Bringing Blurry Alive at High Frame-Rate with an Event Camera

Liyuan Pan\textsuperscript{1,2}, Richard Hartley\textsuperscript{1,2}, Cedric Scheerlinck\textsuperscript{1,2}, Miaomiao Liu\textsuperscript{1,2}, Xin Yu\textsuperscript{1,2}, and Yuchao Dai\textsuperscript{3}

Abstract—Event-based cameras can measure intensity changes (called ‘events’) with microsecond accuracy under high-speed motion and challenging lighting conditions. With the active pixel sensor (APS), the event camera allows simultaneous output of the intensity frames. However, the output images are captured at a relatively low frame-rate and often suffer from motion blur. A blurry image can be regarded as the integral of a sequence of latent images, while the events indicate the changes between the latent images. Therefore, we are able to model the blur-generation process by associating event data to a latent image. Based on the abundant event data and the low frame-rate easily blurred images, we propose a simple and effective approach to reconstruct a high-quality and high frame-rate shape video. Starting with a single blurry frame and its event data, we propose the Event-based Double Integral (EDI) model. Then, we extend it to multiple Event-based Double Integral (mEDI) model to get more smooth and convincing results based on multiple images and their events. We also provide an efficient solver to minimize the proposed energy model. By optimizing the energy model, we achieve significant improvements in removing general blurs and reconstructing high temporal resolution video. The video generation is based on solving a simple non-convex optimization problem in a single scalar variable. Experimental results on both synthetic and real images demonstrate the superiority of our mEDI model and optimization method in comparison to the state-of-the-art.

Index Terms—Event Camera, Motion Blur, High-temporal Resolution Reconstruction, mEDI Model, Fibonacci search.

1 INTRODUCTION

Event cameras (such as the Dynamic Vision Sensor (DVS) \cite{5} and the Dynamic and Active-pixel Vision Sensor (DAVIS) \cite{6}) are sensors that asynchronously measure the intensity changes at each pixel independently with microsecond temporal resolution (if nothing moves in the scene, no events are triggered). The event stream encodes the motion information by measuring the precise pixel-by-pixel intensity changes. Event cameras are more robust to low lighting and highly dynamic scenes than traditional cameras since they are not affected by under/over exposure or motion blur associated with a synchronous shutter.

Due to the inherent differences between event cameras and standard cameras, existing computer vision algorithms designed for standard cameras cannot be applied to event cameras directly. Although the DAVIS \cite{6} can provide the simultaneous output of the intensity frames and the event stream, there still exist major limitations with current event cameras:

- **Low frame-rate intensity images**: In contrast to the high temporal resolution of event data ($\geq 3\mu$s frame rate), the current event cameras only output low frame-rate intensity images ($\geq 5ms$ time resolution).
- **Inherent blurry effects**: When recording highly dynamic scenes, motion blur is a common issue due to the relative motion between the camera and the scene. The output of the intensity image from the APS tends to be blurry.

To address these above challenges, various methods have been proposed by reconstructing high frame-rate videos. The existing methods can be in general categorized as:

1. Event data only solutions \cite{7, 8, 9}, where the results tend to lack the texture and consistency of natural videos, as they fail to use the complementary information contained in the low frame-rate intensity image;
2. Low frame-rate intensity-image-only solutions \cite{2}, where an end-to-end learning framework has been proposed to learn regression between a single blurry image and a video sequence, whereas the rich event data are not used;
3. Jointly exploiting event data and intensity images \cite{3, 10, 11, 12}, building upon the interaction between both sources of information. However, these methods fail to address the blur issue associated with the captured image frame. Therefore, the reconstructed high frame-rate videos can be degraded by blur.

Contrary to the existing methods treat blur as the degree of the image quality, we give an alternative new insight into the problem. Although blurry frames cause undesired image degradation, they also encode the relative motion between the camera and the observed scene. Taking full advantage of the encoded motion information would benefit the reconstruction of high frame-rate videos.

To alleviate these problems, in our previous work \cite{4}, we proposed an Event-based Double Integral (EDI) model to solve the above problems by reconstructing a high frame-rate video from a single image (even blur) and its event sequence, where the blur effects have been reduced in each
reconstructed frame. Our EDI model naturally relates the desired high frame-rate sharp video, the captured intensity frame and event data. Based on the EDI model, high frame-rate video generation is as simple as solving a non-convex optimization problem in a single scalar variable.

As the EDI model only based on a single image, error from the event data will accumulate to reduce the quality of the reconstructed video. Though we integral over small time intervals from the centre of the exposure time to mitigate this error, flickering (un-smooth) sometimes occurs at two endpoints of the video. The flickering usually happens when objects and camera have a large relative motion. Therefore, in this paper, we propose an extended multiple Event-based Double Integral (mEDI) model to handle the discontinuity in the reconstructed videos and further improves the quality of the reconstructed video significantly.

