Research on the Static Response Mechanism of Event Cameras

and Their Applications in Static Scenes

Qiyao Gao

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Mechanical Engineering

University of Washington

2023

Committee:

Xu Chen

Santosh Devasia

Byron Boots

Program Authorized to Offer Degree:

Mechanical Engineering

University of Washington

Abstract

Research on Static Response Mechanism of Event Camera
and Its Application in Static Scenes

Qiyao Gao

Chair of the Supervisory Committee:
Xu Chen
Mechanical Engineering

Event cameras are novel sensors that have gained significant importance in the field of dynamic vision since the first prototype in 2008. Unlike conventional frame-based cameras, event cameras exhibit independent and asynchronous responses at each pixel to changes in brightness. This unique characteristic allows them to have microsecond time resolution and low latency. Currently, event cameras are widely deployed for dynamic vision tasks such as high-speed motion capture. However, their potential in static scenes remains largely unexplored and not well understood. This research aims to address this gap by delving into the principles and applications of event cameras in static themes. This work explains the phenomenon that event cameras respond differently to static objects of varying brightness. Specifically, a theoretical model is derived, linking event generation to photon absorption rate, and a new concept of static sensitivity is introduced to quantitatively analyze the phenomenon. In addition to the theoretical work, we propose two techniques to unleash the potential of event cameras in static scenes. First, a static calibration method is developed for event cameras, achieving a pattern detection ratio of 0.98 and an average reprojection error of 0.13. Second, we propose to utilize event

cameras to detect transparent objects, achieving the successful detection and grasping of a transparent cube. The efficacy and robustness of these techniques are validated through extensive experiments.

TABLE OF CONTENTS

Chapter 1 Introduction 1

1.1 Thesis Introduction 1

1.2 Challenges of Frame-Based Cameras 2

1.3 Introduction to Event Cameras 3

1.4 Experimentation Setup and Workflow 4

Chapter 2 Theoretical Analysis 5

2.1 Static Sensitivity of Event Cameras 5

2.2 Analysis of Circuit Structure 7

2.3 Proposed Static Sensitivity Analysis Formula 8

2.4 Photon Discreteness Theory 10

Chapter 3 Experiment and Analysis 10

3.1 Event Counting Experiment 10

3.2 Experiment on the Influence of Surface Reflectance 12

Chapter 4 Static Calibration of Event Cameras 14

4.1 Existing Calibration Methods for Event Cameras 14

4.2 Proposed Static Calibration for Event Cameras 15

4.3 Advantages of the Static-Calibration Method 18

Chapter 5 Transparent Object Detection Based On Event Cameras 19

5.1 Related Work on Transparent Object Detection 19

5.2 Proposed Detection of Transparent Objects Using Event Cameras 20

Chapter 6 Conclusion 23

Chapter 7 Future Work 24

Reference 25

Acknowledgments

First and foremost, I express deep gratitude to Professor Xu Chen for affording me the privilege of being part of the MACS laboratory. Over the past two years, my journey of research and learning has been enriched with invaluable knowledge and experience. I also extend my heartfelt appreciation to my fellow laboratory members, whose unwavering support significantly propelled the advancement of my projects.

I extend my sincere thanks to Prof. Santosh Devasia and Prof. Byron Boots for their gracious participation in my committee, which means a lot to me. Furthermore, I am indebted to my family and friends, with special mention to my father, whose unwavering encouragement has consistently motivated me to move forward, propelling me to reach higher goals.

Chapter 1 Introduction

1.1 Thesis Introduction

Event cameras are neuromorphic sensors that respond to changes in light intensity within their field of view. They offer several advantages, such as microsecond time resolution and low latency. These impressive characteristics are achieved by departing from the traditional frame-based approach and instead, modeling three crucial properties of biological vision. The initial prototype of the event camera was developed in 2008 using CMOS technology [1], with each pixel being equipped with individual amplification circuits. This unique design renders event cameras highly sensitive to brightness changes, making them exceptionally adept at capturing high-speed motion.

The excellent performance of event cameras in the field of dynamic vision has caused researchers to overlook their potential applications value in static themes. For instance, event cameras possess dynamic response ranges of up to 140 dB [2], allowing them to effortlessly handle challenging lighting conditions. This inspired me to explore the possibility of utilizing event cameras to address difficulties that frame cameras struggle to handle.

This project aims to investigate the static response mechanism of event cameras and unlock their potential applications in static scenes. To achieve this, we introduce a new concept called "static sensitivity" to characterize how event cameras respond to objects with different surface brightness levels. Furthermore, a theory is proposed, establishing the connection between static sensitivity and photon absorption rate, along with a specific calculation formula. Through this research, the crucial role of object surface reflectance in effectively controlling the static response of event cameras is demonstrated, leading to the development of a general method for detecting static objects using event cameras.

In addition to the theoretical work, this thesis presents two novel techniques. The first one involves static calibration of event cameras. Compared to current mainstream calibration methods that rely on active light sources, the static calibration method requires no external equipment. The whole process is accomplished by detecting static calibration targets and optimizing the raw frames through image processing algorithms from OpenCV library. In practice, it realized a pattern detection rate of 0.98 and an average reprojection error of 0.1356, outperforming the current state-of-the-art calibration techniques, which have a pattern detection ratio of 0.65 and a reprojection error of 0.21.

The second technique proposed in this thesis involves utilizing event cameras to detect transparent objects. The presence of specular reflections on the surface of transparent objects can cause loss of events within the event camera's field of view, making them distinguishable from the background. In this project, we integrated an event camera onto the universal robot UR5e manipulator to create a visual-based grasp system and achieved the grasping of a transparent cube.

This exploration of event camera applications in static scenes unleashes their full potential and presents innovative solutions for tasks that prove challenging for frame cameras. I believe that this work can be further expanded upon, and I discuss possible future plans for this project at the end of the thesis.

1.2 Challenges of Frame-Based Cameras

Frame-based cameras have been the primary tool for capturing images since their inception and have become a cornerstone of the field of computer vision. As cameras are deployed in daily life and production, we place more demands on these devices. However, the global shutter mechanism and some inherent design limit their performance in certain fields such as high-speed motion detection. Frame cameras currently face two main challenges: motion blur and low dynamic range.

Motion blur predominantly occurs when capturing high-speed motion. When an object moves too quickly, each frame captured by the camera appears blurred, as shown in Figure 1. This issue is not attributed to the precision of the device itself, but rather to the fact that frame cameras capture the entire image at a fixed frequency, limiting their temporal resolution to the millisecond level. Although increasing the frame rate can alleviate this problem, it introduces another concern: a higher frame rate leads to a significant amount of redundant data in global shooting. Additionally, image distortion caused by equipment often proves difficult for algorithms to restore, resulting in unsatisfactory performance of frame cameras in dynamic vision applications.

