

Article

A Binocular Stereo-Imaging-Perception System with a Wide Field-of-View and Infrared- and Visible Light-Dual-Band Fusion

Youpan Zhu ^{1,2}, Dan Zhang ³, Yongkang Zhou ^{2,4}, Weiqi Jin ^{1,*}, Lingling Zhou ³, Guanlin Wu ¹ and Yong Li ⁵

¹ MOE Key Laboratory of Optoelectronic Imaging Technology and System, Beijing Institute of Technology, Beijing 100081, China; zhuyubit@aliyun.com (Y.Z.); 3120220595@bit.edu.cn (G.W.)

² Kunming Institute of Physics, No. 31, Jiaochang East Road, Wuhua District, Kunming 650223, China; zyk1120102464@163.com

³ Yunnan North Optical & Electronic Instrument Co., Ltd., No. 300, Haikou Town, Xishan District, Kunming 650114, China; zhangdanynoi@163.com (D.Z.); lynzhll@163.com (L.Z.)

⁴ School of Life Science, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China

⁵ Chengdu Zhongke Information Technology Co., Ltd., No. 1369 Kezhi Road, Xinglong Street, Tianfu New District, Chengdu 610042, China; scl2012@163.com

* Correspondence: jinwq@bit.edu.cn

Abstract: With the continuous evolution of autonomous driving and unmanned driving systems, traditional limitations such as a limited field-of-view, poor ranging accuracy, and real-time display are becoming inadequate to satisfy the requirements of binocular stereo-perception systems. Firstly, we designed a binocular stereo-imaging-perception system with a wide-field-of-view and infrared- and visible light-dual-band fusion. Secondly we proposed a binocular stereo-perception optical imaging system with a wide field-of-view of 120.3°, which solves the small field-of-view of current binocular stereo-perception systems. Thirdly, For image aberration caused by the wide-field-of-view system design, we propose an ellipsoidal-image-aberration algorithm with a low consumption of memory resources and no loss of field-of-view. This algorithm simultaneously solves visible light and infrared images with an aberration rate of 45% and 47%, respectively. Fourthly, a multi-scale infrared- and visible light-image-fusion algorithm is used, which improves the situational-awareness capabilities of a binocular stereo-sensing system in a scene and enhances image details to improve ranging accuracy. Furthermore, this paper is based on the Taylor model-calibration binocular stereo-sensing system of internal and external parameters for limit correction; the implemented algorithms are integrated into an NVIDIA Jetson TX2 + FPGA hardware framework, enabling near-distance ranging experiments. The fusion-ranging accuracy within 20 m achieved an error of 0.02 m, outperforming both visible light- and infrared-ranging methods. It generates the fusion-ranging-image output with a minimal delay of only 22.31 ms at a frame rate of 50 Hz.

Keywords: binocular stereoscopic-perception system; wide field-of-view; infrared and visible light fusion; image distortion



Citation: Zhu, Y.; Zhang, D.; Zhou, Y.; Jin, W.; Zhou, L.; Wu, G.; Li, Y. A Binocular Stereo-Imaging-Perception System with a Wide Field-of-View and Infrared- and Visible Light-Dual-Band Fusion. *Sensors* **2024**, *24*, 676. <https://doi.org/10.3390/s24020676>

Academic Editor: Felipe Jiménez

Received: 28 November 2023

Revised: 28 December 2023

Accepted: 11 January 2024

Published: 21 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the field of autonomous driving, the primary methods for perception systems currently include millimeter-wave radar perception, LiDAR perception, millimeter-wave radar, and LiDAR fusion with visual perception, monocular vision, and binocular stereo vision. Radar- and vision sensor-fusion methods involve the calibration of multiple sensors in both time and space, transforming their acquired data into a common coordinate system, followed by information processing. The authors in [1] used the radar sensor and visual sensor approaches to investigate the vehicle environment. This approach is more accurate in terms of the information obtained but it suffers from the complexity of calibrating multiple sensors with inconsistent sensing ranges and becomes increasingly expensive as

