Event-based Visual Data Sets for Prediction Tasks in Spiking Neural Networks

Tingting (Amy) Gibson[1], Scott Heath, Robert P. Quinn, Alexia H. Lee, Joshua T. Arnold, Tharun S. Sonti, Andrew Whalley, George P. Shannon, Brian T. Song, James A. Henderson, and Janet Wiles[2]

The University of Queensland
[1]t.ng1@uq.edu.au [2]j.wiles@uq.edu.au

Abstract. For spiking networks to perform computational tasks, benchmark data sets are required for model design, refinement and testing. Classic machine learning benchmark data sets use classification as the dominant paradigm, however the temporal characteristics of spiking neural networks mean they are likely to be more useful for problems involving sequence data. To support these paradigms, we provide data sets of 11 moving scenes, each with multiple variations, recorded from a dynamic vision sensor (DVS128), comprising high dimensional (16k pixels) and low latency (15 microsecond) events. We also present a novel long range prediction task based on the DVS128 data, and introduce a pilot study of a spiking neural network learning to predict thousands of events into the future.

Keywords: event-based data sets, predictions, spiking neural networks

1 Introduction

For artificial spiking neural networks (SNNs) to move out of the laboratory and into the real world, they need to be able to process real world data, deal with real world noise, and solve real world tasks. We are interested in developing SNNs for embodied robots that face a range of tasks such as sensorimotor coordination, navigation and social interaction at the hundred millisecond time scale.

Real world robot studies are time and energy consuming, and benchmark data can facilitate the design process. While there are many data sets for testing sensory processing algorithms, computer vision in particular, they are not inherently spike based and need to be pre-processed into spikes before using with an SNN. SNNs are well suited for processing asynchronous spiking inputs with high temporal resolution, but the majority of data sets are frame-based with coarse temporal precision. In addition, despite the importance of prediction to robot cognition, the dominant paradigms within currently available data sets are those of classification, tracking and mapping.

Traditional frame-based video data contains a huge amount of redundant information, compared to biological sensory data which only signals new information. Sensory inputs are better represented by event

