Challenges and Perspectives in Neuromorphic-based Visual IoT Systems and Networks

Maria Martini and Nabeel Khan
WMN Research Group
Kingston University London
Kingston upon Thames (London), UK

Yin Bi and Yiannis Andreopoulos
University College London
Electronic and Electrical Engineering Dept.
London, UK

Hadi Saki and M. Shikh-Bahaei
King's College London
CTR.
London, UK

***Abstract*—Neuromorphic sensors, a.k.a. dynamic vision sensors (DVS) or silicon retinas, do not capture full images (frames) at a fixed rate, but asynchronously capture spikes indicating changes of brightness in the scene, following the principles of biological vision and perception in mammals. DVS sensing and processing produces a data representation where the scene can be represented with a very high time resolution with a limited number of bits (an inherent data compression is performed at the time of acquisition). Such representation can be used locally to derive actionable responses and selected parts can be transmitted and then processed in another network location. Due to these features, such sensors represent an excellent choice as visual sensing technology for next-generation Internet-of-Things, e.g. in surveillance, drone technology, and robotics. It is in fact becoming evident that in this framework acquiring, processing, and transmitting frame-based video is inefficient in terms of energy consumption and reaction times, in particular in some scenarios. Hence, we explore here the feasibility of advanced Machine to Machine (M2M) communications systems that directly capture, compress and transmit spike-based visual information to cloud computing services in order to produce content classification or retrieval results with extremely low power and low latency.**

I. Introduction

In the next ten years, most of the envisaged services for the analysis of sensor data for event, action, object or person recognition/classification, and context awareness will utilize advanced communications and networking solutions for the transport of sensor data (or aggregates of such data) to cloud storage and computing servers that perform the analysis, detection or reconstruction tasks. Due to the utilization of the internet protocol stack for this process, see for instance Message Queuing Telemetry Transport (MQTT) application-layer messaging [1] and 6LoWPAN [2] for the network layer, the wider context of such applications has been termed as the Internet of Things (IoT). A key enabler of future IoT deployments is the achievement of the equivalent to high frame-rate visual sensing and processing for surveillance and monitoring (e.g., vehicles, drones) at very low power. Conventional high-speed visual sensing results in very high bandwidth and processing requirements, hence new paradigms inspired by biological vision are now emerging that challenge the classical notion of video frames. The human vision system detects details of reflectance and movement in scenes (predominantly via cone and rod photoreceptor cells in mammals) in an asynchronous manner (and at very high speed), while the visual cortex is filling in the remaining information. Inspired by this behaviour, hardware designs of neuromorphic sensors, a.k.a. dynamic vision sensors (DVS), have been proposed recently. DVS devices output data as coordinates and timestamps of reflectance events, triggering on or off in an asynchronous manner, i.e., when the logarithm of the intensity value at a pixel of a planar CMOS sensor changes beyond a threshold.

Figure 1(a) shows the acquisition of data via both a conventional frame sensor and a DVS. Unlike a conventional frame based sensor, the DVS camera captures the on/off triggering of the fast motion events in the scene. Remarkably, DVS achieves this with an order of magnitude reduction in power consumption (10-20 mW of power consumption instead of hundreds of mW) and two orders of magnitude increase in speed (e.g., when the events are rendered as video frames, 700-2000 frames per second can be achieved). Unlike compressive sensing cameras, DVS cameras are commercially available, (see e.g., the iniVation DVS240, DVS346 and the ProPheSee Onboard device). Exciting new applications have already begun to emerge for implants for visual stimulation of the visually impaired [3], robotic reflexes with superhuman capabilities, visual surveillance capable of detecting very high-speed events, and specific methods for DVS data analysis have been developed, including feature detection and tracking [4].

To accommodate these applications, a scalable and hierarchical representation for multipurpose usage of DVS data, rather than a fixed representation suitable for an individual application (such as motion analysis or object detection) would be required. Instead of constraining applications to on-board processing, in our Internet-of-Silicon Retinas (IoSiRe) framework illustrated in Figure 1(b), we propose layered data representations and adaptive M2M transmission frameworks for DVS data representations, which are mapped to each application's quality metrics, response times, and energy consumption limits, and will enable a wide range of services by selectively offloading the data to the cloud. In this paper, we summarize our results and perspectives in these domains and present indicative application results for spike-based sensing in visual IoT applications.

