

# A Stereo Vision Method for Low Illumination Based on Passive Free-Running SPAD

Jiamu Fu, Ruizhi Chen, Dong Li, Jin Hu, Xiayu Wang, Rui Ma, and Zhangming Zhu

**Abstract**— Single-photon avalanche diodes (SPADs) acquiring images without active illumination in a free-running mode is known as passive free-running SPAD (PF-SPAD) imaging, demonstrating high adaptability in low-light scenes. Passive imaging requires less resource consumption than the direct time-of-flight (dToF) method. However, stereo vision, which typically utilizes digital cameras as light-sensitive devices, fails to function effectively in low-light environments due to quantization and read noise. In this paper, the application of PF-SPAD to stereo vision is proposed to extend the application of binocular vision ranging in environments with low illumination. First, a binocular vision ranging platform based on PF-SPAD is built. Then, we simplify the camera calibration method because the SPAD array pixels are too few to use the conventional camera calibration method. Moreover, because of the number of pixels and noise in the PF-SPAD image, an image preprocessing method is designed, the local block-matching (BM) stereo algorithm is further improved, and various scenes are built to verify the robustness of the stereo matching algorithm. Finally, the adaptability of the proposed platform in low illumination, the coplanarity of the left and right imaging planes, and the adaptability of the baseline-adjustable design for scenarios with different distances are verified by comparing the measured and actual distances of different experimental scenes.

**Index Terms**— Camera calibration, passive imaging, single-photon avalanche diodes (SPADs), stereo matching, stereo vision.

## I. INTRODUCTION

3-D imaging can provide depth information in addition to the planar view limited to the length and width of 2-D imaging. With the continuous development of sensor technology and measurement methods [1], 3-D imaging is widely used in intelligent robotics, autonomous driving, industrial manufacturing, aerospace, and other fields.

Different 3-D imaging methods differ in their observation

This work was supported in part by the National Science and Technology Major Project(2021ZD0114403) and in part by the National Natural Science Foundation of China under Grant (62204181). (Corresponding author: Dong Li).

Jiamu Fu is with the Faculty of Integrated Circuit, Xidian University, Xi'an 710071, China.

Ruizhi Chen is with the School of Telecommunications Engineering, Xidian University, Xi'an 710071, China.

Dong Li, Jin Hu, Xiayu Wang, Rui Ma, and Zhangming Zhu are with the Key Laboratory of Analog Integrated Circuits and Systems (Ministry of Education), School of Integrated Circuits, Xidian University, Xi'an 710071, China. (e-mail: lidong@xidian.edu.cn).

scenes and constraints. The stereo vision method uses two cameras to capture the same scene simultaneously, and its measurement theory is well-established [2]. Stereo vision is simple and inexpensive and can measure long distances by adjusting the baseline length [3]. However, currently, stereo vision mainly uses digital cameras as light-sensitive devices. At low brightness levels, noise caused by analog-to-digital conversion (ADC) affects the imaging quality of shaded areas and reduces the practical texture information [4], [5], [6]. Therefore, under conditions of insufficient illumination, the stereo vision method has difficulty matching the corresponding image blocks of two cameras, making the depth measurement inaccurate. Industrialized stereoscopic depth cameras use projectors to make up for the shortcomings of digital cameras in low illumination, where the left and right cameras can't see without a proper light source [7]. Projectors can consume considerable power, sometimes up to several watts [8]. Direct time-of-flight (dToF) 2-D pixel arrays typically use single-photon avalanche diodes (SPADs) as receivers, which allow for high frame rates and accurate measurements at moderate distances [9], [10]. Sometimes, the power consumption of dToF systems is mainly from the laser diodes (LDs) and their driving circuits [9], [11]. The formation of a mismatch between a diffuse scattering of the laser light and the SPAD array that is partially operational at a given moment causes a loss of optical illumination power [12]. In addition, the bright background light environment causes an increase in shot noise, and the extended measuring range requires more laser energy to enhance the SNR [11], [13], [14]. Recently, some work has used SPAD to acquire images without active illumination in a free-running mode, a process known as passive free-running SPAD (PF-SPAD) imaging. PF-SPAD imaging does not use an ADC, and in low-light and high-speed scenarios, PF-SPAD exhibits strong adaptability [4], [5], [6], [15], [16]. 3.2 Megapixel SPAD arrays have been fabricated with the potential to be used in various machine vision applications [17].

In this paper, we propose to use PF-SPAD for binocular vision ranging. Since PF-SPAD does not use lasers, our method saves resource consumption compared to the dToF scheme. It can also extend the application of binocular vision ranging in environments with insufficient illumination as an alternative in similar scenarios.

The main contributions of this article are summarized as follows.

- 1) We build a binocular vision ranging platform and

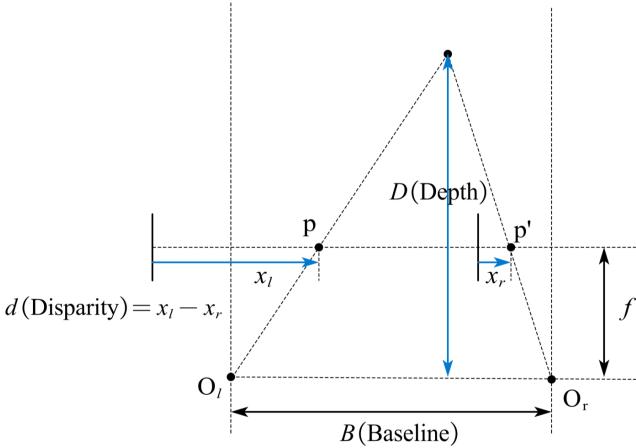


Fig. 1. Geometric relation between disparity and depth.

implement depth feature extraction based on PF-SPAD. In this platform, the hardware is responsible for configuring the working mode and the acquisition, caching, and transmission of PF-SPAD images, and the software is responsible for the selection of configuration files and the display, processing, and storage of PF-SPAD images.

2) Based on the SPAD array with too few pixels to apply a conventional camera calibration method, we simplify the camera calibration method and derive the solution formula for the parameter  $f$  to be calibrated. We also build a calibration experiment scene and give the calculation results of the parameter  $f$  when using different calibration objects.

3) For PF-SPAD images, we design a preprocessing method to improve the SNR and compare it with the built-in preprocessing method of OpenCV to further improve the stereo matching algorithm and calculate the disparity. We verify the adaptability of the algorithm through various scenes.

The rest of the paper is organized as follows. Section II gives the binocular vision ranging platform's hardware design and software processing based on PF-SPAD. Section III compares the measured distance with the ground truth in different scenes to evaluate the measurement accuracy and precision of the proposed platform and to verify the adaptability to different illumination conditions and the coplanarity of the left and right imaging planes, introduces a baseline-adjustable design to adapt to different distance scenarios, and compares the proposed platform with the state-of-the-art, including digital camera-based binocular vision ranging platforms and dToF light detection and ranging (LiDAR) based on SPAD array. Section IV summarizes the work of this paper.

