Chapter 8
Digital Image Correlation with a Neuromorphic Event-Based Imager

Peter Meyerhofer, Andre Green, Alessandro Cattaneo, and David Mascareñas

Abstract Digital image correlation is a well-established method for estimating the full-field displacement of two-dimensional surfaces by comparing pairs of images. The basic idea is to attach a texture or speckle pattern to the surface, and track the features of this pattern through the grayscale values of the image pixels. All this assumes a conventional camera that records image intensity at every pixel in every frame, which is inefficient for sharp transient events. In watching a balloon burst, for example, one needs a high frame rate to capture the details of the burst, but storing such a frame rate for the longer, slower expansion of the balloon before it bursts costs a lot of memory. It would be convenient, in applications such as monitoring structures, to store dense information only during dynamic events, not between them.

Silicon retinas are an alternative sensor type that record events rather than pixels. The raw data for such a device are a sequence of 4-tuples: the time at which each event occurred, the horizontal and vertical pixel coordinates containing the event, and whether the event involved increasing or decreasing intensity. Events can be reassembled into buckets corresponding to each pixel and a time interval corresponding to any desired camera shutter to produce frames analogous to those of a conventional camera. This project assesses whether it is feasible to combine digital image correlation or a similarly developed computer vision technique with the efficient storage of a silicon retina, via such converted frames.

The test article for this experiment was a latex band with a painted speckled pattern, mounted into the stationary and moving ends of a frame and subject to cyclical stretching. The image of this band was recorded simultaneously on conventional and silicon retina detectors through a beam splitter. Analysis of the resulting data showed that the silicon retina frames allow feature tracking with close to the quality of a conventional camera, but the computed displacements are consistently smaller. The surface has to accelerate before there are enough changes to register on the silicon retina, and this initial motion at either end of the oscillation is not recorded in the converted frames. This work demonstrates that silicon retina imagers have potential for persistent surveillance applications where there is a need to record sparsely occurring transient deformations over long time periods.

Keywords Digital image correlation · Neuromorphic · Event-based · Dynamic

8.1 Introduction

One developing alternative to frame-based digital cameras is event-based or neuromorphic cameras, where the data is a list of brightness changes recorded by pixel location, time stamp, and polarity (whether the brightness increased or decreased). Initial development occurred in the 1980s, inspired by the neuron spikes that occur in a biological retina [1, 2]. For a recent review, see [3].

A major advantage of event-based imaging is that little information is stored when the subject is not changing, which saves memory compared to a conventional camera which records every pixel in every frame. This has motivated the application of event-based imagers to structural vibration modes, where damage events occur only sparsely across the structure and in time [4]. Similar savings apply to the computation time to process events, enabling robots to respond to high-speed motion [5] or to perform localization and mapping more robustly [6, 7]. The high speed and high dynamic range of event-based imagers can be used to synthesize fast, high-contrast video from an initial frame [8], or to fill in detailed feature tracking for the time between frames [9].

P. Meyerhofer (✉) · A. Green · A. Cattaneo · D. Mascareñas
Los Alamos National Lab, Los Alamos, NM, USA
e-mail: andre_green@lanl.gov; cattaneo@lanl.gov; dmascarenas@lanl.gov

A notable feature of event-based cameras is that the time stamps can exhibit latencies in the range of 200–1 µs. The low latency of silicon retina imagers is in contrast to a conventional imager where data takes the form of entire frames with a specified frame rate. In an event-based imager, events at different pixels can be processed as they occur. One can also focus on grouping the events into frames and applying established computer vision methods. Previous applications of frame-based processing of event data include stereo vision [10], optical flow [11], and autonomous vehicles [12]. Here we use a formulation in which corner features are identified in the first frame and tracked between frames using frame-based optical flow [13].

8.2 Latex Band Test

We arranged a speckle pattern with speckles of diameter 0.05, 0.1, or 0.2 inches on a latex band gripped by two ends of a translation stage (Fig. 8.1). The right end remained stationary, while the left end oscillated to vary the strain in the latex. A beamsplitter placed 49.5 cm from the plane of the band received the light and diverted it to the pair of image sensors: a FLIR GS3-U3-23S6M (the frame-based camera) and an iniVation DVXPlorer (the event-based camera). The operating software sent commands for both sensors to start recording slightly before the fixture on the left side began to move.