Figure 1: Motion Blur

Figure 2: Low Dynamic Range

Dynamic range refers to the ratio between the lightest and darkest parts of an image, ranging from pure black to the brightest white. Typically, frame cameras have a dynamic range of around 60 dB, which can cause image corruption in photos taken in extremely bright or dark environments, as demonstrated in Figure 2. This drawback cannot be remedied by increasing camera resolution or employing advanced algorithms since it stems from the inherent global shutter mechanism of frame cameras.

1.3 Introduction to Event Cameras

To overcome the challenges mentioned above, researchers have developed the event camera, which measures changes of light intensity. Once the change surpasses the preset global threshold, the active pixel will generate an "event" package that comprises essential information:

timestamp, pixel coordinates, and polarity.

The mainstream application of event cameras is almost concentrated in the field of dynamic vision. For instance, Davide Falanga et al. [3] equipped a drone with an event camera, enabling it to navigate through dynamic obstacles at a speed of 10m/s. Their algorithm compensates for the event camera's ego-motion, allowing it to distinguish between static and dynamic objects effectively. Olivier Bichler et al. [4] proposed to train spiking neural networks using spike-timing-dependent plasticity. This approach enables the extraction of temporally correlated features, like car trajectories, from the event streams. Henri Rebecq et al. [5] have proposed a recurrent network that learns to directly reconstruct videos from event streams. These studies showcase the versatility and potential of event cameras in addressing various dynamic vision challenges.

While event cameras have many advantages, its further development encounters several challenges, the most prominent being the scarcity of compatible algorithms. Traditional computer vision algorithms tailored for image sequences are not well-suited for event cameras since their output is in the form of Address Events (AEs). Currently, two primary approaches are employed to handle event camera data: developing algorithms that are directly compatible with the AEs output or converting event streams into conventional image formats for processing. Both methods represent the mainstream efforts in addressing this challenge.

1.4 Experimentation Setup and Workflow

One of the primary objectives of this project is to successfully grasp a transparent cube. To achieve this, we have combined an event camera, a depth camera, and a UR5e manipulator into a system, as illustrated in Figure 3.

The project workflow, depicted in Figure 4, can be divided into two main phases: data collection and data processing. In the data collection phase, we have installed Metavision SDK 2.3 on Ubuntu 20.04 Focal Fossa, which is a ROS driver provided by Prophesee, enabling the reception of raw data from event cameras and its subsequent publication as ROS topics.

To facilitate the conversion of the sparse event stream output into dense frames, we have developed a frame generator. This component subscribes to event stream-related topics published by the SDK driver package, from which it extracts position information. Through the accumulation of event data over a period of seconds, a complete frame is constructed.

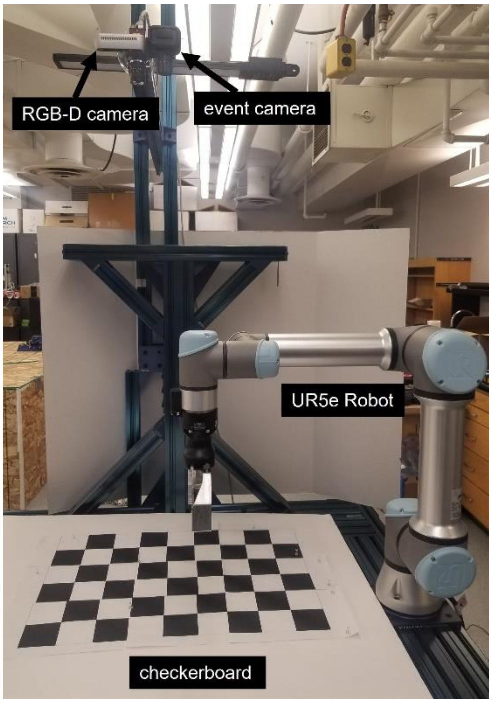

Figure 3: Experimentation setup

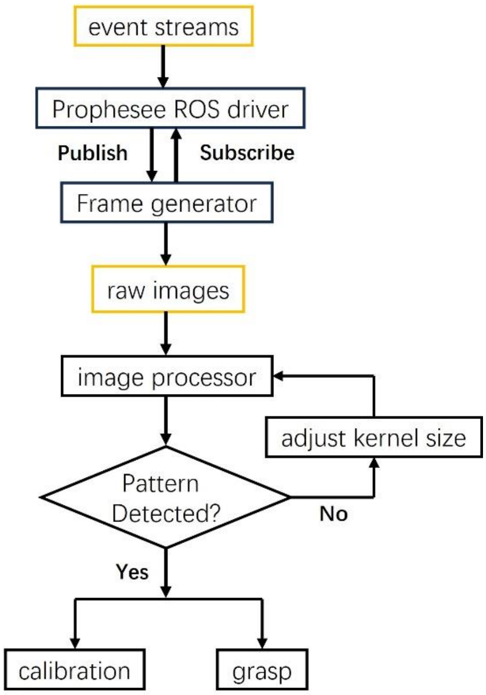

Figure 4: Workflow of this project

In the data processing phase, we designed an image processing algorithm aimed at enhancing the raw frames to fulfill the criteria for pattern detection. It is critical to obtain frames with good detection features for the subsequent calibration and grasping. This phase operates as a cyclic process, wherein the kernel size of each function can be dynamically adjusted to optimize results.

Chapter 2 Theoretical Analysis

2.1 Static Sensitivity of Event Cameras

In the conventional perspective, event cameras were primarily thought to respond solely to changes in brightness. However, through experimentation, we uncovered an interesting phenomenon: the event camera displayed an abnormal sensitivity to dark objects, as showcased in Figure 5. Within the camera's field of view, an array of objects boasting diverse colors and materials were present. Upon utilizing the event camera to simultaneously detect these objects and accumulating the event stream over a span of 10 seconds to create a frame, a remarkable pattern emerged. Namely, objects with relatively lower surface brightness exhibited a heightened capability to activate the event camera, leading to an increased generation of events.

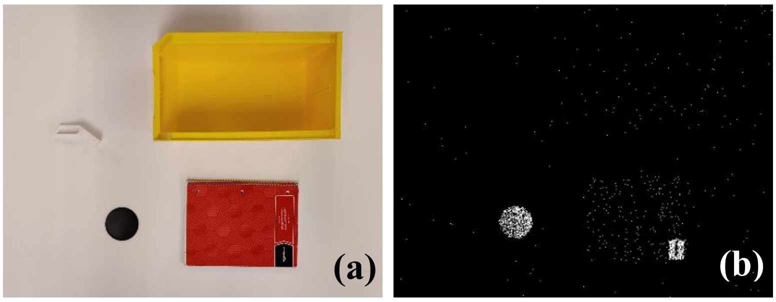

Figure 5: Event-based image generated with 10 seconds of exposure.

