



**Titre:** Event-based Perception with Structured Light  
Title:

**Auteur:** Seyed-Ehsan Marjani-Bajestani  
Author:

**Date:** 2024

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Marjani-Bajestani, S.-E. (2024). Event-based Perception with Structured Light  
Citation: [Thèse de doctorat, Polytechnique Montréal]. PolyPublie.  
<https://publications.polymtl.ca/59466/>

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/59466/>  
PolyPublie URL:

**Directeurs de  
recherche:** Giovanni Beltrame  
Advisors:

**Programme:** Génie informatique  
Program:

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

**Event-based Perception with Structured Light**

**SEYED-EHSAN MARJANI-BAJESTANI**

Département de génie informatique et génie logiciel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*

Génie informatique

Septembre 2024

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

Cette thèse intitulée :

**Event-based Perception with Structured Light**

présentée par **Seyed-Ehsan MARJANI-BAJESTANI**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*

a été dûment acceptée par le jury d'examen constitué de :

**Christopher J. PAL**, président

**Giovanni BELTRAME**, membre et directeur de recherche

**Mohammad HAMDAQA**, membre

**François POMERLEAU**, membre externe

## DEDICATION

*To my parents...*



## ACKNOWLEDGEMENTS

I would like to express my deep and heartfelt gratitude to my supervisor, Prof. Giovanni Beltrame. It was a true pleasure working with him during my PhD. He always provided excellent advice and guidance and believed in me, even when I doubted myself. He showed what a true leader should be like with his patience, great communication skills, and passion for his work. I have learned a lot from his example and hope to follow it in the future.

To the members of the MIST lab., thank you for your teamwork and friendship. Your help made this journey both successful and enjoyable.

To my family, your support and love were my foundation throughout this process. I could not have done it without you.

## RÉSUMÉ

La précision d'un robot mobile dans la perception de son environnement est étroitement liée à la méthode utilisée pour mesurer les distances relatives. Il est crucial d'utiliser un dispositif de mesure 3D rapide et robuste si un robot en mouvement rapide utilise ces données pour créer une carte de la zone tout en se localisant simultanément (SLAM<sup>1</sup>).

Parmi les méthodes de reconstruction environnementale 3D de pointe, les approches basées sur la vision ont été largement développées. Les caméras RGB et RGB-D sont peu coûteuses et couramment utilisées en robotique. Cependant, elles présentent des limitations inhérentes : elles nécessitent un bon éclairage, souffrent de flou de mouvement, ont une plage dynamique relativement faible (pouvant entraîner une saturation en cas de changements de conditions d'éclairage) et peuvent nécessiter une large bande passante en fonction de la résolution et de la fréquence d'image.

Pour pallier ces limitations, des capteurs tels que le Light Detection And Ranging (LiDAR) et les caméras événementielles ont été introduits. Les dispositifs LiDAR émettent un rayonnement laser sur la scène et capturent les signaux réfléchis pour obtenir une représentation 3D de l'environnement, déterminant ainsi les distances des points en 3D. Malgré leur grande précision, les LiDAR ne capturent pas de données couleur et ne peuvent pas ajuster dynamiquement le compromis entre détail, précision et vitesse. Leur faible densité de sortie limite la capacité à obtenir des données plus denses sans augmenter le temps de mesure.

Les caméras événementielles (ECs) sont des capteurs bio-inspirés qui détectent les mouvements rapides et les changements de luminosité de manière asynchrone, similaire à l'œil humain. Bien que les ECs ne capturent pas d'images complètes, elles peuvent détecter le mouvement beaucoup plus rapidement que les capteurs RGB standard, ce qui les rend précieuses pour les projets nécessitant une détection rapide des mouvements. En raison de leurs avantages, les ECs sont particulièrement utiles pour les mesures rapides de profondeur 3D. Cependant, elles ne fournissent pas de données dans des situations statiques.

L'objectif est d'introduire une méthode capable de générer des nuages de points colorés avec des compromis variables de vitesse et de résolution pour créer des cartes de profondeur dans des environnements difficiles (dynamiques et faiblement éclairés), même lorsque la caméra (ou l'objet cible) est stationnaire.

Cette recherche présente une méthode utilisant une caméra événementielle et un projecteur

---

<sup>1</sup>Simultaneous Localization And Mapping

Digital Light Processing (DLP) pour capturer des événements dans l'espace 3D. Le projecteur DLP projette des motifs de lumière structurée sur la scène, variant en type, fréquence et couleur/longueur d'onde. La caméra événementielle capture les réflexions de ces motifs, permettant la création de nuages de points 3D basés sur la triangulation. Cette configuration permet également de capturer la couleur de la scène simultanément, produisant un nuage de points coloré. La commutation dynamique des motifs projetés permet de contrôler la bande passante. Le système bénéficie de l'utilisation d'une caméra monochrome haute résolution et peut incorporer des données couleur au besoin.

En utilisant cette configuration, nous avons atteint des vitesses de balayage couleur jusqu'à 1,4 kHz et des balayages de profondeur basés sur les pixels jusqu'à 4 kHz, résultant en un flux d'événements marqués avec couleur et profondeur, ainsi que des images et une sortie de nuage de points coloré. Cette méthode est applicable dans diverses conditions environnementales, qu'elles soient statiques ou dynamiques. Elle offre des mesures 3D haute résolution comparables aux LiDAR (dans la plage du millimètre), avec des vitesses de mesure plus rapides (capturant des événements en microsecondes) et inclut de manière cruciale des données couleur. Elle offre également un contrôle sur les compromis de résolution et de vitesse d'acquisition.

## ABSTRACT

The accuracy of a mobile robot in perceiving its surroundings is closely related to the method used for measuring relative distances in the environment. It is crucial to use a fast and reliable 3D measurement device when a fast-moving robot relies on this data for Simultaneous Localization and Mapping (SLAM).

Among state-of-the-art 3D environmental reconstruction methods, vision-based approaches have been highly developed. RGB and RGB-D cameras are inexpensive and commonly used in robotics. However, they have inherent limitations: they require good illumination, suffer from motion blur, have a relatively low dynamic range (leading to saturation under changing lighting conditions), and can demand high bandwidth depending on resolution and frame rate.

To address these limitations, sensors such as Light Detection And Ranging (LiDAR) and event-based cameras have been introduced. LiDAR devices emit laser light onto the scene and capture the reflected signals to obtain a 3D representation of the environment, determining distances to 3D points. Despite their high accuracy, LiDARs do not capture color data and cannot dynamically adjust the trade-off between detail, accuracy, and speed. Their sparse output limits the ability to obtain denser data without increasing measurement time.

Event-based cameras (ECs) are bio-inspired sensors that detect rapid movements and changes in brightness asynchronously, similar to the human eye. Although ECs do not capture full images, they can detect motion much faster than standard RGB sensors, making them valuable for projects requiring fast movement detection. Due to their advantages, ECs are particularly useful for fast 3D depth measurements. However, they do not provide data in static situations.

The goal is to introduce a method capable of generating colored point clouds with variable speed/resolution trade-offs for creating depth maps of challenging environments (dynamic and low-light), even when the camera (or target object) is stationary.

This research introduces a method using an EC and a Digital Light Processing (DLP) projector to capture events in 3D space. The DLP projector projects Structured Light (SL) patterns onto the scene, varying in type, frequency, and light color/wavelength. The EC captures reflections of these patterns, enabling triangulation-based 3D point cloud creation. This setup also allows for capturing the color of the scene simultaneously, producing a colorful point cloud. Dynamic switching of projected patterns enables bandwidth control. The

system benefits from using a high-resolution monochrome camera and can incorporate color data as needed.

Using this setup, we achieved color scanning speeds up to 1.4 kHz and pixel-based depth scanning up to 4 kHz, resulting in a stream of events stamped with color and depth, along with frames and a colorful point cloud output. This method is applicable across various environmental conditions, whether static or dynamic. It offers high-resolution 3D measurements comparable to LiDAR (in the millimeter range), with faster measurement speeds (capturing events in microseconds), and crucially includes color data. It also provides control over the resolution and acquisition speed trade-offs.

## TABLE OF CONTENTS

DEDICATION . . . . .	iii
ACKNOWLEDGEMENTS . . . . .	iv
RÉSUMÉ . . . . .	v
ABSTRACT . . . . .	vii
TABLE OF CONTENTS . . . . .	ix
LIST OF TABLES . . . . .	xii
LIST OF FIGURES . . . . .	xiii
LIST OF SYMBOLS AND ACRONYMS . . . . .	xvii
CHAPTER 1 INTRODUCTION . . . . .	1
1.1 Context and Motivation . . . . .	1
1.2 Problem statement . . . . .	3
1.3 Research Objectives . . . . .	4
1.3.1 Spatial resolution: (RQ1) . . . . .	5
1.3.2 Mobility limitation: (RQ2) . . . . .	5
1.3.3 Texture dependency: (RQ3) . . . . .	5
1.3.4 Acquisition speed: (RQ3 and RQ4) . . . . .	5
1.3.5 Light-Efficiency: (RQ3 and RQ4) . . . . .	6
1.3.6 Power consumption: (RQ1, RQ3 and RQ4) . . . . .	6
1.4 Novelty and Impact . . . . .	6
1.5 Research contributions . . . . .	7
1.6 Thesis structure . . . . .	8
CHAPTER 2 LITERATURE REVIEW . . . . .	9
2.1 Monochrome to Color . . . . .	9
2.2 Monocular Depth Sensing . . . . .	9
2.3 Depth Sensing via Triangulation Method . . . . .	11
2.3.1 Event-based depth sensing with multi camera . . . . .	12
2.3.2 Event-based depth sensing with SL . . . . .	12

CHAPTER 3	ARTICLE 1: EVENT-BASED RGB SENSING WITH STRUCTURED LIGHT . . . . .	16
3.1	Introduction . . . . .	17
3.2	Related Work . . . . .	19
3.3	Monochrome to color . . . . .	20
3.3.1	Color detection speed limits . . . . .	21
3.3.2	Advantages over monochrome cameras . . . . .	21
3.3.3	Advantages over color event-based cameras . . . . .	23
3.3.4	White balance and color correction . . . . .	23
3.4	ASL: Adaptive Structured Light . . . . .	26
3.5	Conclusions . . . . .	31
CHAPTER 4	ARTICLE 2: EVENT-BASED VISION FOR ROBOT SOCCER . . . . .	33
4.1	Introduction . . . . .	34
4.2	Related work . . . . .	36
4.3	Event-based camera calibration . . . . .	37
4.4	Dataset . . . . .	37
4.5	Conclusion . . . . .	39
4.6	Acknowledgment . . . . .	39
CHAPTER 5	ARTICLE 3: E-RGB-D: REAL-TIME EVENT-BASED PERCEPTION WITH STRUCTURED LIGHT . . . . .	41
5.1	Introduction . . . . .	42
5.2	Monochrome to Color . . . . .	45
5.3	Depth Sensing via Triangulation . . . . .	45
5.3.1	Event-based depth sensing with SL . . . . .	45
5.4	ASL: Adaptive Structured Light . . . . .	49
5.4.1	Color detection . . . . .	49
5.4.2	Depth detection . . . . .	51
5.5	Experiments . . . . .	54
5.5.1	Setup . . . . .	55
5.5.2	Baseline and Ground Truth . . . . .	56
5.5.3	Results . . . . .	57
5.6	Conclusion . . . . .	59
CHAPTER 6	GENERAL DISCUSSION . . . . .	67
6.1	Features and Advantage of the Proposed Method . . . . .	67

6.1.1	Detecting color with monochrome camera . . . . .	67
6.1.2	Real-time depth detection per pixel . . . . .	67
6.1.3	Controlling trade-off between speed and details . . . . .	67
6.2	Technical Challenges and Solutions . . . . .	68
6.2.1	White balance . . . . .	68
6.2.2	Field of view . . . . .	68
6.2.3	Black objects . . . . .	69
6.3	Potential Commercial Applications . . . . .	69
6.3.1	Augmented and Virtual Reality (AR/VR) Experiences . . . . .	69
6.3.2	Content Creation and Production . . . . .	70
6.3.3	Collaborative Robotics and Automation . . . . .	70
6.4	General Impact of the Proposed Method . . . . .	71
6.4.1	Academia (People and Knowledge) . . . . .	72
6.4.2	Industry and the Economy . . . . .	72
6.4.3	Space Applications . . . . .	72
6.4.4	Society . . . . .	73
CHAPTER 7	CONCLUSION . . . . .	74
7.1	Summary of Works . . . . .	74
7.2	Limitations . . . . .	74
7.3	Future Research . . . . .	75
REFERENCES	. . . . .	76



## LIST OF TABLES

Table 1.1	General comparison of methods across different criteria . . . . .	4
Table 2.1	Summary of previous SL-based systems for depth detection with EC.	15
Table 3.1	Color detection quality w.r.t. ground truth (GT) . . . . .	25
Table 5.1	Summary of Previous SL-based Systems Addressing Depth Estimation with monocular EC. . . . .	47
Table 5.2	Analysis of Variance (ANOVA) for the Color PSNR . . . . .	61
Table 5.3	Analysis of Variance (ANOVA) for the Color RMSE . . . . .	61
Table 5.4	Analysis of Variance (ANOVA) for the Color FR . . . . .	62
Table 5.5	Analysis of Variance (ANOVA) for the Depth RMSE . . . . .	62
Table 5.6	Analysis of Variance (ANOVA) for the Depth FR . . . . .	62

## LIST OF FIGURES

Figure 1.1	The yearly distributions of the number of published papers related to the 3D reconstruction on indoor environments based on the Google Scholar. . . . .	2
Figure 3.1	Color detection of a stable (top row) and spinning (bottom row) colorful paper pinwheel. Left column: monochrome events without structured light. Right column: colorful image reconstructed aided by structured light with two patterns and equivalent speeds of 30 fps (top) and 150 fps (bottom). . . . .	17
Figure 3.2	Left: The experimental setup with a DLP LightCrafter 4500 Evaluation Module and a Prophesee evaluation kit (Gen3-VGA). Middle: Printed color wheel with the logo of the MIST Lab., captured by a frame-based high-resolution camera. Right: Colorful image reconstructed by proposed method captured by monochrome EC aided by SL. . . . .	19
Figure 3.3	Color detection of a printed color wheel. Top left captured by frame-based high-resolution camera, top right is colorful image reconstructed by proposed method, captured by a VGA monochrome EC aided by SL. Bottom from left to right are collected event-frames for each color light (red, green and blue) by monochrome EC. . . . .	22
Figure 3.4	Color detection of a spinning colorful paper pinwheel reconstructed at different frame rates. Top row (static) left: captured by frame-based high resolution camera, right: colorful image reconstructed by proposed method captured by monochrome EC. Bottom row (spinning pinwheel) reconstructed at, from left to right, 30, 100, 120 and 150 fps. . . . .	22
Figure 3.5	Color detection of printed Macbeth color chart. Image captured by a frame-based high-resolution camera (left), the colorful image reconstructed by the proposed method captured by monochrome EC, without (middle) and with (right) white balance. . . . .	26
Figure 3.6	Different types of SL patterns have been used to control the event rate by Adaptive Structured Light. . . . .	27
Figure 3.7	Colorful board scanned by dot patterns with varying CP. The temporal window sizes are 2.5, 4.3, and 7.14 ms from the top. . . . .	29
Figure 3.8	Colorful board scanned by line patterns with varying CP. The temporal window sizes are 6.67ms for the top and 7.14ms otherwise. . . . .	30

Figure 3.9	Comparing patterns with different CPs in speed and quality of color detection. . . . .	31
Figure 4.1	Comparison between frames captured by the DAVIS346 color event-based camera during RoboCup 2023 on the MSL field. The left column shows RGB frames recorded at 30fps, while the right column displays gathered events with a temporal resolution of less than 2ms. In this record, the ball approaches the robot and bounces in front of it. . . .	34
Figure 4.2	Comparison of event-based camera output with frame-based camera in different motion modes. Note: The EC shows no activity in static scenes and experiences less blur than frame-based cameras at high speeds.	35
Figure 4.3	Setup for creating the dataset during RoboCup 2023. Top left: Robot number 5 from the French team, Robot Club Toulon, with our standalone setup on the MSL soccer field. Top right: Calibration circle dot-board used for calibrating the event-based camera and camera-projector setup, with white dots from LEDs and red dots projected from a projector. Bottom left: Color-DAVIS346 EC. Bottom right: Khadas VIM3 . . . . .	38
Figure 4.4	The ROS GUI to control the bias parameters of the EC and calibrating the camera/projector. . . . .	39
Figure 4.5	Various frames of the dataset captured with the Color-DAVIS346. . .	40
Figure 5.1	Color (left) and depth (right) detection of a volleyball ball being thrown up in front of a VGA monocular event-based camera, reconstructed using the proposed method at an equivalent speed of 120 fps from a distance of approximately 1.5 m. . . . .	43
Figure 5.2	Color detection of a printed color wheel in [1]. Top: The proposed procedure involves projecting various light patterns in different wavelengths/colors. Bottom, from left to right: high-resolution Ground Truth (GT), reconstructed images captured by a VGA monochrome EC in the channels of Red, Green, Blue, and the fully reconstructed image. . . . .	46
Figure 5.3	Proposed ASL pattern types for balancing speed and detail in ERGBD scanning. M is greater than N, which means reconstructing more points and have higher CP. We did not use a pseudo-random dot pattern to detect depth, although it is commonly used in similar approaches [2]. However, with our method, it is possible to reconstruct color in addition to detecting depth. . . . .	50

Figure 5.4	Pattern sequence and their exposure times in microseconds. In our experiments, we used mode 3 with two different values for $n$ (23 and 45), and mode 4 with $n=23$ . One ID for each pattern type would be enough, but we could have different IDs for each color or depth pattern mode. While this increases the total scanning time, it makes the system more robust and trackable. . . . .	50
Figure 5.5	<b>Top Left:</b> Output of the D455 RGB camera. <b>Top Right:</b> The RGB image reconstructed by the ERGBD with a Monochrome EC. <b>Middle Row:</b> Depth detection comparison between D455 (Left), ESL (Middle), and ours ERGBD (Right). <b>Bottom Row:</b> Zoomed-in view of the middle row. Note that the color differences are due to defining different minimum and maximum values for the jet-coded colorization; the actual difference in mm is detailed in Section 5.5.3. . . . .	51
Figure 5.6	Temporal map reconstructed by ESL (Left), ours ERGBD (Right). . . . .	54
Figure 5.7	The pattern of one dot with the size of $3 \times 3$ pixels in our dot-based patterns. If we project (A) due to the diamond pixel configuration of the DMD, we will see (B) on the object surface. Which could lead to mislocating the center of the projected dot. Therefore, we should move red-colored pixels to blue ones in (C) and project (D) to achieve the pattern (A) on the object's surface with a 45-degree rotation. . . . .	56
Figure 5.8	<b>Left:</b> The experimental setup includes a DLP LightCrafter 4500 Evaluation Module, a Prophesee evaluation kit (Gen3-VGA), and an Intel RealSense D455 (only used for comparison). <b>Right:</b> The calibration circle dot-board used for calibrating the EC and camera-projector setup, with white dots from LEDs and red dots projected by the projector. . . . .	57
Figure 5.9	Comparison of color and depth detection for Duck setup (M4L23 pattern). . . . .	58
Figure 5.10	Comparison of color detection for all setups at different speeds. . . . .	60
Figure 5.11	Comparison of depth detection for all setups at different speeds. . . . .	61
Figure 5.12	Quantile-Quantile plots of the color detection dataset. . . . .	62
Figure 5.13	Quantile-Quantile plots of the depth detection dataset. . . . .	63

Figure 5.14	Comparison of color and depth detection for static scenes between RealSense D455, ESL, and ERGBD (ours). Noting that the available code for ESL needs to create a temporal map as a numpy array from a recorded raw file, we did not include the time required for these processes in our calculations. We only considered the time needed for calculating depth from those files. The operating system had an NVIDIA(R) GeForce RTX(TM) 2060 6GB GPU and an Intel Core i7-9750H CPU with 16GB of memory. . . . .	64
Figure 5.15	Comparison of color and depth detection for dynamic scenes (setup with ball number one) between RealSense D455 (top row, right and left) and our system (middle, using pattern M3L23). . . . .	65
Figure 5.16	Series of images showing color and depth detection for dynamic scenes, scanned in 7.4 to 17.23 ms using the M3L23 pattern. The images in the first column from the left were captured by the RealSense D455 for comparison, which took 33 ms. . . . .	66
Figure 6.1	Combining projection with event-based cameras to obtain synchronized spatial and temporal perception. . . . .	70

## LIST OF SYMBOLS AND ACRONYMS

2D	Two-dimensional
3D	Three-dimensional
ANOVA	Analysis of Variance
AR	Augmented Reality
ASL	Active Structured Light
ATIS	Asynchronous Time-based Image Sensor
CCD	Charge-Coupled Device
CFA	Color Filter Array
CFM	Color Filter Mosaic
CMOS	Complementary Metal-Oxide-Semiconductor
CMY	Cyan, Magenta, Yellow
CMYK	Cyan, Magenta, Yellow, Black
CNN	Convolutional Neural Network
CP	Coverage Percentage
CPU	Central Processing Unit
CVF	Computer Vision Foundation
CW	Continuous Wave
DAVIS	Dynamic and Active-pixel Vision Sensor
dB	decibel
DLP	Digital Light Projector
DMD	Digital Micromirror Device
DVS	Dynamic Vision Sensor
EC	Event-based Camera
ERGBD	Event-based Red, Green, Blue and Depth
ESL	Event-based Structured Light
FOV	Field Of View
FPP	Fringe Projection Profilometry
fps	Frame per second
FR	Fill Rate
FTDD	Frequency Tagged Dynamic Dots
GPS	Global Positioning System
GPU	Graphics Processing Unit
GT	Ground Truth

GUI	Graphical User Interface
HC	Histogram Correlation
HDR	High Dynamic Range
HFR	High Frame Rate
HSV	Hue Saturation Value
ID	Identification
IEEE	Institute of Electrical and Electronics Engineers
Laser	Light Amplification by Stimulated Emission of Radiation
LED	Light-emitting diode
LiDAR	Light Detection and Ranging
LUT	Look-Up Table
MC3D	Motion Contrast Three-dimensional
MEMS	Micro-Electro-Mechanical Systems
MIST	Making Innovative Space Technology
MSL	Middle Size League
NIR	Near InfraRed
NIR	Near-Infrared
PhD	Doctor of Philosophy
PSNR	Peak Signal-to-Noise Ratio
RGB	Red, Green, Blue
RGBD	Red, Green, Blue and Depth
RGBG	Red, Green, Blue Green
RGBW	Red, Green, Blue, White
RMSE	Root Mean Square Error
ROI	Region of Interest
ROS	Robot Operating System
RYB	Red, Yellow, Blue
SAR	Spatial Augmented Reality
SBC	Single Board Computer
SfM	Structure from Motion
SGE	Structured light system based on Gray code with an Event camera
SL	Structured Light
SLAM	Simultaneous Localization And Mapping
SM	Stereo Matching
SSL	Small Size League
TI	Texas Instruments

TMM	Temporal Matrices Mapping
ToF	Time of Flight
USB	Universal Serial Bus
VGA	Video Graphics Array
VR	Virtual Reality
WACV	Winter Conference on Applications of Computer Vision



## CHAPTER 1 INTRODUCTION

### 1.1 Context and Motivation

One of the main challenges for mobile robots is localizing and understanding their surroundings. Various vision-based localization methods existed with respect to the scene configuration, measurement device, and types of captured elements [3,4]. Localization can be described as identifying the pose (position and orientation) of the robot relative to the initial/world coordinate system, if there is any predefined coordinate system. It also involves detecting the location of other objects or features relative to the initial coordinate, or determining the robot's pose in reference to an object and vice versa. A 3D map represents the locations of objects within an area.

In an unknown environment without a predefined map, various methods have been developed to create a 3D map, including visual Simultaneous Localization and Mapping (SLAM) and Structure from Motion (SfM) [5]. SLAM is the process of detecting, recording, and recovering the geometrical structures of a scene in an active or passive manner [6]. To create a 3D map of an environment or expand a pre-loaded one, the mobile robot first needs to perform 3D measurements of its surroundings. In other words, the initial step in creating a 3D map is conducting 3D measurements [6].

3D measurement is vital for many visual applications where positioning information of all objects in a 3D scene is crucial for localizing and navigating the robot among them. Figure 1.1 shows the yearly distributions of the number of publications related to the "3D reconstruction of indoor environments", clearly indicating the growing importance of the 3D reconstruction topic.

**Cameras** are commonly used to observe the environment and calculate the relative locations of objects or features within their field of view. However, passive cameras typically capture frames in a 2D plane. These 2D outputs can be used to determine the relative locations of known objects by applying camera transformation matrices. However, for unknown objects, additional information is required to accurately determine their relative locations. Since the size of the object is unknown, it is not possible to find the depth using only the camera projection matrix [7].

**LiDAR** is an active device used for 3D measurement and generating a point cloud of the environment. It operates by projecting laser light beams onto the scene and detecting the reflections. Distances are calculated using differences in beam return times and wavelengths.

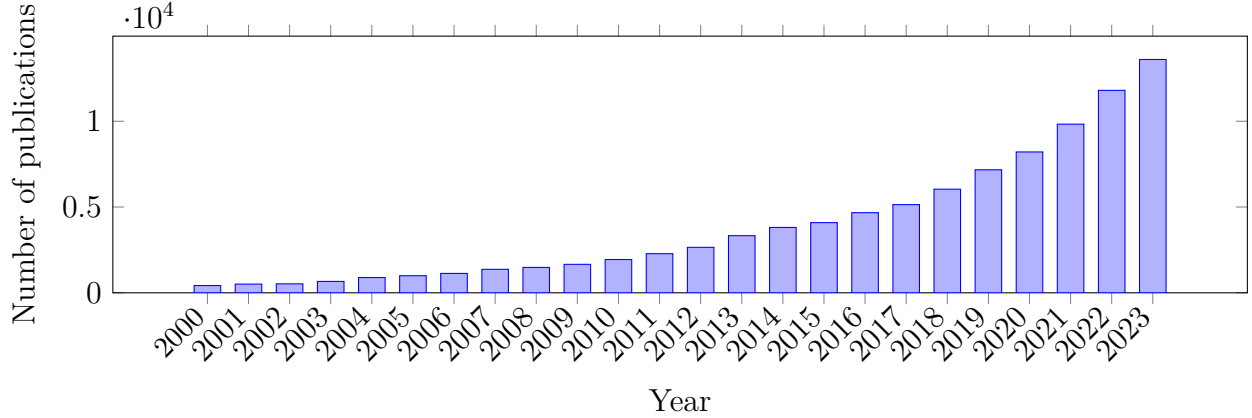


Figure 1.1 The yearly distributions of the number of published papers related to the 3D reconstruction on indoor environments based on the Google Scholar.

LiDARs are valued for their high accuracy and resolution, although they are generally more expensive than cameras and do not capture color information. Additionally, their output tends to be more sparse compared to active cameras.

**RGB-D cameras** are active devices that provide depth information for each pixel alongside color. Unlike LiDARs, which measure distances point-to-point, RGB-D cameras measure distances for multiple points simultaneously by emitting modulated light onto the scene and measuring the wavelengths and travel time of the reflected light beams [8]. However, they are considered low-speed devices with response times typically in the order of tens of milliseconds.

**Stereo vision**, akin to human vision, utilizes two cameras simultaneously for 3D perception. The use of two cameras not only increases energy consumption but also necessitates a more powerful processor to analyze two frames concurrently for stereo matching and finding disparities, thereby further contributing to higher energy consumption. However, utilizing standard cameras in a stereo setup is hindered by their low speed and limited dynamic range (around 60 dB), which can cause saturation in challenging lighting conditions, thereby negatively affecting their functionality. Moreover, achieving fast 3D reconstruction of the environment can be challenging in textureless regions, where accurate stereo matching is necessary to correctly identify corresponding points between the two camera planes [9]. Although there are methods for monocular 3D vision [10, 11], they typically acquire additional information by accumulating frames or estimating depth through learning.

**Structured Light (SL)** scanning systems have been widely used in various applications to acquire more data from a scene using a single camera [12–14]. In this method, a projector emits a known pattern onto one or more objects, and a camera captures these patterns to

measure the relative deformation of the structured light on the target object. This information is then used to construct a 3D model of the objects. However, similar to normal cameras, SL systems still face challenges related to speed and dynamic range.

**Event-based cameras (ECs)**, known by various names such as Dynamic Vision Sensors (DVS) [15], Motion Contrast Sensors [16], Asynchronous Time-based Image Sensors (ATIS) [17], or Asynchronous Transient Vision Sensors [18], are bio-inspired or neuromorphic sensors renowned for their high dynamic range [19, 20].

