ARL-TR-9108 ● Nov 2020

DEVCOM ARMY RESEARCH LABORATORY

Co-calibration and Registration of Color and Event Cameras

by Qifan Zhang, Jinwei Ye, Philip Osteen, and S Susan Young

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-9108 ● Nov 2020

Co-calibration and Registration of Color and Event Cameras

Qifan Zhang and Jinwei Ye
Louisiana State University

Philip Osteen
Vehicle Technology Directorate, DEVCOM Army Research Laboratory

S Susan Young
Computational and Information Sciences Directorate, DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
November 2020	Technical Report	October 2019–October 2020

4. TITLE AND SUBTITLE
Co-calibration and Registration of Color and Event Cameras

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Qifan Zhang, Jinwei Ye, Philip Osteen, and S Susan Young

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DEVCOM Army Research Laboratory
ATTN: FCDD-RLC-CI
2800 Powder Mill Rd
Adelphi, MD 20783

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-9108

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The event camera uses a novel bio-inspired sensor that works radically different from the traditional optical sensor. Registration between color and event images is challenging due to the data heterogeneity. This technical report presents a method for event and color image registration through camera calibration. Specifically, an event and a color camera are co-calibrated with a common calibration target. As the event camera cannot capture static scenes, we change the intensity of the light source around the calibration board and generate the accumulated frame. At the same time, we press the color camera shutter to obtain a color image. According to this method, multiple pairs of calibration board images are obtained. Next, we combine the event camera with the color camera to perform stereo calibration. The intrinsic and extrinsic parameters of each camera can be obtained respectively, and the relative position between the two cameras determined. Based on the two cameras' homography matrix, we can deduce the corresponding position of the pixel in the color image, which is the registration of color and event cameras. The results of experiments performed with a color camera and an event camera pair show the method is effective.

15. SUBJECT TERMS
event camera, color images, camera calibration, image registration, homographic transformation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			S Susan Young
Unclassified	Unclassified	Unclassified	UU	28	19b. TELEPHONE NUMBER (Include area code) (301) 394-0230

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Introduction

Researchers are inspired by the principle of the human retina to design a neuromorphic vision sensor, which is now called dynamic vision sensor (DVS) or an event-based camera. The event-based camera works radically different from the traditional optical sensor (Fig. 1). Instead of capturing images at a fixed rate, it measures per-pixel brightness changes asynchronously. The asynchronous and differential principle of the event camera allows it to capture images at high speed and work in a broad illumination range.

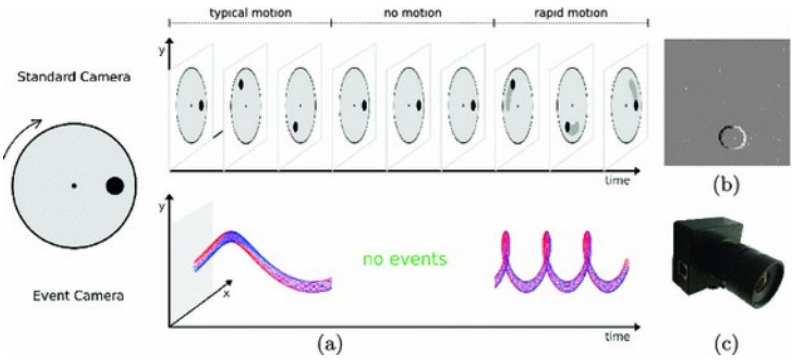

Fig. 1 Comparison between the event camera and the standard camera

When the scene in observation changes, the event camera outputs signals (or "events") that indicate the intensity changes. For each pixel, when the change in intensity exceeds the threshold, this position will produce an event output. An event output can be represented with four elements (X, Y, P, T). (X, Y) is the 2-D coordinate of the event on the image plane. T is the time the event takes. P is the polarity, which is either positive or negative, indicating the direction (increase or decrease) of the intensity change at that pixel and time. In our experiments, we use the Dynamic Vision Platform (DVP) to collect and process the event camera data. The DVP includes the decoding and integration functions to facilitate our data processing.

Since the event camera can perceive dynamic changes, the captured event stream exhibits motion-dependent characteristics. The advantages of event cameras include very high temporal resolution and low latency (both in microseconds), high dynamic range (140 dB vs. 60 dB of standard cameras), and low power consumption. Hence, the event cameras have large potential for robotics and military applications in challenging scenarios for traditional color cameras, such as presence of high-speed moving targets and object capturing in low-light conditions. Event cameras work in a fundamentally different way from standard cameras by measuring per-pixel intensity changes asynchronously rather than measuring

"absolute" brightness at constant rate; therefore, novel algorithms need to be developed to process the event images and unlock their potential.

This report presents a method for event and color image registration through camera calibration. Specifically, an event camera and a color camera are first co-calibrated with a calibration checkerboard. As the event camera cannot capture static scenes, we use a flickering pattern shown on a digital display as the target for event camera calibration and then use the calibrated camera parameters for image registration. The calibration board is moved in front of the event camera to generate the accumulated frame, which can be used to calculate camera calibration with regular color image. Next, we combined the event camera with the color camera to perform stereo calibration. The intrinsic and extrinsic parameters of the two cameras can be obtained respectively, and the relative position between the two cameras can be determined, assuming that a pixel in the event image is known. Based on the two cameras' homography matrix, we can deduce the corresponding position of the pixel in the color image, which is the registration of color and event cameras. We perform experiments by pairing a color camera with an event camera. The results of calibration and registration show that the method is effective.