II. Current challenges

1) Conventional frame-based video and IoT applications: Current Internet of Things (IoT) systems for visual data anal-

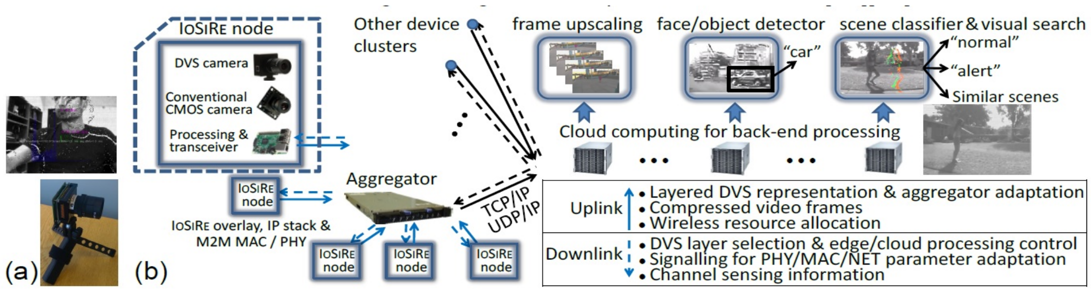

Fig. 1. (a) DVS camera acquisition. (b) The IoSiRe framework bundling multimodal sensors and transceivers into an IP-based framework for sensor-to-aggregator-to-cloud processing.

ysis extend the current video delivery paradigms to wireless networks amenable to IoT deployments and/or apply visual data analytics to low-end conventional cameras (see e.g. [5]). Even though there is potential in such approaches, the severe bandwidth and energy consumption requirements imposed by the needed spatio-temporal resolution become insurmountable even in low frame-rate capturing transmission [6].

2) Low-energy visual data capturing and processing: The recent paradigms of compressive sensing (CS), distributed compressive sensing and distributed video coding (DVC), as well as energy- and information-optimal data acquisition and transmission protocols have been proposed as a solution to the bandwidth and energy burden imposed by conventional frame-based sampling and processing. However, such approaches present two main issues: Despite more than a decade since the introduction of the first CS hardware and recent prototypes [7], no commercially-viable CS cameras have been achieved to date; In addition, data reconstruction requires the solution of a dynamic programming problem (or the execution of an estimation process for DVC), which is computationally intensive, especially when high frame rate is pursued.

3) Dynamic vision sensing and transmission systems: DVS applications have begun to emerge in the last few years, mostly for motion analysis and optical flow estimation but also, to a limited extent, for on-board processing with a coupled neural network for neuromorphic system development [8]. Relevant research works focus on biological aspects of vision and on using the acquired data for computer vision applications, rather than on sensing and transmission of events within IoT-oriented systems. At the moment, encoding and transmission of DVS data to back-end servers for analysis and processing remains an open problem. In particular, in order to design appropriate transmission systems, it is important to be able to estimate the traffic produced by such sensors, as well as to represent the data to be transmitted in the most compact way possible using limited resources.

4) Machine-to-machine communications for visual data gathering: Efficient data aggregation through uplink M2M networks has recently attracted interest; optimisation of uplink transmit power for energy-efficient Machine-to-Data aggregator communication has been investigated under multi-node interference [9]. Physical layer parameters and protocols are designed for hierarchical uplink M2M aggregation networks with sequential and parallel data communication. These PHY-layer methods are oblivious of the higher layers. Also MAC channel access schemes and Network-layer routing protocols, suited for intermittent low-data-rate communication between machines and data collectors, have been considered in the literature [10] [9]. However, these limited works on uplink M2M communication schemes and protocols are not designed or optimised for DVS-based applications.

III. Recent Results and Perspectives

A. Generated data traffic and data compression

Neuromorphic vision sensors asynchronously transmit pixel-level relative intensity (brightness) changes, with micro-second time resolution, using the Address Event Representation (AER) protocol for exchanging data. Following the protocol, each event is represented by a 4-tuple (x, y, t, p), where x and y are the coordinates of the pixel which has undergone a brightness change, t is the timestamp, and p is the polarity of the event. The polarity of the event represents either an increase (positive sign) or a decrease (negative sign) of the gray-level intensity change. Events are triggered whenever there is either motion of the neuromorphic vision sensor or motion / change of illumination in the scene or both. In other words, no data is transmitted for stationary vision sensors and static scenes. These unique properties enable neuromorphic vision sensors to achieve low-bandwidth, low-latency, and low-power requirements.