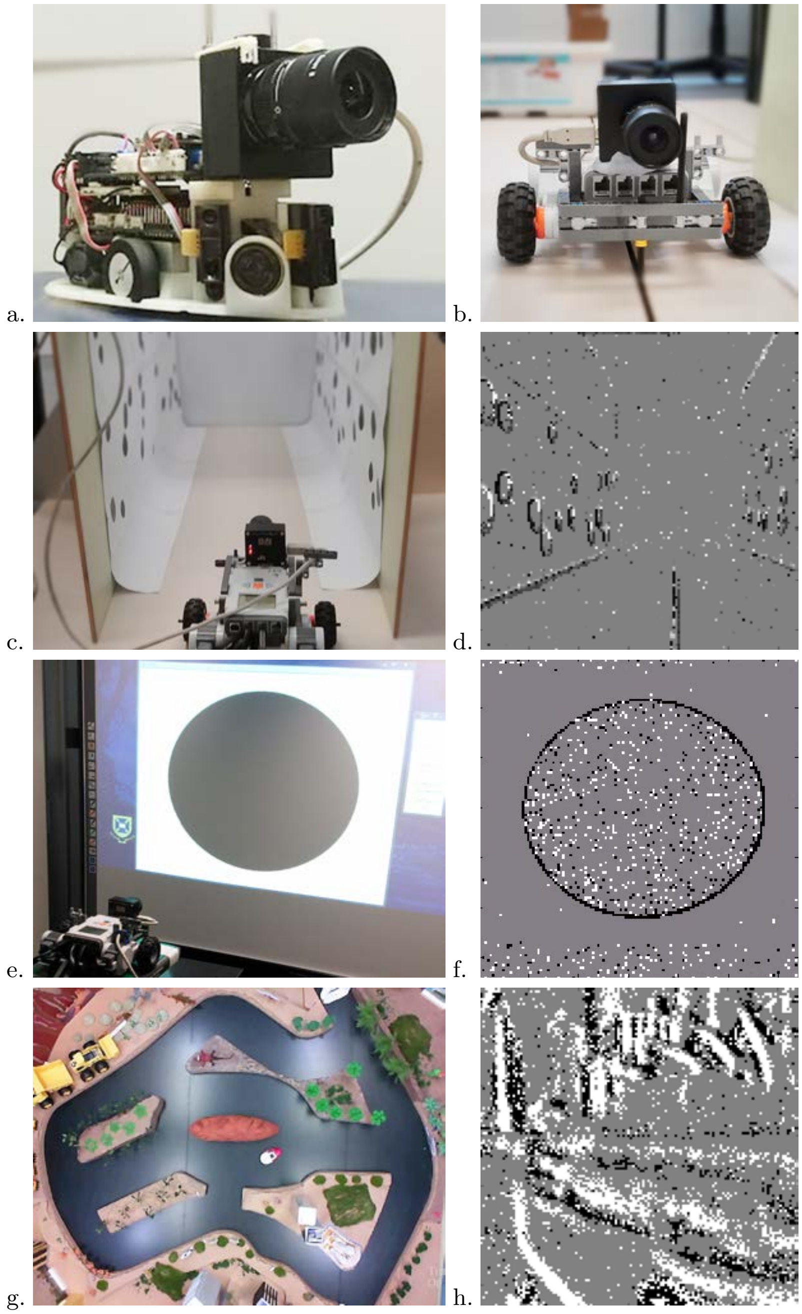


Fig. 1. Experimental setup. The DVS128 mounted on the iRat (a), and the Lego NXT robot (b). Collecting data from self-motion on a linear track (c), and its corresponding data (d). Recording from a static camera facing a projector (e), and its corresponding data (f). Overhead view of a 3x3m Australia maze to collect locomotion data (g), and a scene from the maze recorded by the DVS128 (h) .

streams—sequences of events each indicating a change in state in the environment [1]. The temporal intervals between events in the real world vary, making event streams inherently multi-scale.

High temporal precision event streams can be recorded by event-based neuromorphic sensory devices, such as the silicon retina Dynamic Vision Sensor (DVS) [2] and the silicon cochlea [3]. Such devices require low bandwidth and power, making them ideal for robot embodiment. The DVS has been used for classification tasks including recognising hand gestures [4], classifying the DVS version of the MNIST postcode data set [5] and extracting pulsed laser line for terrain reconstruction [6]. DVS data have also been used in tracking tasks, such as tracking people [7], particles [8], optic flow [9], ball and car trajectories [10, 11] and LED markers with the camera mounted on a robot [12]. Object depth has also been calculated using two DVS cameras for stereo visual inputs [13, 14]. DVS data have been fed to modified simultaneous localization and mapping (SLAM) algorithms for robot self-localisation [15, 16]; and to control robot motors [17, 18]. Most of the studies conducted with the DVS cameras are for classification, tracking and mapping tasks. There is a lack of studies that utilise the event-based nature of the DVS to investigate long range prediction.

This paper presents a novel set of benchmark data sets based on event data, and introduces a new SNN prediction paradigm with learnable time delays in both the dendrites and axons to demonstrate how the event data can be used to learn long range prediction.

2 Methods

2.1 Experimental setup

The DVS128 [2] is a neuromorphic camera that contains a 128x128 array of light sensors, with each emitting a spike asynchronously at a time resolution of 15 microseconds if a change in luminosity is detected. Spikes are encoded using a 32 bit AER (Address Event Representation) containing an *(x, y)* coordinate, a time-stamp, a change in intensity (+1 or -1) and additional synchronising bits to allow multiple cameras to be used together.

Two robot platforms were used for collecting data—UQ's iRat robot [19] and a LEGO NXT (Fig. 1 a and b). The iRat is a small rat animat which has the computational power of a PC on wheels. The iRat has IR sensors and wheel encoders which allow it to maintain distances from walls, making it suitable for recording navigation data. The LEGO robot is the equivalent of adding motors to the DVS camera. It requires a wired connection so it is more limited in its applications. However, it provides more options when mounting the camera. A data projector and a custom track were also used. A python program was written to generate feature rich patterns which were then displayed on the projector or printed out and used as the walls of the track. The general purpose event-stream viewer jAER (Java tools for AER) [20] and the embedded system based cAER (C tools for AER) [21] were used to collect all the data sets in this paper.