## II. IMPLEMENTATION

### A. Platform Overview

Fig. 1 shows the geometric relation in stereo vision.  $O_l$  and  $O_r$  are the optical centers of the two lenses.  $x_r$  and  $x_l$  are the coordinates of the corresponding points  $p$  and  $p'$ , respectively, on the imaging planes of the two cameras.  $f$  is the focal length in pixels, part of the camera intrinsic matrix.  $D$  is the depth information to be calculated, and  $B$  is the baseline length.

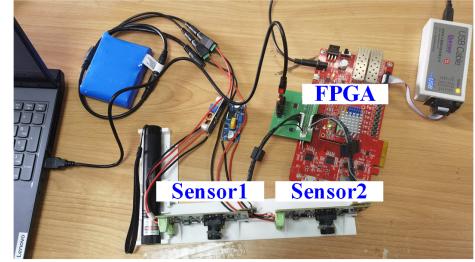


Fig. 2. Binocular vision ranging platform overview. (a) Hardware setup. (b) Platform system block diagram.

To measure the depth of a spatial point, we need to convert the positional difference between the corresponding points captured by the left and right cameras into a depth  $D$  using the geometrical relation in Fig. 1, and this positional difference is denoted by the disparity  $d = x_l - x_r$ . According to the similar triangle theorem, the  $D$  can be expressed as

$$D = \frac{f \cdot B}{d}. \quad (1)$$

Satisfying the geometric relation in Fig. 1 requires ensuring that the optical axes of the two cameras are parallel and that the left and right imaging planes are coplanar, so all epipolar lines are parallel, and the corresponding 2-D points representing the same 3-D point in both images have the same vertical coordinates. If these conditions hold, finding a point's counterpart in the other image requires only looking in the corresponding row. Because of a few pixels, we utilize a physical structure rather than a stereo rectification algorithm to achieve the above requirements and ensure no information is lost.

As shown in Fig. 2, the hardware part of the binocular vision ranging platform based on PF-SPAD consists of two single-photon image sensors (Sensor1 and Sensor2) with  $64 \times 8$  resolution, a control system based on a field-programmable gate array (FPGA), and a display, processing, and storage system. The lenses have a field of view of  $78^\circ \times 60^\circ \times 40^\circ$  (diagonal  $\times$  horizontal  $\times$  vertical) and a focal length of 6 mm. The control system uses Pango Micro's PGL50H FPGA board, which mainly consists of two modules: IIC (Inter-Integrated Circuit) master and USB controller. The IIC is used to configure the working modes and read image data. The USB controller module controls the FT245 for transferring image data to a PC. The display, processing, and storage system runs a LabVIEW program (National Instruments, Austin, TX, USA), which first acquires each frame, performs median filtering with a kernel size of  $3 \times 1$ , and displays result images, then extends

the exposure time and averages multiple frames into a single image to eliminate shot noise, and finally stores the images for subsequent stereo matching. 3-D printing can quickly fabricate the mount for the two single-photon image sensors for a lightweight and stable structure. The installation of the single-photon image sensors ensures that the images are horizontally parallel to each other, allowing for a greater dynamic range of disparity since the images' width is larger than the height. The outgoing direction of the laser pointer is perpendicular to the imaging plane.

In the software part, the stored images are subjected to an improved local block-matching (BM) stereo algorithm [8]. The OpenCV-based algorithm calculates the corresponding points of the two images and generates a disparity image. Then, the disparity image is converted to a depth map using (1). The algorithm details are presented in Section II-C.

### B. Camera Calibration

In order to obtain the intrinsic matrix information from 2-D images, camera calibration is an essential step in 3-D computer vision. We use the pinhole imaging model, and the relationship between the 3-D point  $M = [X, Y, Z, 1]^T$  in the world coordinate system and its image projection point  $m = [\mu, v, 1]^T$  is as follows:

$$sm = A[R \ t]M, \text{ with } A = \begin{bmatrix} \alpha & \gamma & \mu_0 \\ 0 & \beta & \nu_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where  $s$  is the scale factor,  $[R \ t]$  is called the extrinsic parameters, which are the rotation and translation between the world coordinate system and the camera coordinate system, and  $A$  is called the intrinsic matrix, where  $\mu_0$  and  $\nu_0$  are the coordinates of the principal point,  $\alpha$  and  $\beta$  are the focal lengths of the image  $u$ -axis and  $v$ -axis, and  $\gamma$  stands for the degree of tilt of the two image axes. The problem of camera calibration is the solution of the five parameters in the intrinsic matrix.  $\alpha$  has the same meaning as represented by  $f$  in (1).

Traditional digital camera calibrations usually use multiple images containing specific objects, such as checkerboards [18] and Apriltag [19]. The coordinates of multiple feature points in these images must be precisely determined. Calibration methods that require dozens of feature points are meaningless for PF-SPAD images with a few pixels. We use the method proposed by Zhang, which requires a minimum of only three feature points [20], stringing three white balls with a thin cord and sticking the outer one on a black background cloth [Fig. 3(a)]. The positions of the balls' centers, i.e., the coordinates of the feature points, are not easy to estimate because the balls occupy many pixels, and the images' height is too small to move point B with enough range of motion [Figs. 3(b) and (c)]. For images with significant differences in width and height, 1-D objects cannot have an extensive range of motion. In addition, because this method uses the vector products of the three feature point vectors two by two for subsequent processing, errors accumulate as the operation proceeds.

As an improvement, we use a simple method for camera calibration. We only need to solve  $\alpha$  in (2), forgoing the

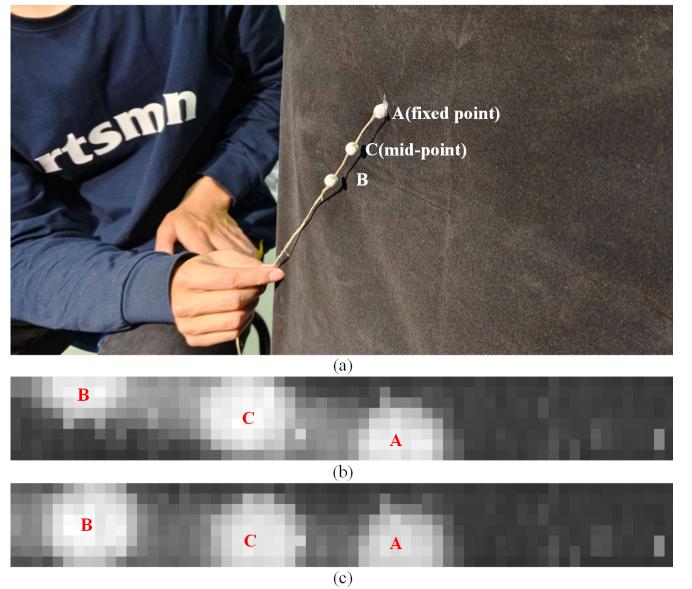


Fig. 3. Calibration using 1-D objects. (a) A string of small balls is fixed at one end and moved at the other, and images are taken at different positions. (b) The number of pixels in the PF-SPAD image is small, and the measurement error of the centers of the small balls is too large for calibration. (c) Moving the position to take a picture, the range of motion visible for all three balls is small.

