

eWand: An extrinsic calibration framework for wide baseline frame-based and event-based camera systems

Thomas Gossard*, Andreas Ziegler*, Levin Kolmar, Jonas Tebbe and Andreas Zell

Abstract—Accurate calibration is crucial for using multiple cameras to triangulate the position of objects precisely. However, it is also a time-consuming process that needs to be repeated for every displacement of the cameras. The standard approach is to use a printed pattern with known geometry to estimate the intrinsic and extrinsic parameters of the cameras. The same idea can be applied to event-based cameras, though it requires extra work. By using frame reconstruction from events, a printed pattern can be detected. A blinking pattern can also be displayed on a screen. Then, the pattern can be directly detected from the events. Such calibration methods can provide accurate intrinsic calibration for both frame- and event-based cameras. However, using 2D patterns has several limitations for multi-camera extrinsic calibration, with cameras possessing highly different points of view and a wide baseline. The 2D pattern can only be detected from one direction and needs to be of significant size to compensate for its distance to the camera. This makes the extrinsic calibration time-consuming and cumbersome. To overcome these limitations, we propose eWand, a new method that uses blinking LEDs inside opaque spheres instead of a printed or displayed pattern. Our method provides a faster, easier-to-use extrinsic calibration approach that maintains high accuracy for both event- and frame-based cameras.

SUPPLEMENTARY MATERIAL

The project’s code and additional resources are available at: <https://cogsys-tuebingen.github.io/ewand/>

I. INTRODUCTION

Before using cameras for robotic applications, one needs to calibrate them first. Camera calibration is the process of estimating the intrinsic camera parameters (the camera matrix \mathbf{K} and the distortion coefficients \mathbf{d}) as well as the extrinsic camera parameters (the rotation \mathbf{R} and translation \mathbf{t} of the camera with respect to a reference frame) when multiple cameras are used. Thus, after calibrating a camera system, we know how a point in the 3D space will be projected into the 2D captured images. This enables the use of computer vision algorithms like triangulation or Perspective-n-Point.

Camera calibration for conventional cameras is a well-studied subject, with literature published over multiple decades [1][2][3]. The current best practice is to capture images of a printed pattern in various positions and orientations. The most popular patterns are the checkerboard, the asymmetrical circle grid, and AprilTags. Calibration with

* Equal contribution

The authors are with the Cognitive Systems Group, Dept. Informatics, University of Tuebingen. Corresponding author thomas.gossard@uni-tuebingen.de.

This work was funded by Sony AI.

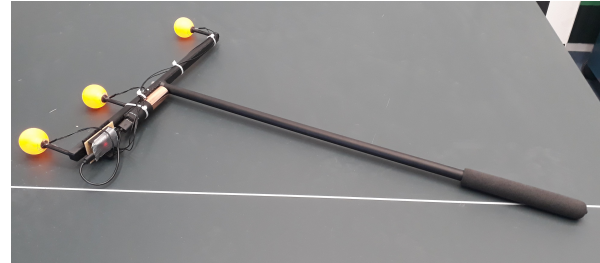


Fig. 1: Our eWand, composed of three of our markers.

such patterns work fine for a setup with a single camera or with cameras with a very similar field of view, e.g., stereo-vision setups. However, calibrating multiple cameras with different viewpoints can be problematic for two reasons. First, capturing images where the pattern is recognized by all cameras is not always possible, requiring calibration of cameras by pairs. When optimizing camera pairs individually, the pairwise optimal solution often does not coincide with the optimal solution using all cameras combined. Second, the pattern often has to be significantly tilted to be visible by multiple cameras, leading to difficulties in pattern detection and might lower the accuracy of the marker localization, potentially resulting in an inaccurate calibration. This is the reason why for such camera setups, 3D markers (such as spheres) are preferred instead of 2D markers. Sphere markers are always identically viewed, independent of the point of view. Commercial motion capture tools such as Optitrack¹ or Vicon² use “wands” to calibrate their cameras. Their infrared (IR) cameras capture the positions of IR-reflective spheres mounted on a stick. Though both Optitrack and Vicon are the industrial standard in terms of accuracy for motion capture, these systems are also quite expensive. Moreover, they require extra steps if they are to be used to calibrate a setup with standard RGB cameras. This concept has been translated to RGB cameras, using colored spheres and color filters to extract the position of the spheres [4]. However, such wands cannot be directly used for event-based cameras for calibration.

