University of Zurich[UZH]

ROBOTICS & PERCEPTION GROUP

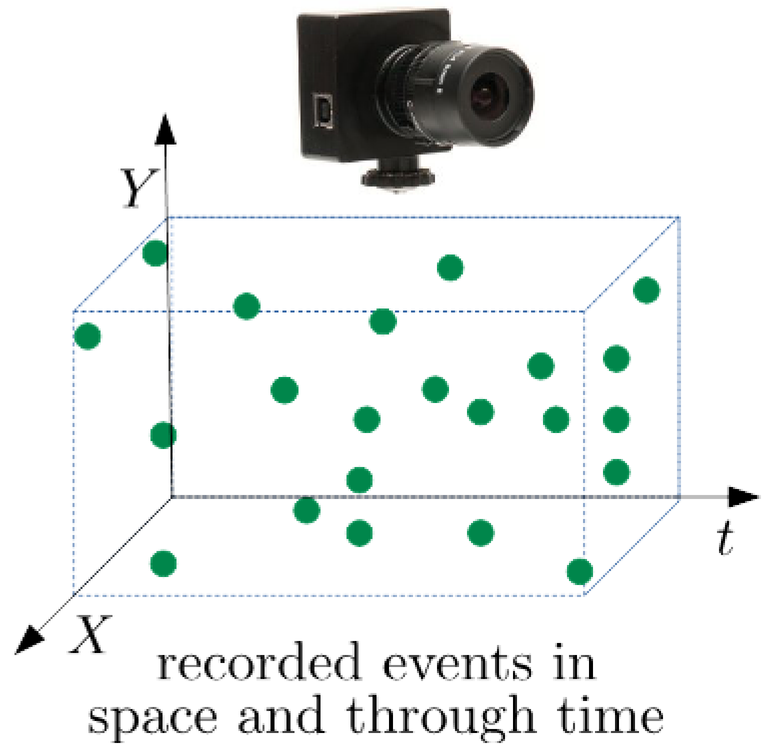
Unsupervised Feature Learning for Event Data: Direct vs Inverse Problem Formulation

IAPR/IEEE, ICPR2021, Milano, Italy

D. Kostadinov and D. Scarammuza

Introduction

- Event-based camera
  - Captures changes in intensity
  - Asynchronous
  - Advantages
    - High temporal resolution,
    - High dynamic range and
    - No motion blur



- Key question
  - How to extract meaningful and useful information?

Prior Work

- Handcrafted feature extraction approaches for event-based data

- Learning based approaches for event-based data
  - Spiking Neural Networks (SNNs) [21] for event-based camera
  - Standard Neural Networks [5], [11], [4], [3] for event-based camera

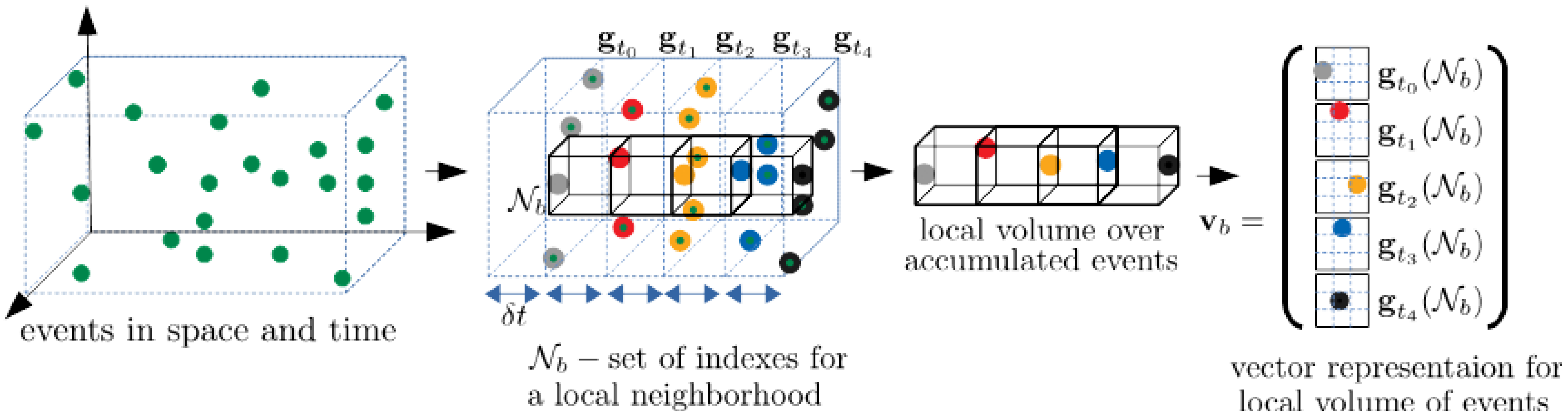
- Applications
  - Gesture recognition [22],
  - Object recognition [27], [4],
  - Face detection [7],
  - Optical flow prediction [37], [34] and
  - Image reconstruction [26]