This peculiar tendency to selectively respond to darker objects while disregarding others seized my attention. My intuition strongly suggests that the object surface's brightness significantly influences the static response outcomes of the event camera.

To substantiate my hypothesis, we devised an experiment. Initially, we generated a series of samples distinguished by varying grayscale values, arranged in accordance with an arithmetic sequence—a representation depicted in Figure 6 (a). Subsequently, employing an event camera, we conducted a simultaneous detection of these samples, and the ensuing outcomes are demonstrated in Figure 6 (b).

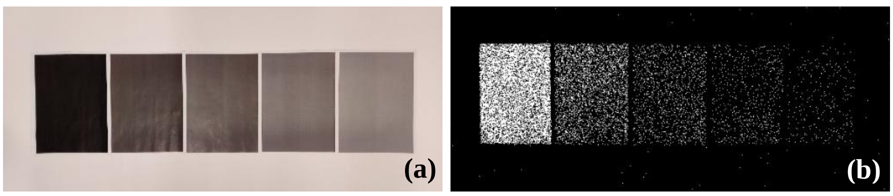

Figure 6: a). Printing papers with zones of different grayscale values.
b). 10-second accumulated output of the event camera.

Evidently, as the sample's grayscale value decreases, it becomes darker and darker, resulting in more events generated. This trend implies a robust correlation between the object surface's grayscale value and the event density.

To facilitate subsequent research, we introduce a novel concept termed "static sensitivity." This concept quantitatively captures the phenomenon where event cameras exhibit heightened sensitivity to dark objects. "Static sensitivity" is defined as the count of events generated within a unit pixel area over a specific time interval. This definition is formalized in Formula 1:

$$S_s = \frac{N_e}{A \times t} \tag{1}$$

where N_e is the number of events, A is the pixel area, and t is the accumulation time.

2.2 Analysis of Circuit Structure

Given that the distinctive attributes of the event camera stem from its unique circuitry design, an analysis of its intrinsic circuitry should be necessary and important.

The basic structure of the event camera, rooted back to the first generation of silicon retina developed during 1986-1992 [6], is a logarithmic pixel circuit structure modeled on the three-layer Kufler retina and encodes pixel coordinates in the form of address events.

The circuit of the event camera consists of a fast logarithmic photoreceptor circuit, a differential circuit for amplifying changes with high precision, and an inexpensive two-

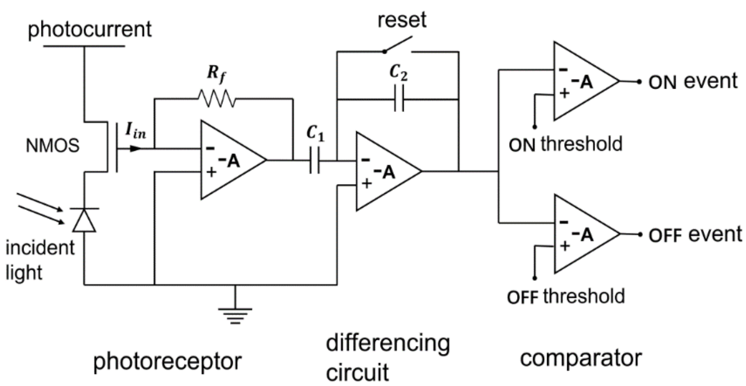

Figure 7: The schematic of the built-in circuits of the event camera (redrawn from [1])

transistor comparator, as shown in Figure 7.

The process of event generation can be divided into three steps. First, the photodiode in the photoreceptor circuit absorbs photons and the built-in saturated NMOS transistor generates a photocurrent, which is logarithmically converted to a voltage through a transimpedance configuration. Second, the differential circuit buffers the output of the photosensor through a source follower, reversely amplifies it through a capacitive feedback amplifier, and then outputs the result to a comparator. Third, the comparator compares the output to a global threshold offset from the reset voltage. If there is a change crossing the threshold, an event will be generated depending on the polarity.

2.3 Proposed Static Sensitivity Analysis Formula

Each pixel of the event camera has an independent gain circuit (through its logarithmic response) around it. The resulting DC mismatch will then be removed by balancing the output of the differential circuit to a reset level. Both the logarithmic conversion in the photoreceptor circuit and the DC elimination in the differential circuit contribute to the result that the pixel is sensitive to the temporal contrast (T_c) [1], which is defined as:

$$T_c = \frac{1}{I(t)} \times \frac{dI(t)}{dt} = \frac{dlnI(t)}{dt} \tag{2}$$

where I is the photocurrent. Temporal contrast stands as an inherent parameter of event cameras and holds significant relevance in dictating event generation rates. While there doesn't exist a precise mathematical formula to precisely estimate the event generation rate, it can be deduced through trial and error, as outlined in Formula 3.

$$f(t) \triangleq \ Event\ Rate(t) \approx \frac{T_c}{\theta} = \frac{1}{\theta}\frac{dln(I)}{dt} \tag{3}$$

where θ is the temporal contrast threshold, and the photocurrent intensity comes from:

$$I = R \times P \tag{4}$$

where R is the detector responsivity (built-in parameter), and P is the optical power, that is, the total energy of photons absorbed per unit of time:

$$P = \frac{N \times h_v}{\Delta T} \tag{5}$$

where N is the number of photons absorbed in a time window of ΔT seconds, and h_v is the energy of a single photon. Consequently, we established a correlation between the event generation rate and the photon absorption rate, which leads to the formula governing static sensitivity, depicted as follows:

$$S_s = \frac{\int_{t_1}^{t_2} f(t)dt}{A \times (t_2 - t_1)} = \frac{\frac{1}{\theta} \ln\left(\frac{N_{t2}}{N_{t1}}\right)}{A \times (t_2 - t_1)} \tag{6}$$

where N_{t1} and N_{t2} are the numbers of photons absorbed at time t_1 and t_2. For a stable light source, the incident light can be regarded as a flow of photons with uniform density, which means that the count of absorbed photons isn't subject to fluctuate dramatically over extremely brief time intervals. As a result, a correlation can be established between N1 and N2 as follows:

$$N_{t2} = N_{t1} + \Delta N \tag{7}$$

where ΔN is the increment and should be small enough to be neglectable compared to N_{t1} and N_{t2}. In this way, the static sensitivity can be simplified as follows:

$$S_s = \frac{\frac{1}{\theta} \cdot ln(1 + \frac{\Delta N}{N_{t1}})}{A \cdot (t_2 - t_1)} \tag{8}$$

2.4 Photon Discreteness Theory

It can be seen from Formula 8 that the static sensitivity of the event camera decreases as the number of absorbed photons increases. This leads me to propose a theory elucidating why event cameras exhibit heightened sensitivity to darker objects.