ECs capture asynchronous measurements of brightness changes in each individual pixel. As mentioned earlier, a single 2D camera does not capture sufficient information from the environment to detect depth in a single shot. Like conventional cameras, there are methods to obtain depth information using only one EC [21–24]. However, these methods are typically learning-based or require movement to gather enough information to create a dense 3D point cloud. Table 1.1 is a general comparison of methods across different criteria.

## 1.2 Problem statement

This research project aims to provide a solution to relieve multiple problems and answer various questions in vision-based 3D perception, including but not limited to the following:

- **RQ1:** How to achieve 3D reconstruction of an environment faster than Time-of-Flight cameras, with adjustable power consumption?
- **RQ2:** How to enable 3D reconstruction for (a) ECs in static situations, and (b) SL-aided methods in dynamic situations?
- **RQ3:** How to obtain a 3D reconstruction of an environment *faster*, *lighter*<sup>1</sup>, and *more accurate* than stereo cameras? Stereo cameras commonly face issues such as blurring in fast movement, saturation in high illumination, blindness in darkness, and inaccuracies in untextured regions.
- **RQ4:** How to achieve a 3D reconstruction of an environment *faster* for detecting dynamic objects and *controllable* to switch between dense and sparse modes, balancing the bandwidth of the system, compared to Structured Light-based methods? Low sensor bandwidth and sensitivity to *specular materials* limit the performance of SL methods.

---

<sup>1</sup>Lighter in terms of processing/computing and subsequently power consumption

Table 1.1 General comparison of methods across different criteria

Method	Color	Range (Short-Normal-Long)	Shadow	Spatial (Low-Normal-High)	Temporal (L-N-H)	Easy Calibration	Sparse or Dense	Power Consumption (L-N-H)	Dynamic Range (L-N-H)	Computational (L-N-H)	Texture need	Dynamic (L-N-H)
Stereo Standard	✓	L <sup>1</sup>	✗	H <sup>2</sup>	L	✓	D	L	L	H	✓	L
ToF	✗ <sup>3</sup>	N	✗	N	L	✗	S	H	N	H	✗	L
LiDAR	✗	L	✗	H	N	✗	S	H	H	H	✗	N
Structured Light	✓ <sup>4</sup>	S	✓	H	L	✓	D	H	L	H	✗	L
EC Monocular	✗	N <sup>5</sup>	✗	N	H	✓	S	L	H	L	✗ <sup>6</sup>	H
Stereo EC	✗	N <sup>5</sup>	✗	N	H	✗	S	L	H	H	✗ <sup>6</sup>	H
EC - SL	✗	S	✓	H	H	✓	D	H	H	H	✗	H
EC - SL ( <b>Ours</b> )	✓	S	✓	H	H	✓	S to D	N to H	H	L	✗	H <sup>7</sup>

<sup>1</sup> Shorter range than LiDAR

<sup>2</sup> Lower resolution than LiDAR

<sup>3</sup> Not available in Monocular

<sup>4</sup> Must capture one frame without SL

<sup>5</sup> Higher range than ToF

<sup>6</sup> Texture could aid in detecting more events

<sup>7</sup> Could detect more high speed movements than the other methods

### 1.3 Research Objectives

The aim of this project is to introduce a new adaptive method for generating a 3D reconstruction of challenging environments suitable for mobile robots. The created 3D colorful point cloud could be utilized for localization and mapping purposes. Color information provides additional data that can enhance tasks like segmentation and recognition. This extra data can also be valuable for loop closure detection in SLAM, improving the system's ability to recognize previously visited locations and enhance mapping accuracy. This new vision-based approach is crucial for a mobile robot when fast and efficient localization and mapping are required in challenging situations. Additionally, the project aims to achieve 3D reconstruction of scenes for mobile robots with adjustable speed and energy consumption. The main

objective is explained through various aspects as follows:

### **1.3.1 Spatial resolution: (RQ1)**

Visual-based 3D scanning systems always face a trade-off between point cloud resolution and scanning speed. Time-of-Flight (ToF) cameras excel in quickly providing depth information but have limitations in fast RGB detection [25]. This research aims to obtain 3D reconstructions with adjustable resolution, allowing us to capture high-resolution colored 3D point clouds of environments (in the order of millimeters, akin to 3D laser/ToF scanners but with color data), as well as high-speed 3D scanning (in the order of milliseconds using more sparse patterns).

### **1.3.2 Mobility limitation: (RQ2)**

A 3D scanner sensor mounted on a mobile robot needs to operate effectively at varying velocities. Event-based cameras (ECs) remain inactive in static situations, while structured light (SL) scanners are sensitive to motion. A method that can adjust the SL pattern and camera bias dynamically, enabling efficient 3D scanning in both static and dynamic environments.

### **1.3.3 Texture dependency: (RQ3)**

Stereo vision systems typically depend on surface textures during the feature/stereo matching process. Feature points are utilized to establish correlations between two camera planes and subsequently compute the depth of these points in the scene. Generally, the depth accuracy of a stereo camera system diminishes on untextured surfaces [3]. One primary objective of this research is to achieve 3D capture regardless of surface texture.

### **1.3.4 Acquisition speed: (RQ3 and RQ4)**

Vision-based scanning systems often face a trade-off between acquisition speed, resolution, and luminous efficacy [16]. The scanning speed is generally related to the capture method, sensor bandwidth, and initial data analysis speed. This project aims to surpass current state-of-the-art methods in speed, allowing for the capture of more points to generate a colored 3D point cloud. It also aims to introduce a method that can adjust the speed to gather more data when high-speed scanning is not required.

### 1.3.5 Light-Efficiency: (RQ3 and RQ4)

Several methods, such as Gray coding or phase-shifting [26], have been developed to enhance the bandwidth and speed of SL-based scanning. However, these techniques are constrained by the power output of the light source. Moreover, standard cameras, which have low dynamic range, face challenges when scanning environments with highly specular materials. While there are approaches to mitigate these challenges using cameras, they inherently operate at a slower pace [27–29]. This research aims to introduce a method capable of handling changes in lighting and operating in dark environments using a high dynamic range sensor. The system is aimed to operate in real-time and includes color detection.

### 1.3.6 Power consumption: (RQ1, RQ3 and RQ4)

The power consumption of onboard sensors is crucial for mobile robots. Lower power consumption allows them to operate missions for longer periods. This capability enables exploration projects to cover wider areas and reach greater distances. Therefore, higher sensor power consumption imposes constraints on exploration missions. This project aims to introduce a specific active 3D reconstruction method that can efficiently manage light source power during the scanning process and minimize energy consumption.

## Experiments and validation

To evaluate the proposed method under static and dynamic conditions, we considered various setups. Different pattern combinations are designed to detect depth and color at various speeds. The quality of the output is compared to the ground truth and baseline. Comparison with recent works highlights differences in scanning methodologies, emphasizing the novel approach of combining color reconstruction with depth estimation

## 1.4 Novelty and Impact

The novelty of this research lies in introducing, for the first time, a method to integrate a monochrome event-based camera (EC) with a projector to achieve super-fast event-based color and depth detection. Moreover, the method is not limited to a single pattern; it allows for pattern adjustment based on the need. This provides the opportunity to control the trade-off between speed and detailed data while maintaining bandwidth.

## 1.5 Research contributions

To the best of the authors' knowledge, this thesis presents significant novel contributions to color reconstruction systems and fast, colorful 3D perception for mobile robots, aligning with our research objectives. The primary contributions of this research, can be summarized as follows: introducing a novel 3D reconstruction technique for rapid area scanning in challenging environments; enhancing scanning performance with adaptive structured light patterns, particularly in motion; demonstrating the reliability of event-based 3D scanning under low-light conditions compared to traditional camera-projector systems; achieving high-resolution 3D scanning; and developing ultra-fast, real-time color and depth measurements per pixel adaptable to diverse scene conditions.

The contributions, categorized and presented per respective article, are as follows:

1. Marjani Bajestani, Seyed Ehsan, and Giovanni Beltrame. Event-based RGB sensing with structured light. In IEEE/CVF Winter Conference on Applications of Computer Vision. 2023 (peer-reviewed)

This work contributes an approach to detect full RGB events using a monochrome EC with a structured light projector. The projector emits rapidly changing RGB patterns, and the EC captures the reflections, allowing for color detection of both static and moving objects. This technique, utilizing a TI LightCrafter 4500 projector and a monochrome Prophesee Gen. 3 EC, enables frameless event-based RGB sensing applications.

2. Marjani-Bajestani, Seyed-Ehsan, and Giovanni Beltrame. Event-based Vision for Robot Soccer. In RoboCup International Symposium 2024. (peer-reviewed)

We created a dataset using event-based cameras from iniVation and Prophesee, recording events in the lab and during Middle Size League matches at RoboCup 2023. Additionally, we developed a ROS-compatible Graphical User Interface (GUI) to simplify camera and camera-projector setup calibration. This GUI allows online control of camera bias parameters and publishes streams of events and event "frames" on ROS topics. These advancements will help RoboCup teams transition to event-based technologies, improving ball detection regardless of color or lighting conditions.

3. Marjani-Bajestani, Seyed-Ehsan, and Giovanni Beltrame. E-RGB-D: Real-Time Event-Based Perception with Structured Light. Submitted to IEEE

Transactions on Pattern Analysis and Machine Intelligence 2024. (peer-reviewed)

This approach introduces a dynamic projection pattern to detect both color and depth for each pixel with ECs. By dynamically adjusting the projection, we can coordinate the devices when needed, managing the overall bandwidth of the system. These adjustments optimize data acquisition, ensuring accurate and colorful point clouds without sacrificing spatial resolution. Specifically, we achieved a color detection speed of 1400 fps and 4 kHz for pixel depth detection, significantly advancing event-based 3D reconstruction methods.

## 1.6 Thesis structure

The rest of this thesis is structured as follows: Chapter 2 reviews various methods in the field of visual 3D perception with a focus on SL scanning and ECs. Chapter 3 is the "Event-Based RGB Sensing With Structured Light", accepted at the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), which describes the proposed method for detecting color with a monochrome EC. Chapter 4 is the "Event-based Vision for Robot Soccer", accepted at the RoboCup International Symposium 2024, which describes the application of ECs in Robot Soccer to detect fast-moving objects like the ball and includes datasets and tools. Chapter 5 is the "E-RGB-D: Real-Time Event-Based Perception with Structured Light", a submitted paper to the IEEE Transactions on Pattern Analysis and Machine Intelligence, which details the 3D reconstruction method and evaluates its outcomes. Chapter 6 contains a general discussion including features and advantages of the proposed method with technical challenges and their solutions. Finally, Chapter 7 outlines concluding remarks and future work.



## CHAPTER 2 LITERATURE REVIEW

### 2.1 Monochrome to Color

Color information plays a crucial role in tasks such as segmentation and recognition [30]. Colorization refers to the creation of a color image from a monochrome sensor or grayscale image while preserving resolution. This method relies on external color data obtained from an external device [31], user input [32], or a trained neural network that integrates scene-specific color information [33, 34]. However, this process can be resource-intensive in terms of time and cost.

Initially, event cameras (ECs) were predominantly monochrome, with color ECs emerging more recently [35–37]. However, color ECs typically feature lower resolution compared to monochrome ECs, attributed to constraints such as sensor size limitations and the incorporation of color filters [38–40].

In order to preserve resolution while increasing bandwidth threefold, Marcireau et al. [41] employed dichroic filters on three ECs. This approach allowed them to merge the outputs from these ECs, facilitating the acquisition of color information across three separate event streams.

### 2.2 Monocular Depth Sensing

One way to detect the depth with monocular camera, is to use active devices. LiDAR scanners and Time-of-Flight (ToF) cameras are both active depth measurement devices. LiDAR scanners project modulated light beams onto the scene and measure the time it takes for the reflected beam to return. This method calculates distance based on differences in light beam return times and wavelengths. LiDAR systems typically use mechanical components to perform 3D scanning tasks. In contrast, ToF cameras perform time-of-flight computations using integrated circuits embedded in standard CMOS (Complementary Metal Oxide Semiconductor) or CCD (Charge-Coupled Device) technologies [42].

In outdoor settings and environments with strong ambient lighting, LiDARs offer superior performance and accuracy compared to other depth measurement devices because they are unaffected by ambient light. However, LiDAR systems requires more computational resources and consume more power than cameras [43]. Additionally, the high cost of 3D LiDAR scanners limits their use in many mobile robot projects.

The point cloud obtained from LiDAR is often very sparse, depending on the number of scan lines, resulting in many missed pixels compared to denser measurement devices. To address this issue, *depth completion* is essential. Generally, neural networks are used for depth completion in depth map processing research, but it remains an open problem [44,45].

Another drawback of LiDAR depth maps arises when the point cloud is projected onto a 2D plane, which can lead to challenges in maintaining the accuracy of the data. This mapping process introduces irregularities in the point cloud, requiring specialized techniques to handle these inconsistencies effectively. Additionally, the irregular spacing of points in the cloud can further complicate data processing and analysis. These limitations, along with the inherent noise and sparsity of LiDAR data, are considered 'intrinsic artifacts' of the acquisition process, reflecting the fundamental limitations of the technology itself. [44].

Pulsed light and Time-of-Flight (ToF) cameras differ from LiDAR in that they measure the distance of many points simultaneously rather than point by point [8]. Accurate time measurement is challenging, and the methods can be direct (measuring time from pulsed light or phase from CW<sup>1</sup> operation) or indirect (using derivative methods). Commercial devices typically calculate distance by scaling the phase with Amplitude Modulated CW [8].

These devices use near-infrared (NIR) light for emitting light but often suffer from poor accuracy and depth resolution. Additionally, they require two calibration procedures: one for the projection matrix (like standard cameras) and one for distance measurement. The low resolution makes feature detection challenging. Another issue is the inaccuracy of depth images, especially at object edges, caused by capturing multiple light reflections due to object concavity [8].

Combining ToF technology with standard RGB cameras can produce colorful point clouds. However, using a standard RGB camera introduces limitations in terms of detection speed and light efficiency, especially for fast RGB detection. [25]. Moreover, although these systems do not use two cameras for triangulation depth measurement, they still require two cameras, disqualifying them as monocular devices.

Another way to detect depth with a monocular device is by using an event-based camera (EC). Like other 2D cameras, ECs do not provide enough information about the scene in a single-shot<sup>2</sup> to calculate depth. However, the events generated by camera movement can be valuable for gathering more information about objects in the scene. The polarity of events is particularly useful in optical flow methods to determine the camera's trajectory [46–

---

<sup>1</sup>Continuous Wave

<sup>2</sup>ECs are not shot-based, so *single-shot* is not accurate. This work views them as data captured at once or as events occurring in a short period.

49]. In other words, various viewpoints are used to discover the depth, noting that clearly this condition is not a single-shot capturing. As in this method moving the camera is a requirement [23, 50–55]. Mostly in these approaches the scene should be static. In [21], authors have used the events created due to the movement of a hand-held EC in a static scene, to estimate the camera pose followed by performing a 3D reconstruction. Methods using global illumination changes, such as turning on a light in a dark room [56, 57], or applying a rotating polarizer [58], are also employed to reconstruct iso-contours or estimate surface normals.

By considering the events in a spatio-temporal neighborhood, *event frames* will be obtained. Event frames will change an unfamiliar event stream to a familiar 2D image with information about scene edges which is compatible with conventional computer vision. By providing an adaptive frame rate signal, event frames enable the camera pose estimation practicable via image alignment [22]. In [22], the authors used a monocular EC to obtain a semi-dense and a pose tracking that will be used in SLAM projects. They applied space sweeping for 3D reconstruction same as the proposed method in [23], however this was achieved by using edge-map alignment for camera motion tracking.

Another way to predict the depth from monocular EC is employing a learning method. In [24] the authors proposed a recurrent architecture that leverages the temporal consistency of the event stream in generating network prediction. However, the predicted depth is provided by a learning method; therefore the training dataset has a critical role in predicting the depth of the new seen objects. It could give false result if the speed of the object is almost identical to the speed of the EC, which will result in the camera to receive very few events related to the object.

### 2.3 Depth Sensing via Triangulation Method

To improve scene comprehension, researchers are exploring fusion methods that incorporate additional sensors. Interference-based methods excel in highly accurate micro-scale measurements, while Time-of-Flight methods are preferred for their broader coverage in large-scale scenes, albeit with lower precision. Triangulation-based approaches, such as stereo vision and Structured Light (SL), strike a balance between these extremes [8], offering reliable depth information particularly suited for shorter distances.

### 2.3.1 Event-based depth sensing with multi camera

Although stereo ECs are significantly faster than frame-based cameras, the stereo matching process is more complex. In this case, the matching process becomes temporal, meaning that matching pixels should be based not only on event appearances across the image planes but also on the similarity of the event timestamps, which are in the range of microseconds [59]. Also, in terms of brightness alterations; in some occasions the two cameras do not receive the same brightness changes in one specific position [60]. In some projects, time surfaces<sup>2</sup> are used to obtain a proxy intensity picture [43,61].

The calibration process for event-based stereo camera is also different from the frame-based stereo camera. In [62], they have implemented a calibration mechanism for calibrating the system. An LED grid with 64 LED lights in two different depth was used. However, for stereo extrinsics calibration, Zhu et al. [63] utilized the Kalibr toolbox.

In [64], they have used two DVS static cameras to achieve the stereo reconstruction of the moving object in front of the camera. A cooperative network fed by captured series of events is employed to incorporate as spatiotemporal context, which calculates the disparity for each incoming event. The network is continually developed in time by capturing more events. This approach matches events by examining their neighborhoods and improved by calculating the initial weights and using window-based matching instead of time-based single event matching [65].

In [43], multi event-based cameras (synchronized) are used to obtain a 3D reconstruction of the scene. By examining the time surfaces used in proxy intensity images for stereo matching in a temporal way (temporal correlations of the events).

In [60], they propose using additional hardware, specifically a mirror-galvanometer-driven laser, to create light spots in the scene. Similarly, in [66], they projected a binary speckle pattern to capture features more easily with a stereo EC setup. In [67], they projected a line laser to generate a known pattern instead of a speckle pattern. All of these contrast changes are captured by a stereo EC, enabling resource-efficient matching and eliminating the need for sensor synchronization.

### 2.3.2 Event-based depth sensing with SL

ECs detect SL events by sensing changes in frequency or contrast. Adjusting the SL frequency enables the camera to distinguish individual points. There are two main ways SL and ECs are combined: the first involves using an SL device to generate events that the camera captures,

---

<sup>2</sup>Employing events in a period of time

similar to the approach proposed in [60]. Another approach involves using an SL device (or other 3D depth sensors) to first generate a depth reconstruction of the scene, which is then combined with the captured events. For example, in [68], the authors employed an external SL projector (Kinect) to acquire depth information about the scene. Simultaneously, an EC captures events (brightness changes in each pixel due to motion), while the SL device records its own data. These outputs are then combined to create a 3D event representation. This method necessitates movement to gather the required data, despite relying on an active external light source. Operating the Kinect independently alongside this process increases power consumption and does not reduce processing time.

Structured Light (SL) functions similarly to active stereo-vision, employing triangulation for depth measurement. Instead of using two cameras, SL projects a pattern onto the surface and captures its deformation with a single camera. While a typical SL setup involves one camera and one projector, multiple cameras can also be utilized to achieve higher resolution or faster measurements [13, 14].

Using an external projector for Structured Light (SL) to overcome stereo challenges with a monocular EC is fundamental in modern systems incorporating event cameras, as seen in works such as [15, 16, 69, 70]. Various SL methods differ based on the types of patterns they utilize. Table 2.1, shows a summary of previous SL-based systems for depth detection with EC.

**Coded Patterns in SL:** Striped SL patterns are utilized for area scanning. Unique patterns enable the system to identify points and compute depth. When these patterns are binary (black and white), multiple patterns are necessary to match corresponding points. Consequently, binary patterns are sensitive to object movement compared to other patterns. Achieving a high frequency of pattern switching aids in reducing this sensitivity. DLP projectors can switch patterns at kilohertz frequencies [71, 72]. Higher resolution is achieved by introducing 2D and hybrid patterns [73].

Leroux et al. [74] propose a 3D reconstruction approach utilizing an Asynchronous Time-based Image Sensor (ATIS) with  $304 \times 240$  pixels and a DLP projector. Instead of traditional line structured light (SL), they employ Frequency-Tagged Dots (FTD) projected onto the scene. By analyzing the known pattern and distances, they calculate pattern deformation and point depth. However, this method is susceptible to movement and changes in ambient light, which can impact the number of events captured.

Mangalore et al. [20] and Li et al. [75] employed a Dynamic Active-pixel Vision Sensor (DAVIS346,  $346 \times 260$  pixels) paired with a DLP LightCrafter 4500 for 3D reconstruction. They developed a Fringe Projection Profilometry (FPP) system using a moving fringe pat-

tern, allowing the EC to scan multiple lines simultaneously, which is faster than traditional line-scanning methods. An advantage of ECs is their ability to detect shadowed areas, unlike frame-based cameras where shadows and dark regions may appear indistinguishable. However, pre-recording the scene without objects is necessary to "inpaint" shadowed areas, and the camera's limited event reporting capacity may result in some events being missed.

**Simple patterns SL:** The basic statistical pattern method involves projecting a random array of dots. This approach is common in commercial devices like Microsoft Kinect V1, Intel RealSense [76], and Orbbec Astra [77] due to its simplicity and compact size. Huang et al. [2] also used a DLP6500 projector to project a single pseudo-random pattern frequently. They extracted event frames from this pattern and employed digital image correlation to compute displacements and create 3D surfaces of objects. While this dot-based approach speeds up scanning, it sacrifices detail compared to the denser information achievable with line patterns. Furthermore, using discrete patterns led to inaccuracies in dot placement and sensitivity to ambient light, resulting in lower-resolution 3D depth measurements [8].

Another approach to achieve higher resolution is by using lines instead of discrete dots. This method allows for highly accurate measurements in one direction. It is often combined with a line laser for short-range scanning. However, to measure depth accurately in all directions, the orientation of the line needs to vary.

Brandli et al. [15] utilized a laser line and an event camera (EC) for surface scanning of objects. The concentrated laser line improved contrast, making it easier for the EC to detect the line. When the environment is stable, structured light (SL) emission aids in detecting relevant pixels only. However, achieving a complete 3D reconstruction using this approach requires varying the line direction or moving the object in relation to the laser line.

Matsuda et al. [16] addressed the speed-resolution trade-off in 3D scanning by employing a line laser and an event camera (EC), known as Motion Contrast 3D Scanning (MC3D). Unlike traditional SL scanners, which can be affected by changes in illumination, their approach using a high dynamic range EC resulted in enhanced final outcomes. While a laser scanner typically required 28.5 seconds of exposure time, their proposed device achieved similar results with just one second of exposure time.

Building upon Matsuda et al.'s research, Muglikar et al. [69] introduced Event-based Structured Light (ESL), where they utilized time maps to synchronize the projector and camera temporally. Initially, they generated depth maps through an epipolar disparity search across rectified projector time maps. They then implemented a subsequent processing step to improve pixel-level consistency and reduce event fluctuations. However, this additional processing requires substantial computational resources, limiting their method's ability to

achieve real-time performance.

Morgenstern et al. [70], in their X-maps approach, introduced a method that converts the projector time map into a rectified X-map. This X-map captures X-axis correspondences for incoming events, enabling direct disparity lookup without the need for additional search processes. Their method supports real-time interactivity, making it suitable for applications like Spatial Augmented Reality (SAR) that demand low latency and high frame rates. They reported that their approach is significantly faster—7 to 100 times faster—compared to the ESL method, which involves row-by-row disparity search and depth calculation for the entire frame. We employed a similar X-mapping technique to compute depth for each pixel, as detailed in the following section.

Table 2.1 Summary of previous SL-based systems for depth detection with EC.

Method	Type	Pattern	EC	Sensor	Projector
Brandli et al. [15]	Monocular	Simple	DVS128	$128 \times 128$	Laser line 500 Hz
MC3D [16]	Monocular	Simple	DVS128	$128 \times 128$	Laser point 60 fps
FTP [74]	Monocular	Coded	ATIS0	$304 \times 240$	DLP TI LightCrafter 3000
FPP [20, 75]	Monocular	Coded	DAVIS346	$346 \times 260$	DLP TI LightCrafter 4500
Martel et al. [60]	Stereo	Both	DAVIS240	$240 \times 180$	Laser beam
ESL [69]	Monocular	Simple	Prophesee Gen3	$640 \times 480$	Laser point 60 fps
X-maps [70]	Monocular	Simple	Prophesee Gen3	$640 \times 480$	Laser point 60 fps
SGE [72]	Monocular	Simple	Prophesee Gen4	$1280 \times 720$	Laser point 60 fps (static scene)
		Coded			DLP projector OPR305185 (dynamic scene)

## CHAPTER 3    ARTICLE 1: EVENT-BASED RGB SENSING WITH STRUCTURED LIGHT

**Preface:** Event-based cameras (ECs) asynchronously report pixel brightness changes, making them beneficial in challenging lighting conditions and for detecting fast movements due to their high dynamic range, pixel bandwidth, temporal resolution, and low power consumption. The first generation of ECs were monochrome, but now include color versions, which have lower resolution due to the use of color filters.

In this chapter, we propose a method to detect full RGB events using a monochrome EC with a structured light projector. The projector emits rapidly changing RGB patterns, and the EC captures the reflections, allowing for color detection of both static and moving objects. This technique, utilizing a TI LightCrafter 4500 projector and a monochrome EC, enables frameless event-based RGB sensing applications.

**Full Citation:** Marjani Bajestani, Seyed Ehsan, and Giovanni Beltrame. "Event-based RGB sensing with structured light." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023. (published October 11, 2022)

**DOI:** <https://doi.org/10.1109/WACV56688.2023.00542>

**Abstract:** Event-based cameras (ECs) are bio-inspired sensors that asynchronously report pixel brightness changes. Due to their high dynamic range, pixel bandwidth, temporal resolution, low power consumption, and computational simplicity, they are beneficial for vision-based projects in challenging lighting conditions and they can detect fast movements with their microsecond response time. The first generation of ECs are monochrome, but color data is very useful and sometimes essential for certain vision-based applications. The latest technology enables manufacturers to build color ECs, trading off the size of the sensor and substantially reducing the resolution compared to monochrome models, despite having the same bandwidth. In addition, ECs only detect changes in light and do not show static or slowly moving objects. We introduce a method to detect full RGB events using a monochrome EC aided by a structured light projector. The projector emits rapidly changing RGB patterns of light beams on the scene, the reflection of which is captured by the EC. We combine the benefits of ECs and projection-based techniques and allow depth and color detection of static or moving objects with a commercial TI LightCrafter 4500 projector and a monocular



monochrome EC, paving the way for frameless RGB-D sensing applications. Our code is available publicly: [github.com/MISTLab/event\\_based\\_rgb\\_d\\_ros](https://github.com/MISTLab/event_based_rgb_d_ros)

### 3.1 Introduction

Event-based cameras (ECs) report pixel brightness changes asynchronously, a behavior inspired by the human eye [78]. When the brightness changes over a certain threshold for a pixel, the camera generates an event containing the coordinates of the pixel  $(x,y)$ , a timestamp, and the polarity of the event (i.e. increasing or decreasing). Although ECs do not capture full images, they can detect movement thousands of times faster than standard frame-based sensors, and since they do not have an external shutter cycle, their output is event-driven and frameless, resulting in very low latency, power, and bandwidth demands.

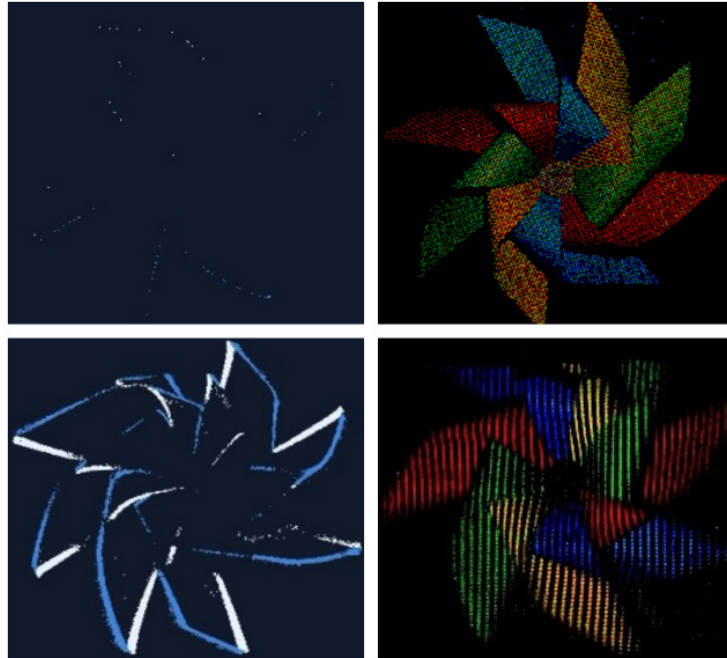


Figure 3.1 Color detection of a stable (top row) and spinning (bottom row) colorful paper pinwheel. Left column: monochrome events without structured light. Right column: colorful image reconstructed aided by structured light with two patterns and equivalent speeds of 30 fps (top) and 150 fps (bottom).