In this paper, we extend the previous work significantly in the following ways:

1) We propose a simple and effective model, named the Event-based Double Integral (EDI) model, and extend it to mEDI model to restore a better high frame-rate sharp video from multiple images (even blur) and their corresponding event data.

2) Using our proposed formulation of mEDI, we propose a stable and general method to generate a sharp video under various types of blur by solving a single variable non-convex optimization problem, especially in low lighting and complex dynamic conditions.
The frame rate of our reconstructed video can theoretically be as high as the event rate (200 times greater than the original frame rate in our experiments). With multiple images as input, the reconstructed video preserves more abundant texture and the consistency of natural image.

2 RELATED WORK

Event cameras such as the DAVIS and DVS [5], [6] report log intensity changes, inspired by human vision. The result is a continuous, asynchronous stream of events that encodes non-redundant information about local brightness change. Estimating image intensity from events is important because it grants computer vision researchers a readily available high-temporal-resolution, high-dynamic-range imaging platform that can be used for tasks such as face-detection [9], SLAM [13], [14], [15], or optical flow estimation [7]. However, even though several works try to explore the advantages of the high temporal resolution provided by event cameras [16], [17], [18], [19], [20], [21], how to make the best use of the non-redundant information about local brightness change is a continuous, asynchronous stream of events that encodes log intensity changes, inspired by human vision. The result of scenes is not trivial. Reinbacher interval (or event batch size) that works robustly for all types to the length of the time window, and choosing a time-based image reconstruction. Kim et al. [14] reconstruct high-quality images from an event camera under a strong assumption that the only movement is pure camera rotation, and later extend their work to handle 6-degree-of-freedom motion and depth estimation [15]. Bardow et al. [7] simultaneously optimise optical flow and intensity estimates within a fixed-length sliding spatio-temporal window using the primal-dual algorithm [22]. Taking a spatio-temporal window of events imposes a latency cost at minimum equal to the length of the time window, and choosing a time-interval (or event batch size) that works robustly for all types of scenes is not trivial. Reinbacher et al. [8] integrate events over time while periodically regularising the estimate on a manifold defined by the timestamps of the latest events at each pixel. Barua et al. [9] generate image gradients by dictionary learning and obtain a logarithmic intensity image via Poisson reconstruction. However, the intensity images reconstructed by the previous approaches suffer from obvious artifacts as well as lack of texture due to the spatial sparsity of event data.

To achieve more image detail in the reconstructed images, several methods trying to combine events with intensity images have been proposed. The DAVIS [6] uses a shared photo-sensor array to simultaneously output events (DVS) and intensity images (APS). Scheerlinck et al. [3] propose an asynchronous event-driven complementary filter to combine APS intensity images with events, and obtain continuous-time image intensities. Brandli et al. [12] combine event frames and event stream from the DAVIS camera to create inter-frame intensity estimates by dynamically estimating the contrast threshold (temporal contrast) of each event. Each new image frame resets the intensity estimate, preventing excessive growth of integration error, but also discarding important accumulated event information. Shedligeri et al. [11] first exploit two intensity images to estimate depth. Then, they use the event data only to construct a pseudo-intensity sequence (using [8]) between the two intensity images and use the pseudo-intensity sequence to estimate ego-motion using visual odometry. Using the estimated 6-DOF pose and depth, they directly warp the intensity image to the intermediate location. Liu et al. [23] assume a scene should have static background. Thus, their method needs an extra sharp static foreground image as input and the event data are used to align the foreground with the background. To alleviate these problems, in our previous work [4], we proposed an EDI model to solve the above problems by reconstructing a high frame-rate video from a single image (even blur) and its event sequence, where the blur effects have been reduced in each reconstructed frame. Our EDI model naturally relates the desired high frame-rate sharp video, the captured intensity frame and event data. Based on the EDI model, high frame-rate video generation is as simple as solving a non-convex optimization problem in a single scalar variable.

Image deblurring. Traditional deblurring methods usually make assumptions on the scenes (such as a static scene) or exploit multiple images (such as stereo, or video) to solve the deblurring problem. Significant progress has been made in the field of single image deblurring. Methods using gradient-based regularizers, such as Gaussian scale mixture [24], $l_1/l_2$ norm [25], edge-based patch priors [26] and $l_0$-norm regularizer [27], have been proposed. Non-gradient-based priors such as the color line-based prior [28], and the extreme channel (dark/bright channel) prior [29], [30] have also been explored. Another family of image deblurring methods tends to use multiple images [31], [32], [33], [34], [35].