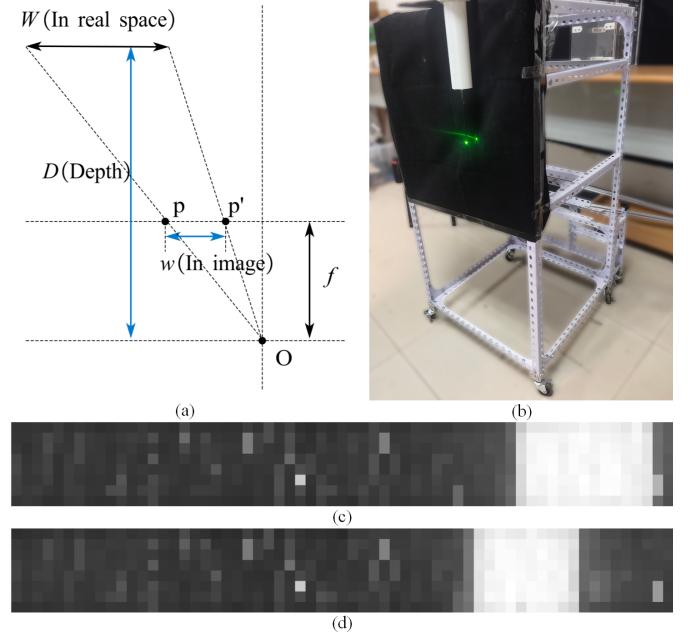


Fig. 4. Calibration in the horizontal direction only. (a) Geometric relation between image and actual width. (b) Device used for calibration. (c) PF-SPAD imaging of a cylinder with precise object edges and small  $w$  error. (d) Image after moving the cylinder far from the camera.

determination of the remaining parameters in the intrinsic matrix  $A$ . In Fig. 4(a), a monocular camera photographs an object with depth  $D$  and width  $W$ , which appears as width  $w$  in the image. It is usually unable to measure the depth  $D$  directly due to uncertainty in the optical center  $O$  coordinates. However, if the coordinates  $Z = D - D_0$  are established with an arbitrary origin in the depth direction, the relationship between  $Z$  and  $w$  can be obtained as follows:

TABLE I  
CALIBRATION RESULTS

Camera	W=40mm	W=20mm	W=10mm	Average
Sensor1	138.18	134.54	98.50	<b>123.74</b>
Sensor2	121.92	139.31	103.65	<b>121.63</b>

$$\frac{1}{w} = \frac{Z + D_0}{f \cdot W} \quad (3)$$

where  $f$  can be estimated by finding the slope of the  $1/w-Z$ , which is independent of  $D_0$ , and the width  $W$  is measurable. This method utilizes only the horizontal coordinates of the images, and the object is free to move within the observable range. In addition, we determine the object's width  $w$  in the images easily, and only its reciprocal is processed, reducing the error when the number of pixels in the images is small.

In Fig. 4(b), a calibration device example is depicted. It consists of a stand with rollers at the bottom, covered with a black background cloth. In front of the background cloth is a white cylinder mounted with its axis perpendicular to the ground. The cylinder is chosen because it has the same width when viewed at different heights. During calibration, keep a single-photon image sensor fixed and its optical axis visually parallel to the ground. Figs. 4(c) and (d) show the gray image with different distances between the cylinder and the sensor. The distance between the white cylinder and the sensor's optical center cannot be measured directly. Since the light from the laser pointer is perpendicular to the imaging plane, we extend a rod from the cylinder's axis and use a tape measure to measure the distance  $Z$  from the rod to the emitting end of the laser pointer when the light shines on the rod, which ensures that the measurement results are accurate and stable.

We calibrate the two single-photon image sensors using cylinders with 40 mm, 20 mm, and 10 mm diameters. The least squares method is used to fit the relationship between  $1/w$  and  $Z$  linearly, and Fig. 5 shows the comparison between the raw data and the fitting results. The difference between the Sensor1 results and the Sensor2 results in the horizontal direction comes from the difference caused by the laser pointer's mounting position in the experiment. When the object widths  $W$  are the same, the slopes of the fitting results of Sensor1 and Sensor2 are almost the same. Table I gives the calibration results  $f$  of two sensors, showing that the calibration results at  $W=10\text{mm}$  significantly differ from the average value because the  $w$  is smaller than other cylinders at the same distance. Moreover,  $w$  can only be taken as an integer, leading to a significant error in  $1/w$ .

### C. Stereo Matching

Stereo matching is a fundamental method in stereo vision systems that utilizes two differently positioned cameras to extract 3-D information by evaluating the relative positions of objects in the two images.

The main window-based local matching criteria commonly used in stereo vision are the sum of absolute or squared differences (SAD / SSD) and normalized cross-correlation (NCC) [21]. The BM algorithm performs image stereo

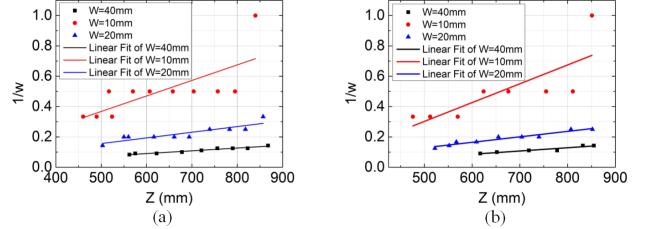


Fig. 5. Straight-line fitting results. (a) The Sensor1 results. (b) The Sensor2 results.