Event-based cameras are a new type of camera that have known rapid growth and increase in popularity in recent years. Instead of synchronously capturing all pixel values at periodic intervals, like frame-based cameras, event-based cameras continuously and asynchronously emit a signal (so-called *events*) on a log-intensity change above a threshold

¹<https://optitrack.com/>

²<https://www.vicon.com/>

for every pixel independently. Because of their low latency, high dynamic range, and non-redundant information flow, they offer many benefits for robotic applications [5]. They have already been used for stereo-vision [6], SLAM [7] and feature tracking [8].

As with frame-based cameras, event-based cameras also need to be calibrated first. The same calibration techniques as for conventional cameras cannot be directly applied. There are two approaches for event-based camera calibration: either frames are reconstructed from the events and then traditional calibration methods can be used [9] or the events are directly used [10][11].

Performing extrinsic calibration for frame-based and event-based cameras *jointly* requires having the pattern be detectable by both types of cameras *simultaneously*. The printed pattern and the LCD screen are the common approaches. The printed pattern is directly observable by the frame-based cameras and needs to be reconstructed for the event-based cameras. Depending on the camera used, the LCD screen can be set to either blinking mode or not. However, both also rely on 2D markers, which makes them inappropriate for a setup with multiple cameras with very different points of view. Moreover, for camera systems with a wide baseline between cameras, the pattern needs to be of significant size to be robustly detected from all cameras. It is obvious that a large LCD screen is heavy and difficult to move around. But this is also the case for printed patterns. The pattern's flatness requirement necessitates a rigid and consequently heavy support to prevent bending. This makes moving the pattern around difficult and the extrinsic calibration process slow. Wand-based calibration approaches are therefore preferred for multi-camera setups with very different points of view and wide baselines.

In order to make the wand detected by both, the frame-based and event-based cameras, we propose the use of blinking LEDs as markers. However, it isn't easy to pinpoint the center of an LED. Their small size and non-spherical shape make their detection and center regression difficult. Additionally, due to the high luminosity of the LED compared to the rest of the scene, smear, or blooming might be observed with the frame-based cameras, which makes it further challenging to estimate the center of the LED. Therefore, LED patterns tend to have a higher reprojection error [9]. To overcome these shortcomings, we propose placing these LEDs inside colored, opaque spheres. The opaque spheres act as a diffuser and are also easier to detect due to their size and symmetry. This enables both event and frame-based cameras to detect the balls. We name our method **eWand**. Using **eWand**, we can capture points over all cameras and use them to calculate the cameras' extrinsic parameters.

Contributions of this work are as follows:

- Robust and accurate markers for frame- and event-based cameras
- A detection method for said markers
- Extending the wand-based extrinsic calibration to event-based cameras

II. RELATED WORK

Since calibration of conventional cameras with printed calibration targets is a well-established topic with several toolboxes freely and commercially available [3][2][12], we will focus on the calibration of event-based cameras and wand-based calibration methods in the related work.

A. Event-camera calibration

As already mentioned, there are two methodologies for event-based camera calibration: Either frames are reconstructed from the event stream, and then traditional calibration methods can be used, or the events are directly used for the calibration. The first approach was introduced by [9]. In this work, E2VID [13] is used to reconstruct frames from the event stream. The reconstruction is a quite computationally heavy step (max frame rate of 11Hz for a resolution of 1280x720 with a NVIDIA RTX 2080 Ti GPU [13]). More recent reconstruction networks such as FireNet[14] provide faster reconstruction, being approximately three times faster than E2VID. However, even with reconstructed frames, pattern detection can be challenging. Even more so with a tilted pattern, which is often the case for patterns observed from highly different points of view. With the reconstructed frames, the kalibr calibration toolbox [3] is used to calibrate in the same way as for conventional cameras.

The second approach uses the events directly without any reconstruction for faster calibration and higher detection rates of the patterns. Early works make use of active markers such as blinking LED patterns [15][16] or blinking screens [17][18][19] to detect and extract the calibration patterns. In a more recent work [10], the authors introduced a dynamic event-based camera calibration algorithm. The approach directly calibrates given events captured during relative motion between the camera and the calibration pattern. However, the approach is not designed to calibrate multiple event-based cameras extrinsically. Although the calibration approach presented in [10] does use a circular pattern, often used for calibration of conventional cameras, it is not straightforward to calibrate event-based and frame-based cameras simultaneously. Another intrinsic calibration method was later introduced in [11]. Their approach is similar to [10] with improvements such as better clustering performance and achieving sub-pixel localization accuracy of the circular pattern.

