FOE-based Regularization for Optical Flow Estimation from an In-vehicle Event Camera

Jun Nagata[a], Yusuke Sekikawa[b], Kosuke Hara[b], and Yoshimitsu Aoki[a]

[a]Keio University, 3-14-1, Hiyoshi, Kohoku Yokohama, Kanagawa, Japan
[b]Denso IT Laboratory Inc., Shibuya Cross Tower 28th Floor, 2-15-1, Sibuya, Tokyo, Japan

ABSTRACT

Optical flow estimation in onboard cameras is an important task in automatic driving and advanced driver-assistance systems. However, there is a problem that calculation is mistakable with high contrast and high speed. Event cameras have great features such as high dynamic range and low latency that can overcome these problems. Event cameras report only the change in the logarithmic intensity per pixel rather than the absolute brightness. There is a method of estimating the optical flow simultaneously with the luminance restoration from the event data. The regularization using the L1 norm of differentiation is insufficient for spatially sparse event data. Therefore, we propose to use the focus of expansion (FOE) for regularization of optical flow estimation in event camera. The FOE is defined as the intersection of the translation vector of the camera and the image plane. The optical flow becomes radial from the FOE excluding the rotational component. Using the property, the optical flow can be regularized in the correct direction in the optimization process. We demonstrated that the optical flow was improved by introducing our regularization using the public dataset.

Keywords: Optical flow, Event camera, Focus of expansion

1. INTRODUCTION

In automatic driving and advanced driver-assistance systems, optical flow (OF) is frequently used to recognize moving objects using in-vehicle camera. However, it is difficult to calculate the OF with high contrast and high speed. Against these problems, bio-inspired image sensors named event cameras have been developed[1].[2] Event cameras have excellent characteristics such as high temporal resolution and high dynamic range. The output of these cameras is not image frames like in a standard camera; rather, these cameras output an asynchronous sequence of reports of intensity change at each pixel. Therefore, computer vision algorithms specialized for event cameras are required. In the conventional OF estimation method,[3] intensity and OF are simultaneously estimated from the event data by optimizing the cost function consisting of the event data term, the constraints of the OF, and the smoothness term with the L1 norm. However, regularization assuming spatial smoothness for sparse event data is insufficient. Therefore, we propose the novel regularization of the OF estimation in event camera utilizing the focus of expansion (FOE).

2. BACKGROUND

The FOE is defined as the intersection of the translation axis of the camera motion and the image plane. By introducing the FOE, it is possible to obtain motion parameters from fewer corresponding points with little calculation under the condition that the monocular camera is fixed to the car body.[4] The FOE is also used as a constraint when determining the OF due to the vehicle's ego-motion.[5] The FOE has an important property that the OF becomes radial from it when the rotations component is removed from the OF of the background point. We utilize this property as regularization for the OF estimation from an In-vehicle event camera.

Further author information:
J. Nagata: E-mail: jnagata@aoki-medialab.jp
Y. Aoki: E-mail: aoki@elec.keio.ac.jp

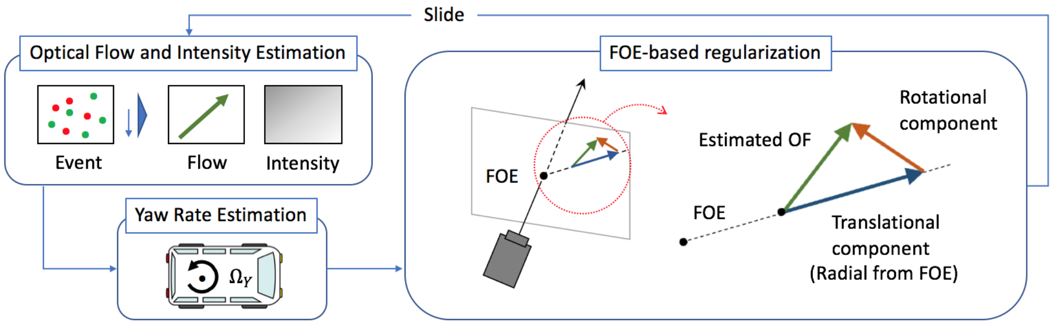

Figure 1. System flow.