2. Related Work

2.1 Event Camera Background

In conventional computer vision, we mainly rely on the frame-based camera architecture. In this architecture, each pixel in the photo is rapidly captured by the shutter.[1] However, such methods have the problem of long latency, which causes the lack of data between adjacent frames. The event camera overcomes this issue by capturing images asynchronously. Because of its low energy consumption and low latency, the event camera has big potential in computer vision applications.[2] Many algorithms based on the event cameras are developed for object tracking,[3] recognition,[4] visual SLAM,[5] and image deblurring.[6] Gallego et al.[2] conducted an overall survey of the current technologies based on the event camera. The survey provides a comprehensive overview of the emerging field of event-based vision and summarizes the application and algorithm to develop the properties of event camera. Gallego et al.[2] presented objective functions to analyze event alignment in motion compensation approaches, calling them focus loss functions. This work concludes that the variance, the gradient, and the Laplacian magnitudes are among the best loss functions. The applicability of the loss functions is shown on several tasks: rotational motion, depth, and optical flow estimation.

2.2 Event Camera Calibration

Feature extraction is a key step to camera calibration. Clady et al.[7] introduce an event-based luminance-free method to detect and match corner events from the output of asynchronous event-based cameras. Kim[8] uses a blinking checkerboard pattern displayed on a computer monitor. The blinking target is at the same frequency as the event-based camera to allow reconstruction of a sequence of image-like frames by accumulating events within a time interval. Some methods calibrate the event camera through combination with other optical features. Song et al.[9] incorporate a 3-D LiDAR when calibrating the event camera. Morales et al.[10] proposed a bio-inspired stereo vision system, supported by other additional hardware. They present a novel calibration technique for an event-based stereo vision system that is implemented with specialized hardware (field-programmable gate array, or FPGA). It reduces the hardware latencies and allows real-time implementation. In our experiment, we use a similar method to Kim's,[8] which constantly changes the light intensity to obtain the accumulated images. We then compare the event frame image with the color images for registration.

2.3 Event Image Registration

Event image registration methods can be categorized into two classes. The first class is the event-by-event registrations. Galanis[11] treats the output data as an event stream for event camera registration. The method identifies the event traces that objects leave in one DVS data stream and finds the corresponding traces in the other DVS data stream. This method can be used to register data from two-event cameras. Similarly, Kalber[12] uses the event stream to perform registration among event images. The method adapts the point set registration algorithm (e.g., coherent point drift) for event stream registration.

The second class uses "groups of events" for registration. Since a single-event image is highly susceptible to noise, the idea is to put together several event images for processing. This class of methods first converts the event image into a frame image through integration, and then applies traditional image registration methods[13] (e.g., feature-based image registration).[14] The accumulated frames contain additional information beyond the trigger events, which includes spatial information[15] and edge structures, thus resulting in richer features for correspondence matching. Some methods use machine learning for event image registration. Aberman et al.[16] proposed neural network-based feature-mapping between two images, where the object of interest in the two input images can have a large difference in shape and appearance.

3. Approach

This section presents the methods for calibrating the color and event cameras and event and color image registration.

3.1 Camera Calibration

The purpose of the camera calibration process is to find a functional model and obtain the parameters in the model that can approximately fit the 3-D to 2-D process. On the other hand, the inverse function of the 2-D image to the 3-D world can also be obtained from this module.

The camera projection model is critical to calibration.[17] Assume that a pinhole camera model (Fig. 2) is suitable for both event cameras and color cameras. Images in the pinhole camera model are formed by light rays that pass through a tiny pinhole. The fixed structure of the camera lens determines the conjugate relationship between the object and its image. After the corresponding object passes through the pinhole, it becomes an upside-down image on the image plane. To better explain the theory, extend the length of the focal length from the focal point to form a virtual plane parallel to the image plane. The upright image on this plane is what we usually see. In the camera calibration model, parameters can be divided into intrinsic matrix and extrinsic matrix. The intrinsic parameters, considered as some properties of the camera itself, include focal length and location of focal point in the image plane. The extrinsic parameters are mainly the relative position between the cameras with the real-world coordinates of the object. After rotation and translation, world coordinates can be converted to camera coordinates. In addition, during the production of cameras, there are some unavoidable error parameters that will also affect the camera calibration, such as the degree of lens distortion. The higher accuracy of the camera calibration we obtain, the more beneficial for subsequent image processing algorithms.

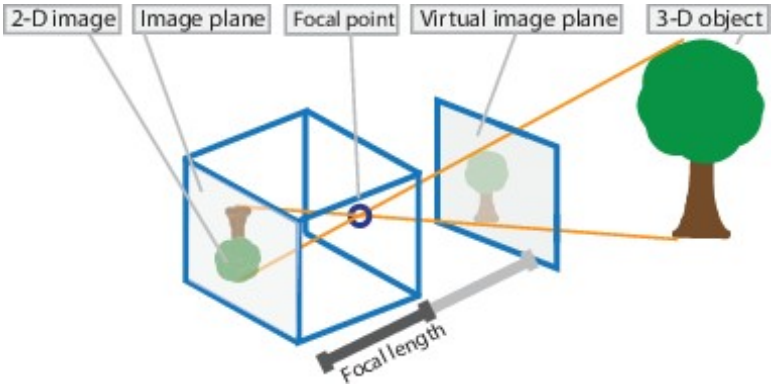

Fig. 2 Pinhole camera model

The pinhole imaging model is simplified into a form of geometric expression (Fig. 3). Before proceeding with the geometric expression, we need to introduce several concepts: world coordinate system (3-D), camera coordinate system (3-D), image coordinate system (2-D), and pixel coordinate system (2-D)[18]:

- World coordinate system: can be freely selected by the surveyor; the camera has its own relative position in the world coordinate system.
- Camera coordinate system: the origin of the coordinates is the optical center of the camera. The plane formed by the x-axis and the y-axis is parallel to the imaging plane, and the optical axis passing through the optical center of the camera is the z-axis. The optical axis is perpendicular to the imaging plane.
- Image coordinate system: the intersection of the optical axis and the virtual imaging plane is used as the origin of the coordinate system, and the x-y plane is the virtual imaging plane.
- Pixel coordinate system: we usually use the upper-left corner of the image as the origin of coordinates.

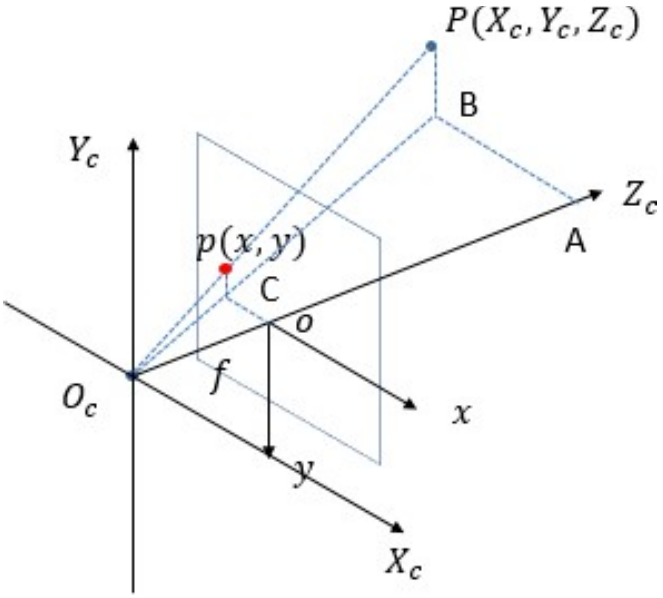

Fig. 3 Camera projection geometry

3.1.1 Convert World Coordinate to Camera Coordinate

We need to map the position of the object in the world coordinate system to the camera coordinate system. There is no deformation of the object during the coordinate conversion process. Thus, the process can be divided into rotation R and translation t as shown in the transformation equation:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{1}$$

The process of rotation can be divided into three parts, rotating *R1*, *R2*, *R3* (keeping the z-axis, x-axis, and y-axis unchanged). Take *R1* as an example, rotate a degree θ around the z-axis.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_1 \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{2}$$

Similarly, get *R2* and *R3*. $R = R_1R_2R_3$

For the process of translation, add a translation vector *t.* Finally get the form of Eq. 1.

3.1.2 Convert Camera Coordinate to Image Coordinate

According to the Fig. 3, $\Delta ABO_c \sim \Delta oCO_c$ and $\Delta PBO_c \sim \Delta pCO_c$ (triangle similarity). Thus, we can get the relational expression:

$$\frac{AB}{{}_OC} = \frac{AO_c}{{}_OO_C} = \frac{PB}{pC} = \frac{X_c}{x} = \frac{Z_c}{f} = \frac{Y_c}{y} \tag{3}$$

$$x = f\frac{X_c}{Z_c}, y = f\frac{Y_c}{y} \tag{4}$$

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{5}$$

3.1.3 Convert Image Coordinate to Pixel Coordinate

On the image plane, the geometric relationship that we established in the previous step is determined by the image coordinate system. The intersection point between image plane and the optical axis (Point *o* in Fig. 3) is used as the origin of the image coordinate system. We cannot get the origin point directly. However, the pixel coordinate system with the upper-left corner of the image plane as the origin is well determined. We introduce the 2-D vector (*u, v*) as the image coordinate. (Eq. 6).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{P_x} & 0 & u_0 \\ 0 & \frac{1}{P_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6}$$

The size of the pixel in world units is (p_x, p_y) and (u_0, v_0) is the displacement between two origin points. Then, combine the conversion Eqs. 1, 5, and 6 to form our final camera calibration model:

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{7}$$

3.1.4 Distortion in Camera Calibration

We also need to consider another important part of the calibration system, the lens. Regular lenses mainly consider radial and tangential distortion. The radial distortion extends outward from the center. Obviously, the distortion and distance have a nonlinear transformation relationship, which can be approximated by polynomials:

$$\begin{cases} x_{\text{distorted}} = x(1 + k_1 r^2 + k_2 r^4 + + k_3 r^6) \\ y_{\text{distorted}} = y(1 + k_1 r^2 + k_2 r^4 + + k_3 r^6) \end{cases} \tag{8}$$

For tangential distortion:

$$\begin{cases} \mathrm{x}_{\text{distorted}} = x + [2p_1 xy + p_2(r^2 + 2x^2)] \\ y_{\text{distorted}} = x + [2p_2 xy + p_1(r^2 + 2y^2)] \end{cases} \tag{9}$$

In the distortion equation, *K1*, *k2*, *k3* are the radial distortion coefficients and p_1, p_2 are the tangential distortion coefficients. $r^2 = x^2 + y^2$ (r is the distortion radius, which is equal to the distance to the origin).

