This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

Globally-Optimal Contrast Maximisation for Event Cameras

Xin Peng, *Student Member, IEEE,* Ling Gao, *Student Member, IEEE,* Yifu Wang, *Student Member, IEEE,* and Laurent Kneip, *Member, IEEE*

Abstract—Event cameras are bio-inspired sensors that perform well in challenging illumination conditions and have high temporal resolution. However, their concept is fundamentally different from traditional frame-based cameras. The pixels of an event camera operate independently and asynchronously. They measure changes of the logarithmic brightness and return them in the highly discretised form of time-stamped events indicating a relative change of a certain quantity since the last event. New models and algorithms are needed to process this kind of measurements. The present work looks at several motion estimation problems with event cameras. The flow of the events is modelled by a general homographic warping in a space-time volume, and the objective is formulated as a maximisation of contrast within the image of warped events. Our core contribution consists of deriving globally optimal solutions to these generally non-convex problems, which removes the dependency on a good initial guess plaguing existing methods. Our methods rely on branch-and-bound optimisation and employ novel and efficient, recursive upper and lower bounds derived for six different contrast estimation functions. The practical validity of our approach is demonstrated by a successful application to three different event camera motion estimation problems.

Index Terms—Event Cameras, Motion Estimation, Optical Flow, Contrast Maximisation, Global Optimality, Branch and Bound.

1 Introduction

Visual perception plays an increasingly important role in a number of fields such as robotics, smart vehicles, and augmented/virtual reality (AR/VR). These are broad and complex areas of application that require the solution of a variety of problems including but not limited to image matching [1], [2], camera motion estimation [3], [4], [5], localisation [6], 3D reconstruction [7], [8], [9], [10], and object segmentation [11], [12]. Over the past several decades, the community has achieved great progress in traditional camera based solutions to these problems [13], [14]. Their robust application to real-world problems remains nonetheless difficult, which is—at least partially—due to the fact that traditional camera measurements are easily affected in situations of high dynamics, low texture or structure distinctiveness, and challenging illumination conditions.

Event cameras—also called Dynamic Vision Sensors (DVS)—represent an interesting alternative to traditional cameras pairing High Dynamic Range (HDR) with high temporal resolution. Different from standard cameras which capture intensity images at a certain frame rate, the pixels of an event camera sense changes of the logarithmic brightness and operate asynchronously and independently of one-another. An *event* is triggered when the absolute difference between the current logarithmic intensity and the one at the time of the most recent event surpasses a given threshold. Due to their special bio-inspired design, event cameras have very low latency ($\sim 1\mu s$) and very low power consumption. Moreover, event cameras possess a high dynamic range (e.g. 140 dB compared to 60 dB for standard cameras) [15].

Although event cameras have the potential to outperform standard cameras in challenging scenarios, the particular asynchronous and discretised nature of their outputs makes it difficult to directly migrate traditional computer vision algorithms to an event-based vision problem. Hence novel algorithms are needed. Recently, Gallego et al. [16] introduced a unifying contrast maximisation (CM) pipeline with applications to various event-based vision problems, such as motion estimation, 3D reconstruction and optical flow estimation. The core idea of contrast maximisation is given by modelling trajectories in a space-time volume for the high-gradient points that generate events. Based on the assumption that the texture is given by a sparse set of sharp edges, the optimal motion parameters are found when the events exhibit maximum alignment with as few as possible point trajectories. Practically, the quality of the alignment is simply judged by measuring the contrast in the so-called Image of Warped Events (IWE). The objective has been successfully used for solving a variety of problems with event cameras such as optical flow [16], [17], [18], [19], [20], [21], moving object segmentation [18], [22], [23], 3D reconstruction [19], [20], [24], [25], and camera motion estimation [16], [26]. However, existing methods mostly rely on local optimisation of the generally non-convex contrast maximisation objective (cf. Fig. 2), and thus fail if no good initial guess is given.

In this work, we present an efficient, globally-optimal contrast maximisation framework (GOCMF) based on the Branch-and-Bound (BnB) optimisation paradigm (epsilon-optimal solution). The work is inspired by our previous

- *X. Peng is with the Mobile Perception Lab of School of Information Science and Technology, ShanghaiTech University, and Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, and University of Chinese Academy of Sciences.*
- *L. Gao and L. Kneip are with the Mobile Perception Lab of School of Information Science and Technology, ShanghaiTech University.*
- *Y. Wang is with the Australian National University.*
- *E-mail: see http://mpl.sist.shanghaitech.edu.cn*

Manuscript received October 1, 2020; revised November 1, 2020.

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

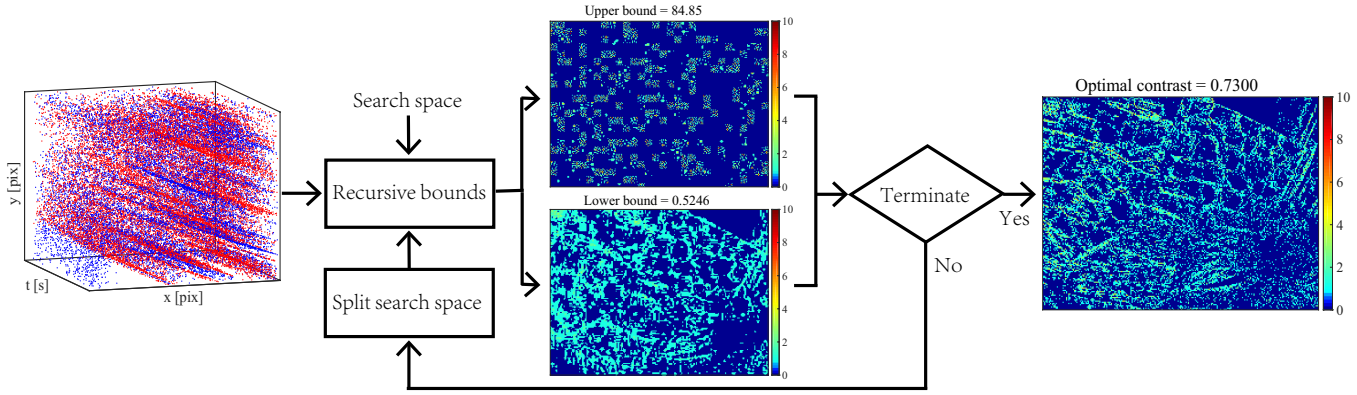


Fig. 1. Globally-Optimal Contrast Maximisation Framework (GOCMF): Given a spatiotemporal event stream $\mathcal{E}$ and a parameter search space Θ, our method applies the branch-and-bound optimisation paradigm with a recursive evaluation on upper and lower bounds' values. The search space is split until upper and lower bounds' values converge, upon which the algorithm returns the globally optimal motion parameters of the considered contrast maximisation problem.

work [27] and relies on recursively evaluated upper and lower bounds of the optimisation objective. These bounds are paramount for the accuracy and efficiency of the algorithm, and we present their full derivation for a total of six different focus loss functions. We furthermore present an application of GOCMF to three different image or camera motion estimation problems. The work also shares commonalities with Liu et al.'s [28] globally optimal event camera rotation estimation algorithm, which however does not calculate the bounds recursively and thus is significantly slower.

Our detailed contributions are as follows:

- We propose a globally optimal contrast maximisation framework—GOCMF—which solves the contrast maximisation problem via Branch and Bound (BnB).
- We derive bounds for six different contrast evaluation functions. The bounds are furthermore calculated recursively, which enables efficient processing.
- We successfully apply this strategy to three common computer vision problems: optical flow estimation, camera rotation estimation, and motion estimation with a downward-facing event camera.

The paper is organised as follows: Section 2 reviews related literature on event-based vision and applications of BnB in computer vision. Section 3 reviews the general idea of contrast maximisation for event cameras. In Section 4, we then introduce our globally optimal contrast maximisation framework and derive the recursive upper and lower bounds. Sections 5.1, 5.2 and 5.3 then present the application to optical flow estimation, downward-facing event camera motion estimation, and camera rotational estimation, respectively. To conclude, Sections 6 and 7 discuss a further analysis of the proposed algorithm as well as final remarks.

2 Related Work

Event cameras (e.g DVS) became commercially available in 2008, and there have been an increasing number of works on event-based computer vision problems in recent years. We first introduce existing works on event-based vision as well as specific algorithms that employ contrast maximisation as a core objective to be optimised for the solution of several different event-based vision problems. We furthermore review existing literature on the application of BnB for the solution of such problems with normal cameras.

2.1 Event-based vision

Most works for event-based motion estimation, focus on homography scenarios [26], [29], [30], [31] or fusion with other sensors [32], [33], [34]. Zhu et al. [35] propose visual odemetry (VO) with an event camera and known map by feature tracking and PnP methods. Zhou et al. [36] come up with the first parallel tracking-and-mapping VO with a stereo event camera. A spiking neural network is adopted by Gehrig et al. [37] for regression of angular velocity. As for optical flow estimation, Benosman et al. [38] estimate normal flow by assuming the events are locally planar. Bardow and Davison [39] show a generic optical flow estimation method which also outputs image intensity by minimising a loss function with smoothness regularisation. Stoffregen and Kleeman [18] employ contrast maximisation to show a simultaneous optical flow and segmentation algorithm. Moreover, learning-based approaches recently became popular to be applied to event-based optical flow estimation [20], [40], [41].

Event cameras also have been used for other computer vision tasks: object tracking [42], [43], [44], pattern recognition [45], [46], [47], 3D reconstruction [48], [49], [50], intensity image reconstruction [51], [52], [53] and so on.

2.2 Event-based vision with contrast maximisation

Gallego and Scaramuzza [26] introduce the contrast of the image of warped events as a valid objective to be maximised in image registration problems. Their method estimates accurate angular velocities even in the presence of high-speed motion. At the same time, Zhu et al. [17] come up with an expectation-maximisation based data tracking approach to find the best alignment of warped events. Rebecq et al. [24], [48] propose Event-based Multi-View Stereo (EMVS) to

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

estimate semi-dense 3D structure from an event camera with known poses. In their later work, they extend it to a real-time 6-DOF tracking and mapping system—Event-based Visual Odometry (EVO) [54]. The contrast maximisation concept is concluded in [16] with applications to motion, depth, and optical flow estimation. Mitrokhin et al. [23] also ultilize a parametric model to estimate the motion of the camera via motion compensation. Moving objects are detected during an iterative process that identifies events that are inconsistent with the primary displacement model. Zhu et al. [20] adopt contrast maximisation as a loss function to train an unsupervised neural network to estimate optical flow, depth and egomotion. Various reward functions that maximise contrast have been presented and analysed in the recent works of Gallego et al. [15] and Stoffregen and Kleeman [55]. However, practically all aforementioned papers apply local contrast maximisation, only.

Liu et al. [28] propose a globally optimal contrast maximisation algorithm via BnB for camera rotation estimation. They derive the upper bound of the contrast objective by relaxing the sum of squares maximisation to an integer quadratic program. Nevertheless, time consumption is a severe problem for globally optimal methods. Our work introduces an alternative globally optimal contrast maximisation framework based on a more efficient, recursively evaluated upper bound. We present the detailed derivations for six different objective functions, and furthermore extend our previous work [27] to optical flow, planar motion and rotational motion estimation.

2.3 Branch and bound in computer vision

Branch-and-bound (BnB) is an optimisation strategy that guarantees global optimality without requiring any priors. There are quite a few solutions to geometric computer vision problems that are grounded on BnB, such as 2D-2D registration [56], [57], 3D-3D registration [58], [59], [60], [61], [62], [63], [64], [65], [66], 2D-3D registration [67], [68], [69], [70], [71], [72], and relative pose estimation [73], [74], [75], [76], [77], [78], [79] methods. One of the earliest applications is proposed by Breuel [56] who analyses its implementation and derives various bounds for 2D-2D point registration. Another pioneering application of BnB is given by Li and Hartley's [58] global 3D point registration method with unknown correspondences, which uses Lipschitz optimisation to search the space of 3D rotations. More modern, practically usable globally optimal 3D-3D registration methods are given by Go-ICP [63] and GOGMA [62], which use points and Gaussian Mixture Models (GMAs) to represent 3D structure. Bustos et al. [64] propose novel bounds to achieve faster rotation search based on cardinality maximisation. Liu et al. [65] propose a new rotation invariant feature (RIF) that allows a prior, efficient, globally optimal estimation of the translation. More recently, Hu et al. [66] propose a simultaneous estimation of a symmetry plane for 3D-3D registration of partial scans with limited overlap. Brown et al. [67], [71] utilize the BnB paradigm for 2D-3D registration without known correspondences. Their globally optimal approach is applicable to both point and line features. Campbell et al. [68], [69] finally present a similar global inlier set cardinality maximisation method for simultaneous pose and correspondence estimation, however introduce novel tighter and approximation-free bounds. Most recently, Liu et al. [72] and Campbell et al. [70] present further solutions to the 2D-3D problem by again relying on density distribution mixtures. With respect to relative pose estimation, Yang et al. [78] extend the globally optimal *rotation space search* by Hartley et al. [73] to essential matrix estimation in the presence of feature mismatches or outliers. Zheng et al. [77] furthermore use BnB for finding a globally optimal fundamental matrix, however with an objective function that assumes no outliers. Ling et al. [79] most recently introduce globally optimal homography matrix estimation for featureless scenarios.

Although BnB has seen many past use cases in geometric vision, our application to event-camera based motion estimation relies on the substantially different objective of contrast maximisation, for which we derive novel recursive upper and lower bounds.

3 Contrast Maximisation

Gallego et al. [16] recently introduced contrast maximisation as a unifying framework allowing the solution of several important problems for dynamic vision sensors, in particular motion estimation problems in which the effect of camera motion may be described by a homography (e.g. motion in front of a plane, pure rotation). Our work relies on contrast maximisation, which we therefore briefly review in the following.

The core idea of contrast maximisation is relatively straightforward: the flow of the events is modelled by a time-parametrised homography. An event camera outputs a sequence of *events* denoting temporal logarithmic brightness changes above a certain threshold. An event $e = \{\mathbf{x},\ t,\ s\}$ is described by its pixel position $\mathbf{x} = [x\ y]^\mathsf{T}$, timestamp t, and polarity s (the latter indicates whether the brightness is increasing or decreasing, and is ignored in the present work). Given its position and time-stamp, every event may therefore be warped back along a point-trajectory into a reference view with timestamp t_{ref}. Since events are more likely to be generated by high-gradient edges, the correct homographic warping parameters are found when the unwarped events align along an as sharp as possible edge-map in the reference view, i.e. the Image of Warped Events (IWE). Gallego et al. [16] simply propose to consider the contrast of the IWE as a reward function to identify the correct homographic warping parameters. Note that homographic warping functions include 2D affine and Euclidean transformations, and thus can be used in a variety of vision problems such as optical flow, feature tracking, or fronto-parallel motion estimation.

