



OPEN

Sparse pixel image sensor

Lukas Mennel, Dmitry K. Polyushkin, Dohyun Kwak & Thomas Mueller

As conventional frame-based cameras suffer from high energy consumption and latency, several new types of image sensors have been devised, with some of them exploiting the sparsity of natural images in some transform domain. Instead of sampling the full image, those devices capture only the coefficients of the most relevant spatial frequencies. The number of samples can be even sparser if a signal only needs to be classified rather than being fully reconstructed. Based on the corresponding mathematical framework, we developed an image sensor that can be trained to classify optically projected images by reading out the few most relevant pixels. The device is based on a two-dimensional array of metal–semiconductor–metal photodetectors with individually tunable photoresponsivity values. We demonstrate its use for the classification of handwritten digits with an accuracy comparable to that achieved by readout of the full image, but with lower delay and energy consumption.

The Nyquist-Shannon sampling theorem¹ establishes a lower bound for the sampling rate required to capture any given signal of finite bandwidth without loss of information. The original signal can then be perfectly reconstructed from the samples. For many practical signals, however, reconstruction may still be possible from far fewer samples—or measurements—than required by the sampling theorem. This can be understood from the fact that these signals may contain much redundant information or, more precisely, are sparse when represented in a proper domain or basis. Natural images, for example, are known to be sparse in the Fourier or Wavelet domains, which is exploited in several types of transform coding schemes, including the JPEG and MPEG standards².

Compressed sensing (CS) is a mathematical framework for the recovery of sparse signals from few measurements^{3–6}. In CS, a signal that is incoherently (e.g. randomly) sampled at the encoder side, can be reconstructed at the decoder by finding the sparsest solution of an underdetermined linear system. Both sampling and compression are performed simultaneously to reduce the number of measurements at the expense of increased computational cost for signal reconstruction.

By combining CS with statistical learning, the number of required measurements can be further reduced, particularly if a given signal only needs to be assigned to one of a few categories, or classes, rather than being fully reconstructed. This can be achieved by using a task-specific basis, learned from data, instead of a generic one such as Fourier or Wavelet. In the sparse sensor placement optimization for classification (SSPOC) algorithm^{7,8}, the data are not sampled randomly, but a few representative measurement locations are identified from training data. Subsequent samples can then be classified with performance comparable to that obtained by processing the full signal.

Several new types of image sensors have been developed in recent years⁹, targeting lower energy consumption and latency than their conventional frame-based counterparts. Many of those devices emulate certain neurobiological functions of the retina, either using complementary metal–oxide–semiconductor (CMOS) technology (silicon retina)^{10–13} or emerging device concepts^{14–19}. CS has likewise led to new types of image acquisition systems, such as single-pixel cameras²⁰, coded aperture imagers²¹, and CMOS CS imaging arrays^{22,23}. SSPOC, on the other hand, has inspired applications in dynamics and control^{24,25}, but has to the best of our knowledge not been employed in an imaging device yet. Here, we present a hardware implementation of this algorithm, based on a two-dimensional array of tunable metal–semiconductor–metal (MSM) photodetectors. Each of these detectors can be addressed individually and their photoresponsivity values can be set by the application of a bias voltage. The device is fully reconfigurable and we demonstrate its use for the classification of handwritten digits from the MNIST dataset with an accuracy comparable to that achieved by readout of the full image, but with substantially lower delay and energy consumption.

Results

Operation principle. Let us first lay out the operation principle of the image sensor (Fig. 1a), exemplified by a simple linear classification problem. We restrict ourselves to binary classification, where an optical image, that is projected onto the chip, is assigned to one of two possible classes. The image is represented by a vector $\mathbf{p} = (P_1, P_2, \dots, P_n)^T$ in an n -dimensional vector space \mathbb{R}^n , where P_k is the optical power at the k -th pixel. Unlike

Institute of Photonics, Vienna University of Technology, Gußhausstraße 27-29, 1040 Vienna, Austria. email: thomas.mueller@tuwien.ac.at

