A Unifying Contrast Maximization Framework for Event Cameras, with Applications to Motion, Depth, and Optical Flow Estimation – Supplementary Material –

Guillermo Gallego, Henri Rebecq, Davide Scaramuzza
Dept. of Informatics and Neuroinformatics, University of Zurich and ETH Zurich
http://rpg.ifi.uzh.ch

5. Multimedia Material

A video showing the application of our framework to solve several computer vision problems with event cameras is available at: https://youtu.be/KFMZFhi-9Aw.

6. Optical Flow Estimation

Our framework seeks for the point trajectories on the image plane that best fit the event data and it is able to take into account all the information contained in the events: space-time coordinates and polarity (i.e., sign) of the brightness changes. More specifically, event polarity is incorporated in the framework during the creation of the image patches of warped events (equation (2) in the paper).

Figure 12 compares the elements of our framework (image patches of warped events H and objective function f) for a set of events whose optical flow we want to estimate, in two scenarios:

1. Using polarity ($b_k = p_k$), i.e.,

$$H(\mathbf{x};\theta) = \sum_{k=1}^{N_e} p_k \, \delta(\mathbf{x} - \mathbf{x}'_k(\theta)). \tag{7}$$

2. Not using polarity ($b_k = 1$), i.e.,

$$H(\mathbf{x};\theta) = \sum_{k=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}'_k(\theta)). \tag{8}$$

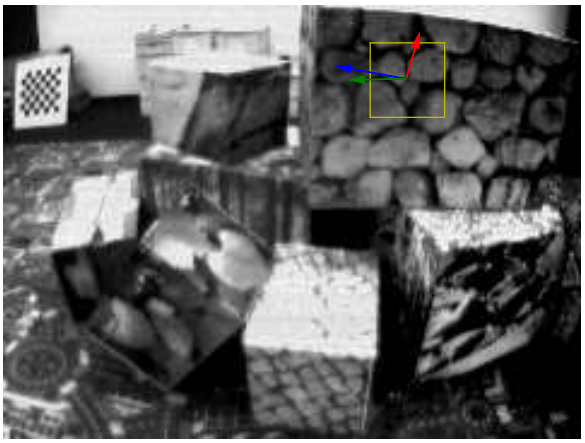
For illustration purposes, the intensity frame in Fig. 12a shows the patch corresponding to the considered events (yellow rectangle). However, such an intensity frame is not used in our framework.

Without using polarity, Figs. 12b and 12c show the contrast function $f(\theta)$ and images of warped events (8) for three candidate point trajectories, specified by the three optical flow vectors $\theta_i, i = 0,1,2$, that are displayed in Fig. 12a. Conversely, Figs. 12d and 12e show the corresponding elements if event polarity is used. The image patches of warped events are color coded from blue (low) to red (high). If polarity is not used (Fig. 12b), blue means absence of events (small values of (8)), whereas red indicates large accumulation of events (large values of (8)). If polarity is used (Fig. 12e), green means absence of events, whereas red and blue indicate large accumulation of positive and negative events, respectively, according to (7).

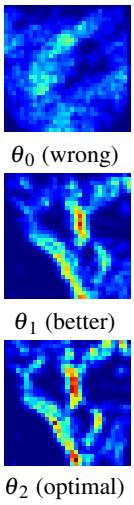
As it can be observed by comparing Figs. 12c and 12d, both objective functions provide approximately the same optimal velocity (peak of the objective function) $\theta \equiv \mathbf{v} \approx (-40, 0)$ pixel/s. However, the basin of attraction of the optimal value is slightly narrower and more pronounced if polarity is used than if it is not taken into account, as can be noted since Figs. 12c and 12d are displayed using the same color range. This can be explained by comparing the image patches of warped events in Figs. 12b and 12e. In case of thin edge structures like the ones in the considered patch, if events are warped so that nearby edges overlap, and therefore their opposite event polarities cancel, then the contrast function $f(\theta)$ greatly decreases (thus reducing the width of the contrast peak, i.e., its basin of attraction). Conversely, if event polarity is not used, the alignment of nearby edges does not produce cancellation, and therefore the contrast decreases more slowly, due to the warped edges becoming further apart.