Figure 2 shows, for the *Diving* scene in dataset [11], the neuromorphic spike events rendered as frames, in the first row, and the corresponding conventional pixel based frames, in the second row. The green color represents events with positive change in brightness, whereas events with negative change in brightness are represented by the red color.

In order to compare the bandwidth requirements for neuromorphic and conventional vision sensors, let us consider the data rate of the *Diving* sequence shown in Figure 2. The eight-second long sequence, with spatial resolution 320×240, results in an average neuromorphic event rate of 28.67 Keps (kilo-events per second). According to the literature, eight bytes are needed to represent the data associated to an event (see e.g. [12]) With such assumption, the data rate of the diving sequence is approximately 1.83 Mbps. On the other hand, the data rate required to transmit raw video (grayscale only), assuming 30 fps, is $320 \times 240 \times 30 \times 8 = 18.43$ Mbps. The rate of DVS data is approximately 10 times lower, with a much higher temporal resolution.

In order to design appropriate transmission strategies, it is important to estimate the expected data rate output by neuromorphic visual sensors. In [13] [14] we proposed a model for the estimation of such data rate by considering the scene content and the motion speed of the visual sensor. The model was developed and tested for scenarios with a moving vision sensor. According to the study, the neuromorphic event rate varies exponentially with the scene complexity, with the Sobel and Prewitt based mean gradient approximations resulting as the best estimators for the complexity of the video sequence. In addition, we studied the relationship between the sensor speed and the event rate, that we found to be linear. Based on this analysis, we proposed a two parameter exponential model for the dependency of the event rate on scene complexity and sensor speed. According to the results, the model has a bit-rate prediction accuracy of approximately 88.4 % for the *Outdoor* dataset and an overall prediction accuracy of approximately 84 %. Such model will enable the selection and design of the appropriate transmission strategy and will also facilitate the performance evaluation of neuromorphic-based IoT systems via simulation.

One of the key challenges for resource constrained networks is managing high data rates, hence the need for compression. As highlighted before, dynamic vision sensors already provide inherent data compression in the acquisition phase. However, further compression is required in many IoT scenarios to reduce the bandwidth requirements. A natural solution is to compress the data at the sensing device. The authors in [12] exploit the unique characteristics of event data and propose a lossless compression strategy where each neuromorphic event is represented after compression by a variable number of bits (lower than the eight bytes needed before compression), depending on the considered dataset . With the scenes considered by the authors (although not all of them are realistic) each event is represented in average with 3.3 bits (see Table 1 in [12]), with a compression ratio of 19.5. However, with the dataset on intelligent driving [11], the same authors report in [15] a compression ratio of approximately 2.65. The authors in [16] compared several lossless compression strategies for static DVS scenarios. These strategies include dictionary based compression, IoT specific compression, Integer based approaches and DVS specific compression algorithms. According to the detailed experimental analysis, the LZMA strategy achieves the best compression ratio among all the considered strategies. On the other hand, the Brotli compression method achieves the best trade-off between (compression and decompression) speed and compression ratio.

While the aforementioned studies provide an excellent starting point for the design of tailored compression strategies, it is important that future approaches for the compression of data acquired by neuromorphic sensors address the complexity and delay limitations of the current approaches, besides aiming at reducing further the data rate. Our current work is addressing both these directions.

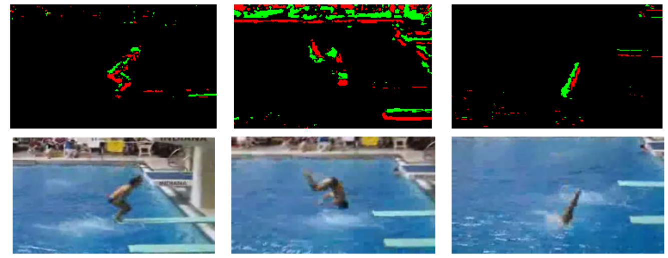

Fig. 2. Rendered frames from DVS data (above) and video frames acquired via camera (below).