ECs have been used in various computer vision applications such as fast movement detection and tracking [49, 79, 80], optical flow, pose tracking and visual-inertial odometry [81, 82], Simultaneous Localization And Mapping (SLAM) [22, 83], pattern recognition [84], depth estimation and stereo vision [19, 23, 61], and many more.

In computer vision, color information has an important role [30] and could be essential to many tasks such as segmentation and recognition [41]. The first generation of ECs are monochromatic, with color ECs only recently becoming available [35–37, 85]. However, due to limitations in terms of sensor size, color ECs have lower resolution than mono ECs because they need to use color filters.

It is worth noting that ECs report pixel brightness changes, meaning that an EC will not report anything when the camera (and/or the object in its field of view) is static or slowly moving (Fig. 3.1, top left), which can be critical in some cases (e.g. for a slow-moving robot). To overcome this issue, one could use an external active device such as a laser, a flashing LED, or a light projector to generate events in static and almost static situations. This external active lighting system could also be used to detect depth by projecting detectable patterns called Structured Light (SL) [15, 16, 69, 86].

We present a method to add color and depth to a monocular, monochrome event-based camera while maintaining fast response time and resolution. We use a Digital Light Processing (DLP) projector that emits patterns of lights that we call Active Structured Light (ASL) on a scene, the reflection of which is captured by the EC which in turn generates events tagged with the color and depth of the scene. It is worth noting that our ASL method could also be used with color ECs, allowing the detection of static scenes. By dynamically adjusting the projection, we have color data when needed, managing the overall bandwidth of the system. For example, we can use the full resolution of the camera to detect static color scenes, or use more sparse patterns for fast moving objects. Projecting patterns also allows triangulation-based measurements to create a colorful 3D point cloud of the scene. Overall, our method generates colorful events from a monochrome EC:

1. with no loss of spatial resolution;
2. with the ability to detect static objects and scenes;
3. optimizing the bandwidth of the EC by detecting the color when and where it is needed;
4. using patterns that allow event-based depth measurement, ultimately generating colorful point clouds.

In this work, we focused on visual light wavelength (emitted by the LED projector) and materials that are not in the category of fluorescence and they do not change the wavelength of the light. We validated our approach in different dynamic conditions: Fig. 3.2 shows the experimental setup with a DLP projector<sup>1</sup> and a Prophesee evaluation kit<sup>2</sup>. With this setup, we achieved full color detection at an equivalent rate of 1400 frames per second (fps) (note

---

<sup>1</sup>LightCrafter 4500 Evaluation Module

<sup>2</sup>Gen3-VGA

that the camera is frameless, we use fps just for the purpose of comparison). Fig. 3.2 also shows the color detection of a static printed color wheel.

The rest of the paper is as follows: Section 3.2 presents related work; Section 3.3 describes our method for color detection; Section 3.4 details the results of our method in several conditions; and finally, Section 3.5 draws some concluding remarks and outlines possible future work.

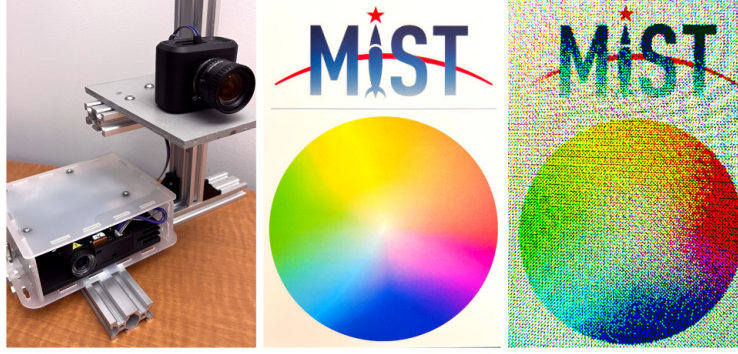


Figure 3.2 Left: The experimental setup with a DLP LightCrafter 4500 Evaluation Module and a Prophesee evaluation kit (Gen3-VGA). Middle: Printed color wheel with the logo of the MIST Lab., captured by a frame-based high-resolution camera. Right: Colorful image reconstructed by proposed method captured by monochrome EC aided by SL.

### 3.2 Related Work

Digital color cameras use various Color Filter Array (CFA) or Color Filter Mosaic (CFM) on their sensors to detect different colors for each pixel, and among them, the Bayer array filter [38] is the most common CFA [40]. The size of a CFA is between 4 to 36 pixels (or sometimes larger [39]), which means that we need several monochrome pixels to generate each color pixel, effectively decreasing the resolution (e.g., 4x with a  $2 \times 2$  CFA).

Colorization is the process to generate a color image based on a monochrome sensor or grayscale image without loss of resolution. Colorization requires either external data about the image colors, user interaction, or a trained neural network embedding the knowledge of the colors on the scene, and can be a time-consuming and expensive task [32]. Levin et al. [32] introduced a method that needs a few initial inputs from a user to generate a full color image and keeps tracking the color on upcoming frames in a video. Zhang et al. [33] introduce a fully automatic colorization approach based on a convolutional neural network (CNN) that can change a grayscale image into a near-real colorful image. Their method successfully deceived 32% of human participants in distinguishing the generated and ground-truth images. In contrast to these colorization approaches, our method does not need initial input data to get

color out of a monochrome camera and it can provide realistic color information faster than CNN models.

Another approach to generate color data without quality loss on a monochrome image is to use separate cameras: the monochrome sensor takes a more detailed and higher contrast image, while a lower resolution RGB camera adds color information. This combination is common, but the image fusion, colorization, or the color transfer process are still a challenge [31].

Event-based cameras have introduced a new field of imaging systems. Due to their advantages compared to standard cameras, many scientists investigated ways to generate and reconstruct images from events to use in frame-based computer vision algorithms. A monochrome EC has been used in many image reconstruction works [87–92]. Also, combining a standard frame-based camera and an EC can produce a deblurred high frame rate (HFR) and high dynamic range (HDR) video [93].

By combining three ECs using dichroic filters Marcireau et al. [41] introduced a prototype to capture a stream of events in RGB separate channels for color segmentation. This method maintains the monochrome resolution but increases the bandwidth 3x.

With the introduction of color event-based cameras [37], some research focused on the reconstruction of images and videos based on color events [94–96]. Scheerlinck *et al.* [97] presented a dataset for color ECs. They also compared the output quality of some image reconstruction methods such as [87–89] in color.

As digital color cameras, current color ECs also use CFA to generate color events, which reduces their output resolution leading to lower bandwidth when compared with Marcireau et al. [41]. Our method reconstructs color data when needed, keeping the bandwidth of the system in check.

### 3.3 Monochrome to color

Compared to frame-based cameras, ECs are faster sensors, however, since they report nothing in a static situation or with slowly moving objects, they require an additional sensor to provide visual perception in these situations. We use an external event generator, namely a DLP projector. By emitting a pattern of light on objects in the scene, not only we can detect their color, but we are also able to detect depth, which makes event-based RGB-D sensing possible. Moreover, since ECs have high dynamic range, a high-power light projector is not necessary in dark environments.

There are many standard color formats for digital color descriptions (additive or subtractive), such as CMY (cyan, magenta, yellow), or with black CMYK, RYB (red, yellow, blue), RGB

(red, green, blue) or with white RGBW and etc. Selecting the color space could depend on the application and the color range of the desired objects in the environment. Without loss of generality, we select the RGB color space which is more common in vision applications.

We use the EC to measure the amount of reflection of the emitted light on an object. To measure the color, we project three different wavelengths (structured light in red, green, and blue) on the environment and measure the amount of reflected light captured by the EC. During each pattern exposure time, the received events are gathered in an appropriate color channel on the initial frame.

To synchronize the DLP projector with the EC, we connect the trigger pins of the camera to the projector. By changing the pattern color, the DLP sends a pulse to the camera which identifies the incoming events as belonging to the appropriate color channel. Fig. 3.3 depicts the output of the color detection of a printed RGB color wheel separated in each color channel. The bottom frame of the Fig. 3.3 shows that the printed color wheel does not have pure green  $(0, 255, 0)$  and blue  $(0, 0, 255)$  colors in 24 bit RGB format. For example, in the red light channel (bottom left), the green circle also reflected some light (although less than the red circle) and as a result, it appears gray.

### 3.3.1 Color detection speed limits

One of the main advantages of the ECs is their response time which is in the range of microseconds. However, with the introduced method, we need to gather events of each color separately, limiting the speed of color detection to the maximum speed of pattern switching of the DLP projector. With the LightCrafter 4500 Evaluation Module, we are able to detect color with an equivalent frame rate up to 1400 fps due to its high frequency (4225 Hz<sup>1</sup>). However, assuming that the color of the object is not changing, we could still use the other methods to track the object only based on the high speed stream of events [49, 80] and use the color detection method for a short period of time. Fig. 3.4 shows the output of the color detection of a spinning colorful paper pinwheel reconstructed at different frame rates.

### 3.3.2 Advantages over monochrome cameras

Monochrome or grayscale cameras have been used in vision-based applications that do not need color information. As mentioned in Section 3.2, the combination of a monochrome camera and a color camera could be challenging for image fusion, colorization, or the color transfer process [31]. A dual-camera consisting of a frame-based camera and an EC can

---

<sup>1</sup>Switching rate for preloaded 1 Bit depth pattern of the LightCrafter 4500 Evaluation Module

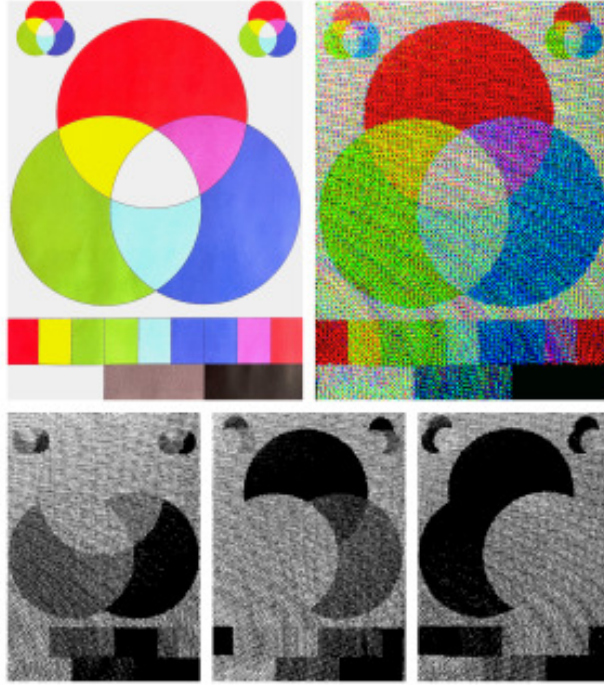


Figure 3.3 Color detection of a printed color wheel. Top left captured by frame-based high-resolution camera, top right is colorful image reconstructed by proposed method, captured by a VGA monochrome EC aided by SL. Bottom from left to right are collected event-frames for each color light (red, green and blue) by monochrome EC.

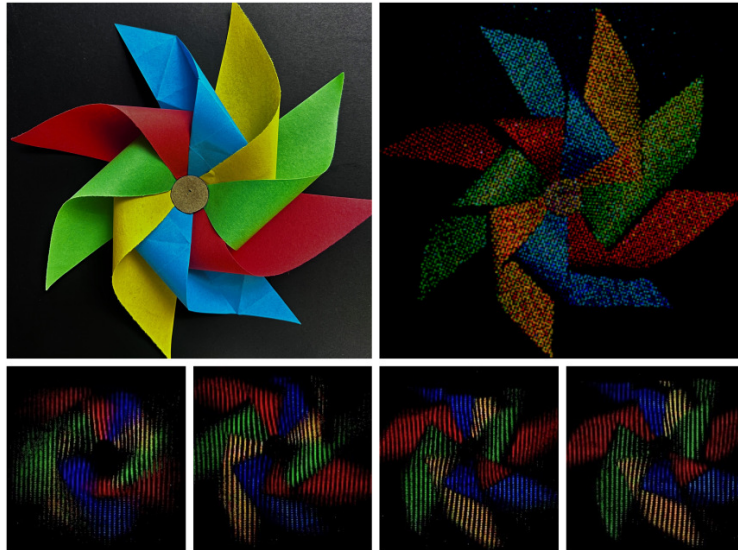


Figure 3.4 Color detection of a spinning colorful paper pinwheel reconstructed at different frame rates. Top row (static) left: captured by frame-based high resolution camera, right: colorful image reconstructed by proposed method captured by monochrome EC. Bottom row (spinning pinwheel) reconstructed at, from left to right, 30, 100, 120 and 150 fps.

produce a deblurred high frame rate (HFR) and high dynamic range (HDR) video [93]. However, adding a second camera increases the required bandwidth. Our method allows us to benefit from ECs’ features and detect/update the color information for a given period of time. Moreover, the camera-projector combination enables depth sensing and simplifies feature detection and matching (w.r.t. stereo cameras) [15, 16, 69, 86].

### 3.3.3 Advantages over color event-based cameras

As mentioned in Section 3.2, digital cameras often use CFA to detect color. For instance, the Color-DAVIS346 [85] is one of the most recent color EC that uses an RGBG Bayer pattern with an output resolution of  $346 \times 260$  pixels. This kind of camera is reporting the stream of events in 3 or 4 different channels, which increase the need for bandwidth. Higher bandwidth requirements can cause bus saturation (as described in Section 3.4). In addition, despite increasing the bandwidth needs and decreasing the resolution, color ECs cannot detect the environment when the camera or object is static or moving very slowly. Our method is useful to efficiently use the bandwidth by detecting the color only when and where it is needed. Moreover, our method also gathers information from the environment from an initially static robot or camera, meaning there is no need to have mechanical parts to move the camera and receive events, which makes the system more reliable. Further, since a high-resolution EC could be subject to more noise in a dark environment compared with a low-resolution EC [98], our method could still get the benefits of the high-resolution monochrome EC in a dark environment.

### 3.3.4 White balance and color correction

White balance and color correction can make the captured image close to its natural color. White balance can be adjusted before or after capturing the image. Generating white light with the DLP projector can change the image white balance, because the color temperature of a light source or the warmth/coolness of the white light can change the white balance directly. The DLP projector has three different LED colors: red, green, and blue. Generating LED-based white light could be challenging with wideband wavelength RGB LEDs [99–101]. Since the DLP projector has narrowband LEDs, the white balance can be adjusted by changing the current of each LED separately.

**Lighting model:** If we consider the DLP projector as a point-sized light source, we can model the lighting with the Lambertian shading model which is one of the simplest bidirectional reflectance distribution functions (BRDF) and an appropriate approximation to many real-world material surfaces [102]. In the Lambertian shading model, R, G, B values of the

resulting pixel are independent of the angle that the viewing ray hits the surface:

$$W = S_{ref} S_{pow} \max(0, n \cdot b), \quad (3.1)$$

where  $W$  is the combination of  $(R, G, B)$  values for a desired pixel, and  $S_{ref}$  is the spectral reflectance of the material,  $S_{pow}$  represents the spectral power distribution of the projector (as the light source),  $n$  is the outward surface normal (of the object) and  $b$  is the light beam vector which is from the surface intersection point to the projector. The dot product of these two unit vectors gives the amount of attenuation based on the angle between the surface to the projector. The  $\max$  function is used to prevent a condition where  $n \cdot b < 0$ , because the projector would be behind the object in this case. This model could be divided for each color, for example the model for red light is:

$$R = S_{ref_R} S_{pow_R} \max(0, n \cdot b) \quad (3.2)$$

To generate white light in an ideal situation, we consider that each color has the same power distribution and  $S_{pow_R} = S_{pow_G} = S_{pow_B}$ . For a white or gray surface we would have  $S_{ref_R} = S_{ref_G} = S_{ref_B}$ . As a result, by controlling the current of each LED ( $S_{pow}$ ) we can have balanced white light.

We can use a gray card, a color wheel, or a Macbeth color chart/color checker to calibrate our system. We used a printed Macbeth color chart to do the calibration, and Fig. 3.5 shows the output of the color detection with the proposed method with and without white balance calibration.

**Absolute error:** To check the quality of the reconstructed image, we need to have a base image and specify an error calculation method. We consider the image captured by a frame-based, high-resolution iPhone 13 Pro camera as the base image (Ground Truth or GT) in Fig. 3.5. To calculate the absolute error, we compared the histogram of two images in Hue Saturation Value (HSV) format based on the correlation metric<sup>1</sup>:

$$c = d(H_o, H_b) = \frac{\sum_I (H_o(I) - \bar{H}_o)(H_b(I) - \bar{H}_b)}{\sqrt{\sum_I (H_o(I) - \bar{H}_o)^2 \sum_I (H_b(I) - \bar{H}_b)^2}}, \quad (3.3)$$

where

$$\bar{H}_k = \frac{1}{N} \sum_J H_k(J), \quad (3.4)$$

---

<sup>1</sup>The OpenCV histogram comparison correlation method.



and  $N$  is the total number of histogram bins, which in our case is 256 (8 bit in each color channel). The  $H_o$  and  $H_b$  are respectively histogram of the output image and the baseline image with a Histogram Correlation ( $HC$ ) between 0 and 1. The  $HC$  between the base image (left) and each reconstructed image is respectively 0.22 and 0.76 for the reconstructed image without white balance (middle) and with white balance (right) in Fig. 3.5. Moreover, to check the difference between each pixel in the reconstructed image and the GT, and calculate the absolute error, we calculated the root mean square error (RMSE) separately for each channel. As an example, the RMSE for the red channel is:

$$RMSE_r = \sqrt{\frac{\sum_N (p_{or} - p_{br})^2}{N}}, \quad (3.5)$$

where  $p_{or}$  and  $p_{br}$  are respectively the pixel value in the red channel of the output frame and the baseline frame.  $N$  is the total number of pixels, i.e.  $640 \times 480 = 307200$ . Table 3.1 shows the quality of the color detection for each image in Fig. 3.5 compared to the GT image. Table 3.1 also shows that, after manual white balance tuning, all three channels had 12% better RMSE on average. To have a more realistic color detection, an online white balance calibration could be helpful in minimizing the average RMSE if needed. To check the quality of each image we computer the Peak Signal-to-Noise Ratio (PSNR), shown in Table 3.1.

Table 3.1 Color detection quality w.r.t. ground truth (GT)

	GT	No WB	WB
$RMSE_{red}$	0	83.89	<b>83.65</b>
$RMSE_{green}$	0	87.79	<b>71.16</b>
$RMSE_{blue}$	0	92.76	<b>76.97</b>
$RMSE$	0	88.15	<b>77.26</b>
$PSNR$	$+\infty$ dB	9.23 dB	<b>10.37 dB</b>
Histogram Correlation (HC)	1	0.22	<b>0.76</b>

Another way to do the color correction is to capture the image with a white channel by RGBW color spaces and perform the correction on four channels similar to RGBW CFA-equipped sensors [103]. This comes at the cost of adding a 4th color light to the SL, adding at least 33% to the length of the capture time.

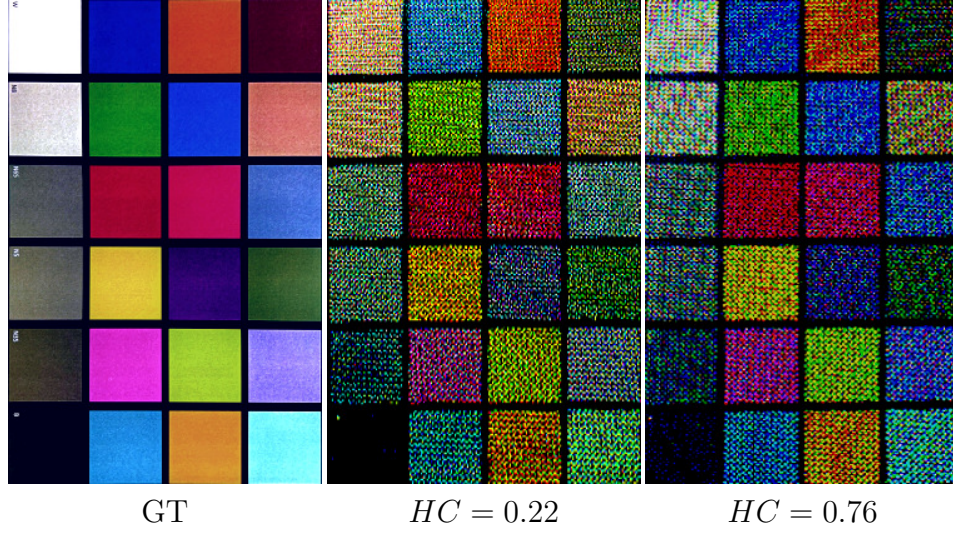


Figure 3.5 Color detection of printed Macbeth color chart. Image captured by a frame-based high-resolution camera (left), the colorful image reconstructed by the proposed method captured by monochrome EC, without (middle) and with (right) white balance.

### 3.4 ASL: Adaptive Structured Light

High-resolution ECs have a higher event rate and need more bandwidth compared to low-resolution ECs, but each EC has a limited data rate (finite bandwidth) on the output interface or bus. If the data rate or the number of events exceeds the limit, bus saturation could happen [78, 98]. Filtering [104] or online event-rate control [105] can mitigate this issue. When using an external event generator such as the DLP projector which emits SL on the scene, controlling the event rate is even more important. One method to control the event rate when using a projector is to define a region of interest (ROI) and project the pattern only where it is needed. Muglikar et al. [86] used one EC camera to detect the ROI (generally the area of the image frame that has more events due to the movement) and then projected the SL on that area followed by detecting the depth with a second EC. Instead of adding a second EC to the system, we introduced ASL to control the event-rate. Fig. 3.6 shows different patterns of the SL which change based on the number of received events. As expected, there is a trade-off between having high-resolution (dense) and high-speed (sparse) color detection. The generated SL patterns are, multiple dots or lines patterns and solid patterns. In static conditions, ASL could also be used with a color EC and white light. However, it should be noted that color ECs need more bandwidth compared to monochrome ECs with the same resolution.

**Bandwidth control:** By frequently projecting SL into the scene, we receive events caused

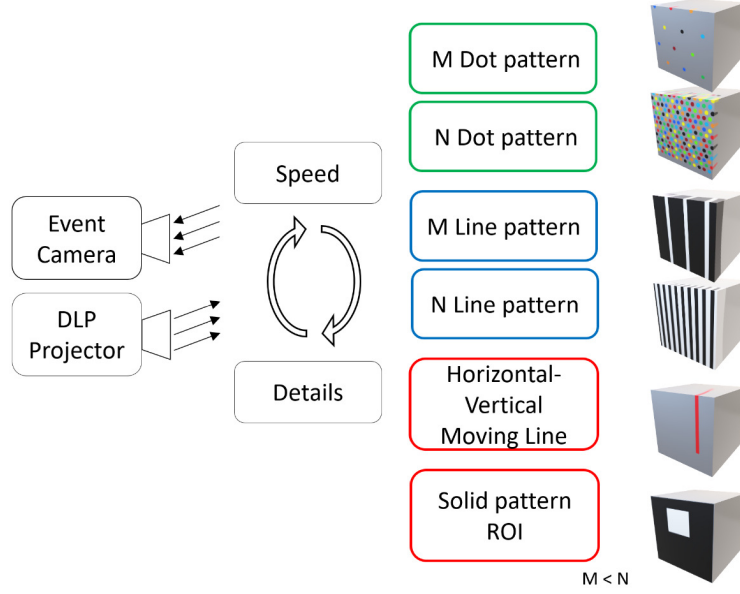


Figure 3.6 Different types of SL patterns have been used to control the event rate by Adaptive Structured Light.

by the SL alongside the events caused by the movement of objects or the camera. We want to control the biases of the camera to prevent bus saturation, but we do not want to lose data by excessively decreasing the camera sensitivity. In general, the number of events must be lower than the bandwidth of the EC:

$$Max.Bandwidth < events_{SL} + events_M, \quad (3.6)$$

where  $events_M$  is the number of events caused by the movement of the EC or any object in the scene (i.e., any other events that have not emerged due to the SL).  $events_{SL}$  is the number of events caused by the SL, and we can control it by changing the pattern and the power of the LED projector.  $events_{SL}$  is not only linked to the color of the object (and its reflectivity/fluorescence percentage, which we do not investigate in this paper), but also it is related to the distance of the camera-projector from the object. Increasing the distance, the spectral power distribution decreases because of the reduction in power density. Unfortunately, there is no information available concerning the variation of power density changes with distance for each LED of the DLP projector. Modelling the DLP projector power density could be useful, but it is out of the scope of this paper. In this work, we make the simplifying assumption that all LEDs have the same power density. As a result, to control  $events_{SL}$ , we need to control  $S_{pow}$  from (3.1).

Considering a one-bit pattern, we can control  $S_{pow}$  by changing the pattern (changing the

number of white pixels in a black and white frame), instead of changing the current of the LEDs. We call the number of white pixels per frame as the coverage percentage (CP), with each pattern type having a different CP. To additionally simplify the problem, we assume that the DLP and the EC are close and we can consider the CP on the DLP frame plane despite the fact that, depending on the relative pose of the camera to the projector, the CP could be different on the camera frame plane. We used a colorful board in our experiments, placed 160cm from the camera-projector, shown in Figs. 3.7 and 3.8.

**Dot pattern** Dot grid and circle patterns are one of the simplest patterns to detect the local depth from SL [106] or even calibrating the camera when it is out of focus [107]. Changing the number of dots (feature points) or their distance affects the depth resolution. However, more feature points lead to additional processing time as well as generating more events, which can lead to bus saturation in ECs. By changing the number of dots dynamically based on the event rate, we are able to control the trade-off between the speed of scanning and the amount of detail. Fig. 3.6’s top two rows show the proposed ASL with dot-grid patterns where  $M$  and  $N$  ( $M < N$ ) are the number of dots on each grid. Fig. 3.7 shows three different dot patterns with different CPs. The top row is generated with a temporal window size of 2.5ms (equivalent to 400fps). Similarly, the second row has a window of 4.34ms or 230fps, and the bottom row for 7.14ms or 140fps. The leftmost column of Fig. 3.7 is a ground truth (GT) frame generated with a one-second temporal window; the middle column is an example frame among the 430 frame samples. We compare each frame pixel by pixel with the GT frame to compute the *RMSE* for each channel, shown in the rightmost column.

**Multiple-lines pattern** Since dot-grid patterns are leading to a sparse image, to generate a dense image, line patterns are preferred in low-speed 3D scanning and multi-shot 3D measurement methods. Sequential projection techniques mostly use strip lines [108]. Since the DLP projector can quickly switch (4225 Hz) between patterns, it is possible to generate a dense graph for some region of the object by projecting lines and measuring the depth with triangulation. Although for the spaces between lines we do not have measurements, increasing the number of lines generates more features and it covers a larger area. Similarly to the dot-grid pattern, increasing the number of lines or dots increases the scanning processing time and event rate, so speed and detail must be traded off. The third and fourth rows from top in Fig. 3.6, show the proposed ASL with the line patterns where  $M$  and  $N$  ( $M < N$ ) are the number of lines in each pattern. As Fig. 3.7, Fig. 3.8 shows line patterns with different CPs.

