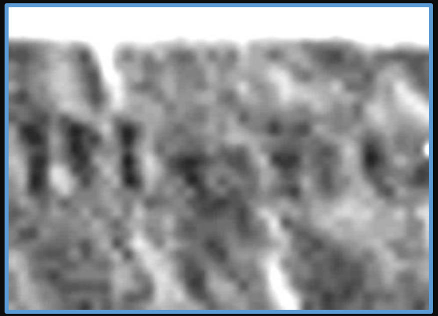


High Dynamic Range Photography

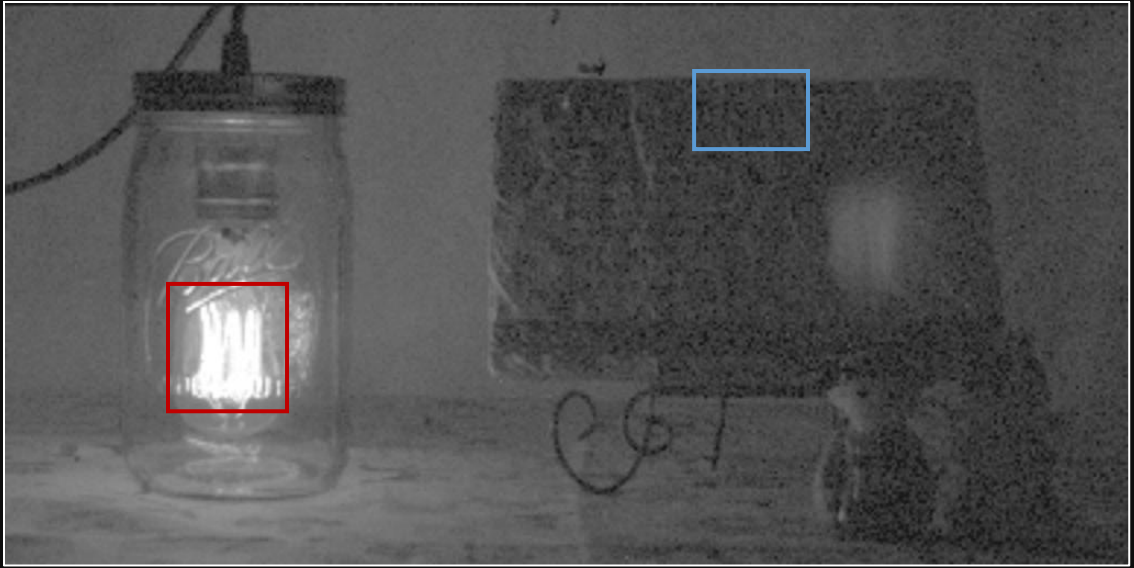
Conventional Camera (Long Exposure)

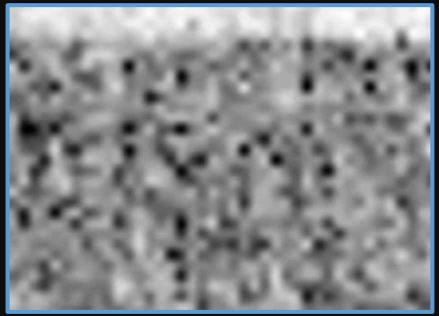


Saturation



Conventional Camera (Short Exposure)





