Paradigm Academic Press
Innovation in Science and Technology
ISSN 2788-7030
NOV. 2023 VOL.2, NO.6

Binocular Vision-Based Fire System for University Laboratories

Pengfei Li[1], Qi Yu[1], Pengjun Zheng[1], Meimei Pan[1] & Liqun Zhou[1]

[1] Chengdu University of Technology, China

Correspondence: Pengfei Li, Chengdu University of Technology, China.

doi:10.56397/IST.2023.11.07

Abstract

With the frequent occurrence of fire accidents in college laboratories, the study of fire localization systems has become particularly important. This paper proposes a fire localization system for college laboratories based on binocular vision. The MATLAB toolbox is utilized to complete the calibration of binocular camera and the stereo correction of binocular image. Then combined with the YOLOv5 target detection algorithm for target detection and recognition of the image, to obtain the type of target and the border coordinate information. Finally, according to the principle of binocular vision ranging, the depth information of the target object is calculated to realize the accurate positioning of the fire source. The system proposed in this paper can effectively localize fires in university laboratories. The system has an important application value, which can quickly respond and take corresponding rescue measures in fire accidents and improve the safety factor of university laboratories.

Keywords: binocular vision, camera calibration, target detection, YOLOv5 algorithm, binocular vision ranging

1. Introduction

The frequent occurrence of laboratory fires in colleges and universities will not only seriously jeopardize the life safety of students, but also cause certain economic losses. Therefore, an effective fire detection model needs to be developed to detect the occurrence of fires and the subsequent prevention of fires. The binocular vision-based fire localization system described in this paper will provide more convenient functions in this regard.

1.1 Background and Context

Fire Risks in College Laboratories, as important places for scientific research and teaching, have fire risks. Laboratories often store a large number of flammable and explosive substances and are equipped with a large number of instruments and equipment, such as circuit equipment, chemical reagents, etc., all of which may cause a fire. Although university laboratories usually take a series of fire prevention measures, such as installing fire extinguishing equipment, setting up sprinkler systems, etc., these means cannot completely avoid the occurrence of fire. Universities also lack convenient and effective means of fire prevention. Therefore, it is imperative to research and develop fire detection systems in university laboratories for fire prevention and detection purposes.

Research on fire detection systems in university laboratories continues to make progress. A number of research organizations and universities are actively exploring new fire detection technologies, which also provides a research background for studying fire detection systems in university laboratories.

1.2 Importance

Laboratory fires can lead to injuries and property damage. The risk of injury or death can be reduced and the safety of laboratory staff can be safeguarded with a fire detection system.

On the other hand, laboratory equipment is costly and often difficult to replace. Fire damage to experimental equipment not only causes economic loss, but may also have a negative impact on scientific research. By establishing an effective fire detection system. Fire can be detected in time and appropriate measures can be adopted to minimize all kinds of losses.

2. Binocular Vision Ranging Algorithm

The principle of the eye camera is similar to that of the human eye. The human eye is able to perceive the distance of an object due to the difference between the images presented by the two eyes of the same object, also known as “parallax”. The farther away the object is, the smaller the parallax is; conversely, the larger the parallax is. The size of the parallax corresponds to the distance between the object and the eyes. Binocular vision ranging is the use of binocular cameras to observe the target, by calculating the parallax of the same target in the left and right views, combined with the parameters of the camera itself, the depth information of the target is deduced.

It can be seen that binocular vision ranging is a method of ranging based on binocular stereo vision, in which the depth information of the target is deduced by calculating the parallax of the target in the left and right camera images.

2.1 Conversion of the Coordinate System

2.1.1 Three Coordinate Systems

(1) World coordinate system

The world coordinate system is the reference system for the target object. In real space, in order to determine the specific coordinates of a point in three-dimensional space, it is necessary to utilize the world coordinate system as a reference coordinate system to represent space, generally a right-handed coordinate system. Regardless of how an object moves or rotates, the position and direction of its coordinate system remain constant.

(2) Camera coordinate system

The camera coordinate system is a three-dimensional coordinate system with the camera as the reference, with the center of the camera’s optical system (lens center) as the coordinate origin and the optical axis as the z-axis. The camera coordinate system is a bridge between the pixel coordinate system and the world coordinate system, and the position and direction of its coordinate system change with the movement or rotation of the camera.

