ROT-Harris: A Dynamic Approach to Asynchronous Interest Point Detection

S. Harrigan(*), S. Coleman,
D. Kerr, P. Yogarajah
School of Computing, Engineering
and Intelligent Systems
Ulster University
Northern Ireland, UK
BT480QR
{sp.harrigan,sa.coleman
d.kerr,p.yogarajah}
@ulster.ac.uk

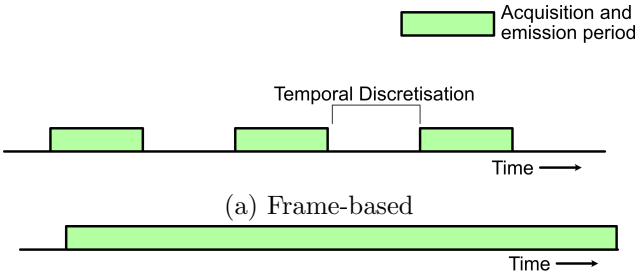
Z. Fang, C. Wu
Faculty of Robot Science
and Engineering
Northeastern University
Liaoning, China
110169
{fangzheng,wuchengdong}
@mail.neu.edu.cn

Abstract

Event-based vision sensors are a paradigm shift in the way that visual information is obtained and processed. These devices are capable of low-latency transmission of data which represents the scene dynamics. Additionally, low-power benefits make the sensors popular in finite-power scenarios such as high-speed robotics or machine vision applications where latency in visual information is desired to be minimal. The core datatype of such vision sensors is the 'event' which is an asynchronous per-pixel signal indicating a change in light intensity at an instance in time corresponding to the spatial location of that sensor on the array. A popular approach to event-based processing is to map events onto a 2D plane over time which is comparable with traditional imaging techniques. However, this paper presents a disruptive approach to event data processing that uses a tree-based filter framework that directly processes raw event data to extract events corresponding to interest point features, which is then combined with a Harris interest point approach to isolate features. We hypothesise that since the tree structure contains the same spatial information as a 2D surface mapping, Harris may be applied directly to the content of the tree, bypassing the need for transformation to the 2D plane. Results illustrate that the proposed approach performs better than other state-of-the-art approaches with limited compromise on the run-time performance.

1 Introduction

Event-based vision (EV) sensors are rapidly growing in use for applications where rapid information access is essential such as high-speed robotics and machine vision applications. EV sensors are bio-inspired emulations of retinal neural behaviour [16, 4] which asynchronously release information on a pixel-by-pixel basis unlike frame-based sensors which produce a complete frame of information using fixed cyclic emission rates (Fig.1). The core datatype of EV sensors is the event $e_i = \langle t_i, x_i, y_i, pol_i \rangle$ where t is the emission time of the event, x and y are spatial identifiers of the emitting pixel in the sensor array, $pol = \pm 1$ indicates an increase or decrease in light intensity and i is the index of the event within $E = \{e_1, \ldots, e_i\}$. EV is still in it's infancy within computer vision research and its potentials have yet to be fully realised.



(a) Frame-based

(b) Event-based

Figure 1: Illustration of 1(a) frame-based emission showcasing the fix cyclic emission rate with temporal discretisation and 1(b) showing the asynchronous emission of EV sensors

The prevalent approach to processing EV data is to project it onto a 2D plane via integration over time to produce an event-frame analogous to a frame from classical image sensor. Such event-based planes are often accompanied by processing mechanisms to maintain the temporal relevance normally lost through integration such as the time-surface (TS) [27] and surface-of-active-events (SAE) [3] which exponentially decays the event's contribution on the surface over time. This approach can be considered analogous to motion history images [1]. The obvious downsides to these mapping processes are (1) sensitivity to the effects of sparse event data, (2) loss of event time resolution and (3) reduction in temporal precision. Research has focussed

on minimising these issues [11, 12, 25, 26] but they remain inherent due to the frame-based mapping.

A binary tree is a data structure which consists of nodes and edges. The term binary refers to a rule that a node can only ever possess two child nodes. Binary trees have been used in computer vision tasks, for example [24] where a form of binary tree known as a binary partition tree is used for image segmentation. Similarly in [22] a binary tree is used for image compression and in [5] binary trees are used in decomposition of image data for shape analysis and pattern recognition as well as shape matching [14]. More recently binary trees have been used for panorama construction [10] and visual localisation [13]. The approach proposed here is based on a novel binary tree data structure and the corresponding algorithms. This data representation and feature extraction framework does not require a 2D mapping of the EV data but rather it can directly process the raw EV data.