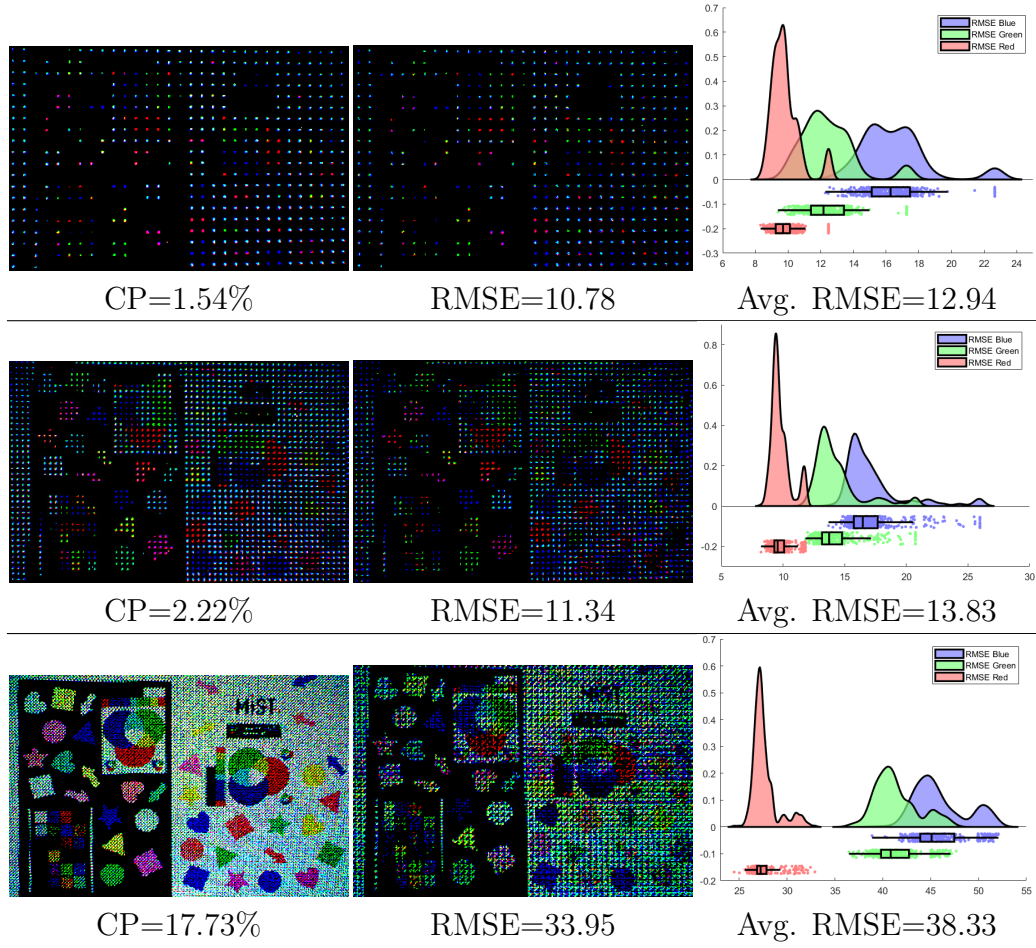


Figure 3.7 Colorful board scanned by dot patterns with varying CP. The temporal window sizes are 2.5, 4.3, and 7.14 ms from the top.

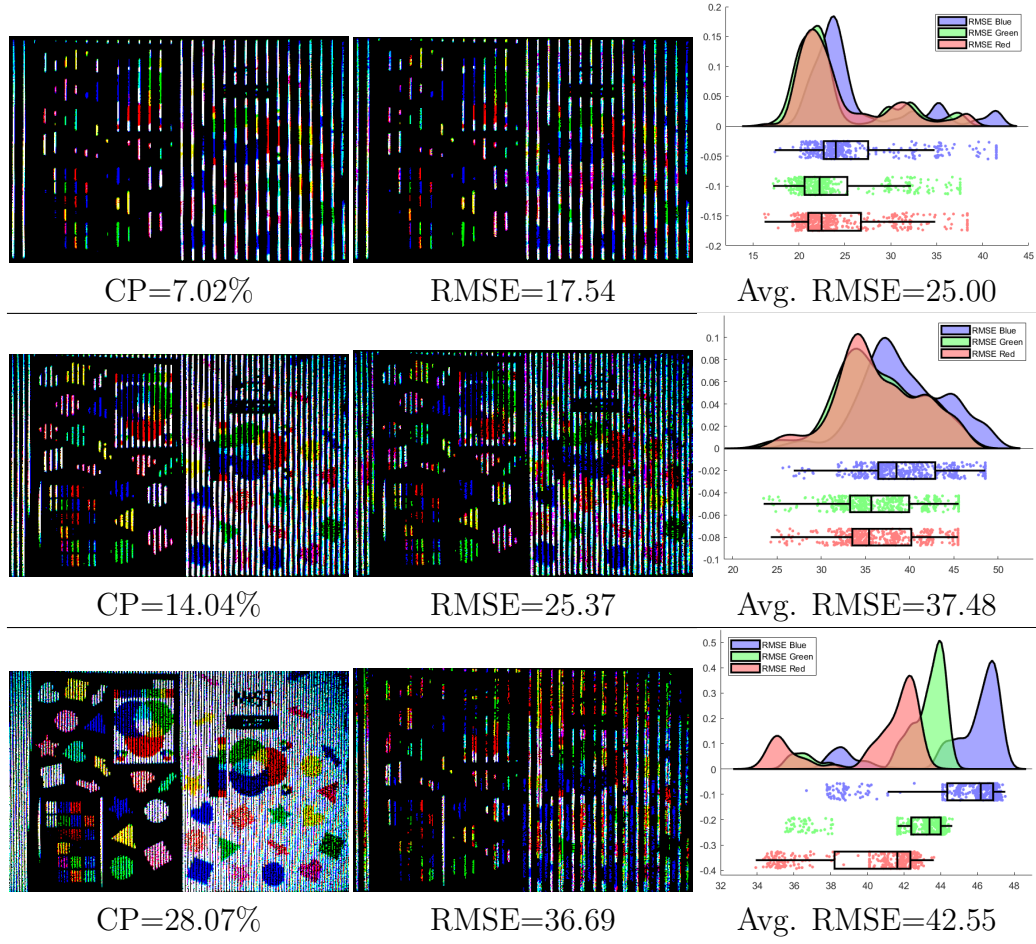


Figure 3.8 Colorful board scanned by line patterns with varying CP. The temporal window sizes are 6.67ms for the top and 7.14ms otherwise.



**Moving-line pattern** To have a full dense scanning in 3D, a line pattern is very common [15, 16, 69]. We propose to use a moving line pattern (horizontal or vertical depending on the offset between the camera and the projector), when the event rate is lower than the bandwidth limits, providing dense scanning.

**Solid pattern** Whenever the 3D scanning is performed, or when we need the color information only for a specific area (the region of interest), we can use the ROI mode. As described in Section 3.4, Muglikar et al. [86] defined an ROI dynamically based on the situation of the scene, then scanned that area with more laser points. The bottom row of Fig. 3.6, shows ASL with the solid pattern for the ROI mode.

To compare different pattern and speed of scanning, we projected patterns with various CPs onto the colorful board. Fig. 3.9 shows the trade-off between details, speed, and the quality of the reconstructed colorful image. It shows that to have a more detailed image, we need to spend more time switching patterns to cover more area. Also, for high speed scanning, a sparse pattern (lower CP with fewer details) is needed. Note that a sparse pattern does not decrease the quality of the color detection even with high speed sampling. Fig. 3.9 has been generated by using  $\sim 24000$  frames.

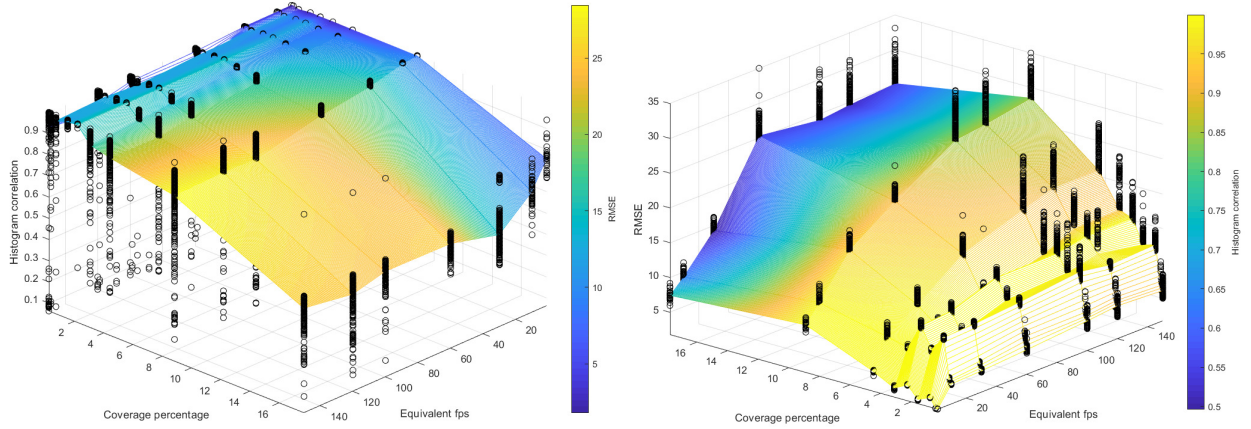


Figure 3.9 Comparing patterns with different CPs in speed and quality of color detection.

### 3.5 Conclusions

We present a method to add color and depth to a monocular, monochrome event-based camera while maintaining fast response time and resolution. Our method reconstructs colorful events and frames using a monochrome EC aided by adaptive structured light (ASL). By

dynamically adjusting the projection, we have color data when needed, managing the overall bandwidth of the system.

We achieved a color detection speed equivalent of 1400 fps with a Texas Instrument’s DLP LightCrafter 4500 projector. Our method could be used in event-based depth measurement and perception projects. Advantages of ECs, could makes the colorful depth detection much faster than RGBD cameras.

Although color detection is related to the lighting conditions and material properties at the intersection point (object surface), the scope of this work was the color detection on common materials that are generally matte and not too shiny (with high reflection) or fluorescence. Some materials can interact with light: they can be absorbing, scattering or emitting light [109]. In this work we focused on visual light wavelength (emitted by the LED projector) and materials that are not in the category of fluorescence and they do not change the wavelength of the light. However, the use of the event-based camera with a different type of light source and materials could be investigated in future works. Also, without considering color detection, static reflective materials can be scanned more effectively with ECs when compared to the other depth measurement devices [16]. To detect the color of these kind of materials, a Blinn-Phong shading model [110] could be considered in future works.



## CHAPTER 4 ARTICLE 2: EVENT-BASED VISION FOR ROBOT SOCCER

**Preface:** Object tracking is a major challenge for soccer-playing robots, especially for goalkeepers due to the fast-moving soccer ball. To improve tracking speed, we propose using Event-based Cameras (ECs), which report pixel brightness changes asynchronously. With their high dynamic range, pixel bandwidth, temporal resolution, low power consumption, and microsecond response time, ECs are ideal for challenging lighting conditions and fast movements.

In this chapter, we created a dataset using event-based cameras from iniVation and Prophesee, recording events in the lab and during Middle Size League matches at RoboCup 2023. Additionally, we developed a ROS-compatible Graphical User Interface (GUI) to simplify camera and camera-projector setup calibration. This GUI allows online control of camera bias parameters and publishes streams of events and event "frames" on ROS topics. These advancements will help RoboCup teams transition to event-based technologies, improving ball detection regardless of color or lighting.

**Full Citation:** Marjani-Bajestani, Seyed-Ehsan, and Giovanni Beltrame. "Event-based Vision for Robot Soccer." Proceedings of the 27th Robot World Cup (RoboCup 2024), Springer Nature Switzerland. (published June 7, 2024)

**Abstract:** Object tracking is one of the main challenges in soccer-playing robots. Due to its fast movement, detecting and tracking the soccer ball is challenging for goalkeepers in both humanoid and wheeled robots. To speed up object tracking, we propose the use of Event-based Cameras (ECs). ECs are bio-inspired sensors that asynchronously report changes in brightness for each pixel. Because of their high dynamic range, pixel bandwidth, temporal resolution, low power consumption, and computational simplicity, they are beneficial for vision-based projects in challenging lighting conditions and can detect fast movements with their microsecond response time. We created a dataset using two different event-based cameras from iniVation and Prophesee that recorded events in the lab and during Middle Size League matches at RoboCup 2023. Additionally, we created a Graphical User Interface (GUI) working with the Robot Operating System (ROS) to simplify camera and camera-projector setup calibration for RoboCup participants. The proposed ROS GUI is able to control the camera bias parameters online and publish a stream of events in addition to event "frames"

on ROS topics. These advancements will help all RoboCup teams shift from frame-based to event-based technologies, enhancing ball detection regardless of color or lighting. The dataset is available publicly: [github.com/MISTLab/event\\_based\\_data](https://github.com/MISTLab/event_based_data).

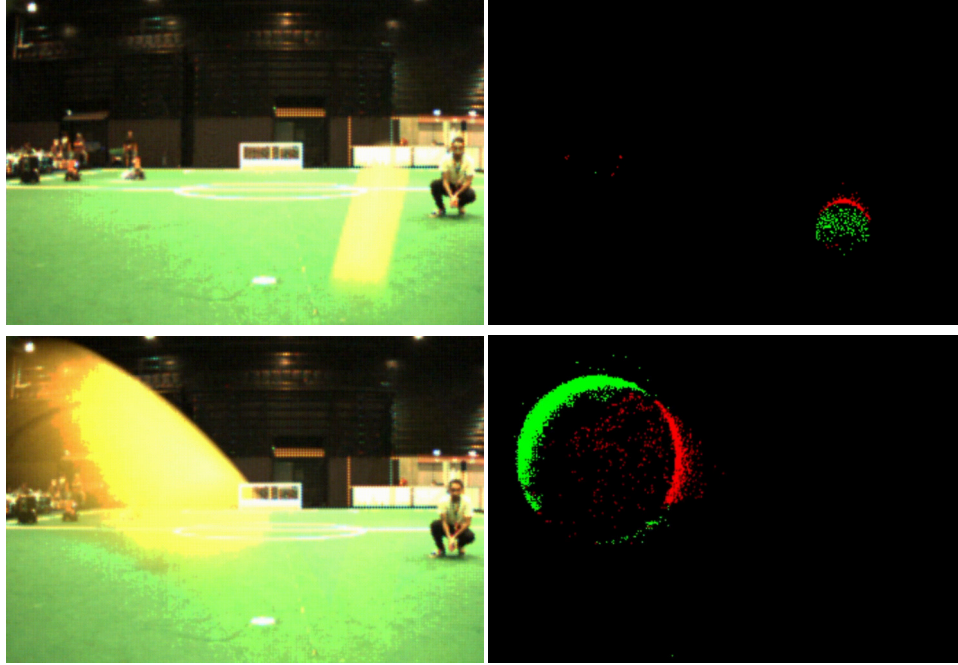


Figure 4.1 Comparison between frames captured by the DAVIS346 color event-based camera during RoboCup 2023 on the MSL field. The left column shows RGB frames recorded at 30fps, while the right column displays gathered events with a temporal resolution of less than 2ms. In this record, the ball approaches the robot and bounces in front of it.

## 4.1 Introduction

Frame-based standard cameras are widely used on soccer-playing robots in the RoboCup. Typically, robots use these cameras to detect and locate the ball and other robots on the soccer field. Additionally, vision-based sensors provide other important information, such as the field lines and their relative distance to localize the robot. However, frame-based cameras have inherent limitations. They require good illumination, they are susceptible to blurring during fast movements (see Fig. 4.1), have a relatively low dynamic range (leading to saturation in changing illumination conditions), and may require high bandwidth depending on the output resolution and frame rate.

In the RoboCup soccer leagues, real robots mainly fall into two types: humanoid and wheeled robots [111]. Even though robots in the Small Size League (SSL) do not have individual

cameras (mounted cameras above the field capture frames, and a central computer reports the locations), detecting and tracking a fast-moving ball individually or cooperatively is challenging for all players in the Middle Size League (MSL) and the Humanoid Leagues (Kid, Teen, and Adult sizes). In recent years, teams have investigated various types of vision-based sensors to achieve the fastest available tracking methods in RoboCup. However, to the authors’ knowledge, no research has been conducted on the advantages of the ECs in the RoboCup. ECs (also known as neuromorphic cameras, Dynamic Vision Sensors (DVS), motion contrast sensors, Asynchronous Time-based Image Sensors (ATIS), and asynchronous transient vision sensors) are bio-inspired sensors that produce a “paradigm shift” in the way visual data is obtained [78]. Fig. 4.2, compares standard cameras with ECs.

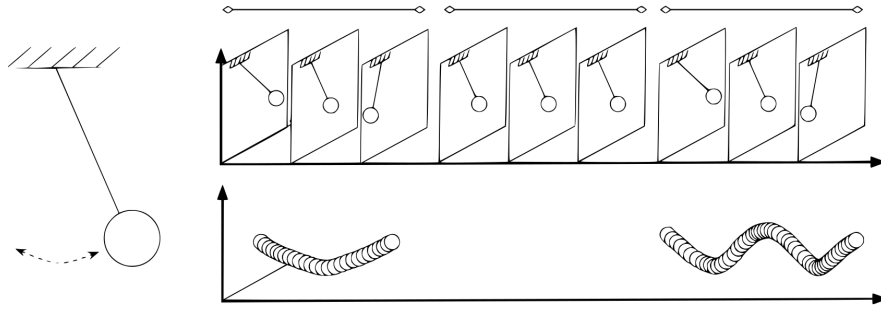


Figure 4.2 Comparison of event-based camera output with frame-based camera in different motion modes. Note: The EC shows no activity in static scenes and experiences less blur than frame-based cameras at high speeds.

While these cameras are commercially available, they tend to be relatively more expensive compared to frame-based cameras. Also, due to their neuromorphic nature, employing conventional convolutional algorithms with event-based cameras is not feasible. Overcoming these obstacles is crucial for advancing the adoption and efficacy of event-based technologies in the RoboCup. The main contribution and benefits of this research for the RoboCup participants are as follows:

1. Introduction of Event-based Cameras (ECs) to RoboCup soccer leagues, expanding the technological scope. Which is beneficial in fast object detection and tracking while maintaining computational simplicity, improving overall performance.
2. Providing a dataset containing a recorded stream of events during a RoboCup match, facilitating experimentation and analysis.
3. Development of a Graphical User Interface (GUI) on the Robot Operating System (ROS) [112], enabling the generation of event frames compatible with conventional algorithms.

4. Streamlining camera calibration processes through the ROS GUI, enhancing accuracy and efficiency in camera setup.

Our method provides teams with an open-source code that can be used to feed their current ball detection and tracking algorithm with a faster frame rate. By generating frames gathered from events on ROS topics, our open-source framework seamlessly integrates with existing computational vision-based algorithms, ensuring teams do not need to modify their primary algorithms. Additionally, the provided dataset allows teams to test their code before investing in ECs, aiding in their transition from frame-based cameras to event-based ones.

## 4.2 Related work

ECs do not capture full images; instead, they report a stream of events. However, they can detect movement thousands of times faster than standard RGB frame-based sensors. They are particularly useful in fast-moving detection projects, such as dodging multiple dynamic obstacles with a quadrotor [113]. In [113], the authors used ECs to estimate ego-motion and the motion of moving objects, achieving a 70% success rate, including the detection of objects with unknown shapes in low-light conditions<sup>1</sup>.

Researchers have integrated frame-based cameras with ECs to avoid receiving blurred frames during high-speed motions. In [93], they combined these two types of cameras to generate a deblurred high frame rate (HFR) and high dynamic range (HDR) video.

In [114] they presented a perception system for 6-DOF localization during high-speed maneuvers. Their method had robust motion tracking with angular speeds up to 1,200°/sec.

In [79], the authors used powerlink *IEEE61158* industrial network, communicating the FPGA with a controller connected to a self-developed two-axis servo-controlled robot. And they compared frame-based cameras and ECs in terms of the response time and robustness to the variable lighting conditions. By utilizing ECs, they achieved 85% data reduction and 99 ms faster position detection on average compared to the frame-based camera.

The authors of [49], used a DVS to detect and track a fast-moving object. They approximated the 3D geometry of the event stream to motion-compensate for the camera and detect unknown moving objects. They reached over 84% of success rate in detecting multiple unknown moving objects in various lighting conditions<sup>1</sup>.

The authors of [115], used an EC on a robotic arm (as a goalkeeper) to block upcoming shots. They achieved an 80% blocking capability even against fast shots, with update rates of 550

---

<sup>1</sup><https://youtu.be/k1uzsiDI4hM>

<sup>1</sup><https://youtu.be/UCAJi0ZFaz8>

Hz and low latencies of  $2.2 \pm 2$  ms. These results were achieved with a peak CPU load of less than 4% and standard USB buses, showcasing practical viability under real-world operating conditions.

### 4.3 Event-based camera calibration

To calibrate the camera, we used our Event-based RGBD ROS Wrapper [1]. The provided ROS GUI is able to control the camera setting online and publish a stream of color-stamped events in addition to RGB frames on ROS topics. The GUI is to calibrate the camera with a Symmetric Circles Grid LED Pattern (Fig. 4.3 top right). Fig. 4.4, shows the provided ROS GUI during the camera-projector calibration.

The ROS GUI distinguishes between events generated by the LEDs on the calibration board and events generated by the projected pattern dots from the projector. As shown in Fig. 4.4, the GUI has been used first to calibrate the camera by detecting various relative positions of the LED board. Then it has been used to track the calibration board and calculate its relative position to the camera, to calibrate the projector and determine the relative position of the projector to the camera. The camera-projector calibration is integrated into the code, making it convenient for structured light 3D scanning projects. While the projector calibration may not be applicable in RoboCup scenarios, the camera calibration section would be useful alongside the other available software [116].

### 4.4 Dataset

To create the dataset, we mounted a Color-DAVIS346 [117] EC on top of an MSL robot<sup>1</sup> and gathered data during the RoboCup 2023 (Fig. 4.3 top left). The camera reports events and frames simultaneously, which can be useful for teams to compare the proposed method with their current camera. It also has a 6-axis (Gyro and Accelerometer) with up to 8 kHz sampling rate. The power consumption of the camera is less than 180 mA at 5 VDC (USB). To record events we used Khadas VIM3 single-board computer (SBC) [118]. The VIM3 features an Amlogic A311D processor.

The dataset is available publicly: [github.com/MISTLab/event\\_based\\_data](https://github.com/MISTLab/event_based_data). The provided dataset contains a diverse range of scenarios, including the ball approaching the goalkeeper, leaving the goal area, penalty kicks, dribbling in front of the goalkeeper, and even scenarios involving human dribbling across the soccer field. Fig. 4.5 shows various frames of the dataset captured with the Color-DAVIS346.

---

<sup>1</sup>Robot Club Toulon

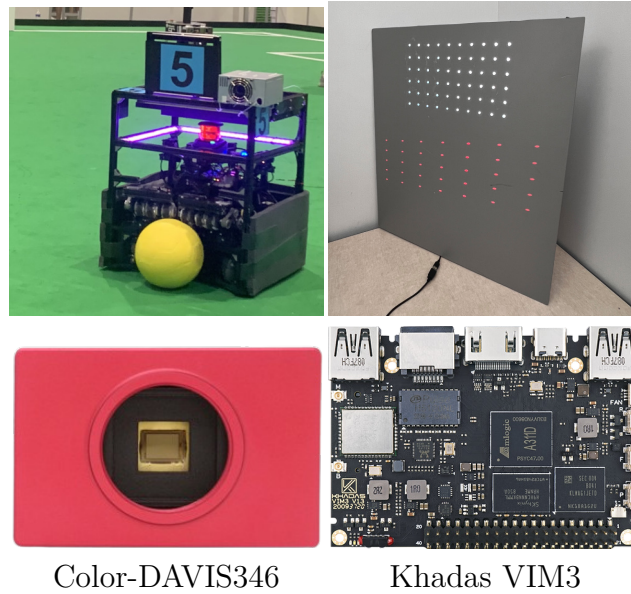


Figure 4.3 Setup for creating the dataset during RoboCup 2023. Top left: Robot number 5 from the French team, Robot Club Toulon, with our standalone setup on the MSL soccer field. Top right: Calibration circle dot-board used for calibrating the event-based camera and camera-projector setup, with white dots from LEDs and red dots projected from a projector. Bottom left: Color-DAVIS346 EC. Bottom right: Khadas VIM3

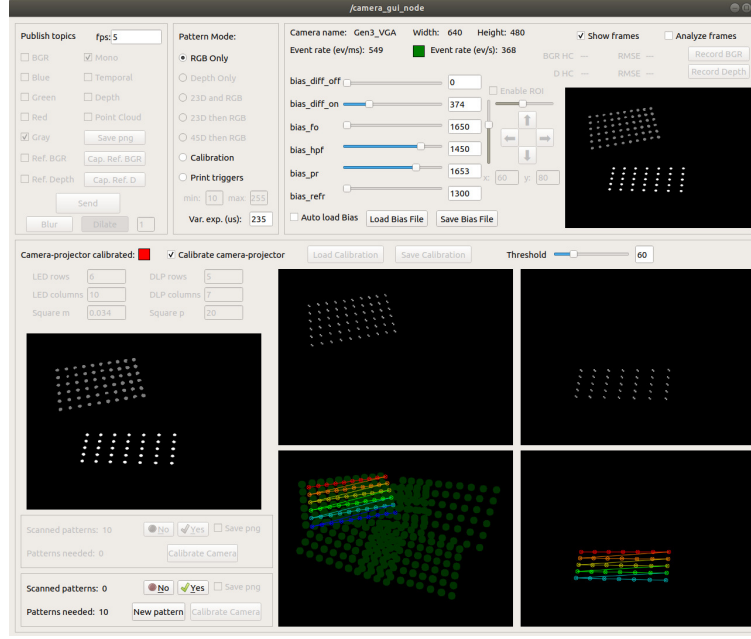


Figure 4.4 The ROS GUI to control the bias parameters of the EC and calibrating the camera/projector.

## 4.5 Conclusion

In conclusion, our work provides tools for calibrating event-based cameras and offers a dataset for RoboCup participants. Teams can utilize their current convolutional image processing algorithm by generating frames from the event-based camera using our ROS wrapper. This approach simplifies object detection and tracking while maintaining compatibility with existing algorithms.

## 4.6 Acknowledgment

We extend our appreciation to the iniVation company for providing a DAVIS346 color event-based camera in 2023, which facilitated the creation of the dataset and our presentation of this camera to members of the RoboCup community participating in the soccer robot league (Middle Size League) in Bordeaux, France.

We also would like to acknowledge the collaboration of the French team Robot Club Toulon for their assistance in affixing the camera onto their robot during intermissions between matches.

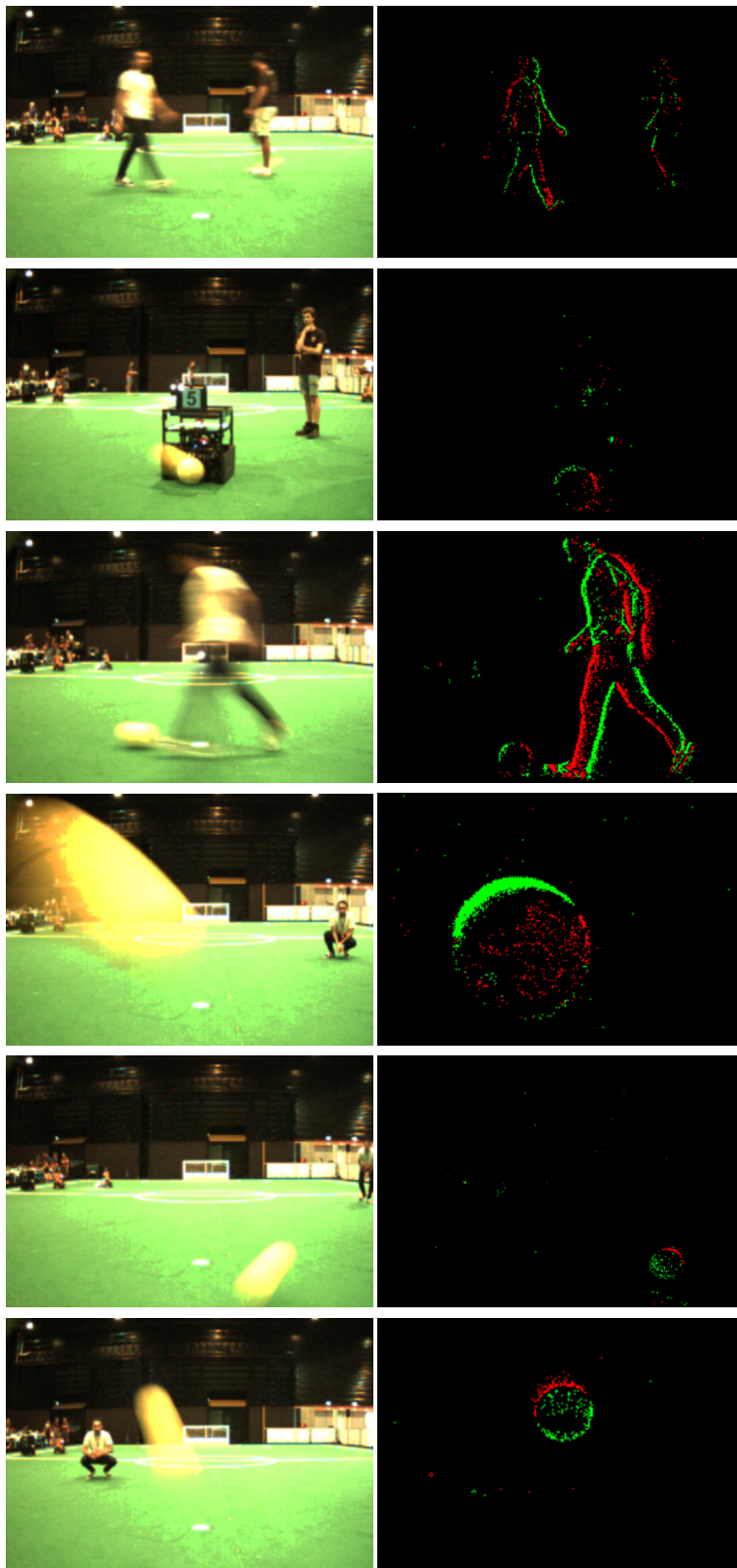


Figure 4.5 Various frames of the dataset captured with the Color-DAVIS346.