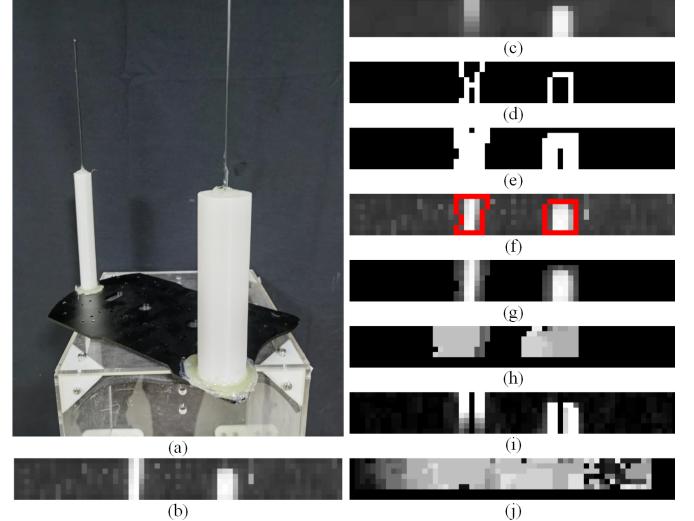


Fig. 6. Preprocessing effects. (a) Original object scene. (b) The raw data image received from the sensor on the left. (c) The result after a morphological open operation. (d) The result after edge detection using the Canny operator. Object contour and texture information are effectively extracted. (e) The result after a morphological expansion operation. (f) The result after contour finding. (g) The result after ROI extraction. (h) Effect of SAD matching on preprocessing results. (i) Horizontal Sobel operator preprocessing result. (j) Effect of SAD matching using Sobel operator preprocessing results.

matching using SAD. We improve the BM algorithm to adapt the PF-SPAD image. The algorithm can be categorized into three steps: 1) preprocessing, 2) disparity computing, and 3) disparity image optimization.

1) *Preprocessing*: The raw image data obtained by PF-SPAD can be divided into valid object information and noise, mainly dark counts, afterpulsing noise, and shot noise. Noise significantly interferes with calculating the disparity image, thus affecting the final depth result. The preprocessing of the image can reduce the impact of noise and the angle of the illumination light on the image.

We preprocess the images using OpenCV. This open-source computer vision library has a variety of built-in filters and differential operators to simplify the complexity of image processing. The preprocessing is divided into five steps: morphological open operation, edge detection, morphological expansion operation, contour finding, and region of interest (ROI) extraction. We use the example of a scene with two white cylinders on a black background [Fig. 6(a)]. There is much noise in PF-SPAD imaging [Fig. 6(b)]. The morphological open operation can remove the isolated noise and smooth the image [Fig. 6(c)]. We use the Canny operator, a second-order differential operator, for edge detection, i.e., we determine the image's edge pixels by the image function's maximum value

[22]. We use processing steps such as non-maximum suppression and hysteresis thresholding algorithm to improve the edge detection accuracy and obtain the image's texture information [Fig. 6(d)]. Our program sets the size of the Canny operator to 3. The morphological expansion operation can expand the range of the white area and keep the contour as continuous as possible [Fig. 6(e)]. The morphological open and morphological expansion operations use  $3 \times 3$  rectangular-structure elements. If the scene is dark and edge detection is affected by noise, the morphological open operation is further performed using  $3 \times 3$  cross-structure elements. Contour finding uses texture information to find valid object contour boundaries [Fig. 6(f)]. Finally, within the outer boundary of the contour is selected as the ROI, and the noise information outside the ROI is eliminated [Fig. 6(g)], and the ROI is saved for subsequent algorithms to process. If the scene is dark, use the original image after  $3 \times 3$  size kernel Gaussian filtering. The preprocessing results are directly SAD-matched, where small speckles can be filtered out by subsequent connective domain detection [Fig. 6(h)], and the object has smooth disparity values. The built-in BM algorithm of OpenCV uses a horizontally oriented Sobel filtering for preprocessing. Sobel filtering computes horizontal partial derivatives using the Sobel operator, a first-order differential operator. To compare with our preprocessing method, the size of the Sobel operator is also set to 3. The white pixel band in the resultant image is the region with a large absolute gradient value, and the noise is effectively filtered out. However, the effective object information reduces [Fig. 6(i)], the edges of the image are sharper, but the internal information of the object is lost, and the variability between the edges cannot be distinguished. Using the Sobel filtering results for SAD matching, the image is noisy, and the object disparity is not easy to distinguish [Fig. 6(j)], the internal disparity of the object is not smooth, and the external low-texture region is not removed.

2) *Disparity Computing*: The basic idea of the SAD algorithm is to sum the absolute differences between the corresponding pixels of the left and right rectangular image blocks and match the image block with the minimum value of the similarity measure function [23]. The SAD similarity measure function can be expressed as

$$C_{SAD} = \sum_{i \in w_1} \sum_{j \in w_2} |IL - IR| \quad (4)$$

$$IL = I_L(p+i, q+j) \quad (5)$$

$$IR = I_R(p+i+d, q+j)$$

where  $I_L(p,q)$  and  $I_R(p,q)$  are the grey values of the pixels at  $(p,q)$  in the left and right images, respectively,  $d$  is equal to the disparity when  $C_{SAD}$  is minimum in a given disparity search range  $w_3$ , and  $w_1$  and  $w_2$  are the sets of integers from 0 to the width and height of the image block, respectively.

We use the preprocessed images for disparity computing. We match each image block in the left image with the block in the corresponding row in the right image to obtain the disparity image. In addition, we regard the black regions in morphological expansion resultant images of the left image as

---

**Algorithm 1** Disparity Map Hole Filling Algorithm

---

**Require:** The ROI in the disparity image (based on the left image),  $R_{mn}$ ;

**Ensure:** The optimized ROI,  $P_{mn}$ ;

- 1: Initialize a  $m \times n$  array  $P_{mn}$  with all 0 values;
- 2: Calculate the mode of the internal disparities of the ROI and define the mode as  $M$ ;
- 3: **for**  $R_{i,j}$  such that  $R_{i,j} \in R_{mn}$  **do**
- 4:     Define  $p_{i,j}$  as the disparity value in  $R_{i,j}$ ;
- 5:     **if**  $p_{i,j} = 0$  **then**
- 6:         **if**  $p_{i-1,j} > 0$  or  $p_{i,j-1} > 0$  or  $p_{i+1,j} > 0$  or  $p_{i,j+1} > 0$  **then**
- 7:              $P_{i,j} \leftarrow M$ ;
- 8:         **end if**
- 9:     **end if**
- 10: **end for**
- 11: Perform Gaussian filtering on  $P_{mn}$  with a kernel size of  $3 \times 1$ ;
- 12: **return**  $P_{mn}$ ;

---

regions with less texture information. If the left image block is in these regions, we set the corresponding disparity value to *new\_Value*, which represents an inaccurate disparity calculation. Our program sets *new\_Value* to 0 and uses the asymmetric image block of size  $2 \times 5$ . The algorithm's time complexity is  $O(WHD)$  (width×height×disparity search range).

3) *Disparity Image Optimization*: Influenced by the image's noise, occlusion, weak texture, and repetitive texture, the disparity value obtained after the disparity computing usually has a significant deviation. We must detect the connective domain and fill the holes in the disparity image.