(3) Image coordinate system

An image coordinate system is a coordinate system that expresses the position of pixels in physical units, reflecting the specific dimensions of the points in the image, and is also usually established as a coordinate system with the origin at the upper left of the image.

(4) Pixel coordinate system

A pixel coordinate system is the position of a pixel in an image, established in an image coordinate system with pixels as the unit of the coordinate system, usually established with the origin at the upper left of the image, generally in front of the camera coordinate system.

2.1.2 Conversion of Three Coordinate Systems

The projection of the target object on the image plane is the link between the object in the camera coordinate system and the image coordinate system, and ultimately need to restore the position coordinates in the real world need to be calibrated to get the camera internal and external parameter information, to realize the conversion between the four coordinate axes to complete the depth measurement of the target object. Among them, the relationship between the four coordinate systems is as follows:

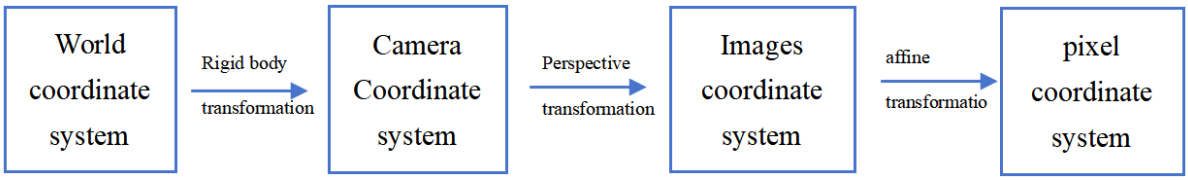

Figure 1. Conversion diagram of coordinate system

If the coordinates of the center of the image are (u_0, v_0), the relation from the imaging plane to the image plane is:

$$\begin{cases} u - u_0 = \alpha_x x \\ v - v_0 = \alpha_y y \end{cases}$$

Among them α_x, α_y is the magnification from the imaging plane to the image plane on the X, Y axes.

According to the similarity theorem there is:

$$\begin{cases} u - u_0 = \alpha_x f\frac{X}{Z} = f_x\frac{X}{Z} \\ v - v_0 = \alpha_y f\frac{Y}{Z} = f_y\frac{Y}{Z} \end{cases}$$

where f_x is the equivalent focal length in the X, Y direction f_y is the equivalent focal length in the X, Y direction, and f is the focal length of the lens.

The image coordinate system is related to the world coordinate system by the equation:

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} R & t \\ 0^t & 1 \end{bmatrix}\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_1M_2\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

where M_1 is the camera internal parameter matrix, and M_2 is the camera external parameter matrix.

In addition, there exists an error that cannot be eliminated because the manufacturing process cannot be perfected — aberration (two types of aberration are usually considered: radial and tangential aberration). Aberrations affect the accuracy of the image and need to be corrected in camera calibration and image processing.

2.2 Principle of Binocular Vision Measurement

Binocular stereo vision ranging utilizes two viewpoints of a binocular camera to observe a target, and by calculating the parallax of the target in the left and right views, the three-dimensional coordinate information of the target is deduced. In binocular vision system, the target is projected onto the image plane by imaging. By calculating the parallax, the three-dimensional coordinates of the target can be deduced, the principle of which is shown in the figure below.

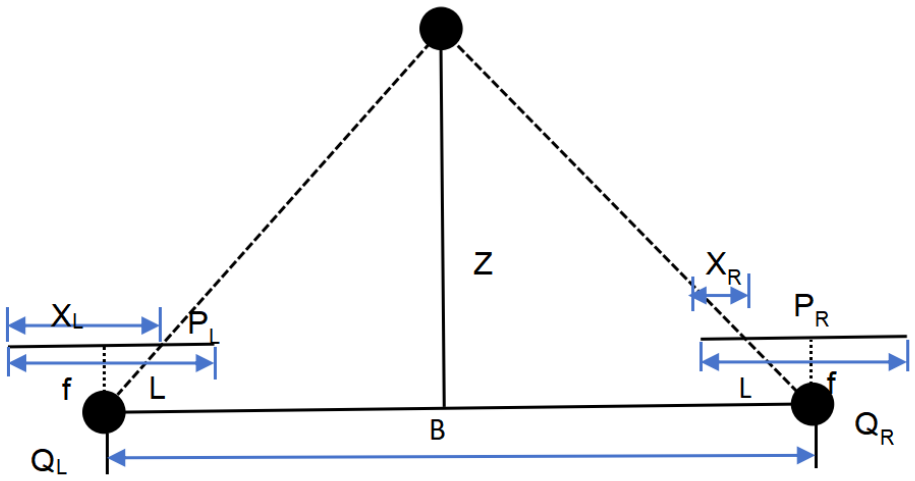

Figure 2. Schematic diagram of binocular ranging principle

B for the baseline is the distance between the two camera projection center of the line, PR and PL respectively three-dimensional space at any point P in the left and right camera imaging point, then the point P in the left and right camera parallax is:

$$d = |x_L - x_R|$$

From the theory of similar triangles it follows that the distance Z from the point P to the plane of the center of projection is:

$$Z = \frac{b \times f}{d}$$

As a result, if the parallax of a point is known, the depth scene information of that point can be known. Based on the imaging of any point in 3D space in the camera and its parallax on different images and camera parameters,

the 3D coordinates of the point can be known.

2.2.1 Steps in Binocular Ranging

Binocular ranging actually operates in 4 steps: camera calibration — binocular calibration — binocular matching — calculation of depth information.

(1) Camera calibration: Camera calibration is inside and outside the parameter matrix in order to establish a mathematical simulation model of the camera, in order to eliminate aberrations, correct the size and determine the actual size of the object in the image and the specific location, to achieve the matching of the image with reality. For the calibration of fire recognition mainly Tsai calibration and Zhang Zhengyou calibration method are used in the traditional calibration method. Among them, Zhang Zhengyou calibration method has the advantages of simplicity, high precision and good robustness, which is widely used for 3D reconstruction in the field of computer vision, and it is a method often used for camera calibration in the field of computer vision.

Zhang Zhengyou calibration method is to shoot the calibration object at different angles, take out its corner points, resolve the camera's distortion parameters and internal and external parameters, and optimize the parameters according to the great likelihood method.

Camera due to the characteristics of the optical lens makes the image has a radial aberration, can be determined by three parameters k1, k2, k3; due to assembly errors, the sensor and the optical lens is not completely parallel to each other, so the image has a tangential aberration, can be determined by two parameters p1, p2. The calibration of a single camera is mainly to calculate the internal parameters of the camera (focal length f and imaging origin, five aberration parameters (generally only need to calculate k1, k2, p1, p2, for the fisheye lens and other radial aberration is particularly large before the need to calculate k3)) and external parameters (the world coordinates of the calibrated object). In contrast, binocular camera calibration requires not only deriving the internal parameters of each camera, but also measuring the relative position between the two cameras (i.e., the rotation matrix R of the right camera with respect to the left camera, and the translation vector t) through calibration.

(2) Binocular correction: Binocular correction is based on the camera calibration of the monocular internal parameter signal values (focal length, imaging origin, distortion coefficient) and binocular relative position relationship (rotation matrix and translation vector), respectively, the left and right view of the aberration and row alignment elimination, so that the left and right images of the imaging origin position is the same, the two sides of the camera optical axis is parallel to the front and back of the image planes are coplanar, and the rows of the polarization line are equal. In this way, any point of a pixel must have the same row number as its counterpart in another pixel, and only a one-dimensional lookup in that row is needed to match the neighboring points.

(3) Binocular Matching: The function of binocular matching is to match the corresponding image points of the same scene on the left and right views, and the purpose of doing so is to get the parallax map.

(4) Calculate the depth information: get the parallax data, through the above principle in the formula can easily calculate the depth information.

3. Flame Target Recognition Based on YOLOv5

YOLO artificial neural network is an algorithm for target detection using convolutional neural networks. It can directly predict the relationship between input and output variables — learning a model parameter from a given training dataset to make predictions about unknown images. YOLO has a wide range of applications in target detection and image classification applications. YOLOv5 is based on the optimization and improvement of YOLO. YOLOv5 has ultra-fast detection speed and small and compact network model. It can detect video images in real time.

3.1 Brief Description of the Principle of YOLOv5 Target Detection

The model structure of YOLOv5 consists of the following four main parts: Input end, Backbone, Neck, and Head.

3.1.1 Input End

When the image enters the input end, it usually goes through an image preprocessing phase. This part mainly consists of Adaptive Anchor Frame Calculation, Mosaic, and Adaptive Image Scaling.

Mosaic is a data enhancement approach that enables stitching of different images by random scaling, random cropping, and random arranging. It can avoid overfitting caused by too few images in the dataset. At the same time, Mosaic can improve the training speed of the model and network accuracy.

