Asynchronous Planar Motion Estimation using Address Event Sensors

Charbel Azzi University of Waterloo, Canada
Adel Fakih University of Waterloo, Canada
Eihab Abdel-Rahman University of Waterloo, Canada
Email: {cazzi, afakih, eihab}@uwaterloo.ca

Abstract

Motion estimation is a central problem of computer vision essential to many applications, such as optical flow and egomotion estimation. In traditional frame-based cameras, motion estimation relies on the brightness constancy assumption. An inherent limitation of this assumption is that its temporal resolution is bounded by the fixed frame-rate of the camera. Address Event Sensors (AES) are bio-inspired vision sensors characterized by low latency, high dynamic range and high resilience to motion blur. Contrary to traditional frame-based cameras which output frames at fixed time intervals, AES generates asynchronous events at microsecond resolution each time the local brightness of a pixel changes. However, most of the current Address Event (AE) approaches to estimate motion have not been effective at exploiting these characteristics. They mostly rely on spatial smoothness that require accumulating events into grid-like representations for processing, eliminating most of the AES advantages. We conjecture that processing events asynchronously as they arrive should lead to better use of the camera's temporal resolution and hence result in motion estimates that are more resilient to rapid and shaky motions. In this paper, we present an asynchronous particle filter approach using BCE-based likelihood function, to solve for planar motion velocities using AES. It uses the AE data as the only source of information relying on a single event track, while freeing events from the spatial smoothness assumption. It is, thus, capable of exploiting the advantages offered by AES for motion estimation. Our results for general planar motion estimation are on par with state-of-the-art results.

1 Introduction

Motion estimation based on computer vision has significantly improved over the past years. It is a key technology for many applications, such as optical flow, camera egomotion, and tracking. In traditional frame-based cameras, the reliance on the brightness constancy assumption in motion estimation [1] results an inherent limitation. It bounds the temporal resolution by the frame-rate of the camera. On the other hand, motion estimation is one of the main applications of Address Event Sensors (AES) [2], owing to the fact that, under constant illumination, AES are motion sensors. Furthermore, the low latency (microsecond resolution) and high dynamic range (140 dB) of AES are additional advantages with the potential to alleviate the limitations imposed by the frame-based cameras on motion estimation.

AES approaches for motion estimation rely mostly on an assumption of spatial smoothness, where events in a spatial neighborhood are assumed to have similar velocities. Additionally, asynchronous methods [3] for motion estimation are sensitive to differentiation errors and require additional information for event-by-event processing, resulting in inconclusive estimates. On the other hand, the dominant stream of approaches accumulates events into grid-like groups. [4], which suffer from limitations including accumulation windows that do not contain enough information, adapting to fast motion, the aperture problem, inaccurate estimation due to violation of the constant velocity assumption within a temporal window (typically in ms close to the traditional cameras frame rate). These limitation end up undermining the AES advantages in motion estimation .

We conjecture that to exploit AES to their full advantages in solving the motion estimation problem, events should be processed asynchronously, on an event-by-event as soon as they fire. To realize that, one of the main challenges to address is making use of AE information to achieve single event tracking without spatial smoothness. Probabilistic Filtering techniques are the most common way used for asynchronous processing of events. In AES, these methods [5, 6] focus on designing likelihood functions for the filter. However, filtering techniques in AES rely on the availability of additional information (3D maps, grayscale images, external sensors, etc).

We present an asynchronous particle filter approach using a BCE-based likelihood function to solve for planar motion velocities using AES. The filter uses AE data as the only source of information. The BCE is used to derive an objective (likelihood) function for a given velocity by projecting backwards an incoming event, and minimizing the distance between this velocity and all possible candidates generated by that constraint. Our results on planar motion estimation show that our approach achieves state-of-the-art results by estimating optical flow accurately.

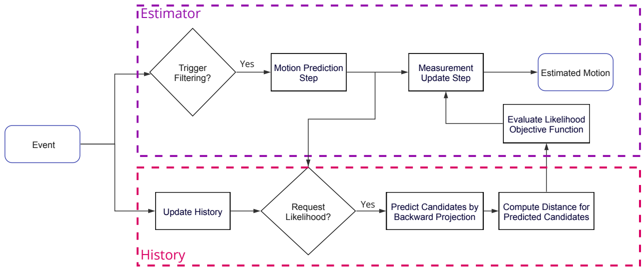

Fig. 1: Flow chart of an asynchronous particle filter algorithm for planar motion estimation.