3.2 Event Camera Calibration

The pinhole camera theory is also applicable for the event camera. However, due to its properties—the change of light intensity reaches the threshold to trigger an event and form an event stream—we cannot directly calibrate this event stream. The problem is focused on the acquisition of event images and the need to acquire multiple sets of such a frame to improve accuracy. For our experiment, we mainly introduce the method based on the accumulation structure of the DVP. This requires a module, called accumulator. The DVP can directly use the accumulator module. The module takes events and prints them onto a frame within a fixed time interval. Removal of the information is done with a configurable decay function. We can set the time window to accumulate events before sending out a new frame in milliseconds.

The whole structure (Fig. 4) of the event camera calibration contains these modules: capture, accumulator, and calibration. The capture module contains the DVXplorer

camera, which continuously captures the changes of intensity as output event information. Then the event data enter the accumulator module.

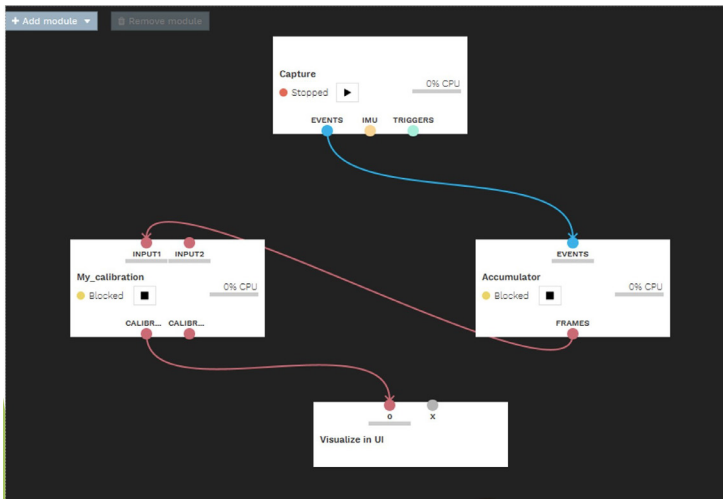

Fig. 4 Event camera calibration pipeline

According to the accumulation algorithm, the output frame image is the input of calibration module. In the calibration process, we detect the corner points of the calibration board (the experiment uses a checkerboard as the calibration board), select the appropriate detected event images, and substitute these images into the calibration theory mentioned previously (Section 3.1, Camera Calibration). Finally, we calculate the intrinsic and extrinsic parameters of the event camera.

In the accumulation module, we need to adjust accumulator settings. The first is the accumulation time (ms). Our experimental method is to stimulate the event trigger by changing the position and brightness of the light source (illuminating the checkerboard with a flashlight), which will not produce high-frequency changes. We also need to set the decay function, which describes how each pixel value returns to the neutral potential. Decay functions include three types: linear, exponential, and step. Our experiment uses the exponential decay function with the decay parameter set as 1,000,000 (μs). The function applies an exponential decay of the form (I is the pixel value after decay process in time t):

$$I = I_0 * \exp(-(\mathrm{t}/\ \mathrm{decayparam}\ )) \tag{10}$$

The information of a frame image requires the contribution of events. When an event arrives at a position (x,y), the pixel value in the frame at (x,y) gets increased or decreased by the value of event contribution, based on the event's polarity. At the same time, the accumulated event information of the pixel will continue to decrease according to the selected decay function (Eq. 10). Eventually in this way, one accumulated event frame image is formed as time changes.

3.3 Registration with Color Image

Stereo camera calibration is usually closely related to image registration. Two cameras, event and regular, require stereo calibration to measure the relative pose between each other. Through the intrinsic and extrinsic parameters of the two cameras, the correspondence between the different images generated by the two cameras can be found.

The previous camera calibration can help improve the accuracy value and promote the next image registration experiment. Image registration, a basic step in computer vision, finds the mapping relationship of an image pixel between two kinds of images. Specifically, for two images in a pair of images, one image (floating image) is mapped to another image (reference image) by searching for geometric transformations or local displacements. Through the registration, the two images correspond to the points of the same position in the space, so as to achieve the purpose of information fusion. This pair of images can be obtained at different times or acquired by different camera sensors.

Our experiment considers images taken by two kinds of cameras in different positions. The first camera is the event camera, and the other camera color, which can obtain ordinary RGB information (red, green, blue [RGB]). Stereo camera calibration is usually closely related to image registration. Two cameras, event camera and color camera, can be co-calibrated to determine the relative position between each other. Through the intrinsic and extrinsic parameters of the two cameras, the correspondence of pixels between the different images generated by the two cameras can be found. Because the acquisition of event images requires accumulation of events within a fixed time interval, the images are in the form of gray-scale images. According to the properties of the event camera, if a certain pixel does not reach the threshold of brightness change within a period of time, it is impossible to obtain the effective information of the pixel. On the other hand, the accumulated image is easily affected by the length of the time interval and intensity of light change during the time period. The frame image may have some information missing. Some image areas will also appear distorted and blurry. Nevertheless, event cameras still have many practical advantages, such as low power consumption and low latency, which is not available in traditional color cameras. In this case, we intend to combine the respective advantages of the two types of images, fuse their image information, perform the registration for two types of images, and find the mapping between the key pixels.

3.3.1 Feature-Based Registration

A popular class of methods of image registration is based on features. These methods have three steps: key point detection and feature description, feature matching, and image transformation. In a word, we select the key points in the two images and associate the equivalent key points. Key points represent important or unique content (corners, edges) in the image, that is, some features. Scale-invariant feature transform (SIFT) is the original algorithm for key point detection. Now, more and more algorithms focus on feature point detection, from SIFT to deep learning. Once the key points are identified in a pair of images, we need to associate or match the corresponding key points in the two images.