Figure 8.2 shows sample frames from the conventional and silicon retina sensors. Part (a) is a conventional frame (1920 × 1080), where the left and right boundaries are dark because the 25-mm lens did not fit inside the beamsplitter. Part (b) shows a sample silicon retina frame (640 × 480), converted from raw event data by the following process. First, events were sorted by the pixel at which they occurred, and we kept only those events which occurred at pixels with between 200 and 750 events (in about 10 s of recording). Higher event counts indicated a "hot" pixel which often recorded an event regardless of light input, and lower event counts indicated an area where little happened. Since the frame rate throughout this test was 30 fps, the remaining events were grouped into time intervals of 1/30 s, and their counts at each pixel in each interval were scaled to an 8-bit gray value. Finally, the image was inverted so that large event counts were indicated in black and small event counts in white.

Note that the silicon retina frames contain ovals that emphasize the right and left edges, since the speckles created change events by moving in that direction, so that events were less frequent at the top and bottom of the speckles. We included increasing and decreasing events on equal terms, since if we only used one polarity (say, increasing intensity), the converted frame would show only the right edge alternating with only the left edge, potentially confusing the tracking algorithm.

Since a silicon retina records information only when the subject of the video is moving, periods when the band was stationary (at either end of the oscillation) mostly show noise. Therefore, we computed the variance of pixel values in each converted frame and retained for digital image correlation (DIC) only the 40% of frames with the highest variance. 40% was a fraction chosen to be sure that we excluded frames not showing the speckle pattern; a few usable frames were excluded with them.
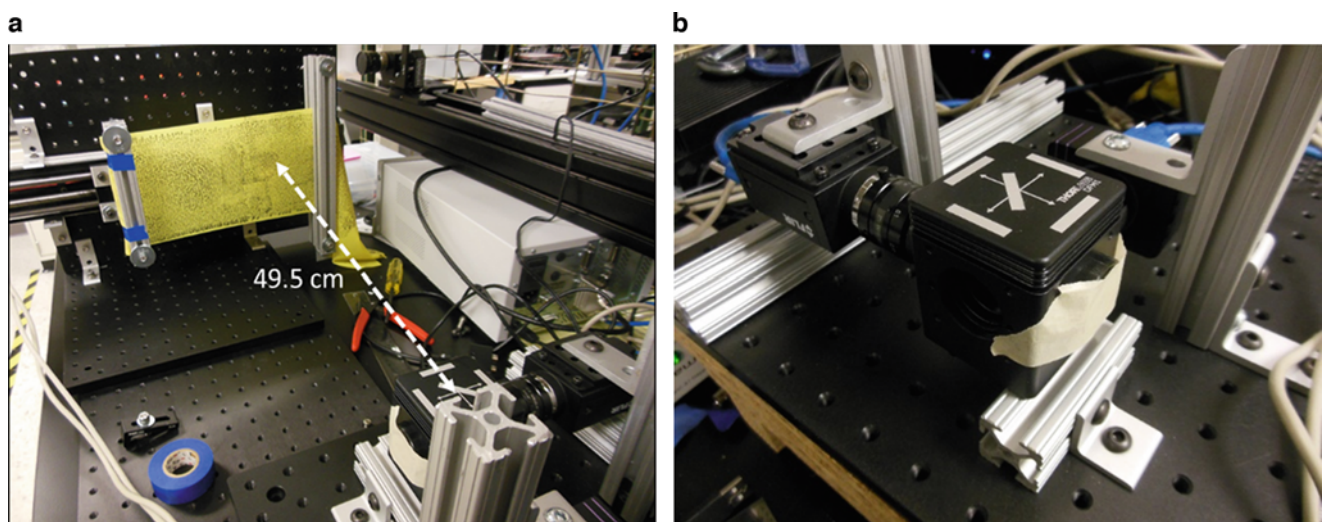