The first step is to analyze the connective domains in the disparity image to detect speckles due to noise and remove them. The connected pixel regions where the difference between any neighboring disparity values does not exceed *max\_Diff* are judged to be connective domains. *max\_Diff* is set to 1 in the experiment, which can be modified according to the tolerance for differences in disparity within a single object. In addition, the program removes connective domains with the number of pixels smaller than a threshold *max\_Speckle\_Size* and sets the threshold *max\_Speckle\_Size* to 10, which needs to be adjusted according to the size of speckles to be removed, setting the internal disparity value to *new\_Value*. The meaning of *new\_Value* is as described in the previous section.

The second step is to fill the holes. Unreliable disparities *new\_Value* will appear in the region with effective object information, i.e., they will create data holes in the disparity image. We developed a filling algorithm based on calculating the mode of the disparity values in the ROI. The mode of disparities is used to fill the data holes since the disparities within the same contour are similar for PF-SPAD images. Because the SAD matching is performed based on the left image, the ROI is still selected for hole filling based on the preprocessing contours in the left image. The ROI needs to be widened according to the width of the SAD image block. The Gaussian filter is used to filter out the noise of the ROI to obtain the final disparity image. We set the kernel size for Gaussian filtering to  $3 \times 1$ . Algorithm 1 shows the process of holes-filling

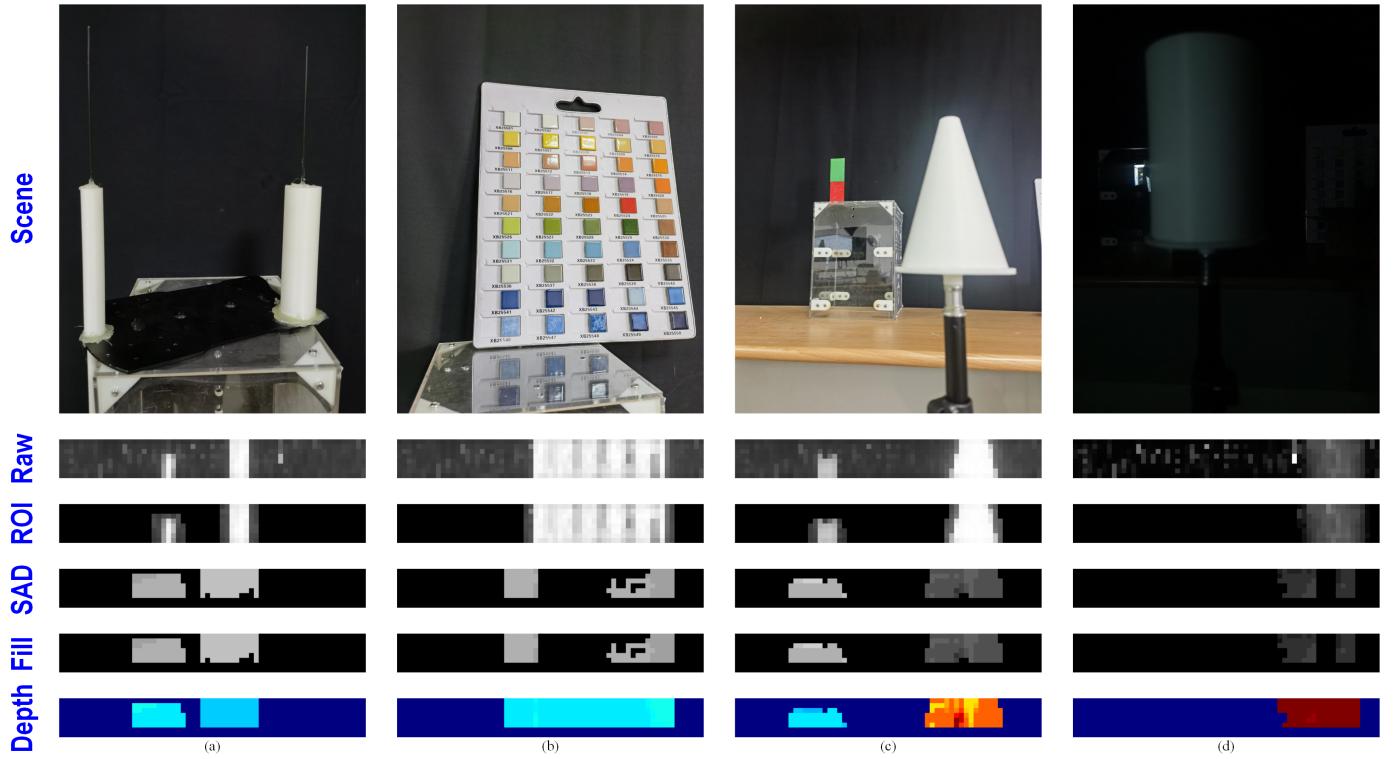


Fig. 7. Validation of stereo matching effects. (a) Multi-object scene. (b) Repeated texture scene. (c) Scene with objects at different distances and varying widths in the vertical direction. (d) Night scene.

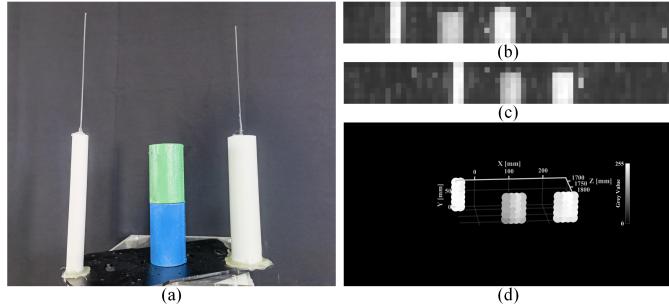


Fig. 8. 3-D imaging procedure. (a) Experimental scene. (b) Left image. (c) Right image. (d) Point cloud diagram. Larger grey values of the dots represent more significant photon fluxes.

in the disparity image. After filling the holes, we convert the disparity image into a depth map using (1).

The stereo matching algorithm in this section is adaptive and robust in different scenes. Fig. 7 shows the processing of the stereo matching algorithm in four scenes: cylinders with a small distance gap, a flat plate inlaid with ceramic blocks of different colors, a circular truncated cone and a triangular prism with a large distance gap, and a cylinder indoors at night (with a scene illumination of 5.0 lux). Each of the four groups of images contains a scene image, a PF-SPAD raw image, an image after preprocessing, an image after disparity computing and connective domain detection, an image after holes filling, and a depth image. The improved BM algorithm gives good depth results for scenes with multiple targets, repeated textures, variations in vertical orientation, and different illumination conditions. Fig. 7(a) shows a multi-object scene. In the preprocessed image (ROI), noise and objects are well separated, and the noise part is eliminated. In the image after disparity computing and connective domain detection (SAD),

the SAD matching causes the object contour to widen. Disparity results are multiplied by 16 and converted to 8-bit unsigned integers for display convenience. In the image after hole filling (Fill), the holes-filling algorithm fills and recovers the data holes of the disparity image, and the Gaussian filter smoothes the disparity image to suppress local noise. In the depth map (Depth), warm colors represent less depth, cool colors represent more depth, and dark blue represents undetectable depth. Fig. 7(b) shows a repeated texture scene. In the preprocessed image, as much texture information as possible is preserved. Fig. 7(c) shows a scene with objects at different distances and varying widths in the vertical direction. Fig. 7(d) shows a night scene. In the raw data image (Raw), the contrast is enhanced so that the noise is more pronounced.