It is well known that the only condition required to trigger an event is the change in photocurrent of a pixel exceeds the preset global threshold. Event cameras absorb photons reflected by objects and generate photocurrents. For a dark object, it has lower reflectivity, reflecting fewer photons to the camera, causing the photocurrent to be small. In this case, any fluctuation or even noise becomes significant compared to the small photocurrent generated, making it easier to exceed the threshold and generate events.

On the other hand, bright objects reflect a larger number of photons, resulting in a large and stable photocurrent, it becomes challenging to exceed the threshold, so there will be no events being generated.

Chapter 3 Experiment and Analysis

3.1 Event Counting Experiment

The theory shows that the static sensitivity of event cameras is related to object surface reflectance. However, modeling the reflectance of physical objects directly is challenging due to various factors such as color, material, and surface topography. To facilitate quantitative analysis, we introduce grayscale values as an engineering approximation [7].

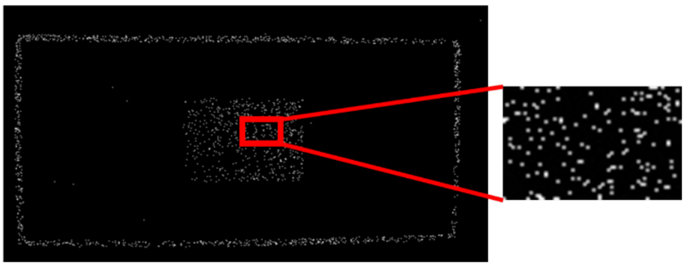

Figure 8: Print dark patches with different grayscale values and obtain the event density.

In this experiment, we first printed dark blocks with varying grayscale values in an arithmetic sequence order on a white paper (cf. each gray box in Figure 6). To ensure consistency, we utilized papers of the same material and shape across four repeated trials. From the event camera, we monitored these static surfaces, accumulated the output over a duration of 0.3 seconds, and generated an image representing the events obtained (as illustrated in Figure 8). Subsequently, we selected regions at identical locations within each image, determined the number of events contained within these regions, and calculated the event density, representing

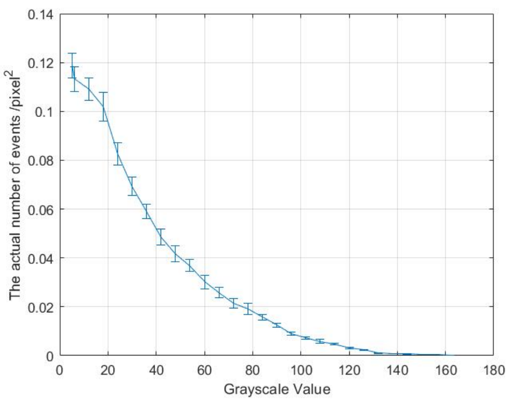

Figure 9: Relation between the measured event density and the grayscale value.

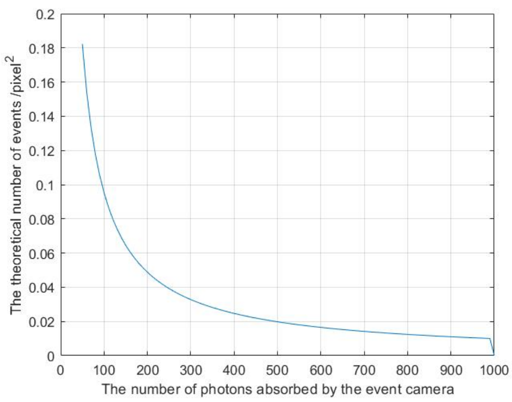

Figure 10: relationship between static sensitivity and the number of photons absorbed.

the number of events per unit time per unit pixel area. Finally, we plotted the relationship between the event density and the grayscale value and compared it with the theoretical sensitivity curve. The corresponding figures are presented in Figure 9 and Figure 10.

The coherence between the two figures is evident, and this alignment further confirms that the event density is directly affected by the grayscale value, which strengthens the reliability of our theory and formula. One thing to note is that there is a data mismatch between the two figures, which arises due to my utilization of grayscale values in lieu of the count of photons. This decision holds merit, given that both metrics symbolize the object's surface brightness. Yet, in real-world applications, the employment of grayscale values proves more expedient for quantitative assessment.

3.2 Experiment on the Influence of Surface Reflectance

With the established relationship between the grayscale value and the static sensitivity, it is possible to find a method for effectively controlling the surface brightness of objects, thereby controlling the static output of the event camera. Given that objects illuminated by the same light source can exhibit varying surface brightness due to differences in reflectivity, we suggest a strategy for proficiently managing the static output of the event camera. This involves exerting control over the reflectivity of the object's surface.

First, we printed three samples with the pixel intensity of the red, green, and blue channels set to 255, respectively, and then monitored the three paper samples using an event camera. The response results are presented in Figure 11. Compared with the red paper, the blue paper causes the event camera to generate more events.

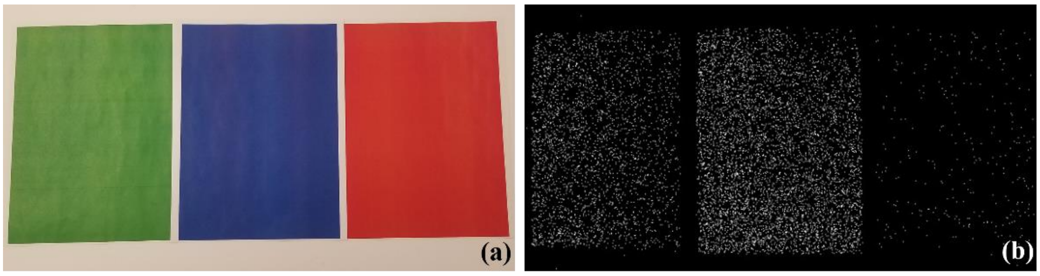

Figure 11: Response results of the event camera to different color papers with the same grayscale value.

This can be understood from the perspective of light absorption and reflection. Objects with specific colors can only reflect the light of the corresponding wavelength and absorb the rest [8], which means that under the same light source, the blue paper will reflect fewer photons because blue light owns a narrower wavelength band (450 nm to 485 nm), making the event camera more sensitive. In contrast, the green paper reflects light with wavelengths from 500 nm to 565 nm, and the red paper reflects light with wavelengths from 625 nm to 750 nm, which progressively weakens the sensitivity of the event camera.