Suppose we are given a set of N events $\mathcal{E} = \{e_k\}_{k=1}^{N}$. We define a general warp function $\mathbf{x}'_k = W(\mathbf{x}_k, t_k; \boldsymbol{\theta})$ that returns the position $\mathbf{x}'_k$ of an event e_k in the reference view at time t_{ref}. $\boldsymbol{\theta}$ is a vector of warping parameters. The IWE is generated by accumulating warped events at each discrete pixel location:

$$I(\mathbf{p}_{ij}; \boldsymbol{\theta}) = \sum_{k=1}^{N} \mathbf{1}(\mathbf{p}_{ij} - \mathbf{x}'_k) = \sum_{k=1}^{N} \mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_k, t_k; \boldsymbol{\theta}))\,, \tag{1}$$

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

where $\mathbf{1}(\cdot)$ is an indicator function that counts 1 if the absolute value of $(\mathbf{p}_{ij} - \mathbf{x}'_k)$ is less than a threshold ϵ in each coordinate, and otherwise 0. $\mathbf{p}_{ij}$ is a pixel in the IWE with coordinates $[i\ j]^\mathsf{T}$, and we refer to it as an *accumulator* location. We set $\epsilon = 0.5$ such that each warped event will increment one accumulator only.

Existing approaches replace the indicator function with a Gaussian kernel to make the IWE a smooth function of the warped events, and thus solve contrast maximisation problems via local optimisation methods (cf. [15], [16], [26]). In contrast, we propose a method that is able to find the global optimum of the above, discrete objective function.

As introduced in [15], [55], reward functions for event un-warping all rely on the idea of maximising the contrast or sharpness of the IWE (they are also denoted as *focus loss functions*). They proceed by integration over the entire set of accumulators, which we denote $\mathcal{P}$. The most relevant six objective functions for us are introduced in Section 4 and Table 1.

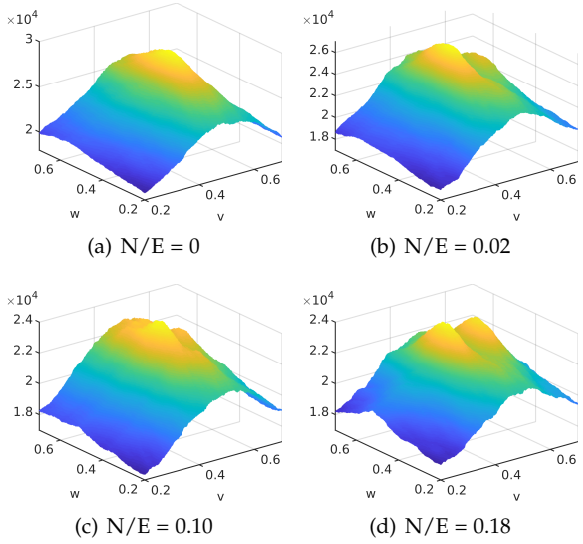


Fig. 2. Visualization of the sum-of-squares contrast function. The camera is moving in front of a plane, and the motion parameters are given by translational and rotational velocity (cf. Section 5). The sub-figures from (a) to (d) are functions with increasing Noise-to-Events (N/E) ratios. Note that contrast functions are non-convex.

4 Globally Maximised Contrast using Branch and Bound

Fig. 2 illustrates how contrast maximisation for motion estimation is in general a non-convex problem, meaning that local optimisation may be sensitive to the initial parameters and not find the global optimum. We tackle this problem by introducing a globally optimal solution to contrast maximisation using Branch and Bound (BnB) optimisation. BnB is an algorithmic paradigm in which the solution space is subdivided into branches in which we then find upper and lower bounds for the maximal objective value. The globally optimal solution is isolated by an iterative search in which entire branches are discarded if their upper bound for the maximum objective value remains lower than the lower bound in another branch. The most important factor deciding the effectiveness of this approach is given by the tightness of the bounds.

Our core contribution is given by a recursive method to efficiently calculate upper and lower bounds for the maximum value of a contrast maximisation function over a given branch. In short, the main idea is given by expressing a bound over $(N+1)$ events as a function of the bound over N events plus the contribution of one additional event. The strategy can be similarly applied to all six relevant contrast functions, which is why we limit the exposition in Section 4.2 to the derivation of bounds for the objective function "sum of squares (L_{SoS})". Detailed derivations for all further loss functions are provided in Section 4.3.

4.1 Objective Function

In the following, we assume that $L_N = L_{\mathrm{SoS}}$. The maximum objective function value over all N events in a given time interval $[t_{\mathrm{ref}}, t_{\mathrm{ref}} + \Delta T]$ is given by

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_N = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{\mathbf{p}_{ij}\in\mathcal{P}} \left[\sum_{k=1}^{N} \mathbf{1}\left(\mathbf{p}_{ij} - W(\mathbf{x}_k, t_k; \boldsymbol{\theta})\right)\right]^2 , \quad (2)$$

where $\boldsymbol{\Theta}$ is the search space (i.e. branch or sub-branch) over which we want to maximise the objective. For alternative classical problems the measurement samples can be evaluated in parallel (e.g. inlier/outlier decisions for pairwise correspondences), whereas the contrast of the IWE is a measure that depends on all events. Note furthermore that events are processed in a given order, which—in our work—is set to the temporal order in which the events are captured. Intuitively speaking, our discrete problem can be described as finding optimal parameters that allocate N events to a grid pattern such that the sum of squares of each cell's value is maximised (cf. Fig. 3).

4.2 Upper and Lower Bound

We calculate the bounds recursively by processing the events one-by-one, each time updating the IWE. The events are notably processed in temporal order with increasing timestamps.

For the lower bound, it is readily given by evaluating the contrast function at an arbitrary point on the search space interval $\boldsymbol{\Theta}$, which is commonly picked as the interval center $\boldsymbol{\theta}_0$. We present a recursive rule to efficiently evaluate the lower bound.

Theorem 1. *For search space $\boldsymbol{\Theta}$ centered at $\boldsymbol{\theta}_0$, the lower bound of SoS-based contrast maximisation may be given by*

$$\underline{L_N} = \underline{L_{N-1}} + 1 + 2I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)\,, \quad (3)$$

where $I^{N-1}(\mathbf{p}_{ij}; \boldsymbol{\theta}_0)$ is the incrementally constructed IWE, its exponent $(N-1)$ (where $N \geq 1$) denotes the number of events that have already been taken into account, and

$$\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0} = \mathrm{round}(W(\mathbf{x}_N, t_N; \boldsymbol{\theta}_0)) \quad (4)$$

returns the accumulator closest to the warped position of the N-th event.

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 1
Recursive Upper and Lower Bounds for six focus loss functions

	Upper Bound $\overline{L_N}$	**Lower Bound** $\underline{L_N}$	L_0
SoS	$\overline{L_{N-1}} + 1 + 2Q$	$\underline{L_{N-1}} + 1 + 2I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)$	0
Var	$\overline{L_{N-1}} + \frac{1}{N_p} - \frac{2\mu_I}{N_p} + \frac{2}{N_p}Q$	$\underline{L_{N-1}} + \frac{1}{N_p} - \frac{2\mu_I}{N_p} + \frac{2}{N_p}I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)$	μ_I^2
SoE	$\overline{L_{N-1}} + (e-1)e^Q$	$\underline{L_{N-1}} + (e-1)e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)}$	N_p
SoSA	$\overline{L_{N-1}} + (e^{-\delta}-1)e^{-\delta \cdot Q}$	$\underline{L_{N-1}} + (e^{-\delta}-1)e^{-\delta \cdot I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)}$	N_p
SoEaS	$\overline{L_{N-1}} + w_1(e-1)e^Q + w_2 + 2w_2Q$	$\underline{L_{N-1}} + w_1(e-1)e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)} + w_2 + 2w_2 I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)$	$w_1 N_p$
SoSAaS	$\overline{L_{N-1}} + w_1(e^{-\delta}-1)e^{-\delta Q} + w_2 + 2w_2Q$	$\underline{L_{N-1}} + w_1(e^{-\delta}-1)e^{-\delta I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)} + w_2 + 2w_2 I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)$	$w_1 N_p$

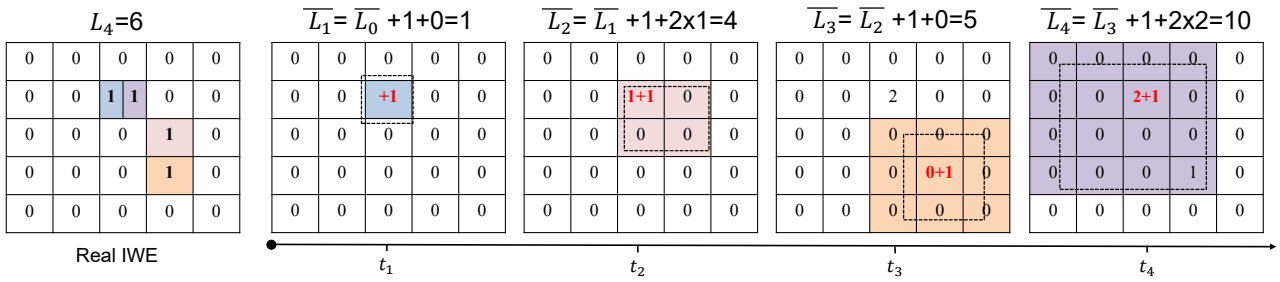


Fig. 3. An example of the incremental update of the upper bound IWE $\overline{I}^N$. For this example, the event number $N = 4$, and different events are indicated by different colors. The left matrix shows the constructed IWE with ground truth motion parameters. The incremental update of the upper bound IWE $\overline{I}^N$ is shown in the right matrices. For each new event e, we choose and increment the currently maximal accumulator in the bounding box $\mathcal{P}^{\boldsymbol{\Theta}}$ (the rectangle bounding the dashed line formed by all possible locations $W(\mathbf{x}, t; \boldsymbol{\theta} \in \boldsymbol{\Theta})$). We simply increment the center of the bounding box if no other accumulator exists. It is easy to see that the upper bound is bigger than the optimal result.

Proof 1. According to the definition of the sum of squares focus loss function,

$$\begin{aligned}\underline{L_N} &= \sum_{\mathbf{p}_{ij}\in\mathcal{P}} \left[\sum_{k=1}^{N} \mathbf{1}\left(\mathbf{p}_{ij} - W(\mathbf{x}_k, t_k; \boldsymbol{\theta}_0)\right)\right]^2 \\ &= \sum_{\mathbf{p}_{ij}\in\mathcal{P}} \left[I^{N-1}(\mathbf{p}_{ij}; \boldsymbol{\theta}_0) + \mathbf{1}\left(\mathbf{p}_{ij} - W(\mathbf{x}_N, t_N; \boldsymbol{\theta}_0)\right)\right]^2 \\ &= a + b + c\,, \text{ where}\end{aligned} \tag{5}$$

$$I^{N-1}(\mathbf{p}_{ij}; \boldsymbol{\theta}_0) = \sum_{k=1}^{N-1} \mathbf{1}\left(\mathbf{p}_{ij} - W(\mathbf{x}_k, t_k; \boldsymbol{\theta}_0)\right)$$

$$a = \sum_{\mathbf{p}_{ij}\in\mathcal{P}} I^{N-1}(\mathbf{p}_{ij}; \boldsymbol{\theta}_0)^2\,,$$

$$b = 2\sum_{\mathbf{p}_{ij}\in\mathcal{P}} \left[\mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_N, t_N; \boldsymbol{\theta}_0)) I^{N-1}(\mathbf{p}_{ij}; \boldsymbol{\theta}_0)\right]\,,$$

$$c = \sum_{\mathbf{p}_{ij}\in\mathcal{P}} \left[\mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_N, t_N; \boldsymbol{\theta}_0))\right]^2\,.$$

It is clear that $a = \underline{L_{N-1}}$. In c, owing to the definition of our indicator function, only the $\mathbf{p}_{ij}$ which is closest to $W(\mathbf{x}_N, t_N; \boldsymbol{\theta}_0)$ makes a contribution, thus we have $c = 1$. For b, the term $\mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_N, t_N; \boldsymbol{\theta}_0))$ is simply zero unless we are considering an accumulator $\mathbf{p}_{ij} = \boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}$, which gives $b = 2I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0)$. Thus we obtain (3). Note that the IWE is iteratively updated by incrementing the accumulator which locates closest to $\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}$.

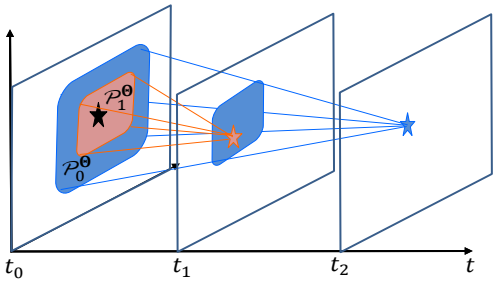


Fig. 4. Bounding boxes of two events generated by a same point. Given two events e_1 (orange) and e_2 (blue) with timestamps t_1 and t_2 generated by a same 3D point, as illustrated in the figure. Uncertainty typically increases with the timestamp, hence the bounding box $\mathcal{P}_1^{\boldsymbol{\Theta}} \subseteq \mathcal{P}_2^{\boldsymbol{\Theta}}$.

We now proceed to our main contribution, a recursive upper bound for the contrast maximisation problem. Let us define $\mathcal{P}_k^{\boldsymbol{\Theta}}$ as the bounding box around all possible locations $W(\mathbf{x}_k, t_k; \boldsymbol{\theta} \in \boldsymbol{\Theta})$ of the un-warped event. Lemma 1 is introduced as follows.

Lemma 1. *Given a search space $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, for a small enough time interval, if $W(\mathbf{x}_i, t_i; \boldsymbol{\theta}) = W(\mathbf{x}_j, t_j; \boldsymbol{\theta})$, $t_i \leq t_j$ and $0 < i < j \leq N$, we have $\mathcal{P}_i^{\boldsymbol{\Theta}} \subseteq \mathcal{P}_j^{\boldsymbol{\Theta}}$. An intuitive explanation is given in Fig. 4. Note that "a small enough interval" here simply denotes an interval for which the constant velocity assumption is sufficiently valid.*

Lemma 1 enables deriving our recursive upper bound.

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

Theorem 2. *The upper bound of the objective function L_N for SoS-based contrast maximisation satisfies*

$$\begin{aligned} L_N &= L_{N-1} + 1 + 2I^{N-1}(\boldsymbol{\eta}_N^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}}) && (6)\\ &\leq \overline{L_{N-1}} + 1 + 2Q^{N-1} = \overline{L_N}, && (7)\\ \text{where } Q^{N-1} &= \max_{\mathbf{p}_{ij} \in \mathcal{P}_N^{\boldsymbol{\Theta}}} \overline{I}^{N-1}(\mathbf{p}_{ij}) \geq I^{N-1}(\boldsymbol{\eta}_N^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}}) \end{aligned}$$

$\mathcal{P}_N^{\boldsymbol{\Theta}}$ is a bounding box for the N-th event. $\hat{\boldsymbol{\theta}}$ is the optimal parameter set that maximises L_N over the interval $\boldsymbol{\Theta}$. $\overline{I}^{N-1}(\mathbf{p}_{ij})$ is the value of pixel $\mathbf{p}_{ij}$ in the upper bound IWE, a recursively constructed image in which we always increment the maximum accumulator within the bounding box $\mathcal{P}_N^{\boldsymbol{\Theta}}$ (i.e. the one that we used to define the value of Q^{N-1}. The incremental construction of $\overline{I}^{N-1}(\mathbf{p}_{ij})$ is illustrated in Fig. 3.