2 A Particle Filter for Asynchronous Planar Motion Estimation

In planar (2D) motion scenes where the camera moves parallel to a plane of motion, i.e. with a known depth Z and no rotational movement, the optical flow (u,v) is related to the camera's translational velocity (T_x,T_y) by

$$(u = f\frac{T_x}{Z}, v = f\frac{T_y}{Z}). \tag{1}$$

To estimate the planar motion of the camera over time, we use a particle filter with a multi-hypothesis capability to cope with the AE data on an event-by-event basis as soon as an event fires. The system overview is illustrated in Figure 1. For an event $e=(x,y,t,p)$ firing at time t, our filter at time t is represented by a list of N particles, $P^{(t)} = \{P_1^{(t)}, P_2^{(t)}, ...P_N^{(t)}\}$, where $1 \leq i \leq N$. Each particle consists of a hypothesis of the current state $X_i^t = U_i^{(t)}$ which estimates the OF vector $U_i^{(t)} = (u_i^{(t)}, v_i^{(t)}) \in \mathbb{R}^2$. Note that estimating the OF vector $U_i^{(t)}$ is equivalent to estimating the camera's 2D translation velocity vector $T_i^{(t)} = (T_{x_i}^{(t)}, T_{y_i}^{(t)}) \in SO(2)$ for planar motion.

For the motion prediction step of our filter, we adopt a standard constant motion model where the estimated average motion optical flow of a particle at any given moment is stable but the variance of the estimation is variable.

In the measurement update step, we update the weights of the perturbed particles by applying Bayes rule to each particle knowing that the weights will be normalized afterwards. We present a BCE-based likelihood $P(z|\ \mathrm{U}_i^{(t)})$ for each particle in the form of an exponential decay function:

$$P(z \mid U_i^{(t)}) = \exp(-\alpha_p L_{d_i}^{(t)}(e)), \tag{2}$$

where α_p is a scaling decay parameter, and $L_{d_i}^{(t)}(e))$ is our asynchronous distance objective function derived from the BCE and evaluated at each particle i. For each input event $e(x,y,t,p)$ let $\mathcal{E}(\hat{e},d_s)$ represent the set of events within a maximum spatial radial distance d_s of $\hat{e}$, where $\hat{e}$ is the predicted event at location $(\hat{x},\hat{y})$ and time $t-dt$, denoted by $\hat{e} = (\hat{x},\hat{y},t-dt,p)$. $(\hat{x},\hat{y})$ is obtained by back projecting e from time t to time $t-dt$ using a constant velocity model. The spatial search window is 3×3 and the temporal search window is $50\mu s$. Then the distance cost corresponding to a particular input event can be written as:

$$L_d(e) = \min_{e \in \mathcal{E}(\hat{e}, d_s)} (D(e, \hat{e})), \quad (3)$$

where D is a distance metric in space, time and polarity between two events e_i and e_j defined as follows:

$$D(x_i, y_i, t_i, t_{pi}, x_j, y_j, t_j, t_j) = \begin{cases} r_t|t_i - t_j| + ||(x_i, y_i), (x_j, y_j)||, & \text{if } p_i = p_j \\ d_{max} & \text{otherwise} \end{cases} \quad (4)$$

where r_t penalizes the deviation in time, d_{max} is a maximum spatio-temporal distance parameter.

Initially, all particles are uniformly initialized to cover a large range of image velocities, with the same weight. The likelihood score is computed using a history component, which represents the memory of our particle filter as shown in Figure 1. The history stores and updates two separate channels one for each polarity, i.e. an event with positive polarity uses a the history of all positive polarity measurements. When the filter requires a likelihood to update its measurement step, it will use the history component to find the minimum distance score for a certain hypothesis (velocity) given the current measurement event. After assigning a likelihood to each particle, we normalize the distribution, then apply standard systematic resampling. Finally, a mean OF estimation, for every event measurement, is saved in a shape of weighted average over all the particles.

3 Results

We first show our results on synthetic dataset generated by an event-based data simulator we created to simulate the AES. Figure 2 shows a camera moving with a linearly increasing velocities in the x and y directions. Each plot show the optical flow prediction (red), ground truth OF (green), and particles (black) vs Events. For better visualization we plot a prediction for every 100 events. It is clear that our approach can accurately estimate the planar motion flow.

