Niklas Huhs*, Jens Kraitl, Christoph Hornberger, Olaf Simanski

Event based acceleration measurement and fall detection

https://doi.org/10.1515/cdbme-2025-0221

Abstract: Fall detection is essential for elderly and disabled individuals, as undetected falls can be life-threatening. Traditional methods use acceleration sensors and neural networks, but body-worn sensors can be intrusive. This study explores neuromorphic cameras for fall detection with minimal data processing and lightweight neural networks. Denoising techniques were applied to event data, followed by statistical analysis to estimate position, velocity, and acceleration. This method produced patterns similar to accelerometers. Neural network architectures were evaluated, from simple one-dimensional convolutional networks to hybrid models combining convolutional layers with Long-Short-Term-Memory units. Training data were generated by converting video-based fall datasets (le2i, MCFD) into event data using the v2e-toolbox. Data augmentation resulted in 2,610 samples (1,314 falls, 1,296 daily activities). The best model, a three-layer 1D convolution combined with a two-layer LSTM (hidden size 64, 125k trainable parameters), achieved 97% accuracy. Live inference on streamed videos and a DVXplorer event camera was possible without noticeable delay. Our approach matches state-of-the-art acceleration sensor methods while offering a non-intrusive, real-time monitoring solution, potentially improving response times and user comfort.

Keywords: Neuromorphic vision, machine vision, machine learning, ambient assisted living, fall detection

1 Introduction

Acceleration sensors have been used for fall detection for a long time and are able to produce highly accurate results, if processed by a capable neural network. These sensors are usually worn on the body of the subject, making them invasive and particularly problematic if the system relies on the person to remember wearing them every day. Especially in assisted living environments, where patients are highly vulnerable to falls and may also experience cognitive impairments.

One approach to solve this problem is to use non-invasive techniques like cameras paired with video processing or other contactless sensors like infrared sensors, radar or a combination of those. The camera based approaches make use of normal RGB-cameras [1] or more advanced cameras like time-of-flight 3D cameras [2]. What most of them have in common is the use of multi-dimensional convolutional and recurrent neural networks which are significantly larger and more complex than the neural networks used to detect falls in one-dimensional time-based accelerometer-data.

Another novel camera technology, neuromorphic vision, offers new ways to approach the problem of fall detection. Neuromorphic cameras do not capture frames like a standard camera but report changes in illumination for every individual pixel. This way, data are only transmitted when events occur in the frame, also giving them the name event-cameras.

In this work, event cameras were used to extract the movement and acceleration of a body, before a simple neural network was used to detect fall events in this data

2 Methods

To measure the acceleration of a person or body moving through the view of a neuromorphic camera, the object has to be detected first. Due to the nature of neuromorphic cameras, events only occur on the pixels in the frame, where movement happens. This can be leveraged to enable very lightweight object detection. The events that are transmitted by the event camera are pooled in time slices, a common technique for handling the asynchronous events of neuromorphic cameras. Every ten milliseconds, the last batch of events is transmitted and processed. The mean value of the x- and y-coordinates of all activated events and the immediate area around them in each time slice is monitored, while also using the nearest 10 time slices to recognize if the ROI is moving in a plausible manner. This way, the movement of an object is tracked in the spatiotemporal domain.

***Niklas Huhs:** Hochschule Wismar, Philipp-Müller-Str. 14, Wismar, Germany, e-mail: niklas.huhs@hs-wismar.de
Jens Kraitl, Christoph Hornberger, Olaf Simanski: Hochschule Wismar, Philipp-Müller-Str. 14, Wismar, Germany

The centre point of this ROI was tracked and through differentiation, the velocity and the acceleration of the body was calculated. Before the differentiation, a moving average filter is applied to smooth out the noise caused by non-uniform distribution of the events. The general data processing is visualized in figure 1.

Figure 1: Visualization of data flow.

Because the relative amount of noise in the calculated acceleration increases when the number of activated events decreases, an activation function has been introduced which takes on values between 0 and 1, depending on the number of activated events relative to the total number of possible events:

$$\boldsymbol{\sigma} = \frac{1}{1+e^{-\left(e^{\frac{n}{\varepsilon}}-e^{2}\right)}} \quad (1)$$

Where n represents the number of activated events in a time slice and ε represents a constant based on the frame size of the neuromorphic camera. ε can be tuned according to the size of the room and the expected number of events caused by a human moving through this room. By using this activation function, actual human activity gets highlighted, while inactivity in the frame gets represented less. An example of this activation function over a 30 second video is seen in figure 2.

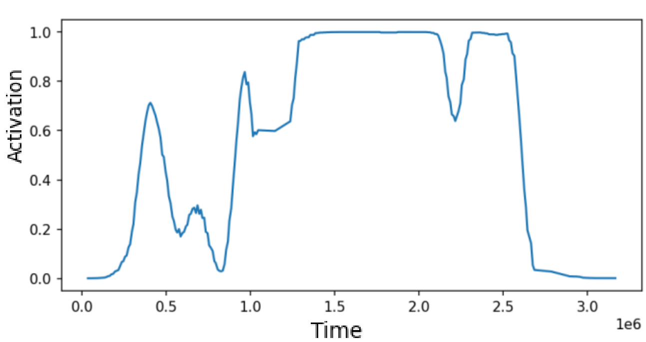

Figure 2: Activation over time in a video sample.