The identification of interest points is one of the fundamental building blocks of many machine vision applications; therefore it has been an area of interest since the early development of EV such as in [6, 28, 18, 15]. One example of an interest point detector is the Harris [9] detector, an established method for extracting salient points from an image using characteristics of surface gradient. The Harris approach has been successfully applied to EV data via 2D plane mapping [29, 15, 19]. However this inherits the key downsides of the mapping process as highlighted above.

This paper presents a novel approach to interest point detection using the Reduction-Over-Time (ROT) tree [8] and a novel event-driven adaptation of the Harris interest point detector [9] called the ROT-Harris. The ROT tree offers a fast means of processing EV data and has been shown to be beneficial for noise reduction. In Sec. 2 we present the event-driven ROT-Harris framework followed by experimentation and comparative analysis with leading state-of-the-art interest point detection methods in Sec. 3. Conclusions in Sec. 4 discuss how the ROT tree approach enables interest point extraction without the need for 2D plane mapping.

2 ROT-Harris

The ROT-Harris is a novel framework for interest point detection using ROT spatially indexed trees $RS(e_i)$. ROT makes use of a data reduction model $P(k) \longrightarrow 0.184/\log_{10} k^{1.25} + 0.184$ (where k is the time difference between the current event and previous events), which emulates information retention over time, performs self-pruning (using a threshold τ) and uses the same self-balancing approaches adopted by binary trees such as the Red-Black tree [7].

ROT trees are self-pruning, based on the $P(k)$ retention model which produces a metric of relative distance between e_i and $\forall e_j$ where $j = 1, \ldots, i-1$, and self-balancing data structures optimised for EV data and capable of custom indexing schemes. RS indexes use x, y co-ordinates and maintain temporal precision of EV data based on $P(k)$. We propose to use two ROT trees to represent EV data polarity pol where $RS_+ \forall e : pol = 1$ and $RS_- \forall e : pol = -1$. A salient event is then deemed to be an event which occurs in both RS_+ and RS_- and the temporal difference in events is $<< 0.01$ as illustrated in Fig. 2. This eliminates EV data noise which can be described as isolated events [30] that occur in one tree or the other, but not both. We define the resulting tree RE of salient events as $RE = RS_+ \cap RS_-$. We hypothesise that RS_+ and RS_- contain the same spatial information as surface-of-active-events (SAE) under the same conditions whereas RE, contains a refined subset of only salient events. Therefore, Harris is applicable to RE using local event of interest transformations to the 2D plane, rather than the global transformation used in [29, 15, 19]. In order to apply the Harris operator to RE we create a local binary patch d of size $L \times L$, centred on the event of interest $e_i = \langle t_i, x_i, y_i, pol_i \rangle$ as denoted in Eq. 1.

$$d(x_{i+n}, y_{i+n}) = \begin{cases} 1 & \text{if } e_{i+n} \in RE \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

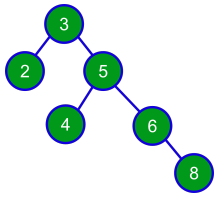
where $d(x_{i+n}, y_{i+n})$ is the spatial neighbours surrounding e_i and $n = (-\frac{L-1}{2}, ..., \frac{L-1}{2})$. The approximation symmetry matrix M used by Harris is defined as

$$M = \begin{bmatrix} \sum\limits_d g(d) B_h^2 & \sum\limits_d g(d) B_h B_v \\ \sum\limits_d g(d) B_h B_v & \sum\limits_d g(d) B_v^2 \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \tag{2}$$

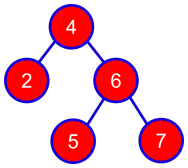
where d is the local binary patch from Eq.1, $g(d)$ is a Gaussian filter, B_h and B_v are the horizontal and vertical gradients computed using the Sobel operator. The Harris response $R = (ab - b^2) - k(a+b)^2$ where $k = [0.04, 0.06]$ is computed for each e_i. This calculation is known to be computationally expensive when applied to 2D plane surfaces such as SAE and TS but the nature of the ROT tree structure ensures sparse and minimal processing as only events of interest are processed.