## CHAPTER 5    ARTICLE 3: E-RGB-D: REAL-TIME EVENT-BASED PERCEPTION WITH STRUCTURED LIGHT

**Preface:** Event-based cameras (ECs) are high-speed sensors that detect pixel brightness changes asynchronously, ideal for efficient vision sensing. While traditional monochrome ECs lack color detection and struggle with static objects, our approach integrates a Digital Light Processing (DLP) projector to form Active Structured Light (ASL) for RGB-D sensing. This combines EC advantages with projection techniques to separately capture color and depth per pixel. Dynamic projection adjustments optimize data acquisition, ensuring accurate, colorful point clouds without sacrificing spatial resolution. Specifically, we achieved a color detection speed equivalent to 1400 fps and 4 kHz of pixel depth detection, significantly advancing the realm of event-based 3D reconstruction methods.

**Full Citation:** Marjani-Bajestani, Seyed-Ehsan, and Giovanni Beltrame. "E-RGB-D: Real-Time Event-Based Perception with Structured Light", IEEE Transactions on Pattern Analysis and Machine Intelligence 2024 (submission July 14, 2024).

**Abstract:** Event-based cameras (ECs) have emerged as bio-inspired sensors that report pixel brightness changes asynchronously, offering unmatched speed and efficiency in vision sensing. Despite their high dynamic range, temporal resolution, low power consumption, and computational simplicity, traditional monochrome ECs face limitations in detecting static or slowly moving objects and lack color information essential for certain applications. To address these challenges and extend upon previous work [1], we present a novel approach that integrates a Digital Light Processing (DLP) projector, forming Active Structured Light (ASL) for RGB-D sensing. By combining the benefits of ECs and projection-based techniques, our method enables the detection of color and the depth of each pixel separately. Dynamic projection adjustments optimize bandwidth, ensuring selective color data acquisition and yielding colorful point clouds without sacrificing spatial resolution. This integration, facilitated by a commercial TI LightCrafter 4500 projector and a monocular monochrome EC, not only enables frameless RGB-D sensing applications but also achieves remarkable performance milestones. With our approach, we achieved a color detection speed equivalent to 1400 fps and 4 kHz of pixel depth detection, significantly advancing the realm of computer vision across diverse fields from robotics to 3D reconstruction methods. Our code is available

publicly: [github.com/MISTLab/event\\_based\\_rgbd\\_ros](https://github.com/MISTLab/event_based_rgbd_ros)

## 5.1 Introduction

Event-based cameras (ECs) are innovative sensors that detect changes in pixel brightness asynchronously. Unlike traditional frame-based cameras, ECs do not capture full images, they generate events containing pixel coordinates, timestamps, and the polarity of brightness changes whenever a change exceeds a certain threshold. These sensors enable researchers to detect movement and changes at extremely high speeds with very low latency, minimal power consumption, and low bandwidth requirements. It makes ECs highly suitable for high-speed vision-based applications such as depth estimation.

Depth estimation is a vital element in both computer vision and robotics, used in applications like 3D modeling, augmented reality, and navigation. Structured Light (SL) systems, which project known patterns onto a scene and observe the deformations with a camera, have traditionally been used for this purpose [119]. While accurate, these systems are limited by factors such as device bandwidth and projector light power, impacting acquisition speed and performance under various lighting conditions. The high temporal resolution and high dynamic range (HDR) of ECs can address these limitations, as their asynchronous nature allows for fast, efficient data capture without the redundancy seen in frame-based systems. However, despite their advantages, traditional monochrome ECs have limitations in capturing color information.

In our previous work [1], we proposed combining an EC with a Digital Light Processing (DLP) projector to form an Active Structured Light (ASL) system for color sensing with a monochrome camera. We extended our previous work and explained how we can achieve event-based color detection and depth measurements for each pixel separately. Integrating DLP projectors with ECs overcomes the constraints of traditional SL systems. The ECs' ability to suppress temporal redundancy and their high dynamic range enables effective operation in diverse lighting conditions, enhancing the depth estimation process. Additionally, the dynamic projection adjustments allow for selective color data acquisition, ensuring efficient use of bandwidth while maintaining spatial resolution. Our method enhances depth and color detection, providing robust E-RGB-D sensing at 1.4 to 4 kHz per pixel. The main contribution is achieving ultra-fast and real-time, event-based color and depth measurements per pixel that can work in different situations within the scene. Various aspects of the main contribution are explained as follows:

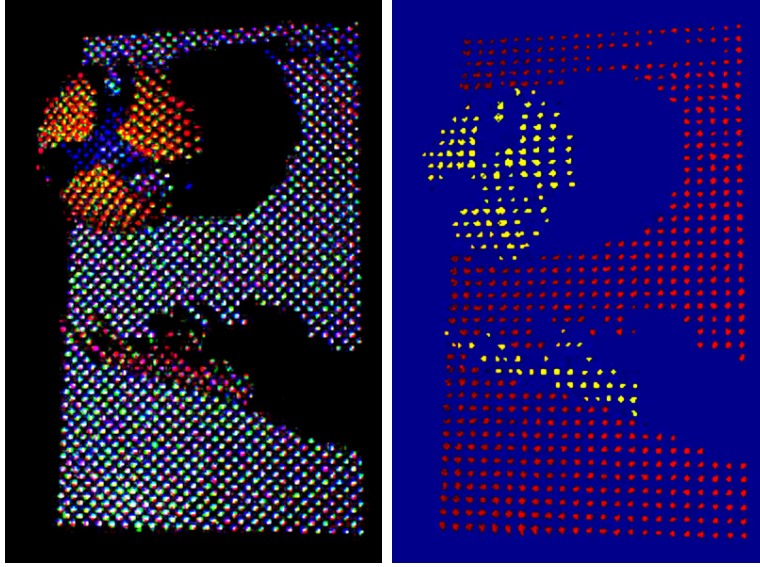


Figure 5.1 Color (left) and depth (right) detection of a volleyball ball being thrown up in front of a VGA monocular event-based camera, reconstructed using the proposed method at an equivalent speed of 120 fps from a distance of approximately 1.5 m.

**Spatial resolution** There is always a trade-off between point cloud resolution and scanning process speed in visual-based 3D scanning systems. Time-of-Flight (ToF) cameras based 3D scanners can quickly report depth information but have limitations in fast RGB detection [25]. We obtained 3D reconstructions with variable resolution, allowing us to acquire a high-resolution colored 3D point cloud of an environment (in the order of millimeters, similar to 3D laser/ToF scanners but color data included), as well as high-speed 3D scanning (in the order of milliseconds with more sparse patterns).

**Mobility limitation** A 3D scanner sensor on a mobile robot should be capable of operating at different velocities. ECs do not report anything in a static situation, and SL scanners are sensitive to motion. We introduced a method that adapts to the speed of movement by adjusting the SL pattern and camera bias, allowing for effective 3D scanning in both static and dynamic situations.

**Texture dependency** Stereo vision systems usually rely on surface textures in the feature/stereo matching process. Feature points are used to find the correlation between two camera planes and subsequently calculate the depth of those feature points in the scene. Generally, the depth accuracy of a stereo camera system decreases in the case of an untextured surface [3]. Our proposed method performs 3D capture independently of surface texture and

color.

**Acquisition speed** Vision-based scanning systems often face a trade-off between acquisition speed, resolution, and luminous efficacy [16]. The scanning speed is mainly related to the capture method, sensor bandwidth, and speed of raw data primary analysis. In our proposed method, depth parametrization and stereo matching are combined in a single step, increasing scanning speed to real-time and operating on an event-based principle by eliminating the need to search for corresponding pixels in the full captured frame. The proposed method is faster than available state-of-the-art methods, enabling it to capture more points to create the colored 3D point cloud.

**Light-Efficiency** Various methods, such as Gray coding or phase-shifting [26], have been introduced to improve the bandwidth and speed of SL-based scanning; however, these methods are limited by the power of the light source. Additionally, since standard cameras are low dynamic range devices, scanning in an environment with highly specular materials is challenging. Although there are methods to make the use of cameras feasible, they are inherently slow [27–29]. The proposed method can handle light alterations and work in dark environments by employing a high dynamic range sensor similar to [15, 16, 69, 70], but in real time and with color detection included.

**Power consumption** The power consumption of onboard sensors is crucial for mobile robots. Lower power consumption allows for longer mission durations, enabling wider area coverage and greater distances. High sensor power consumption, however, limits exploration missions. By using a high dynamic range sensor, we developed a specific active 3D scanner that manages light source power efficiently during scanning. This device minimizes energy use while still detecting object colors and distances to the camera in low-light conditions. Although it is an active method that could consume more energy than passive methods, utilizing EC allows us to control the current and power of the projector light source.

The rest of the paper is structured as follows: Section 5.2 discusses color detection methods; Section 5.3 covers event-based triangulation-based depth measurement; Section 5.4 describes our color and depth detection method; Section 5.5 details our results under various conditions; and Section 5.6 outlines concluding remarks and future work.

## 5.2 Monochrome to Color

Color information is essential for tasks like segmentation and recognition [30]. Colorization involves creating a color image from a monochrome sensor or grayscale image without sacrificing resolution. This process relies on external data about the image’s colors obtained from external device [31], user input [32], or a trained neural network that incorporates the scene’s color information [33,34]. It can be both time-consuming and costly.

The first generation of event cameras (ECs) were monochrome, with color ECs only recently becoming available [35–37]. However, color ECs have a lower resolution than monochrome ECs due to sensor size limitations and the need for color filters [38–40].

To maintain resolution while tripling the bandwidth requirements, Marcireau et al. [41] utilized dichroic filters on three ECs. They combined the output of three ECs, enabling the capture of color information in three distinct event streams.

In our initial work [1], we utilized a DLP projector to emit light patterns, referred to as ASL, onto a scene. The EC then captured the reflections of these patterns, generating events that were tagged with the scene’s color information. As Fig. 5.2 shows the proposed procedure, we are able to reconstruct full color image from a monochrome EC. Section 5.4 describes our method for color and depth detection in details in terms of pattern frequency and coverage.

## 5.3 Depth Sensing via Triangulation

In order to gather more information about the scene, fusion methods with an additional sensor are considered. While the interference-based methods are known as extremely accurate for micro-scale measuring, Time-of-flight methods are well known as low-accuracy measuring methods for a large-scale scene. Triangulation-based methods would be in the middle of them [8]. Triangulation-based techniques, including stereo vision and SL, have been demonstrated to provide precise depth information at short distances. In this paper, we focused on event-based depth measuring via SL.

### 5.3.1 Event-based depth sensing with SL

ECs can identify SL related events due to their frequency or contrast changes. By changing the frequency of SL, the camera can detect points individually. It required to be expressed that there are two different fusions of SL and EC. The first type is when a SL device is used to *create* events and capturing those events by the camera (same as the proposed method in [60]). The second method is to use a SL device (or any other 3D depth sensors) to obtain

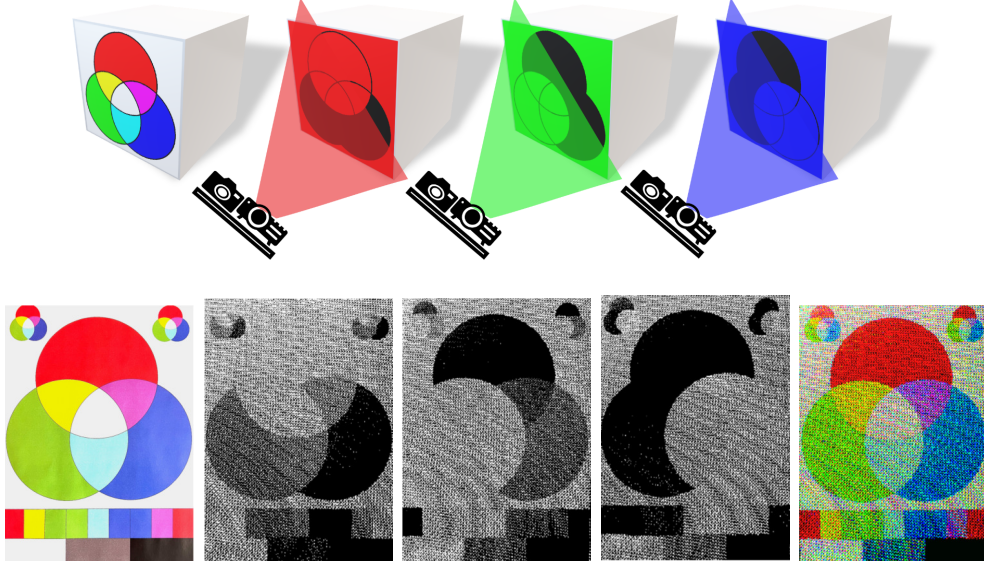


Figure 5.2 Color detection of a printed color wheel in [1]. Top: The proposed procedure involves projecting various light patterns in different wavelengths/colors. Bottom, from left to right: high-resolution Ground Truth (GT), reconstructed images captured by a VGA monochrome EC in the channels of Red, Green, Blue, and the fully reconstructed image.

a depth pre-reconstruction of the scene and subsequently adding the depth information to the captured events. For instance, in [68], the authors have used an external SL projector (Kinect) to receive the depth information of the scene. In parallel, an EC is used to obtain events (the brightness changes of each pixel due to movements), and the two outputs are merged to construct the 3D events. Therefore, despite having an active external lighting source, movement is required to generate the necessary data. Using the Kinect separately in parallel is power-consuming, and the process time would not be shortened.

In general, SL can be considered as an active stereo-vision method which uses the same concept (triangulation) for depth measurement. However, in this case, a pattern is emitted on the surface and the camera will capture the pattern deformation. Consequently, the structured pattern replaces one of the cameras in stereo-camera system. A typical SL system uses one camera and one projector, however, to reach a higher resolution or a faster full measurement more cameras can be utilized as well [13, 14].

The concept of using an external projector to utilize SL and prevent stereo challenges with a monocular EC is foundational to modern systems utilizing event cameras including [15, 16, 69, 70]. Table 5.1 summarizes previous structured light (SL) systems that have tackled the issue of depth estimation using event cameras. The SL methods are different in terms of patterns types.

Table 5.1 Summary of Previous SL-based Systems Addressing Depth Estimation with monocular EC.

Method	EC	Sensor	Projector
Brandli et al. [15]	DVS128	$128 \times 128$	Laser line 500 Hz
MC3D [16]	DVS128	$128 \times 128$	Laser point 60 fps
FTD [74]	ATIS0	$304 \times 240$	DLP TI LightCrafter 3000
FPP [20, 75]	DAVIS346	$346 \times 260$	DLP TI LightCrafter 4500
ESL [69]	Prophesee Gen3	$640 \times 480$	Laser point 60 fps
X-maps [70]	Prophesee Gen3	$640 \times 480$	Laser point 60 fps
SGE [72]	Prophesee Gen4	$1280 \times 720$	Laser point 60 fps (for static scene) DLP projector OPR305185 (for dynamic scene)
<b>Ours</b> ERGBD	Prophesee Gen3	$640 \times 480$	DLP TI LightCrafter 4500

**Coded patterns SL:** Strips SL are introduced to perform the area scanning. If they are unique, the system can identify points and calculate the depth. If these strips are in black and white (*binary coding*), series of patterns are needed to identify the corresponding points. Thus compared to the other patterns, binary patterns are sensitive to object movement. Consequently, a high frequency of switching the binary patterns is assisting. DLP projectors have the ability to switch patterns in the order of kilo Hertz [71, 72]. To have a higher resolution, 2D and hybrid patterns are introduced [73].

Leroux et al. [74] present a 3D reconstruction method using an Asynchronous Time-based Image Sensor (ATIS) with  $304 \times 240$  pixels and a DLP projector. Instead of line structured light (SL), Frequency-Tagged Dots (FTD) are projected onto the scene. The known pattern and distances allow calculation of pattern deformation and point depth. This method, however, is highly sensitive to movement and ambient light changes, which affect the number of captured events.

Mangalore et al. [20] and Li et al. [75] used a Dynamic Active-pixel Vision Sensor (DAVIS346,  $346 \times 260$  pixels) with a DLP LightCrafter 4500 for 3D reconstruction. They developed a Fringe Projection Profilometry (FPP) system using a moving fringe pattern, allowing the EC to scan multiple lines simultaneously, which is faster than line-scanning methods. The EC's advantage is detecting shadowed areas, unlike frame-based cameras where shadows and dark regions appear the same. However, a pre-recording of the scene without the object is needed to "inpaint" shadowed areas, and the camera's limited event reporting capacity can lead to some events being eliminated.

**Simple patterns SL:** The simplest method, known as the statistical pattern, involves a random distribution of dots. This method is used in various commercial devices such as Microsoft Kinect V1, Intel RealSense [76], and Orbbec Astra [77] due to its simplicity and small footprints. Huang et al. [2] also frequently projected a single pseudo-random pattern

using a DLP6500 projector. They generated event frames from the event stream and utilized a digital image correlation method to calculate displacements and derive 3D surfaces of target objects. While using a single dot-based pattern increased scanning speed, it sacrificed detailed information compared to dense information achievable with line-based patterns. Moreover, implementing a discrete pattern led to inaccurate dot location and sensitivity to ambient light, resulting in low-resolution 3D depth measurements [8].

Another way to acquire higher resolution is to use line instead of discrete dots. In that way, high accuracy measurement in one direction will be attained. This method is also combined with a line laser for short range scanning. However, to measure the depth in all directions the line direction needs to vary.

Brandli et al. [15] used a laser line and an EC to scan the surface of an object. Using a concentrated light line (laser) helped achieve more contrast and detect the line more easily with the EC. Additionally, if the environment is static, emitting the structured light (SL) helps detect only the relevant pixels. However, to obtain a 3D reconstruction of the scene using this method, it is necessary to change the line direction or move the object in front of the laser line.

By using the line laser and the EC, Matsuda et al. [16] (Motion Contrast 3D Scanning or MC3D), resolved the speed-resolution trade-off issue present in traditional SL scanners. Traditional SL scanners become inoperative when illumination changes occur in the environment; however, using a high dynamic range EC yields an improved final result. The laser scanner had an exposure time of 28.5 seconds, but their proposed device had an exposure time of one second.

Expanding on the Matsuda et al.'s work, Muglikar et al. [69] (Event-based Structured Light or ESL), employed time maps to establish a temporal link between the projector and camera. Initially, they produced depth maps by conducting an epipolar disparity search within rectified projector time maps. Following this, an additional processing stage was implemented to enhance pixel-level coherence and reduce event fluctuations. However, this stage demands significant computational resources, preventing their method from achieving real-time performance.

Morgenstern et al. [70] (X-maps), introduced a method that converts the projector time map into a rectified X-map, capturing X-axis correspondences for incoming events and enabling direct disparity lookup without additional search. This method supported real-time interactivity, making it suitable for Spatial Augmented Reality (SAR) experiences requiring low latency and high frame rates. They claimed that their method is 7 to 100 times faster than considering the entire frame, as in the ESL method, because there is no need to do a row-by-



row disparity search and calculate the depth for the whole frame. We used a partially similar X-mapping method to calculate the depth for each pixel. A detailed description is provided in the next section.

#### 5.4 ASL: Adaptive Structured Light

The proposed method is to adjust the SL pattern to balance the trade-off between scanning speed and detail. Achieving a denser output requires more pixel data, but there is a limit to the number of events an EC can process simultaneously, as it may become bus-saturated. The percentage of Ground Truth (GT) points estimated by the proposed method, relative to the total number of pixels in the GT that contain data, is referred to as Fill Rate (FR), completeness, or depth map completion [69]. However, to avoid reaching the camera’s bus-saturation limits, we control the number of projected points by changing pattern. We refer to the ratio of ON to OFF pixels in a pattern as the Coverage Percentage (CP) [1]. Each pattern type has a different CP, and each experiment is evaluated based on its FR.

The high-resolution method, which uses line-based scanning and is less sensitive to ambient light [16, 120], is given the highest priority among the proposed patterns, while the dot-based pattern is assigned the lowest priority. Figure 5.3 represents the various patterns that ERGBD can work with, including line-based scanning, dot-based scanning, and pseudo-random dot patterns. For instance, the SL pattern can be changed according to either the remaining battery charge or the robot’s motion speed. The robot can switch to the line method and decrease the LED current (decrease the power of the projector) when the energy level reaches its critical state.

To reconstruct the color and depth, we introduced different pattern sequences, we will describe them separately in detail in the following subsections.

##### 5.4.1 Color detection

As mentioned in Section 5.2 and shown in Figure 5.2, we project each pattern three times onto the scene with different wavelengths/colors, followed by capturing the reflection with the EC. These patterns can be part of a depth measuring procedure (moving line or dots) or a single pattern just to detect the color. Figure 5.4 shows the pattern sequence and their exposure times in microseconds. There are 4 different types of sequences to obtain color and/or depth. For instance, when measuring depth first and then color, any pattern from Figure 5.3 could be used to capture the color, even if the depth pattern is different. So, based on the needs, one pattern could be denser and have a higher CP than the other. One could

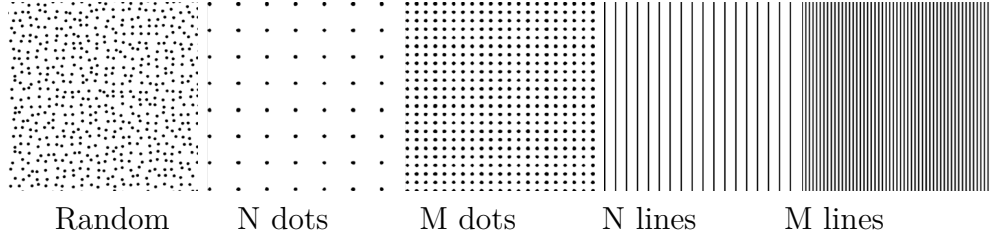


Figure 5.3 Proposed ASL pattern types for balancing speed and detail in ERGBD scanning.  $M$  is greater than  $N$ , which means reconstructing more points and have higher CP. We did not use a pseudo-random dot pattern to detect depth, although it is commonly used in similar approaches [2]. However, with our method, it is possible to reconstruct color in addition to detecting depth.

**Mode One:** Color only

ID	$R$	$G$	$B$
250	235	235	235

**Mode Two:** Depth only

ID	$D_1$	...	$D_n$
260	235	...	235

**Mode Three:** Depth then Color

ID	$D_1$	...	$D_n$	$R$	$G$	$B$
270	235	...	235	235	235	235

**Mode Four:** Depth and Color

ID	$D_1$	...	$D_n$	$D_1$	...	$D_n$	$D_1$	...	$D_n$
280	235	...	235	235	...	235	235	...	235

Figure 5.4 Pattern sequence and their exposure times in microseconds. In our experiments, we used mode 3 with two different values for  $n$  (23 and 45), and mode 4 with  $n=23$ . One ID for each pattern type would be enough, but we could have different IDs for each color or depth pattern mode. While this increases the total scanning time, it makes the system more robust and trackable.

use a solid pattern and decrease the Region of Interest (ROI) of the camera to prevent bus saturation while still obtaining a fully dense, colorful image for a specific area.

Since the projector and camera are connected through the external trigger pins, we projected a blank pattern with a specific exposure time as the ID, allowing the software to detect which pattern sequence mode is being projected by the projector. Both the ESL and X-maps methods utilize a Micro Electro-Mechanical System (MEMS) laser projector limited to 60 Hz,

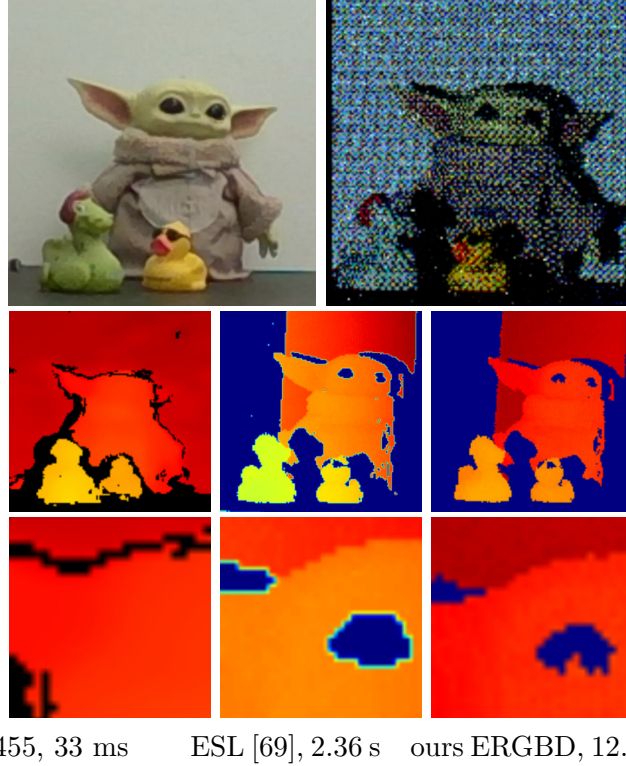


Figure 5.5 **Top Left:** Output of the D455 RGB camera. **Top Right:** The RGB image reconstructed by the ERGBD with a Monochrome EC. **Middle Row:** Depth detection comparison between D455 (Left), ESL (Middle), and ours ERGBD (Right). **Bottom Row:** Zoomed-in view of the middle row. Note that the color differences are due to defining different minimum and maximum values for the jet-coded colorization; the actual difference in mm is detailed in Section 5.5.3.

making them suitable for raster scanning patterns where rows are projected sequentially. In contrast, our work employs a Digital Micromirror Device (DMD projector) capable of projecting all pixels simultaneously, with a capacity exceeding 4 kHz. The top row of Figure 5.5 shows the RGB reconstructed by our method compared to the output color image of the RealSense D455 camera.

#### 5.4.2 Depth detection

Like X-maps [70], we provided a lookup table for disparity values to eliminate the need for disparity search. Because computing scene disparity by aligning time entries of the map along epipolar lines with an idealized projector time map is computationally intensive [69]. However, our DMD projector does not exhibit raster printing behavior, so we did not store the projector’s  $x$  coordinates in relation to  $y$  and time  $t$ , and we do not use a temporal map either. Instead, we determine the disparity by knowing the column of the projected line. We

will describe it in detail in this section.

**Direct disparity lookup table:** We formulate the problem of depth estimation using epipolar lines. After calibrating the system and setting up the stereo configuration, we create a lookup table  $LUT_c$  that assigns each pixel on the camera plane  $P_c(x_c, y_c)$  to its corresponding pixel on the rectified camera's image  $P_{c_r}(x_{c_r}, y_{c_r})$ . Additionally, a lookup table  $LUT_p$  assigns each pixel  $P_p(x_p, y_p)$  on the projector plane to  $P_{p_r}(x_{p_r}, y_{p_r})$  on the rectified projector image.

$$LUT_c(P_c(x_c, y_c)) = P_{c_r}(x_{c_r}, y_{c_r}) \quad (5.1)$$

$$LUT_p(P_p(x_p, y_p)) = P_{p_r}(x_{p_r}, y_{p_r}) \quad (5.2)$$

Considering that we are projecting different numbers of lines on the object, we could have an array *Columns* that carries the column number  $x_p$  of each line on the projector plane. As shown in Figure 5.4, the size of this lookup array could vary from 1 to  $n$  based on the pattern mode.

$$Columns = [x_{p_1} \quad x_{p_2} \quad \dots \quad x_{p_n}] \quad (5.3)$$

By identifying which pattern or columns are being projected and knowing the coordinates of the captured event, we can directly determine the disparity of the incoming event. Imagine at the time of receiving the  $Event_c(x_c, y_c)$ , we are projecting line number  $m$ ; then, we can determine two points of this line on the projector rectified map as:

$$Top_{p_r} = LUT_p(Columns[m], H) = (x_T, y_T) \quad (5.4)$$

$$Bottom_{p_r} = LUT_p(Columns[m], 0) = (x_B, y_B) \quad (5.5)$$

where  $H$  represents the height of the projector resolution.