However, the event camera cannot directly generate a frame of image. It can only form a special image type through event stream accumulation. This accumulated image is relatively blurred, and some pixels will lack information or have inaccurate output. The feature-based method is not suitable for the event camera. Therefore, our experiment proposes a novel co-calibration registration model.

3.3.2 Calibration-Based Registration

This image registration method is based on the stereo camera calibration. In other words, the prerequisite is that we must have two cameras' information. If we only have images, we cannot do that. In Fig. 5, a point P in the world is imaged on P_1 and P_2 on the image planes of the two different cameras. Our image registration is to find the mapping relationship between P_1 and P_2. In the calibration, we can obtain the intrinsic and extrinsic parameters of the two cameras, respectively. The intrinsic parameters represent the relationship between the world coordinates of the space point and the camera coordinates. Converting the two 3-D coordinates involves rotation and translation. R matrix represents rotation, and t vector represents translation. Then R_1 and t_1 represent the converting between the camera C_1 with the world coordinate system, and R_2 and t_2 represent the converting between the camera C_2 and the world coordinate system.

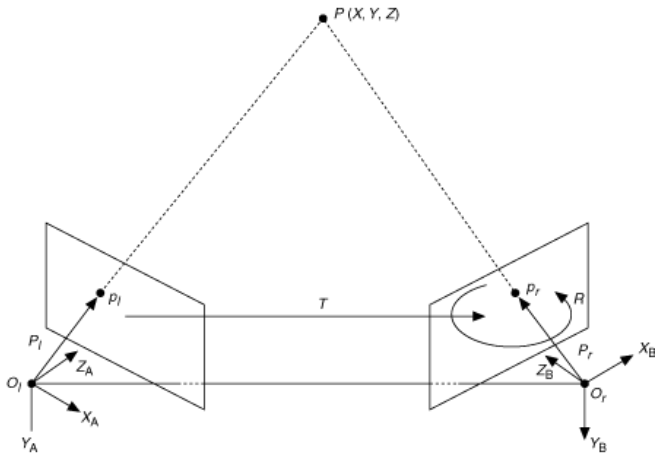

Fig. 5 Stereo camera geometry

$$\vec{C_1} = R_1\overrightarrow{W} + \overrightarrow{t_1} \tag{11}$$

$$\vec{C_2} = R_2\overrightarrow{W} + \overrightarrow{t_2} \tag{12}$$

Combine Eqs. 11 and 12:

$$\vec{C_1} = R_1R_2^{-1}\vec{C_2} + \vec{t_1} - R_2^{-1}\overrightarrow{t_2} \tag{13}$$

Then we get the rotation and translation between these two cameras:

$$R = R_1R_2^{-1}, \vec{t} = \vec{t_1} - R_2^{-1}\vec{t_2} \tag{14}$$

$\vec{C_1}$, $\vec{C_2}$ is the camera coordinate, $\overrightarrow{W}$ is the world coordinate, and we keep the world coordinates unchanged. Thus, if we obtain the extrinsic parameters of the color and event cameras, the relative position between the two cameras can be determined. We want to determine the mapping between the corresponding pixels in the image; that is, find the pixel coordinates of P_1 in the color image and the pixel coordinates of P_2 in the event image, and establish a conversion between these two coordinates.

$$Z_1\begin{bmatrix}u_1\\v_1\\1\end{bmatrix} = \begin{bmatrix}f_x & 0 & u_0 & 0\\0 & f_y & v_0 & 0\\0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}R & t\\0 & 1\end{bmatrix}\begin{bmatrix}x_w\\y_w\\z_w\end{bmatrix} = A_1B_1\overrightarrow{W} \tag{15}$$

$$Z_2\begin{bmatrix}u_2\\v_2\\1\end{bmatrix} = \begin{bmatrix}f_x & 0 & u_0 & 0\\0 & f_y & v_0 & 0\\0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}R & t\\0 & 1\end{bmatrix}\begin{bmatrix}x_w\\y_w\\z_w\end{bmatrix} = A_2B_2\overrightarrow{W} \tag{16}$$

In the experiment, we fix the world coordinate system on the checkerboard. (The plane composed by x-axis and y-axis coincides with the checkerboard plane, and the z-axis is perpendicular to the checkerboard plane). The coordinates of any point on the checkerboard are $Z_w = 0$. Therefore, the original single-point imaging model can be transformed into the following formula:

$$Z\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} (R_a \quad R_b \quad t) \begin{pmatrix} x_w \\ y_w \\ 1 \end{pmatrix} \tag{17}$$

R_a and R_b are the first two columns of the extrinsic matrix; A_1, A_2 is the intrinsic matrix; and B_1, B_2 is the extrinsic matrix. The intrinsic and extrinsic parameter matrices are both in the form of 3 × 3. Combine the two equations of camera calibration modules. Obtain the relation between two pixel-coordinates:

$$\vec{P}_1 = (Z_2/Z_1)A_1B_1B_2^{-1}A_2^{-1}\vec{P}_2 \tag{18}$$

A_1, B_1, B_2^{-1} and A_2^{-1} all can be obtained by calibrations of two cameras. Their product $A_1B_1B_2^{-1}A_2^{-1}$ results in a 3 × 3 homography matrix H. Combine Eqs. 15–18. The final equation can be written as

$$\begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = \frac{Z_2}{Z_1} H \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} = \frac{Z_2}{Z_1} \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} \tag{19}$$