B. Low latency data transmission

Future wireless communication systems are required to be able to comply with emerging data traffic with diversified functionality characteristics, in order to promote use cases like massive machine-type communications (MTC) and ultra-reliable and low-latency communications (URLLC). Unlike throughput-oriented legacy wireless networks, the new network needs to support a range of applications, such as industrial spiking dynamic vision sensors (DVS) and virtual reality with ultra-reliable and low-latency demands. Specifically, in 5G MTC URLLC use cases, the end-to-end (E2E) latency is required to be less than 1 milliseconds. The E2E network latency is affected by several network components. Therefore, several techniques are required to be applied and involved at a different levels of the network to enable the low latency target. Significant latency improvement can be achieved at the processing level by employing proactive computing, and coded computing where redundant on-demand computing is avoided and dependency of task processing is eliminated respectively. At the communication level, minimizing the spatial distance between the edge and the application is a key low latency enabler. This concept is inspired by the idea that reducing the transmitter and the receiver distance will be effective in improving the network capacity. Network densification, the principle of dense small cell deployment, relay heads (RH), remote radio units (RRU), mobility assisted multi-edge computing (MEC), Computing location swapping are some of the attractive proximity-based computing methods.

High capacity link is another communication layer low latency enabler that plays significant roles in decreasing offloading latency from application to servers by offering large bandwidth. Motivated by the sub 6 GHz spectrum shortage, the spectrum from 30 to 300 GHz, or the mmWave band, has been acquiring growing attention, to the level that it is currently regarded as the most significant enabler to offer the 10 Gbps data rates expected for the 5G networks. However, at these frequencies signal propagation is severely affected and naturally distinctive from that at the sub 6 GHz band frequencies. These differences include higher penetration losses, higher path loss for equivalent antenna gains and higher transmit power required to maintain a similar signal-to-noise ratio (SNR) at lower bands unless enhanced signal processing techniques that feature massive input massive output (MIMO) and beamforming (BF) are deployed effectively.

To this end, machine learning methods can be implemented for latency-oriented joint antenna selection and beamforming algorithms. In these solutions the learning-based antenna assignment and beamforming power control policy can be adaptive to the users' energy state information (ESI), the users' queue state information (QSI) and/or the uplink/downlink channel state information (CSI). The work in [17] developed an analytical framework for the latency-optimal control problem based on the theory of infinite horizon partially observable Markov decision process (POMDP). In this POMDP method, the proposed optimal protocols optimize the transposition delay of the user while limiting the energy consumption by users and the aggregator. Furthermore, this work presents a multi-objective optimization framework to study the trade-off between uplink transmit power minimization, downlink power minimization, and latency minimization. To reduce the complexity, the infinite-horizon POMDP problem is transformed into an equivalent value Bellman program and solved by the near-optimal point-based heuristic search value iteration (PB-HSVI) method under specific standard conditions.

C. Object recognition

Object recognition finds numerous applications in visual surveillance, human-machine interfaces, image retrieval and visual content analysis systems. With the rapid evolution of CMOS active pixel sensing (APS) and advancements in deep learning [18], [19], researchers have already achieved impressive gains in APS-based object recognition accuracy. However, neuromorphic vision sensing (NVS) provides alternative to APS for energy-saving and computation-efficient object recognition systems [20].

Feature descriptors with classifiers for object recognition have been widely used by the NVS community. These descriptors include corner detectors and line/edge extraction [21]–[23], optical flow [24]–[26], spatio-temporal time-surface feature descriptors [27]–[29]. While these efforts were promising early attempts for NVS-based object recognition, their performance does not scale well when considering complex datasets such as N-Caltech101 and CIFAR10-DVS [29]. Moreover, optical flow and time-surfaces feature extraction have very high computational requirements, which diminishes their usability in real-time applications.

Another avenue for NVS-based object recognition is via frame-based methods, i.e., converting the neuromorphic events into synchronous frames of spike events, on which conventional computer vision techniques can be applied [30]–[32]. For example, Zhu [30] introduced a four-channel image form with the same resolution as the neuromorphic vision sensor: the first two channels encode the number of positive and negative events that have occurred at each pixel, while last two channels as the timestamp of the most recent positive and negative event; then CNNs are leveraged to perform the downstream recognition task by inputting constructed images. However, these methods do not offer the compact and asynchronous nature of NVS, as the frame sizes that need to be processed are substantially larger than those of the original NVS streams.

The third type of neuromorphic object recognition consists of event-based methods. The most commonly used architecture is based on spiking neural networks (SNNs) [33]–[36]. While SNNs are theoretically capable of learning complex representations, they have still not achieved the performance of gradient-based methods because of lack of suitable training algorithms. Essentially, since the activation functions of spiking neurons are not differentiable, SNNs are not able to leverage on popular training methods such as backpropagation. To address this, researchers currently follow an intermediate step [37]–[40]: a neural network is trained off-line using continuous/rate-based neuronal models with state-of-the-art supervised training algorithms and then the trained architecture is mapped to an SNN. However, until now, despite their substantial implementation advantages at inference, the obtained solutions are complex to train and have typically achieved lower performance than gradient-based CNNs.