Noisy, indiscernible

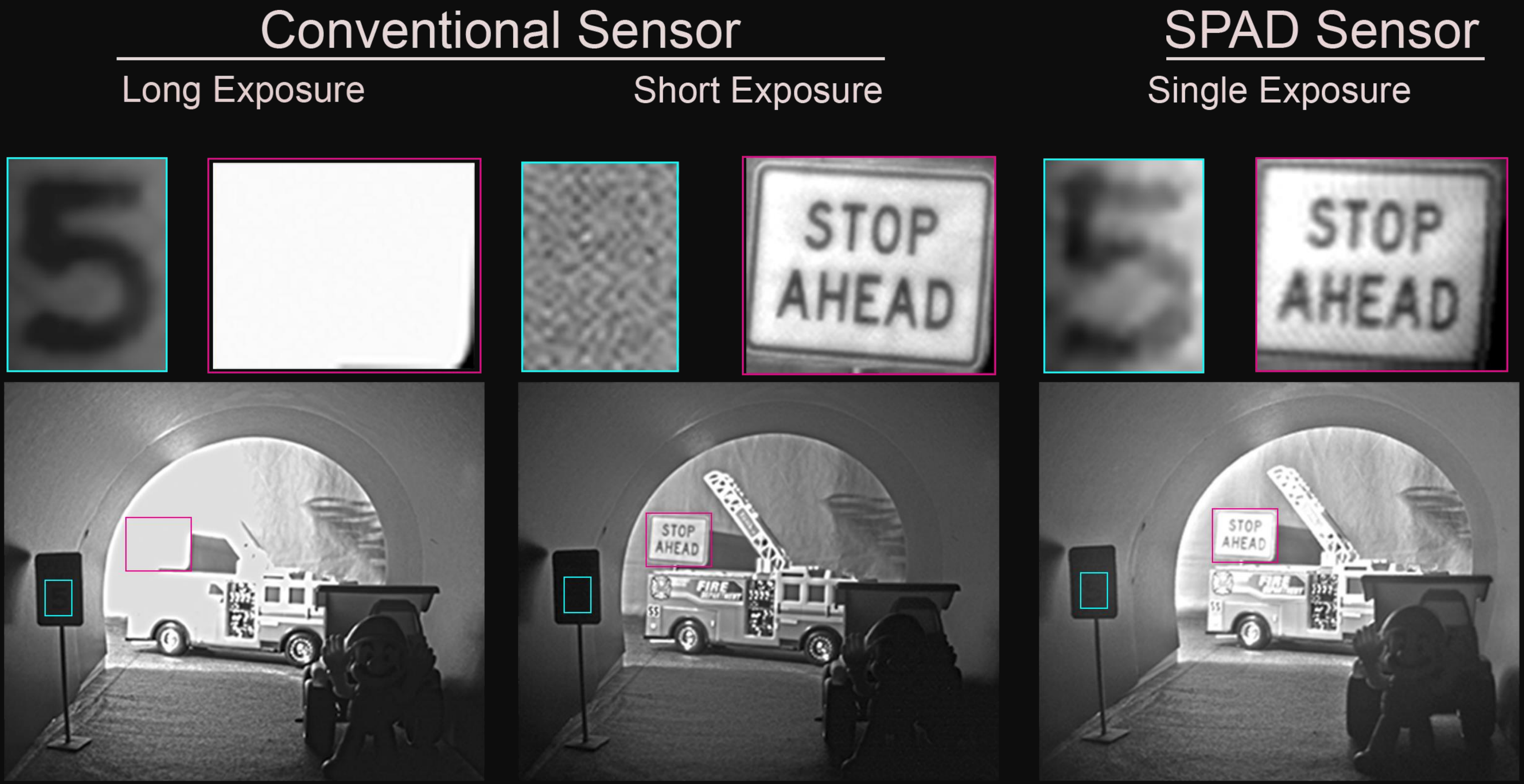
Simulated from captured SPAD binary images

High Dynamic Range Photography

Quanta Burst Photography

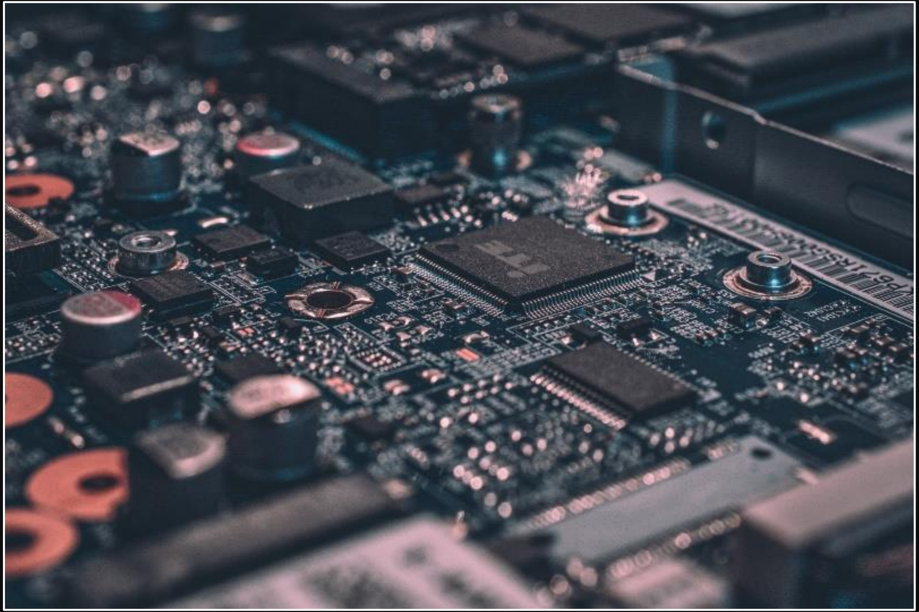
Recovers both bright and dark regions

Comparison with Conventional CMOS Sensor



Is Quanta Vision Ready for Prime Time?