Driven by the success of deep neural networks, Sun et al. [36] propose a convolutional neural network (CNN) to estimate locally linear blur kernels. Gong et al. [37] learn optical flow from a single blurry image through a fully-convolutional deep neural network. The blur kernel is then obtained from the estimated optical flow to restore the sharp image. Nah et al. [38] propose a multi-scale CNN that restores latent images in an end-to-end learning manner without assuming any restricted blur kernel model. Tao et al. [1] propose a light and compact network, SRN-DeblurNet, to deblur the image. However, deep deblurring methods generally need a large dataset to train the model and usually require sharp images provided as supervision. In practice, blurry images do not always have corresponding ground-truth sharp images.

Blurry image to sharp video. Recently, two deep learning based methods [2], [39] propose to restore a video from a single blurry image with a fixed sequence length. However, their reconstructed videos do not obey the 3D geometry of the scene and camera motion. Although deep-learning based methods achieve impressive performance in various scenarios, their success heavily depend on the consistency between the training datasets and the testing datasets, thus hinder the generalization ability for real-world applications.

3 FORMULATION

In this section, we first introduce our EDI model and then develop a mEDI model of the relationships between the events, the latent images and the blurry images. Our goal is to reconstruct a high frame-rate, sharp video from single and multiple images and its corresponding events. This
model can tackle various blur types and work stably in highly dynamic sceneries and low lighting conditions.

### 3.1 Event Camera Model

Event cameras are bio-inspired sensors that asynchronously report logarithmic intensity changes [3], [6]. Unlike conventional cameras that produce the full image at a fixed frame-rate, event cameras trigger events whenever the change in intensity at a given pixel exceeds a preset threshold. Event cameras do not suffer from the limited dynamic ranges typical of sensors with synchronous exposure time, and are able to capture high-speed motion with microsecond accuracy.

Inherent in the theory of event cameras is the concept of the latent image $L_{xy}(t)$, denoting the instantaneous intensity at pixel $(x, y)$ at time $t$, related to the rate of photon arrival at that pixel. The latent image $L_{xy}(t)$ is not directly output by the camera. Instead, the camera outputs a sequence of events, denoted by $(x, y, t, \sigma)$, which record changes in the intensity of the latent image. Here, $(x, y)$ are image coordinates, $t$ is the time the event takes place, and polarity $\sigma = \pm 1$ denotes the direction (increase or decrease) of the intensity change at that pixel and time. Polarity is given by,

$$\sigma = T \left( \log \left( \frac{L_{xy}(t)}{L_{xy}(t_{ref})} \right), c \right),$$  

where $T(\cdot, \cdot)$ is a truncation function,

$$T(d, c) = \begin{cases} 
+1, & d \geq c, \\
0, & d \in (-c, c), \\
-1, & d \leq -c.
\end{cases}$$

Here, $c$ is a threshold parameter determining whether an event should be recorded or not, $L_{xy}(t)$ is the intensity at pixel $(x, y)$ in the image $L(t)$ and $t_{ref}$ denotes the timestamp of the previous event. When an event is triggered, $L_{xy}(t_{ref})$ at that pixel is updated to a new intensity level.

### 3.2 Intensity Image Formation

In addition to the event sequence, event cameras can provide a full-frame grey-scale intensity image, at a much slower rate than the event sequence. The grey-scale images may suffer from motion blur due to their long exposure time. A general model of image formation is given by,

$$B = \frac{1}{T} \int_{f-T/2}^{f+T/2} L(t) \, dt.$$  

where $B$ is a blurry image, equal to the average value of the latent image during the exposure time $[f-T/2, f+T/2]$. This equation applies to each pixel $(x, y)$ independently, and subscripts $x, y$ denoting pixel location are often omitted henceforth.

### 3.3 Event-based Double Integral Model

We aim to recover a sequence of latent intensity images by exploiting both the blur model and the event model. We define $e_{xy}(t)$ as a function of continuous time $t$ such that

$$e_{xy}(t) = \sigma \, \delta_{t_0}(t),$$

wherever there is an event $(x, y, t_0, \sigma)$. Here, $\delta_{t_0}(t)$ is an impulse function, with unit integral, at time $t_0$, and the sequence of events is turned into a continuous time signal, consisting of a sequence of impulses. There is such a function $e_{xy}(t)$ for every point $(x, y)$ in the image. Since each pixel can be treated separately, we omit the subscripts $x, y$.
During an exposure period \([f - T/2, f + T/2]\), we define \(E(t)\) as the sum of events between time \(f\) and \(t\) at a given pixel,

\[
E(t) = \int_f^t e(s)ds,
\]

which represents the proportional change in intensity between time \(f\) and \(t\). Except under extreme conditions, such as glare and no-light conditions, the latent image sequence \(L(t)\) is expressed as,

\[
L(t) = L(f) \exp(cE(t)) = L(f) \exp(cE(t)),
\]

In particular, an event \((x, y, t, \sigma)\) is triggered when the intensity of a pixel \((x, y)\) increases or decreases by an amount \(c\) at time \(t\). We put a tilde on top of things to denote logarithm, e.g. \(\tilde{L}(t) = \log(L(t))\).

\[
\tilde{L}(t) = \tilde{L}(f) + cE(t).
\]
Given a sharp frame, we can reconstruct a high frame-rate video from the sharp starting point \( \mathbf{L}(f) \) by using Eq. (3). When the input image is blurry, a trivial solution would be to first deblur the image with an existing deblurring method and then to reconstruct the video using Eq. (3) (see Fig.8 for details). However, in this way, the event data between intensity images is not fully exploited, thus resulting in inferior performance. Instead, we propose to reconstruct the video by exploiting the inherent connection between event and blur, and present the following model.