Proof 2. (6) is derived in the same manner as Theorem 1. The proof of inequation (7) then proceeds by mathematical induction.

For $N = 0$, it is obvious that $L_0 = \overline{L_0} = 0$. Similarly, for $N = 1$, $L_1 = 1 \leq \overline{L_0} + 1 + 0$, and $Q^0 = I^0(\boldsymbol{\eta}_1^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}}) = 0$ (which satisfies Theorem 2). We now assume that L_n as well as the corresponding upper bound IWE $\overline{I}^n$ are given for all $0 < n \leq N$. We furthermore assume that they satisfy Theorem 2. Our aim is to prove that (7) holds for the $(N+1)$-th event. It is clear that $\overline{L_N} \geq L_N$, and we only need to prove that $Q^N \geq I^N(\boldsymbol{\eta}_{N+1}^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}})$, for which we will make use of Lemma 1. There are two cases to be distinguished:

- The first case is if there exists an event e_k with $0 < k < N + 1$ and for which $\boldsymbol{\eta}_k^{\hat{\boldsymbol{\theta}}} = \boldsymbol{\eta}_{N+1}^{\hat{\boldsymbol{\theta}}}$. In other words, the k-th and the $(N + 1)$-th event are warped to a same accumulator if choosing the locally optimal parameters. Note that if there are multiple previous events for which this condition holds, the k-th event is chosen to be the most recent one. Given our assumptions, $\overline{L_{k-1}}$ as well as the $(k - 1)$-th constructed upper bound IWE satisfy Theorem 2, which means that $Q^{k-1} \geq I^{k-1}(\boldsymbol{\eta}_k^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}})$. Let $\mathbf{p}_k \in \mathcal{P}_k^{\boldsymbol{\Theta}}$ now be the pixel location with maximum intensity in $\overline{I}^{k-1}(\mathbf{p}_k)$. Then, the k-th updated IWE satisfies $\overline{I}^k(\mathbf{p}_k) = Q^{k-1} + 1 \geq I^{k-1}(\boldsymbol{\eta}_k^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}}) + 1$. According to Lemma 1, we have $\mathcal{P}_k^{\boldsymbol{\Theta}} \subseteq \mathcal{P}_{N+1}^{\boldsymbol{\Theta}}$, therefore $\mathbf{p}_k \subseteq \mathcal{P}_{N+1}^{\boldsymbol{\Theta}}$, and $Q^N \geq \overline{I}^k(\mathbf{p}_k) \geq I^{k-1}(\boldsymbol{\eta}_k^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}}) + 1$. With optimal warp parameters $\hat{\boldsymbol{\theta}}$, events with indices from $k+1$ to N will not locate at $\boldsymbol{\eta}_{N+1}^{\hat{\boldsymbol{\theta}}}$, and therefore $I^{k-1}(\boldsymbol{\eta}_k^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}}) + 1 = I^N(\boldsymbol{\eta}_{N+1}^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}}) \leq Q^N$.
- If there is no such a event, it is obvious that $Q^N \geq I^N(\boldsymbol{\eta}_{N+1}^{\hat{\boldsymbol{\theta}}}; \hat{\boldsymbol{\theta}})$.

With the basic cases and the induction step proven, we conclude our proof that Theorem 2 holds for all natural numbers N.

4.3 Bounds for Other Objective Functions

We further apply the proposed strategy to derive upper and lower bounds for the other five aforementioned contrast functions including variance (L_{Var}), sum of exponentials (L_{SoE}), sum of suppressed accumulations (L_{SoSA}), sum of exponentials and squares (L_{SoEaS}) and sum of suppressed accumulations and squares (L_{SoSAaS}). For convenience, we employ L_N as a unified representation of each function.

Variance (Var)

The variance of an IWE is used in [16], [26] as an objective function. We define $\mu_I = N/N_p$ as the mean value of $I^N(\mathbf{p}_{ij}; \boldsymbol{\theta})$ over all pixels, where N is the number of events and N_p is the total number of accumulators in an image plane. Hence μ_I may be approximated to be constant, which renders L_{SoS} and L_{Var} essentially equivalent (also implied in [26]). The function reads as follows:

$$\begin{aligned} L_N &= \frac{1}{N_p} \sum_{\mathbf{p}_{ij} \in \mathcal{P}} (I^N(\mathbf{p}_{ij}; \boldsymbol{\theta}) - \mu_I)^2 \\ &= \frac{1}{N_p} \sum_{\mathbf{p}_{ij} \in \mathcal{P}} \left[\sum_{k=1}^{N} \mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_k, t_k; \boldsymbol{\theta})) - \mu_I \right]^2 \\ &= L_{N-1} + a + b + c, \end{aligned} \tag{8}$$

where

$$\begin{aligned} a &= \frac{2}{N_p} \sum_{\mathbf{p}_{ij} \in \mathcal{P}} \left\{ I^{N-1}(\mathbf{p}_{ij}; \boldsymbol{\theta}) \mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_N, t_N; \boldsymbol{\theta})) \right\} \\ b &= -\frac{2\mu_I}{N_p} \sum_{\mathbf{p}_{ij} \in \mathcal{P}} \mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_N, t_N; \boldsymbol{\theta})) = -\frac{2\mu_I}{N_p} \\ c &= \frac{1}{N_p} \sum_{\mathbf{p}_{ij} \in \mathcal{P}} \left[\mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_N, t_N; \boldsymbol{\theta})) \right]^2 = \frac{1}{N_p}. \end{aligned} \tag{9}$$

The term $\mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_N, t_N; \boldsymbol{\theta}))$ is simply zero unless we are considering an accumulator $\mathbf{p}_{ij} = \boldsymbol{\eta}_N^{\hat{\boldsymbol{\theta}}}$, which makes b and c in equation (9) constant. Thus, given a search space $\boldsymbol{\Theta}$ centered at $\boldsymbol{\theta}_0$, the lower bound is

$$\underline{L_N} = \underline{L_{N-1}} + \frac{1}{N_p} - \frac{2\mu_I}{N_p} + \frac{2}{N_p} I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0). \tag{10}$$

Furthermore, according to Theorem 2, the upper bound becomes

$$\overline{L_N} = \overline{L_{N-1}} + \frac{1}{N_p} - \frac{2\mu_I}{N_p} + \frac{2}{N_p} Q^{N-1}. \tag{11}$$

Note that the initial upper and lower bounds $\overline{L_0} = \underline{L_0} = L_0 = \mu_I^2$.

Sum of Exponentials (SoE)

The sum of exponentials objective reads

$$\begin{aligned} L_N &= \sum_{\mathbf{p}_{ij} \in \mathcal{P}} e^{I^N(\mathbf{p}_{ij}; \boldsymbol{\theta})} \\ &= \sum_{\mathbf{p}_{ij} \in \mathcal{P}} e^{\sum_{k=1}^{N} \mathbf{1}(\mathbf{p}_{ij} - W(\mathbf{x}_k, t_k; \boldsymbol{\theta}))}. \end{aligned} \tag{12}$$

It sums up the exponential intensity at each pixel in the IWE. Note that the sum of exponentials is generally much bigger

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

than L_{Var} and L_{SoS}. We leverage a similar recursive strategy to calculate the loss

$$\begin{aligned} L_N &= \sum_{\mathbf{p}_{ij}\in\mathcal{P}} e^{\sum_{k=1}^{N-1}\mathbf{1}(\mathbf{p}_{ij}-W(\mathbf{x}_k,t_k;\boldsymbol{\theta}))+\mathbf{1}(\mathbf{p}_{ij}-W(\mathbf{x}_N,t_N;\boldsymbol{\theta}))} \\ &= \sum_{\mathbf{p}_{ij}\in\mathcal{P}} e^{\sum_{k=1}^{N-1}\mathbf{1}(\mathbf{p}_{ij}-W(\mathbf{x}_k,t_k;\boldsymbol{\theta}))} e^{\mathbf{1}(\mathbf{p}_{ij}-W(\mathbf{x}_N,t_N;\boldsymbol{\theta}))} \\ &= \sum_{\substack{\mathbf{p}_{ij}\in\mathcal{P} \\ \mathbf{p}_{ij}\neq\boldsymbol{\eta}_N^{\boldsymbol{\theta}}}} e^{I^{N-1}(\mathbf{p}_{ij};\boldsymbol{\theta})} + a \\ &= \sum_{\mathbf{p}_{ij}\in\mathcal{P}} e^{I^{N-1}(\mathbf{p}_{ij};\boldsymbol{\theta})} - e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})} + a \\ &= L_{N-1} - e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})} + a, \end{aligned} \tag{13}$$

where, according to the property of the indicator function,

$$\begin{aligned} a &= e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})} e^{\mathbf{1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}}-W(\mathbf{x}_N,t_N;\boldsymbol{\theta}))} \\ &= e\cdot e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})}. \end{aligned}$$

Thus, using a similar strategy than in Theorem 1 and Theorem 2, the recursive bounds become

$$\begin{aligned} \underline{L_N} &= \underline{L_{N-1}} + (e-1)e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0};\boldsymbol{\theta}_0)} \\ \overline{L_N} &= \overline{L_{N-1}} + (e-1)e^{Q}. \end{aligned} \tag{14}$$

Note that the initial upper and lower bounds $\overline{L_0} = \underline{L_0} = L_0 = N_p$.

Sum of Suppressed Accumulations (SoSA)

For L_{SoSA}, δ is a design parameter called the *shift factor*. Different from other objective functions, locations with few accumulations will contribute more to the L_{SoSA}. The intuition here is that more empty locations again mean more events that are concentrated at fewer accumulators, and thus a higher-contrast IWE. The focuss loss function reads

$$\begin{aligned} L_N &= \sum_{\mathbf{p}_{ij}\in\mathcal{P}} e^{-I(\mathbf{p}_{ij};\boldsymbol{\theta})\cdot\delta} \\ &= \sum_{\mathbf{p}_{ij}\in\mathcal{P}} e^{-\delta\cdot\sum_{k=1}^{N}\mathbf{1}(\mathbf{p}_{ij}-W(\mathbf{x}_k,t_k;\boldsymbol{\theta}))}. \end{aligned} \tag{15}$$

The derivation is analogous to the derivation for the sum of exponentials, and leads to

$$\begin{aligned} L_N &= \sum_{\mathbf{p}_{ij}\in\mathcal{P}} e^{-\delta\cdot I^{N-1}(\mathbf{p}_{ij};\boldsymbol{\theta})-\delta\cdot\mathbf{1}(\mathbf{p}_{ij}-W(\mathbf{x}_N,t_N;\boldsymbol{\theta}))} \\ &= L_{N-1}(\boldsymbol{\theta}) - e^{-\delta\cdot I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})} + a, \end{aligned} \tag{16}$$

where

$$\begin{aligned} a &= e^{-\delta\cdot I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})} e^{-\delta\cdot\mathbf{1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}}-W(\mathbf{x}_N,t_N;\boldsymbol{\theta}))} \\ &= e^{-\delta} e^{-\delta\cdot I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})}. \end{aligned}$$

Thus, the bounds are

$$\begin{aligned} \underline{L_N} &= \underline{L_{N-1}} + (e^{-\delta}-1)e^{-\delta\cdot I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0};\boldsymbol{\theta}_0)} \\ \overline{L_N} &= \overline{L_{N-1}} + (e^{-\delta}-1)e^{-\delta\cdot Q}. \end{aligned} \tag{17}$$

Note that the initial upper and lower bounds $\overline{L_0} = \underline{L_0} = L_0 = N_p$.

Sum of Exponentials and Squares (SoEaS)

L_{SoEaS} is actually a hybrid function, which is a weighted sum of L_{SoE} and L_{SoS}. We therefore have

$$L_N = w_1 L_{\mathrm{SoE}} + w_2 L_{\mathrm{SoS}}, \tag{18}$$

where w_1 and w_2 are linear combination weights. The equation can still be simplified into recursive form, which leads us to

$$\begin{aligned} L_N &= w_1\left[L_{N-1}^{\mathrm{SoE}} + (e-1)e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})}\right] \\ &+ w_2\left[L_{N-1}^{\mathrm{SoS}} + 1 + 2I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})\right] \\ &= L_{N-1} + w_1(e-1)e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})} \\ &+ w_2 + 2w_2 I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})). \end{aligned} \tag{19}$$

It follows immediately that the bounds are given as

$$\begin{aligned} \underline{L_N} &= \underline{L_{N-1}} + w_1(e-1)e^{I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0};\boldsymbol{\theta}_0)} \\ &+ w_2 + 2w_2 I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0};\boldsymbol{\theta}_0)) \\ \overline{L_N} &= \overline{L_{N-1}} + w_1(e-1)e^{Q} + w_2 + 2w_2 Q. \end{aligned} \tag{20}$$

Note that the initial upper and lower bounds $\overline{L_0} = \underline{L_0} = L_0 = w_1 N_p$. In this paper, we choose $w_1 = 1.0$ and $w_2 = 1.0$.

Sum of Suppressed Accumulations and Squares (SoSAaS)

L_{SoSAaS} is another hybrid function, which is a weighted sum of L_{SoSA} and L_{SoS}, i.e.

$$L_N = w_1 L_{\mathrm{SoSA}} + w_2 L_{\mathrm{SoS}}. \tag{21}$$

w_1 and w_2 are again linear combination weights. The recursive form is given by

$$\begin{aligned} L_N &= w_1\left[L_{N-1}^{\mathrm{SoSA}} + (e^{-\delta}-1)e^{-\delta I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})}\right] \\ &+ w_2\left[L_{N-1}^{\mathrm{SoS}} + 1 + 2I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})\right] \\ &= L_{N-1} + w_1(e^{-\delta}-1)e^{-\delta\cdot I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})} \\ &+ w_2 + 2w_2 I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}};\boldsymbol{\theta})). \end{aligned} \tag{22}$$

It again immediately follows that the bounds are given by

$$\begin{aligned} \underline{L_N} &= \underline{L_{N-1}} + w_1(e^{-\delta}-1)e^{-\delta\cdot I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0};\boldsymbol{\theta}_0)} \\ &+ w_2 + 2w_2 I^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0};\boldsymbol{\theta}_0)) \\ \overline{L_N} &= \overline{L_{N-1}} + w_1(e^{-\delta}-1)e^{-\delta Q} + w_2 + 2w_2 Q. \end{aligned} \tag{23}$$

Note that the initial upper and lower bounds $\overline{L_0} = \underline{L_0} = L_0 = w_1 N_p$. In this paper, we set $w_1 = 1.0$ and $w_2 = 1.0$.