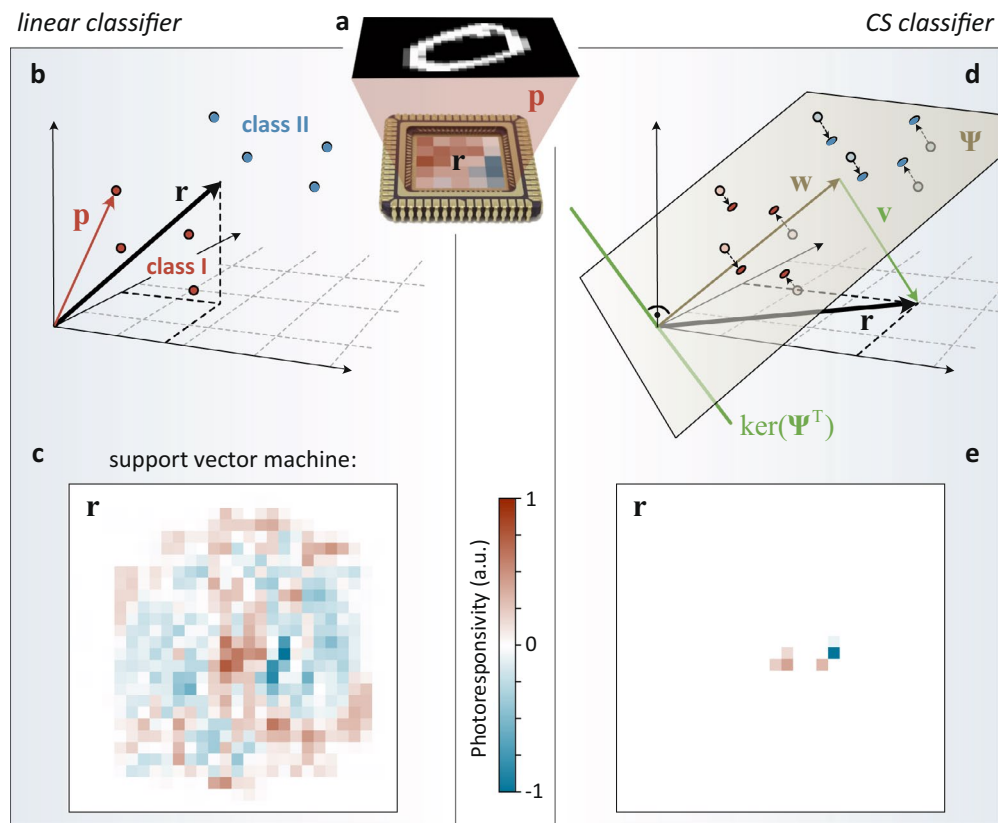


Figure 1. Theoretical background and operation principle. (a) Schematic illustration of the setup. An optical image \mathbf{p} is projected onto the face of the image sensor with photoresponsivity values \mathbf{r} that vary from pixel to pixel. (b) A binary linear classifier assigns an image to one of two possible classes I or II, depending whether or not the inner product $\mathbf{r}^T \mathbf{p}$ is larger than some threshold. In our implementation, the inner product is realized by summing up the photocurrents produced by all detector elements. (c) Photoresponsivities for a sensor that has been trained as a linear SVM for the classification of zeros and ones from the MNIST dataset. Almost all pixels exhibit non-zero photoresponsivity values. (d) Natural images have low-dimensional structure. This allows to construct a sparse photoresponsivity vector \mathbf{r} for classification. (e) Results for the same binary classification task as in c. Comparable performance is achieved with 99.2% of the detector elements having zero responsivity.

in conventional imagers, the photoresponsivity of each pixel is not fixed, but varies over the face of the chip. We aggregate the photoresponsivity values into a vector $\mathbf{r} = (R_1, R_2, \dots, R_n)^T \in \mathbb{R}^n$, where R_k denotes the responsivity of the k -th detector. A linear classifier is a predictor of the form²⁶

$$y = \sigma(\mathbf{r}^T \mathbf{p}), \quad (1)$$

where σ is a threshold function that maps all values of the inner product $\mathbf{r}^T \mathbf{p}$ below a certain threshold (bias) to the first class and all other values to the second class (Fig. 1b). Physically, the inner product is implemented by simply summing up the photocurrents produced by all n detector elements, $I_{\text{tot}} = \sum_{k=1}^n I_k = \sum_{k=1}^n R_k P_k = \mathbf{r}^T \mathbf{p}$. By thresholding I_{tot} , a binary output is obtained that is representative of the two classes. \mathbf{r} is learned from a set of labeled training data. A generalization to multi-class problems can be achieved by splitting pixels into subpixels^{14,27} which allows for a physical implementation of a responsivity matrix \mathbf{R} .