7. Depth Estimation

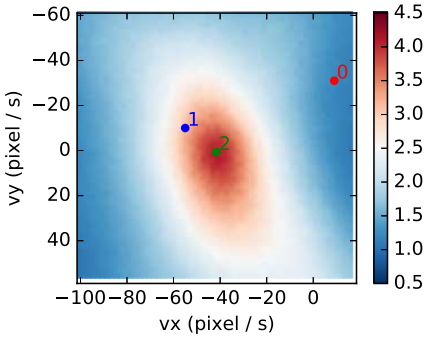
To illustrate how depth estimation improves as more events are processed, we carried out an experiment with the `slider_depth` sequence from the dataset [1]. We reconstructed the scene with the same steps as those used for Fig. 7 in the paper, but varying the number of events processed, N_e, between 20 000 and 1 million. The results are displayed in Fig. 13. As it can be seen, as more events are processed (corresponding to a larger camera baseline), the reconstructed point cloud becomes more accurate and less noisy. This effect is also visible in the semi-dense depth maps overlaid on the grayscale frame of the reference view.



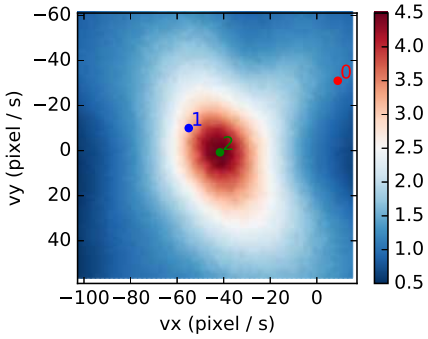
(a) Intensity frame showing the patch corresponding to the considered events (yellow rectangle) and three candidate velocities (optical flow): θ_0 (red arrow), θ_1 (blue arrow) and θ_2 (green arrow).



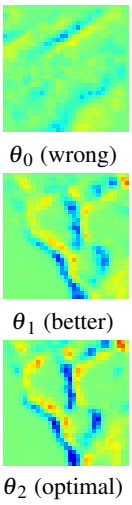
(b) Warped events $H(\mathbf{x};\theta)$, for $\{\theta_i\}_{i=0}^{2}$ in Fig. 12c.



(c) Contrast function $f(\theta)$ for H in (8). For illustration, we show in Fig. 12b the images of warped events (8) corresponding to the three candidate velocities in Fig. 12a.



(d) Contrast function $f(\theta)$ for H in (7). For illustration, we show in Fig. 12e the images of warped events (7) corresponding to the three candidate velocities in Fig. 12a.



(e) Warped events $H(\mathbf{x};\theta)$, for $\{\theta_i\}_{i=0}^{2}$ in Fig. 12d.

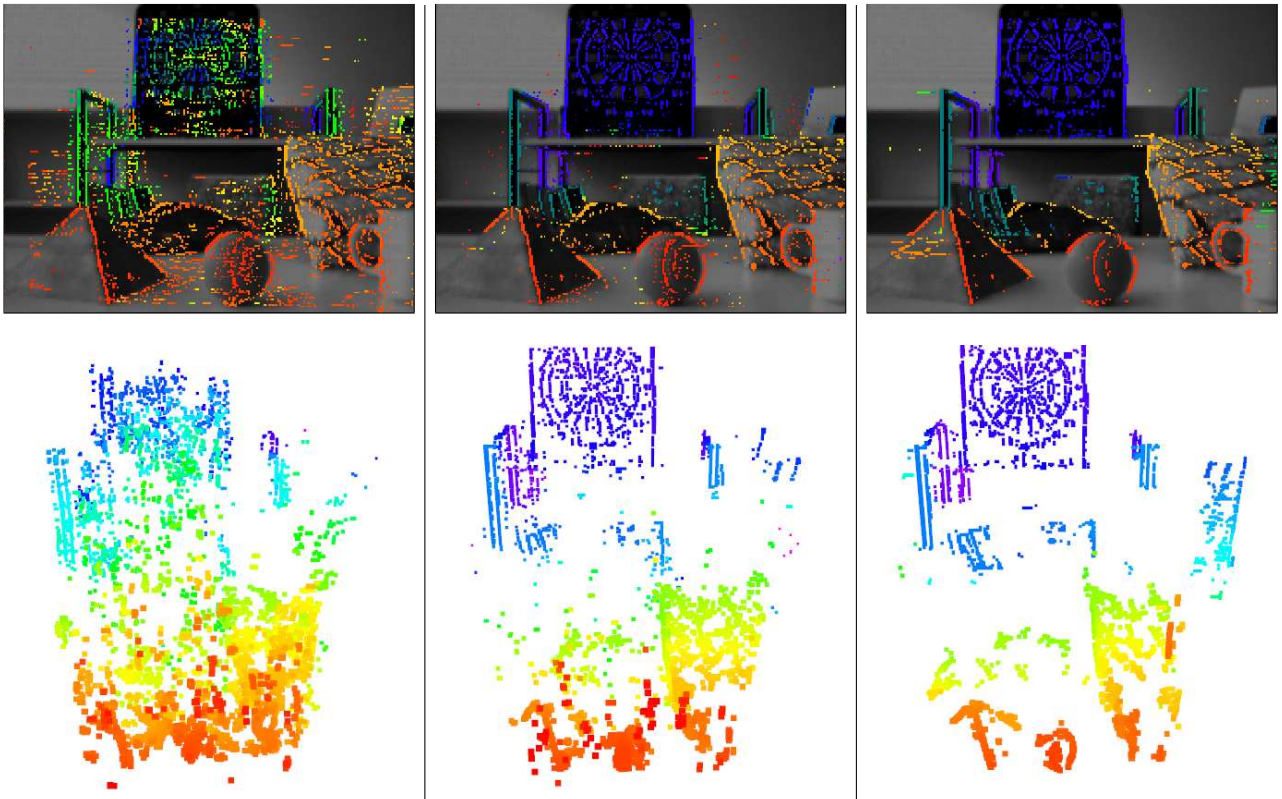
Figure 12: *Optical Flow* (patch-based) estimation. Comparison of objective functions and images of warped events using event polarity (Figs. 12e and 12d) or not using it (Figs. 12b and 12c). In either case, the optical flow is estimated by finding the maximizer of the contrast $f(\theta)$.