3. METHOD

Our method is roughly divided into three steps and executed in the flow shown in Fig. 1. Similar to the OF estimation method,[3] we have a time window whose temporal length is T, which we discretize into K cells, which are each of δt length. The second and third steps are executed for all frames in the sliding time window after estimation of the OF and intensity. The cost function is described as (1):

$$\begin{aligned} E(L,\boldsymbol{u}) = &\int_S \int_T (\lambda_1 \|\boldsymbol{u}_{\boldsymbol{x}}\|_1 + \lambda_2 \|\boldsymbol{u}_t\|_1 + \lambda_3 \|L_{\boldsymbol{x}}\|_1 + \lambda_4 \|\langle L_{\boldsymbol{x}} \cdot \delta t \boldsymbol{u}\rangle + L_t\|_1 + \lambda_5 h_\theta (L - L(t_p)))dtd\boldsymbol{x} \\ &+ \int_I \sum_{i=2}^{\phi(\boldsymbol{x})} \|L(t_i) - L(t_{i-1}) - \theta\rho_i\|_1 d\boldsymbol{x} + \lambda_{foe} \|\boldsymbol{u}' - \boldsymbol{u}'_{foe}\|_2^2 \end{aligned} \tag{1}$$

The last term is our proposed regularization term, and the other terms are same as conventional ones.[3] Considering convexity, we optimized the two respectively. The notation of the last regularization term is explained in Sec. 3.3.

3.1 Intensity and optical flow estimation

In this step, we optimize a part of the cost function excluding the last term by the primal dual algorithm. The cost function in Bardow's method[3] consists roughly three types of terms: one for spatial and temporal smoothness, one for OF, and one for fitting to event data. The smoothness term is widely used in variational methods, such as TV-L[1], and has been used successfully in computer vision algorithms[6,7] (for example, image de-noising, OF estimation and segmentation). The second term, for OF, evaluates the consistency between the intensity and the motion vector; it reflects the classical definition of OF.[8] The third term is unique for event cameras and evaluates whether the intensity conforms to the events generation model. More specifically, if the difference between the luminance value at the time when an event occurred and the value at a previous time are not equal to the threshold, a penalty is given by that amount.

3.2 Yaw rate estimation

In this step, we estimate the yaw angular velocity Ω_Y from the estimated OF. Assuming a rotational motion Ω_Y around the y-axis by steering the wheel operation when turning the corner and introducing the FOE coordinate (x_0, y_0), Ω_Y is determined by the following equation:

$$\Omega_{Y,(x,y)} = f \frac{u_{x,(x,y)}(y - y_0) - u_{y,(x,y)}(x - x_0)}{(f^2 + x^2)(y - y_0) - xy(x - x_0)} \tag{2}$$

In theory, Ω_Y can be obtained from only one correspondence point.[4] Since the correspondence point is a form of dense OF estimated from the event, we propose the proprietary yaw estimation method. In our proposal method, Ω_Y is determined as the mode value in the place where the event occurred in the cells before and after the frame. The place where the event occurred to be more reliable because the OF is estimated based on the luminance restored from the event data depending on the OF term in (1). Conversely, since other parts are estimated depending on the smoothness term in (1), the reliability is considered to be low.

3.3 FOE-based regularization

This step is the main part of our method; FOE-based regularization that constrain estimated the OF to follow the FOE. By using the components due to the rotation of the OF $\boldsymbol{u}^{\Omega}_{(x,y)} = \left((f^2+x^2)\hat{\Omega}_Y/f, xy\hat{\Omega}_Y/f\right)$, the relationship between the OF and the FOE is written as the following equation (3).

$$\frac{u_{x,(x,y)} - u^{\Omega}_{x,(x,y)}}{u_{y,(x,y)} - u^{\Omega}_{y,(x,y)}} = \frac{x - x_0}{y - y_0} \tag{3}$$