Motivation

- Data-adaptive, learned single-layer architectures for event-based data not studied extensively
- Unknown to which extend a single-layer model could be useful for event-based data and
- How the spatial and temporal resolution of the event-based data impacts performance for a recognition task

Unsupervised Feature Learning

- Local volume of events extraction

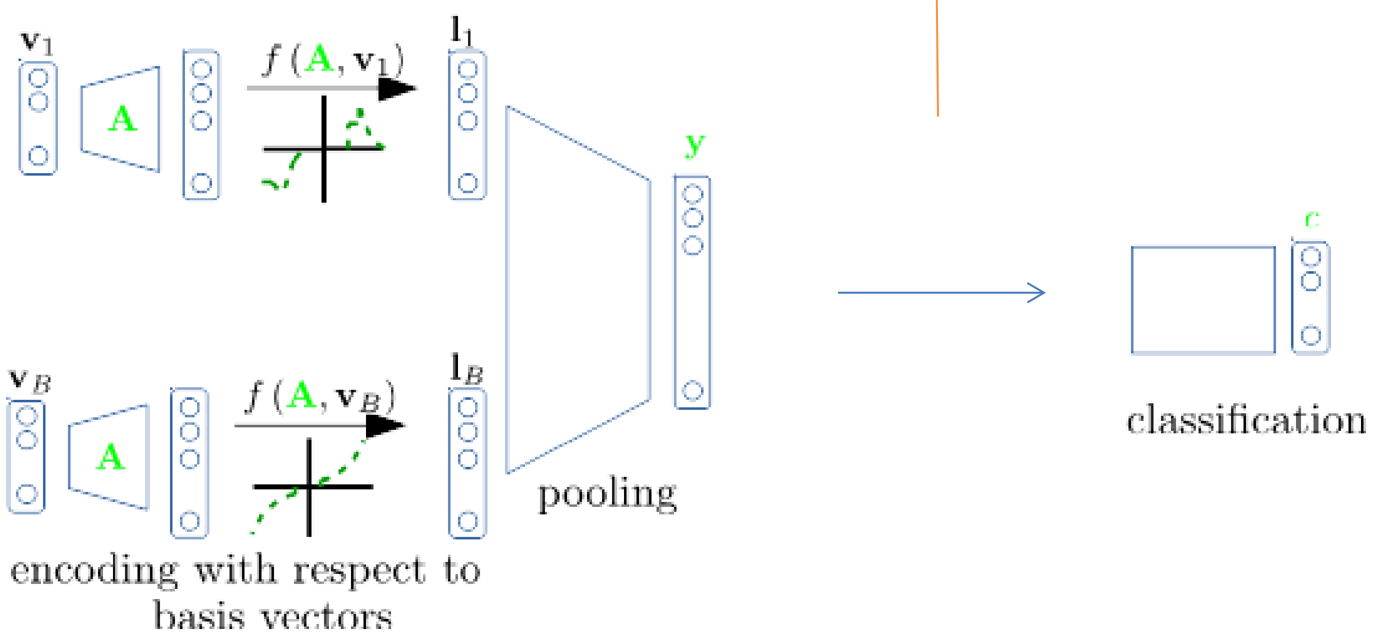


Unsupervised Feature Learning

- Two part recognition pipeline

Part one: unsupervised feature learning + pooling
(learning a mapping with basis vectors as parameters)
-Direct problem formulation
-Inverse problem formulation