As for the blurred image,

\[
\mathbf{B} = \frac{1}{T} \int_{f-T/2}^{f+T/2} \mathbf{L}(t) dt
\]

\[
= \frac{\mathbf{L}(f)}{T} \int_{f-T/2}^{f+T/2} \exp \left( c \mathbf{E}(t) \right) dt
\]

\[
= \frac{\mathbf{L}(f)}{T} \int_{f-T/2}^{f+T/2} \exp \left( c \int_{t} e(s) ds \right) dt,
\]

In this manner, we construct the relation between the captured blurry image \( \mathbf{B} \) and the latent image \( \mathbf{L}(f) \) through the double integral of the event. We name Eq. 4 the Event-based Double Integral (EDI) model.

We denote

\[
\mathbf{J}(c) = \frac{1}{T} \int_{f-T/2}^{f+T/2} \exp(c \mathbf{E}(t)).
\]

Taking the logarithm on both sides of Eq. 4 and rearranging, yields

\[
\tilde{\mathbf{L}}(f) = \mathbf{\tilde{B}} - \log \left( \frac{1}{T} \int_{f-T/2}^{f+T/2} \exp(c \mathbf{E}(t)) dt \right)
\]

\[
\tilde{\mathbf{L}}(f) = \mathbf{\tilde{B}} - \mathbf{\tilde{J}}(c).
\]

which shows a linear relation between the blurry image, the latent image and the integral of the events in the log space.

### 4 Using More Than One Frames

If two or more blurry images are available, then we can solve the equations directly, in least-squares, or some method for solving an over-constrained system. In particular, if there are \( n \) frames, with \( N \) pixels, then this gives \( Nn \) equations in \( N \) unknowns, assuming that \( c \) is known. If \( c \) is not known, there are sufficiently many equations, but they are not linear in \( c \).

#### 4.1 mEDI Model

Consider now an event camera supplying a continuing sequence of events, and also images, \( \mathbf{B}_i \) for \( i = 0, \ldots, n \) called the ‘blurry images’. Suppose that the image \( \mathbf{B}_i \) is captured between time \([f_i - T/2, f_i + T/2]\). Therefore, we write \( \mathbf{E}(t) \) and \( \mathbf{J}(c) \) as

\[
\mathbf{E}_i(t) = \int_{t}^{f_i} e(s) ds
\]

\[
\mathbf{J}_i(c) = \frac{1}{T} \int_{f_i - T/2}^{f_i + T/2} \exp(c \mathbf{E}_i(t)) dt
\]

Then EDI model in section 3 becomes:

\[
\tilde{\mathbf{L}}_i(t) = \tilde{\mathbf{L}}_i(f) + c \mathbf{E}_i(t)
\]

\[
\tilde{\mathbf{B}}_i(f) = \tilde{\mathbf{L}}_i(f) + \tilde{\mathbf{J}}_i(c)
\]

Let \( \mathbf{L}_i(t) \) denote the \( i^{th} \) frame in the video reconstructed from the \( i^{th} \) blurry image. We use \( \mathbf{L}_i \) and \( \mathbf{B}_i \) to represent \( \tilde{\mathbf{L}}_i(f) \) and \( \tilde{\mathbf{B}}_i(f) \) separately in the following section.

\[
\tilde{\mathbf{L}}_i = \tilde{\mathbf{B}}_i + a_i
\]

\[
\tilde{\mathbf{L}}_{i+1} = \tilde{\mathbf{L}}_i + c \mathbf{E}(f_{i+1})
\]

\[
\tilde{\mathbf{L}}_i = \tilde{\mathbf{B}}_i + a_i
\]

Then we would have the new mEDI model based on multiple images and their events.

\[
\tilde{\mathbf{L}}_i = \tilde{\mathbf{B}}_i - a_i
\]