Most of these methods could also be used for joint frame- and event-based camera calibration, if the event-based pattern extraction outputs the pattern in a data structure which can later be optimized together with the pattern extracted from the frame-based cameras. Reconstruction-based methods already rely on a pattern that can be directly used by the frame-based cameras. But as we explained, printed 2D patterns are not suited for calibrating multiple cameras with highly different points of view. Printed patterns are moreover not as straightforward as methods based on active markers. They can be set to always on for frame-based cameras or to blinking for event-based cameras. But the pattern then needs

to be perfectly static between each capture, which makes the calibration much more time-consuming and needs more care.

B. Wand-based calibration

While the majority of the camera calibration literature and available calibration tool boxes use printed calibration targets for their ease of use, there also exist methods that use a wand as calibration target [4][20]. Using a wand as a calibration target allows rapid, flexible, and accurate calibration of multiple frame-based camera systems and does not require the cameras to all face in the same direction or to have overlapping fields of view [4]. They can be used to calibrate extrinsics and the intrinsics to some degree, as shown by the Optitrack or Vicon systems (it is, however, unclear how the focal length of the cameras is estimated). While these methods were successfully applied for motion tracking with multiple infrared cameras, a wand-based calibration has never been tested to calibrate frame and event-based cameras jointly.

III. EVENT CAMERA MARKER

As previously mentioned, calibration targets are crucial for the calibration of cameras. In this section, we will describe our marker, which can be detected by frame- and event-based cameras.

Blinking LEDs have already been tested and used as markers for event cameras [9][21]. However, due to their shape, the identified position of the LED will depend on the point of view. Moreover, their small size makes detecting them from long distances difficult, as they can be mistaken for noise. We solve both of these issues by placing the LEDs in a spherical diffuser. The spherical shape of the diffuser makes the center of the marker independent of the point of view. This enables us to find the center of the marker more accurately. To ensure homogeneous and powerful lighting inside the diffuser, we used 2 high-intensity LEDs (7500MCD) oriented orthogonally. We tested the detection of our marker for different ambient luminosity, shown in fig. 2. The detection rate is the ratio of successful wand detections to the total number of observations. The marker

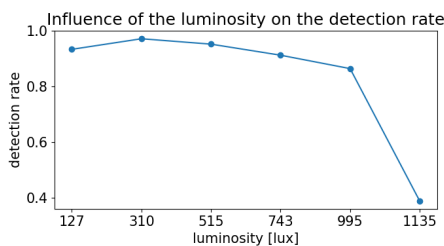


Fig. 2: Influence of the luminosity on the detection rate of the marker.

can be robustly detected for illuminations up to 1000 lux, which is the standard illumination of an office.

The blinking frequency of the blinking LED is controlled with a pulse-width modulation (PWM) signal from a micro-

controller. The duty cycle of the PWM signal was set to 50% to evenly space the on/off events for the event camera.

As for the blinking frequency of the LEDs, we tuned it to obtain the best detection rate with the event cameras' default settings. Intuitively, we want the highest possible frequencies in order to distinguish events generated from the blinking (high frequency) and events generated from the marker's movement (low frequency). However, because of the refractory period of event-based cameras, the blinking frequency cannot reach the kHz levels achievable by the microcontroller. In fig. 3, we show that frequencies higher than 600 Hz lead to lower detection rates of the markers.

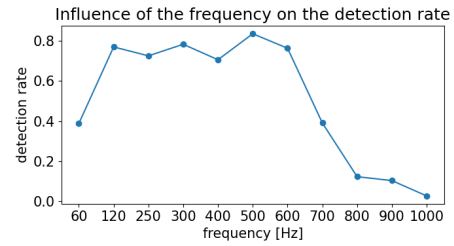


Fig. 3: Influence of the frequency on the detection rate of the marker.

A. Marker position extraction

In this section, we will describe the extraction of the marker position for frame-based cameras in section III-A.1 and for event-based cameras in section III-A.2.

1) *Frame-based camera*: For frame-based cameras, we identify the markers using a combination of color segmentation and blob detection, using the *OpenCV* implementation. However, false positives can still occur. To further filter out the detected blobs, we only keep blobs close to each other and aligned. An example of a simultaneously detected wand in each of the four frame-based cameras is visualized in fig. 4. It is important that the camera's exposure time is longer than a blinking period to avoid brightness inconsistencies in the marker's color. We also keep the exposure as low as possible to make the blinking marker stand out compared to the darker background.