8. Rotational Motion Estimation

A comparison between the types of warped event images H obtained depending on whether they store the event count ($b_k = 1$) or the balance of polarities ($b_k = p_k$) is shown in Fig. 14.

In the top row of Fig. 14, polarity is not used. For visualization purposes, event images in this row (Figs. 14a and 14b) are displayed in negative form (bright means lack of events and dark means abundance of events)[1]. As it is observed, per-pixel event accumulation (Fig. 14a) produces a motion-blurred image since events are triggered by moving edges. Contrarily, the image of warped events using the estimated motion parameters (Fig. 14b) presents a higher contrast and sharpness than the image in Fig. 14a, which indicates a better *alignment of the events* along the candidate

[1]Figs. 14a and 14b are the same type of images as in Fig. 12b, but with a different color scheme: from white to black instead of from blue to read.



(a) $N_e = 20\,000$ events
Baseline: 1.9 cm
Time span: 0.06 s.

(b) $N_e = 100\,000$ events
Baseline: 9.5 cm
Time span: 0.33 s.

(c) $N_e = 1\,000\,000$ events
Baseline: 85 cm
Time span: 2.93 s.

Figure 13: *Depth Estimation* for different subsets of events, with increasing time span and baseline. Top row: depth map overlaid on grayscale frame. Bottom row: 3D reconstruction (point cloud). Depth is color-coded, from red (close) to blue (far), in the range of 0.45 m to 2.4 m.

point trajectories on the image plane. The effect of having a higher contrast can also be noticed by comparing the distribution of values (i.e., histogram) of the images of warped events, as shown in Fig. 14c. The image with larger contrast (Fig. 14b) has a larger *range of values* than the image with lower contrast (i.e., darker pixels and larger amount of dark pixels in Fig. 14b with respect to Fig. 14a), and, since the range of values is non-negative and with a peak at zero, this means that the mass distribution of values shifts toward larger (positive) numbers as the contrast increases (i.e., the red curve in Fig. 14c becomes the blue curve as contrast increases).

The previous observations are also applicable to the second row of Fig. 14, where *event polarity* is used (cf. Figs. 14d and 14e). The average gray level corresponds to pixels where no events were generated; dark regions correspond to negative events, and bright regions correspond to positive events. Indeed, the image of warped events H obtained with the optimal parameters (Fig. 14e) has a larger contrast than the one with per-pixel polarity accumulation (Fig. 14d). The larger contrast of Fig. 14e over Fig. 14d is evidenced by the larger range of values and larger amount of brighter and darker pixels, as reported in the comparison of the distributions (Fig. 14f) of pixel values in both images.