Because our setup is a horizontal stereo, the epipolar lines in the rectified images are horizontal and have the same y-coordinate. The corresponding point is determined by calculating the intersection of the line that passes through these two points and the horizontal line that passes through the event's pixel on the rectified camera plane.

$$Event_{c_r} = LUT_c(x_c, y_c) = (x_E, y_E) \quad (5.6)$$

$$x_{p_r} = x_T + (y_E - y_T) \cdot slope \quad (5.7)$$

where  $slope = \frac{x_B - x_T}{y_B - y_T}$ . And the disparity would be:

$$disparity = x_{p_r} - x_E \quad (5.8)$$

The depth  $Z$  of a point in a scene can be calculated using the formula:

$$Z = \frac{f \cdot B}{\text{disparity}} \quad (5.9)$$

where  $f$  is the focal length of the camera,  $B$  is the baseline. Algorithm 1 provides a concise overview of the entire process.

We were able to generate a temporal map similar to the ones produced by using raster printing projectors. Figure 5.6 displays the temporal map of Figure 5.5 setup, created by ESL and ERGBD (ours). Although we do not use the temporal map, this figure shows that we could generate one simply by knowing which column is being projected by the DMD projector and assigning a temporal index/color to that specific receiving event. However, in the other methods, they need to receive all events and normalize the temporal map based on the time of receiving the first and the last event.

Moreover, in X-maps, to obtain values for all possible measurements from the camera and create the lookup reference, they needed to record events at least for one time. However, we do not need to record anything and we can directly find out the depth by knowing the column of the projected line on the rectified projector image.

In practice, the projector’s resolution is usually higher than that of the camera sensor, which means the EC may not capture individual projector columns or may see overlapping columns. To address this, we introduced a gap between each line in our fully dense patterns, ensuring a solid, dense temporal map with the camera.

Our method publishes events, and a separate ROS node aggregates these events and publishes frames at various speeds. Since our system operates on an event-based model rather than a frame-based one, it does not require complete pattern captures. However, as a consequence of this approach, some patterns may not be fully captured at higher frequencies. This contrasts with traditional methods that rely on complete pattern captures to generate data. One metric for assessing output quality is the Fill Rate (FR), which compares the number of pixels containing data in the current frame to those in the ground truth frame. For instance, patterns with 23 lines complete faster compared to those with 45 lines, allowing quicker coverage of the field of view and achieving a higher FR. However, reducing the number of lines sacrifices detail. This trade-off between speed and detail is a deliberate aspect of our approach, which is not achievable with other methods. Raster-based projectors frequently project a solid pattern, while methods using coded patterns like grayscale or binary codes require capturing all patterns to report depth. They cannot increase speed by reducing detail, nor can they enhance detail by sacrificing speed.

---

Algorithm 1 Stamping events with depth and color

```

1: Initialize  $LUT_p$ ,  $LUT_c$ , and  $Columns$  based on calibration and pattern mode
2: while camera triggers events do
3:   if external trigger then
4:     Increment  $m$  and/or indicate the new color
5:   else
6:     Capture  $Event_c(x_c, y_c)$ 
7:     Retrieve  $Event_{c_r}(x_E, y_E)$  from  $LUT_c(x_c, y_c)$ 
8:     Retrieve  $x_p$  from  $Columns[m]$ 
9:     Retrieve  $Top_{p_r}(x_T, y_T)$  and  $Bottom_{p_r}(x_B, y_B)$  from  $LUT_p(x_p, H)$  and  $LUT_p(x_p, 0)$ 
10:    Calculate  $x_{p_r} = x_T + (y_E - y_T) \cdot \frac{x_B - x_T}{y_B - y_T}$ 
11:    Compute  $depth = \frac{f \cdot B}{x_{p_r} - x_E}$ 
12:    Publish depth and color-stamped events alongside the colored point cloud on ROS topics.
13:   end if
14: end while

```

---

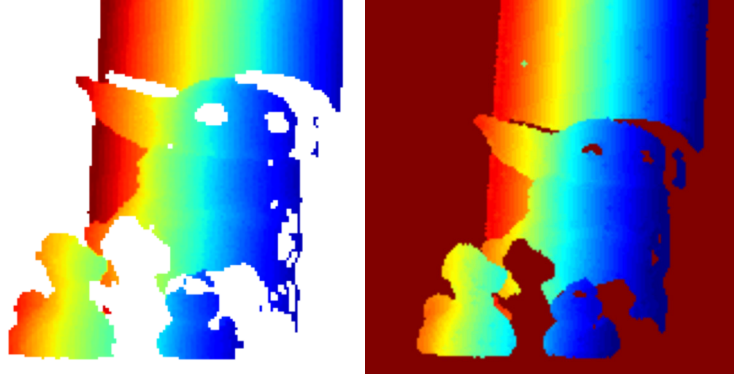


Figure 5.6 Temporal map reconstructed by ESL (Left), ours ERGBD (Right).

## 5.5 Experiments

This section assesses the performance of our event-based SL system for depth and color estimation. We begin by introducing the hardware setup (Section 5.5.1), along with the baseline methods and ground truth used for comparison (Section 5.5.2). Subsequently, we conduct experiments on static scenes to quantify the accuracy of the proposed method and on dynamic scenes to demonstrate its high-speed acquisition capabilities (Section 5.5.3).

### 5.5.1 Setup

**Camera:** A Prophesee Evaluation Kit 1 (EVK1) [121] with a Gen 3.0 sensor is used for the event camera. This dynamic vision sensor offers a resolution of  $640 \times 480$  pixels with a  $15 \mu\text{m}$  pixel pitch and only detects contrast changes. It features a dynamic range greater than 120 dB, an average latency of  $200 \mu\text{s}$ , and timestamps events with microsecond precision (Figure 5.8).

**Projector:** To project a binary pattern at high speed (over 4 kHz), the DLP LightCrafter 4500 [122] projector utilized, it can project patterns at  $912 \times 1140$  resolution in diamond pixel configuration with a  $235 \mu\text{s}$  exposure period, allowing a 4.225 kHz switching rate, easily captured by the EC (Figure 5.8). Because of the diamond pixel array of the DMD, the pixel data does not appear on the DMD exactly as it would in an orthogonal pixel arrangement. Figure 5.7 shows the pattern that we used to project dots in our dot-based patterns.

**Calibration:** To calibrate the system, we introduced a camera-projector calibration approach specifically designed for event-based SL. Our method involves calibrating the intrinsic parameters of the event camera by utilizing a standard calibration tool (OpenCV [123]) on the images generated after converting events into images. This conversion is done while observing a flickering circle-grid pattern from various angles. We chose for circle patterns over checkerboards due to their superior performance in terms of both the quality and stability of the final calibration results across multiple iterations [124]. Unlike checkerboards, which may lead to uncertainties in corner detection, circle-grid patterns allow for more precise extraction of circle centers, for instance, through the calculation of the center of gravity of all circle pixels.

Our approach for this calibration is straightforward and adaptable. We begin by setting up a circle grid of flickering LEDs on a flat surface and projecting another circle grid pattern using the projector beside it on the board. After calibrating the intrinsic parameters of the event camera using the LED circle-grid pattern, we proceed to calibrate the extrinsic parameters of the camera-projector setup and the intrinsic parameters of the projector.

Once the camera calibration is complete, we determine the board’s relative position, which in turn allows us to calibrate the projector. This streamlined method enables the simultaneous calibration of all parameters through simple operations, significantly reducing the complexity typically associated with calibration processes. Furthermore, it is universally applicable to all event-based SL systems. Figure 5.8 shows the calibration circle dot-board used for calibrating the system.

Luo et al. [125] proposed an alternative approach for calibrating the SL camera-projector

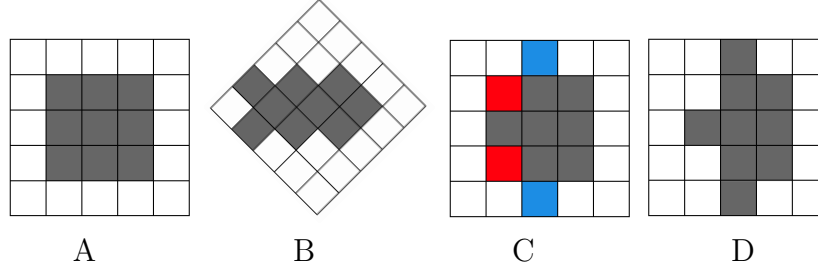


Figure 5.7 The pattern of one dot with the size of  $3 \times 3$  pixels in our dot-based patterns. If we project (A) due to the diamond pixel configuration of the DMD, we will see (B) on the object surface. Which could lead to mislocating the center of the projected dot. Therefore, we should move red-colored pixels to blue ones in (C) and project (D) to achieve the pattern (A) on the object’s surface with a 45-degree rotation.

system. Their method involves introducing four reference planes and generating lookup tables for pixel correspondences. While this method can enhance calibration quality, it adds complexity to the procedure. For this work, we chose a simpler approach, but their method could be used to achieve better accuracy if needed.

Similarly, Wang et al. [126] presented a calibration technique based on Temporal Matrices Mapping (TMM). They utilized two temporal matrices to establish pixel correspondences between the SL projector plane and the event camera (EC) plane. Although this approach can improve calibration accuracy, it also increases the overall complexity of the calibration process.

**Software:** We have implemented our method on the Robot Operating System (ROS). The designed Qt-based Graphical User Interface (GUI) enables us to control the camera settings online, and the ROS nodes to publish color and depth-stamped events alongside RGB frames and colorful point clouds on ROS topics. More details can be found on our Event-based RGBD ROS Wrapper [1] Git repository at [github.com/MISTLab/event\\_based\\_rgbd\\_ros](https://github.com/MISTLab/event_based_rgbd_ros).

### 5.5.2 Baseline and Ground Truth

As shown in Table 5.1, the most recent related works are the ESL [69], the X-maps [70], and the SGE [72]. Since all of them used a laser point projector with a scanning speed of 60 Hz, they were automatically excluded from comparison in high-speed real-time event-based scanning. However, we have chosen the ESL work for comparison to the final result since both the ESL and the X-maps use the same type of pattern, which is raster scanning, and this could be considered partly similar to line scanning. In addition, the ESL used a raw recorded event file and pre-data-processing to synchronize and convert the raw event data



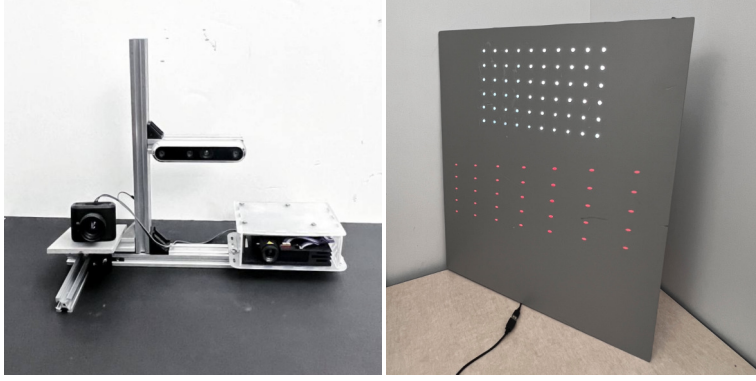


Figure 5.8 **Left:** The experimental setup includes a DLP LightCrafter 4500 Evaluation Module, a Prophesee evaluation kit (Gen3-VGA), and an Intel RealSense D455 (only used for comparison). **Right:** The calibration circle dot-board used for calibrating the EC and camera-projector setup, with white dots from LEDs and red dots projected by the projector.

file in the absence of triggers. This makes it more accurate (but slower) than the X-maps method because it considers all events together rather than processing them in real-time. More importantly, we can still generate the time-maps even by using a DLP. We did not use the SGE method in dynamic conditions because their code is not yet publicly available.

In this task, obtaining the GT is challenging due to the lack of methods capable of producing dense depth with accuracy above the millimeter level for natural scenes. Therefore, ESL [69] adopts averaged MC3D [16] as their ground truth, whereas X-maps [70] employs optimized ESL. While we acknowledge the potential for accuracy enhancement through averaging operations in depth estimation, our approach differs due to reconstructing color alongside depth. Given the utilization of a distinct point scanning method and projector type, we developed our method to generate the ground truth by accumulating more events over an extended time window (one second) and performing averaging across 10 scans.

### 5.5.3 Results

**Static Environments:** To evaluate the outcome of the proposed method in static situations, we have designed seven different setups (see Figure 5.14). We utilized line and dot patterns in modes 3 and 4 (see Figure 5.3). It is necessary to mention that in mode 4, the color and depth patterns are the same. However, in mode 3, different patterns can be used (e.g., in mode 3, we utilized dot/line patterns to detect depth but used dot patterns to detect color). To provide abbreviated names for each pattern sequence, we used the following method: for example, 'M3L45' indicates mode 3 with 45 lines in the pattern for depth detection.

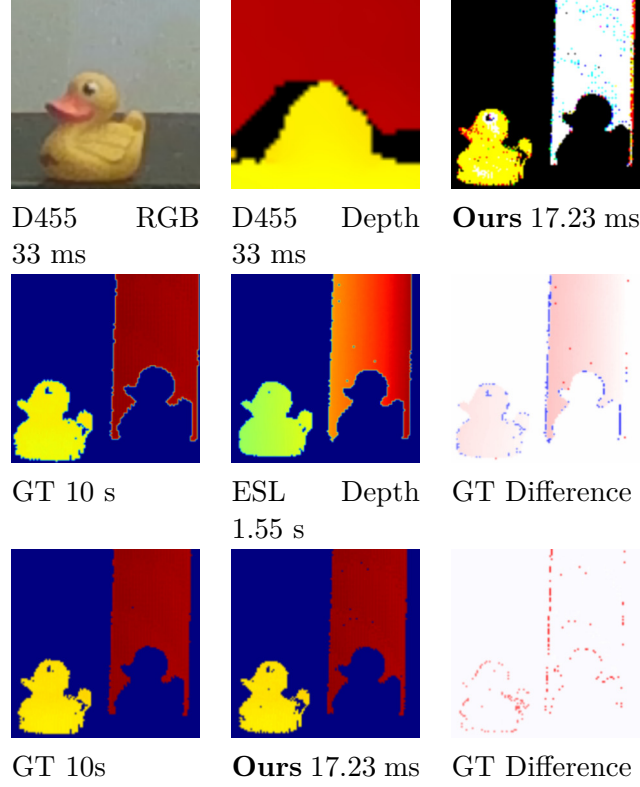


Figure 5.9 Comparison of color and depth detection for Duck setup (M4L23 pattern).

Projecting 23 patterns, along with color patterns, in mode 3 takes 7.4 ms, while projecting 23 patterns three times in mode 4 takes 17.23 ms. Projecting 45 patterns along with the color patterns in mode 3 takes 12.57 ms. To scan static objects, we used the patterns M3L45, M3D45, and M4L23.

Each pattern falls into one of three categories: wide, normal, and dense. To indicate the category, we used Coverage Percentage (CP). To calculate the CP, we consider the area size between the first and last lines rather than the entire projector plane. Otherwise, considering 45 lines dense or sparse would yield the same CP. In general, a higher CP in our case indicates a denser pattern with a smaller field of view (FOV), while a lower CP suggests a more sparse pattern that covers a larger area with a higher FOV.

Our evaluation spanned three different speeds: 1, 26, and 58 fps. However, as mentioned in Subsection 5.4.2, our system is an event-based model rather than a frame-based one. To create rain cloud figures (Figures 5.10 and 5.11), we used raincloud plots [127] over more than 25k frames of data.

The Analysis of Variance (ANOVA) tables (Tables 5.2 to 5.6) demonstrate significant effects of both the Speed and CP factors (as well as their interaction), on the dependent variables

(PSNR, RMSE, and FR). Residuals meet the normality assumption, supported by Q-Q plots in Figures 5.12 and 5.13, validating the model. This enhances the reliability of the ANOVA results for RMSE, PSNR, and FR, confirming statistical validity under normality assumptions. To generate ANOVA tables, we used JASP [128].

Figure 5.9 illustrates color and depth detection for the Duck setup using pattern M4L23. Our method achieves depth frame reconstruction about 90 times faster than the ESL method, while also simultaneously reconstructing color. Compared to RealSense, our method provides sharper and more accurate results. For a comprehensive comparison of all setups in a static scenario, refer to Figure 5.14.

**Dynamic Environments:** To evaluate the outcome of the proposed method in dynamic situations, we have designed five different setups. To scan dynamic objects, we used the patterns M3D45, M3L23, and M3D23. We did not utilize pattern M4L23, although it has better output in terms of color in low-speed scanning. This is because considering the number of patterns that it needs to project, the total scanning speed could decrease, which is critical in dynamic conditions.

Since there is no GT as mentioned in Subsection 5.5.2, we could not provide FR or RMSE for dynamic experiments. Figure 5.15 shows the color and depth detection of a volleyball ball being thrown up in front of the setup at a distance of approximately 1.5 m, using pattern M3L23. It also shows that the D455 sensor captured a slightly blurred RGB image, and the depth detection is not sharp and accurate.

## 5.6 Conclusion

This study introduces a method for generating colored point clouds with adjustable speed and resolution, enabling depth map creation in challenging environments like dynamic and low-light conditions, even with stationary cameras or objects.

Results show that denser scanning patterns provide more accurate data for smaller fields of view (FOV), while sparser patterns cover larger areas. Using fewer patterns, we achieved a scanning time of 7.4 ms, significantly faster than the RealSense D455 (33 ms) and the ESL method (2 to 5 seconds). This method balances detail and speed effectively, offering improved speed control and enabling color reconstruction.

Using this setup, we achieved color scanning speeds up to 1.4 kHz (Mode 1, as investigated in our previous work) and pixel-based depth scanning speeds up to 4 kHz (Mode 2 with  $n = 1$ ). This provides a continuous stream of annotated events with color and depth, along with detailed, colorful point cloud output. The method shows versatility across static and dynamic

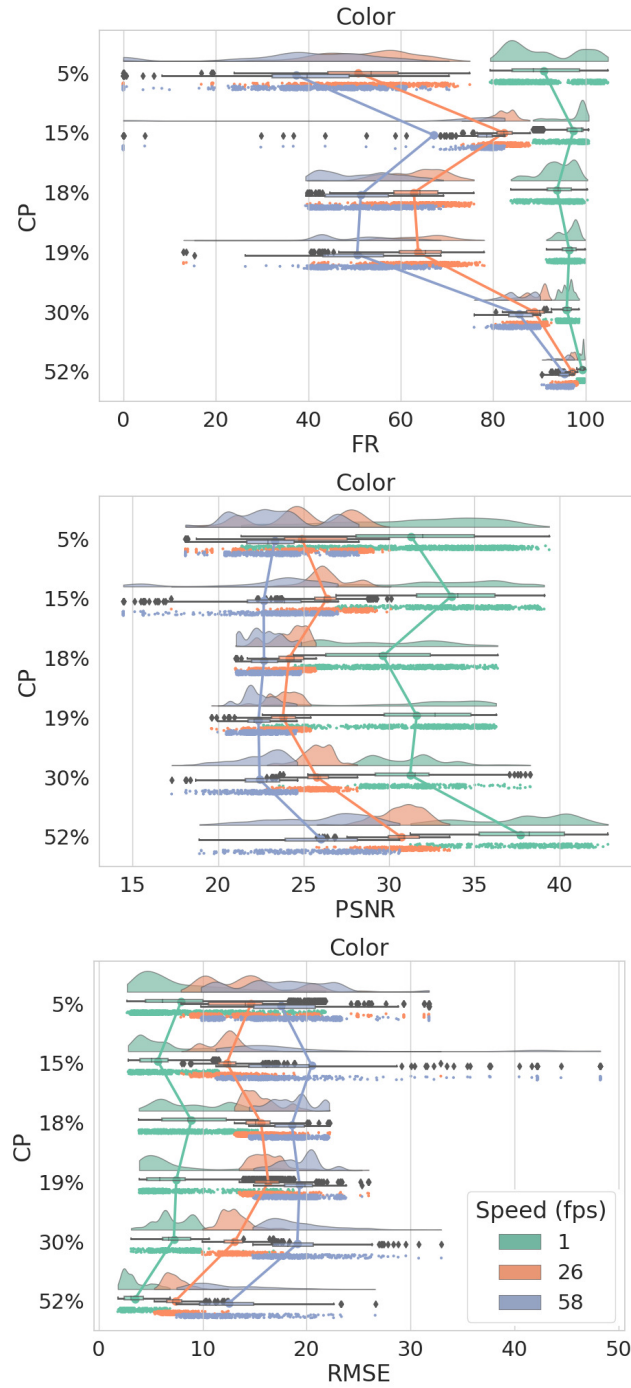


Figure 5.10 Comparison of color detection for all setups at different speeds.

environments, introducing an innovative strategy for balancing resolution and acquisition speed.

In light of the findings and outcomes of this study, several promising avenues for future research have emerged:

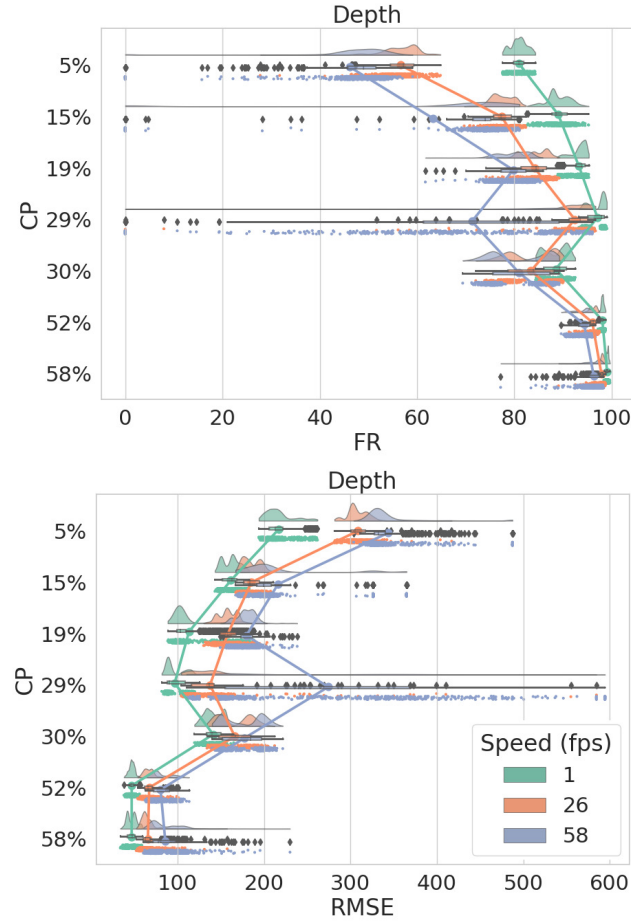


Figure 5.11 Comparison of depth detection for all setups at different speeds.

Table 5.2 Analysis of Variance (ANOVA) for the Color PSNR

Cases	Sum of Squares	df	Mean Square	F	p
Speed	164027.653	2	82013.826	11569.337	< .001
CP	33584.097	5	6716.819	947.513	< .001
Speed * CP	6169.135	10	616.913	87.025	< .001
Residuals	89192.490	12582	7.089		

Table 5.3 Analysis of Variance (ANOVA) for the Color RMSE

Cases	Sum of Squares	df	Mean Square	F	p
Speed	225230.139	2	112615.070	7267.745	< .001
CP	42003.544	5	8400.709	542.150	< .001
Speed * CP	12562.719	10	1256.272	81.075	< .001
Residuals	194960.457	12582	15.495		

Table 5.4 Analysis of Variance (ANOVA) for the Color FR

Cases	Sum of Squares	df	Mean Square	F	p
Speed	$1.814 \times 10^{+6}$	2	906934.152	8884.279	< .001
CP	$1.897 \times 10^{+6}$	5	379446.459	3717.037	< .001
Speed * CP	658311.623	10	65831.162	644.879	< .001
Residuals	$1.284 \times 10^{+6}$	12582	102.083		

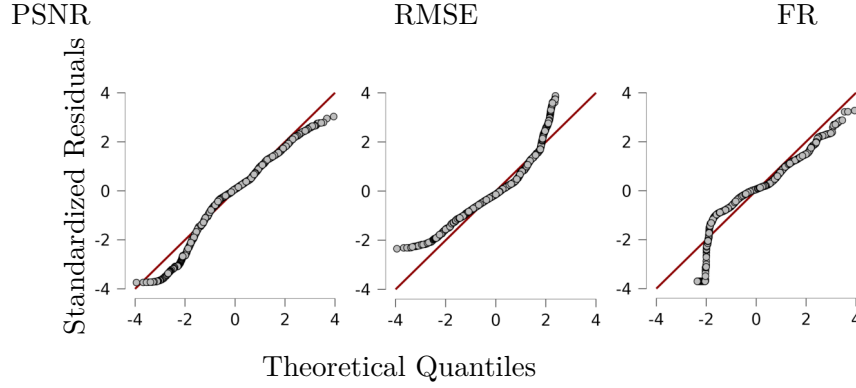


Figure 5.12 Quantile-Quantile plots of the color detection dataset.

Table 5.5 Analysis of Variance (ANOVA) for the Depth RMSE

Cases	Sum of Squares	df	Mean Square	F	p
Speed	$1.136 \times 10^{+7}$	2	$5.680 \times 10^{+6}$	4288.971	< .001
CP	$6.380 \times 10^{+7}$	6	$1.063 \times 10^{+7}$	8028.498	< .001
Speed * CP	$6.595 \times 10^{+6}$	12	549596.715	414.970	< .001
Residuals	$1.666 \times 10^{+7}$	12579	1324.425		

Table 5.6 Analysis of Variance (ANOVA) for the Depth FR

Cases	Sum of Squares	df	Mean Square	F	p
Speed	518315.917	2	259157.958	3401.744	< .001
CP	$1.689 \times 10^{+6}$	6	281427.178	3694.053	< .001
Speed * CP	329420.234	12	27451.686	360.335	< .001
Residuals	958316.716	12579	76.184		

1. **Optimizing resolution of color and depth detection** by projecting a denser pattern in a specific area could be achieved through the implementation of a movement detection algorithm during SL light downtime or by using a secondary camera to detect movement in parallel.
2. **Increasing scanning speed** at the cost of higher bandwidth usage, while reducing

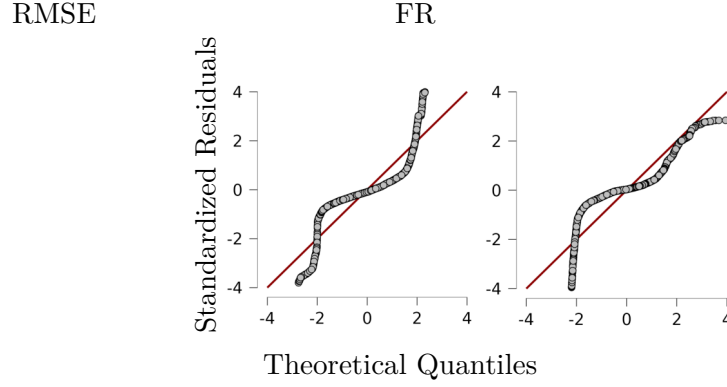


Figure 5.13 Quantile-Quantile plots of the depth detection dataset.

resolution, can be achieved by projecting white-colored patterns when utilizing a color event camera (EC).

3. **Optimizing depth measurement** based on the SL wavelength and the object's surface color by controlling the current of each LED of the projector.
4. **Increasing the scanning range** by a few meters while maintaining accuracy could be achieved by using a Near-IR projector instead of capturing color.