2) *Event-based camera*: By tuning the biases of the event cameras (high and low pass filter mostly), it is possible to filter out most of the events except those generated by the marker.

However, this approach is not appropriate for tasks that also involve capturing events generated by other movements. Instead, we chose to extract the frequency of the generated events to filter out any frequencies that are not that of the blinking LEDs.

We rely on FrequencyCam [22] for that purpose. FrequencyCam takes an event stream as input and determines the frequency of every pixel in the field of view of the event-based camera. A visualization of the output of FrequencyCam can be seen in fig. 5b. We modified FrequencyCam to record only pixels corresponding to the frequency of our blinking LEDs,

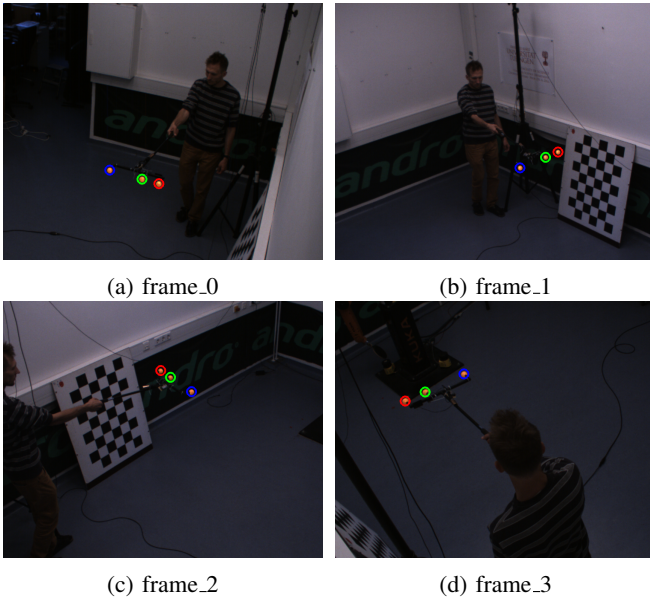


Fig. 4: Wand observed with the frame-based cameras. The position of the balls is marked with the red, green and blue circles.

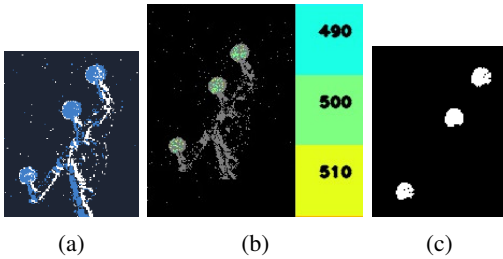


Fig. 5: (a) Accumulated events from the wand. (b) The detected frequencies of the marker positions by Frequency-Cam [22]. (c) Output of the blob detector.

employed blob detection (as shown in fig. 5c) for precise marker center determination, and exported both position data and timestamps.

IV. EWAND

In this section, we provide a general overview of our calibration method, visualized in fig. 6. We describe our calibration target in section IV-A. How we performed the intrinsic camera calibration is outlined in section IV-B. In section IV-C we detail the extrinsic calibration.

A. Calibration target

The wand is composed of three of our markers, described in section III, mounted in a line onto a stick, as shown in fig. 1. These spheres constitute the pattern's markers for the calibration. Marker size and spacing were chosen considering construction constraints and camera resolution. The spheres are aligned such that it is possible to distinguish them by the distance ratio between them. Indeed, the ratio is conserved regardless of the viewing angle. This is necessary to have marker correspondence between all cameras.

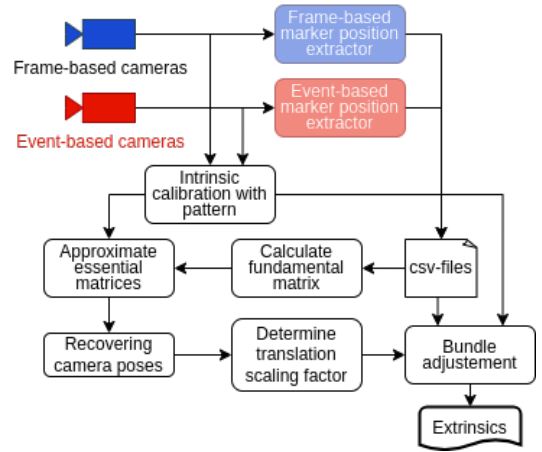


Fig. 6: An overview of our proposed calibration method.