The important property of the FOE that we previously mentioned is represented in (3). Here the rotational component is removed ($\boldsymbol{u}' = \boldsymbol{u} - \boldsymbol{u}^{\Omega}$). An arrow extending from the FOE to each coordinate point is denoted by $\boldsymbol{u}_{foe,(x,y)} = (x - x_0, y - y_0)$, and an OF in which the length is made equal to $\boldsymbol{u}'$ is represented by $\boldsymbol{u}'_{foe} = \boldsymbol{u}_{foe} \cdot \|\boldsymbol{u}'\|/\|\boldsymbol{u}_{foe}\|$. The regularized OF $\hat{\boldsymbol{u}}'$ can be obtained by internally dividing $\boldsymbol{u}'$ and $\boldsymbol{u}'_{foe}$ by a ratio α ($\hat{\boldsymbol{u}}' = (1-\alpha)\boldsymbol{u}' + \alpha\boldsymbol{u}'_{foe}$). It is possible to regularize the orientation of the OF to follow the FOE while maintaining this size. After this operation, we returned the excluded rotation components ($\hat{\boldsymbol{u}} = \hat{\boldsymbol{u}}' + \boldsymbol{u}^{\Omega}$). The FOE-based regularization of the OF is performed for each frame after the luminance reconstruction, and the three steps are repeated across the sliding time window.

4. EXPERIMENT

In this section, we evaluate the effectiveness of our proposed method based on the experiment results. The dataset we used is Multi-Vehicle Stereo Event Camera dataset (MVSEC),[9] which includes driving scenes in urban areas in day and night. An Event-based camera used in the dataset is mDAVIS-346B, which is the improved version of DVS128[1] and DVS240;[2] it can record event data with a spatial resolution 260 × 346. The ground truth OF can be obtained by combining depth data and rotational motion data from velodyne in this dataset.[10]

4.1 Outline of Experiment

The evaluation metric is the average end-point error (AEE), defined as the desitance between the end points of estimated and ground truth flow vectors:

$$AEE = \frac{1}{N}\sum_{x,y}\left\|\begin{pmatrix} u_x(x,y)_{est} \\ u_y(x,y)_{est}\end{pmatrix} - \begin{pmatrix} u_x(x,y)_{gt} \\ u_y(x,y)_{gt}\end{pmatrix}\right\|_2 \tag{4}$$

As with what is done in EV-FlowNet,[10] when computing this metric, we used only the pixels where the event occurred. We used outdoor driving sequences named outdoor_day1, whose length corresponds to 30,000 frames. In order to show the effectiveness of our regularization, comparison was made in the following two cases.

- Bardow[3]'s method
- Our method, including FOE-based regularization

We set δt to 7.5 ms, α to 0.5, and K to 128. For the sake of simplicity, it was assumed that the optical axis of the camera coincided with the traveling direction of the car and the coordinate of the FOE was located at the center of the image.

4.2 Experimental result

The AEE of the OF with our regularization is 24.82 pix/s, which is smaller than one without our regularization 28.36 pix/s. and it can be said that the more correct OF was obtained by our effective regularization. Since the estimation is performed with a considerably small time width of 7.5 ms, the error becomes large when evaluating with metric of 1 s unit, Fig. 2 is visualization of the result. As you can see from the mixture of different colors, the direction of the flow estimated in Bardow's method[3] is mistaken.