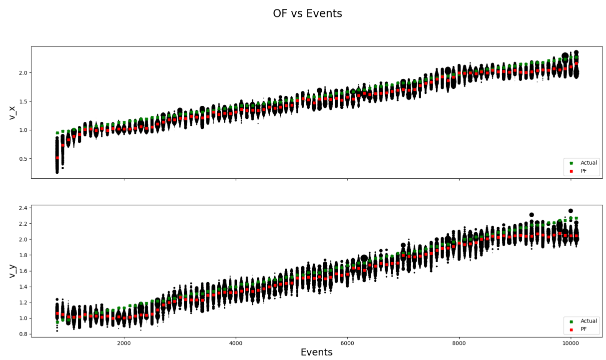

Fig. 2: Our asynchronous approach accurately tracks a camera moving with linearly increasing velocities in the x and y directions. Comparison of the ground truth (green) and our predicted (red) velocities. Particle distributions are shown in black.

For real events data evaluation, we used the state-of-the-art Event-Camera Dataset and Simulator [7]. It contains a series planar static scenes which were collected via the DAVIS AES [2]. Table 1 shows our results against real event data for planar motion estimation, evaluated against one of the main AES approaches for optical flow [4], and the MATLAB implementation of the LK frame-based approach using the grayscale images from the provided with dataset. Our approach clearly outperforms the synchronous main stream of AES relying on the spatial smoothness assumption. The latter further highlights the importance of exploiting AES to their full advantages. Additionally, AES significantly outperforms the frame-based approaches, especially on the high dynamic range (hdr) sets, where the LK approach was failing due to the quality of the grayscale images affected by the large intensity differences in those scenes. Figure 3 shows the Relative Endpoint Error (EE) vs events for the *slider far* sequence over a sample of $100,000$ events. It also shows the relative EE for the first few thousands of events used in the initial estimates.

Table 1: Evaluation of our asynchronous planar motion approach against the main TS approach called SAE [4], and the frame-based Lucas and Kanade [1]. Relative Average Endpoint Errors (AEE) are reported.

Sequence	slider far	slider close	slider hdr far	slider hdr close
	AAE$_{rel}[\%]$	AAE$_{rel}[\%]$	AAE$_{rel}[\%]$	AAE$_{rel}[\%]$
Ours	**0.95**	**1.1**	**0.97**	**1.08**
SAE [4]	3.9	3.7	4.9	4.6
Frame-based LK [1]	14.7	13.9	-	-

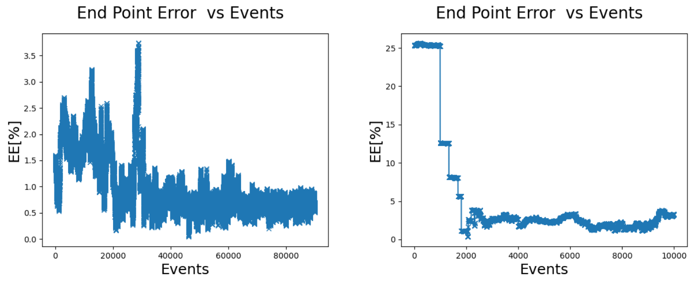

Fig. 3: Relative Endpoint Error (EE) vs a sample of 10^5 events for the *slider close* sequence.

4 Conclusion & Future Work

In conclusion, we presented an asynchronous particle filter approach to solve for planar motion estimation using Address Event Sensors. We introduced a BCE-based likelihood distance function for the filter, which is computed via a history of events created for our particle filter. Our results showed that our approach is capable of accurately estimating the planar motion for simulated data, and achieving state-of-the-art results on real events data. We are working on extending this approach to estimation the general motion using Address Event Sensors.

References

[1] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision." Vancouver, British Columbia, 1981.

[2] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, "A 240$\times$ 180 130 db 3 μs latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014.

[3] R. Benosman, S.-H. Ieng, C. Clercq, C. Bartolozzi, and M. Srinivasan, "Asynchronous frameless event-based optical flow," *Neural Networks*, vol. 27, pp. 32–37, 2012.

[4] R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, "Event-based visual flow," *IEEE transactions on neural networks and learning systems*, vol. 25, no. 2, pp. 407–417, 2013.

[5] E. Mueggler, B. Huber, and D. Scaramuzza, "Event-based, 6-dof pose tracking for high-speed maneuvers," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on.* IEEE, 2014, pp. 2761–2768.

[6] H. Kim, A. Handa, R. Benosman, S.-H. Ieng, and A. J. Davison, "Simultaneous mosaicing and tracking with an event camera," *J. Solid State Circ*, vol. 43, pp. 566–576, 2014.

[7] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam," *The International Journal of Robotics Research*, vol. 36, no. 2, pp. 142–149, 2017.