We quantified the effect of using or not using the event polarity for rotational motion estimation on sequences from the dataset [1]. Each sequence has a 1 minute length and contains about 100-200 million events. Ground truth camera motion is provided by a sub-millimeter motion capture system. Each rotational motion sequence starts with rotations around each camera axis, and then is followed by rotations in all 3-DOFs. In addition, the speed of the motion increases as the sequence progresses. Fig. 15 shows the comparison of the results of our framework, not using event polarity ($b_k = 1$), against ground truth on the `poster_rotation` sequence. The curves corresponding to

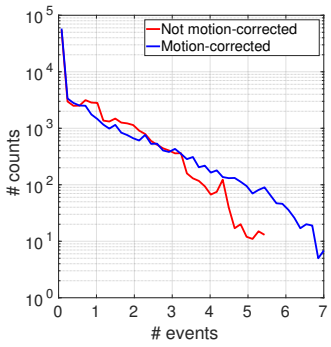
the 3-DOFs of the event camera on the entire sequence are shown in Fig. 15a. This plot shows the increasing speed of the motion, with excitations close to $\pm 1000\,^{\circ}/\mathrm{s}$. Figures 15b and 15c are zoomed-in versions of Fig. 15a, with rotations dominantly around each axis of the event camera (Fig. 15b) or in arbitrary axes (Fig. 15c), respectively. Our framework provides very accurate results, as highlighted by the very small errors: the lines of our method and those of the ground truth are almost indistinguishable at this scale. These errors are better noticed in the boxplots of Fig 16a, where errors are reported in sub-intervals of 15 s, in accordance with the increasing speed of the motion in the sequence. Fig. 16b reports the boxplot errors in case of using the event polarity ($b_k = p_k$) to build the image of warped events. As it can be observed by comparing both boxplots (Figs. 16a and 16b), using event polarity does not significantly change the results in this scenario. We measured Root Mean Square (RMS) angular velocity errors over the entire sequence of: $25.96\,^{\circ}/\mathrm{s}$ (without using polarity) and $24.39\,^{\circ}/\mathrm{s}$ (using polarity). Both are relatively small, compared to the peak velocities close to $1000\,^{\circ}/\mathrm{s}$, i.e., in the order of 2.5 % error.

References

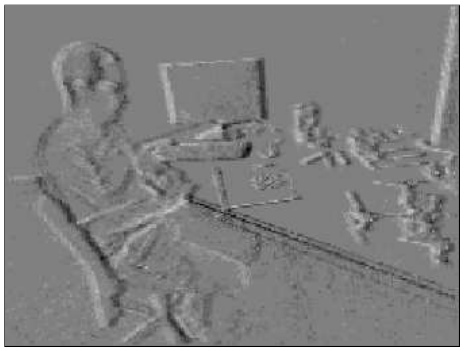
[1] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *Int. J. Robot. Research*, 36:142–149, 2017. 1, 3, 5, 6

(a) Warped events for angular velocity $\theta = \mathbf{0}$ (i.e., no motion correction). Using $b_k = 1$ in the image of warped events H.

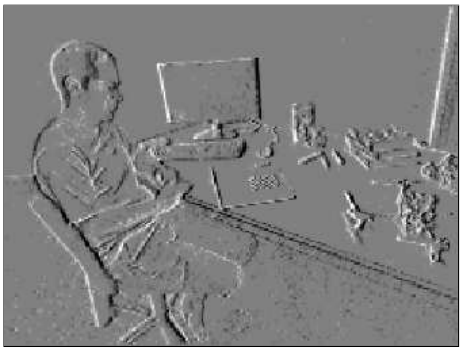
(b) Warped events using the estimated angular velocity θ^*, which produces motion-corrected, sharp edges.



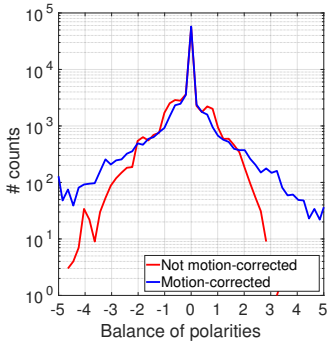
(c) Histograms of the negative of the images in Figs. 14a and 14b. The peak at zero corresponds to the white pixels.



(d) Warped events angular velocity $\theta = \mathbf{0}$ (i.e., no motion correction). Using polarity, $b_k = p_k$, in the image of warped events H.