In Fig. 1c we plot \mathbf{r} for a linear support vector machine (SVM) that is trained to classify handwritten zeros (“0”) and ones (“1”) from the MNIST dataset. 90% of randomly picked images are used for training and the remaining 10% for assessment. Almost all photodetectors are active, with varying responsivity values, and a classification accuracy of 99.8% is reached.

We now aim to obtain a comparable performance by selecting a small, optimal subset of detectors, or pixels. Figure 1d provides a geometrical interpretation of the algorithm^{7,8}. A d -dimensional feature space, that spans the $d \ll n$ most significant variations among the training data, is calculated using principal component analysis (PCA)²⁶ and the principal component vectors \mathbf{u} are assembled in a matrix $\Psi = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_d)$. The choice of d is a tradeoff between the number of relevant pixels and accuracy for a specific classification task, as shown in Supplementary Figure S1. For categorical decisions, a measurement \mathbf{p} is projected into this low-dimensional subspace ($\Psi^T : \mathbb{R}^n \rightarrow \mathbb{R}^d$) and a linear classifier, described by the weight vector $\mathbf{w} \in \mathbb{R}^d$, is then applied therein: $y = \sigma(\mathbf{w}^T \Psi^T \mathbf{p})$. In image space coordinates, this expression resembles Eq. (1) with a photoresponsivity vector $\mathbf{r} = \Psi \mathbf{w}$. Note, however, that there exists an infinite number of solutions for \mathbf{r} , because adding any vector \mathbf{v} in

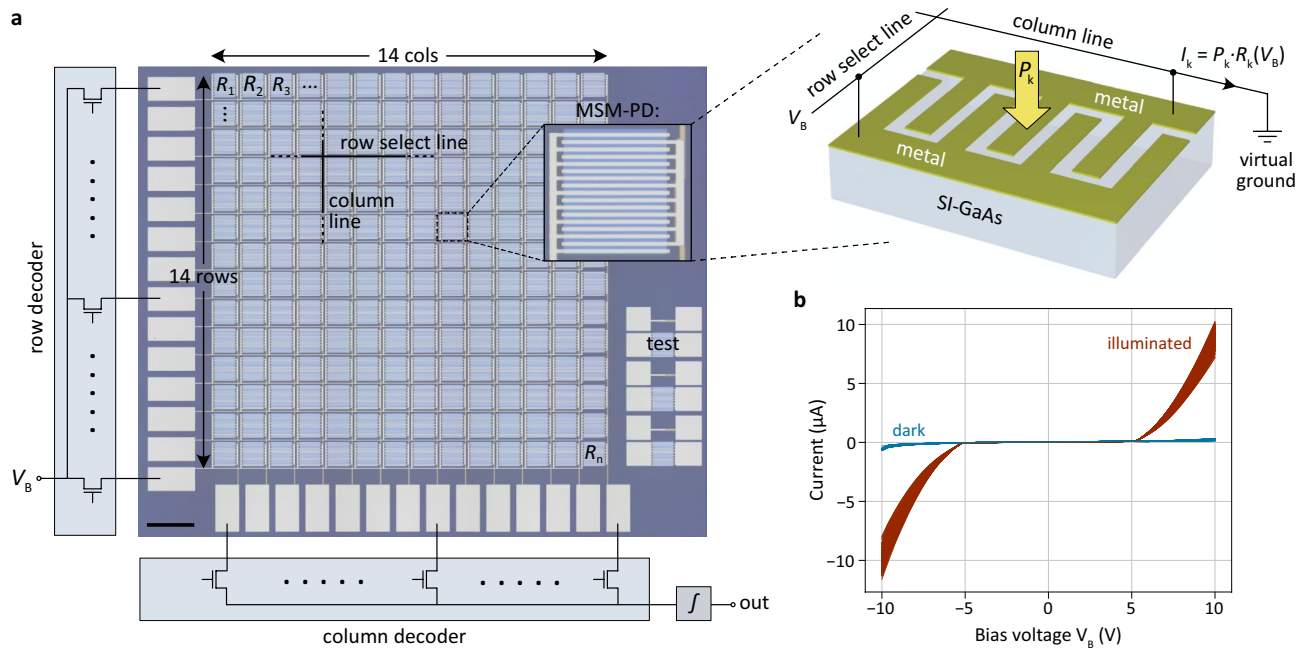


Figure 2. Image sensor architecture and characterization. **(a)** Microscope image of the sensor, with schematic illustrations of the external row/column decoders and integrating output (left). Scale bar, 200 μm . The chip size is 2.75 mm^2 . A detailed view of one of the MSM photodetectors is presented in the inset and a schematic illustration is in the picture to the right. Each of the detector elements is 90 \times 90 μm^2 in size. Details regarding the electrical measurement setup can be found in Supplementary Figure S5. **(b)** Bias voltage dependent device currents for all 196 detectors with (red lines) and without (blue lines) optical illumination ($\sim 160 \text{ W/m}^2$). The detectors are operated in the range $\pm 5 \text{ V}$ to $\pm 10 \text{ V}$.