In target detection algorithms, anchor frames are some predefined rectangular frames defined on the input image for detecting targets of different sizes. Adaptive anchor frames are used to automatically calculate the most suitable anchor frame parameters for the input image by learning, which improves the accuracy

and robustness of target detection. YOLOv5 uses an adaptive anchor frame calculation method called ATSS (Adaptive Training Sample Selection). ATSS can automatically select the positive and negative samples based on the match between the sample and the anchor frame (i.e., the cross concurrency ratio IoU), it can automatically select positive and negative samples, which can effectively improve the detection accuracy without bringing additional computation and parameters.

Before YOLOv5, the common way for target detection models is to uniformly scale the original image to a standard size. However, this approach can lead to image distortion or loss of some information. Therefore, Adaptive Image Scaling (AIS) is proposed in YOLOv5. It is an image scaling method based on the target scale, which can adaptively scale the size of the target image to adapt to the detection of different scale targets and ensure that the target is not deformed. It is also known as uniform scaling using gray fill — scaling the image in equal proportions and filling the blanks with gray.

For example, to adaptively scale the original 800x400 image to the original size of 200x200:

The aspect ratio between the two dimensions is first calculated to obtain two scaling factors of 0.25 and 0.5. Choose the smaller scaling factor for scaling;

Then according to the scaling factor to calculate the size of the original image after scaling by equal proportion, the scaled size is 200x100. At this time, the length becomes 200, the width becomes 100.

3.1.2 Backbone

Backbone is usually a network of some high-performing classifier species used to extract some generalized feature representations. The CSPDarknet53 structure and the Focus structure are mainly used as the benchmark networks in YOLOv5.

The Focus structure is a convolutional neural network layer used for feature extraction. The Focus structure compresses and combines the information in the input feature map to extract a higher level representation of the features. An example is shown in the following figure:

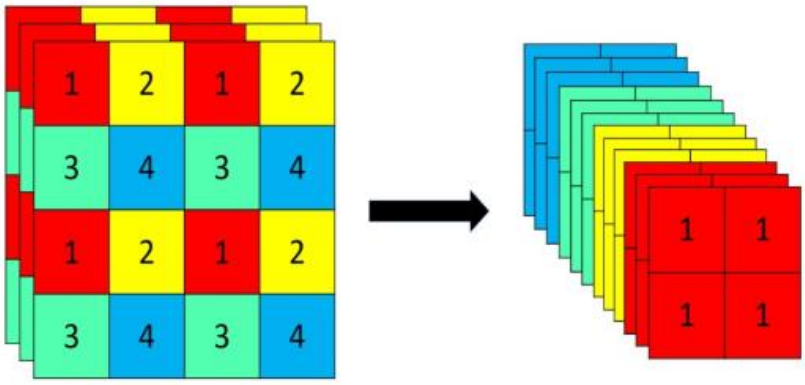

Figure 3. Schematic diagram of the image

Description: in YOLOv5s, the input 608x608x3 image is sliced using a slicing operation, which first turns it into a 304x304x12 feature map, and then undergoes a convolution operation with 32 convolution kernels, which finally turns it into a 304x304x32 feature map.

The CSP (Cross Stage Partial) structure can effectively reduce the network parameters and computation, while improving the efficiency of feature extraction.

The core idea of the CSP structure is to split the input feature map into two parts, one part is processed by a small convolutional network (called a sub-network) and the other part is directly processed in the next layer. The feature maps obtained from the two parts are later stitched together and used as input for the next layer. This structure combines low-level detailed features with high-level abstract features to improve the efficiency of feature extraction.

3.1.3 Neck

Neck is usually in the middle of the whole structure, which can be utilized to further improve the diversity and robustness of the features. YOLOv5 adopts the FPN +PANet structure for multi-scale fusion. The FPN (Feature Pyramid Network) continuously reduces the image by top-down sampling, and improves target detection by fusing the features of high and low layers; the PANet (Path Aggregation Network) makes full use of the shallow

layer of the network for segmentation by bottom-up feature fusion, and the top layer can also receive the location information from the bottom layer. The PANet is a bottom-up feature fusion, which makes full use of the shallow features of the network for segmentation, and the top layer can also receive the position information brought by the bottom layer.