The experimental result demonstrates that, despite having the same grayscale values on the surface of an object, the static response outcomes of the event camera will differ when the surface reflectance of the object is adjusted to produce varying numbers of reflected photons.

The second experiment consists of three control groups. In the first group, we detected a black plastic plate with a rough surface using an event camera, and the static response result exhibited a uniform event domain, as depicted in Figure 12 (d). In the second group, we applied scotch tapes to the plate, altering the surface reflectance while keeping the grayscale value unchanged. The response results are presented in Figure 12 (e), revealing a lack of events in the region covered by the scotch tape. This phenomenon becomes more pronounced in the third group, where we covered the plastic plate with a transparent bag, resulting in a significant loss of events, as depicted in Figure 12 (f). It is obvious that within the same object, different areas with varying degrees of reflectance will exhibit distinct static response outcomes.

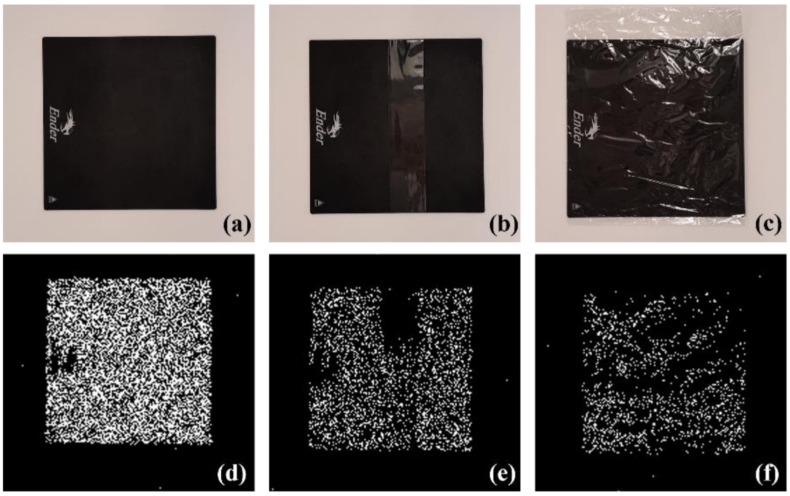

Figure 12: Upper row: a) standard blackboard b) blackboard with scotch tape in the middle, and c) blackboard covered by a transparent bag. Bottom row: corresponding response results.

The two experiments demonstrate that the static output of the event camera can be effectively controlled by modifying the reflectance of the object surface. Moreover, it opens up possibilities for utilizing event cameras in static scenes.

Chapter 4 Static Calibration of Event Cameras

4.1 Existing Calibration Methods for Event Cameras

Calibrating cameras becomes not only necessary but also critical when deploying them for vision-based control tasks. Existing calibration techniques for event cameras can be broadly categorized into two groups: those dependent on active light sources and methods that involve the conversion of event camera output formats.

For the calibration approach based on active light sources, researchers employ blinking LED lights to stimulate the event camera, inducing the generation of a sufficient quantity of events for calibration purposes. Manuel J. Domínguez-Morales et al. [9] developed an approach based on active light sources for DVS implemented over an FPGA. They utilized a grid of flashing LEDs at different depths to stimulate AER DVS-retinas, enabling the reception of 3D point information. Similarly, Elias Mueggler et al. [10] proposed a calibration approach for DVS using a computer screen with blinking patterns. They employed a blinking dots pattern for intrinsic-parameter calibration and a concentric black-and-white squares pattern for focus adjustment.

Within the methodology of converting event camera output formats, a pivotal aspect revolves around the transformation of the sparse event stream into a denser frame. A related work along this path is [11], where Manasi Muglikar et al. proposed a calibration based on image reconstruction. They employed a recurrent network to reconstruct images from event data over a certain duration and used a calibration toolbox to perform the task. This method allows event cameras to benefit from standard camera calibration techniques. This approach achieves a high pattern detection rate and low reprojection error compared to previous methods, at the cost of needing sufficient data during network training.

4.2 Proposed Static Calibration for Event Cameras

In the past few decades, many researchers have proposed a variety of frame camera calibration techniques, among which checkerboard detection is the most widely used one [12]. By identifying the inner corners of a checkerboard pattern in multiple images, we can solve the homography matrix that relates the calibration plane to the pixel plane, and then obtain essential camera parameters for accurate calibration. It is challenging to adapt this well-established technique to event cameras because event cameras do not generate frames with detailed information. However, our theoretical work has demonstrated that event cameras respond differently to objects with varying surface brightness, which enables to integrate frame camera calibration technology into the domain of event cameras.

In the proposed approach, we print an oversized checkerboard pattern with square sizes of 75mm and employ an event camera to detect this calibration target. While 75mm is not a mandatory number, considering the potential edge blur caused by subsequent image processing operations, the size of the checkerboard patterns should not be too small. The response result is depicted in Figure 13.Raw frames cannot be used directly for calibration as they are noisy and have incoherent patterns. we proposed an algorithm to process these images by calling functions from OpenCV image processing library, as shown on the left.

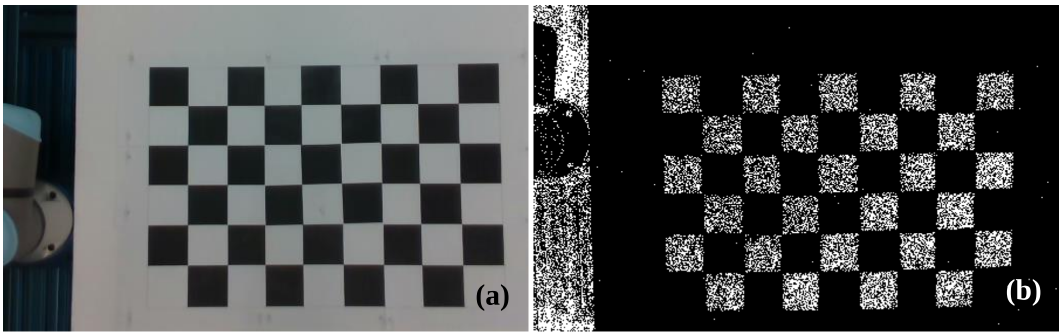

Figure 13: Checkerboard calibration images: a) RGB image from a standard camera, b) Static response result from an event camera.