The knowns are \( \tilde{\mathbf{B}}_i \), whereas the unknowns are the \( \tilde{\mathbf{L}}_i, a_i \) and \( \tilde{a}_i \). There is a linear set of equations in the \( \tilde{\mathbf{L}}_i, \) the equations become

\[
\begin{bmatrix}
1 & -1 & -1 & \cdots & -1 \\
1 & \ddots & \ddots & \ddots & \ddots \\
1 & \ddots & \ddots & \ddots & \ddots \\
1 & \ddots & \ddots & \ddots & \ddots \\
1 & \ddots & \ddots & \ddots & \ddots \\
\end{bmatrix}
\begin{bmatrix}
\tilde{\mathbf{L}}_0 \\
\tilde{\mathbf{L}}_1 \\
\vdots \\
\tilde{\mathbf{L}}_n \\
\end{bmatrix}
= 
\begin{bmatrix}
-b_0 \\
-b_1 \\
\vdots \\
-b_{n-1} \\
\tilde{\mathbf{B}}_0 - a_0 \\
\tilde{\mathbf{B}}_1 - b_1 + b_0 \\
\vdots \\
\tilde{\mathbf{B}}_{n-1} - a_{n-1} - b_{n-1} + b_{n-2} \\
\tilde{\mathbf{B}}_n - a_n + b_{n-1}
\end{bmatrix}
\]

Expanding this out specifically, gives

\[
\begin{bmatrix}
2 & -1 & \ddots & \ddots & \ddots & \ddots & 1 \\
-1 & 3 & 1 & \ddots & \ddots & \ddots & 2 \\
\ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\ddots & \ddots & \ddots & 1 & -1 & 3 & 1 \\
-1 & -1 & 3 & 1 & 2 & \ddots & \ddots \\
\end{bmatrix}
\begin{bmatrix}
\tilde{\mathbf{L}}_0 \\
\tilde{\mathbf{L}}_1 \\
\vdots \\
\tilde{\mathbf{L}}_n \\
\end{bmatrix}
= 
\begin{bmatrix}
\tilde{\mathbf{B}}_0 - a_0 - b_0 \\
\tilde{\mathbf{B}}_1 - a_1 - b_1 + b_0 \\
\vdots \\
\tilde{\mathbf{B}}_{n-1} - a_{n-1} - b_{n-1} + b_{n-2} \\
\tilde{\mathbf{B}}_n - a_n + b_{n-1}
\end{bmatrix}
\]

This is a particularly easy set of equations to solve. Since it has to be solved for each pixel, it is important to do it efficiently.
4.2 LU Decomposition

The best way is to take the LU decomposition of the left-hand-side matrix, which has a particularly simple form. The LU decomposition of the above (with appropriate recording of rows) is given by

$$ LU = \begin{pmatrix} -2 & -5 & -13 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & 0 & \cdots & \cdots & -1 \\ \end{pmatrix} \begin{pmatrix} -1 & 3 & -1 & \cdots & \cdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -1 & 3 & -1 \\ \cdots & \ddots & \cdots & -1 & 2 \\ \end{pmatrix} $$

more precisely, if the Fibonacci sequence is $1, 2, 3, 5, 8, \cdots$ and $\phi_k$ denotes the $k$-th entry of this sequence (thus $\phi(2) = 2$), then the top line of the left-hand matrix is

$$ \begin{pmatrix} \phi_2 & \phi_4 & \cdots & \phi_{2(n-1)} & 1 \end{pmatrix}, $$

consisting of the even numbered entries of the Fibonacci sequence, and the entry at the bottom left of the right-hand matrix is $\phi_{2(n+1)}$, the next odd-numbered Fibonacci number, which is also the determinant of the original matrix.

Solving equations by LU decomposition by back-substitution is particularly simple in this case. The procedure in solving equations $LUx = r$ is done by solving

$$ Ly = r $$

$$ Ux = y $$

The solution of $Ly = r = (r_1, r_2, \cdots, r_n)^T$ is simply

$$ y = (r_1, r_2, \cdots, r_n, \sum_{i=1}^{n-1} r_i \phi_{2i})^T $$

The solution of $Ux = y$ is given by back-substitution from the bottom:

$$ x_n = y_n / \phi_{2n+1} = \sum_{i=1}^{n-1} r_i \phi_{2i} / \phi_{2n+1} $$

$$ x_{n-1} = 2x_n - r_n $$

$$ x_{n-2} = 3x_{n-1} - x_n - r_{n-1} $$

$$ x_{n-3} = 3x_{n-2} - x_{n-1} - r_{n-2} $$

$$ \cdots $$

4.3 High Frame-Rate Video Generation

The right-hand side of Eq. (8) is known, apart from perhaps the value of the contrast threshold $c$, the first term from the grey-scale image, the second term from the event sequence, it is possible to compute $L_i$, and hence $L_i$ by exponentiation. Subsequently, from Eq. (6) the latent image $L_i(t)$ at any time may be computed.