2.1 Dataset

Due to the limited availability of event-based fall-detection datasets, the v2e-toolbox [3] has been used to convert the video-based datasets le2i [4] and Multiple cameras fall dataset (MCFD) [5] into event-based data. V2e converts low-framerate videos into low-latency event-based data by first interpolating frames into the original video and then generating events for each pixel, also introducing noise, usually found in event-cameras.

The le2i dataset consists of videos containing fall events and videos containing no fall events and a corresponding csv-file for every video, containing frame numbers for the start and end of each fall event. Using these frame numbers and the v2e-toolbox, the videos are split and converted into four-second-long aedat4-event-files containing either a fall event or a non-fall-event, in this case named as an activity of daily living (ADL). The MCFD dataset provided similar labelling of the fall events and non-fall events.

Data augmentation was achieved by adding different levels of noise and different filters to the event-data, utilizing event polarity filters and background noise activity filters. This way a dataset of 2,610 samples, consisting of 1,314 falls and 1,296 ADL, has been created.

2.2 Training

Different neural network architectures were tested to achieve the necessary accuracy while being reasonably sized and able to be run on minimal hardware.

Common architectures for training on time series data are one-dimensional convolutional neural networks which are able to learn local patterns in time series data and recurrent neural networks which introduce a form of memory into neural networks by utilizing recurring connections between layers. A special kind of recurrent neural networks are LSTM (Long Short Term Memory)-Networks which are better at learning long-term dependencies compared to traditional RNNs. Convolutional neural networks and LSTM-networks can also be combined to improve performance for time series classification. The CNN might learn local patterns in the data while the LSTM learns the long term dependencies of these patterns.

The data was split into training (60%), validation (20%), and test (20%) sets to ensure balanced model evaluation. A batch size of 32 was used for all architectures, and training was conducted over 200 epochs on an Nvidia RTX 3080 Ti GPU. Each network employed a ReLU activation function with an AdamW optimizer, and a learning rate scheduler dynamically adjusted the rate from an initial 10^{-4} to a minimum 10^{-9} when validation loss plateaued.

2.3 Testing

To test the network in lifelike conditions, an uncut event-video sample from the dataset has been chosen, which was not included in the training data. The file in aedat4-format is read in real-time and fed to the neural network. Figure 3 shows the visualization of the event stream where every coloured pixel represents an event. When a fall is detected, a fall event is triggered and the acceleration of the body is displayed as in figure 4, showing rapid acceleration and deceleration.

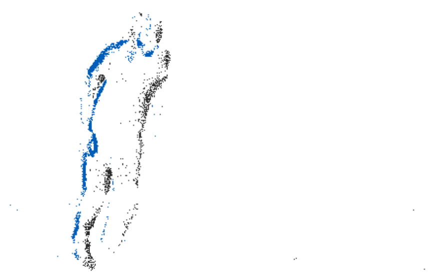

Figure 3: Fall event visualized.

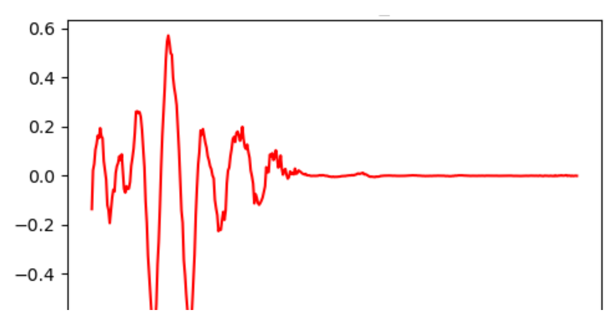

Figure 4: Estimated acceleration of fall.

Furthermore, a DVXplorer camera was set up in a working environment for testing and data acquisition. Multiple samples with a length of 2 minutes were acquired which consisted of a person performing activities of daily living like working on a desk, walking, sitting down and standing up. While false negatives after inference were rare, false positives due to sudden movement of the person were present. The constant ε (see eq. 1) was tuned according to the new environment for optimal results.

3 Results and discussion

The extraction of a bodies acceleration by utilizing the event based nature of neuromorphic cameras shows the potential to enable a new method for contactless fall detection and in this state is comparable to state of the art techniques based on accelerometer readings, where accuracies between 97 % and 99.5 % were achieved. [6] The proposed method reaches an accuracy of 97 % and through more complex calculation of the estimated position of the person, these results can be improved as the largest factor of noise and faulty readings remains in the non-uniform distribution of events over the position of the body, which causes the average calculated position to fluctuate, mostly in the y-direction.