Fig. 8.1 Pictures of the experimental setup: the frame holding the latex band (**a**) and the pair of image sensors (**b**)
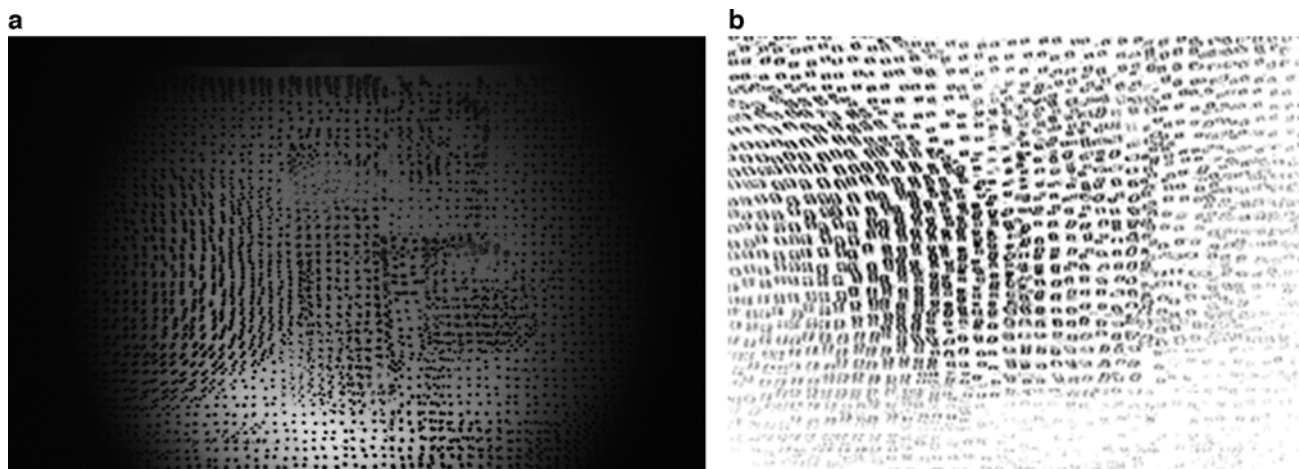
Fig. 8.2 Sample frames from the conventional camera (**a**) and converted from the silicon retina data (**b**)

To begin the tracking algorithm, we used OpenCV to select up to 400 corner features in a rectangular region of interest in the first frame and followed them throughout the frame sequence with Lucas-Kanade optical flow. Any features that failed to track (there were few to none of these in each sequence) were excluded from later steps. With this set of features spanning all frames, we used SciPy to form a Delaunay triangulation based on their locations in the first frame (Fig. 8.3), and filled an array with the horizontal and vertical displacements (with respect to the first frame) of every feature in every frame.

Finally, we estimated the horizontal and vertical displacements at every pixel in the region of interest using barycentric interpolation on the triangulation (Fig. 8.4). Applied to each frame, this method produced a video of the horizontal and vertical displacements of the latex band. Sample displacement maps are shown for the conventional frames in Fig. 8.5 and for the converted silicon retina frames in Fig. 8.6. Pixel displacements were converted to metric displacements by a scale factor: in the conventional frames, a 0.1-inch speckle spans about 20 pixels, and the ratio of speckle size in pixels is about 10 (silicon retina) to 7 (conventional). Frame 36 in the conventional camera corresponds to an extremum in the speckle motion, and Frame 50 in the silicon retina data is the first usable time after the band begins to return from this extremum. Pixels outside the triangulation are plotted as white and disregarded.

The silicon retina results are slightly noisier than the conventional results and notably smaller in magnitude. Both differences are probably increased by the inability of a silicon retina to observe stationary objects. Tracking only begins after the speckles have accelerated to a sufficient speed, so that the earliest displacements are not recorded and the total displacements are less than for the conventional camera.

8.3   Conclusion

Deformation estimates using DIC over event-based frames were compared to estimates made using DIC over traditional imager data. Although noisier, the estimates from the silicon retina are of sufficient resolution and accuracy to be qualitatively comparable, and the process for obtaining said estimates poses several advantages over the traditional method.

Insofar as event-formed frames are a post-processing function over the events, the ideal exposure time and flutter pattern need not be known prior to observing the structure: for real-world structures whose dynamics change with age, use, and in response to varying environmental conditions, this is particularly useful. Although a similarly adaptable data stream could be obtained through the use of high-speed imagers, the redundancy of information captured and consequent energy, memory, and processing requirements to obtain and form equivalent frames is anticipated to be significantly higher.