To avoid accumulated errors of constructing a video from many frames of a blurred video, it is more suitable to construct each frame $L_i(t)$ using the closest blurred frame. Theoretically, we could generate a video with frame-rate as high as the DVS’s eps (events per second). However, as each event carries little information and is subject to noise, several events must be processed together to yield a reasonable image. We generate a reconstructed frame every 50-100 events, so for our experiments, the frame-rate of the reconstructed video is usually 200 times greater than the input low frame-rate video. Furthermore, as indicated by Eq. (8), the challenging blind motion deblurring problem has been reduced to a single variable optimization problem of how to find the best value of the contrast threshold $c$. In the following section, we use $L_i(c, t)$ to present the latent sharp image $L_i(t)$ with different $c$.

5 Optimization

The unknown contrast threshold $c$ represents the minimum change in log intensity required to trigger an event. By choosing an appropriate $c$ in Eq. (8), we can generate a sequence of sharper images. Here, we propose two different methods to estimate the unknown variable $c$: manually chosen and automatically optimized.

5.1 Manually Chosen $c$

According to our mEDI model in Eq. (8), given a value for $c$, we can obtain a sharp image. Therefore, we develop a
Quantitative comparisons on the Synthetic dataset [38]. This dataset provides videos can be used to generate not only blurry images but also event data. All methods are tested under the same blurry condition, where methods [1], [2], [38], [41] use GoPro dataset [38] to train their models. Jin [2] achieves their best performance when the image is down-sampled to 45% mentioned in their paper.

### Table 1

| Method          | PSNR(dB) | SSIM     |
|-----------------|---------|----------|
| Fan et al. [29] | 23.50   | 0.8336   |
| Sun et al. [30] | 25.30   | 0.8511   |
| Gong et al. [37]| 26.05   | 0.8632   |
| Jin et al. [2]  | 26.98   | 0.8922   |
| Tao et al. [1]  | 30.26   | 0.9342   |
| Zhang et al. [41]| 29.18  | 0.9306   |
| Nah et al. [38] | 29.08   | 0.9135   |
| Ours            | 29.06   | 0.9430   |

Average result of the deblurred images on dataset [38]

### Table 2

| Method          | PSNR(dB) | SSIM     |
|-----------------|---------|----------|
| Baseline 1 [1] + [3] | 25.52   | 0.7685   |
| Baseline 2 [3] + [1] | 26.34   | 0.8090   |
| Scheerlinck et al. [3] | 25.84   | 0.7904   |
| Jin et al. [2]  | 25.62   | 0.8556   |
| EDI             | 28.49   | 0.9199   |

Average result of the reconstructed videos on dataset [38]

Fig. 6. An example of the reconstructed result on our synthetic event dataset based on the GoPro dataset [38]. [38] provides videos to generate the blurry images and event data. (a) The blurry image. The red close-up frame is for (b)-(e), the yellow close-up frame is for (f)-(g). (b) The deblurring result of Jin et al. [2]. (c) Our deblurring result. (d) The crop of their reconstructed images and the frame number is fixed at 7. Jin et al. [2] uses the GoPro dataset added with 20 scenes as training data and their model is supervised by 7 consecutive sharp frames. (e) The crop of our reconstructed images. (f) The crop of Reinbacher [8] reconstructed images from only events. (g) The crop of Scheerlinck [3] reconstructed image, they use both events and the intensity image. For (e)-(g), the shown frames are the chosen examples, where the length of the reconstructed video is based on the number of events. (Best viewed on screen).

5.2 Automatically Chosen $c$

To automatically find the best $c$, we need to build an evaluation metric (energy function) that can evaluate the quality of the deblurred image $L_i(c, t)$. Specifically, we propose to exploit different prior knowledge for sharp images and the event data.

#### 5.2.1 Energy function

For an unknown parameter $c$, the values on the right-hand side of Eq. (8) are dependent on $c$. In particular, we may write

$$b_i = c \int_{f_i} e(s)ds$$

$$a_i = \log \left( \frac{1}{T} \int_{f_i - T/2}^{f_i + T/2} \exp(c E(t))dt \right)$$

(11)
Fig. 7. Examples of reconstruction result on real event dataset. (a) The intensity image from the event camera. (b) Reconstruction result of Pan et al. [4] from combining events and a single blurry frame. (c) Our reconstruction result from combining events and multiple blurry frame. The new mEDI model is based on multiple images and can get better results compared with our previous one based a single image. (Best viewed on screen).

Refer to eq. 9, we denote it as \(Ax = r(c)\), the least-squares solution is given by solving \(A^TAx = A^Tr(c)\).

\[
\hat{x} = (A^TA)^{-1}A^Tr(c).
\]

The optimal \(c\) can be estimate by solving Eq. (12),

\[
\min_c ||Ax - r(c)||,
\]

Examples show that as a function of \(c\), the residual error in solving the equations is not convex. However, in most cases (empirically) it seems to be convex, or at least it has a single minimum.

5.2.2 Fibonacci search

Finding the minimum of a function along a single line is easy if that function has a single minimum. In the case of least-squares minimization problems, various strategies for determining the line-search direction are currently used, such as conjugate gradient methods, gradient descent, Levenberg-Marquardt.