After obtaining the depth results, the coordinates of the 3-D points corresponding to the 2-D pixels can be calculated based on (2). Without loss of generality, we choose the camera coordinate system to represent the 3-D point so that  $R=I$  and  $t=0$  in (2). We use the dimensions of the objects in the scenes described above and the positions of the objects relative to the sensors to roughly estimate the other parameters in the intrinsic matrix  $A$  except  $\alpha$ . To demonstrate the 3-D imaging effect, the scene shown in Fig. 8(a) is constructed, where two cylinders have the same depth, and the other has a slightly lower depth than the previous two. Use Sensor1 on the left and Sensor2 on the right to capture this scene, and the acquired images are Fig. 8(b) and Fig. 8(c), respectively. The point cloud image is shown in Fig. 8(d). The diameters of the three objects in the point cloud image are 15 mm, 45 mm, and 45 mm from left to right, and the ground truth is 20 mm, 50 mm, and 40 mm, respectively. The point cloud image can be seen to reflect the

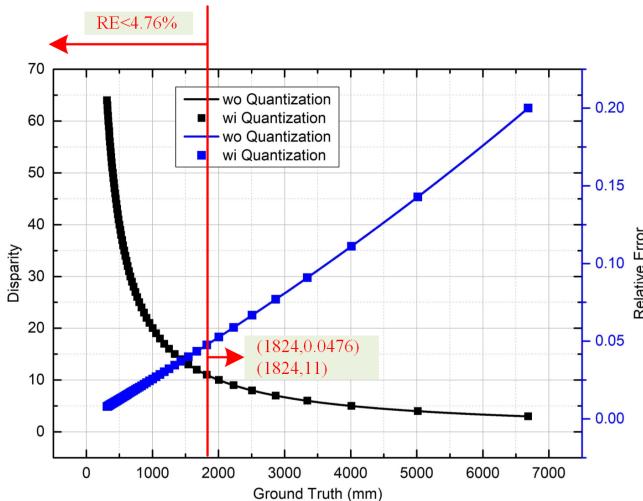


Fig. 9. Distance versus disparity (black line) and maximum relative error (blue line) under ideal conditions.

actual 3-D space.

### III. EXPERIMENTS

#### Evaluation of Low-Light-Ranging Performance

In this section, the effect of low illumination on the measurement results of the binocular vision ranging platform based on PF-SPAD is demonstrated. It is worth mentioning that because the disparity  $d$  in (1) takes a finite range of integers, there is a quantization error in the binocular vision ranging system. Assume  $d$  is quantized by rounding that is,

$$d = i, \quad \text{when } \frac{B \cdot f}{i + 0.5} < D < \frac{B \cdot f}{i - 0.5}. \quad (6)$$

Equation (6) represents an imaging process that fully conforms to the pinhole model, and an algorithm accurately accomplishes matching at the pixel level. Under the ideal conditions described above, given disparity  $i$  have specific depth measurement relative error ( $RE$ ) maxima as follows

$$RE < \frac{\frac{B \cdot f}{i - 0.5} - \frac{B \cdot f}{i}}{\frac{B \cdot f}{i}} = \frac{0.5}{i - 0.5}. \quad (7)$$

Fig. 9 shows the relationship between measurement distance and disparity/maximum relative error according to (1) and (7). The dots represent points of practical significance after quantification of the disparity. The horizontal coordinates of the green and red dots represent possible distance measurements, and the horizontal spacing between neighboring dots reflects the magnitude of the distance resolution per disparity unit. As the distance increases, the distance resolution per disparity unit and the maximum relative error increase. Due to the limitation of image width, the maximum relative error cannot be arbitrarily close to 0. The maximum measurement range with suitable accuracy is 1824 mm, which is limited by the focal length in pixels and the baseline length, and the maximum relative error under ideal conditions is 4.76% in this range.

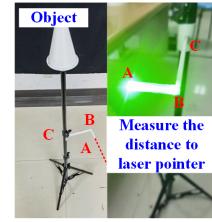


Fig. 10. Setup for low-light-ranging performance evaluation.

The experimental setup is shown in Fig. 10, which evaluates the measurement accuracy and precision. The object is placed on a carrier plate on a tripod, and a right-angled part is mounted on the tripod's support column to help define the tripod's position. A slot is engraved on the part, and the tripod is placed where the laser can illuminate the slot uniformly to make AB parallel to the light. The length of BC ensures that the object is simultaneously in the field of view of both cameras. Measure the distance of point A from the emitting end of the laser pointer. The scene is arranged so the ranging platform has ample space from the wall to ensure the measuring range. The wall is covered with a black background cloth. A white cylinder and a circular truncated cone are the objects to be observed, and the left and right images of the objects at different distances can be obtained by moving the tripod. We set up two lighting conditions, one for a bright indoor scene and the other for a nighttime indoor scene. The bright indoor scene is illuminated with an ample fill light, and the scene's illumination ranges from 167 lux to 327 lux, with the maximum and minimum values corresponding to the closest and farthest points, respectively. The nighttime indoor scene has no direct light source, and the scene illuminance is at 5.0 lux. In both scenes, measurements are taken at different distances within 1824 mm, with one shot of the cylinder and one shot of the circular truncated cone at each measurement point.