In the future, a hybrid approach wherein a data-fusion filter (e.g., Kalman or complementary) combines lower-temporal resolution event-frame cluster position estimates with higher-temporal resolution event-based cluster motion estimates may further reduce the energy requirements for low-latency and high temporal resolution deformation estimation and open up an avenue for real-time stress and strain approximation.
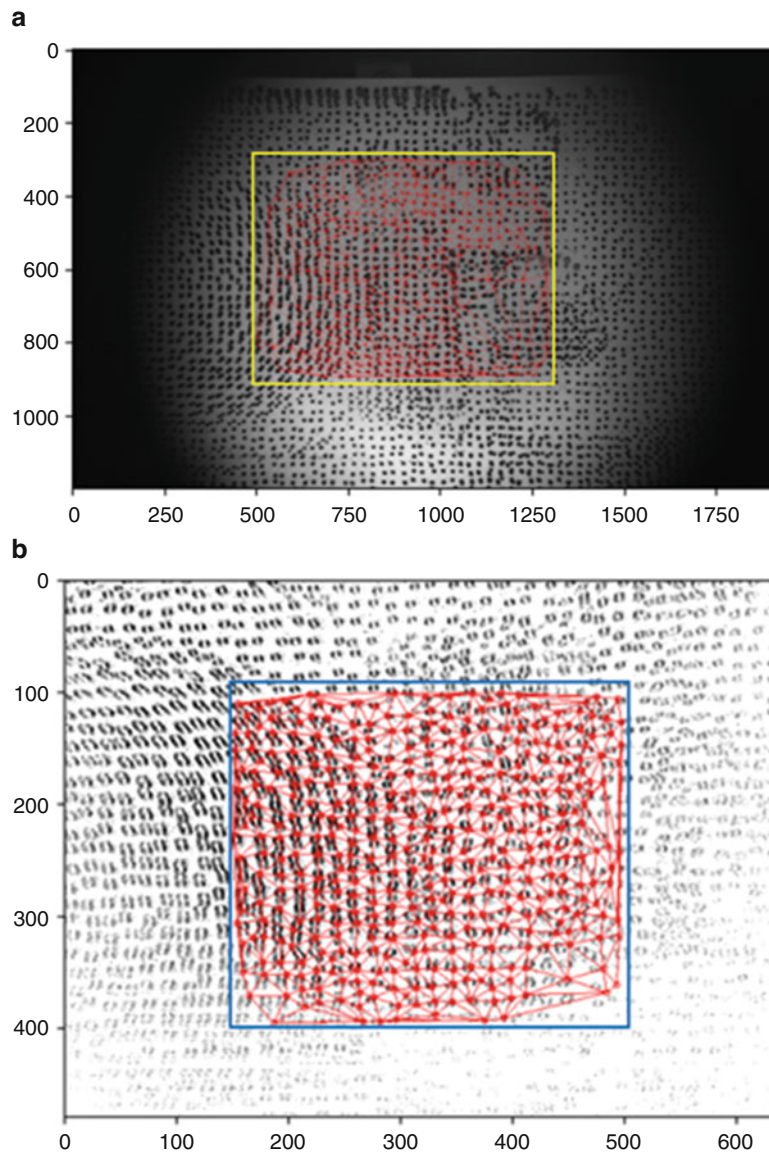
Fig. 8.3 Sample frames with tracked corner features and the corresponding triangulation from the conventional camera (**a**) and converted from the silicon retina data (**b**). The rectangular regions of interest are also drawn
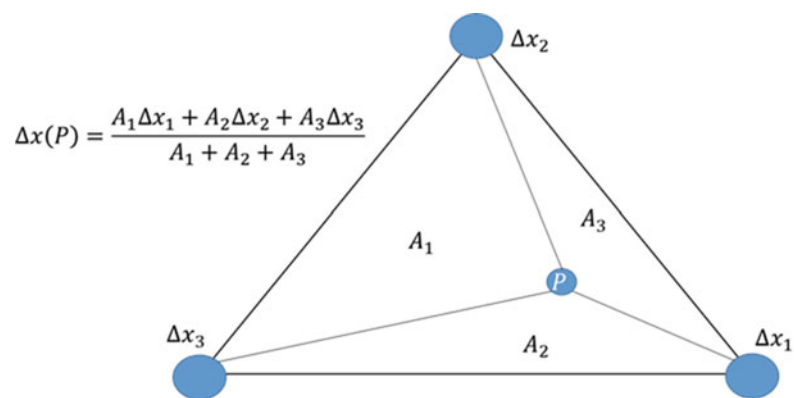

$$\Delta x(P) = \frac{A_1 \Delta x_1 + A_2 \Delta x_2 + A_3 \Delta x_3}{A_1 + A_2 + A_3}$$