The derived upper and lower bounds for all aforementioned six loss functions are listed and summarized in Table 1. Note that the initial case varies depending on the considered loss functions.

4.4 Algorithm

Our complete globally-optimal contrast maximisation framework (GOCMF) is outlined in Algorithm 1 and Algorithm 2. We propose a nested strategy for calculating upper bounds, in which the outer layer **RB** evaluates the objective function, while the inner layer **BB** estimates the bounding box $\mathcal{P}_N^{\boldsymbol{\Theta}}$ and depends on the specific motion parametrisation.

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

Algorithm 1 GOCMF: globally optimal contrast maximisation framework

Input: event set $\mathcal{E}$, initial search space $\boldsymbol{\Theta}$, termination threshold τ
Output: optimal warping parameters $\hat{\boldsymbol{\theta}}$

Initialise $\boldsymbol{\theta}_0$ with the center of $\boldsymbol{\Theta}$
$\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_0$
Initialise priority queue Q
$\{\overline{L}, \underline{L}\} \leftarrow \mathbf{RB}(\mathcal{E}, \boldsymbol{\Theta})$, $\hat{L} \leftarrow \underline{L}$
Push $\boldsymbol{\Theta}$ into Q with priority $\overline{L}$
while Q is not empty **do**
  Pop $\boldsymbol{\Theta}$ form Q
  if $\overline{L} - \underline{L} \leq \tau$, **then** terminate
  $\boldsymbol{\theta}_0 \leftarrow$ Center of $\boldsymbol{\Theta}$
  if $\underline{L} \geq \hat{L}$, **then** $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_0$, $\hat{L} \leftarrow \underline{L}$
  Subdivide $\boldsymbol{\Theta}$ into subspaces $\boldsymbol{\Theta}_j$
  for all subspaces $\boldsymbol{\Theta}_j$ **do**
    $\{\overline{L}, \underline{L}\} = \mathbf{RB}(\mathcal{E}, \boldsymbol{\Theta}_j)$
    if $\overline{L} \geq \hat{L}$ **then** Insert $\boldsymbol{\Theta}_j$ into Q with priority $\overline{L}$
return $\hat{\boldsymbol{\theta}}$

Algorithm 2 RB: recursive bounds calculation

Input: event set $\mathcal{E}$, search space $\boldsymbol{\Theta}$
Output: lower bound $\underline{L}$, upper bound $\overline{L}$

Initialise accumulator image matrices $\overline{I}$ and $\underline{I}$ with zeros
Initialise $\underline{L}$, $\overline{L}$ according to Table 1
$\boldsymbol{\theta}_0 \leftarrow$ center of $\boldsymbol{\Theta}$
for each event $e_k \in \mathcal{E}$ **do**
  $\mathcal{P}_k^{\boldsymbol{\Theta}} \leftarrow \mathbf{BB}(W(\cdot), \boldsymbol{\Theta}, e_k)$
  $Q = \max_{\mathbf{p}_{ij} \in \mathcal{P}_k^{\boldsymbol{\Theta}}} \overline{I}(\mathbf{p}_{ij})$
  $\boldsymbol{\eta}_k^{\boldsymbol{\theta}_0} = \text{round}(W(\mathbf{x}_k, t_k; \boldsymbol{\theta}_0))$
  $\boldsymbol{\nu}_k = \text{argmax}_{\mathbf{p}_{ij} \in \mathcal{P}_k^{\boldsymbol{\Theta}}} \overline{I}(\mathbf{p}_{ij})$
  Update $\underline{L}$, $\overline{L}$ (cf. Table 1)
  $\overline{I}(\boldsymbol{\nu}_k) = \overline{I}(\boldsymbol{\nu}_k) + 1$
  $\underline{I}(\boldsymbol{\eta}_k^{\boldsymbol{\theta}_0}) = \underline{I}(\boldsymbol{\eta}_k^{\boldsymbol{\theta}_0}) + 1$
return $\underline{L}$, $\overline{L}$

5 Applications

As introduced by Gallego et al. [16], contrast maximisation can be applied to several event-based vision problems. We first apply GOCMF to a simple optical flow estimation problem (Section 5.1). We then apply it to fronto-parallel motion estimation (Section 5.2) in front of noisy or feature-poor, fast-moving textures, and demonstrate the potential of outperforming regular camera alternatives. We finally apply our framework to the problem of camera rotation estimation, where we compare our algorithm against the recently proposed, alternative globally-optimal framework by Liu et al. [28] (Section 5.3).

5.1 Optical Flow Estimation

Optical flow plays a vital role in object tracking, image registration, visual odometry and other navigation tasks [2]. Horn-Schunck [80] and Lucas-Kanade [81] are classical algorithms for optical flow estimation with standard cameras. Contrast maximisation methods for event cameras are good alternatives to estimate optical flow in challenging scenarios in which regular cameras would suffer from motion blur. We start by applying our globally optimal framework (Algorithm 1) to event-based optical flow estimation. More specifically, the goal is to estimate a two-dimensional velocity vector for a given point in the image plane by considering the contrast in a small bounding box around that point. This first-order model simply assumes that the trajectory of a pixel in the image plane is a straight, constant-velocity line over short time intervals. Complete flow fields can be computed by repeating the operation for each location in the image plane.

The warping function that takes events back to the reference view is hence given by

$$W(\mathbf{x}_k, t_k; \mathbf{v}) = \mathbf{x}'_k = \mathbf{x}_k + \mathbf{v} t_k, \tag{24}$$

where $\mathbf{v} = [v^x \; v^y]^\mathsf{T}$ is the velocity (optical flow) at the considered point, $\mathbf{x}_k$ the location where the k-th event occurred, and t_k the elapsed time since the time of the reference view.

It is intuitively clear that—given a branch of the search space $\mathcal{V} = [v^x_{\text{min}}, v^x_{\text{max}}] \times [v^y_{\text{min}}, v^y_{\text{max}}]$—the warped event may locate in a rectangular bounding box defined by

$$\begin{aligned} \underline{x'_k} &= x_k + v^x_{\text{min}} t_k, \; \underline{y'_k} = y_k + v^y_{\text{min}} t_k \\ \overline{x'_k} &= x_k + v^x_{\text{max}} t_k, \; \overline{y'_k} = y_k + v^y_{\text{max}} t_k. \end{aligned} \tag{25}$$

We can then employ the derived bounding box to GOCMF to explore the globally optimal optical flow. Note that the objective function we used in the experiments is L_{SoS} which is a common loss used in previous contrast maximisation work.

5.1.1 Results and Discussion

We test our globally optimal optical flow estimation algorithm on the two sequences *Circle* and *Line* collected by ourselves with a downward-facing event camera mounted on an Autonomous Ground Vehicle (AGV) (cf. Fig. 7). Ground truth optical flow is obtained by using the ground truth camera motion parameters, and calculating the first-order differential of the image motion. We find the optical flow at a image point by considering the contrast within a 40×40 pixel patch centered around that point. We thus assume that the optical flow of all the pixels in a patch is identical. We divide the *Circle* and *Line* sequences into sub-sequences of 0.04s.

We compare the proposed GOCMF with CMGD (local optimisation method using Matlab's `fmincon` function). Fig. 5 displays the optical flow estimated by GOCMF and CMGD as well as ground-truth. The estimated results match well with ground-truth. Quantitative results are exhibited in Table 2 showing the Average Endpoint Error ($AEE = \frac{1}{N} \sum \|\mathbf{x}' - \mathbf{x}'^*\|_2$ with $\mathbf{x}'$ and $\mathbf{x}'^*$ being the warped location with ground truth and estimated optical flows, and N being the number of events) and the runtimes. GOCMF is more accurate than the locally optimising CMGD, while CMGD runs faster than our method.

Qualitative results are shown in Fig. 6. It illustrates the frame captured at the reference time and the images of warped events by GOCMF and CMGD. It is clear that IWEs with parameters estimated by GOCMF are much sharper than IWEs generated by CMGD.

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

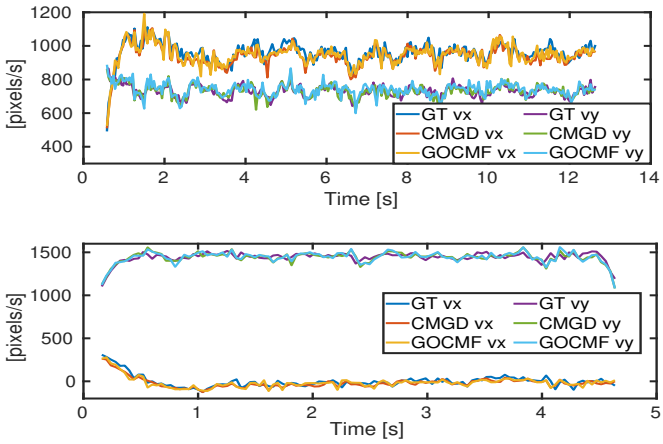


Fig. 5. Estimated pixel velocities compared against ground truth on sequences *Circle* (top) and *Line* (bottom).

TABLE 2
Runtimes and Average Endpoint Errors (AEE) for GOCMF and CMGD

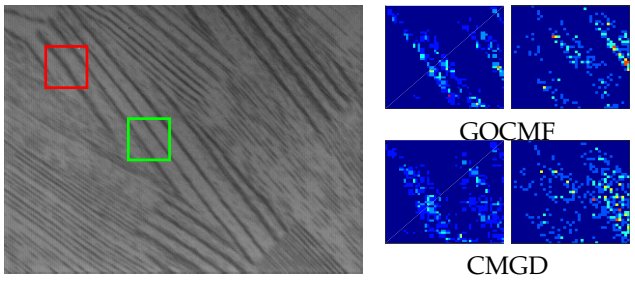
Method	*Cicle*		*Line*	
	AEE	time [s]	AEE	time [s]
CMGD	1.22	12.12	1.12	9.59
GOCMF	1.15	34.23	0.98	28.73

5.2 Visual odometry with a downward-facing event camera

Motion estimation for planar Autonomous Ground Vehicles (AGVs) is an important problem in intelligent transportation. An interesting alternative is given by employing a downward instead of a forward facing camera, which turns the image-to-image warping into a homographic mapping with known depth (cf. Fig. 7(a)). A traditional camera based method would be affected by the following, potentially severe challenges: a) reliable feature matching or even extraction may be difficult for certain noisy ground textures, b) fast motion may easily lead to motion blur, and c) stable appearance may require artificial illumination. We therefore consider an event camera as a highly interesting and much more dynamic alternative visual sensor for this particular scenario.

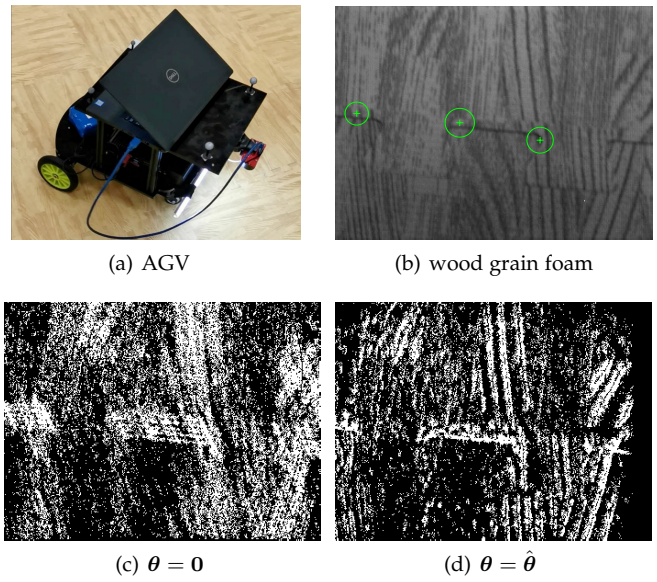
5.2.1 Homographic Mapping and Bounding Box Extraction

We rely on a recently published globally-optimal BnB solver [79] for correspondence-less AGV motion estimation with a normal, downward facing camera. We employ the two-dimensional Ackermann steering model [79], [82], [83] describing the commonly non-holonomic motion of an AGV. Employing this 2-DoF model leads to benefits in BnB, the complexity of which strongly depends on the dimensionality of the solution space. The Ackermann model constrains the motion of the vehicle to follow a circular-arc trajectory about an Instantaneous Centre of Rotation (ICR). The motion between successive frames can be conveniently described at the hand of two parameters: the half-angle of



(a) Two patches on the reference view (b) Warped events

Fig. 6. (a) Frame captured at the reference time and two patches (green and red) for optical flow estimation. (b) Images of warped events with optical flow estimated by GOCMF (top) and CMGD (bottom). Our method finds global optima and leads to significantly sharper IWEs than CMGD.



(a) AGV (b) wood grain foam

(c) $\boldsymbol{\theta} = \mathbf{0}$ (d) $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$

Fig. 7. (a) AGV equipped with a downward facing event camera for vehicle motion estimation . (b) collected image with detectable corners, (c) image of warped events with $\boldsymbol{\theta} = \mathbf{0}$, and (d) image of warped events with optimal parameters.

the relative rotation angle ϕ, and the baseline between the two views ρ. However, the alignment of the events requires a temporal parametrisation of the relative pose, which is why we employ the angular velocity $\omega = \frac{\theta}{t} = \frac{2\phi}{t}$ as well as the translational velocity $v = \omega r = \omega\rho\frac{1}{2\sin(\phi)}$ in our model. The relative transformation from vehicle frame v' back to v is therefore given by

$$\mathbf{R}_v(t) = \begin{bmatrix} \cos(\omega t) & -\sin(\omega t) & 0 \\ \sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{t}_v(t) = \frac{v}{\omega}\begin{bmatrix} 1-\cos(\omega t) \\ \sin(\omega t) \\ 0 \end{bmatrix}. \tag{26}$$

In practice the vehicle frame hardly coincides with the camera frame. The orientation and the height of the origin can be chosen to be identical, and the camera may be laterally mounted in the centre of the vehicle. However, there is likely to be a displacement along the forward direction, which we denote by the signed variable l. In other words, $\mathbf{R}_v^c = \mathbf{I}_{3\times3}$ and $\mathbf{t}_v^c = \begin{bmatrix} 0 \; l \; 0 \end{bmatrix}^\mathsf{T}$. As illustrated in Fig. 8, the

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

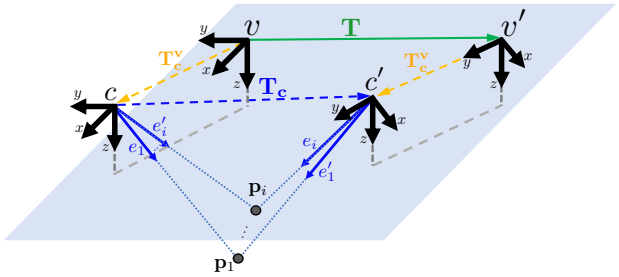


Fig. 8. Connections between vehicle displacement, extrinsic transformation, and relative camera pose.