2.2 Data set collection: Event-based visual data

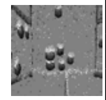
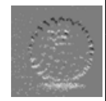
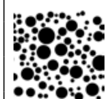
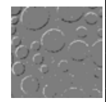
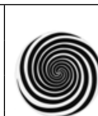
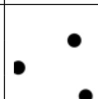
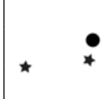
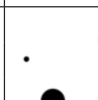
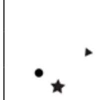
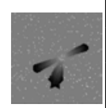
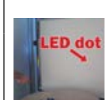
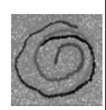
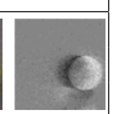
The task of interest is to learn from the high dimensional features presented in the event-based data and predict future sequences of events. A robust prediction paradigm requires diverse data sets to test its abilities. We created a wide range of data sets including 11 moving scenes, each with 1-180 variations, and 5-10 instances per variation (see Table 1). These data sets were collected using a static camera and embodied autonomous robots, focusing on self-motion (M1-M4) and object flow (M1, M3, S1a, S1c, S2-S7). The data sets contain single (M2, S1, S2-S6, S7) or multiple (M1, M3, S2-S5, S7) objects; some scenes are simple with few primitive shapes (M2a, S1a, S1c, S2, S4, S6, S7b) while the others are complex with irregular or dense objects (M1, M3, S1b, S3, S5, S7a); and a special data set for environment navigation (M4). Note that most SLAM data sets (eg. [22]) contain a single instance, here we are interested in sensory prediction, and hence we have provided multiple instances of each motion trajectory. Details of the data sets are listed in Table 1 (full data sets are available from www.itee.uq.edu.au/cis), which shows examples of the collection sources (left column) and the DVS128 data accumulated over a short period (right column):

1. Ego-motion (M1-M4)—perspective of stationary scenes from a moving agent:
   (a) The LEGO robot was moving along a straight track in which patterns of dots and shapes were attached to the walls of the track.
   (b) The LEGO robot was facing a large screen displayed with patterns of dots and shapes. The robot was moving towards and away from the projected images to collect data generated by ego-motion.
   (c) The iRat mounted with the DVS128 was navigating through the UQ Australia maze to collect data of complex scenery change.
2. Static camera (S1-S7)—perspective on object motion from a static agent:
   (a) The stationary DVS128 was facing a projector screen displaying a wide range of transforming patterns, vary from simple to complex trajectories.
   (b) The stationary DVS128 facing real moving objects.

2.3 Performance measurement of sensory data stream prediction

To evaluate the prediction paradigms, their performance needs to be measured and assessed. An advantage of tasks that predict the future of sensory data streams is that the "ground truth" is self-contained within the data. Performance measurement of a prediction paradigm depends on the representation of the event streams. If these events are treated as spikes and the predicting model is also outputting spikes, binned spike count differences, iterative closest point (ICP) [23] or any distance metric (Euclidean, Hamming, mean squared error) could be used to compare the predictions to the data. If the events are represented as probability densities (as used in Section 3 or [7]) the accuracy of the predicted densities can be compared to the actual event densities by using a range of divergences including chi-squared divergence or Kullback-Liebler divergence [24].

Table 1. The benchmark data sets (www.itee.uq.edu.au/cis)

Self-motion datasets	**Trials**	**Sources**	**Data**
M1. Moving forward and backward on a track			
M1a Variable dots, 4 spacings x 3 speeds	12x10=120		
M1b Various shapes, 3 spacings x 3 speeds	9x10=90		
M2. Facing objects, moving forward and backward			
M2a Smooth ball, 3 speeds	3x10=30		
M2b Spiky ball, 3 speeds	3x10=30		
M3. Facing projector, moving forward and backward			
M3a Same sized dots, 2 sizes x 3 speeds	6x10=60		
M3b Variable dots, 3 variations x 3 speeds	9x10=90		
M3c Variable sized shapes, 3 speeds	3x10=30		
M4. iRat navigating Australia maze	1x5=5	(see Fig. 1 g & h)	
Static camera data sets			
S1. Shape transformations			
S1a Circle expanding and contracting	1x10=10		
S1b Spinning spirals	1x10=10		
S1c Disc moving 1 in circle & 1 in spiral	2x10=10		
S2. Uniform circles moving in trajectories			
S2a 1-4 circles x 3 sizes x 3 speeds x 5 paths	180x10=1800		
S2b Variable sizes & speeds, 1-4 circles x 5 paths	20x10=200		
S3. Various shapes moving in trajectories			
S3a 2-5 shapes x 3 sizes x 3 speeds x 5 paths	180x10=1800		
S3b Variable sizes & speeds, 2-5 shapes x 5 paths	20x10=200		
S4. Circles radiating out			
S4a 1-5 circles x 2 sizes x 3 speeds x 5 paths	150x10=1500		
S4b Variable sizes & speeds, 1-5 circles x 5 paths	25x10=250		
S5. Various shapes radiating out			
S5a 2-4 shapes x 2 sizes x 3 speeds x 5 paths	90x10=900		
S5b Variable sizes & speeds, 2-4 shapes x 5 paths	15x10=150		
S6. LED pointer drawing			
S6a Circles; S6b Figure 8; S6c Sideway movement; S6d Inward spirals; S6e Squares; S6f Stars	6x10 = 60		
S7. Real objects			
S7a Remote car moving, 4 directions	4x5 = 20		
S7b Ball rolling	1x5 = 5		