The tracking of an object using nothing but the x- and y-coordinates of activated events nonetheless proves to be effective for estimating the movement of a single object in the frame of a neuromorphic camera. To handle multiple bodies, detection of ROI and the object itself can be achieved by utilizing machine learning models for object detection, which would come at the expense of higher computational cost. Considering this, the proposed method is a valid alternative. Table 1 shows the architectures that were used to classify the custom neuromorphic fall dataset into fall and no-fall events, their number of trainable parameters and the achieved accuracy. At a number of 125,000 trainable parameters and an accuracy of 97 %, the network consisting of a one-dimensional convolutional network followed by an LSTM was the most successful. Real-time inference on live event data is made possible by an inference time of less than 50 milliseconds, enabling constant supervision of a patient with limited hardware resources and power consumption.

Table 1: Comparison of tested models

Architecture	Trainable parameters	Achieved accuracy
1D-CNN	175 k	93 %
LSTM + CNN	88.3 k	94 %
CNN + LSTM	125 k	97 %

Figure 5 shows the validation loss over 200 epochs on the CNN + LSTM model, where the effect of the learning rate scheduler is seen on the decreasing fluctuations. The simple one-dimensional CNN and the LSTM + CNN model perform worse on the same data with the training also being unstable due to the network only learning local patterns due to missing understanding of long term dependencies, causing overfitting after a sufficient amount of epochs.

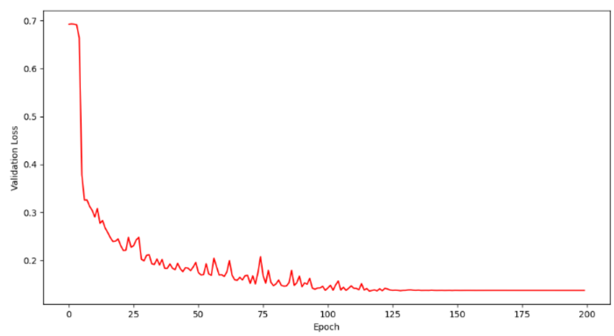

Figure 5: Validation loss over 200 epochs on CNN+LSTM model

In addition to these results the use of neuromorphic cameras for camera based fall detection also offers the advantage of increased privacy, due to the face of the person, as well as the surroundings and living space not being visible.

Figure 6 Shows an event from a video acquired in the testing enviromnent, mentioned in 2.3, where the subject simulates a fall, followed by a lack of movement.

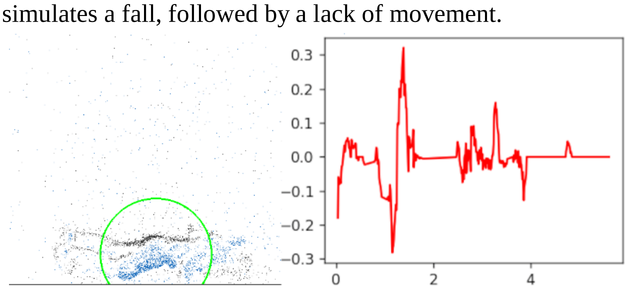

Figure 6: Fall event in testing environment

Most activities of daily living were correctly classified as non-fall events and smaller objects like a set of keys falling to the ground were not picked up as relevant objects. Classification difficulties arose when a person made abrupt movements, such as running and coming to a sudden stop.

4 Conclusion

In this work a novel approach for tracking humans and estimating their acceleration with neuromorphic cameras has been proposed and proven to be effective in controlled environments. The goal of this study, being minimizing the computational costs and therefore the necessary energy consumption of a fall detection system by reducing the classification model to a one dimensional convolutional neural network paired with a recurrent LSTM network, has been achieved and real-time inference on less powerful hardware has been achieved compared to more complex three-dimensional convolutional neural networks, which usually require more powerful hardware or simply take too much time for inference.

To improve results, more work can be done to accurately calculate the position of the body in the frame and therefore increase accuracy of the estimated acceleration and the dataset will be expanded with real world samples, taken in a working environment,

Author Statement
Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] De Miguel K, Brunete A, Hernando M, Gambao E. Home Camera-Based Fall Detection System for the Elderly. Sensors. 2017; 17(12):2864. https://doi.org/10.3390/s17122864
[2] Xu Y, Chen J, Yang Q, Guo Q. Human posture recognition and fall detection using kinect V2 camera. In: 2019 Chinese Control Conference (CCC). IEEE, pp 8488–8493
[3] Hu Y, Shihi-Chii L, Delbruck T. v2e: From Video Frames to Realistic DVS Events 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) arXiv:2006.07722 .
[4] Charfi I, MiteranJ , Dubois J, Atri M, Tourki R. Definition and performance evaluation of a robust SVM based fall detection solution, 8th International Conference on Signal ImageTechnology and Internet Based Systems, SITIS 2012r, pp. 218–224, 2012, doi:10.1109/SITIS.2012.155.
[5] Auvinet E, Rougier C, Meunier J, St-Arnaud A, Rousseau J. Multiple cameras fall data set, 2011.
[6] Singh A, Rehman S, Yongchareon S, Chong P. Sensor Technologies for Fall Detection Systems: A Review. IEEE Sensors Journal, vol. 20, no. 13, pp. 6889-6919, 1 July1, 2020, doi: 10.1109/JSEN.2020.2976554.