The minimum may be reliably located using Fibonacci search [42]. In this work, we use Fibonacci search for the value of \(c\) that gives the least error is indicated. It helps to reduce the running time three times fast than using gradient descent in our case. In Fig. 5, we illustrate the clearness of the reconstructed image against the value of \(c\). Meanwhile, we also provide the PSNR of the corresponding reconstructed image. As demonstrated in the figure, our proposed reconstruction metric could locate/identify the best-deblurred image with peak PSNR properly.

6 EXPERIMENT

6.1 Experimental Setup

Synthetic dataset. In order to provide a quantitative comparison, we build a synthetic dataset based on the GoPro blurry dataset [38]. It supplies ground truth videos which are used to generate the blurry images. Similarly, we employ the ground-truth images to generate event data based on the methodology of event camera model.

Real dataset. We evaluate our method on a public Event-Camera dataset [40], which provides a collection of sequences captured by the event camera for high-speed robotics. Furthermore, we present our real blurry event dataset, where each real sequence is captured with the DAVIS [6] under different conditions, such as indoor, outdoor scenery, low lighting conditions, and different motion patterns (e.g., camera shake, objects motion) that naturally introduce motion blur into the APS intensity images.

Implementation details. For all our real experiments, we use the DAVIS that shares photosensor array to simultaneously output events (DVS) and intensity images (APS). The framework is implemented by using MATLAB with C++ wrappers. It takes around 1.5 second to process one image on a single i7 core running at 3.6 GHz.

6.2 Experimental Results

We compare our proposed approach with state-of-the-art blind deblurring methods, including conventional deblurring methods [29], [30], deep based dynamic scene deblurring methods [1], [2], [36], [38], [41], and event-based image reconstruction methods [3], [8]. Moreover, Jin et al. [2] can

1. To be released with codes
Fig. 8. Deblurring and reconstruction results on our real blurry event dataset. (a) Input blurry images. (b) Deblurring result of [2]. (c) Baseline 1 for our method. We first use the state-of-the-art video-based deblurring method [2] to recover a sharp image. Then use the sharp image as input to a state-of-the-art reconstruction method [3] to get the intensity image. (d) Baseline 2 for our method. We first use method [3] to reconstruct an intensity image. Then use a deblurring method [2] to recover a sharp image. (e) Samples from our reconstructed video from $L(0)$ to $L(150)$. (Best viewed on screen).

In order to prove the effective of our model, we show some baseline comparisons in Fig. 8 and Table 1. For baseline 1, we first apply a state-of-the-art deblurring method [1] to recover a sharp image, and then the recovered image as an input is then fed to a reconstruction method [3]. For baseline 2, we first use the video reconstruction method to construct a sequence of intensity images, and then apply the deblurring method to each frame. As seen in Table 1, our approach obtains higher PSNR and SSIM in comparison to both baseline 1 and baseline 2. This also implies that our approach better exploits the event data to not only recover sharp images but also reconstruct high frame-rate videos.

In Table 1 and Fig. 6, we show the quantitative and qualitative comparisons with the state-of-the-art image deblurring approaches [1], [2], [29], [36], [37], [38], [41], and the video reconstruction method [3] on our synthetic dataset, respectively.

As indicated in Table 1, our approach achieves the best performance on SSIM and competitive result on PSNR compared to the state-of-the-art methods, and attains significant performance improvements on high-frame video reconstruction.

In Fig. 6, we compare our generated video frames with the state-of-the-art deblurring methods [1], [2], [29], [38] qualitatively. Furthermore, image reconstruction methods [3], [8] are also included for comparisons. Fig. 6 shows that our method can generate more frames from a single blurry image and the recovered frames are much sharper.

We also report our reconstruction results on the real dataset, including text images and low-lighting images, in Fig. 1, Fig. 3, Fig. 4 and Fig. 9. Compared with state-of-the-art deblurring methods, our method achieves superior results. In comparison to existing event-based image reconstructed methods [3], [8], our reconstructed images are not only more realistic but also contain richer details.

7 LIMITATION

Thought an event camera has record a continuous, asynchronous stream of events that encodes redundant information for our EDI model, there still have some limitations.

1) Extreme lighting changes, such as suddenly turn on/off the light, move from dark indoor scenes to outdoor scenes;
2) Event error accumulation will reduce the quality of reconstructed images. However, we integrate over small time intervals from the centre of the exposure time to mitigate this error.

8 CONCLUSION

In this paper, we propose a multiple Event-based Double Integral (mEDI) model to naturally connect intensity images and event data captured by the event camera, which also takes the blur generation process into account. In this way, our model can be used to not only recover latent sharp images but also reconstruct intermediate frames at high frame-rate. We also propose a simple yet effective method to solve our mEDI model. Due to the simplicity of our optimization process, our method is efficient as well. Extensive experiments show that our method can generate high-quality high frame-rate videos efficiently under different conditions, such as low lighting and complex dynamic scenes.