3 Performance Evaluation

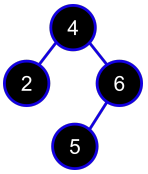
To evaluate ROT-Harris we use publicly available datasets [21]; specifically we use the shapes, boxes, walking and run datasets. The datasets are captured using a 240×180 resolution event-based and frame-based hybrid camera known as the DAVIS240-C [4]. Regarding the base RS, the nodes are pruned when the $P(k) = 0$ as the trees evolve over time. For comparison we use the following state-of-the-art methods for interest point detection using EV data: eHarris [29], FA-Harris [15], TLF-Harris [19] and Arc* [2]. Each of these methods make use of a 2D plane such as SAE. The eHarris applies the original Harris method directly to an SAE, FA-Harris applies the Harris operator to the localised SAE to determine an interest point. An interest point is then checked for saliency over time, and if determined to be salient is mapped to the global SAE. TLF-Harris is a modification of Harris designed for



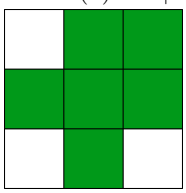
(a) RS_+



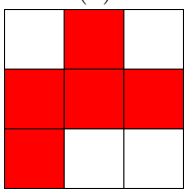
(b) RS_-



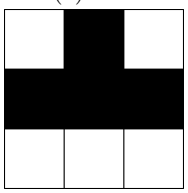
(c) RE



(d) 2D image of 2(a)



(e) 2D image of 2(b)



(f) 2D image of 2(c)

1	2	3
4	5	6
7	8	9

(g) The 2D surface mapping

Figure 2: Illustration of the ROT showing 2(a) RS_+, 2(b) RS_-, 2(c) RE, 2(d) a 2D mapping of 2(a), 2(e) a 2D mapping of 2(b) and 2(f) a 2D mapping of 2(c) where the mapping is illustrated in 2(g) (best viewed in colour)

minimised resource usage by applying approximations to local binary patches using various SAEs and event lifetime adjustments (keeping active events on a surface over a period of time based on a spatial and temporal analysis of neighbours or pruning inactive events). Arc* applies a modified version of the Features from Accelerated Segment Test (FAST) [23, 20] on an SAE surface. By incrementally increasing the search operation in arcs, Arc* overcomes the 180° limitations of FAST to achieve full 360° coverage.

Data contained in RS_+ and RS_- are presented in Fig.3(a) where we visually represent the events on a 2D plane. Fig.3(b) shows the resulting RE tree of salient events. Here we can see that significant noise removal has occurred. Fig. 3(c) shows the output from ROT-Harris (red) overlaid on the visual representation of events in Fig.3(a). For comparative purposes Fig.3(d) illustrates the interest points extracted using TLF-Harris using an SAE. By visually comparing Fig. 3(c) with Fig.3(d), we can see that our proposed approach ROT-Harris obtains several more accurately located interest points than TLF-Harris.

For quantitative evaluation, we use the technique proposed in [2] to evaluate the accuracy of EV interest point detection in terms of true and false interest points (TP and FP respectively). This has been widely adopted as an architecture for comparison between interest point detection methods. To facilitate this, ground-truth (GT) interest points are detected using the Kanade-Lucas-Tomasi (KLT) [17] tracking algorithm with the Harris interest point detector [9] operating on the intensity images from the dataset alongside cubic spline matching over a defined temporal period to track interest points. Cylinders are centred on the GT interest points and detected interest points are labelled as True-Positives (TP) if they fall within a cylindrical region defined with a 3-pixel spatial radius and the height of the cylinder is defined as a period of time corresponding to 5ms. Similarly, interest points are defined as False Positives (FP) if they are detected outside the 3-pixel GT cylindrical area but within a 5-pixel GT cylindrical area. Accuracy is then calculated as

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

Tab.1 reports the accuracy for each method using each dataset. Highlighted in bold are the highest accuracy results for each of the four datasets and an additional fifth column is provided showing the overall average accuracy across all the datasets. Each row corresponds to the detection method used. The proposed ROT-Harris is the best performing method in terms of accuracy for all datasets except for the walking dataset. This can be attributed to the pixel distance sensitivity (i.e., proximity to GT interest points) that the ROT trees maintain compared to the 2D surfaces of the other methods. Additionally, overall ROT-Harris has the highest average accuracy across all datasets. Tab.2 reports the execution times in nanoseconds for ROT-Harris, TLF-Harris (selected as it was the second best in terms of accuracy), and Harris operating over RE which has been mapped to a 2D plane; we use the eHarris [29] adaptation of Harris, with thresholding set as in [20], as a base metric to demonstrate the computational speed of our approach.