Using the previous formula, eliminate the scale factor Z_2/Z_1:

$$\begin{aligned} u_1 &= \frac{H_{11}u_2 + H_{12}v_2 + H_{13}}{H_{31}u_2 + H_{32}v_2 + H_{33}} \\ v_1 &= \frac{H_{21}u_2 + H_{22}v_2 + H_{23}}{H_{31}u_2 + H_{32}v_2 + H_{33}} \end{aligned} \tag{20}$$

3.3.3 Specific Operations

First, fix the event camera and the color camera and keep them unchanged during the experiment. Then place the checkerboard in front of the two cameras. There is a prerequisite that the checkerboard must completely appear in the view field of both cameras. The event camera requires changes in light intensity to accumulate event stream into image frames. We can change the surrounding light source to illuminate the checkerboard in different directions and stimulate the event camera trigger. The accumulated image may miss some information. To ensure the success rate, the calibration module of DVP detects three to five checkerboard images. Select the best one of them as the output image of the event camera. At the same time, the output image of the color camera is captured. These two images (event image, color image) form an image pair. Then continue to keep the positions of the two cameras. Change the position of the checkerboard, and follow the above steps to obtain 10 to 20 image pairs. According to the co-calibration image registration model, the color camera is used as C_1 (camera 1), and the event camera is used as C_2 (camera 2). The homography matrix can be calculated.

4. Experimental Results

The event camera used in the experiment was the DVXplorer, manufactured by company Inivation, with a resolution of 640 × 480. A FLIR Flea 3 USB camera was used as the color camera with a resolution 1600 × 1200. According to the previous experimental approach, we collected 12 pairs of event images and color images. A checkerboard pattern was used for calibration. The checkerboard was a 9 × 6 square and each square was 12 × 12 mm. The position of the checkerboard was fixed for each pair of images. Its position was changed through rotation and translation. Altogether, 12 pairs of checkerboard images were captured in different poses.

4.1 Color Camera Calibration

We first calibrated the color camera using the MATLAB calibration toolbox to calculate the camera parameters. Figure 6 shows the detected points of the color image. The intrinsic parameter matrix is

$$A_1 = \begin{bmatrix} 2742.0 & 0 & 768.9 \\ 0 & 2742.0 & 586.6 \\ 0 & 0 & 1 \end{bmatrix}$$

The focal length is 2742.0 pixels. The optical center (the principal point) is (768.9, 586.6), in pixels.

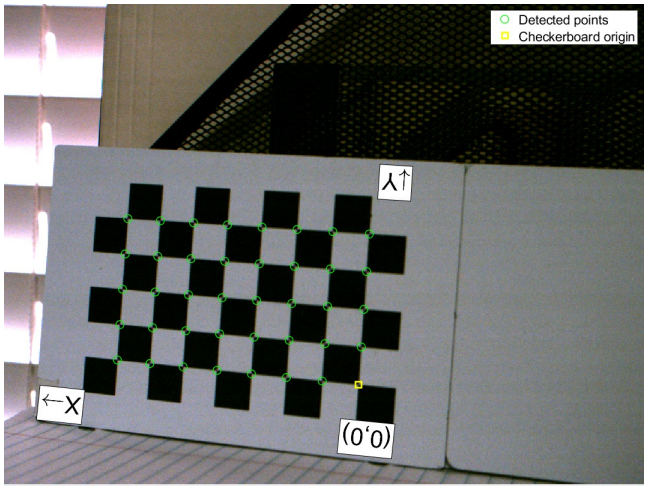

Fig. 6 Detected feature points in the color image

For the calibration evaluation, reprojection errors provide a qualitative measure of accuracy. A reprojection error is the distance between a pattern key point detected in a calibration image and a corresponding world point projected into the same image. The mean reprojection error here of 0.1109 provides an average reprojection

error for these detection images. The reprojection error (0.1109) is less than 1 pixel, which indicates that the calibration experiment has been successfully completed. The acquired camera intrinsic parameters can be used for subsequent registration experiments.

4.2 Event Camera Calibration

The experiment used the DVP to perform the accumulation operation to draw the event stream into a frame of image output as follows: 1) call the calibration module of DVP to detect the points of checkerboard, 2) configure the parameters, and 3) change the light to obtain the accumulated image frames. For the accumulation module, the experiments used exponential decay function with the decay parameter set as 1,000,000 (μs). The accumulation time was 26 ms. Event contribution was 0.04. Through calculation, the intrinsic parameter matrix of the event camera is

$$A_2 = \begin{bmatrix} 598.1 & 0 & 323.0 \\ 0 & 598.2 & 237.9 \\ 0 & 0 & 1 \end{bmatrix}$$

In the same way, the detected points of the event frame are obtained (Fig. 7). The focal length $f_x = 598.1$, $f_y = 598.2$. (323.0, 237.9) is the optical center (the principal point) in pixels. We get the mean reprojection error of 0.1313, which provides an average reprojection error for these detection images. The reprojection error is less than 1 pixel, which indicates that the calibration experiment has been successfully completed. The acquired camera intrinsic parameters can be used for subsequent registration experiments.