The wand was 3D printed in one piece to ensure maximal accuracy of the markers' position. However, because it was hand-built, it is still prone to inaccuracies ($\pm 2mm$).

B. Intrinsic calibration

In order to estimate the camera matrix \mathbf{K} and the distortion coefficients \mathbf{d} , we need a 3D or 2D pattern with known geometry. Multiple methods [23][24][25] are available, but for all of them, the 3D pattern requires more than three points. The intrinsic parameters' initialization method [26] requires planar patterns and cannot be applied to the wand, which is just a line.

Moreover, the inaccuracies in the marker positions of the wand, since it is self-manufactured, make it ill-suited for intrinsic calibration, which is very sensitive to bias and noise. For these reasons, we do not rely on the wand for the intrinsic calibration. We use the intrinsic parameters calculated with the *calibrateCamera* function from OpenCV with the asymmetric circle pattern and reconstruction with E2VID[13]. Camera intrinsics remain constant and only require one-time calibration, unlike extrinsics, which must be recalculated whenever a camera is moved even slightly. This is why the ease of use of eWand makes it preferable over other methods.

C. Extrinsic calibration

For the extrinsic calibration, we start by recording images and events while moving the wand around. After capturing the data from the event- and frame-based cameras, we extract the position of the markers, explained in section III-A.

We then rely on bundle adjustment for extrinsic calibration. It can be seen as an optimization problem that requires finding the best extrinsics and 3D marker positions to minimize the reprojection errors

$$\operatorname{argmin}_{\mathbf{P}_j, \hat{\mathbf{X}}_i, \mathbf{K}, \mathbf{d}} \sum_{i=1}^n \sum_{j=1}^m (x_{i,j} - \hat{x}_{i,j}(\mathbf{P}_j, \hat{\mathbf{X}}_i, \mathbf{K}, \mathbf{d}))^2 \quad (1)$$

where $x_{i,j}$ is the i th observed marker with camera j , $\hat{\mathbf{X}}_i$ is the estimated 3D position of the i th marker observation, \mathbf{P}_j is the

projection matrix of the j^{th} camera, and $\hat{x}_{i,j}$ is the prediction of $x_{i,j}$ using \mathbf{P}_j , $\hat{\mathbf{X}}_i$ and the intrinsic camera parameters \mathbf{K} and \mathbf{d} . The projection matrix is the product of the camera matrix, that we know, and of the camera pose, that we want to estimate. We rely on Ceres [27] to perform the optimization part of the bundle adjustment.

Additional costs are added to the optimization problem to use the extra information we have from the wands' structure as a constraint. We assume here that index 0, 1 and 2 belong to the same wand.

$$\text{Linearity: } (\hat{\mathbf{X}}_0 - \hat{\mathbf{X}}_1) \times (\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_2) = 0 \quad (2)$$

$$\text{Distance: } \|\|\hat{\mathbf{X}}_0 - \hat{\mathbf{X}}_1\| - l_{ref,0}\| + \|\|\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_2\| - l_{ref,1}\| = 0 \quad (3)$$

where $l_{ref,0}$ and $l_{ref,1}$ are the real distances between the wand's markers. Non-linear optimization problems, such as bundle adjustment, require a good initialization of the extrinsic parameters so as not to get stuck in the wrong local minimum. To do so, we first need the fundamental matrix \mathbf{F}_j of a pair of cameras. The fundamental matrix is obtained using a RANSAC version of the 7-point algorithm. Once we have \mathbf{F}_j , we can calculate the essential matrix \mathbf{E}_j with the camera matrix of the two cameras \mathbf{K}_j and \mathbf{K}_k using the formula: $\mathbf{E}_j = (\mathbf{K}_k)^T \mathbf{F}_j \mathbf{K}_j$. The essential matrix can then be decomposed into possible camera rotation \mathbf{R}_j and translation \mathbf{t}_j , using SVD. However, the estimated translation vector \mathbf{t}_j is only defined up to scale, but since we know the distance between the markers on the wand, we can retrieve the correct scale k using a golden-section search.

V. EXPERIMENT

A. Cameras

Our setup contains a total of four frame-based and two event-based cameras with baselines of 3m to 5m, organized as shown in fig. 7 (schematic is up to scale).