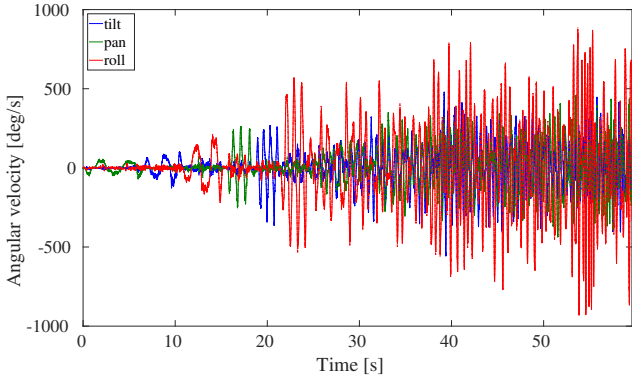


(e) Warped events using the estimated angular velocity θ^*, which produces motion-corrected, sharp edges. Using polarity, $b_k = p_k$, in H.

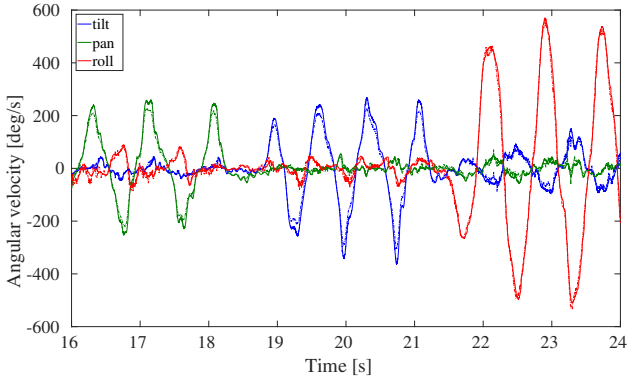


(f) Histograms of Figs. 14d and 14e. The peak at zero corresponds to the gray pixels.

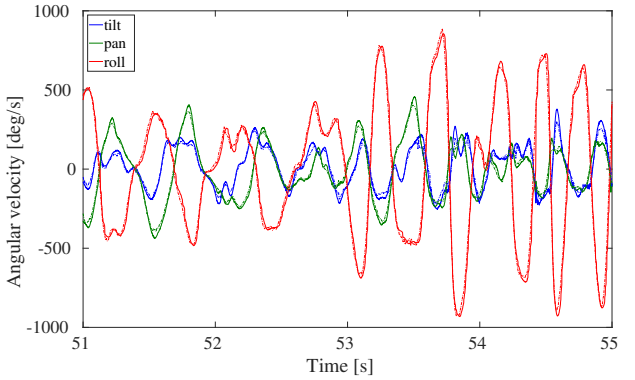
Figure 14: *Rotational Motion Estimation*. Images of warped events, displayed in grayscale to better visualize the motion blur due to event misalignment and the sharpness due to event alignment. Top: Not using polarity ($b_k = 1$ in the image of warped events H); bottom: using polarity ($b_k = p_k$ in H). Sequence `dynamic_rotation` from the dataset [1].



(a) Whole sequence. Rotational motion with increasing velocity, reaching speeds close to $\pm 1000\,^{\circ}/\mathrm{s}$.

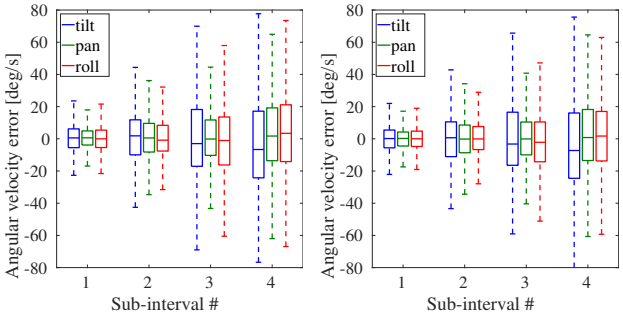


(b) Zoom of Fig. 15a, showing a series of rotations dominantly along one axis: pan (rotation around Y axis), tilt (rotation around X axis) and roll (rotation around Z axis).



(c) Zoom of Fig. 15a, showing rotations in arbitrary directions, with speed close to $1000\,^{\circ}/\mathrm{s}$.

Figure 15: *Rotational Motion Estimation*. Comparison of the estimated angular velocity (solid line) using our framework with $b_k = 1$ (i.e., without event polarity) against ground truth (dashed line). Sequence `poster_rotation` in the dataset [1].



(a) Without polarity (b) With polarity

Figure 16: *Rotational Motion Estimation*. Angular velocity error (estimated vs. ground truth) for the same sequence, with or wihout taking into account event polarity in the image of warped events H. Sequence `poster_rotation` from the dataset [1].