transformation from camera pose c' (at an arbitrary future timestamp) to c (at the initial timestamp t_{ref}) is therefore given by

$$\begin{aligned}\mathbf{R}_c(t) &= \mathbf{R}_v^{c\mathsf{T}}\mathbf{R}_v(t)\mathbf{R}_v^c\,,\\ \mathbf{t}_c(t) &= -\mathbf{R}_v^{c\mathsf{T}}\mathbf{t}_v^c + \mathbf{R}_v^{c\mathsf{T}}\mathbf{t}_v(t) + \mathbf{R}_v^{c\mathsf{T}}\mathbf{R}_v(t)\mathbf{t}_v^c\,.\end{aligned} \tag{27}$$

Using the known plane normal vector $\mathbf{n} = [0\,0\,-1]^\mathsf{T}$ and depth-of-plane d, the image warping function $W(\mathbf{x}_k, t_k; [\omega\ v]^\mathsf{T})$ that permits the transfer of an event $e_k = \{\mathbf{x}_k, t_k, s_k\}$ into the reference view at t_{ref} is finally given by the planar homography equation

$$\mathbf{H}(t_k - t_{\text{ref}})\begin{bmatrix}\mathbf{x}_k\\1\end{bmatrix} = \mathbf{K}(\mathbf{R}_c(t_k - t_{\text{ref}}) - \frac{\mathbf{t}_c(t_k - t_{\text{ref}})\mathbf{n}^\mathsf{T}}{d})\mathbf{K}^{-1}\begin{bmatrix}\mathbf{x}_k\\1\end{bmatrix}. \tag{28}$$

Note that $\mathbf{K}$ here denotes a regular perspective camera calibration matrix with homogeneous focal length f, zero skew, and a principal point at $[u_0\ v_0]^\mathsf{T}$. Note further that the result needs to be dehomogenised. After expansion, we easily obtain

$$\begin{aligned}\mathbf{x}_k' &= W(\mathbf{x}_k, t_k; [\omega\ v]^\mathsf{T}) = [x_k'\ y_k']^\mathsf{T}, \text{where} \qquad (29)\\ x_k' &= -[y_k - v_0 + l\cdot\frac{f}{d}]\sin(\omega(t_k - t_{\text{ref}}))\\ &\quad +[x_k - u_0 - \frac{f}{d}\cdot\frac{v}{w}]\cos(\omega(t_k - t_{\text{ref}})) + \frac{f}{d}\cdot\frac{v}{w} + u_0,\\ y_k' &= [x_k - u_0 - \frac{f}{d}\cdot\frac{v}{w}]\sin(\omega(t_k - t_{\text{ref}}))\\ &\quad +[y_k - v_0 + l\cdot\frac{f}{d}]\cos(\omega(t_k - t_{\text{ref}})) - l\cdot\frac{f}{d} + v_0.\end{aligned}$$

Finally, the bounding box $\mathcal{P}_k^\Theta$ is found by bounding the values of x_k' and y_k' over the intervals $\omega \in \mathcal{W} = [\omega_{\min}, \omega_{\max}]$ and $v \in \mathcal{V} = [v_{\min}, v_{\max}]$. The bounding is easily achieved if simply considering monotonicity of functions over given sub-branches. For example, if $\omega_{\min} \geq 0$, $v_{\min} \geq 0$, $x_k \geq u_0$, and $y_k \geq v_0 - l\cdot\frac{f}{d}$, we obtain

$$\begin{aligned}\underline{x_k'} &= -[y_k - v_0 + l\cdot\frac{f}{d}]\sin(\omega_{\max}t)\\ &\quad +[x_k - u_0 - \frac{f}{d}\cdot\frac{v_{\min}}{\omega_{\max}}]\cos(\omega_{\max}t) + \frac{f}{d}\cdot\frac{v_{\min}}{\omega_{\max}} + u_0\,,\\ \overline{x_k'} &= -[y_k - v_0 + l\cdot\frac{f}{d}]\sin(\omega_{\min}t)\\ &\quad +[x_k - u_0 - \frac{f}{d}\cdot\frac{v_{\max}}{\omega_{\min}}]\cos(\omega_{\min}t) + \frac{f}{d}\cdot\frac{v_{\max}}{w_{\min}} + u_0\,,\\ \underline{y_k'} &= [x_k - u_0 - \frac{f}{d}\cdot\frac{v_{\max}}{\omega_{\min}}]\sin(\omega_{\min}t)\\ &\quad +[y_k - v_0 + l\cdot\frac{f}{d}]\cos(\omega_{\max}t) - l\cdot\frac{f}{d} + v_0\,,\ \text{and}\\ \overline{y_k'} &= [x_k - u_0 - \frac{f}{d}\cdot\frac{v_{\min}}{\omega_{\max}}]\sin(\omega_{\max}t)\\ &\quad +[y_k - v_0 + l\cdot\frac{f}{d}]\cos(\omega_{\min}t) - l\cdot\frac{f}{d} + v_0\,.\end{aligned} \tag{30}$$

We kindly refer the reader to the appendix for all further cases.

5.2.2 Results and Discussion

We apply our method to real data collected by a DAVIS346 event camera, which outputs events streams with a maximum time resolution of $1\mu s$ as well as regular frames at a frame rate of 30Hz. Images have a resolution of 346×260. The camera is mounted on the front of a XQ-4 Pro robot and faces downwards (see Figure 7(a)). The displacement from the non-steering axis to the camera is $l = -0.45$m, and the height difference between camera and ground is $d = 0.23$m. We recorded several motion sequences on a wood grain foam which has highly self-similar texture and poses a challenge to reliably extract and match features. Ground truth is obtained via an Optitrack optical motion tracking system. Our algorithm is working in undistorted coordinates, which is why normalisation and undistortion are computed in advance. The following aspects are evaluated:

Event-based vs frame-based: GOVO [79] and IFMI [84] are frame-based algorithms specifically designed for planar AGV motion estimation under featureless conditions. Fig. 7 shows an example frame of the wood grain foam texture, and Fig. 9 the results obtained for all methods. As can be observed, GOVO finds as little as three corner features for some of the images, thus making it difficult to accurately recover the vehicle displacement despite the globally-optimal correspondence-less nature of the algorithm. Both IFMI and GOVO occasionally lose tracking (especially for linear motion), which leaves our proposed globally-optimal event-based method using L_{SoSAaS} as the clearly outperforming method. Table 3 lists the RMS errors over the angular velocity and linear velocity. Note that the average runtime of GOCMF over *Line*, *Circle* and *Curve* is about 55s, 85s and 71s respectively.

BnB vs Gradient Ascent: We apply both gradient descent as well as BnB to the *Foam* dataset with curved motion. For the first temporal interval and the local search method, we vary the initial angular velocity ω and linear velocity v between -1 and 0.8 with steps of 0.2 (rad/s or m/s, respectively). For later intervals, we use the previous local optimum. Fig. 10 illustrates the estimated trajectories for all initial values, compared against ground truth and a BnB

Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

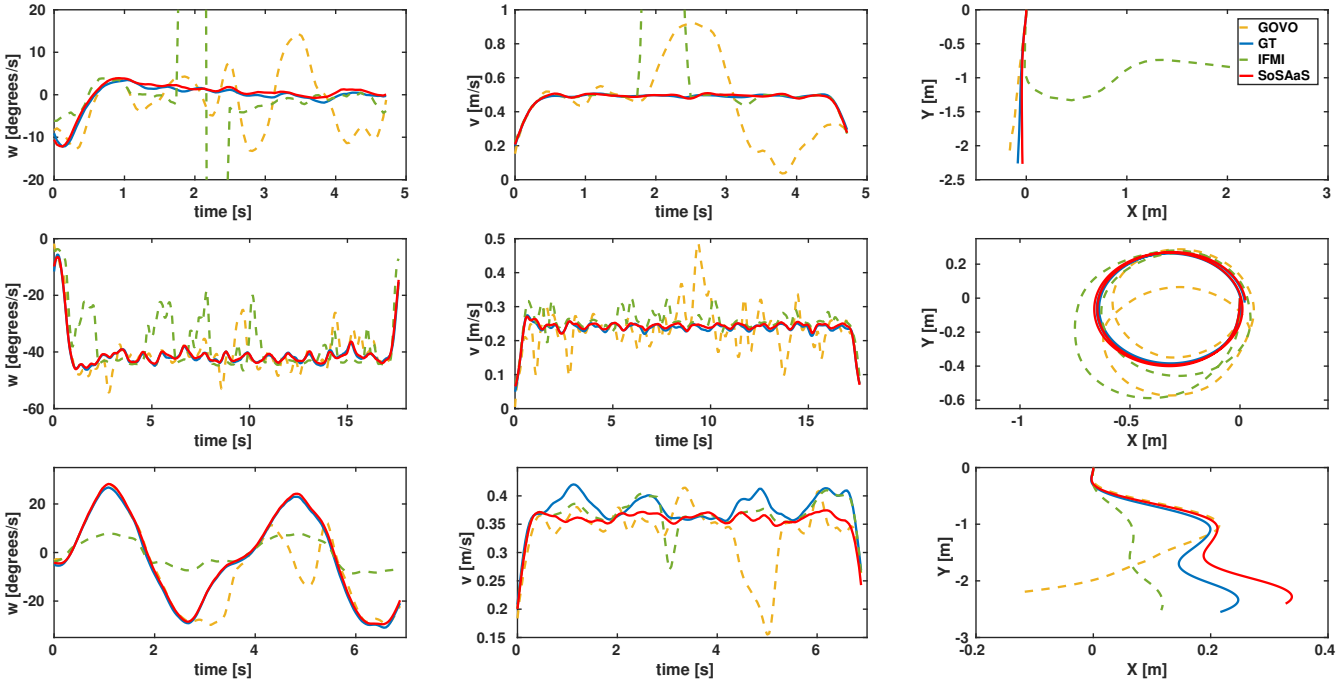


Fig. 9. Results for all methods over different datasets. The first two columns are errors over time for ω and v, and the third column illustrates a bird's eye view onto the integrated trajectories. Both IFMI and GOVO occasionally lose tracking (especially for linear motion), which leaves our proposed globally-optimal event-based method using L_{SoSAaS} as the clearly outperforming method.

TABLE 3
RMS errors for Event-based and frame-based methods

Method	*Line* w [°/s]	*Line* v [m/s]	*Circle* w [°/s]	*Circle* v [m/s]	*Curve* w [°/s]	*Curve* v [m/s]
GOCMF	**0.517**	**0.008**	**0.529**	**0.004**	**0.554**	**0.018**
IFMI	145.37	1.059	8.109	0.024	12.804	0.019
GOVO	6.970	0.240	4.550	0.064	9.865	0.059

TABLE 4
RMS errors for gradient ascent and SoS

Method	w [°/s]	v [m/s]
SoS	3.0091	0.0208
GA	11.5023	0.0379

TABLE 5
RMS errors for the different textures

Scene	w [°/s]	v [m/s]
Carpet	4.730	0.034
Poster	3.122	0.030

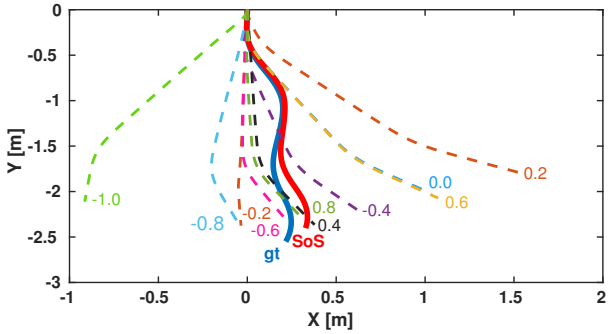


Fig. 10. Estimated trajectories by our method (SoS), gradient ascent with various initializations, and ground truth (gt). Obviously, good initializations are important for local optimisation, while GOCMF has no such requirements.

search using L_{SoS}. RMS errors are also indicated in Table 4. As clearly shown, even the best initial guess eventually diverges under a local search strategy, thus leading to clearly inferior results compared to our globally optimal search.

Various textures: To further analyse the robustness, we test our algorithm on datasets collected with various textures. Fig. 11 presents frames from two further datasets named *Carpet* and *Poster*. The *Carpet* sequences are collected on a carpet with non-repetitive almost featureless texture, while the *Poster* sequences are collected on a poster with characters and figures for which it is easy to extract features. The estimated errors are summarised in Table 5. As can be observed, our algorithm consistently shows a similar level of accuracy for the various textures in the datasets.

5.3 Rotational motion estimation

Our final application of our contrast maximisation framework considers event-based pure rotation estimation [16], [26], [28]. The latter is a recurrent topic as general motion estimation often depends on a non-zero baseline assumption, and thus may break in situations of negligibly small camera translation. The technique is furthermore commonly applied to video stabilisation [85] and panoramic image creation [86]. Our technique is applicable as the flow of the events in a purely rotational displacement situation can also be characterized by a homographic warping (i.e. the homography at infinity).

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

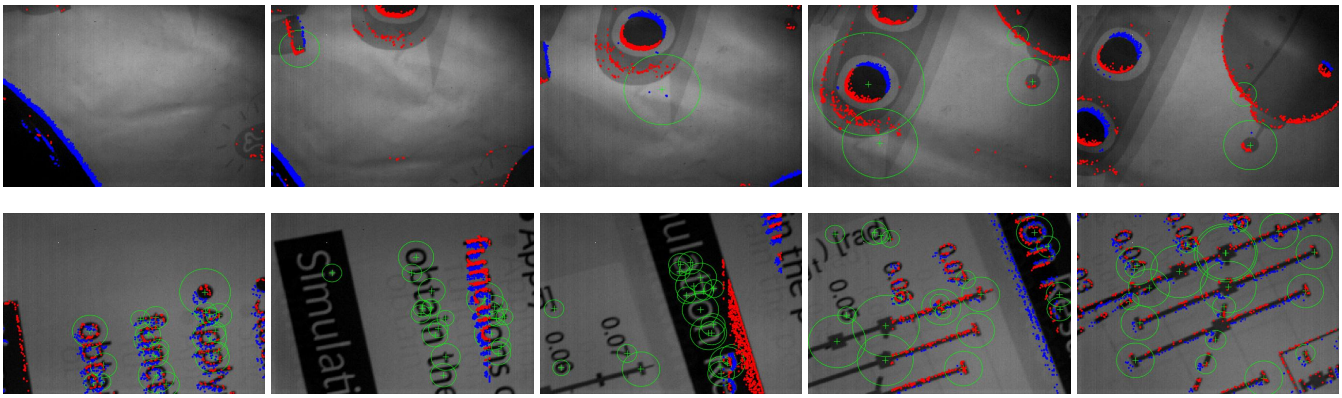


Fig. 11. Frames from dataset *Carpet* (first row) and *Poster* (second row).