In the above image processing, we first utilize Gaussian Blurring from the OpenCV smoothing images library to eliminate Gaussian noise in the image and blur the pattern edges. Next, we apply the Closing function from the morphological transformations library to fill small black holes within the object. Following that, we employ the Bilateral Filtering function from the smoothing images library to sharpen the pattern edges. This filter incorporates an additional Gaussian filter based on pixel intensity differences, allowing the preservation of intensity variations between edge pixels. Finally, we invert the intensity of each pixel in the processed image to obtain a black-and-white reversed image, resembling the results obtained with standard cameras. The processing results at each step are illustrated in Figure 14.

Algorithm 1: Image Processing Algorithm

```
Input  : The original static response results
Output: Processed images; Detection success rate
read in raw images;
while in the raw image folder do
    image 1 = cv2.GaussianBlur (raw image);
    image 2 = cv2.morphologyEx (image 1);
    image 3 = cv2.bilateralFilter (image 2);
    image 4 = ~ (image 3);
    bool, corners = cv2.findChessboardCorners()
    if bool then
        save image 4 to a new folder;
        successful detections += 1;
    else
        failed detections += 1;
    end
end
Return the new folder; the detection success rate
```

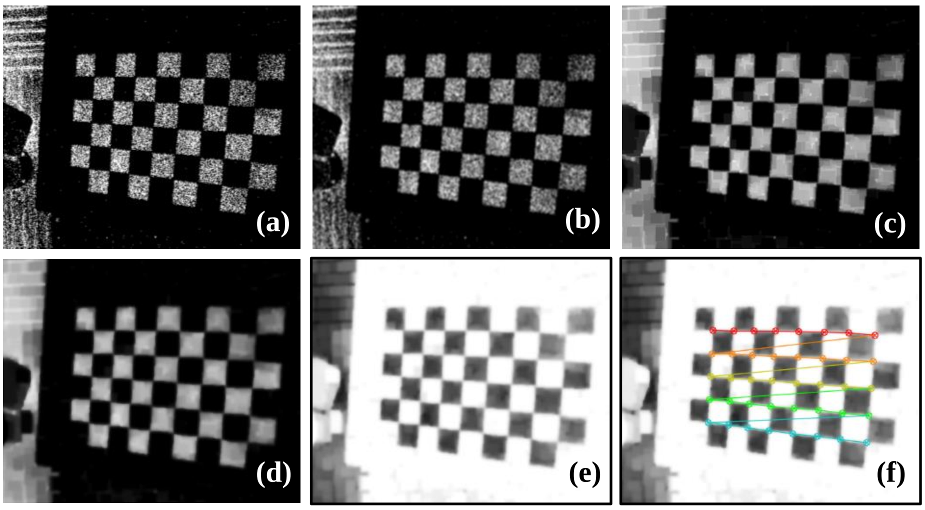

Figure 14: The processing results at each step: a) the original image; b) after applying a Gaussian Blurring function with a kernel size of 7 by 7; c) after applying a Closing function with a kernel size of 15 by 15; d) after applying a Bilateral Filtering function; e) after inverting the pixel intensities of the image; f) finding the positions of the internal corners of the checkerboard using the function

By implementing this preprocessing procedure, we enhance the raw frame and obtain images with prominent detection features. In practice, we gathered a total of 50 checkerboard images with different poses by accumulating events in a period of 20 seconds. These images were then divided into three separate test groups to assess the effectiveness of the algorithm proposed.

In the first group, we kept the kernel size constant for each function in the algorithm, namely, we used this algorithm to process all raw images. In the second group, we adaptively adjusted the kernel size of each function based on the varying characteristics of each image. In the third group, we incorporated additional image processing operations into the algorithm to enhance detection, such as applying the $erosion$ function to refine image edges. The evaluation results are presented in Table 1. In the first group, only 41 processed images were successfully detected, resulting in a detection success rate of 82%. In the second group, 47 images were successfully detected, yielding a detection success rate of 94%. Lastly, the third group achieved a detection success rate of 98%, with 49 images successfully detected.

Group	**Number of detections**	**Success rate**
Fixed kernel	41	82%
Adaptive kernel	47	94%
Additional operation	49	98%

Table 1: Detection rate of Algorithm 1

We acquired approximately 15 processed images displaying dependable features and employed Zhang's method [12] as a benchmark for calibrating the event camera. In this work, we conducted four distinct calibrations for the Prophesee EVK1-VGA event camera, and the intrinsic camera parameters are detailed in Table 2.

Parameters	α	β	γ	u_0	v_0	k_1	k_2	*reprojection error*
Group 1	557.95	574.11	0.4928	309.05	252.31	0.02417	-0.26078	0.14240
Group 2	552.18	568.18	0.2305	312.30	250.64	0.02165	-0.23314	0.12879
Group 3	553.87	569.74	0.4440	312.24	251.58	0.02216	-0.24003	0.13728
Group 4	551.92	567.82	0.2345	311.41	250.59	0.02438	-0.25761	0.13374

Table 2: Prophesee EVK1-VGA Event Camera Intrinsic Parameter Estimation. α and β are the scale factors in the image plane, γ is the parameter describing the skewness of the axes, u_0 and v_0 are the principal point coordinates, k_1 and k_2 are distortion coefficients. All these parameters were optimized using the Levenberg-Marquardt method. The reprojection error was calculated by considering the distortion.

In comparison to the research conducted by Manasi Muglikar et al. [11], who achieved a pattern detection rate of 0.65 and a reprojection error of 0.21, our algorithm demonstrates a superior performance, surpassing their results by approximately 38%.

4.3 Advantages of the Static-Calibration Method

Static calibration techniques for event cameras differ from traditional methods which rely on flashing pattern devices to trigger events, it leverages the unique static response mechanism of event cameras, getting rid of the drag of external equipment.

By converting sparse event streams into dense static response frames, we effectively apply well-established calibration techniques developed for standard cameras to event cameras, resulting in highly satisfactory outcomes. The advantages of this technique can be summarized as follows:

1. It does not require any external devices, such as LED lights or flashing screens, for calibration. The calibration can be conducted under any stable light source.
2. (Static calibration) There is no need to move cameras or targets. Sufficient events for calibration are obtained by directly detecting objects with low surface reflectance.
3. The method requires only small amounts of data to achieve a high accuracy. For the full calibration, we achieved a detection rate of 98% and a reprojection error of 0.1356.

Chapter 5 Transparent Object Detection Based On Event Cameras

5.1 Related Work on Transparent Object Detection

Transparent objects are omnipresent in our daily lives; however, achieving precise detection and localization of such objects has remained a persistent challenge in the field. This predicament is attributed not only to the ease with which transparent objects can blend into backgrounds due to their high light transmission but also to the substantial presence of specular reflections on their surfaces.