When processing the  $n$ th set of images, the depth  $D_n$  corresponding to the mode of the disparity values at the time of hole filling is recorded as the distance measurement, which has a systematic error  $D_0$  with the distance  $Z_n$  from point A to the emitting end of the laser pointer as affected by the length of AB and the laser pointer mounting position. We use the least squares method (LSM) to obtain the systematic error  $D_0$  that can be expressed as

$$LSM = \min \sum_n (D_n - D_0 - Z_n)^2. \quad (8)$$

Setting the LSM's derivation to 0, we can get

$$D_0 = \bar{D} - \bar{Z} \quad (9)$$

where  $\bar{D}$  and  $\bar{Z}$  are the mean values of  $D_n$  and  $Z_n$ , respectively. The  $D_0$  is constant when the testing platform is fixed, and the  $RE$  can be obtained using (10).

$$RE_n = \frac{|D_n - D_0 - Z_n|}{Z_n + D_0}. \quad (10)$$

If there is an outlier  $RE_n > 4.76\%$ ,  $D_n$  and  $Z_n$  of the  $n$ th set of images are excluded from the calculation of the mean values, and then use (9) to calculate. If there is no outlier, use the  $D_0$  to correct  $Z_n$ , including the outliers; otherwise, repeat the

TABLE II  
PERFORMANCE SUMMARY AND COMPARISON WITH OTHER STATE-OF-THE-ART 3D IMAGERS

parameters	unit	THIS WORK	[24]	[25]	[26]	[27]	[28]
3D Imaging Method		stereo vision	stereo vision	structured light	dToF	dToF	dToF
CMOS Technology		110 nm	-	-	-	160 nm	150 nm
Pixels Resolution		59×6	1280×720	1280×720	16×16	40×10	50×40
Light-sensitive Device		SPAD	CMOS	CMOS	SPAD	SPAD	SPAD
Horizontal Field of View	Deg	18/60	65	40	27	-	-
Measurement Range	m	1.84/10	5	0.9	4	2	3
Average Laser Power	μW	0	0	30000	40	14	40
Minimum Background	lux	5	25	0.8	-	-	-

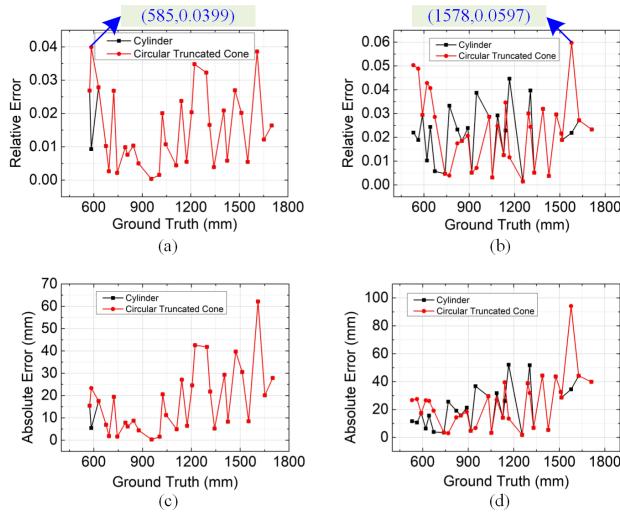


Fig. 11. Accuracy evaluation. (a) The relative error for the 167 lux to 327 lux illumination scenes. (b) The relative error for the 5.0 lux illumination scenes. (c) The absolute error for the 167 lux to 327 lux illumination scenes. (d) The absolute error for the 5.0 lux illumination scenes.

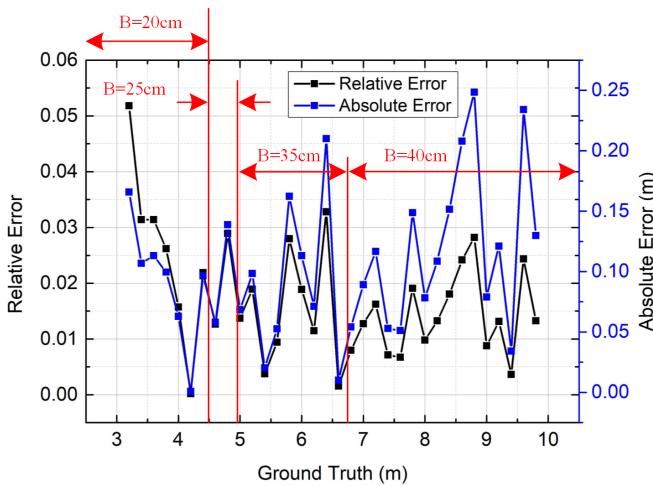


Fig. 12. Baseline-adjustable ranging evaluation. (a) The relative error. (b) The absolute error.

exclusion step. The corrected result is the actual distance.

Fig. 11 shows the relationship between relative and absolute errors and ground truth. When the ground truth is less than 700mm, the relative error is prominent, caused by image distortion and assembly error. When the ground truth is more significant than 700mm, the relative and absolute errors tend to

increase with the increase of the ground truth. In the bright indoor scene, the measured values of the cylinder and circular truncated cone are nearly the same, indicating that the coplanarity of the imaging planes is basically satisfied. However, in the nighttime indoor scene, the measured values of the cylinder and the circular truncated cone are different due to the non-uniformity of light on the surfaces of the two objects. The maximum relative error in the bright indoor scene is 3.99%, less than the reference value of 4.76%, caused by fewer data at large distances. The maximum relative error in the nighttime indoor scene is 5.97%, which is slightly higher than the bright indoor scene because the SNR of the image is lower in low light. To summarize, the adaptability of the proposed platform in low illumination and the coplanarity of the left and right imaging planes are verified.

#### A. Evaluation of Baseline-Adjustable Ranging Performance

In this section, a baseline-adjustable design is introduced, and the performance of the binocular vision ranging platform based on PF-SPAD is evaluated for different distance scenarios. According to (7), the measurement relative error can only be small enough when the disparity is large enough, so when the baseline is fixed, high accuracy can only be obtained at a specific distance interval. If the baseline is adjustable, stereo vision can be adapted to different target distance scenarios. A single-photon image sensor with 64×8 resolution is fixed in a shell with pulleys that can be moved on a slide rail. However, the entire assembly can also be bolted to the slide rail. The sensor can be moved on the slide rail up to 40 cm. Since stationary scenes are captured, one sensor is sufficient, and the left and right images are captured by moving the sensor by one baseline length. The scale below the slide rail can read off the baseline length  $B$  in (1). The sensor was calibrated using the method described in Section II-B to obtain  $f$ . The experimental scene is a stand with a long white bar mounted on it. The stand was moved to different distances, and the appropriate baseline length was adjusted to ensure that the disparity  $d$  was taken within the proper range, thus obtaining distance measurements.

The relative and absolute errors of distance measurements as a function of distance are shown in Fig. 12. For the baseline-adjustable design, when the actual distance is in the range of 4 m to 10 m, the relative error does not exceed 3.28%, and the absolute error does not exceed 0.25 m, fluctuating steadily within a specific interval. Compared to the

baseline-fixed design, the rising trend of baseline-adjustable relative and absolute errors with distance is less pronounced because the disparity is limited to a suitable range, and maximum relative error does not change much under ideal conditions. As described in Section III-A, the baseline-adjustable design still presents high relative errors at low actual distances. In conclusion, the baseline-adjustable design can adapt to scenarios with different target distances compared to the baseline-fixed one.