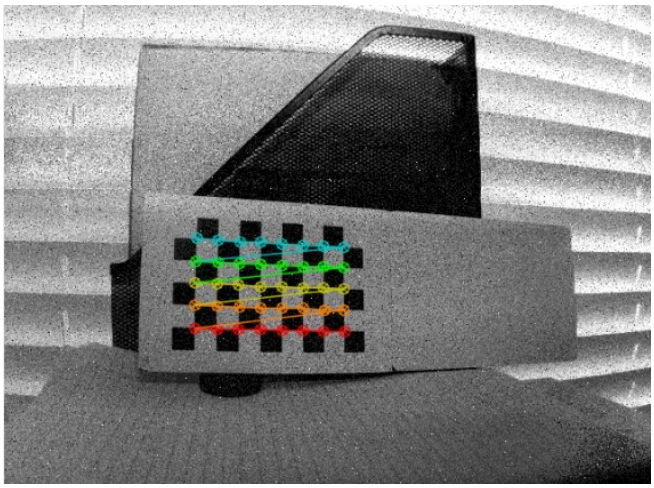

Fig. 7 Detected feature points in the event frame image

4.3 Registration of Color and Event Cameras

With the intrinsic and extrinsic parameters of the two cameras estimated, the homography matrix that transforms the pixel coordinates between the color and event images could be calculated. The camera's intrinsic parameters were fixed, obtained from 12 images. The extrinsic parameters' different values correspond to each image. For each camera, we could generate 12 sets of camera extrinsic parameters. That is, for each pair of images, we could get a homography matrix as the image registration. In this experiment, the first pair of images was used to show the experimental results. The homograph matrix is

$$H = \begin{bmatrix} 4.08 & -0.62 & -228.96 \\ -0.20 & 4.62 & -543.25 \\ 0.00 & 0.00 & 1.19 \end{bmatrix}$$

The corner points of the checkerboard in the event image are registered to the ones in the color image. With the help of MATLAB, we connected the key points of the event image with the corresponding key points of the color image (Fig. 8). The error between the registered points with the actual detection points of the color image can be calculated and used as a standard for registration evaluation:

$$E_{reg} = \sum_{i=1}^{n} (\mathbf{P}_i - H\mathbf{P}_i)/N \tag{21}$$

E_{reg} is the mean registered error. P_i is the key points in event image. H is the homography matrix. On the checkerboard, we detected 40 key points (N = 40). For the first pair of images, mean registered error is 2.1688 pixels. The low-resolution event image maps to the high-resolution color image, which causes the corresponding error generated by the camera calibration to increase several times. In this case, the registered error will become significantly larger.

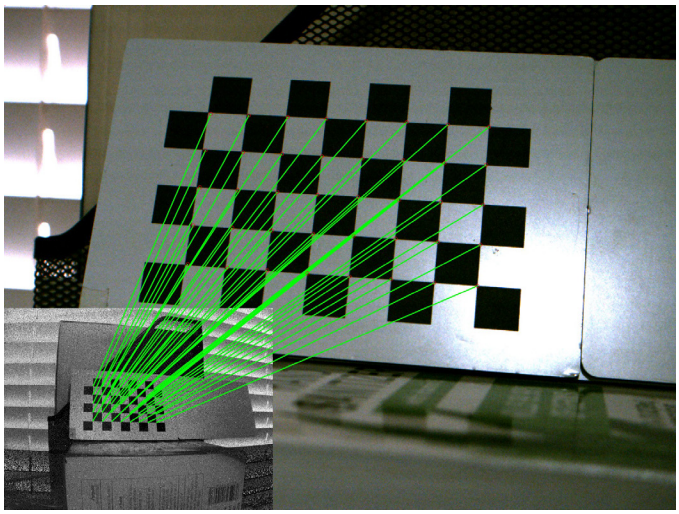

Fig. 8 Mapping key points between the event frame image and the color image

We can combine the event image with the color image and replace part of the checkerboard plate in the color image with the mapping of the relevant pixels of the event image. The registration fusion image in Fig. 9 is created by mapping the image of the event camera to the red channel in the color image (event image is single channel; RGB color image has three channels).