3 Spiking neural network prediction of DVS128 data

We have developed a novel spiking neuron model, in which a spike can be viewed as a link between a past temporal pattern recorded by the neuron's dendritic arbour and a future pattern effected by the neuron's axonal arbour [25]. A network of such prediction neurons trained on data set M2a (Table 1) predicted the motion trajectory of the object into the far future (Fig. 2). The input event stream was first approximated as self-scaling Gaussian distributions, and the network captured the temporal relationship of these events using learnable temporal delays at both dendrites and axons (Fig. 2-left). The dendrites (blue and magenta lines) learnt from inputs of both positive (blue contour lines) and negative (magenta contour lines) event distributions and the axons (black and red lines) predicted into the future a sequence of both positive (black contour lines) and negative (red contour lines) event distributions; see [25] for details of the delays and learning mechanism. After training the network with 43.13s of the data, the network was tested with an unseen section (0.66s) of the data (blue contour lines in Fig. 2-right are the Gaussian representations of the input data) and it predicted the future trajectory (for 14.13s) of the moving object (predicted positive events distributions in black and that of the negative events in red).

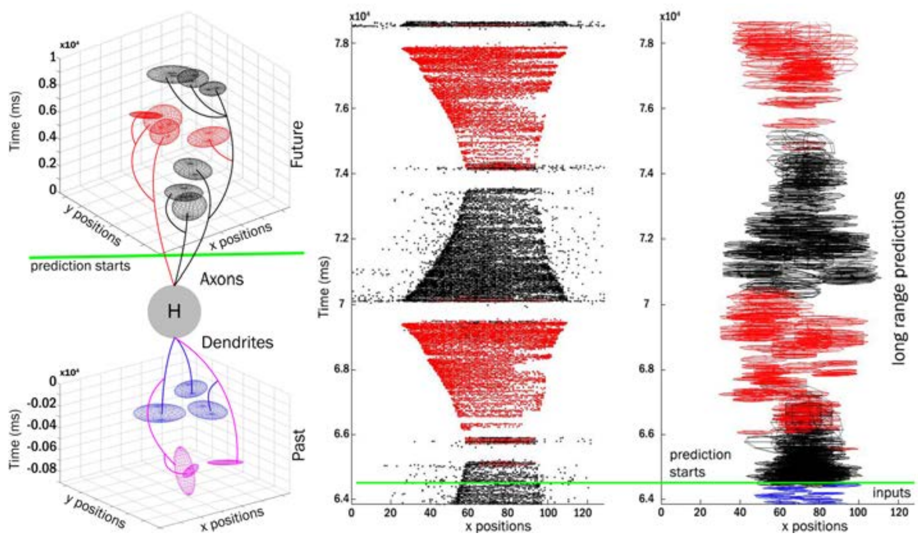


Fig. 2. The SNN prediction paradigm, with contour lines indicating the Gaussian representation of the data. Left: A prediction spiking neuron with dendrites that sample the past (blue: positive; magenta: negative) and axonal branches which predict into the future (black: positive; red: negative). Middle: The raw data in spikes of data set M2a (black dots: positive events; red dots: negative events). Right: The network's prediction of the future events in the form of Gaussians (blue: positive input; black: predicted positive; red: predicted negative).

4 Summary

The DVS128 data sets were developed as an SNN community resource, to facilitate the work of our group's bio-inspired research and to provide easily accessible data for other modellers. Key design features include (i) ego-motion (resulting in whole field dynamics; sets M1-4) and object motion (local field dynamics; sets S1-7); (ii) trajectories suitable for prediction tasks based on episodic memory (rather than categorisation tasks); (iii) event-based encoding which does not require uniform time sampling or video frames, comparable to the clock-free processes of biological neurons; (iv) real-world recordings (rather than simulated data); (v) a range of simple and complex trajectories. The main challenge for collecting the data was creating untethered ego-motion, which was addressed by mounting the DVS128 on the UQ iRat robot and using cAER for direct data capture. Practical challenges result from the size of DVS data. For model development, we have used self-scaling Gaussians to estimate space-time densities, which enables rapid prototyping, with a view to increasing fidelity with more computational power.

Current projects using this data include an event-based SNN to predict future motion trajectories (see Section 3 and [25]); an SNN with a deep learning architecture designed to abstract different levels of features from the temporal flow; and vector flow analysis for prediction of motion trajectories.