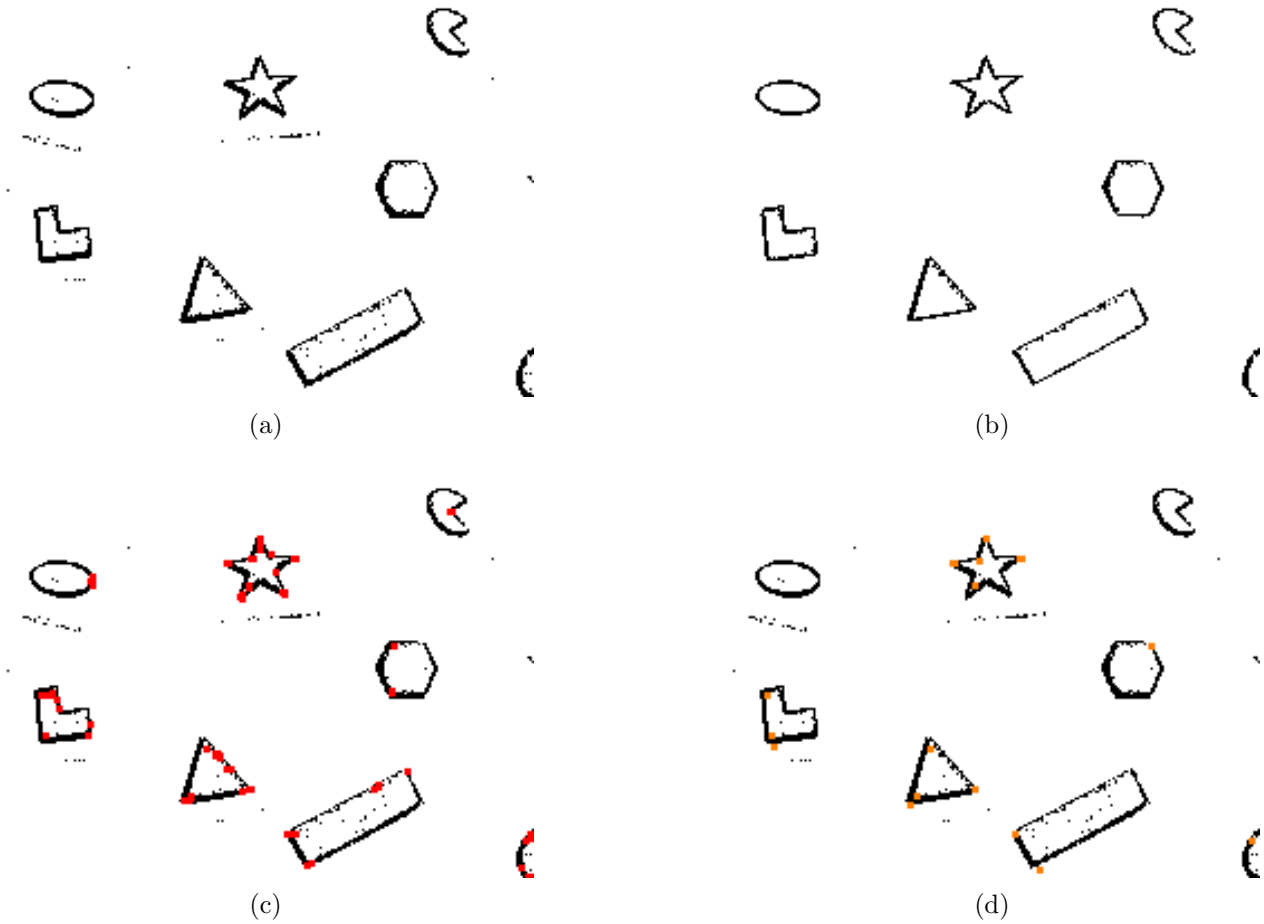


Figure 3: Visual representation of: (a) Data contained in RS_+ and RS_- visually represented on a 2D plane; (b) Resulting RE tree of salient events; (c) Output from ROT-Harris (red) overlaid on the visual representation of events in (a); (d) Interest points extracted using TLF-Harris using an SAE.