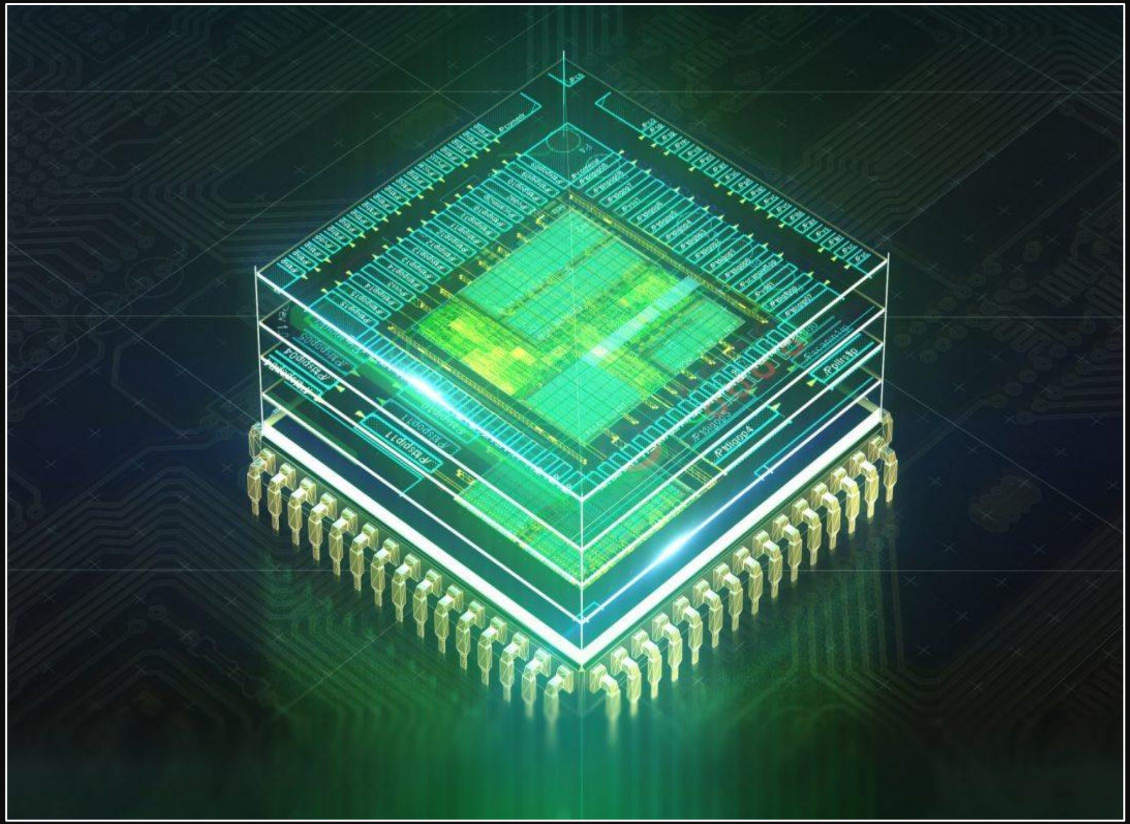
Photon Data Deluge: Bandwidth

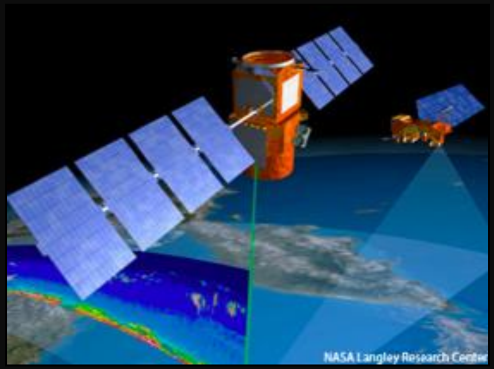


Efficient Computation

Is Quanta Vision Ready for Prime Time?

3D stacking, on-chip computation





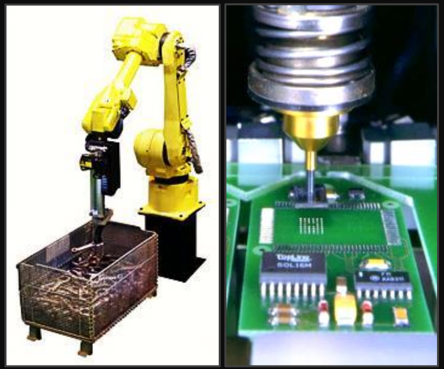
Long range

Low power

Single-Photon Cameras

High dynamic range

High depth resolution



Extreme sensitivity

http://wisionlab.cs.wisc.edu/project/quanta-burst-photography/