The final type is our recently-proposed end-to-end graph-based framework [41], which represents events as graphs and couples this with graph convolution neural networks for object recognition. By representing events as graphs, we are able to maintain event asynchronocity and sparsity and, by using graph convolutional neural networks, we can perform training with traditional gradient-based backpropagation. As shown in Table 1 and Table 2 of [41], graph-based methods acquired superior results to the state-of-the-art in various datasets. Moreover, Table 3 of [41] compares the complexity of graph convolutional networks with conventional deep CNNs in terms of the number of floating-point operations (FLOPs) and the number of parameters, and the results show that proposed graph CNNs have a smaller number of weights and reduces the computation to one-fifth of ResNet_50 [19]. Therefore, graph-based object recognition approach for NVS can be seen as a way to bridge the compact, spike-based, asynchronous nature of NVS with the power of well-established learning methods for graph neural networks.

IV. Conclusion

The paper presented recent developments in spike-based compression, transmission and recognition aspects for visual IoT systems. The description of the various solutions proposed by the authors and collaborators can help as a reference point for further developments in the field of spike-based visual sensing and processing.

Acknowledgment

The authors acknowledge the support of EPSRC via grants EP/P022715/1, EP/P02243X/1 and EP/P022723/ (The Internet of Silicon Retinas: Machine to machine communications for neuromorphic vision sensing data (IoSiRe)).

References

[1] F. Xiong, A. Li, H. Wang, and L. Tang, "An SDN-MQTT based communication system for battlefield UAV swarms," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 41–47, 2019.

[2] C. Gomez, J. Paradells, C. Bormann, and J. Crowcroft, "From 6LoWPAN to 6Lo: Expanding the universe of IPv6-supported technologies for the Internet of Things," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 148–155, 2017.

[3] J. O. Mills, A. Jalil, and P. E. Stanga, "Electronic retinal implants and artificial vision: journey and present," *Eye*, vol. 31, no. 10, pp. 1383–1398, 2017.

[4] D. Tedaldi, G. Gallego, E. Mueggler, and D. Scaramuzza, "Feature detection and tracking with the dynamic and active-pixel vision sensor," in *IEEE International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP), Krakow, Poland, June 2016*.

[5] J. Xu, Y. Andrepoulos, Y. Xiao, and M. van Der Schaar, "Non-stationary resource allocation policies for delay-constrained video streaming: Application to video over internet-of-things-enabled networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 782–794, 2014.

[6] C. Tan, S. Lallee, and G. Orchard, "Benchmarking neuromorphic vision: lessons learnt from computer vision," *Frontiers in neuroscience*, vol. 9, p. 374, 2015.

[7] J. Zhang, T. Xiong, T. Tran, S. Chin, and R. Etienne-Cummings, "Compact all-CMOS spatiotemporal compressive sensing video camera with pixel-wise coded exposure," *Optics express*, vol. 24, no. 8, pp. 9013–9024, 2016.

[8] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis *et al.*, "Event-based vision: A survey," *arXiv preprint arXiv:1904.08405*, 2019.

[9] A. Bakshi, L. Chen, K. Srinivasan, C. E. Koksal, and A. Eryilmaz, "EMIT: An efficient MAC paradigm for the internet of things," *IEEE/ACM Transactions on Networking*, vol. 27, no. 4, pp. 1572–1583, 2019.

[10] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[11] J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, "DDD17: End-to-end DAVIS driving dataset," in *IEEE International Conference on Machine Learning (ICML), Sydney, Australia, August 2017*.

[12] Z. Bi, S. Dong, Y. Tian, and T. Huang, "Spike coding for dynamic vision sensors," in *IEEE Data Compression Conference (DCC), Utah, USA, March 2018*.

[13] N. Khan and M. G. Martini, "Data rate estimation based on scene complexity for dynamic vision sensors on unmanned vehicles," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna, Italy, September 2018*.

[14] ——, "Bandwidth modeling of silicon retinas for next generation visual sensor networks," *Sensors*, vol. 19, no. 8, p. 1751, 2019.