Table 1: Accuracy [%] gained from each method (row) for each data-set (column)

Method \ Dataset	Shapes	Boxes	Walking	Run	Overall
eHarris[29]	56.89	49.16	66.40	62.07	58.63
FA-Harris[15]	57.66	49.66	65.32	49.66	55.58
TLF-Harris[19]	63.20	53.27	**72.13**	68.74	64.34
Arc*[2]	55.38	49.01	52.41	53.41	52.55
ROT-Harris (**ours**)	**71.88**	**61.13**	70.21	**70.56**	**68.45**

Table 2: Average execution time [nSec] compared with Harris for each data-set (column)

Method \ Dataset	Shapes	Boxes	Walking	Run
ROT-2D-eHarris[9, 29]	89	205	93	94
TLF-Harris[19]	**56**	**110**	**57**	**55**
ROT-Harris (**ours**)	68	137	67	61

Execution time was calculated on an Intel© Core i7-3770 CPU @ 3.40GHz with 16GB memory system. In Tab. 2 the fastest run-times are in bold. In Tab. 2 the TLF-Harris approach is the best performing with respect to run-time, this is due to use of an approximation of the Harris response for each patch rather than using the actual Harris response measurement. The proposed approach, ROT-Harris does not perform this approximation but has improved interest point detection accuracy compared to TLF-Harris and it still faster than Harris. This is a trade-off between accuracy and speed, depending on the needs of the machine vision application.

4 Conclusion

In this work we present the novel ROT-Harris which is an event-driven framework for interest point detection which uses the dynamic ROT binary tree to perform interest point detection on EV data. Our original hypothesis was that the ROT trees contain the same spatial information as traditional 2D visual information mapping and the presented experiments allow us to demonstrate that by approximating local binary patches around the nodes within the trees, we can

apply the established and popular Harris interest point detector. Performance evaluation against state-of-the-art EV interest point detectors demonstrates that this hypothesis holds and using ROT-Harris enables an increase in interest point detection accuracy with only a slight reduction in computational performance when compared with the leading SOTA Harris-based approaches.

References

[1] AHAD, M. A. R., TAN, J. K., KIM, H., AND ISHIKAWA, S. Motion history image: its variants and applications. *Machine Vision and Applications 23*, 2 (2012), 255–281.

[2] ALZUGARAY, I., AND CHLI, M. Asynchronous Corner Detection and Tracking for Event Cameras in Real-Time. *IEEE Robotics and Automation Letters* (2018), 1–1.

[3] BENOSMAN, R., CLERCQ, C., LAGORCE, X., IENG, S. H., AND BARTOLOZZI, C. Event-based visual flow. *IEEE Transactions on Neural Networks and Learning Systems 25*, 2 (2014), 407–417.

[4] BRANDLI, C., BERNER, R., YANG, M., LIU, S. C., AND DELBRUCK, T. A 240 x 180 130 dB 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits 49*, 10 (2014), 2333–2341.

[5] CHAUDHURI, B. Applications of quadtree, octree, and binary tree decomposition techniques to shape analysis and pattern recognition. *IEEE transactions on pattern analysis and machine intelligence 7*, 6 (1985), 652–661.

[6] CLADY, X., IENG, S. H., AND BENOSMAN, R. Asynchronous event-based corner detection and matching. *Neural Networks 66* (jun 2015), 91–106.

[7] GUIBAS, L. J., AND SEDGEWICK, R. A dichromatic framework for balanced trees. In *19th Annual Symposium on Foundations of Computer Science (sfcs 1978)* (1978), IEEE, pp. 8–21.

[8] HARRIGAN, S., COLEMAN, S., KERR, D., PRATHEEPAN, Y., FANG, Z., AND WU, C. Reducing-Over-Time Tree for Event-based Data. In *The 25th IEEE International Conference on Pattern Recognition* (2021).

[9] HARRIS, C., AND STEPHENS, M. A Combined Corner and Edge Detector. *Procedings of the Alvey Vision Conference 1988* (1988), 23.1–23.6.

[10] HERNANDEZ-LOPEZ, F. J., TREJO-SÁNCHEZ, J. A., AND RIVERA, M. Panorama construction using binary trees. *Signal, Image and Video Processing* (2020), 1–8.