We assume a constant angular velocity $\boldsymbol{\omega} \in \mathbb{R}^3$ to parameterize the rotational motion over a sufficiently small time interval. The rotation is given by

$$\mathbf{R}(t;\boldsymbol{\omega}) = \exp(\boldsymbol{\omega}^{\wedge} t), \tag{31}$$

where $\boldsymbol{\omega}^{\wedge}$ is the 3×3 screw symmetric matrix form of $\boldsymbol{\omega}$, and exp is the exponential map of the rotation group $SO(3)$ [87]. The warping function is finally given by

$$\mathbf{f}'_k = \exp(\boldsymbol{\omega}^{\wedge} t_k)\mathbf{f}_k, \tag{32}$$

where $\mathbf{f}_k = normalize(\mathbf{K}^{-1}\begin{bmatrix}\mathbf{x}_k^{\mathsf{T}} & 1\end{bmatrix}^{\mathsf{T}})$ is the bearing vector of the event $\mathbf{x}_k$ at time t_k, and $\mathbf{f}'_k$ is the rotated bearing vector expressed in the reference frame at time t_{ref}. Hence we project $\mathbf{f}'_k$ to the image plane to obtain the location of the warped event $\mathbf{x}'_k$.

As mentioned in Algorithm 1 and Algorithm 2, it is again essential to derive a bounding box for each warped event in the IWE. We adopt the bounding box derived in [28]. Given a search space $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ with center $\boldsymbol{\omega}_0$ and bounds $\boldsymbol{\omega}_{\min}$ and $\boldsymbol{\omega}_{\max}$, define

$$\alpha_k(\boldsymbol{\Omega}) := 0.5 \parallel \boldsymbol{\omega}_{\min} t_k - \boldsymbol{\omega}_{\max} t_k \parallel_2 . \tag{33}$$

According to [76] and [28], all the possible rotated bearing vectors $\mathbf{f}'_k$ will then lie in a cone

$$\mathcal{V}_k(\boldsymbol{\Omega}) := \{\mathbf{f} \in \mathbb{R}^3 | \angle(\exp(\boldsymbol{\omega}_0^{\wedge} t_k)\mathbf{f}_k, \mathbf{f}) \leq \alpha_k(\boldsymbol{\Omega})\}. \tag{34}$$

As illustrated in Fig. 12, the projection of bearing vectors in cone $\mathcal{V}_k(\boldsymbol{\Omega})$ yields an elliptical region $\mathcal{L}_k(\boldsymbol{\Omega})$ on the pixel plane. The derivation of the center and semi-major axes of $\mathcal{L}_k(\boldsymbol{\Omega})$ can be found in the work by Liu et al. [28].

5.3.1 Results and Discussion

We proceed to our comparison between the proposed algorithm GOCMF and the alternative globally-optimal algorithm CMBNB presented by Liu et al. [28]. We furthermore compare our results against a local optimiser, denoted CMGD (`fmincon` function from Matlab). We use the publicly available sequences *poster*, *boxes* and *dynamic* from [26], [88], which were recorded using a Davis240C [89] under rotational motion over a static indoor scene. The ground truth motion was captured using a motion capture system. We utilized the last 15s of each sequence and split them into 10ms subsequences to run the algorithms. For each subsequence from *poster* and *boxes*, the events size is

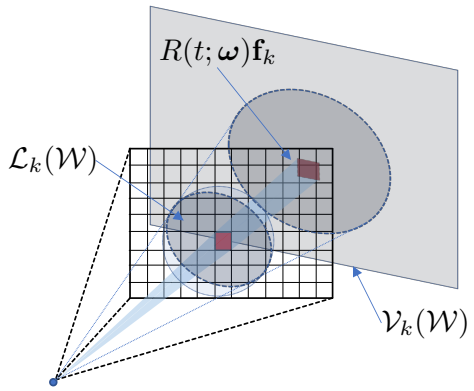


Fig. 12. The cone $\mathcal{V}_k(\boldsymbol{\Omega})$ that $\mathbf{R}(t;\boldsymbol{\omega})\mathbf{f}_k$ might locate in with search space $\boldsymbol{\Omega}$. We also illustrate the elliptical region $\mathcal{L}_k(\boldsymbol{\Omega})$ on the pixel plane.

$N \approx 50k$, while for *dynamic*, the size is $N \approx 25k$. To speed up the algorithm, we downsampled the event stream by a factor 2, which means we dropped half the number of events in each subsequence. Note that the objective function we used in the experiments is L_{SoS} on which CMBNB [28] is based. We ran algorithms on a standard desktop with a Intel Core i7-7700 and CPU @ 3.60GHz $\times$ 8. Moreover, GOCMF and CMBNB were both run with 8 threads.

We evaluated the algorithms using two error metrics. One is

$$\epsilon = \|\omega_{\text{gt}} - \omega_{\text{est}}\|_2, \tag{35}$$

where ω_{gt} and ω_{est} are the ground truth and estimated angular velocities, respectively. The other error metric is

$$\phi = |\|\omega_{\text{gt}}\|_2 - \|\omega_{\text{est}}\|_2|. \tag{36}$$

Table 6 presents the average μ and standard deviation σ of ϵ and ϕ as well as the average runtime over all subsequences. Due to its exact nature, globally optimal approaches have lower errors than CMGD. Moreover, our algorithm performs slightly better than CMBNB over datasets *dynamic* and *poster*, whereas CMBNB performs slightly better over *boxes*. Most interestingly, the average runtime of GOCMF is 3.34, 3.07 and 2.49 times faster than CMBNB respectively over *dynamic*, *poster* and *boxes*. The reason is that our efficient, recursively evaluated upper bound proves to be tighter than the upper bound proposed in [28] which is further analysed in Section 6.2.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 6
Rotational Motion Estimation Errors

Method	*dynamic*					*boxes*					*poster*				
	$\mu(\phi)$	$\delta(\phi)$	$\mu(\epsilon)$	$\delta(\epsilon)$	time	$\mu(\phi)$	$\delta(\phi)$	$\mu(\epsilon)$	$\delta(\epsilon)$	time	$\mu(\phi)$	$\delta(\phi)$	$\mu(\epsilon)$	$\delta(\epsilon)$	time
CMGD	161.6	127.9	168.7	124.1	**6.96**	61.75	103.6	71.82	101.0	**9.08**	43.64	74.29	59.29	79.29	**16.63**
CMBNB	9.88	6.94	**17.02**	8.93	76.46	**19.41**	**13.34**	**30.21**	**15.49**	92.48	22.05	23.18	32.94	23.87	221.0
GOCMF	**9.75**	**6.66**	17.05	**8.58**	22.88	20.19	16.19	31.27	21.10	30.08	**21.98**	**23.02**	**32.39**	**23.44**	88.48

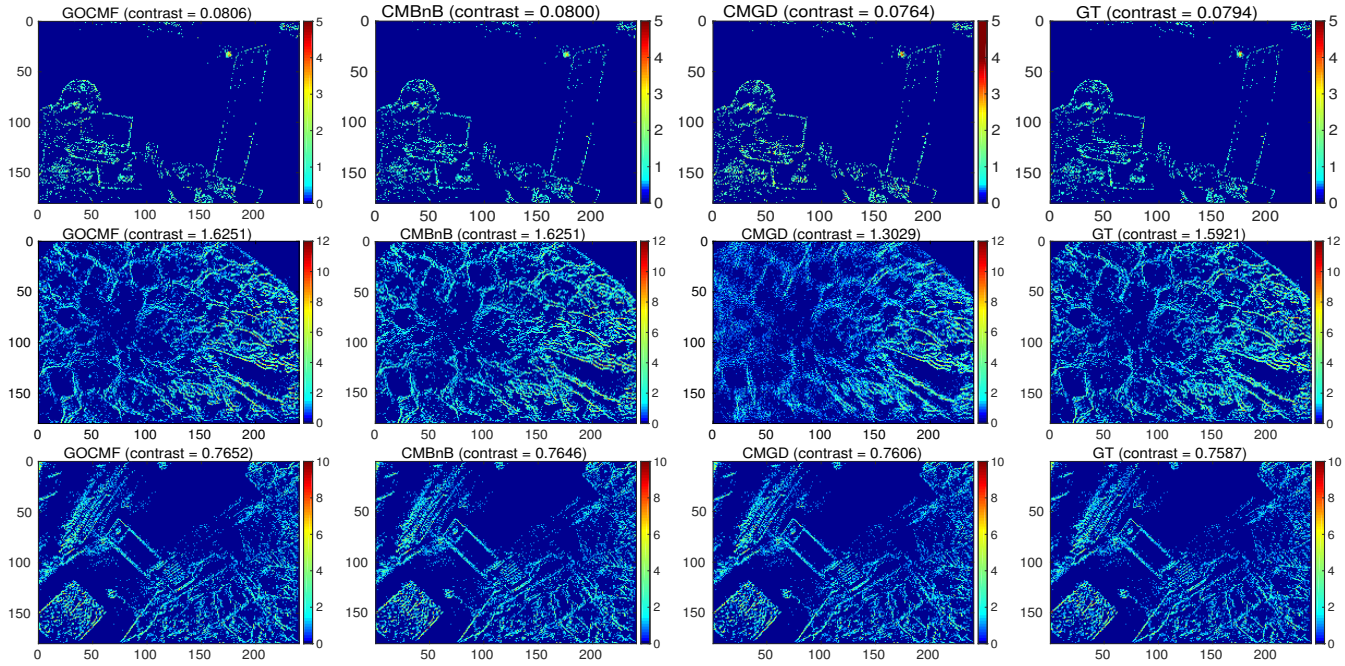


Fig. 13. Images of warped events with anguler velocities estimated by GOCMF, CMBNB, CMGD and groudtruth (GT). The three rows are images using events stream from *dynamic*, *poster* and *boxes* respectively. It is clear that CMGD can indeed often converge to bad local solutions.

Fig. 13 displays the IWEs with optimal parameters estimated by different approaches.

6 Analysis

We analyse multiple further aspects of our GOCMF algorithm. We first analyse the precision and the robustness of our contrast maximisation framework. Second, we compare the bound convergence of GOCMF and CMBNB. Third, we speed up our algorithm by implementing a downsampling mechanism. To conclude, we furthermore compare the accuracy of GOCMF under all six possible objective functions.

6.1 Precision and robustness

We start by evaluating the precision of motion estimation with the contrast maximisation function L_{SoS} over synthetic data. As already implied in [15], L_{SoS} can be considered as a solid starting point for the evaluation. Our synthetic data consists of randomly generated horizontal and vertical line segments on a plane at a depth of 2.0m. We consider Ackermann motion with an angular velocity $\omega = 28.6479^\circ/\mathrm{s}$ (0.5rad/s) and a linear velocity $v = 0.5\mathrm{m/s}$. Events are generated by randomly choosing a 3D point on a line, and reprojecting it into a random camera pose sampled by a random timestamp within the interval $[0, 0.1\mathrm{s}]$. The result of our method is finally evaluated by running BnB over the search space $\mathcal{W} = [0.4, 0.6]$ and $\mathcal{V} = [0.4, 0.6]$, and comparing the retrieved solution against the result of an exhaustive search with sampling points every $\delta\omega = 0.001\mathrm{rad/s}$ and $\delta v = 0.001\mathrm{m/s}$. The experiment is repeated 1000 times.

Figs. 14(a) and 14(b) illustrate the distribution of the errors for both methods in the noise-free case. The standard deviation of the exhaustive search and BnB are $\sigma_\omega = 1.0645^\circ/\mathrm{s}$, $\sigma_v = 0.0151\mathrm{m/s}$ and $\sigma_\omega = 1.305^\circ/\mathrm{s}$, $\sigma_v = 0.0150\mathrm{m/s}$, respectively. While this result suggests that BnB works well and sustainably returns a result very close to the optimum found by exhaustive search, we still note that the optimum identified by both methods has a bias with respect to ground truth, even in the noise-free case. Note however that this is related to the nature of the contrast maximisation function, and not our globally optimal solution strategy.

In order to analyse robustness, we randomly add salt and pepper noise to the event stream with noise-to-event

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

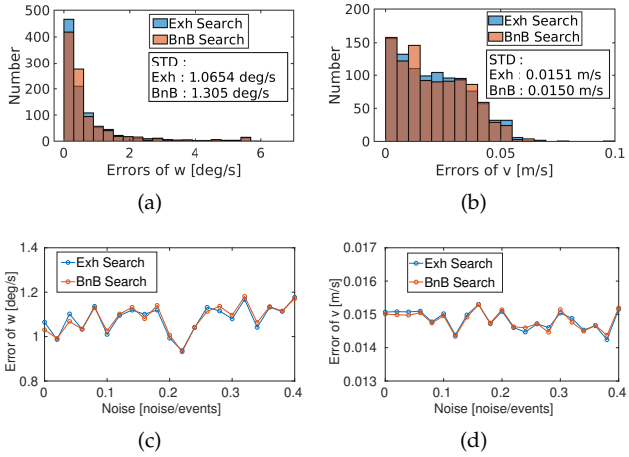


Fig. 14. Simulation Results. (a) and (b) indicate the error distribution for ω and v over all experiments for both our proposed method as well as an exhaustive search. (c) and (d) visualise the average error of the estimated parameters caused by additional salt and pepper noise on the event stream. Results are averaged over 1000 random experiments. Note that our proposed method has excellent robustness even for N/E ratios up to 40%.

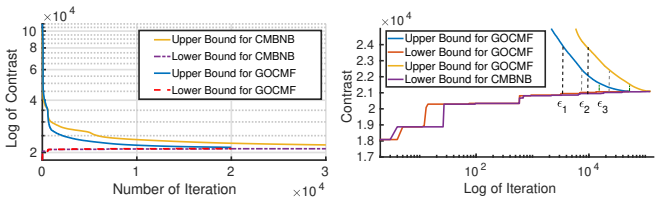


Fig. 15. Upper and lower bound evolution in GOCMF and CMBNB. It is obvious that GOCMF converges faster than CMBNB which terminated with at least 30000 iterations, while GOCMF terminated at about 20000 iterations, illustrated in the left figure. The difference in performance is due to the much tighter bounding of the GOCMF. To indicate the iteration of the lower bounds, we exhibit the right figure with log of iteration. ϵ means the chosen convergence criterion - gap between the upper and lower bound. The smaller the ϵ, the more iteration and more accurate.

(N/E) ratios between 0 and 0.4 (Example objective functions for different N/E ratios have already been illustrated in Figs. 2). Fig. 14(c) and 14(d) show the error for each noise level again averaged over 1000 experiments. As can be observed, the errors are very similar and behave more or less independently of the amount of added noise. The latter result underlines the high robustness of our approach.