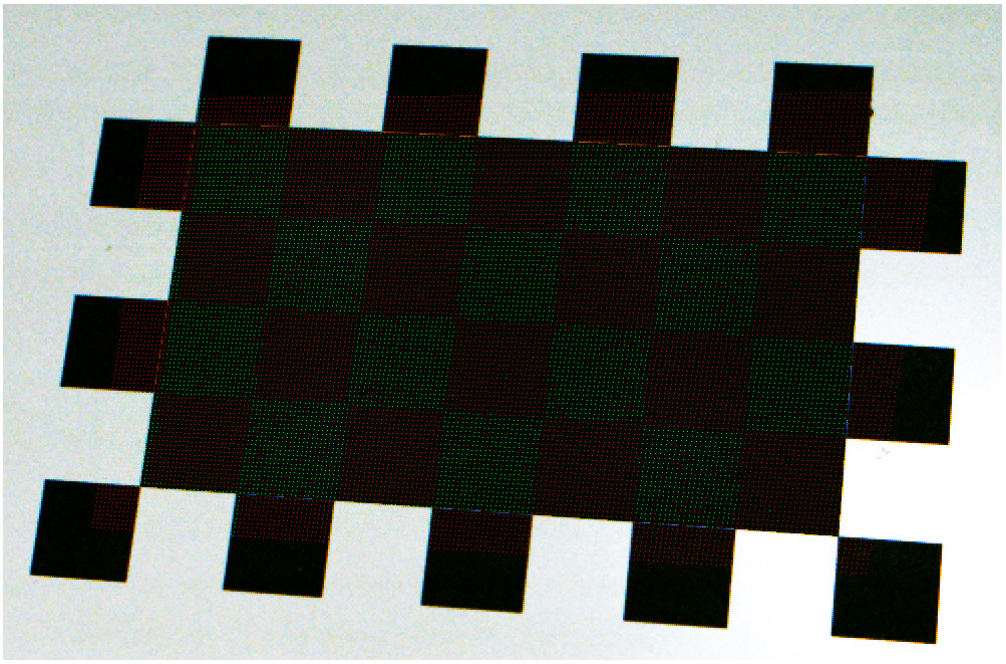

Fig. 9 Event frame image and color image registration result

5. Conclusions

The experiments show we successfully completed the calibration of the event camera and the color camera, and used the co-calibration method to find the homography matrix, which corresponds to the conversion between the key points of the two images. We completed the registration between the event image and the color image. The error generated by this registration method is 2.1688, at the level of a single pixel, which is more accurate than the regular feature-based registration method. Figure 9 shows that the missing part of the color image can be supplemented by the event image, which verifies that our registration method is effective.

This registration method has certain constraints, as it requires two cameras that are pre-calibration determined. Further, during the derivation process of the homography matrix, the z-axis of these key points is 0 in the world coordinate. Therefore, this homography matrix is only suitable for mapping when the key points are on the same plane.

Our experiment shows that by registering the event and color images, we are able to take advantage of their complementary information and develop other useful applications. For example, when imaging in low-light conditions, we can use the registered event image to reduce motion blur in the color image.

6. References

1. Ciermochowski MF, inventor; Holley Carburetor Company, assignee. Photo-optic transducer using apertured shade and moveable shutter. United States patent US 3628024A. 1971 Dec 14.

2. Gallego G, Delbruck T, Orchard G, Bartolozzi C, Taba B, Censi A, Leutenegger S, Davison AJ, Conradt J, Daniilidis K, et al. Event-based vision: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020. doi: 10.1109/TPAMI.2020.3008413.

3. Chin T-J, Bagchi S, Eriksson A, van Schaik A. Star tracking using an event camera. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 June 16–20; Long Beach, CA.

4. Ghosh R, Mishra A, Orchard G, Thakor NV. Real-time object recognition and orientation estimation using an event-based camera and CNN. Proceedings of the IEEE 2014 Biomedical Circuits and Systems (BIOCAS); 2014 Oct 22–24; Lausanne, Switzerland.

5. Weikersdorfer D, Hoffmann R, Conradt J. Simultaneous localization and mapping for event-based vision system. International Conference on Computer Vision Systems (ICVS) 2013; 2013 Sep; St. Petersburg, Russia. doi: 10.1007/978-3-642-39402-7_14.

6. Pan L, Scheerlinck C, Yu X, Hartley R, Liu M, Dai Y. Bringing a blurry frame alive at high frame-rate with an event camera. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 June 16–20; Long Beach, CA. p. 6820–6829.

7. Clady X, Leng S-H, Benosman R. Asynchronous event-based corner detection and matching. Neural Networks. 2015;66:259–275.

8. Kim H. Real-time visual SLAM with an event camera [PhD thesis]. [London (UK)]: Imperial College of London; 2017.

9. Song R, Jiang Z, Li Y, Shan Y, Huang K. Calibration of event-based camera and 3D LiDAR. 2018 WRC Symposium on Advanced Robotics and Automation (WRC SARA); 2018 Aug 16; Beijing, China.

10. Dominguez-Morales M, Jimenez-Fernandez A, Jimenez-Moreno G, Conde C, Cabello E, Linares-Barranco A. J. Bio-inspired stereo vision calibration for dynamic vision sensors. IEEE Access. 2019;7:138415–138425.

11. Galanis M. DVS event stream registration [bachelor's thesis]. [Munich (Germany)]: Technical University of Munich; 2016.

12. Kalber F. A probabilistic method for event stream registration [bachelor's thesis]. [Munich (Germany)]: Technical University of Munich; 2016.

13. Zitova B, Flusser J. Image registration methods: a survey. Image Vision and Computing. 2003;21(11):977–1000.

14. Huang L, Li Z. Feature-based image registration using the shape context. International Journal of Remote Sensing. 2010;31(8):2169–2177.

15. Weikersdorfer D, Adrian DB, Cremers D, Conradt J. Event-based 3D SLAM with a depth-augmented dynamic vision sensor. Proceedings of the 2014 IEEE International Conference on Robotics & Automation (ICRA); 2014 May 31–June 7; Hong Kong, China. p. 359–364.

16. Aberman K, Liao J, Shi M, Lischinski D, Chen B, Cohen-Or D. Neural best-buddies: sparse cross-domain correspondence. ACM Transactions on Graphics. 2018;37(4):69.

17. Lindberg DC. A reconsideration of Roger Bacon's theory of pinhole images. Archive for History of Exact Sciences. 1970;6:214–223.

18. Zhang Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;22(11):1330–1334.

List of Symbols, Abbreviations, and Acronyms

2-D	two-dimensional
3-D	three-dimensional
ARL	Army Research Laboratory
DEVCOM	US Army Combat Capabilities Development Command
DVP	Dynamic Vision Platform
DVS	Dynamic Vision Sensor
LiDAR	light detection and ranging
RGB	red, green, blue
SIFT	scale-invariant feature transform
USB	Universal Serial Bus

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLD DCI TECH LIB
2 (PDF)	DEVCOM ARL FCDD RLC CI S S YOUNG FCDD RLV A P OSTEEN