[11] KIM, H., HANDA, A., BENOSMAN, R., IENG, S.-H., AND DAVISON, A. Simultaneous Mosaicing and Tracking with an Event Camera. *Proceedings of the British Machine Vision Conference 2014* (2014), 26.1—-26.12.

[12] KIM, H., LEUTENEGGER, S., AND DAVISON, A. J. Real-time 3D reconstruction and 6-DoF tracking with an event camera. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9910 LNCS* (2016), 349–364.

[13] LE, H., HOANG, T., AND MILFORD, M. J. Btel: A binary tree encoding approach for visual localization. *IEEE Robotics and Automation Letters 4*, 4 (2019), 4354–4361.

[14] LEU, J.-G., AND HUANG, I.-N. Planar shape matching based on binary tree shape representation. *Pattern Recognition 21*, 6 (1988), 607–622.

[15] LI, R., SHI, D., ZHANG, Y., LI, K., AND LI, R. Fa-harris: A fast and asynchronous corner detector for event cameras. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2019), IEEE, pp. 6223–6229.

[16] LICHTSTEINER, P., POSCH, C., AND DELBRUCK, T. A 128 x 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits 43*, 2 (2008), 566–576.

[17] LUCAS, B. D., KANADE, T., ET AL. An iterative image registration technique with an application to stereo vision. *Proceedings DARPA image Understanding Workshop* (1981).

[18] MANDERSCHEID, J., SIRONI, A., BOURDIS, N., MIGLIORE, D., AND LEPETIT, V. Speed invariant time surface for learning to detect corner points with event-based cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 10245–10254.

[19] MOHAMED, S. A., YASIN, J. N., HAGHBAYAN, M.-H., MIELE, A., HEIKKONEN, J., TENHUNEN, H., AND PLOSILA, J. Dynamic resource-aware corner detection for bio-inspired vision sensors. *arXiv preprint arXiv:2010.15507* (2020).

[20] MUEGGLER, E., BARTOLOZZI, C., AND SCARAMUZZA, D. Fast event-based corner detection. In *Müggler, Elias; Bartolozzi, Chiara; Scaramuzza, Davide (2017). Fast event-based corner detection. In: British Machine Vision Conference (BMVC), London, 4 September 2017 - 7 September 2017, 1-8.* (2017), pp. 1–8.

[21] MUEGGLER, E., REBECQ, H., GALLEGO, G., DELBRUCK, T., AND SCARAMUZZA, D. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *The International Journal of Robotics Research 36*, 2 (2017), 142–149.

[22] ROBINSON, J. A. Efficient general-purpose image compression with binary tree predictive coding. *IEEE Transactions on Image Processing 6*, 4 (1997), 601–608.

[23] ROSTEN, E., AND DRUMMOND, T. Machine learning for high-speed corner detection. *Lecture Notes in Computer Science* (2006), 430–443.

[24] SALEMBIER, P., AND GARRIDO, L. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE transactions on Image Processing 9*, 4 (2000), 561–576.

[25] SCHEERLINCK, C., BARNES, N., AND MAHONY, R. Continuous-time intensity estimation using event cameras. In *Asian Conference on Computer Vision* (2018), Springer, pp. 308–324.

[26] SCHEERLINCK, C., REBECQ, H., GEHRIG, D., BARNES, N., MAHONY, R., AND SCARAMUZZA, D. Fast image reconstruction with an event camera. In *The IEEE Winter Conference on Applications of Computer Vision* (2020), pp. 156–163.

[27] SIRONI, A., BRAMBILLA, M., BOURDIS, N., LAGORCE, X., AND BENOSMAN, R. HATS: Histograms of Averaged Time Surfaces for Robust Event-based Object Classification. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (mar 2018), 1731–1740.

[28] TEDALDI, D., GALLEGO, G., MUEGGLER, E., AND SCARAMUZZA, D. Feature detection and tracking with the dynamic

and active-pixel vision sensor (DAVIS). In *2016 2nd International Conference on Event-Based Control, Communication, and Signal Processing, EBCCSP 2016 - Proceedings* (jun 2016), IEEE, pp. 1–7.

[29] VASCO, V., GLOVER, A., AND BARTOLOZZI, C. 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)* (2016), IEEE, pp. 4144–4149.

[30] WANG, Y., DU, B., SHEN, Y., WU, K., ZHAO, G., SUN, J., AND WEN, H. Ev-gait: Event-based robust gait recognition using dynamic vision sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 6358–6367.