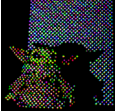

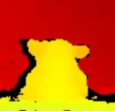
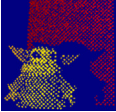
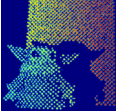
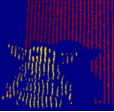

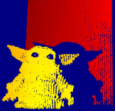
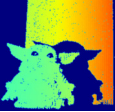




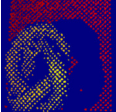
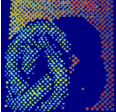
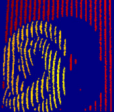
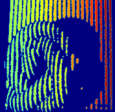
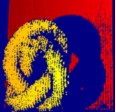
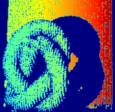
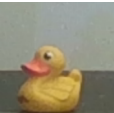
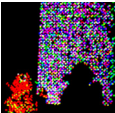


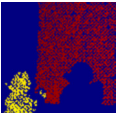
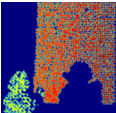

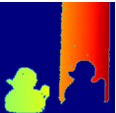

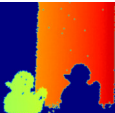

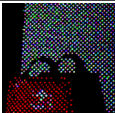
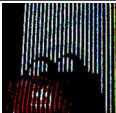

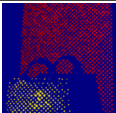
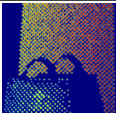
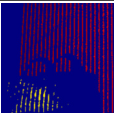
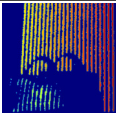
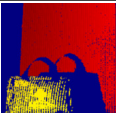
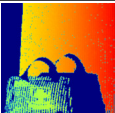
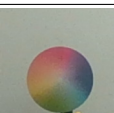
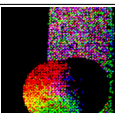
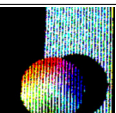

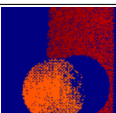
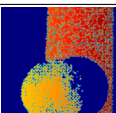
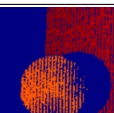
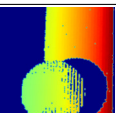
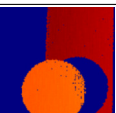
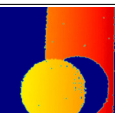

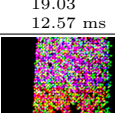
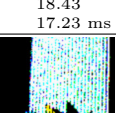

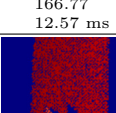
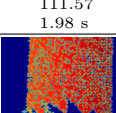
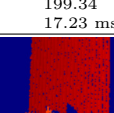
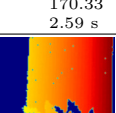
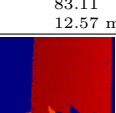
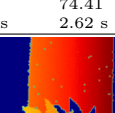

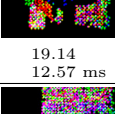
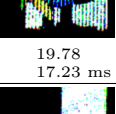

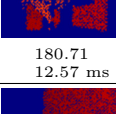
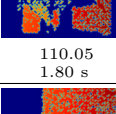
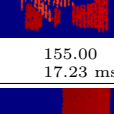
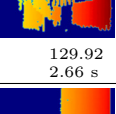
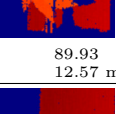
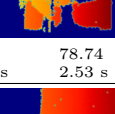
	D455 30fps RGB	Ours M3D45 RGB	Ours M4L23 RGB	D455 30fps Depth	Ours M3D45 Depth	ESL M3D45 Depth	Ours M4L23 Depth	ESL M4L23 Depth	Ours M3L45 Depth	ESL M3L45 Depth
Yoda										
	RMSE: Time:	20.04 12.57 ms	22.42 17.23 ms	RMSE: Time:	330.23 12.57 ms	300.84 2.35 s	228.03 17.23 ms	241.12 1.93 s	335.03 12.57 ms	199.83 5.37 s
Foam										
	RMSE: Time:	19.11 12.57 ms	17.51 17.23 ms	RMSE: Time:	328.98 12.57 ms	334.30 1.97 s	198.69 17.23 ms	196.33 2.70 s	161.86 12.57 ms	213.26 4.33 s
Duck										
	RMSE: Time:	18.64 12.57 ms	12.48 17.23 ms	RMSE: Time:	189.59 12.57 ms	238.92 2.35 s	77.56 17.23 ms	74.10 1.55 s	85.28 12.57 ms	87.03 5.37 s
Bag										
	RMSE: Time:	22.38 12.57 ms	21.53 17.23 ms	RMSE: Time:	371.73 12.57 ms	385.64 2.33 s	222.13 17.23 ms	189.69 2.14 s	326.28 12.57 ms	153.63 4.71 s
Circle										
	RMSE: Time:	19.03 12.57 ms	18.43 17.23 ms	RMSE: Time:	166.77 12.57 ms	111.57 1.98 s	199.34 17.23 ms	170.33 2.59 s	83.11 12.57 ms	74.41 2.62 s
Pinwheel										
	RMSE: Time:	19.14 12.57 ms	19.78 17.23 ms	RMSE: Time:	180.71 12.57 ms	110.05 1.80 s	155.00 17.23 ms	129.92 2.66 s	89.93 12.57 ms	78.74 2.53 s
Stand										
	RMSE: Time:	20.46 12.57 ms	12.53 17.23 ms	RMSE: Time:	181.56 12.57 ms	121.65 2.19 s	83.32 17.23 ms	100.22 1.60 s	85.25 12.57 ms	71.84 2.79 s

Figure 5.14 Comparison of color and depth detection for static scenes between RealSense D455, ESL, and ERGBD (ours). Noting that the available code for ESL needs to create a temporal map as a numpy array from a recorded raw file, we did not include the time required for these processes in our calculations. We only considered the time needed for calculating depth from those files. The operating system had an NVIDIA(R) GeForce RTX(TM) 2060 6GB GPU and an Intel Core i7-9750H CPU with 16GB of memory.



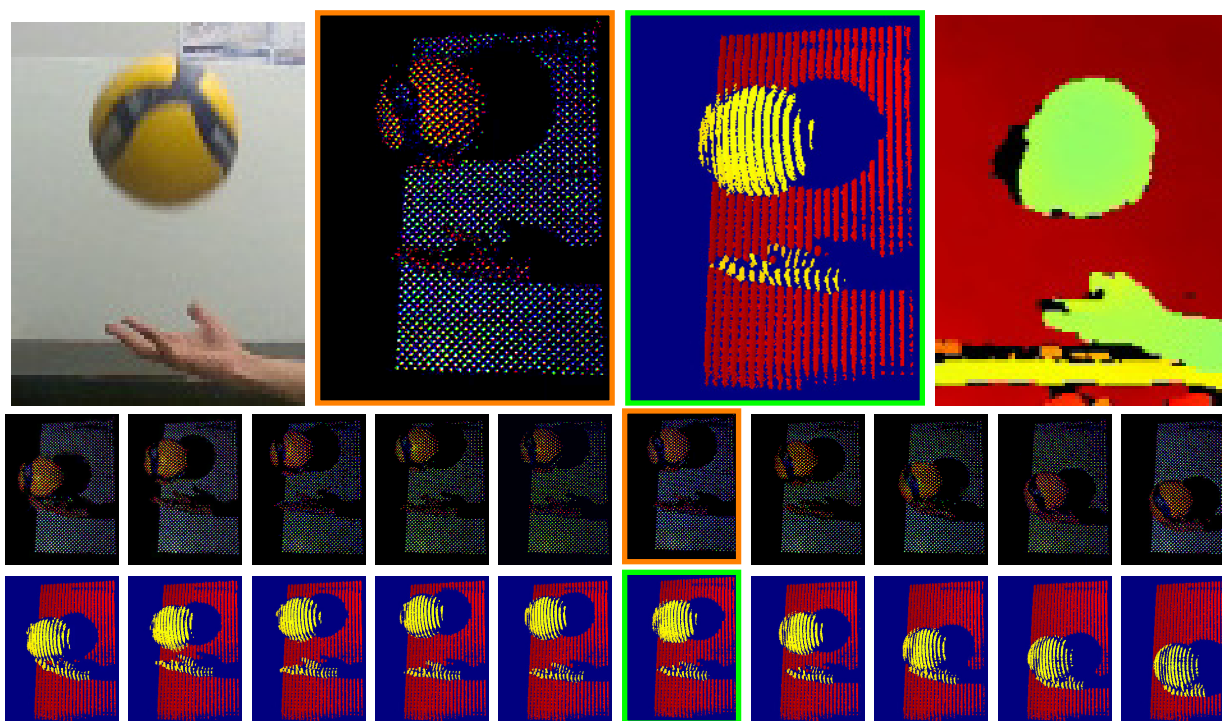


Figure 5.15 Comparison of color and depth detection for dynamic scenes (setup with ball number one) between RealSense D455 (top row, right and left) and our system (middle, using pattern M3L23).

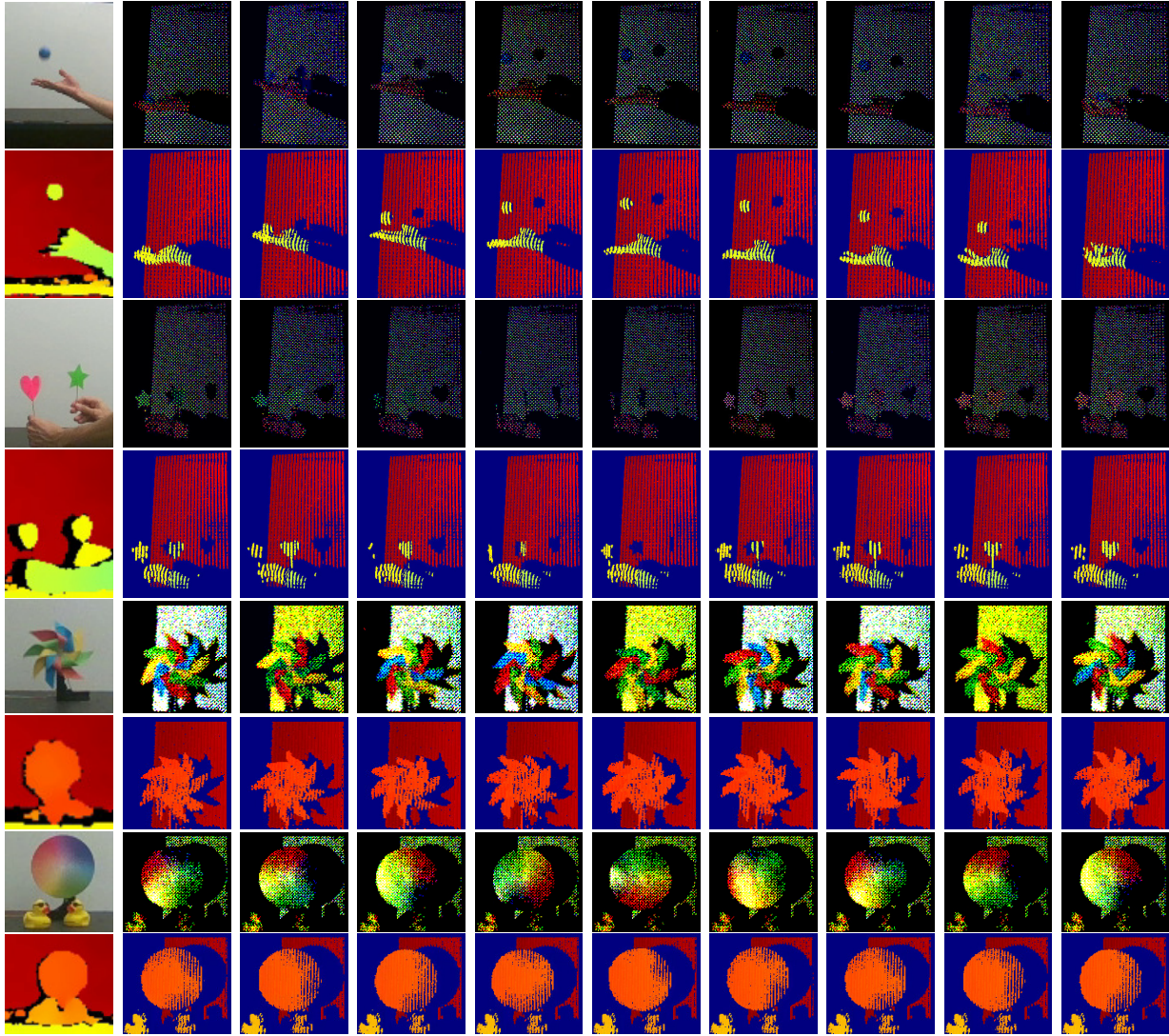


Figure 5.16 Series of images showing color and depth detection for dynamic scenes, scanned in 7.4 to 17.23 ms using the M3L23 pattern. The images in the first column from the left were captured by the RealSense D455 for comparison, which took 33 ms.

## CHAPTER 6 GENERAL DISCUSSION

### 6.1 Features and Advantage of the Proposed Method

#### 6.1.1 Detecting color with monochrome camera

Color information is essential for tasks like segmentation and recognition [30]. Colorization creates a color image from a monochrome or grayscale sensor while maintaining resolution. This method uses external color data from a device [31], user input [32], or a trained neural network incorporating scene-specific colors [33,34]. However, it can be time-consuming and costly.

Compared to color ECs, despite higher bandwidth needs and reduced resolution, color ECs struggle to detect changes in a static or slowly moving environment.

Our approach optimizes bandwidth usage by activating color detection only when necessary and in specific areas. Additionally, it gathers environment data without requiring mechanical movement of the camera or robot, enhancing system reliability. Moreover, our method can leverage the advantages of high-resolution monochrome ECs in low-light conditions, despite the potential for increased noise compared to low-resolution ECs [98].

#### 6.1.2 Real-time depth detection per pixel

In Chapter 5, we discussed how current SL-based methods for depth detection require projecting various patterns (such as binary or coded patterns) to calculate the depth of each pixel. We cannot call it event-based depth detection when it requires projecting several patterns to measure the depth of a single pixel.

Our proposed method detects the depth of each pixel individually, eliminating the need to aggregate all events. It also avoids stereo matching searches and provides depth for upcoming events directly, without requiring complex computations. With our approach, we achieve per-pixel depth detection ten times faster than current methods, such as ESL [69].

#### 6.1.3 Controlling trade-off between speed and details

Methods like X-maps [70] or ESL [69], cannot adjust pattern density for wider fields of view, which limits their ability to balance speed and detail effectively.

Our approach features dynamic projection adjustments to optimize bandwidth and effectively

capture color and depth data, enabling the creation of detailed, colorful 3D maps without compromising spatial clarity. Color information is selectively detected only when needed. Introducing Adaptive Structured Light (ASL) allows us to balance speed and detail: dense, colorful point clouds for narrow fields of view (FOV) or faster, sparser point clouds for wider FOV by projecting shorter patterns—all achievable with our proposed method.

## 6.2 Technical Challenges and Solutions

### 6.2.1 White balance

As described in Chapter 3, adjusting white balance and color correction are crucial for achieving accurate colors in captured images. Generally, white balance can be adjusted before or after taking a photo. Using a DLP projector to generate white light directly impacts white balance because the color temperature or warmth/coolness of the light can modify it. DLP projectors utilize red, green, and blue LEDs, which can present challenges in producing LED-based white light with wideband wavelength RGB LEDs [99–101]. However, since DLP projectors use narrowband LEDs, adjusting white balance is possible by individually changing the current of each LED.

However, detecting color through reflection means that the intensity of the light source and its reflection can influence white balance. For example, the number of received events from reflections of different colored lights may vary based on the object’s distance. To address this issue, one solution could be adjusting the light intensity according to the object’s distance to maintain consistent white balance in varying conditions.

### 6.2.2 Field of view

As mentioned in Chapter 5, we used a horizontal stereo setup, which resulted in our narrow and dense pattern being oriented vertically. This setup provided us with more data above and below the object in front. However, in some projects, having data on the left and right sides of the object in front of the camera is more critical. To address this, one could use a vertical stereo setup by placing the camera on top of the projector and projecting horizontal lines instead of vertical ones. In this configuration, the field of view of the narrow, dense patterns would cover a wider angle in yaw instead of pitch.

### 6.2.3 Black objects

Since the proposed method relies on visible light and its reflection, detecting black objects can be challenging. One solution to this issue is to periodically vary the projector’s intensity for the same pattern. This ensures that events are detected even for black objects in front of the camera.

Alternatively, if color detection is not necessary, near-infrared projectors from DLP can be used. They can project the same pattern with the same configuration without concern for color, enabling depth detection of all objects in the environment regardless of their color or the intensity of the light source.

## 6.3 Potential Commercial Applications

The proposed event-based RGB-D method opens doors to diverse commercial applications, leveraging its unique capabilities in augmented and virtual reality (AR/VR) experiences, content creation and production, and collaborative robotics and automation.

### 6.3.1 Augmented and Virtual Reality (AR/VR) Experiences

Event-based RGB-D technology enhances AR/VR gaming and simulations by providing real-time depth perception and spatial mapping. This enables interactive experiences where users can seamlessly interact with virtual objects and environments. For instance, in gaming, precise depth data allows for accurate object placement and interaction, enhancing realism and user immersion. Simulations benefit from dynamic environment modeling, enabling instant capture and integration of changes in lighting or object positions for a more responsive user experience.

Another potential commercial application in this field is to integrate multiple devices and use one as the leader. By projecting a specific pattern, we could achieve synchronized spatial and temporal perception. Figure 6.1 illustrates a potential setup for this application. In Figure 6.1, all robots equipped with our method can generate spatial perception. One robot can project a pattern to synchronize spatial and temporal perception among all surrounding robots. This capability is enabled by our method, which employs EC and potentially allows for detection of the projected pattern frequency.



Figure 6.1 Combining projection with event-based cameras to obtain synchronized spatial and temporal perception.

### 6.3.2 Content Creation and Production

Event-based RGB-D technology will revolutionize content creation in film, video production, and virtual set design by enabling super-fast real-time 3D object modeling. Filmmakers and designers will be able to capture detailed models of objects with precision using event-driven RGB-D data. This technology will facilitate the rapid integration of virtual elements into live-action scenes and empower designers to create virtual environments with accurate spatial dimensions.

### 6.3.3 Collaborative Robotics and Automation

Robots equipped with this technology gain real-time awareness of their surroundings and can recognize objects, improving their ability to work safely alongside humans in shared spaces. For example, collaborative robots (cobots) can navigate through complex environments without colliding with humans and can adjust their actions based on real-time changes. This technology supports tasks like handling and assembling objects by providing precise depth information, ensuring accurate interaction with tools and components.



## 6.4 General Impact of the Proposed Method

The premise of this work is that combining the strengths of Event-based Cameras (ECs) and Active Structured Light (ASL) can significantly advance RGB-D sensing technology. As the demand for real-time, high-resolution 3D data increases in fields like robotics and 3D reconstruction, our novel approach addresses the limitations of current methods in balancing detail and speed. By utilizing dynamic projection adjustments and a commercial TI LightCrafter 4500 projector with a single monochrome EC, we achieve accurate and vivid point clouds without compromising spatial resolution.

This project develops systems that optimize bandwidth and enhance the adaptability of RGB-D sensing in both static and dynamic environments. The ability to selectively detect color and depth data ensures efficient performance, paving the way for continuous streams of annotated events that provide both color and depth information. This innovation in RGB-D sensing technology not only improves data acquisition speeds but also maintains high accuracy, making it a significant contribution to the advancement of computer vision applications.

We also have implemented our method on the Robot Operating System (ROS), and it is publicly available online as the Event-based RGBD ROS Wrapper [1] on our Git repository at [github.com/MISTLab/event\\_based\\_rgbd\\_ros](https://github.com/MISTLab/event_based_rgbd_ros). This software extends the Prophesee ROS wrapper [129] by enabling online control of camera settings and publishing color and depth-stamped events alongside RGB frames and colorful point clouds on ROS topics. All functionalities are controllable through Graphical User Interfaces (GUI) designed with Qt software.

The provided driver package includes the following nodes:

1. **publisher:** Publishes color and depth stamped events from the Prophesee sensor on ROS topics, aided by the DLP projector.
2. **frame\_generator:** Generates and publishes frames based on published events from the publisher node.
3. **viewer:** Visualizes published frames from the frame\_generator node.
4. **camera\_gui\_node:** Controls camera parameters, manages frame publishing, and calibrates the camera.

The additional features of the provided software are as follows:

1. Includes message types for event and frame data.
2. Offers various services to adjust camera settings, control publishing topics, and manage regions of interest, enhancing the overall flexibility and functionality of the system.

This software package enables researchers and developers to use event-based RGB-D sensing effectively in their ROS projects. It helps process data in real-time, visualize information, and interact with it. These capabilities drive innovation in fields like self-driving robots, identifying objects, and human-robot teamwork. The software's ease of use and powerful features make it a key tool for pushing forward RGB-D sensing technology.

In addition to the aforementioned general impacts, this work will impact on four dimensions:

#### **6.4.1 Academia (People and Knowledge)**

By pioneering the integration of Dynamic Projection and Event-based Camera technologies for RGB-D sensing, this project will advance the academic agenda in computer vision and robotics. It establishes new methodologies for real-time color and depth detection at the pixel level, influencing future research in 3D reconstruction and spatial perception. This novel approach bridges traditional camera-based imaging with dynamic projection techniques, providing insights into efficient data acquisition and processing methods essential for next-generation robotics.

#### **6.4.2 Industry and the Economy**

The innovations developed in this project are essential for advancing RGB-D sensing capabilities in robotics and automation sectors. By achieving real-time, high-resolution color and depth mapping with reduced computational demands, this technology enhances the reliability and efficiency of autonomous systems used in mobile robot applications within industry. These robots require accurate perception of their environment to navigate safely and effectively, which contributes to significant cost savings through increased operational speed and accuracy.

#### **6.4.3 Space Applications**

Integrating advanced RGB-D sensing capabilities into robotic systems holds significant promise for space exploration projects, particularly in lunar missions aimed at creating detailed point clouds. This technology enhances depth perception and color mapping capabilities, promising more accurate and comprehensive mapping of environments, such as lunar surfaces. Such advancements could revolutionize the way we explore and understand celestial bodies, enabling safer navigation and more efficient resource utilization on the Moon and beyond.



#### 6.4.4 Society

Integrating advanced RGB-D sensing into robotic systems significantly enhances their ability to assist in households or aid humans at home. This technology improves depth perception and color mapping, ensuring safer and more efficient operations in dynamic environments. By seamlessly integrating autonomous systems into daily life, it enhances overall safety and efficiency. Robots equipped with these capabilities can navigate complex household environments with precision, ensuring tasks are performed accurately while minimizing potential risks to users. This innovation not only revolutionizes how robots interact with human environments but also encourages trust and acceptance of autonomous technologies in society, accelerating their integration into everyday life.

## CHAPTER 7 CONCLUSION

This study presents a way to create colored point clouds with customizable speed and detail. It allows for depth map creation in difficult settings, such as moving environments or low-light conditions, even when cameras or objects are not moving.

### 7.1 Summary of Works

We introduced a novel method utilizing a DMD projector to generate Active Structured Light (ASL) alongside a monochrome Event-based Camera (EC) for RGB-D sensing. By integrating the strengths of ECs and projection-based techniques, we achieved real-time detection of color and depth for every pixel. Dynamic adjustments in projection optimized bandwidth, facilitating selective color data detection and producing vivid point clouds without compromising spatial resolution. Utilizing a commercial TI LightCrafter 4500 projector and a single monochrome EC, we accomplished frameless RGB-D sensing with reliable results.

The results demonstrate that denser scanning patterns yield denser and more accurate data but cover smaller fields of view, whereas sparser patterns are capable of covering larger areas. By employing different numbers of patterns, our method achieves a scanning time of 7.4 ms, significantly faster than both the RealSense D455 (33 ms) and the ESL method (2 to 5 seconds). This balance between detail and speed effectively enhances speed control and enables color reconstruction.

Our setup achieves color scanning speeds of up to 1.4 KHz and pixel-based depth scanning speeds of up to 4 KHz. This enables a continuous stream of annotated events providing both color and depth information, generating detailed, colorful point clouds. The method demonstrates adaptability in static and dynamic environments, introducing an innovative strategy that balances resolution and acquisition speed. This advancement significantly enhances computer vision applications, spanning fields from robotics to 3D reconstruction.

### 7.2 Limitations

One primary advantage of ECs lies in their rapid response time, typically in the microsecond range. However, with the method introduced here, we must collect events for each color individually, thereby constraining the speed of color detection to the maximum rate at which the DLP projector can switch patterns. Similarly, the speed of depth detection across the entire field of view is restricted by the maximum pattern-switching speed of the DLP projector.

Another limitation of the proposed method arises from its reliance on visible light. This restricts both depth and color detection to materials that are optically straightforward and do not alter or absorb light wavelengths. Materials such as fluorescent substances could pose challenges for detection using this approach.

Because this method operates actively, the detection range may be constrained by the power of the projector. This limitation can impact the accuracy of color detection and, especially, depth detection for darker materials at longer distances. Using a higher-power projector could potentially resolve these issues for long-range scanning.

### 7.3 Future Research

Based on the findings and outcomes of this study, several promising directions for future research have surfaced:

- **Enhancing resolution in color and depth detection:** This could be achieved by projecting denser patterns in targeted regions, possibly integrating a motion detection algorithm during SL light idle periods or employing a secondary camera for concurrent motion detection.
- **Increasing scanning speed:** Achieve higher speeds, albeit with increased bandwidth usage and reduced resolution, by projecting white-colored patterns when using a color EC.
- **Optimizing depth measurement:** Adjust based on the SL wavelength and object surface color by controlling the current of each LED of the projector.
- **Increasing scanning range:** Extend the scanning range by deploying a Near-IR projector instead of capturing color.
- **Investigating the output of the projector:** Explore the color-encoded point cloud generated by the projector in SLAM applications to utilize color information for loop closure detection.