Fig. 8.4 Illustration of barycentric interpolation within a triangle
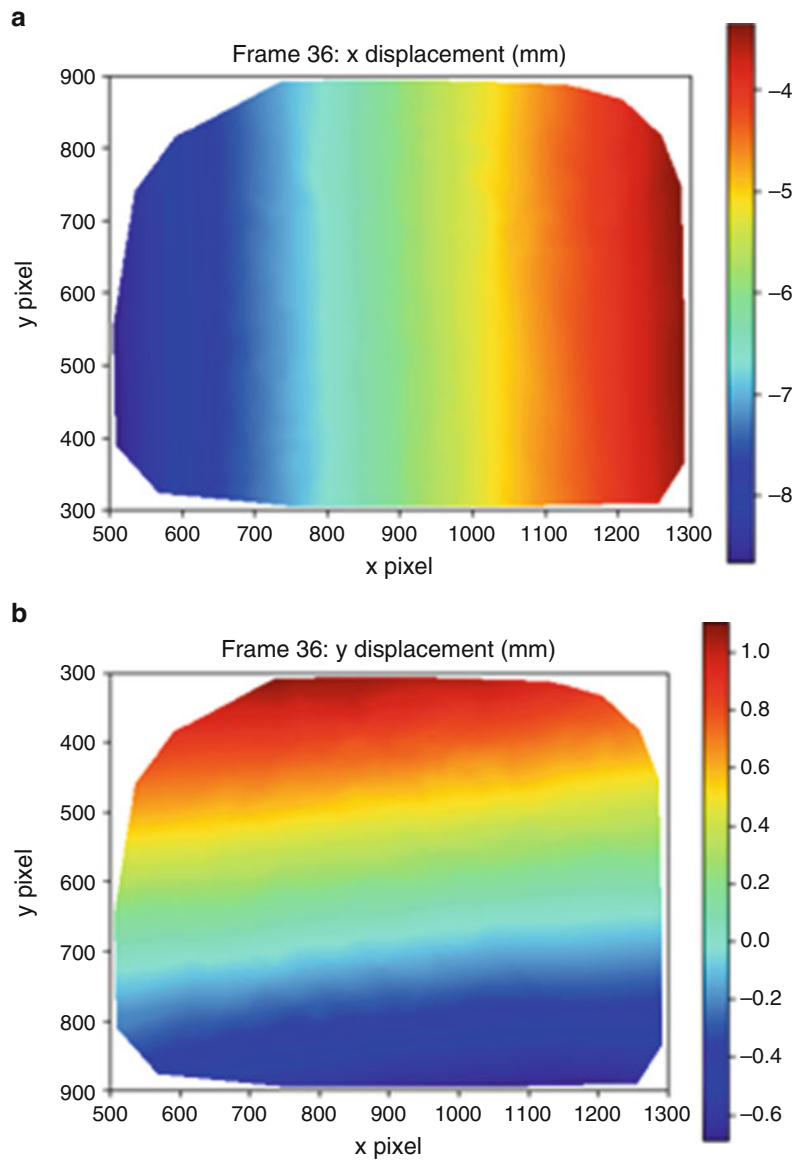
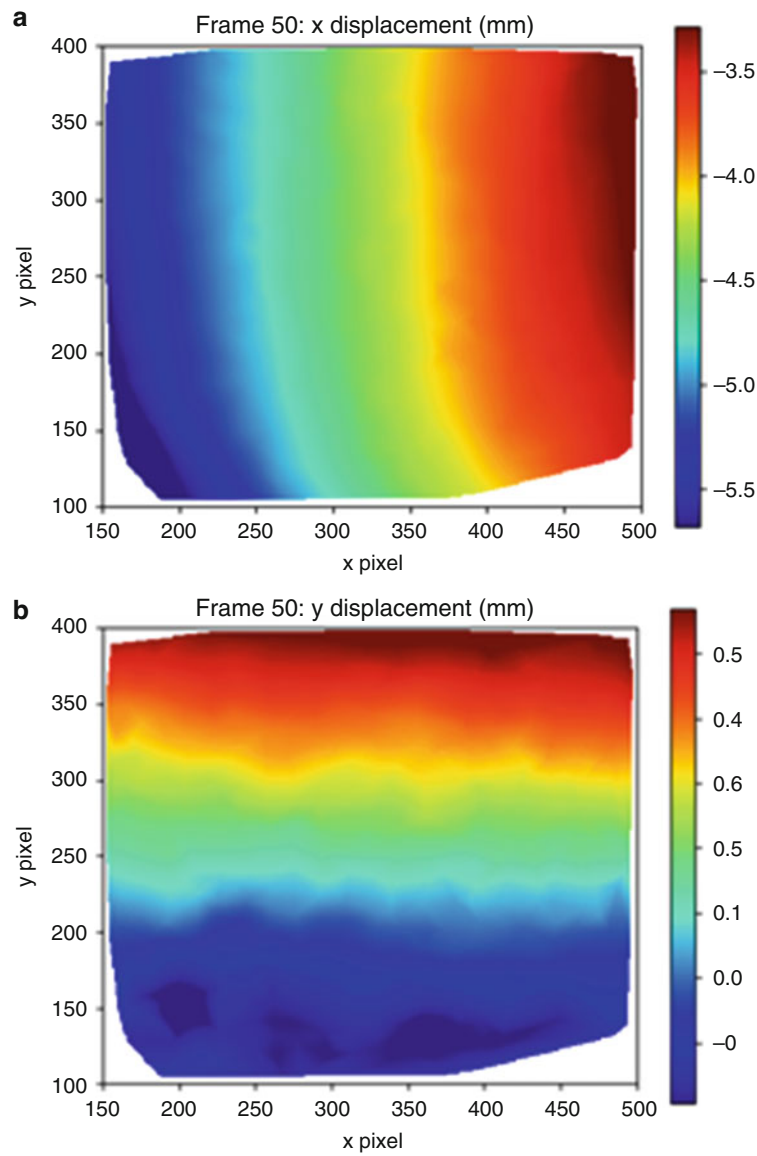**Fig. 8.5** Interpolated horizontal (**a**) and vertical (**b**) displacement for the conventional frames