Much research on 3-D imaging based on stereo vision or dToF is being carried out. Table II compares the binocular vision ranging platform based on PF-SPAD with the state-of-the-art, including digital camera-based binocular vision ranging platforms and dToF LiDAR based on SPAD array. Our method works in darker scenes, unlike stereo vision methods that use digital cameras. Our method saves active illumination power compared to binocular structured light methods using digital cameras and dToF LiDAR.

#### IV. CONCLUSION

This work presents a binocular vision ranging platform based on PF-SPAD. This work demonstrates that even though conventional camera calibration methods do not apply to the PF-SPAD images with a few pixels, it is still possible to estimate the focal length  $f$  in pixels by a simple method. The calibration experimental setup needs to be simplified in the future, and a feasible approach is to utilize the dToF method to measure the distance  $Z$  and, thus, remove the laser pointer and the associated setup. In addition, this work designs an improved BM algorithm for stereo matching of PF-SPAD images. This work verifies the algorithm's robustness in scenes with multiple targets, repetitive textures, variations in vertical orientation, and different illumination conditions. Finally, this work evaluates the ranging performance of our platform. Within a distance of 1824mm, the maximum relative error is 3.99% in the 167 lux to 327 lux illumination scenes and 5.97% in the 5.0 lux illumination scenes. The adaptability of the platform to low-light illumination and the coplanarity of the left and right imaging planes are verified. Our platform has the potential for binocular vision-ranging applications in low-light environments. This work introduces a baseline-adjustable design adapted to different distance scenarios, using which the maximum relative error in the range of 4 m to 10 m is 3.28%.

#### REFERENCES

- [1] J. Ma *et al.*, "A 64×64 Pixel Image Sensor With Gain-Configurable Photodiodes and Combined Subrange Method," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1-10 (Art no. 2000810), 2024.
- [2] H. Lin *et al.*, "Depth Measurement Based on Stereo Vision With Integrated Camera Rotation," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
- [3] P. Hu *et al.*, "Design and Implementation of Binocular Vision System with an Adjustable Baseline and High Synchronization," *IEEE ICIVC*, Chongqing, CHN, pp. 566–570, 2018.
- [4] A. Ingle *et al.*, "High Flux Passive Imaging With Single-Photon Sensors," *IEEE CVPR*, Long Beach, CA, USA, pp. 6753–6762, 2019.
- [5] A. Ingle *et al.*, "Passive Inter-Photon Imaging," *IEEE CVPR*, Nashville, TN, USA, pp. 8581–8591, 2024.
- [6] S. Ma, *et al.*, "Burst Vision Using Single-Photon Cameras," *IEEE WACV*, Waikoloa, HI, USA, pp. 5364–5374, 2023.
- [7] L. Keselman *et al.*, "Intel® RealSense™ Stereoscopic Depth Cameras," *IEEE CVPR*, Honolulu, HI, USA, pp. 1267–1276, 2017.
- [8] K. Konolige, "Projected texture stereo," *IEEE CRA*, Anchorage, AK, US pp. 148–155, 2010.
- [9] B. M. Amin *et al.*, "3-D Raman Imaging Using Time-Resolving CMOS SPAD Line Sensor and 2-D Mapping," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1-10 (Art no. 6006210), 2023.
- [10] T. Xu *et al.*, "A Near-Infrared Single-Photon Detector for Direct Time-of-Flight Measurement Using Time-to-Amplitude-Digital Hybrid Conversion Method," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1-9 (Art no. 4500309), 2024.
- [11] C. Niclass *et al.*, "A 0.18-μm CMOS SoC for a 100-m-range 10-frame/s 200 × 96-pixel time-of-flight depth sensor," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 315–330, 2014.
- [12] S. Jahromi *et al.*, "A 32 × 128 SPAD-257 TDC Receiver IC for Pulsed TOF Solid-State 3-D Imaging," *IEEE J. Solid-State Circuits*, vol. 55, no. 7, pp. 1960–1970, 2020.
- [13] M. Perenzoni *et al.*, "A 64 × 64-Pixels Digital Silicon Photomultiplier Direct TOF Sensor With 100-MPhotons/s/pixel Background Rejection and Imaging/Altimeter Mode With 0.14% Precision Up To 6 km for Spacecraft Navigation and Landing," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 151–160, 2017.
- [14] D. Li *et al.*, "Optimization of System Design and Calibration Algorithm for SPAD-Based LiDAR Imager," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-10 (Art no. 2006010), 2022.
- [15] M. Wei *et al.*, "Passive Ultra-Wideband Single-Photon Imaging," *IEEE ICCV*, Paris, France, pp. 8101–8112, 2023.
- [16] T. Seats *et al.*, "Motion Adaptive Deblurring with Single-Photon Cameras," *IEEE WACV*, Virtual, pp. 1944–1953, 2021.
- [17] K. Morimoto *et al.*, "3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing," *IEEE IEDM*, San Francisco, CA, USA, p. 20.2.1-20.2.4, 2021.
- [18] Zhengyou Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," *IEEE ICCV*, Kerkyra, Greece, pp. 666–673, 1999.
- [19] E. Olson, "AprilTag: A robust and flexible visual fiducial system," *IEEE ICRA*, Shanghai, pp. 3400–3407, 2011.
- [20] Zhengyou Zhang, "Camera calibration with one-dimensional objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 7, pp. 892–899, 2004.
- [21] K. Mikolajczyk *et al.*, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [22] W. Zhao *et al.*, "An Insulator in Transmission Lines Recognition and Fault Detection Model Based on Improved Faster RCNN," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-8, 2021.
- [23] K. Ioannidis *et al.*, "A Digital Image Stabilization Method Based on the Hilbert–Huang Transform," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 9, pp. 2446–2457, 2012.
- [24] A. Grunnet-Jepsen *et al.*, "Projectors for Intel® RealSense™ Depth Cameras D4xx," *Intel Support*, Intel Corporation, Santa Clara, CA, USA 2018.
- [25] Y. Ou *et al.*, "Binocular Structured Light 3-D Reconstruction System for Low-Light Underwater Environments: Design, Modeling, and Laser-Based Calibration," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.
- [26] L. J. Koerner, "Models of Direct Time-of-Flight Sensor Precision That Enable Optimal Design and Dynamic Configuration," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
- [27] V. Sesta *et al.*, "Range-Finding SPAD Array With Smart Laser-Spot Tracking and TDC Sharing for Background Suppression," *IEEE Open J. Solid-State Circuits Soc.*, vol. 2, pp. 26–37, 2022.
- [28] M. Perenzoni *et al.*, "A Fast 50 × 40-Pixels Single-Point DTOF SPAD Sensor With Photon Counting and Programmable ROI TDCs, With  $\sigma < 4$  mm at 3 m up to 18 klux of Background Light," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 86–89, 2020.