3.1.4 Head

It is used to complete the output of the target detection model. This module mainly uses loss function with NMS (Non-maximum suppression). The loss function alleviates the problem of class imbalance in target detection and improves the performance of the model. The loss of YOLOv5 consists of three main parts: Classes loss, Confidence loss, and Location loss. In the target detection task, an object may be detected by multiple prediction frames, and in order to avoid multiple detections of the same object, the duplicate prediction frames need to be filtered, and this process is Non-maximum suppression (NMS).

3.2 Datasets

The program is written in python language and the dataset is built by labeling the images with label tool. Multiple runs of the resulting model can improve the precision and recall of the model. Common datasets include PASCAL VOC and MS COCO. And the data such as precision, recall, mAP@0.5, mAP@0.5:0.95 can be obtained.

3.3 Performance Indicators

3.3.1 Detection Accuracy

The metrics for detecting model accuracy can be measured by Precision, Recall, AP, mAP, etc.

(1) Confusion matrix

The confusion matrix summarizes the predictions for a classification problem. The number of correct and incorrect predictions is summarized using count values and broken down by class.

Table 1. Schematic diagram of the confusion matrix

		Prediction	
		Positive	Negative
Actual	True	TP	FN
	False	FP	TN

Description: The table breaks down the prediction results into four categories. T and F denote the rightness or wrongness of the prediction, and P and N denote the outcome of the prediction. TP denotes that samples that are actually in the positive category are predicted to be in the positive category; FP denotes that samples that are actually in the positive category are predicted to be in the negative category; FP denotes that samples that are actually in the negative category are predicted to be in the positive category; and TN denotes that samples that are actually in the negative category are predicted to be in the negative category.

The corresponding formula is as follows:

Precision = TP /(TP + FP)

Recall = TP /(TP + FN)

Accuracy = (TP+TN)/(TP+TN+FP+FN)

$$\text{F1-score}=\frac{2\times Precision\times Recall}{Precision+Recall}$$

Precision denotes the proportion of correctly predicted detection frames (number of positive samples) among all the detection frames predicted by the model; Recall denotes the proportion of correctly predicted detection frames among all the detection frames predicted by the model to the actual true frames. F1-score is used to comprehensively judge a model.

1) IoU

IoU is the ratio of the overlapping range of the bounding box to the total range of the bounding box. The category of the prediction result can be judged by setting the threshold of IoU.

2) AP and mAP

Ap is a measure of the effectiveness of the detection of a certain category. According to different confidence and IOU thresholds, it corresponds to different accuracy and recall, and then the area of the two-dimensional graph composed of accuracy and recall is the Ap value. The average value of Ap for different categories is mAP, which is a measure of the detection effect of multiple categories.

$$\mathrm{m}AP = \frac{\sum AP}{N(\mathrm{classes})}$$

4. Fusion of Two Algorithms

By combining the algorithmic program of yoov5 with the program of the binocular vision algorithm, and by acquiring real-time images through the binocular camera, it is possible to display the detection and positional information of the flame target on the image.

Acknowledgements

We would like to thank our supervisor, without his careful guidance and selfless help, we could not have completed this thesis. His professional knowledge and experience played an important role in promoting our research. We would also like to thank him for his patient guidance and motivation for me, which filled us with confidence and motivation. Special thanks to our friends for their support and encouragement. They gave us unlimited courage and motivation when we encountered difficulties and setbacks, and were a strong backing for us to move forward.

We would like to express our heartfelt thanks to all those who have helped us. Your help and support meant a lot to us and enabled us to complete this thesis. We hope that our research can make some contribution to the development of fire localization systems in laboratories. Thank you all!

References

Gao R., (2015). Research on infrared binocular vision localization system for tunnel fire. Chang'an University.

Jihui Huang, (2023). Research and application of stereo matching algorithm based on binocular vision system. Guizhou University. DOI:10.27047/d.cnki.ggudu.2022.002269.

Shina Jia, (2023). Research on small target detection algorithm based on improved YOLOv5. Nanchang University. DOI:10.27232/d.cnki.gnchu.2022.004449.

Wanyue Zhao, (2022). Research on target detection algorithm based on YOLOv5. Xi'an Electronic Science and Technology University. DOI:10.27389/d.cnki.gxadu.2021.002918.

Yin WX., (2022). Research on the calibration method of binocular vision system. Xi'an University of Technology. DOI:10.27398/d.cnki.gxalu.2021.000719.