Part two: Classifier learning
-Linear SVM



Numerical Evaluation

- Data sets
  - N-MNIST (10 digits, around 1000 samples per digit)
  - N-CARS (foreground/background, around 10000 per class)
  - N-Caltech101 (101 objects, dis-balanced number of samples per object)
- Learned unsupervised feature mapping as proposed using
  - Direct and
  - Inverse problem formulation
- Classifier
  - Linear SVM
- Used quantitative measures
  - Acc. (measured as average of the classifier prediction accuracy over the test set)

Numerical Evaluation

- Accuracy on all of the data sets

Data set	Acc. % (inverse)	Acc. % (direct)
N-MNIST	98.1	96.8
N-Calteh101	78.4	77.1
N-CARS	84.7	81.3

- Comparison with state-of-the-art on the Caltech101 dataset

Method	Acc. %
Hfist [14]	06.0
HOTS [2]	21.0
Garbor-SNN [41]	19.2
HATS [4]	64.2
DART [30]	70.3

Method	Acc. %
Proposed (inverse)	**78.4**
Proposed (direct)	**77.1**

single-layer architectures

Method	Acc. %
EST [16]	81.7
VID2E [10]	90.1

multi-layer architectures

Numerical Evaluation

- Accuracy under varying: number of basis vectors, size of the local volume and number of accumulation intervals

Method	Number of Basis Vectors			
	1000	1500	1700	2000
Proposed (inverse)	73.2	74.5	78.4	76.0
Proposed (direct)	74.3	77.0	77.1	75.5

Method	Size of the Local Volume			
	4×4×4	4×12×12	4×16×16	4×21× 21
Proposed (inverse)	69.6	78.4	76.4	75.2
Proposed (direct)	64.8	77.1	74.5	75.1

Method	Number of Accumulation Intervals			
	2	4	7	10
Proposed (inverse)	61.7	72.4	78.4	76.3
Proposed (direct)	63.2	69.1	77.1	74.1

References

[28] Jonathan Binas Matthew Cook Shih-Chii Liu Peter U. Diehl, Daniel Neiland Michael Pfeiffer. Fast-classifying, high accuracy spiking deepnetworks through weight and threshold balancing.In Int. Joint Conf.Neural Netw. (IJCNN), 4:933–2940, 2015.

[13] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedes-trian detection: An evaluation of the state of the art.IEEE Trans. PaternAnal. Mach. Intell., pages 743–761, 2012.

[39] Margarita Chli Stefan Leutenegger and Roland Y. Siegwart. Brisk: Binaryrobust invariant scalable keypoints.In Int. Conf. Comput. Vis. (ICCV),page 2548–2555, Nov. 2011. 2.

[24] David G. Lowe. Distinctive image features from scaleinvariant keypoints.page 91–110, 2004.

[38] Josef Sivic and Andrew Zisserman. Efficient visual search of videos castas text retrieval.IEEE Trans. Pattern Anal. Mach. Intell., page 591–606,Apr. 2009.

[43] Paul Viola and Michael J. Jones. Robust real-time face detection.Int. J.Comput. Vis., page 57(2):137–154, May 2004.

[22] Paul K. J. Park Michael Pfeiffer Chang-Woo Shin-Hyunsurk Ryu Jun-haeng Lee, Tobi Delbruck and Byung Chang Kang. Live demonstration:Gesture-based remote control using stereo pair of dynamic vision sensors.In IEEE Int. Symp. Circuits Syst. (ISCAS), 2012.