the null space (kernel) of Ψ^T projects to the very same \mathbf{w} in feature space. We seek the sparsest solution for \mathbf{r} , that is the one that has at most d nonzero elements: $\|\mathbf{r}\|_0 \leq d$. As shown by the CS community^{3–6}, ℓ_1 -minimization leads to a convex optimization problem that can be efficiently solved with standard methods (here, orthogonal matching pursuit):

$$\min_{\mathbf{r}} \|\mathbf{r}\|_1, \text{ s.t. } \Psi^T \mathbf{r} = \mathbf{w}. \quad (2)$$

Figure 1e presents the results for the same binary MNIST classification task as before. Here, the data are projected into a six-dimensional PCA subspace ($d = 6$) in which a SVM is trained for classification. The photoreponsivity vector \mathbf{r} is calculated by ℓ_1 -minimization of (2) using the PySensors package²⁸ in Python (see source code in Supplementary Figure S9) and is plotted in Fig. 1e. The result is intuitive: Four of the six active pixels are located in the center of the image, where they spatially overlap with handwritten “1”s and thus produce a positive photocurrent due to their positive responsivity values. The remaining two pixels are located more to the right and overlap mostly with handwritten “0”s. Their responsivities are negative, so is their output current. The sign of the sum over all photocurrents hence allows to discriminate between the two digits. Geometrically, \mathbf{r} can be interpreted as the spatial locations of pixels that matter most.

Although less than 0.8% of the total pixels (6 out of 784) exhibit a responsivity $R \neq 0$, the classifier performs nearly as well as the SVM applied to the full image, and an accuracy of 99.1% is achieved. Energy consumption and delay, however, are substantially reduced, as both scale linearly with the number of detector elements being read out. We stress that it is not possible to obtain this result by merely thresholding \mathbf{r} in Fig. 1c, as can be seen from Supplementary Figure S2. Similar performance is obtained for the (more complex) Fashion-MNIST dataset, as demonstrated in Supplementary Figure S3.

Device implementation. In Fig. 2a we present the actual device implementation. The sensor is fabricated on a semi-insulating gallium arsenide (SI-GaAs) wafer, with two metal layers for routing of the electrical signals, using standard technology and without high temperature process steps. Details are provided in the Methods section. GaAs is preferred over silicon (Si) because of its shorter absorption and diffusion lengths, which both reduce cross-talk between neighboring pixels and allow for a relatively simple planar device structure. However, with some minor modifications, the sensor concept can be transferred to the Si platform, which also provides the opportunity for low-cost monolithic integration of the electronic driver circuits, that are currently implemented off-chip. Our sensor consists of a two-dimensional array of $n = 14 \times 14 = 196$ pixels, each containing an MSM photodetector²⁹ that converts incident light into photocurrent. Each detector comprises interdigitated metal fingers on the SI-GaAs semiconductor. Photoexcited electrons and holes drift under an electrical field applied