6.2 Convergence

Next, we give a proof that the derived bounds converge as the branches decrease in size. Given a search space $\boldsymbol{\Theta} = [\boldsymbol{\theta}_0 - \delta\boldsymbol{\theta}, \boldsymbol{\theta}_0 + \delta\boldsymbol{\theta}]$, mathematically correct bounds should converge when $\delta\boldsymbol{\theta} \to 0$. The convergence of the bounds depends on the tightness of the bounding boxes $\mathcal{P}_N^{\boldsymbol{\Theta}}$, which varies depending on the considered scenario. However, when $\delta\boldsymbol{\theta} \to 0$, $\mathcal{P}_N^{\boldsymbol{\Theta}}$ will tend to be equal to the single pixel given by $W(\mathbf{x}_N, t_N; \boldsymbol{\theta}_0)$. Hence—according to equation (7)—the upper bound becomes

$$\begin{aligned}\lim_{\delta\boldsymbol{\theta}\to 0} \overline{L_N} &= \lim_{\delta\boldsymbol{\theta}\to 0} \overline{L_{N-1}} + 1 + 2Q^{N-1} \\ &= \lim_{\delta\boldsymbol{\theta}\to 0} \overline{L_{N-1}} + 1 + 2 \lim_{\delta\boldsymbol{\theta}\to 0} \max_{\mathbf{p}_{ij}\in\mathcal{P}_N^{\boldsymbol{\Theta}}} \overline{I}^{N-1}(\mathbf{p}_{ij}) \\ &\approx \lim_{\delta\boldsymbol{\theta}\to 0} \overline{L_{N-1}} + 1 + 2\overline{I}^{N-1}(\boldsymbol{\eta}_N^{\boldsymbol{\theta}_0}; \boldsymbol{\theta}_0) \qquad (37)\end{aligned}$$

By using mathematical induction, it is again relatively straightforward to prove that $\lim_{\delta\boldsymbol{\theta}\to 0} \overline{L_N} \to \underline{L_N}$ and $\overline{I}^N \to \underline{I}^N$.

The convergence speed of the upper and lower bounds indicates the efficiency of the BnB paradigm and the tightness of the bounds themselves. To visualize and empirically compare the convergence of the two globally optimal algorithms GOCMF and CMBNB, we plot the evolution of the upper and lower bounds within a 10 ms subsequence of the *boxes* data. The result is shown in Fig. 15. From a theoretical point of view, the BnB framework terminates when the gap between the upper and lower bound reduces to zero. As we can see in Fig. 15, the gap between the upper and lower bounds decreases slowly. Hence, for practical applications we set the convergence criterion to a non-zero gap ϵ between the upper and lower bound. The parameter ϵ generates a trade-off between accuracy and computational efficiency. The specific value for epsilon is different in each experiment and chosen automatically as a function of the average number of events in a given time interval. For example, the gap is set to 3000 for the sequence *boxes*.

As illustrated in Fig. 15, it is obvious that GOCMF converges significantly faster than CMBNB. The latter terminates after at least 30000 iterations, while GOCMF terminates already after about 20000 iterations. This is consistent with Table 6 in which the runtime indicated for GOCMF is about 3 times faster than the one for CMBNB. The difference in performance indicates tighter (recursive) bounds for GOCMF.

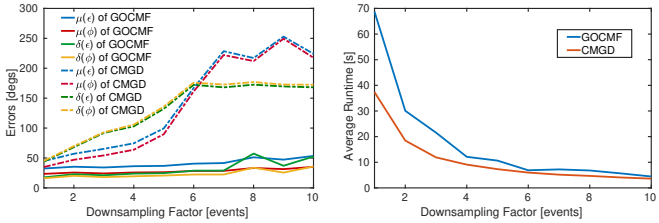
6.3 Event Downsampling

An event camera is a low-latency sensor that asynchronously outputs several million events per second which is challenging to process, especially for a CPU-based implementation. This adds to the anyway expensive nature of the globally-optimal BnB algorithm. We therefore implement a preprocessing step to speed up the algorithm. Contrast maximisation evaluates the sharpness of the IWE, a property that owns a certain robustness against downsampling of the event stream. We evaluate the accuracy of local optimisation (CMGD) and GOCMF over the last 15s of sequence *boxes* with different downsampling factors. A downsampling factor of m simply means that—within the temporally ordered sequence of events—only every m-th event is maintained. In this experiment, we test downsampling factors ranging from 1 to 10.

Fig. 16(a) illustrates the errors of GOCMF and local optimisation with increasing downsampling factors. It is clear that errors increase as the downsampling factor increases. However, GOCMF remains much more robust than CMGD when compared in terms of the two error metrics ϵ and ϕ. The mean error ϕ of CMGD is in the range of $[30°, 250°]$, while GOCMF stays within the range $[25°, 55°]$. The error of GOCMF is stable for a downsampling factor between 1 and 6. Fig. 16(b) furthermore indicates the average runtime of GOCMF and CMGD. The runtime decreases exponentially as the sampling rate decreases. An interesting phenomenon

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence



(a) errors with decreasing sample rate (b) average runtime with decreasing sample rate

Fig. 16. Downsampling results. (a) shows the errors of GOCMF and local optimisation with increasing downsampling factors. GOCMF always proves more robust than CMGD. (b) illustrates the average runtime of GOCMF and CMGD. An interesting phenomenon is that the runtime of the two approaches tends to be consistent.

TABLE 7
RMS errors for different datasets and methods

Method	*Line* w [°/s]	*Line* v [m/s]	*Circle* w [°/s]	*Circle* v [m/s]	*Curve* w [°/s]	*Curve* v [m/s]
SoE	2.408	0.015	2.212	0.025	3.628	0.026
SoEaS	2.405	0.015	2.017	0.024	3.628	0.026
SoS	**0.512**	**0.008**	1.088	0.008	3.009	0.021
SoSA	1.961	0.028	4.249	0.073	9.290	0.072
SoSAaS	**0.517**	**0.008**	**0.529**	**0.004**	**0.554**	**0.018**
Var	**0.512**	**0.008**	1.088	0.008	3.009	0.020

is that the runtime of the two approaches tends to be identical. As a tradeoff between accuracy and runtime, all aforementioned experiments in this paper use a downsampling factor of 2.

6.4 Different objective functions

The proposed recursive bounds for our globally-optimal contrast maximisation framework handle a total number of six different focus loss functions. We test our algorithm with all aforementioned six contrast functions over various types of motions, including a straight line, a circle, and an arbitrarily curved trajectory (datasets used in Section 5.2). Table 7 shows the RMS errors of the estimated dynamic parameters, and compares the accuracy of all six alternatives. As can be observed, L_{SoS} and L_{Var} perform well, but the best performance is given by L_{SoSAaS}.

7 Conclusion

We have introduced a novel globally optimal solution to contrast maximisation for homographically rectified event streams. We have successfully applied this to three different scenarios, including optical flow and pure rotation estimation. To the best of our knowledge, we are the first to apply the idea of homography estimation via contrast maximisation to the real-world case of non-holonomic motion estimation with a downward facing event camera mounted on an AGV. The challenging conditions in these scenarios advertise the use of event cameras. Our globally optimal solutions are crucial for successful contrast maximisation, and significantly outperform incremental local refinement. As shown by our results, the planar motion estimation scenario ultimately favors dynamic vision sensors over regular frame-based cameras.

Appendix
Application to Visual Odometry with a Downward-Facing Event Camera

Let $t = t_k - t_{\mathrm{ref}}$, we recall that

$$\begin{aligned}
\mathbf{x}'_k &= W(\mathbf{x}_k, t_k; [\omega\ v]^{\mathsf{T}}) = [x'_k \quad y'_k]^{\mathsf{T}} && (38)\\
x'_k &= -[y_k - v_0 + l\cdot\frac{f}{d}]\sin(\omega t)\\
&\quad +[x_k - u_0 - \frac{f}{d}\cdot\frac{v}{w}]\cos(\omega t) + \frac{f}{d}\cdot\frac{v}{w} + u_0,\\
&= a_x + b_x + c_x + u_0\\
y'_k &= [x_k - u_0 - \frac{f}{d}\cdot\frac{v}{w}]\sin(\omega t)\\
&\quad +[y_k - v_0 + l\cdot\frac{f}{d}]\cos(\omega t) - l\cdot\frac{f}{d} + v_0.\\
&= a_y + b_y + c_y - l\cdot\frac{f}{d} + v_0 && (39)
\end{aligned}$$

where

$$\begin{aligned}
a_x &= -[y_k - v_0 + l\cdot\frac{f}{d}]\sin(\omega t),\\
b_x &= [x_k - u_0]\cos(\omega t),\\
c_x &= \frac{f}{d}\cdot\frac{v}{\omega}[1 - \cos(\omega t)],\\
a_y &= [x_k - u_0]\sin(\omega t),\\
b_y &= -\frac{f}{d}\cdot\frac{v}{\omega}\sin(\omega t),\\
c_y &= [y_k - v_0 + l\cdot\frac{f}{d}]\cos(\omega t). && (40)
\end{aligned}$$

The bounding box $\mathcal{P}_k^{\Theta}$ over the intervals $\omega \in \mathcal{W} = [\omega_{\min}, \omega_{\max}]$ and $v \in \mathcal{V} = [v_{\min}, v_{\max}]$. Here we only consider the case $|\omega t| < \pi/2$. The bounding box is easily achieved if simply considering the monotonicity and different cases. There are 17 cases in total. One special case is when $\omega = 0$. Given the Ackermann motion model, we then obtain

$$\underline{x'_k} = x_k,\ \overline{x'_k} = -x_k, \qquad (41)$$

$$\underline{y'_k} = y_k + \frac{f}{d}\cdot v_{\min} t,\ \overline{y'_k} = y_k + \frac{f}{d}\cdot v_{\max} t. \qquad (42)$$

The other 16 cases are based on the monotonicity of functions. For example, if $\omega_{\min} \geq 0$, $v_{\min} \geq 0$ and $x_k \geq u_0$, $y_k \geq v_0 - l\cdot\frac{f}{d}$, the lower bound of x'_k is

$$\begin{aligned}
\underline{x'_k} &= \min_{\omega} a_x + \min_{\omega} b_x + \min_{\omega,v} c_x + u_0\ \text{, with} && (43)\\
\min_{\omega} a_x &\geq -[y_k - v_0 + l\cdot\frac{f}{d}]\sin(\omega_{\max} t),\\
\min_{\omega} b_x &\geq [x_k - u_0]\cos(\omega_{\max} t),\\
\min_{\omega,v} c_x &\geq \frac{f}{d}\cdot\frac{v_{\min}}{\omega_{\max}}[1 - \cos(\omega_{\max} t)]. && (44)
\end{aligned}$$

Table 8 lists $\underline{x'_k}$ and $\underline{y'_k}$ with ω and v arguments when the search space is $\omega_{\min} > 0$. Meanwhile $\overline{x'_k}$ and $\overline{y'_k}$ are obtained by $\omega_{\min}$ against $\omega_{\max}$, and $v_{\min}$ against $v_{\max}$. The other 8 cases with $\omega_{\max} < 0$ are derived by a similar strategy.

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 8
Bounding Box Cases

Conditions			x'_k			y'_k		
			a_x	b_x	c_x	a_y	b_y	c_y
$\omega_{\min} > 0$	$v_{\min} \geq 0$	$x_k \geq u_0$, $y_k \geq v_0 - l \cdot \frac{f}{d}$	$\omega_{\max}$	$\omega_{\max}$	$\omega_{\max}$, $v_{\min}$	$\omega_{\min}$	$\omega_{\min}$, $v_{\max}$	$\omega_{\max}$
		$x_k < u_0$, $y_k \geq v_0 - l \cdot \frac{f}{d}$	$\omega_{\max}$	$\omega_{\min}$	$\omega_{\max}$, $v_{\min}$	$\omega_{\max}$	$\omega_{\min}$, $v_{\max}$	$\omega_{\max}$
		$x_k < u_0$, $y_k < v_0 - l \cdot \frac{f}{d}$	$\omega_{\min}$	$\omega_{\min}$	$\omega_{\max}$, $v_{\min}$	$\omega_{\max}$	$\omega_{\min}$, $v_{\max}$	$\omega_{\min}$
		$x_k \geq u_0$, $y_k < v_0 - l \cdot \frac{f}{d}$	$\omega_{\min}$	$\omega_{\max}$	$\omega_{\max}$, $v_{\min}$	$\omega_{\min}$	$\omega_{\min}$, $v_{\max}$	$\omega_{\min}$
	$v_{\min} < 0$	$x_k \geq u_0$, $y_k \geq v_0 - l \cdot \frac{f}{d}$	$\omega_{\max}$	$\omega_{\max}$	$\omega_{\min}$, $v_{\min}$	$\omega_{\min}$	$\omega_{\min}$, $v_{\max}$	$\omega_{\max}$
		$x_k < u_0$, $y_k \geq v_0 - l \cdot \frac{f}{d}$	$\omega_{\max}$	$\omega_{\min}$	$\omega_{\min}$, $v_{\min}$	$\omega_{\max}$	$\omega_{\max}$, $v_{\max}$	$\omega_{\max}$
		$x_k < u_0$, $y_k < v_0 - l \cdot \frac{f}{d}$	$\omega_{\min}$	$\omega_{\min}$	$\omega_{\min}$, $v_{\min}$	$\omega_{\max}$	$\omega_{\max}$, $v_{\max}$	$\omega_{\min}$
		$x_k \geq u_0$, $y_k < v_0 - l \cdot \frac{f}{d}$	$\omega_{\min}$	$\omega_{\max}$	$\omega_{\min}$, $v_{\min}$	$\omega_{\min}$	$\omega_{\max}$, $v_{\max}$	$\omega_{\min}$

REFERENCES

[1] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *International Journal of Computer Vision*, pp. 1–57, 2020.

[2] K. R. Aires, A. M. Santana, and A. A. Medeiros, "Optical flow using color information: preliminary results," in *Proceedings of the 2008 ACM symposium on Applied computing*, 2008, pp. 1607–1611.

[3] L. Kneip, M. Chli, and R. Y. Siegwart, "Robust real-time visual odometry with a single camera and an imu," in *Proceedings of the British Machine Vision Conference 2011*, 2011.

[4] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE robotics & automation magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[5] L. Kneip and P. Furgale, "Opengv: A unified and generalized approach to real-time calibrated geometric vision," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1–8.

[6] A. Ben-Afia, L. Deambrogio, D. Salós, A.-C. Escher, C. Macabiau, L. Soulier, and V. Gay-Bellile, "Review and classification of vision-based localisation techniques in unknown environments," *IET Radar, Sonar & Navigation*, vol. 8, no. 9, pp. 1059–1072, 2014.

[7] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[8] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 1. IEEE, 2006, pp. 519–528.

[9] J. Butime, I. Gutierez, L. Corzo, and C. Espronceda, "3d reconstruction methods, a survey," in *Proceedings of the First International Conference on Computer Vision Theory and Applications*, 2006, pp. 457–463.

[10] H. Ham, J. Wesley, and H. Hendra, "Computer vision based 3d reconstruction: A review," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4, p. 2394, 2019.

[11] A. Papazoglou and V. Ferrari, "Fast object segmentation in unconstrained video," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1777–1784.

[12] E. Che, J. Jung, and M. J. Olsen, "Object recognition, segmentation, and classification of mobile laser scanning point clouds: A state of the art review," *Sensors*, vol. 19, no. 4, p. 810, 2019.