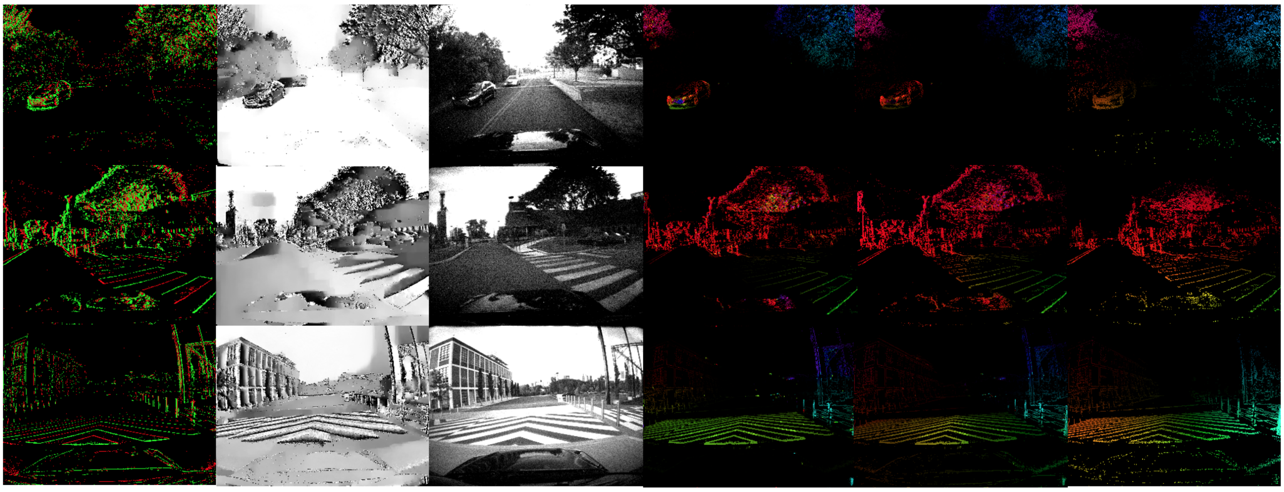

Figure 2. Result images. The table entries from left to right: input events, estimated image, ground truth image, Bardow's OF, estimated OF with our regularization, ground truth OF.

5. CONCLUDION

We proposed the FOE-based regularization of the OF to compensate for the imperfection of estimating from event data. For our regularization, we utilized the important properties of the FOE, which is derived from the definition of the FOE and the features of in-vehicle camera movement. It is possible to correct the flow direction by removing the rotational component of the OF and following the FOE.

In this paper, we showed the effect of regularization using the FOE while considering only the background point; however, our method has the disadvantage of neglecting moving objects that are not background points. As a future prospect, it is necessary to have an algorithm that judges whether a pixel is a background point to be corrected or a point included in a moving object that should not be modified in regularization.

REFERENCES

[1] Lichtsteiner, P., Posch, C., Delbruck, T., and Member, S., "A 128 × 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," *Work* **43**(2), 566–576 (2008).

[2] Brandli, C., Berner, R., Yang, M., Liu, S. C., and Delbruck, T., "A 240 × 180 130 dB 3 μs latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits* **49**(10), 2333–2341 (2014).

[3] Bardow, P., Davison, A. J., and Leutenegger, S., "Simultaneous Optical Flow and Intensity Estimation from an Event Camera," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* , 884–892 (2016).

[4] Hu, Z., Uchiyama, K., and Kawaji, S., "Determining Motion Parameters for Vehicle-Mounted Camera Using Focus of Expansion," *T. IEE Japan* **119-D**(1), 50–57 (1999).

[5] Yamaguchi, K., McAllester, D., and Urtasun, R., "Robust monocular epipolar flow estimation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , 1862–1869 (2013).

[6] Chambolle, A. and Pock, T., "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision* **40**(1), 120–145 (2011).

[7] Zach, C., Pock, T., and Bischof, H., "A Duality Based Approach for Realtime TV-L 1 Optical Flow," *Pattern Recognition* **1**(1), 214–223 (2007).

[8] Horn, B. K. B. and Schunck, B. G., "Determining Optical Flow," *Artificial Intelligence Memo 572* **319**, 185–203 (1980).

[9] Zhu, A. Z., Thakur, D., Ozaslan, T., Pfrommer, B., Kumar, V., and Daniilidis, K., "The Multi Vehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception," **19104**, 1–8 (2018).

[10] Zhu, A. Z., Yuan, L., Chaney, K., and Daniilidis, K., "EV-FlowNet: Self-Supervised Optical Flow Estimation for Event-based Cameras," *Robotics: Science and Systems* (2018).