Significant advancements have unfolded within the field over the past few decades, particularly with the emergence and evolution of neural networks. For instance, Shreeyak Sajjan et al. introduced a network named ClearGrasp in their work [13]. This network employs deep convolutional networks to deduce precise 3D geometry of transparent objects using just a single RGB-D image. However, this approach comes with a tradeoff, as it necessitates the accumulation of an extensive amount of data to effectively train the network.

Guohua Chen et al. [14] proposed a method for reconstructing 3D transparent objects using stereo matching of object contours extracted from sets of RGB and infrared (IR) images. Similarly, Ulrich Klank et al. [15] developed a binocular stereo vision system that reconstructs an approximate surface of the original transparent object by matching intensity channels from time-of-flight (ToF) cameras in two viewpoints.

It is important to note that technologies relying on the perspective principle, such as the aforementioned methods, can be complex and lack robustness due to system structure and inherent errors caused by sensor position deviations. This complexity arises from the need to align and match data from different sensors to obtain accurate results.

5.2 Proposed Detection of Transparent Objects Using Event Cameras

The previous surface reflectance experiments have substantiated the profound impact of object surface reflectance on the static response outcomes of event cameras. Given that transparent objects possess higher surface reflectance compared to conventional objects due to the presence of specular reflections, we propose utilizing low-reflectance objects as backgrounds to enhance the event camera's ability to detect transparent materials.

Specifically, we position a transparent cube on a black sponge, as Figure 15 (a) depicts. After accumulating the output of the event camera for 20 seconds, a frame is generated, revealing a distinct outline of the transparent cube in the uniform event domain, as depicted in Figure 15 (b).

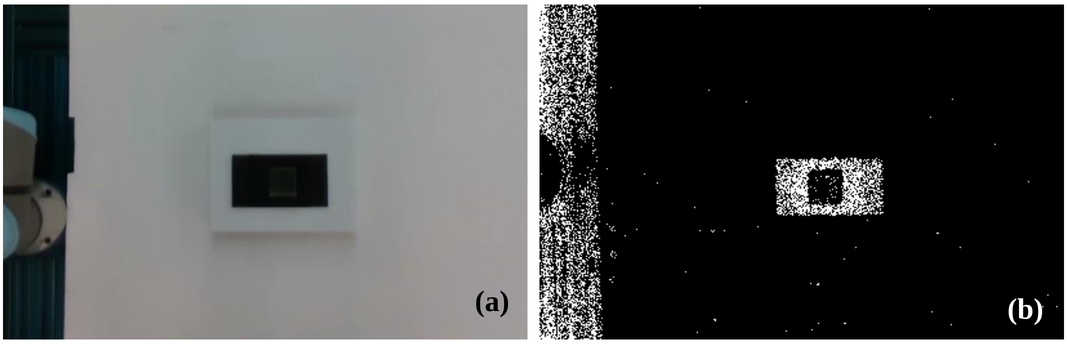

Figure 15: Detection of a transparent cube on a black sponge. a) RGB image from a standard camera b) Static response result from the event camera.

There are two important points to highlight. First, there are multiple options available for background objects. Any object that can create a contrast with the transparent object is suitable. However, for optimal results in subsequent image processing operations, it is advisable to select objects with low reflectance that can generate high-density event domains.

Second, the accumulation time is not fixed. While longer accumulation times result in denser event domains in the background and a more pronounced contrast effect, it is essential to take into consideration of the ultra-high light transmittance of transparent objects and detection efficiency. Prolonged accumulation times may lead to the reappearance of events in the areas covered by transparent objects, and thus weaken the detection effect. We documented the effects of varying accumulation times, which are depicted in Figure 16.

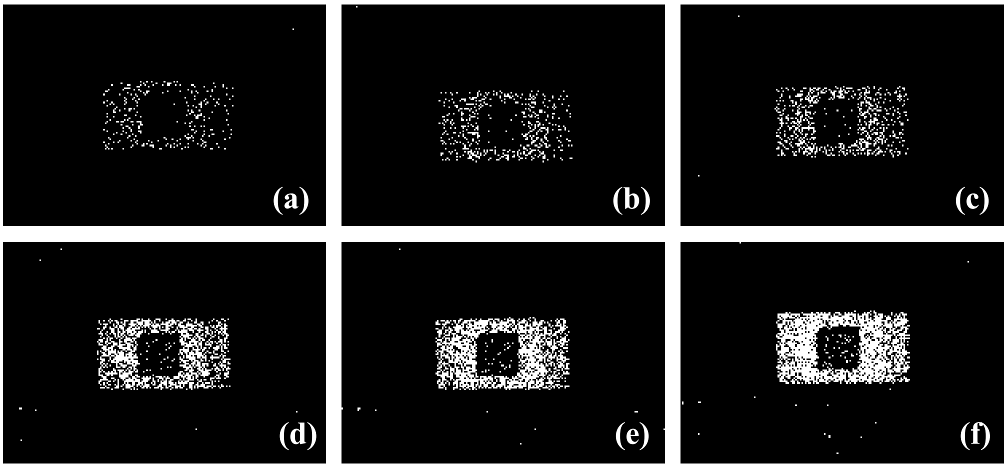

Figure 16: Static response results of transparent cube under different accumulation times: a) 0.5s b) 1s c) 2s d) 10s e) 20s f) 40s. When the cumulative time reaches 40 seconds, a significant number of events are generated in the area covered by the cube.

The acquired raw frame can be refined using Algorithm 1, followed by the application of the Hough Transform to identify the edges of the transparent cube. The outcome of this process is illustrated in Figure 17.

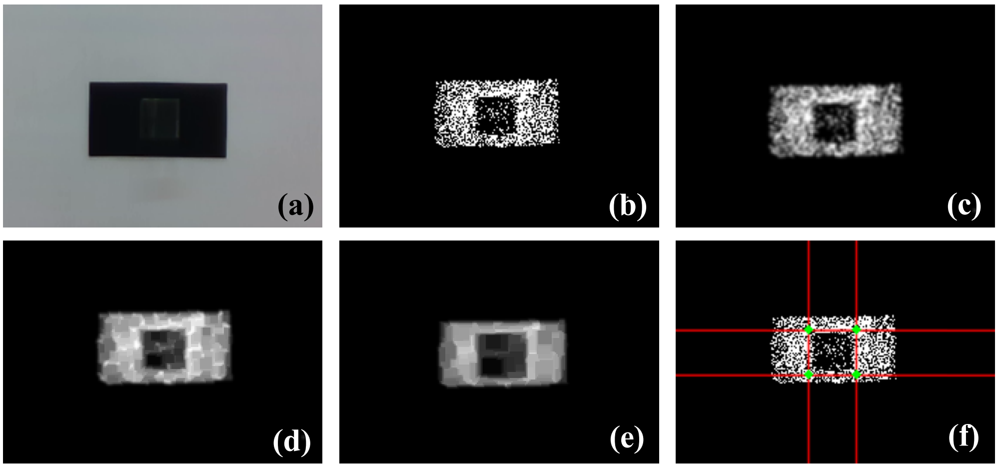

Figure 17: Processing of the static response result of the event camera for the transparent cube: a) RGB image from a standard camera, b) static response result from the event camera, c) after applying a *Gaussian Blurring* function with a kernel size of 15 by 15, d) after applying a *Closing* function with a kernel size of 15 by 15, e) after applying an *Erode* function with a kernel size of 5 by 5, f) lines are detected through *Hough Line Transform* function.