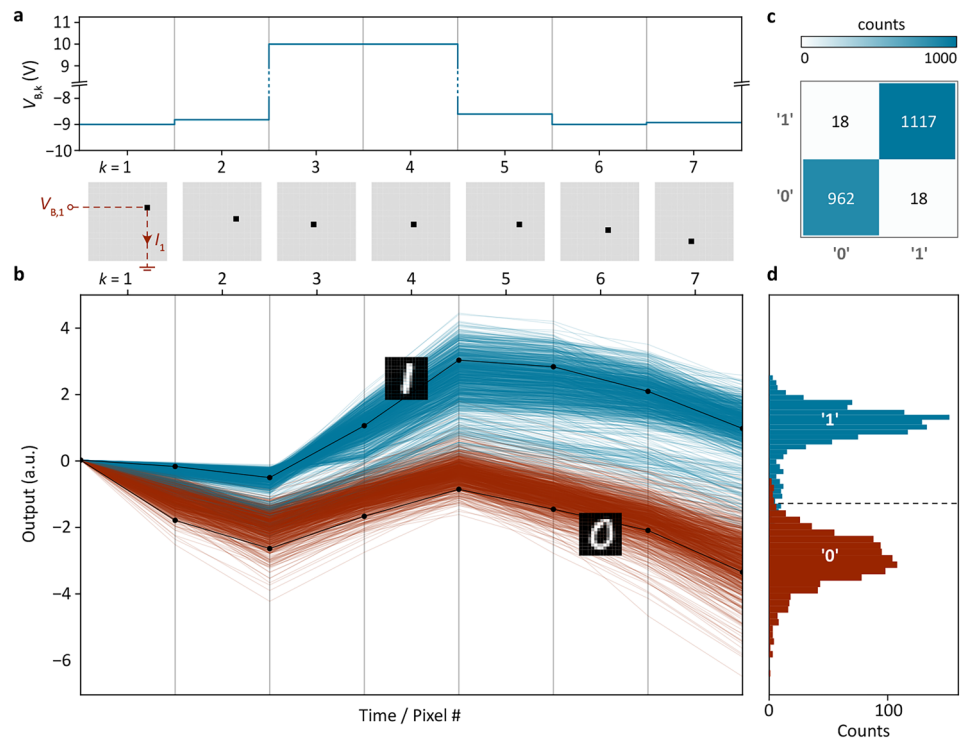


Figure 3. Image sensor operation and performance evaluation. **(a)** Relevant pixel locations k (bottom) and applied bias voltages $V_{B,k}$ (top) for the binary classification task discussed in the main text. **(b)** Temporal evolution of the sensor output for more than 2000 samples from the dataset. Red (blue) lines show cases in which a “0” (“1”) has been projected onto the sensor. The black lines show two representative examples with corresponding MNIST digits. **(c)** Experimental confusion matrix. A classification accuracy of 98.3% is achieved. **(d)** Histogram of sensor output as determined from the measurements in **b**. The dashed line indicates the decision threshold.

between the fingers, giving rise to an external current. The photoresponsivity of the device can be controlled by application of a bias voltage in the range ± 5 V to ± 10 V, as shown in Fig. 2b, where the negative sign of the responsivity indicates a reversed current flow direction. A less conservative design of the gap between the metal fingers (currently ~ 2 μm) could be considered to reduce the operation voltage. The low background carrier concentration of the SI-GaAs wafer ($\sim 8 \times 10^6$ cm^{-3}) ensures full depletion of majority carriers. As a result, the electric field drops homogeneously in the space-charge region between the metal fingers, so that photogenerated carriers are efficiently swept out of the device. Low residual doping is also required to suppress dark current and reduce cross-talk between neighboring detectors. Under 10 V bias, the photoresponsivity reaches values as high as ~ 5 A/W, exceeding 100% quantum efficiency (0.52 A/W at 650 nm wavelength). Such photoconductive gain is often observed in MSM photodetectors³⁰ and can be attributed to Schottky barrier lowering due to trapping of photoexcited holes in localized surface or bulk trap states. The sensor is hence well suited for applications that require high sensitivity. Finally, we verified an approximately linear illumination intensity-dependence of the photocurrent (Supplementary Figure S3), as required by Eq. (1).

As in CMOS sensor technology, detectors are addressed by row and column decoders. The readout is performed one pixel at a time, with relevant pixel locations k and corresponding photoresponsivity values R_k being determined from Eq. (2). Pixel-to-pixel variations and the nonlinear R_k -versus- $V_{B,k}$ behavior in Fig. 2b are accounted for as discussed in Supplementary Figure S4. For details regarding the optical apparatus used for image projection, we refer to the Methods section.