Four FLIR Chameleon3 frame-based cameras (1280×1024 pixels) are mounted in the room's corners. As for the event-based cameras, two Prophesee EVK4 cameras (1280×720 pixels) are placed at two corners of the room. The event-based cameras are "paired" with a frame-based camera to have both type of data with comparable points of view.

The parameters of both types of cameras were set to record the best quality of data. For the frame-based cameras, we want distinguishable ball outline compared to the rest of the scene. For the event-based cameras, we want to avoid noise and capture all the marker blinks. For that, we adjusted the trigger thresholds (defining the level of the illumination change necessary to trigger an event) to filter out noise but keep the events from the blinking markers.

B. Time synchronization

One of the most critical aspects of multi-camera calibration is the time synchronization of the data capture. Because the wand is constantly moving, a frame offset could lead to erroneous triangulation of the balls. In our setup, we use a microcontroller to trigger the frame-based cameras and the

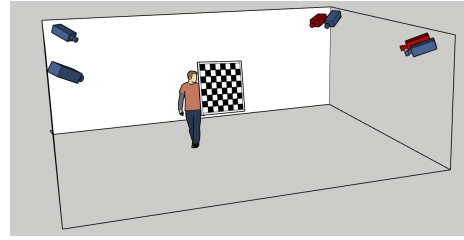


Fig. 7: Our camera setup consisting of four frame-based cameras (in blue) and two event-based cameras (in red) with baselines of 3m to 5m. Schematic is up to scale.

primary event-based camera with 50 Hz. The clocks of the primary and secondary event-based camera are synchronized via a synchronization cable between them.

C. Comparison

We compare the extrinsics estimated by our proposed calibration method, described in section V-F with two other calibration approaches. On the one hand, we use the popular calibration framework kalibr [3], described in section V-D. On the other hand, an intrinsic calibration method based on OpenCV [28] and an extrinsic calibration method using OpenCV and the optimization framework Ceres [27] to solve the calibration as a bundle adjustment problem, described in section V-E. In section V-G, we describe our evaluation of this experiment.

D. Calibration with kalibr

We performed the intrinsic and extrinsic calibration with one dataset for each calibration target, following [9].

E. Calibration via bundle adjustment with OpenCV/Ceres

For the calibration with OpenCV and bundle adjustment, we chose to perform the intrinsic and extrinsic calibration separately and on two different data sets.

Doing the calibration in two stages (first intrinsics and afterward extrinsics) has the advantages that recordings can be used which are specifically targeted for one of the two tasks. For the intrinsic calibration, the calibration target should be recorded in the cameras' whole field of view. On the other hand, for the extrinsic calibration, the calibration target has to be observed by all cameras simultaneously, not able to cover the individual cameras' whole field of view. We use two standard patterns: checkerboard and asymmetric circle grid. We chose not to include AprilTags as they tend to have a poor detection rate after reconstruction from events [9]. For the event-based cameras, E2VID [13] was used to generate reconstructed frames from the events.

a) *Intrinsic calibration:* For the intrinsic calibration, we used OpenCV functions to detect the calibration targets, followed by the *cameraCalibrate* function.

b) *Extrinsic calibration:* To perform the extrinsic calibration, we first calculated an initial estimate of the extrinsics and run a bundle adjustment afterward. For the initial estimate, we used *solvePnP*, given the observed target for every camera to get the position and rotation of the target

Camera	BA eWand (our)	kalibr (circleboard)	kalibr (checkerboard)	BA circleboard	BA checkerboard
frame_0	0.325 ± 0.257	0.274 ± 0.060	0.390 ± 0.325	0.218 ± 0.219	0.225 ± 0.204
frame_1	0.320 ± 0.271	0.311 ± 0.070	0.398 ± 0.335	0.201 ± 0.178	0.201 ± 0.185
frame_2	0.401 ± 0.361	0.168 ± 0.072	0.320 ± 0.320	0.178 ± 0.178	0.213 ± 0.193
frame_3	0.396 ± 0.389	0.159 ± 0.062	0.301 ± 0.243	0.199 ± 0.188	0.180 ± 0.165
event_0	0.603 ± 0.448	0.269 ± 0.068	0.513 ± 0.396	0.398 ± 0.365	0.459 ± 0.363
event_1	0.550 ± 0.431	0.295 ± 0.070	0.515 ± 0.405	0.469 ± 0.447	0.402 ± 0.370

TABLE I: Reprojection error (MAE) in pixels (mean and std.) for each calibration approach and camera, after the calibration. "BA" represents the bundle adjustment approach using OpenCV and Ceres, explained in section V-E.

in the camera frame. With these positions and rotation, we calculated the position and rotation of every camera with respect to the main camera. These initial estimates were then used as the starting point for a bundle adjustment problem, where we used Ceres [27] to optimize the intrinsics and extrinsics of all the cameras.