[27] Jun Haeng Lee Byungkon Kang-Chang-Woo Shin Jooyeon Woo Jun-SeokKim Yunjae Suh Sungho Kim Saber Moradi Ogan Gurel Paul K. J. Park,Kyoobin Lee and Hyunsurk Ryu. Computationally efficient, real-timemotion recognition based on bio-inspired visual and cognitive processing.In IEEE International Conference on Image Processing (ICIP). IEEE,Sept. 2015.

[4] Nicolas Bourdis Xavier Lagorce Amos Sironi, Manuele Brambilla andRyad Benosman. Hats: Histograms of averaged time surfaces for robust event-based object classification. In IEEE Conf. Comput. Vis. PatternRecog. (CVPR), page 1731–1740, 2018.

[7] Souptik Barua, Yoshitaka Miyatani, and Ashok Veeraraghavan. Direct face detection and video reconstruction from event cameras. In IEEE Winter Conf. Appl. Comput. Vis. (WACV), pages 1–9, 2016.

[34] Ryad Benosman Sio-Hoi Ieng Paul Rogister and Christoph Posch.Asynchronous event-based hebbian epipolar geometry.IEEE Trans.Neural Netw., page 22(11):1723–1734, 2011.

[26] Andrew J. Davison Patrick Bardow and Stefan Leutenegger. Simultaneousoptical flow and intensity estimation from an event camera.In IEEEConf. Comput. Vis. Pattern Recog. (CVPR), page 884–892, 2016.

[21] Tobi Delbruck Jun Haeng Lee and Michael Pfeiffer. Training deepspiking neural networks using backpropagation.Front. Neurosci., page10:508, 2016.

[15] Ralph Etienne-Cummings Christoph Posch Nitish Thakor Garrick Or-chard, Cedric Meyer and Ryad Benosman. Hfirst: A temporal approachto object recognition.IEEE Trans. Pattern Anal. Mach. Intell., page2028–2040, 2015.

[20] Carmen Serrano Bego Na Acha Teresa Serrano-GotarredonaShouchun Chen Jose A. Perez-Carrasco, Bo Zhao and BernabeLinares-Barranco. Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidenceprocessing–application to feedforward convnets.IEEE Trans. PatternAnal. Mach. Intell., page 2706–2719, Nov. 2013.

[6]David Berg Timothy Melano Jeffrey McKinstry Carmelo Di Nolfo-Tapan Nayak Alexander Andreopoulos-Guillaume Garreau Marcela MendozaJeff Kusnitz Michael Debole Steve Esser Tobi Delbruck Myron FlicknerArnon Amir, Brian Taba and Dharmendra Modha. A low power, fully event-based gesture recognition system. In IEEE Conf. Comput. Vis.Pattern Recog. (CVPR), page 7388–7397, 2017.

[37] Charles Clercq Chiara Bartolozzi Ryad Benosman, Sio-Hoi Ieng andMandyam Srinivasan. Asynchronous frameless event-based optical flow.Neural Netw., page 27:32–37, 2012.

[5]Guillermo Gallego Narciso Garcıa Ana I. Maqueda, Antonio Loquercio and Davide Scaramuzza. Event-based vision meets deep learning on steering prediction for self-driving cars.In IEEE Conf. Comput. Vis.Pattern Recog. (CVPR), page 5419–5427, 2018.

[11] Michael Pfeiffer Daniel Neil and Shih-Chii Liu. Phased lstm: Acceleratingrecurrent network training for long or event-based sequences.In Conf.Neural Inf. Process. Syst. (NIPS), page 3882–3890, 2016.

[3] Kenneth Chaney Alex Zihao Zhu, Liangzhe Yuan and Kostas Daniilidis Ev-flownet: Self-supervised optical flow estimation for event-based cameras. In Robotics: Science and Systems (RSS), 2018.