[15] S. Dong, Z. Bi, Y. Tian, and T. Huang, "Spike coding for dynamic vision sensor in intelligent driving," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 60–71, 2019.

[16] K. Iqbal, N. Khan, and M. G. Martini, "Performance comparison of lossless compression strategies for dynamic vision sensor data," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE, May. 2020.

[17] H. Saki, T. Peng, and M. S. Bahae, "Energy efficient delay sensitive optimization in swipt-mimo," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–7.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[20] T. Delbruck, "Neuromorphic vision sensing and processing," in *Europ. Solid-State Dev. Res. Conf*, 2016, pp. 7–14.

[21] V. Vasco, A. Glover, and C. Bartolozzi, "Fast event-based harris corner detection exploiting the advantages of event-driven cameras," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4144–4149.

[22] E. Mueggler, C. Bartolozzi, and D. Scaramuzza, "Fast event-based corner detection," in *British Machine Vis. Conf.(BMVC)*, vol. 1, 2017.

[23] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam," *The International Journal of Robotics Research*, vol. 36, no. 2, pp. 142–149, 2017.

[24] X. Clady, J.-M. Maro, S. Barré, and R. B. Benosman, "A motion-based feature for event-based pattern recognition," *Frontiers in neuroscience*, vol. 10, p. 594, 2017.

[25] F. Barranco, C. Fermuller, and Y. Aloimonos, "Bio-inspired motion estimation with event-driven sensors," in *International Work-Conference on Artificial Neural Networks*. Springer, 2015, pp. 309–321.

[26] T. Brosch, S. Tschechne, and H. Neumann, "On event-based optical flow detection," *Frontiers in neuroscience*, vol. 9, p. 137, 2015.

[27] G. Orchard, C. Meyer, R. Etienne-Cummings, C. Posch, N. Thakor, and R. Benosman, "Hfirst: a temporal approach to object recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 10, pp. 2028–2040, 2015.

[28] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "Hots: a hierarchy of event-based time-surfaces for pattern recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 7, pp. 1346–1359, 2017.

[29] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "Hats: histograms of averaged time surfaces for robust event-based object classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1731–1740.

[30] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Ev-flownet: self-supervised optical flow estimation for event-based cameras," *arXiv preprint arXiv:1802.06898*, 2018.

[31] M. Cannici, M. Ciccone, A. Romanoni, and M. Matteucci, "Attention mechanisms for object recognition with event-based cameras," *arXiv:1807.09480*, 2018.

[32] A. Chadha, Y. Bi, A. Abbas, and Y. Andreopoulos, "Neuromorphic vision sensing for cnn-based action recognition," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 7968–7972.

[33] F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G.-J. Nam *et al.*, "Truenorth: Design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 10, pp. 1537–1557, 2015.

[34] P. U. Diehl and M. Cook, "Unsupervised learning of digit recognition using spike-timing-dependent plasticity," *Frontiers in computational neuroscience*, vol. 9, p. 99, 2015.

[35] J. H. Lee, T. Delbruck, and M. Pfeiffer, "Training deep spiking neural networks using backpropagation," *Frontiers in neuroscience*, vol. 10, p. 508, 2016.

[36] E. Neftci, S. Das, B. Pedroni, K. Kreutz-Delgado, and G. Cauwenberghs, "Event-driven contrastive divergence for spiking neuromorphic systems," *Frontiers in neuroscience*, vol. 7, p. 272, 2014.

[37] P. O'Connor, D. Neil, S.-C. Liu, T. Delbruck, and M. Pfeiffer, "Real-time classification and sensor fusion with a spiking deep belief network," *Frontiers in neuroscience*, vol. 7, p. 178, 2013.

[38] P. U. Diehl, D. Neil, J. Binas, M. Cook, S.-C. Liu, and M. Pfeiffer, "Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[39] J. A. Pérez-Carrasco, B. Zhao, C. Serrano, B. Acha, T. Serrano-Gotarredona, S. Chen, and B. Linares-Barranco, "Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing–application to feedforward convnets," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2706–2719, 2013.

[40] E. Stromatias, D. Neil, F. Galluppi, M. Pfeiffer, S.-C. Liu, and S. Furber, "Scalable energy-efficient, low-latency implementations of trained spiking deep belief networks on spinnaker," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[41] Y. Bi, A. Chadha, A. Abbas, , E. Bourtsoulatze, and Y. Andreopoulos, "Graph-based object classification for neuromorphic vision sensing," in *2019 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2019.