We evaluated the sensor performance by the same binary classification task as discussed above. During a measurement, a bias voltage $V_{B,k}$ is applied to the k -th pixel via the row select line and the generated photocurrent is read out via the respective column line (Fig. 3a). The resulting photocurrent is integrated over a period of time (\sim ms), before the next relevant detector is addressed by its row and column, and its output is added. We conducted this measurement for more than 2000 images of zeros and ones from the dataset. The bundle of curves, displayed in Fig. 3b, shows the temporal evolution of the output for each of those samples. Traces in red show cases where a zero has been projected onto the chip; traces in blue correspond to a one. Two representative examples with corresponding images are shown as black lines. With each additional pixel measured, the red and blue traces separate further and the classification accuracy improves. At the end of a cycle (here, after 7 pixels), the output signal is compared to a threshold value and assigned to one of the two classes. Then the next cycle commences. Figure 3d shows a histogram of the sensor outputs after each cycle/sample, as determined from the measurements in Fig. 3b. From the experimental confusion matrix, presented in Fig. 3c, we determine

a classification accuracy of 98.3%. The small deviation from the theoretical expectation (99.7%) and the shift of the decision threshold to below zero are attributed to device imperfections.

Discussion

In summary, we presented a sensor that can be trained to classify images with an accuracy comparable to that of frame-based cameras by reading out the few most relevant pixels. The use of MSM detectors with tunable photoresponsivity values results in a particularly lean and simplistic sensor design. In contrast, conventional imaging devices, such as CMOS cameras, employ photodiodes with fixed responsivities, determined by the doping profile in the semiconductor. An implementation of the here described algorithm can then be achieved by implementation of the inner product $\mathbf{r}^T \mathbf{p}$ with additional electronics that is placed on the chip (e.g. a tunable amplifier). We propose that such a system could be operated in a low-power mode, running the algorithm outlined above, and once a certain scene or gesture is detected, the system switches into a full-frame mode for further analysis.

An extension to multi-class problems can in the simplest case be implemented by treating each class separately which, however, results in a $c - 1$ fold increase of required pixel locations, where c denotes the number of categories. In Ref. 7 it has thus been suggested to introduce a regularization term in Eq. (2) that penalizes the total number of measurements or – in case of an image sensor – pixels. Results for a three-class problem with regularization are presented in Supplementary Figure S8.

Methods

Sensor fabrication. As a substrate we used a semi-insulating gallium arsenide (SI-GaAs) wafer, covered by 20 nm atomic layer deposition (ALD) grown Al_2O_3 . A first metal layer was fabricated by evaporating Ti/Au (3/25 nm) through a mask created by electron-beam lithography (EBL). A 30-nm-thick Al_2O_3 oxide was then deposited using ALD, followed by a second lithography step and wet chemical etching in potassium hydroxide to define via-holes that connect the bottom and top metal layers and top metal with the GaAs substrate where necessary. Lastly, a top metal layer was added by another EBL process and Ti/Au (5/80 nm) evaporation. Prior to the metal evaporation, we removed the GaAs native oxide by a short dip of the sample in concentrated hydrochloric acid. We confirmed the continuity and solidity of the electrode structure by optical microscopy and electrical measurements in a wafer probe station. The sample was finally mounted in a chip carrier and wire-bonded.

Optical setup. A collimated, linearly polarized light beam (650 nm wavelength), produced by a semiconductor laser diode, illuminates a spatial light modulator (SLM, Hamamatsu), operated in intensity-modulation mode. On the SLM, the MNIST digits are displayed and the polarization of the light is rotated according to the pixel value. A polarizer with its optical axis oriented perpendicular to the polarization direction of the incident light acts as analyzer. The generated optical image is then projected onto the image sensor with a lens. A schematic illustration of the apparatus is provided in Supplementary Figure S6.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Received: 12 December 2021; Accepted: 24 March 2022