[13] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, "Visual simultaneous localization and mapping: a survey," *Artificial intelligence review*, vol. 43, no. 1, pp. 55–81, 2015.

[14] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[15] G. Gallego, M. Gehrig, and D. Scaramuzza, "Focus is all you need: loss functions for event-based vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 280–12 289.

[16] G. Gallego, H. Rebecq, and D. Scaramuzza, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3867–3876.

[17] A. Z. Zhu, N. Atanasov, and K. Daniilidis, "Event-based feature tracking with probabilistic data association," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 4465–4470.

[18] T. Stoffregen and L. Kleeman, "Simultaneous optical flow and segmentation (sofas) using dynamic vision sensor," *arXiv preprint arXiv:1805.12326*, 2018.

[19] C. Ye, A. Mitrokhin, C. Parameshwara, C. Fermüller, J. A. Yorke, and Y. Aloimonos, "Unsupervised learning of dense optical flow and depth from sparse event data," *CoRR*, vol. abs/1809.08625, 2018.

[20] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based learning of optical flow, depth, and egomotion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 989–997.

[21] A. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Ev-flownet: Self-supervised optical flow estimation for event-based cameras," in *Robotics: Science and Systems*, 2018.

[22] T. Stoffregen, G. Gallego, T. Drummond, L. Kleeman, and D. Scaramuzza, "Event-based motion segmentation by motion compensation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7244–7253.

[23] A. Mitrokhin, C. Fermüller, C. Parameshwara, and Y. Aloimonos, "Event-based moving object detection and tracking. in 2018 ieee," in *RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–9.

[24] H. Rebecq, G. Gallego, E. Mueggler, and D. Scaramuzza, "Emvs: Event-based multi-view stereo—3d reconstruction with an event camera in real-time," *International Journal of Computer Vision*, vol. 126, no. 12, pp. 1394–1414, 2018.

[25] A. Z. Zhu, Y. Chen, and K. Daniilidis, "Realtime time synchronized event-based stereo," in *European Conference on Computer Vision*. Springer, 2018, pp. 438–452.

[26] G. Gallego and D. Scaramuzza, "Accurate angular velocity estimation with an event camera," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 632–639, 2017.

[27] X. Peng, Y. Wang, and L. Kneip, "Globally optimal event camera motion estimation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[28] D. Liu, A. Parra, and T.-J. Chin, "Globally optimal contrast maximisation for event-based motion estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6349–6358.

[29] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger, "Inter-

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

acting maps for fast visual interpretation," in *The 2011 International Joint Conference on Neural Networks*. IEEE, 2011, pp. 770–776.

[30] 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, 2008.

[31] D. Weikersdorfer, R. Hoffmann, and J. Conradt, "Simultaneous localization and mapping for event-based vision systems," in *International conference on computer vision systems*. Springer, 2013, pp. 133–142.

[32] H. Kim, S. Leutenegger, and A. J. Davison, "Real-time 3d reconstruction and 6-dof tracking with an event camera," in *European Conference on Computer Vision*. Springer, 2016, pp. 349–364.

[33] A. Zihao Zhu, N. Atanasov, and K. Daniilidis, "Event-based visual inertial odometry," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5391–5399.

[34] E. Mueggler, G. Gallego, H. Rebecq, and D. Scaramuzza, "Continuous-time visual-inertial odometry for event cameras," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1425–1440, 2018.

[35] D. Zhu, Z. Xu, J. Dong, C. Ye, Y. Hu, H. Su, Z. Liu, and G. Chen, "Neuromorphic visual odometry system for intelligent vehicle application with bio-inspired vision sensor," in *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2019, pp. 2225–2232.

[36] Y. Zhou, G. Gallego, and S. Shen, "Event-based stereo visual odometry," *arXiv preprint arXiv:2007.15548*, 2020.

[37] M. Gehrig, S. B. Shrestha, D. Mouritzen, and D. Scaramuzza, "Event-based angular velocity regression with spiking networks," *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[38] 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.

[39] P. Bardow, A. J. Davison, and S. Leutenegger, "Simultaneous optical flow and intensity estimation from an event camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 884–892.

[40] C. Lee, A. Kosta, A. Z. Zhu, K. Chaney, K. Daniilidis, and K. Roy, "Spike-flownet: Event-based optical flow estimation with energy-efficient hybrid neural networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[41] D. R. Kepple, D. Lee, C. Prepsius, V. Isler, and I. Memming, "Jointly learning visual motion and confidence from local patches in event cameras," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[42] J. Wu, K. Zhang, Y. Zhang, X. Xie, and G. Shi, "High-speed object tracking with dynamic vision sensor," in *China High Resolution Earth Observation Conference*. Springer, 2018, pp. 164–174.

[43] J. P. Rodríguez-Gómez, A. G. Eguíluz, J. Martínez-de Dios, and A. Ollero, "Asynchronous event-based clustering and tracking for intrusion monitoring in uas," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 8518–8524.

[44] H. Chen, D. Suter, Q. Wu, and H. Wang, "End-to-end learning of object motion estimation from retinal events for event-based object tracking." in *AAAI*, 2020, pp. 10 534–10 541.

[45] H. Li, G. Li, and L. Shi, "Classification of spatiotemporal events based on random forest," in *International Conference on Brain Inspired Cognitive Systems*. Springer, 2016, pp. 138–148.

[46] A. N. Belbachir, M. Hofstätter, M. Litzenberger, and P. Schön, "High-speed embedded-object analysis using a dual-line timed-address-event temporal-contrast vision sensor," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 3, pp. 770–783, 2010.

[47] A. Chadha, Y. Bi, A. Abbas, and Y. Andreopoulos, "Neuromorphic vision sensing for cnn-based action recognition," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 7968–7972.

[48] H. Rebecq, G. G. Bonet, and D. Scaramuzza, "Emvs: Event-based multi-view stereo," in *British Machine Vision Conference (BMVC)*, 2016.

[49] G. Haessig, X. Berthelon, S.-H. Ieng, and R. Benosman, "A spiking neural network model of depth from defocus for event-based neuromorphic vision," *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[50] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, "Semi-dense 3d reconstruction with a stereo event camera," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 235–251.

[51] 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," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6820–6829.

[52] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[53] S. Lin, J. Zhang, J. Pan, Z. Jiang, D. Zou, Y. Wang, J. Chen, and J. Ren, "Learning event-driven video deblurring and interpolation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[54] 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 Robotics and Automation Letters*, vol. 2, no. 2, pp. 593–600, 2016.

[55] T. Stoffregen and L. Kleeman, "Event cameras, contrast maximization and reward functions: an analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 300–12 308.

[56] T. M. Breuel, "Implementation techniques for geometric branch-and-bound matching methods," *Computer Vision and Image Understanding*, vol. 90, no. 3, pp. 258–294, 2003.

[57] F. Pfeuffer, M. Stiglmayr, and K. Klamroth, "Discrete and geometric branch and bound algorithms for medical image registration," *Annals of Operations Research*, vol. 196, no. 1, pp. 737–765, 2012.

[58] H. Li and R. Hartley, "The 3d-3d registration problem revisited," in *2007 IEEE 11th international conference on computer vision*. IEEE, 2007, pp. 1–8.

[59] C. Olsson, F. Kahl, and M. Oskarsson, "Branch-and-bound methods for euclidean registration problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 783–794, 2008.

[60] J.-C. Bazin, Y. Seo, and M. Pollefeys, "Globally optimal consensus set maximization through rotation search," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 539–551.

[61] H. Bülow and A. Birk, "Fast and robust photomapping with an unmanned aerial vehicle (uav)," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 3368–3373.

[62] D. Campbell and L. Petersson, "Gogma: Globally-optimal gaussian mixture alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5685–5694.

[63] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-icp: A globally optimal solution to 3d icp point-set registration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 11, pp. 2241–2254, 2016.

[64] B. A. Parra, T. Chin, A. Eriksson, H. Li, and D. Suter, "Fast rotation search with stereographic projections for 3d registration." *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 11, pp. 2227–2240, 2016.

[65] Y. Liu, C. Wang, Z. Song, and M. Wang, "Efficient global point cloud registration by matching rotation invariant features through translation search," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 448–463.

[66] L. Hu, H. Shi, and L. Kneip, "Globally optimal point set registration by joint symmetry plane fitting," *arXiv preprint arXiv:2002.07988*, 2020.

[67] M. Brown, D. Windridge, and J.-Y. Guillemaut, "Globally optimal 2d-3d registration from points or lines without correspondences," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2111–2119.

[68] D. Campbell, L. Petersson, L. Kneip, and H. Li, "Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1–10.

[69] D. J. Campbell, L. Petersson, L. Kneip, and H. Li, "Globally-optimal inlier set maximisation for camera pose and correspondence estimation," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[70] D. Campbell, L. Petersson, L. Kneip, H. Li, and S. Gould, "The alignment of the spheres: Globally-optimal spherical mixture alignment for camera pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 796–11 806.

[71] M. Brown, D. Windridge, and J.-Y. Guillemaut, "A family of globally optimal branch-and-bound algorithms for 2d–3d correspondence-free registration," *Pattern Recognition*, vol. 93, pp. 36–54, 2019.

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.

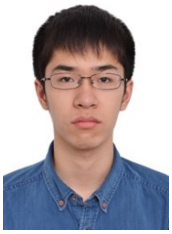
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2021.3053243, IEEE Transactions on Pattern Analysis and Machine Intelligence

[72] Y. Liu, Y. Dong, Z. Song, and M. Wang, "2d-3d point set registration based on global rotation search," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2599–2613, 2018.
[73] R. I. Hartley and F. Kahl, "Global optimization through searching rotation space and optimal estimation of the essential matrix," in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.
[74] J.-H. Kim, H. Li, and R. Hartley, "Motion estimation for multi-camera systems using global optimization," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.
[75] O. Enqvist and F. Kahl, "Two view geometry estimation with outliers." in *British Machine Vision Conference (BMVC)*, vol. 2, 2009, p. 3.
[76] R. I. Hartley and F. Kahl, "Global optimization through rotation space search," *International Journal of Computer Vision*, vol. 82, no. 1, pp. 64–79, 2009.
[77] Y. Zheng, S. Sugimoto, and M. Okutomi, "A branch and contract algorithm for globally optimal fundamental matrix estimation," in *CVPR 2011*. IEEE, 2011, pp. 2953–2960.
[78] J. Yang, H. Li, and Y. Jia, "Optimal essential matrix estimation via inlier-set maximization," in *European Conference on Computer Vision*. Springer, 2014, pp. 111–126.
[79] L. Gao, J. Su, J. Cui, X. Zeng, X. Peng, and L. Kneip, "Efficient globally-optimal correspondence-less visual odometry for planar ground vehicles," in *2020 International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2696–2702.
[80] B. K. Horn and B. G. Schunck, "Determining optical flow," in *Techniques and Applications of Image Understanding*, vol. 281. International Society for Optics and Photonics, 1981, pp. 319–331.
[81] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision," 1981.
[82] X. Peng, J. Cui, L. Kneip *et al.*, "Articulated multi-perspective cameras and their application to truck motion estimation," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019.
[83] K. Huang, Y. Wang, and L. Kneip, "Motion estimation of non-holonomic ground vehicles from a single feature correspondence measured over n views," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
[84] Q. Xu, A. G. Chavez, H. Bülow, A. Birk, and S. Schwertfeger, "Improved fourier mellin invariant for robust rotation estimation with omni-cameras," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 320–324.
[85] T. Delbruck, V. Villanueva, and L. Longinotti, "Integration of dynamic vision sensor with inertial measurement unit for electronically stabilized event-based vision," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2014, pp. 2636–2639.
[86] H. Kim, A. Handa, R. Benosman, S. H. Ieng, and A. J. Davison, "Simultaneous mosaicing and tracking with an event camera," in *British Machine Vision Conference (BMVC)*, 2014.
[87] D. E. Holmgren, "An invitation to 3-d vision: from images to geometric models," *Photogrammetric Record*, vol. 19, no. 108, pp. 415–416, 2004.
[88] 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.
[89] C. Brandli, R. Berner, M. Yang, S. Liu, and T. Delbruck, "A 240 x 180 130 db 3 us latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014.

ACKNOWLEDGMENTS

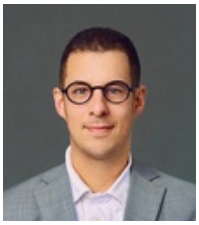
The authors would like to thank Prof. Kyros Kutulakos for his kind advice and Mr. Daqi Liu who kindly offered his code. The authors would also like to thank the fundings sponsored by Natural Science Foundation of Shanghai (grant number: 19ZR1434000) and Natural Science Foundation of China (grant number: 61950410612).

Xin Peng is a Ph.D. student in the Mobile Perception Lab at ShanghaiTech University, China. She received a B.Eng. degree in opto-electronic technology in 2015 from the University of Electronic Science and Technology, China. Her research interests include event-based vision, geometric vision, pose estimation, global optimisation and SLAM.



Ling Gao is a master student in the Mobile Perception Lab at ShanghaiTech University. His research interests consist of Event-based Vision, Robotic Perception, and Visual SLAM. He received a Bachelor of Engineering degree from ShanghaiTech University in 2019. He was a visiting student under the Yale Visiting International Student Program (Y-VISP). *Is he a robot?*

Yifu Wang is a Ph.D. candidate in Research School of Engineering, the Australian National University. He received the B.Eng. degree in Automation from Beijing Institute of Technology, Beijing, China, and B.Eng.(Hons) degree from Australian National University in 2015. His research interests include visual odometry / SLAM, geometric vision, and dynamic vision sensors. He is a recipient of the Best Student Paper award at the 11th International Conference on Computer Vision Systems (ICVS 2017).



Laurent Kneip Laurent Kneip graduated as a Diplom-Ingenieur in mechatronical engineering from the Friedrich-Alexander University Erlangen/Nürnberg in 2008. He received a doctoral degree in robotics and computer vision from ETH Zurich in 2013, where he was a member of the Autonomous Systems Lab (ASL) under the direction of Prof Roland Siegwart. His co-advisors were Prof Marc Pollefeys and Prof Davide Scaramuzza. He then served as a lecturer and senior researcher at the Research School of Engineering at the Australian National University, where he collaborated with Prof Richard Hartley and Prof Hongdong Li. In 2015, he was awarded the prestigious Discovery Early Career Researcher Award (DECRA) from the Australian Research Council (ARC), and he also served as an Associate Investigator of the ARC Centre of Excellence for Robotic Vision. His contribution at ICCV 2017 received the Marr Prize award (honorable mention). Laurent Kneip joined the School of Information Science and Technology at ShanghaiTech University in 2017, where he currently holds a tenured position as an Associate Professor. He is the founder of the Mobile Perception Lab and director of the ShanghaiTech Automation and Robotics Center.

0162-8828 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: ShanghaiTech University. Downloaded on January 22,2021 at 11:55:07 UTC from IEEE Xplore. Restrictions apply.