2. For more deblurring results and high-temporal resolution videos please go to our home page.
Fig. 9. Examples of reconstruction result on our real blurry event dataset in low lighting and complex dynamic conditions (a) Input blurry images. (b) The event information. (c) Deblurring results of [29]. (d) Deblurring results of [30]. (e) Deblurring results of [1]. (f) Deblurring results of [38]. (g) Deblurring results of [2] and they use video as training data. (h) Reconstruction result of [4] from combining events and frames. (i) Reconstruction result of [6] from only events. (j)-(k) Reconstruction results of [3], (j) from only events, (k) from combining events and frames. (l) Our reconstruction result. Results in (c)-(g) show that real high dynamic settings and low light condition is still challenging in the deblurring area. Results in (i)-(j) show that while intensity information of a scene is still retained with an event camera recording, color, and delicate texture information cannot be recovered. (Best viewed on screen).
This research was supported in part by the Australian Research Council through the Australian Centre of Excellence for Robotic Vision CE140100016, the Natural Science Foundation of China grants (61871325, 61420106007, 61671387, 61603303) and the Australian Research Council (ARC) grants (DE140100180, DE180100628).

References

[1] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, “Scale-recurrent network for deep image deblurring,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), June 2018.

[2] M. Jin, G. Meishvili, and P. Favaro, “Learning to extract a video sequence from a single motion-blurred image,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), June 2018.

[3] C. Scheerlinck, N. Barnes, and R. Mahony, “Continuous-time intensity estimation using event cameras,” arXiv e-prints, Nov. 2018. [Online]. Available: https://arxiv.org/abs/1811.00386

[4] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai, “Bringing a blurry frame alive at high frame-rate with an event camera,” arXiv preprint arXiv:1811.10180, 2018.

[5] P. Lichtsteiner, C. Posch, and T. Delbruck, “A 128×128 120 db 15 µs latency asynchronous temporal contrast vision sensor,” IEEE journal of solid-state circuits, vol. 43, no. 2, pp. 566–576, 2008.

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

[7] P. Bardow, A. J. Davison, and S. Leutenegger, “Simultaneous optical flow and intensity estimation from an event camera,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2016, pp. 884–892.

[8] C. Inbrchner, G. Graber, and T. Pock, “Real-time intensity-image reconstruction for event cameras using manifold regularisation,” in British Machine Vis. Conf. (BMVC), 2016.

[9] S. Barua, Y. Miyatani, and A. Veeraraghavan, “Direct face detection and video reconstruction from event cameras,” in IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2016, pp. 1–9.

[10] C. Scheerlinck, N. Barnes, and R. Mahony, “Computing spatial image convolutions for event cameras,” IEEE Robotics and Automation Letters, 2019.

[11] P. A. Shedgimer, K. Shah, D. Kumar, and K. Mitra, “Photorealisti c image reconstruction from hybrid intensity and event based sensor,” arXiv preprint arXiv:1805.06140, 2018.

[12] C. Brandli, L. Muller, and T. Delbruck, “Real-time, high-speed video deconvolution using a frame- and event-based DAVIS sensor,” in IEEE Int. Symp. Circuits Syst. (ISCAS), Jun. 2014, pp. 680–683.

[13] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger, “Interacting maps for fast visual interpretation,” in Int. Joint Conf. Neural Netw. (IJCNN), 2011, pp. 770–776.

[14] H. Kim, A. Handa, R. Benosman, S.-H. Ieng, and A. J. Davison, “Simultaneous mosaicing and tracking with an event camera,” in British Machine Vis. Conf. (BMVC), 2014.

[15] H. Kim, S. Leutenegger, and A. J. Davison, “Real-time 3D reconstruction and 6-DoF tracking with an event camera,” in Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 349–364.

[16] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, “Semi-dense 3d reconstruction with a stereo event camera,” arXiv preprint arXiv:1807.07429, 2018.

[17] H. Rebecq, T. Horstschäfer, G. Gallego, and D. Scaramuzza, “EVO: A geometric approach to event-based 6-DoF parallel tracking and mapping in real-time,” IEEE Robot. Autom. Lett., vol. 2, 2017.

[18] A. Z. Zhu, N. Atanasov, and K. Daniilidis, “Event-based visual inertial odometry,” in IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2017, pp. 5816–5824.

[19] A. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, “Ev-flownet: Self-supervised optical flow estimation for event-based cameras,” in Proceedings of Robotics: Science and Systems, Pittsburgh, Pennsylvania, June 2018.

[20] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, “Asynchronous, photometric feature tracking using events and frames,” in Eur. Conf. Comput. Vis. (ECCV), 2018.