Published online: 05 April 2022

References

- Shannon, C. E. Communication in the presence of noise. *Proc. IRE* **37**, 10–21 (1949).
- Bhaskaran, V. & Constantinides, K. *Image and Video Compression Standards* (Springer, Berlin, 1997).
- Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
- Romberg, J. Imaging via compressive sampling. *IEEE Signal Process. Mag.* **25**, 14–20 (2008).
- Candes, E. J. & Wakin, M. B. An Introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
- Baraniuk, R. A lecture on compressive sensing. *IEEE Signal Process. Mag.* **24**, 1–9 (2007).
- Brunton, B. W., Brunton, S. L., Proctor, J. L. & Kutz, J. N. Sparse sensor placement optimization for classification. *SIAM J. Appl. Math.* **76**, 2099–2122 (2016).
- Manohar, K., Brunton, B. W., Kutz, J. N. & Brunton, S. L. Data-driven sparse sensor placement for reconstruction. *IEEE Control Syst. Mag.* **38**, 63–86 (2018).
- Liao, F., Zhou, F. & Chai, Y. Neuromorphic vision sensors: principle, progress and perspectives. *J. Semicond.* **42**, 013105 (2021).
- Galleo, G. *et al.* Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 154–180 (2020).
- Leñero-Bardallo, J. A., Serrano-Gotarredona, T. & Linares-Barranco, B. A 3.6 μs latency asynchronous frame-free event-driven dynamic-vision-sensor. *IEEE J. Solid-State Circuits* **46**, 1443–1455 (2011).
- Posch, C., Matolin, D. & Wohlgenannt, R. A. QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. *IEEE J. Solid-State Circuits* **46**, 259–275 (2011).
- Brandli, C., Berner, R., Yang, M., Liu, S.-C. & Delbruck, T. A 240×180 130 dB 3 μs latency global shutter spatiotemporal vision sensor. *IEEE J. Solid-State Circuits* **49**, 2333–2341 (2014).
- Mennel, L. *et al.* Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).
- Zhou, F. *et al.* Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol.* **14**, 776–782 (2019).
- Seo, S. *et al.* Artificial optic-neural synapse for colored and color-mixed pattern recognition. *Nat. Commun.* **19**, 5106 (2018).
- Wang, C.-Y. *et al.* Gate-tunable van der Waals heterostructure for reconfigurable neural network vision sensor. *Sci. Adv.* **6**, eaba6173 (2020).
- Jang, H. *et al.* An atomically thin optoelectronic machine vision processor. *Adv. Mater.* **32**, 2002431 (2020).
- Wang, Y. *et al.* An in-memory computing architecture based on two-dimensional semiconductors for multiply-accumulate operations. *Nat. Commun.* **12**, 3347 (2021).
- Duarte, M. F. *et al.* Single-pixel imaging via compressive sampling. *IEEE Signal Process. Mag.* **25**, 83–91 (2008).

21. Arce, G. R., Brady, D. J., Carin, L., Arguello, H. & Kittle, D. S. Compressive coded aperture spectral imaging. *IEEE Signal Process. Mag.* **31**, 105–115 (2014).
22. Robucci, R., Gray, J. D., Chiu, L. K., Romberg, J. & Hasler, P. Compressive sensing on a CMOS separable-transform image sensor. *Proc. IEEE* **98**, 1089–1101 (2010).
23. Oike, Y. & El Gamal, A. CMOS image sensor with per-column $\Sigma\Delta$ ADC and programmable compressed sensing. *IEEE J. Solid-State Circuits* **48**, 318–328 (2013).
24. Kaiser, E. *et al.* Sparsity enabled cluster reduced-order models for control. *J. Comput. Phys.* **352**, 388–409 (2018).
25. Mohren, T. L., Daniel, T. L., Brunton, S. L. & Brunton, B. W. Neural-inspired sensors enable sparse, efficient classification of spatiotemporal data. *Proc. Natl. Acad. Sci. USA* **115**, 10564–10569 (2018).
26. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, Berlin, 2017).
27. Mennel, L., Molina-Mendoza, A.J., Paur, M., Polyushkin, D.K., Kwak, D., Giparakis, M., Beiser, M., Andrews, A.M. & Mueller, T. A photosensor employing data-driven binning for ultrafast image recognition. [arXiv:2111.10612](https://arxiv.org/abs/2111.10612).
28. de Silva, B.M., Manohar, K., Clark, E., Brunton, G., Brunton, S. & Kutz, J. PySensors: a Python package for sparse sensor placement. [arXiv:2102.13476](https://arxiv.org/abs/2102.13476) (2021).
29. Ito, M. & Wada, O. Low dark current GaAs metal-semiconductor-metal (MSM) photodiodes using WSi contacts. *IEEE J. Quantum Electron.* **22**, 1073–1077 (1986).
30. Klingenstein, M. *et al.* Photocurrent gain mechanisms in metal-semiconductor-metal photodetectors. *Solid State Electron.* **37**, 333–340 (1994).

Acknowledgements

We thank Max Andrews for providing a GaAs wafer. We acknowledge financial support from the Austrian Science Fund FWF (START Y 539-N16).

Author contributions

T.M. conceived the device concept. D.K.P., L.M. and D.K. fabricated the sensor. L.M. and T.M. programmed the machine learning algorithms. L.M. designed and built the experimental setup, carried out the measurements and analyzed the data. T.M. and L.M. prepared the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09594-y>.

Correspondence and requests for materials should be addressed to T.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022