Acknowledgments This work was supported by an APA to TG, AOARD Grant FA2386-12-1-4050 and UQ summer scholarships. The authors are grateful to David Tingley for his support and Tobi Delbrück for his valuable advice and for the jAER software infrastructure.

References

1. K. M. Chandy, "Event-driven applications: costs, benefits and design approaches," *Presented at the Gartner Application Integration and Web Services Summit, San Diego, CA*, 2006.
2. P. Lichtsteiner *et al.*, "A 128x128 120 db 15 s latency asynchronous temporal contrast vision sensor," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 2, pp. 566–576, 2008.
3. S.-C. Liu *et al.*, "Event-based 64-channel binaural silicon cochlea with q enhancement mechanisms," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 2010, pp. 2027–2030.
4. E. Y. Ahn *et al.*, "Dynamic vision sensor camera based bare hand gesture recognition," in *Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2011 IEEE Symposium on*, April 2011, pp. 52–59.
5. P. O'Connor *et al.*, "Real-time classification and sensor fusion with a spiking deep belief network," *Frontiers in Neuroscience*, vol. 7, no. 178, 2013.
6. C. Brandli *et al.*, "Adaptive pulsed laser line extraction for terrain reconstruction using a dynamic vision sensor," *Frontiers in Neuroscience*, vol. 7, no. 275, 2014.

7. E. Piatkowska *et al.*, “Spatiotemporal multiple persons tracking using dynamic vision sensor,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on.* IEEE, 2012, pp. 35–40.
8. D. Drazen *et al.*, “Toward real-time particle tracking using an event-based dynamic vision sensor,” *Experiments in fluids*, vol. 51, no. 5, pp. 1465–1469, 2011.
9. F. Koeth *et al.*, “Self-organisation of motion features with a temporal asynchronous dynamic vision sensor,” *Biologically Inspired Cognitive Architectures*, vol. 6, no. 0, pp. 8 – 11, 2013.
10. O. Bichler *et al.*, “Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity,” *Neural Networks*, vol. 32, pp. 339–348, 2012.
11. F. Gomez-Rodriguez *et al.*, “Real time multiple objects tracking based on a bio-inspired processing cascade architecture,” in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, May 2010, pp. 1399–1402.
12. A. Censi *et al.*, “Low-latency localization by active led markers tracking using a dynamic vision sensor,” in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on.* IEEE, 2013, pp. 891–898.
13. P. Rogister *et al.*, “Asynchronous event-based binocular stereo matching,” *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 23, no. 2, pp. 347–353, 2012.
14. E. Pikatkowska and A. N. Belbachir, “Asynchronous stereo vision for event-driven dynamic stereo sensor using an adaptive cooperative approach,” pp. 45–50, 2013.
15. D. Weikersdorfer *et al.*, “Simultaneous localization and mapping for event-based vision systems,” in *Computer Vision Systems.* Springer, 2013, pp. 133–142.
16. D. Weikersdorfer and J. Conradt, “Event-based particle filtering for robot self-localization,” in *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on.* IEEE, 2012, pp. 866–870.
17. F. Perez-Peña *et al.*, “Neuro-inspired spike-based motion: From dynamic vision sensor to robot motor open-loop control through spike-vite,” *Sensors*, vol. 13, no. 11, pp. 15 805–15 832, 2013.
18. T. Delbrück and M. Lang, “Robotic goalie with 3ms reaction time at 4% cpu load using event-based dynamic vision sensor,” *Frontiers in Neuroscience*, vol. 7, no. 223, 2013.
19. D. Ball *et al.*, “irat: Intelligent rat animat technology,” in *Proceedings of the 2010 Australasian Conference on Robotics and Automation*, 2010, pp. 1–3.
20. T. Delbrück, “jaer open source project,” 2007. [Online]. Available: http://jaer.wiki.sourceforge.net
21. ——, “caer,” 2013. [Online]. Available: http://sourceforge.net/projects/jaer/files/cAER/
22. D. Ball *et al.*, “Openratslam: an open source brain-based slam system,” *Autonomous Robots*, vol. 34, no. 3, pp. 149–176, 2013.
23. P. Besl and N. D. McKay, “A method for registration of 3-d shapes,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 239–256, Feb 1992.
24. S. Kullback and R. A. Leibler, “On information and sufficiency,” *The Annals of Mathematical Statistics*, pp. 79–86, 1951.
25. T. Gibson *et al.*, “Predicting temporal sequences using an event-based spiking neural network incorporating learnable delays,” in *2014 The International Joint Conference on Neural Networks (IJCNN)*, 2014 in press.