F. Calibration with the wand

We recorded a data set where we moved the wand around, covering the whole working space. The blinking frequency of the LEDs were set to 500 Hz. Since the wand is easier to detect, we were able to cover a bigger space compared to the other calibration targets. For the calibration we proceeded as explained in section IV.

G. Evaluation

For the evaluation we used two metrics: the commonly used reprojection error and a 3D position error.

1) *Reprojection errors*: Table I lists the reprojection errors obtained after calibrating with each method.

It is important to note that the measurements used differ for eWand and the other methods.

As can be seen, our method achieves similar reprojection errors as the other methods. Considering the standard deviation, the error ranges do overlap. The lower reprojection error of kalibr with the circleboard might come from the fact that kalibr does some outlier rejection, while our proposed method and OpenCV do not.

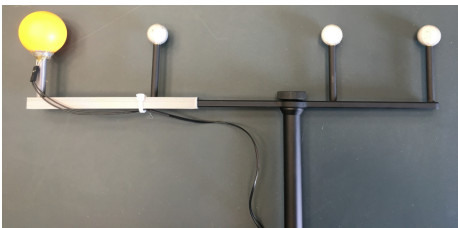


Fig. 8: Modified OptiTrack wand used for benchmarking.

2) *3D error*: For the evaluation, we attached one of our custom marker to an OptiTrack wand, as shown in fig. 8. The marker is the same as the one described in section III. We moved the OptiTrack wand with the attached marker within the workspace. Taking the 2D observations of the marker in all the cameras, we calculate the 3D position by solving a bundle adjustment problem, similar to eq. (1), where only the 3D position of the marker is optimized. Finally, we

compared the 3D position from the OptiTrack system, which we take as ground truth, to the ones we got from the bundle adjustment. The transformation between the Optitrack frame and our setup's frame is calculated by matching both generated point clouds. For the time synchronization of both, we determine the time offset by minimizing the error with the previously found transformation. As we can see in table II,

Method	3D positional error [mm] (mean and std)
BA (checkerboard)	47.0 ± 35.8
BA (circleboard)	84.9 ± 35.8
kalibr (checkerboard)	8.4 ± 10.5
kalibr (circleboard)	8.6 ± 10.4
BA eWand (our)	9.3 ± 10.4

TABLE II: The 3D position MAE [mm] (mean and std) of the different calibration methods.

the positional error of our proposed method (eWand) is very close to the results of kalibr. Considering the standard deviation, the error ranges do overlap. The bundle adjustment approach using OpenCV and Ceres performs a magnitude worse than kalibr and our approach. While kalibr does some outlier rejection, the bundle adjustment approach does not. The positional error with the checkerboard is the lowest in our experiments. One possible explanation is that on the frame reconstructions from the event data, the checkerboard detection might perform better compared to the circleboard.

VI. CONCLUSION

Calibrating camera systems is crucial, but the standard method of using printed patterns cannot be applied to event-based cameras. State-of-the-art methods include reconstructing frames or using active markers. Furthermore, calibration with the widely used calibration targets becomes challenging for multi-camera setups with a wide baseline. Our **eWand** method extends wand-based calibration to event-based cameras, using blinking LEDs to eliminate the need for frame reconstruction. **eWand** offers a simpler way to calibrate extrinsic parameters of multiple cameras, maintaining accuracy with less effort. While achieving good results in this first step, our bundle adjustment can be improved, e.g., by introducing outlier rejection. In the future, we also plan to make an improved wand with better manufacturing, for estimating intrinsic parameters as well. Overall, **eWand** aims to make setting up camera recording systems with frame- and event-based cameras more accessible and user-friendly.