a

b

Fig. 8.6 Interpolated horizontal (**a**) and vertical (**b**) displacement for the converted silicon retina frames

References

1. Mead, C., Mahowald, M.: A silicon model of early visual processing. Neural Netw. **1**, 91–97 (1988)
2. Mahowald, M., Mead, C.: The silicon retina. Sci. Am., 76–83 (May 1991)
3. Gallego, G., et al.: Event-based vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. (2020)
4. Dom, C., et al.: Efficient full-field vibration measurements and operational modal analysis using neuromorphic event-based imaging. J. Eng. Mech. **144**, 04018054–04011-12 (2018)
5. Delbruck, T., Lang, M.: Robotic goalie with 3 ms reaction time at 4% CPU load using event-based dynamic vision sensor. Front. Neurosci. **7**, 223–1-7 (2013)
6. Rebecq, H., Horstschaefer, T., Gallego, G., Scaramuzza, D.: EVO: a geometric approach to event-based 6-DOF parallel tracking and mapping in real time. IEEE Robot. Autom. Lett. **2**, 593–600 (2017)

 7. Vidal, A., Rebecq, H., Horstschaefer, T., Scaramuzza, D.: Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios. IEEE Robot. Autom. Lett. **3**, 994–1001 (2018)
 8. Rebecq, H., Ranftl, R., Koltun, V., Scaramuzza, D.: High speed and high dynamic range video with an event camera. IEEE Trans. Pattern Anal. Mach. Intell. **43**, 1964–1980 (2021)
 9. Gehrig, D., Rebecq, H., Gallego, G., Scaramuzza, D.: EKLT: asynchronous photometric feature tracking using events and frames. Int. J. Comput. Vis. **128**, 601–618 (2020)
10. Kogler, J., Sulxbachner, C., Kubinger, W.: Bio-inspired stereo vision system with silicon retina imagers. In: Fritz, M., Schiele, B., Piater, J. (eds.) Lecture Notes in Computer Science, vol. 5815, pp. 174–183. Springer, Berlin (2009)
11. Liu, M., Delbruck, T.: Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors. In: British Machine Vision Conference (2018)
12. Macqueda, A., et al.: Event-based vision meets deep learning on steering prediction for self-driving cars. In: Computer Vision and Pattern Recognition (2018)
13. Barron, J., Fleet, D., Beauchemin, S.: Performance of optical flow techniques. Int. J. Comput. Vis. **12**, 43–77 (1994)