In practical implementation, we established a visual-based grasping system, comprising an event camera for precise transparent cube localization, a depth camera for distance measurement, and a UR5e manipulator for grasping operations. Through this setup, we successfully achieved the task of grasping the transparent cube, as represented in Figure 18.

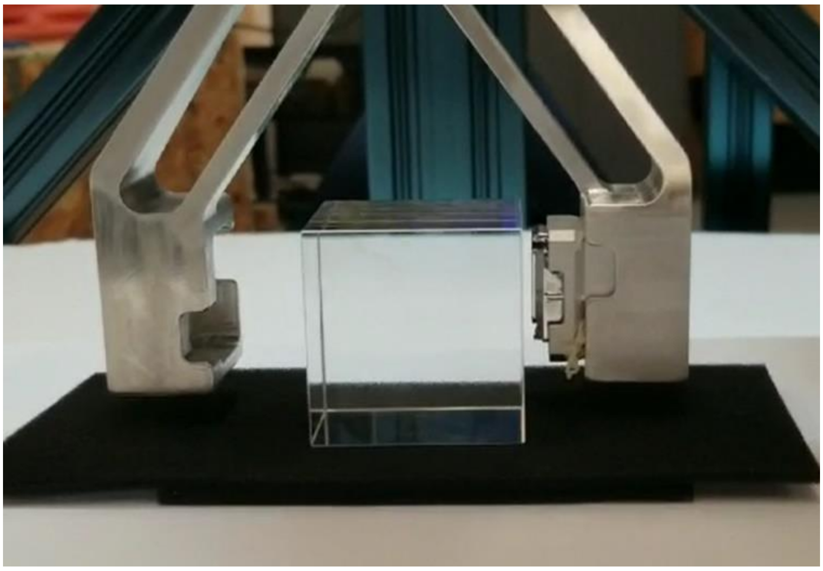

Figure 18: Grab a transparent cube using the event camera.

Chapter 6 Conclusion

This paper presented a comprehensive investigation into the response mechanism of event cameras towards static objects, marking the first systematic exploration in this area. We introduced a novel Photon Discretization Theory, which elucidates the heightened sensitivity of event cameras to darker objects. Additionally, we proposed a static sensitivity metric for event cameras and showed how to compute and verify it experimentally. Moreover, we highlighted the crucial role of object surface reflectance in regulating the static output of event cameras. Based on the theoretical work, we developed two techniques to unleash the application potential of event cameras in static scenes.

Our proposed static calibration technique enables efficient calibration of event cameras with a minimal number of steps. It achieved an impressive detection rate of 0.98 and an average reprojection error of 0.1356, surpassing the current state-of-the-art approaches. Furthermore, we proposed a novel static detection technique using event cameras based on reflectance difference, which addresses challenges that are intractable to standard cameras, such as transparent object detection, hidden object distinction, etc. We demonstrated the principle, method, effect, and experimental results of this technology.

Chapter 7 Future Work

Scene reconstruction is a important research branch of event cameras. As event cameras primarily capture variations in light intensity, their outputs inherently depict changes between the current and previous frames. Consequently, the mainstream technique for utilizing event cameras to reconstruct images or videos involves a continuous accumulation of gain frames atop an initial frame. However, generating this initial frame poses a persistent challenge.

My research has substantiated that event cameras exhibit heightened sensitivity to dark objects, presenting a viable approach for statically generating the initial frame. To elaborate, we intend to enhance the event camera's capabilities by incorporating a variable aperture. By intentionally creating a low-light environment through the limitation of incident light levels, we can harness the event camera's heightened sensitivity to dark objects to stimulate a substantial event generation. Subsequent filtering of extraneous data, including noise, will enable the extraction of pertinent information. This information can then be utilized in subsequent scene reconstruction processes, facilitating the attainment of the original frame.

Reference

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

[2] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis et al., “Event-based vision: A survey,” IEEE transactions on pattern analysis and machine intelligence, vol. 44, no. 1, pp. 154–180, 2020.

[3] D. Falanga, K. Kleber, and D. Scaramuzza, “Dynamic obstacle avoidance for quadrotors with event cameras,” Science Robotics, vol. 5, no. 40, p. eaaz9712, 2020.

[4] O. Bichler, D. Querlioz, S. J. Thorpe, J.-P. Bourgoin, and C. Gamrat, “Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity,” Neural networks, vol. 32, pp. 339–348, 2012.

[5] 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, vol. 43, no. 6, pp. 1964–1980, 2019

[6] M. Mahowald, An analog VLSI system for stereoscopic vision. Springer Science & Business Media, 1994, vol. 2

[7] M. Stevens, C. A. Párraga, I. C. Cuthill, J. C. Partridge, and T. S. Troscianko, “Using digital photography to study animal coloration,” Biological Journal of the Linnean society, vol. 90, no. 2, pp. 211–237, 2007.

[8] H. D. Young, R. A. Freedman, T. Sandin, and A. L. Ford, University physics. Addison-Wesley Reading, MA, 1996, vol. 9.

[9] M. J. Dominguez-Morales, A. Jimenez-Fernandez, G. Jiménez-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.

[10] 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, 2014, pp. 2761–2768.

[11] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza, "How to calibrate your event camera," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1403–1409.

[12] 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.

[13] S. Sajjan, M. Moore, M. Pan, G. Nagaraja, J. Lee, A. Zeng, and S. Song, "Clear grasp: 3d shape estimation of transparent objects for manipulation," in 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020, pp. 3634–3642.

[14] C. Guo-Hua, W. Jun-Yi, and Z. Ai-Jun, "Transparent object detection and location based on rgb-d camera," in Journal of Physics: Conference Series, vol. 1183, no. 1. IOP Publishing, 2019, p. 012011.

[15] U. Klank, D. Carton, and M. Beetz, "Transparent object detection and reconstruction on a mobile platform," in 2011 IEEE International Conference on Robotics and Automation. IEEE, 2011, pp. 5971–5978.