REFERENCES

- [1] T. A. Clarke and J. G. Fryer, "The development of camera calibration methods and models," *The Photogrammetric Record*, vol. 16, no. 91, pp. 51–66, Apr. 1998. [Online]. Available: <https://doi.org/10.1111/0031-868x.00113>
- [2] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*. IEEE, 1999. [Online]. Available: <https://doi.org/10.1109/iccv.1999.791289>
- [3] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2016. [Online]. Available: <https://doi.org/10.1109/icra.2016.7487628>
- [4] J. Mitchelson and A. Hilton, "Wand-based multiple camera studio calibration," *Center Vision, Speech and Signal Process*, 2003.
- [5] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, 2022.
- [6] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, "Semi-dense 3d reconstruction with a stereo event camera," in *Computer Vision – ECCV 2018*. Springer International Publishing, 2018, pp. 242–258. [Online]. Available: https://doi.org/10.1007/978-3-030-01246-5_15
- [7] J. Hidalgo-Carrio, G. Gallego, and D. Scaramuzza, "Event-aided direct sparse odometry," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2022. [Online]. Available: <https://doi.org/10.1109/cvpr52688.2022.00569>
- [8] P. Chibber, E. Perot, A. Sironi, and V. Lepetit, "Detecting stable keypoints from events through image gradient prediction," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, June 2021. [Online]. Available: <https://doi.org/10.1109/cvprw53098.2021.00153>
- [9] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza, "How to calibrate your event camera," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, June 2021. [Online]. Available: <https://doi.org/10.1109/cvprw53098.2021.00155>
- [10] K. Huang, Y. Wang, and L. Kneip, "Dynamic event camera calibration," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 7021–7028.
- [11] M. Salah, A. Ayyad, M. Humais, D. Gehrig, A. Abusafieh, L. Seneviratne, D. Scaramuzza, and Y. Zweiri, "E-calib: A fast, robust and accurate calibration toolbox for event cameras," 2023.
- [12] "Tvcval: Deploy multi-sensor calibration," <https://www.tangramvision.com/sdk/multimodal-calibration>, 2023.
- [13] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI)*, 2019. [Online]. Available: http://rpg.ifi.uzh.ch/docs/TPAMI19_Rebecq.pdf
- [14] C. Scheerlinck, H. Rebecq, D. Gehrig, N. Barnes, R. Mahony, and D. Scaramuzza, "Fast image reconstruction with an event camera," in *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2020, pp. 156–163.
- [15] "Calibration toolbox by rpg group," https://github.com/uzh-rpg/rpg-dvs_ros/tree/master/dvs_calibration, 2014.
- [16] M. J. Dominguez-Morales, A. Jimenez-Fernandez, G. Jimenez-Moreno, C. Conde, E. Cabello, and A. Linares-Barranco, "Bio-inspired stereo vision calibration for dynamic vision sensors," *IEEE Access*, vol. 7, pp. 138 415–138 425, 2019. [Online]. Available: <https://doi.org/10.1109/access.2019.2943160>
- [17] E. Mueggler, B. Huber, and D. Scaramuzza, "Event-based, 6-DOF pose tracking for high-speed maneuvers," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Sept. 2014. [Online]. Available: <https://doi.org/10.1109/iros.2014.6942940>
- [18] "Calibration toolbox by vlo group," <https://github.com/VLOGroup/dvs-calibration>, 2017.
- [19] "Calibration toolbox by prophesee," https://docs.prophesee.ai/metavision_sdk/modules/calibration/guides/intrinsics.html, 2022.
- [20] K.-Y. Shin and J. H. Mun, "A multi-camera calibration method using a 3-axis frame and wand," *International Journal of Precision Engineering and Manufacturing*, vol. 13, no. 2, pp. 283–289, Feb. 2012. [Online]. Available: <https://doi.org/10.1007/s12541-012-0035-1>
- [21] M. LLC. (2023) Active marker 2d tracking sample - metavision sdk docs 4.3.0 documentation. [Online]. Available: https://docs.prophesee.ai/stable/samples/modules/cv/active_marker_2d_tracking_cpp.html
- [22] B. Pfrommer, "Frequency cam: Imaging periodic signals in real-time," 2022. [Online]. Available: <https://arxiv.org/abs/2211.00198>
- [23] R. I. Hartley, "Extraction of focal lengths from the fundamental matrix," 2001.
- [24] S. Bougnoux, "From projective to euclidean space under any practical situation, a criticism of self-calibration," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. Narosa Publishing House. [Online]. Available: <https://doi.org/10.1109/iccv.1998.710808>
- [25] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. USA: Cambridge University Press, 2003.
- [26] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000. [Online]. Available: <https://doi.org/10.1109/34.888718>
- [27] S. Agarwal, K. Mierle, and T. C. S. Team, "Ceres Solver," 3 2022. [Online]. Available: <https://github.com/ceres-solver/ceres-solver>
- [28] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.