## REFERENCES

- [1] S.-E. Marjani-Bajestani and G. Beltrame, “Event-Based RGB Sensing With Structured Light,” in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2023, pp. 5458–5467. [Online]. Available: <https://doi.org/10.1109/WACV56688.2023.00542>
- [2] X. Huang, Y. Zhang, and Z. Xiong, “High-speed structured light based 3D scanning using an event camera,” *Optics Express*, vol. 29, no. 22, pp. 35 864–35 876, 2021. [Online]. Available: <https://doi.org/10.1364/OE.437944>
- [3] A. Morar, A. Moldoveanu, I. Mocanu, F. Moldoveanu, I. E. Radoi, V. Asavei, A. Gradinaru, and A. Butean, “A Comprehensive Survey of Indoor Localization Methods Based on Computer Vision,” *Sensors*, vol. 20, no. 9, p. 2641, 2020. [Online]. Available: <https://doi.org/10.3390/s20092641>
- [4] E. Garcia-Fidalgo and A. Ortiz, “Vision-based topological mapping and localization methods: A survey,” *Robotics and Autonomous Systems*, vol. 64, pp. 1–20, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0921889014002619>
- [5] M. R. U. Saputra, A. Markham, and N. Trigoni, “Visual SLAM and structure from motion in dynamic environments: A survey,” *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, pp. 1–36, 2018. [Online]. Available: <https://dl.acm.org/doi/10.1145/3177853>
- [6] Z. Kang, J. Yang, Z. Yang, and S. Cheng, “A Review of Techniques for 3D Reconstruction of Indoor Environments,” *ISPRS International Journal of Geo-Information*, vol. 9, no. 5, p. 330, 2020. [Online]. Available: <https://www.mdpi.com/2220-9964/9/5/330>
- [7] Y.-Y. Chuang, “Camera projection models, camera calibration, bundle adjustment,” 2005. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.707.5643&rep=rep1&type=pdf>
- [8] A. G. Marrugo, F. Gao, and S. Zhang, “State-of-the-art active optical techniques for three-dimensional surface metrology: a review,” *J. Opt. Soc. Am. A (JOSA A)*, vol. 37, no. 9, pp. B60–B77, 2020. [Online]. Available: <http://doi.org/10.1364/JOSAA.398644>

- [9] T. Xue, A. Owens, D. Scharstein, M. Goesele, and R. Szeliski, “Multi-frame stereo matching with edges, planes, and superpixels,” *Image and Vision Computing*, vol. 91, p. 103771, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885619300745>
- [10] L. He, G. Wang, and Z. Hu, “Learning depth from single images with deep neural network embedding focal length,” *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4676–4689, 2018. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8360460>
- [11] A. Saxena, S. H. Chung, and A. Y. Ng, “3-d depth reconstruction from a single still image,” *International journal of computer vision*, vol. 76, no. 1, pp. 53–69, 2008. [Online]. Available: <https://link.springer.com/article/10.1007%2Fs11263-007-0071-y>
- [12] J. Wilm, “Real Time Structured Light and Applications,” Ph.D. dissertation, 2016. [Online]. Available: <https://orbit.dtu.dk/en/publications/real-time-structured-light-and-applications>
- [13] F. Willomitzer and G. Häusler, “Single-shot 3D motion picture camera with a dense point cloud,” *Optics express*, vol. 25, no. 19, pp. 23 451–23 464, 2017. [Online]. Available: <http://doi.org/10.1364/OE.25.023451>
- [14] Z. Zhao, F. Gu, P. Xie, H. Cao, and Z. Song, “Miniature 3D Depth Camera for Real-time Reconstruction,” in *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2019, pp. 1769–1776. [Online]. Available: <https://doi.org/10.1109/ROBIO49542.2019.8961795>
- [15] C. Brandli, T. Mantel, M. Hutter, M. Höpfinger, R. Berner, R. Siegwart, and T. Delbruck, “Adaptive pulsed laser line extraction for terrain reconstruction using a dynamic vision sensor,” *Frontiers in neuroscience*, vol. 7, p. 275, 2014. [Online]. Available: <https://doi.org/10.3389/fnins.2013.00275>
- [16] N. Matsuda, O. Cossairt, and M. Gupta, “MC3D: Motion Contrast 3D Scanning,” in *2015 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2015, pp. 1–10. [Online]. Available: <https://doi.org/10.1109/ICCPHOT.2015.7168370>
- [17] G. Rohan, M. Abhishek, H. Yu, and N. V. Thakor, “Depth estimation and object recognition in dark environments using ATIS,” in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE, 2014, pp. 371–376. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7064334>

- [18] M. Litzenberger, C. Posch, D. Bauer, A. N. Belbachir, P. Schon, B. Kohn, and H. Garn, “Embedded vision system for real-time object tracking using an asynchronous transient vision sensor,” in *2006 IEEE 12th Digital Signal Processing Workshop & 4th IEEE Signal Processing Education Workshop*. IEEE, 2006, pp. 173–178. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/4041053>
- [19] L. Steffen, D. Reichard, J. Weinland, J. Kaiser, A. Roennau, and R. Dillmann, “Neuromorphic stereo vision: A survey of bio-inspired sensors and algorithms,” *Frontiers in Neurorobotics*, vol. 13, p. 28, 2019. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnbot.2019.00028/full>
- [20] A. R. Mangalore, C. S. Seelamantula, and C. S. Thakur, “Neuromorphic Fringe Projection Profilometry,” *IEEE Signal Processing Letters*, vol. 27, pp. 1510–1514, 2020. [Online]. Available: <https://doi.org/10.1109/LSP.2020.3016251>
- [21] 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*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Springer, 2016, pp. 349–364. [Online]. Available: [https://doi.org/10.1007/978-3-319-46466-4\\_21](https://doi.org/10.1007/978-3-319-46466-4_21)
- [22] 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, 2017. [Online]. Available: <https://doi.org/10.1109/LRA.2016.2645143>
- [23] 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. [Online]. Available: <https://doi.org/10.1007/s11263-017-1050-6>
- [24] J. Hidalgo-Carrió, D. Gehrig, and D. Scaramuzza, “Learning Monocular Dense Depth from Events,” in *2020 International Conference on 3D Vision (3DV)*. IEEE, 2020. [Online]. Available: <http://doi.org/10.1109/3DV50981.2020.00063>
- [25] LUCID Vision Labs, “Helios Time-of-Flight (ToF) Camera,” <https://thinklucid.com/helios-time-of-flight-tof-camera/>.
- [26] M. Gupta and S. K. Nayar, “Micro phase shifting,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 813–820. [Online]. Available: <https://doi.org/10.1109/CVPR.2012.6247753>

- [27] K. Wu, J. Tan, H. Xia, and C. Liu, “An exposure fusion-based structured light approach for the 3D measurement of a specular surface,” *IEEE Sensors Journal*, 2020. [Online]. Available: <https://doi.org/10.1109/JSEN.2020.3027317>
- [28] S.-K. Tin, J. Ye, M. Nezamabadi, and C. Chen, “3D reconstruction of mirror-type objects using efficient ray coding,” in *2016 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2016, pp. 1–11. [Online]. Available: <https://doi.org/10.1109/ICCPHOT.2016.7492867>
- [29] Y. Li, C. Xu, and L. Liu, “Learning from General Diffuse Surfaces: An Event-driven Approach for High Dynamic Range Industrial Optical Measurement,” *Optics & Laser Technology*, vol. 177, p. 111183, 2024. [Online]. Available: <https://doi.org/10.1016/j.optlastec.2024.111183>
- [30] A. Trémeau, S. Tominaga, and K. Plataniotis, “Color in image and video processing: most recent trends and future research directions,” *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–26, 2008. [Online]. Available: <https://jivp-urasipjournals.springeropen.com/articles/10.1155/2008/581371>
- [31] H. W. Jang and Y. J. Jung, “Deep Color Transfer for Color-Plus-Mono Dual Cameras,” *Sensors*, vol. 20, no. 9, p. 2743, 2020. [Online]. Available: <https://doi.org/10.3390/s20092743>
- [32] A. Levin, D. Lischinski, and Y. Weiss, “Colorization using optimization,” in *ACM SIGGRAPH 2004 Papers*, 2004, pp. 689–694. [Online]. Available: <https://doi.org/10.1145/1186562.1015780>
- [33] R. Zhang, P. Isola, and A. A. Efros, “Colorful image colorization,” in *European conference on computer vision*. Springer, 2016, pp. 649–666. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-319-46487-9\\_40](https://link.springer.com/chapter/10.1007/978-3-319-46487-9_40)
- [34] H. Cohen Duwek and E. Ezra Tsur, “Colorful image reconstruction from neuromorphic event cameras with biologically inspired deep color fusion neural networks,” *Bioinspiration & Biomimetics*, vol. 19, no. 3, p. 036001, mar 2024. [Online]. Available: <https://doi.org/10.1088/1748-3190/ad2a7c>
- [35] C. Li, C. Brandli, R. Berner, H. Liu, M. Yang, S.-C. Liu, and T. Delbruck, “Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor,” in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2015, pp. 718–721. [Online]. Available: <https://doi.org/10.1109/ISCAS.2015.7168734>

- [36] D. P. Moeys, F. Corradi, C. Li, S. A. Bamford, L. Longinotti, F. F. Voigt, S. Berry, G. Taverni, F. Helmchen, and T. Delbruck, “A Sensitive Dynamic and Active Pixel Vision Sensor for Color or Neural Imaging Applications,” *IEEE transactions on biomedical circuits and systems*, vol. 12, no. 1, pp. 123–136, 2017. [Online]. Available: <https://doi.org/10.1109/TBCAS.2017.2759783>
- [37] D. P. Moeys, C. Li, J. N. Martel, S. Bamford, L. Longinotti, V. Motsnyi, D. S. S. Bello, and T. Delbruck, “Color temporal contrast sensitivity in dynamic vision sensors,” in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2017, pp. 1–4. [Online]. Available: <https://doi.org/10.1109/ISCAS.2017.8050412>
- [38] B. E. Bayer, “Color imaging array,” *United States Patent 3,971,065*, 1976.
- [39] D. Khashabi, S. Nowozin, J. Jancsary, and A. W. Fitzgibbon, “Joint Demosaicing and Denoising via Learned Nonparametric Random Fields,” *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 4968–4981, 2014. [Online]. Available: <https://doi.org/10.1109/TIP.2014.2359774>
- [40] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, “Color image processing pipeline,” *IEEE Signal Processing Magazine*, vol. 22, no. 1, pp. 34–43, 2005. [Online]. Available: <https://doi.org/10.1109/MSP.2005.1407713>
- [41] A. Marcireau, S.-H. Ieng, C. Simon-Chane, and R. B. Benosman, “Event-Based Color Segmentation With a High Dynamic Range Sensor,” *Frontiers in neuroscience*, vol. 12, p. 135, 2018. [Online]. Available: <https://doi.org/10.3389/fnins.2018.00135>
- [42] M. Zollhöfer, P. Stotko, A. Görlitz, C. Theobalt, M. Nießner, R. Klein, and A. Kolb, “State of the Art on 3D Reconstruction with RGB-D Cameras,” in *Computer graphics forum*, vol. 37, no. 2. Wiley Online Library, 2018, pp. 625–652. [Online]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1111/cgf.13386>
- [43] S.-H. Ieng, J. Carneiro, M. Osswald, and R. Benosman, “Neuromorphic Event-Based Generalized Time-Based Stereovision,” *Frontiers in Neuroscience*, vol. 12, p. 442, 2018. [Online]. Available: <https://doi.org/10.3389/fnins.2018.00442>
- [44] M. M. Ibrahim, Q. Liu, R. Khan, J. Yang, E. Adeli, and Y. Yang, “Depth map artefacts reduction: a review,” *IET Image Processing*, vol. 14, no. 12, pp. 2630–2644, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9222809>
- [45] N. Smolyanskiy, A. Kamenev, and S. Birchfield, “On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural



- network approach,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1007–1015. [Online]. Available: <https://ieeexplore.ieee.org/document/8575301>
- [46] 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. [Online]. Available: <https://doi.org/10.1109/ICRA40945.2020.9197341>
- [47] A. Glover and C. Bartolozzi, “Robust visual tracking with a freely-moving event camera,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 3769–3776. [Online]. Available: <https://doi.org/10.1109/IROS.2017.8206226>
- [48] M. Gehrig, S. B. Shrestha, D. Mouritzen, and D. Scaramuzza, “Event-Based Angular Velocity Regression with Spiking Networks,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 4195–4202. [Online]. Available: <https://doi.org/10.1109/ICRA40945.2020.9197133>
- [49] A. Mitrokhin, C. Fermüller, C. Parameshwara, and Y. Aloimonos, “Event-Based Moving Object Detection and Tracking,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1–9. [Online]. Available: <https://doi.org/10.1109/IROS.2018.8593805>
- [50] Z. Wang, K. Chaney, and K. Daniilidis, “EvAC3D: From Event-Based Apparent Contours to 3D Models via Continuous Visual Hulls,” in *European Conference on Computer Vision*. Springer Nature Switzerland, 2022, pp. 284–299. [Online]. Available: [https://doi.org/10.1007/978-3-031-20071-7\\_17](https://doi.org/10.1007/978-3-031-20071-7_17)
- [51] I. Hwang, J. Kim, and Y. M. Kim, “Ev-NeRF: Event based Neural Radiance Field,” in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023, pp. 837–847. [Online]. Available: <https://doi.org/10.1109/WACV56688.2023.00090>
- [52] Q. Ma, D. P. Paudel, A. Chhatkuli, and L. Van Gool, “Deformable Neural Radiance Fields using RGB and Event Cameras,” in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 3590–3600. [Online]. Available: <https://doi.org/10.1109/ICCV51070.2023.00332>

- [53] W. F. Low and G. H. Lee, “Robust e-NeRF: NeRF from Sparse & Noisy Events under Non-Uniform Motion,” in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 18 335–18 346. [Online]. Available: <https://doi.org/10.1109/ICCV51070.2023.01681>
- [54] Y. Qi, L. Zhu, Y. Zhang, and J. Li, “E2NeRF: Event Enhanced Neural Radiance Fields from Blurry Images,” in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 13 254–13 264. [Online]. Available: <https://doi.org/10.1109/ICCV51070.2023.01219>
- [55] V. Rudnev, M. Elgharib, C. Theobalt, and V. Golyanik, “EventNeRF: Neural Radiance Fields from a Single Colour Event Camera,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 4992–5002. [Online]. Available: <https://doi.org/10.1109/CVPR52729.2023.00483>
- [56] J. Han, Y. Asano, B. Shi, Y. Zheng, and I. Sato, “High-fidelity Event-Radiance Recovery via Transient Event Frequency,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 20 616–20 625. [Online]. Available: <https://doi.org/10.1109/CVPR52729.2023.01975>
- [57] B. Yu, J. Ren, J. Han, F. Wang, J. Liang, and B. Shi, “EventPS: Real-Time Photometric Stereo Using an Event Camera,” in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 9602–9611. [Online]. Available: [https://openaccess.thecvf.com/content/CVPR2024/html/Yu\\_EventPS\\_Real-Time\\_Photometric\\_Stereo\\_Using\\_an\\_Event\\_Camera\\_CVPR\\_2024\\_paper.html](https://openaccess.thecvf.com/content/CVPR2024/html/Yu_EventPS_Real-Time_Photometric_Stereo_Using_an_Event_Camera_CVPR_2024_paper.html)
- [58] M. Muglikar, L. Bauersfeld, D. P. Moeys, and D. Scaramuzza, “Event-Based Shape from Polarization,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 1547–1556. [Online]. Available: <https://doi.org/10.1109/CVPR52729.2023.00155>
- [59] F. Eibensteiner, H. G. Brachtendorf, and J. Scharinger, “Event-driven stereo vision algorithm based on silicon retina sensors,” in *2017 27th International Conference Radioelektronika (RADIOELEKTRONIKA)*. IEEE, 2017, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/RADIOELEK.2017.7937602>
- [60] J. N. Martel, J. Müller, J. Conradt, and Y. Sandamirskaya, “An Active Approach to Solving the Stereo Matching Problem using Event-Based Sensors,” in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/ISCAS.2018.8351411>

- [61] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, “Semi-dense 3D Reconstruction with a Stereo Event Camera,” in *European Conference on Computer Vision (ECCV)*. Springer International Publishing, 2018, pp. 235–251. [Online]. Available: [https://doi.org/10.1007/978-3-030-01246-5\\_15](https://doi.org/10.1007/978-3-030-01246-5_15)
- [62] M. J. Domínguez-Morales, Á. Jiménez-Fernández, 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. [Online]. Available: <https://doi.org/10.1109/ACCESS.2019.2943160>
- [63] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, “The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception,” *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, 2018. [Online]. Available: <https://doi.org/10.1109/LRA.2018.2800793>
- [64] E. Piatkowska, A. N. Belbachir, and M. Gelautz, “Cooperative and asynchronous stereo vision for dynamic vision sensors,” *Measurement Science and Technology*, vol. 25, no. 5, p. 055108, 2014. [Online]. Available: <https://doi.org/10.1088/0957-0233/25/5/055108>
- [65] E. Piatkowska, J. Kogler, N. Belbachir, and M. Gelautz, “Improved Cooperative Stereo Matching for Dynamic Vision Sensors with Ground Truth Evaluation,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 53–60. [Online]. Available: <https://doi.org/10.1109/CVPRW.2017.51>
- [66] C. Xiao, X. Chen, J. Xi, Z. Li, and B. He, “Speckle-Projection-Based High-Speed 3D Reconstruction Using Event Cameras,” in *2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2023, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/CISP-BMEI60920.2023.10373309>
- [67] L. Zhenglei, H. Bowen, X. Chunyuan, C. Xiaobo, and X. Juntong, “Laser Scanning Measurement based on Event Cameras,” in *2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2023, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/CISP-BMEI60920.2023.10373321>
- [68] D. Weikersdorfer, D. B. Adrian, D. Cremers, and J. Conradt, “Event-based 3D SLAM with a depth-augmented dynamic vision sensor,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 359–364. [Online]. Available: <https://doi.org/10.1109/ICRA.2014.6906882>

- [69] M. Muglikar, G. Gallego, and D. Scaramuzza, “ESL: Event-based Structured Light,” in *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021, pp. 1165–1174. [Online]. Available: <https://doi.org/10.1109/3DV53792.2021.00124>
- [70] W. Morgenstern, N. Gard, S. Baumann, A. Hilsmann, and P. Eisert, “X-maps: Direct Depth Lookup for Event-based Structured Light Systems,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 4006–4014. [Online]. Available: <https://doi.org/10.1109/CVPRW59228.2023.00418>
- [71] H. Wang, T. Liu, C. He, C. Li, J. Liu, and L. Yu, “Enhancing Event-based Structured Light Imaging with a Single Frame,” in *2022 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2022, pp. 1–7. [Online]. Available: <https://doi.org/10.1109/MFI55806.2022.9913845>
- [72] X. Lu, L. Sun, D. Gu, Z. Xu, and K. Wang, “SGE: Structured light system based on gray code with an event camera,” *arXiv preprint arXiv:2403.07326*, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2403.07326>
- [73] J. Yu, N. Gao, Z. Zhang, and Z. Meng, “High Sensitivity Fringe Projection Profilometry Combining Optimal Fringe Frequency and Optimal Fringe Direction,” *Optics and Lasers in Engineering*, vol. 129, p. 106068, 2020. [Online]. Available: <http://doi.org/10.1016/j.optlaseng.2020.106068>
- [74] T. Leroux, S.-H. Ieng, and R. Benosman, “Event-based structured light for depth reconstruction using frequency tagged light patterns,” *arXiv preprint arXiv:1811.10771*, 2018. [Online]. Available: <https://doi.org/10.48550/arXiv.1811.10771>
- [75] Y. Li, H. Jiang, C. Xu, and L. Liu, “Event-driven Fringe Projection Structured Light 3D Reconstruction based on Time-frequency Analysis,” *IEEE Sensors Journal*, 2024. [Online]. Available: <https://doi.org/10.1109/JSEN.2024.3349432>
- [76] Intel RealSense. Stereo depth solutions. [Online]. Available: <https://www.intelrealsense.com/stereo-depth/>
- [77] Orbbec, Astra Series. Intelligent computing for everyone, everywhere. [Online]. Available: <https://orbbec3d.com/product-astra-pro/>
- [78] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis *et al.*, “Event-based vision: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9138762>

- [79] J. Barrios-Avilés, T. Iakymchuk, J. Samaniego, L. D. Medus, and A. Rosado-Muñoz, “Movement detection with event-based cameras: Comparison with frame-based cameras in robot object tracking using powerlink communication,” *Electronics*, vol. 7, no. 11, p. 304, 2018. [Online]. Available: <https://www.mdpi.com/2079-9292/7/11/304>
- [80] I. Alzugaray and M. Chli, “Asynchronous corner detection and tracking for event cameras in real time,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3177–3184, 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8392795>
- [81] 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. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2017/html/Zhu\\_Event-Based\\_Visual\\_Inertial\\_CVPR\\_2017\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2017/html/Zhu_Event-Based_Visual_Inertial_CVPR_2017_paper.html)
- [82] 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. [Online]. Available: <https://ieeexplore.ieee.org/document/8432102>
- [83] C. Reinbacher, G. Munda, and T. Pock, “Real-time panoramic tracking for event cameras,” in *2017 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2017, pp. 1–9. [Online]. Available: <https://ieeexplore.ieee.org/document/7951488>
- [84] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, “Hats: Histograms of averaged time surfaces for robust event-based object classification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1731–1740. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Sironi\\_HATS\\_Histograms\\_of\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Sironi_HATS_Histograms_of_CVPR_2018_paper.html)
- [85] G. Taverni, D. P. Moeys, C. Li, C. Cavaco, V. Motsnyi, D. S. S. Bello, and T. Delbruck, “Front and back illuminated dynamic and active pixel vision sensors comparison,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 5, pp. 677–681, 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8334288>
- [86] M. Muglikar, D. P. Moeys, and D. Scaramuzza, “Event guided depth sensing,” in *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021, pp. 385–393. [Online]. Available: <https://ieeexplore.ieee.org/document/9665844>
- [87] G. Munda, C. Reinbacher, and T. Pock, “Real-time intensity-image reconstruction for event cameras using manifold regularisation,” *International Journal of*

- Computer Vision*, vol. 126, no. 12, pp. 1381–1393, 2018. [Online]. Available: <https://link.springer.com/article/10.1007/s11263-018-1106-2>
- [88] C. Scheerlinck, N. Barnes, and R. Mahony, “Continuous-time intensity estimation using event cameras,” in *Asian Conference on Computer Vision*. Springer, 2018, pp. 308–324. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-030-20873-8\\_20](https://link.springer.com/chapter/10.1007/978-3-030-20873-8_20)
- [89] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, “Events-to-video: Bringing modern computer vision to event cameras,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3857–3866. [Online]. Available: [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Rebecq\\_Events-To-Video\\_Bringing\\_Modern\\_Computer\\_Vision\\_to\\_Event\\_Cameras\\_CVPR\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2019/html/Rebecq_Events-To-Video_Bringing_Modern_Computer_Vision_to_Event_Cameras_CVPR_2019_paper.html)
- [90] C. Scheerlinck, N. Barnes, and R. Mahony, “Asynchronous spatial image convolutions for event cameras,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 816–822, 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8613800>
- [91] 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/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6820–6829. [Online]. Available: [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Pan\\_Bringing\\_a\\_Blurry\\_Frame\\_Alive\\_at\\_High\\_Frame-Rate\\_With\\_an\\_CVPR\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2019/html/Pan_Bringing_a_Blurry_Frame_Alive_at_High_Frame-Rate_With_an_CVPR_2019_paper.html)
- [92] C. Haoyu, T. Minggui, S. Boxin, W. YIzhou, and H. Tiejun, “Learning to deblur and generate high frame rate video with an event camera,” *arXiv preprint arXiv:2003.00847*, 2020. [Online]. Available: <https://arxiv.org/abs/2003.00847>
- [93] N. Messikommer, S. Georgoulis, D. Gehrig, S. Tulyakov, J. Erbach, A. Bochicchio, Y. Li, and D. Scaramuzza, “Multi-bracket high dynamic range imaging with event cameras,” *arXiv preprint arXiv:2203.06622*, 2022. [Online]. Available: <https://arxiv.org/abs/2203.06622>
- [94] 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. [Online]. Available: <https://ieeexplore.ieee.org/document/8946715>

- [95] L. Pan, R. Hartley, C. Scheerlinck, M. Liu, X. Yu, and Y. Dai, “High frame rate video reconstruction based on an event camera,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9252186>
- [96] M. Mostafavi, L. Wang, and K.-J. Yoon, “Learning to reconstruct hdr images from events, with applications to depth and flow prediction,” *International Journal of Computer Vision*, vol. 129, no. 4, pp. 900–920, 2021. [Online]. Available: <https://link.springer.com/article/10.1007/s11263-020-01410-2>
- [97] C. Scheerlinck, H. Rebecq, T. Stoffregen, N. Barnes, R. Mahony, and D. Scaramuzza, “Ced: Color event camera dataset,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0. [Online]. Available: [https://openaccess.thecvf.com/content\\_CVPRW\\_2019/html/EventVision/Scheerlinck\\_CED\\_Color\\_Event\\_Camera\\_Dataset\\_CVPRW\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPRW_2019/html/EventVision/Scheerlinck_CED_Color_Event_Camera_Dataset_CVPRW_2019_paper.html)
- [98] D. Gehrig and D. Scaramuzza, “Are high-resolution event cameras really needed?” *arXiv preprint arXiv:2203.14672*, 2022. [Online]. Available: <https://arxiv.org/abs/2203.14672>
- [99] S. Muthu, F. J. Schuurmans, and M. D. Pashley, “Red, green, and blue led based white light generation: issues and control,” in *Conference Record of the 2002 IEEE Industry Applications Conference. 37th IAS Annual Meeting (Cat. No. 02CH37344)*, vol. 1. IEEE, 2002, pp. 327–333. [Online]. Available: <https://ieeexplore.ieee.org/document/1044108>
- [100] S. Muthu and J. Gaines, “Red, green and blue led-based white light source: implementation challenges and control design,” in *38th IAS Annual Meeting on Conference Record of the Industry Applications Conference, 2003.*, vol. 1. IEEE, 2003, pp. 515–522. [Online]. Available: <https://ieeexplore.ieee.org/document/1257549>
- [101] A. David and L. A. Whitehead, “Led-based white light,” *Comptes Rendus Physique*, vol. 19, no. 3, pp. 169–181, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S163107051830029X>
- [102] M. Pharr, W. Jakob, and G. Humphreys, *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016. [Online]. Available: <https://www.sciencedirect.com/book/9780128006450/physically-based-rendering>

- [103] W. Choi, H. S. Park, and C.-M. Kyung, “Color reproduction pipeline for an rgbw color filter array sensor,” *Optics Express*, vol. 28, no. 10, pp. 15 678–15 690, 2020. [Online]. Available: <https://opg.optica.org/oe/fulltext.cfm?uri=oe-28-10-15678&id=431622>
- [104] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff *et al.*, “5.10 a 1280×720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86  $\mu\text{m}$  pixels, 1.066 gepts readout, programmable event-rate controller and compressive data-formatting pipeline,” in *2020 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 2020, pp. 112–114. [Online]. Available: <https://ieeexplore.ieee.org/document/9063149>
- [105] T. Delbruck, R. Graca, and M. Paluch, “Feedback control of event cameras,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1324–1332. [Online]. Available: [https://openaccess.thecvf.com/content/CVPR2021W/EventVision/html/Delbruck\\_Feedback\\_Control\\_of\\_Event\\_Cameras\\_CVPRW\\_2021\\_paper.html](https://openaccess.thecvf.com/content/CVPR2021W/EventVision/html/Delbruck_Feedback_Control_of_Event_Cameras_CVPRW_2021_paper.html)
- [106] H. Levy, “Determining local depth from structured light using a regular dot grid,” *Technical Disclosure Commons*, 2019. [Online]. Available: [https://www.tdcommons.org/dpubs\\_series/2536/](https://www.tdcommons.org/dpubs_series/2536/)
- [107] Y. Wang, X. Chen, J. Tao, K. Wang, and M. Ma, “Accurate feature detection for out-of-focus camera calibration,” *Applied optics*, vol. 55, no. 28, pp. 7964–7971, 2016. [Online]. Available: <https://opg.optica.org/ao/fulltext.cfm?uri=ao-55-28-7964&id=350405>
- [108] S. Van der Jeught and J. J. Dirckx, “Real-time structured light profilometry: a review,” *Optics and Lasers in Engineering*, vol. 87, pp. 18–31, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0143816616000166>
- [109] G. G. Guilbault, *Practical fluorescence*. CRC Press, 2020.
- [110] J. F. Blinn, “Models of light reflection for computer synthesized pictures,” in *Proceedings of the 4th annual conference on Computer graphics and interactive techniques*, 1977, pp. 192–198. [Online]. Available: <https://dl.acm.org/doi/abs/10.1145/563858.563893>
- [111] “RoboCup Federation.” [Online]. Available: <https://www.robocup.org/domains/1>
- [112] Stanford Artificial Intelligence Laboratory et al., “Robotic operating system.” [Online]. Available: <https://www.ros.org>



- [113] N. J. Sanket, C. M. Parameshwara, C. D. Singh, A. V. Kuruttukulam, C. Fermüller, D. Scaramuzza, and Y. Aloimonos, “EVDodgenet: Deep Dynamic Obstacle Dodging with Event Cameras,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 651–10 657. [Online]. Available: <https://ieeexplore.ieee.org/document/9196877>
- [114] 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. [Online]. Available: <https://ieeexplore.ieee.org/document/6942940>
- [115] T. Delbruck and M. Lang, “Robotic goalie with 3 ms reaction time at 4% CPU load using event-based dynamic vision sensor,” *Frontiers in neuroscience*, vol. 7, p. 223, 2013.
- [116] “The iniVation’s documentation page.” [Online]. Available: <https://docs.inivation.com/>
- [117] “The Color-DAVIS346 event-based camera specifications.” [Online]. Available: <https://inivation.com/wp-content/uploads/2022/10/2022-09-iniVation-devices-Specifications.pdf>
- [118] “The Khadas VIM3 single-board computer specifications.” [Online]. Available: <https://www.khadas.com/vim3>
- [119] J. Geng, “DLP-based structured light 3D imaging technologies and applications,” in *Emerging Digital Micromirror Device Based Systems and Applications III*, M. R. Douglass and P. I. Oden, Eds., vol. 7932, International Society for Optics and Photonics. SPIE, 2011, p. 79320B. [Online]. Available: <https://doi.org/10.1117/12.873125>
- [120] M. Gupta, Q. Yin, and S. K. Nayar, “Structured light in sunlight,” in *2013 IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 545–552. [Online]. Available: <https://doi.org/10.1109/ICCV.2013.73>
- [121] “Prophesee, Evaluation Kit.” [Online]. Available: <https://www.prophesee.ai/event-based-evk/>
- [122] “Texas Instrument, Digital Light Processing LightCrafter4500.” [Online]. Available: <https://www.ti.com/tool/DLPLCR4500EVM>

- [123] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000. [Online]. Available: <https://github.com/opencv>
- [124] A. Kaehler and G. Bradski, *Learning OpenCV 3*. O’Reilly Media, 2017, iISBN-10: 1491937998. [Online]. Available: <https://www.oreilly.com/library/view/learning-opencv-3/9781491937983/>
- [125] H. Luo, J. Xu, N. H. Binh, S. Liu, C. Zhang, and K. Chen, “A simple calibration procedure for structured light system,” *Optics and Lasers in Engineering*, vol. 57, pp. 6–12, 2014. [Online]. Available: <http://doi.org/10.1016/j.optlaseng.2014.01.010>
- [126] G. Wang, C. Feng, X. Hu, and H. Yang, “Temporal Matrices Mapping Based Calibration Method for Event-Driven Structured Light Systems,” *IEEE Sensors Journal*, 2020. [Online]. Available: <https://doi.org/10.1109/JSEN.2020.3016833>
- [127] M. Allen, D. Poggiali, K. Whitaker, T. R. Marshall, J. van Langen, and R. A. Kievit, “Raincloud plots: a multi-platform tool for robust data visualization,” *Wellcome open research*, vol. 4, 2019. [Online]. Available: <https://doi.org/10.12688/wellcomeopenres.15191.2>
- [128] JASP Team, “JASP (Version 0.18.3),” Computer software, 2024. [Online]. Available: <https://jasp-stats.org/>
- [129] Prophesee AI. (2022) Prophesee ROS Wrapper. [Online]. Available: [https://github.com/prophesee-ai/prophesee\\_ros\\_wrapper](https://github.com/prophesee-ai/prophesee_ros_wrapper)