accuracy increases. Monocular stereo vision utilizes a single camera to project the three-dimensional world onto a two-dimensional plane, resulting in a loss of depth information. Monocular stereo vision systems assume static ground and solve for a dynamic camera pose to achieve distance-measurement functionality [2–4]. R. Mur-Artal et al. [5] introduced the ORB-SLAM monocular simultaneous localization and mapping (SLAM) system, which exhibits robustness against significant motion noise and allows for wide baseline-loop closure and relocalization but requires further improvements in depth accuracy. Raul Mur-Artal et al. [6] presented ORB-SLAM2, a monocular SLAM system that employs bundle adjustment to enhance depth accuracy and features a lightweight localization mode for the efficient reutilization of disabled maps. Carlos Campos et al. [7] introduced ORB-SLAM3, which employs comprehensive short-term, medium-term, and long-term data associations to achieve zero drift in mapped areas. Compared to other methods, monocular stereo vision systems offer the advantage of being small and lightweight. However, such systems can only capture two-dimensional images and cannot determine distances to objects, leading to issues with low ranging accuracy and poor environmental adaptability. Binocular stereo vision, on the other hand, emulates the human visual system and leverages the principle of disparity to acquire target features from different positions using detectors. This approach establishes the relationship between corresponding pixel points in the image based on the three-dimensional geometric positions of certain features on the spatial-target surface. This process involves constructing a binocular stereo vision-imaging model to obtain model parameters [8,9]. Binocular stereo vision is also referred to as passive binocular depth sensing and offers improved robustness in depth computation compared to monocular stereo vision. This technology also enables three-dimensional object recognition. Various methods exist for obtaining real-world depth information in binocular stereo perception systems, such as passive stereo [10], active stereo [11], time-of-flight imaging [12], and defocus depth [13]. In this paper, the passive stereo method of binocular stereo systems is employed.

With the advancement of industrial automation and machine vision, binocular stereo perception systems have been widely researched and applied in various fields, including vehicle driving [14], intelligent transportation [15], 3D reconstruction [16], virtual reality [17], surveying, and rapid positioning [18]. In the realm of autonomous and assisted driving [19], the first assisted-driving vehicle based on a binocular stereo vision system was developed in Japan. This system processes images captured by two onboard cameras. Subsequently, in Europe and the United States, autonomous driving vehicles were developed based on binocular stereo vision. Notable examples include the NavLab series developed at Carnegie Mellon University in the United States [20] (NavLab-1, NavLab-5, and NavLab-11). These vehicles utilize binocular stereo vision for road environment detection and focus on addressing challenging visual perception issues in complex environments. Silicon Valley chip company Ambarella has developed binocular ADAS (Advanced Driver Assistance Systems) and autonomous driving chips [21], as well as binocular stereo-specific chips and solutions tailored to binocular vision systems, which act as valuable complements to integrated chips. These specialized chips can handle a portion of the perception tasks at the edge, creating a smaller and more efficient perception-decision loop.

With the evolution of deep learning, many researchers have proposed more efficient and accurate binocular stereo algorithms. Sun et al. [22] introduced a multi-path Viterbi (MPV) multi-scale fast stereo matching algorithm for generating dense disparity information. Jure Bontar and others [23] trained a convolutional neural network (CNN) to predict the matching degree between two image patches and used this CNN for binocular stereo matching. The authors incorporated left–right consistency checks to eliminate errors in occluded areas. However, their method was found to be ineffective in textureless regions. Li et al. [24] introduced Stereo R-CNN, which takes binocular images as network inputs. The authors extended Faster R-CNN to simultaneously detect and associate objects in left and right images. Through the binocular FPN network, the authors predicted object key points and sizes to generate coarse 3D object detection boxes. Finally, the authors used

the region-based photometric correspondence method to refine fine-grained 3D detection boxes. Tang et al. [25] used the YOLOv4-tiny model to generate boundary frameworks and employed an adaptive stereo matching approach based on these boundaries. This approach meets real-time detection requirements and exhibits high localization stability and robustness under varying lighting conditions.

Autonomous driving vehicle is a kind of autonomous and automatic unmanned vehicle, the vehicle not only needs to grasp the dynamic situation of nearby vehicles and obstacles in real-time, but also needs to deal with the complex environment of night, haze, and obstacles. The traditional driving system based on visible light is easy to fail in the area of poor lighting conditions, unable to meet the all-weather needs of automatic driving vehicles, infrared images are not subject to the interference of illumination, sun glare, can penetrate smoke, haze, and other characteristics. The perception ability of automatic driving vehicles can be further improved by fusion of visible light and infrared. Current autopilots commonly use small field-of-view cameras, usually with a field-of-view of 40–60°, which limits the system's perception of the external environment, and the field-of-view and perception can be further improved to a wide-field-of-view technology with infrared- and visible light fusion [26].

The field of autonomous driving places significant emphasis on important metrics such as the field-of-view, ranging accuracy, and real-time display in binocular stereo perception systems. This paper integrates the characteristics of visible light and infrared sensors to design a wide-field-of-view binocular stereo imaging perception system using infrared- and visible light-fusion. This system combines a wide-field-of-view binocular stereo optical imaging system with a low-storage image distortion correction algorithm to achieve imaging with a wide field-of-view and minimal distortion. Focusing on the features of visible light and infrared images, this study employs a multi-scale fusion approach for infrared- and visible light images. This work explores a straightforward yet precise method for object ranging based on the Taylor model for calibrating the internal and external parameters of the binocular stereo perception system for extreme correction, thereby enhancing the measurement accuracy of the binocular stereo perception system. The algorithms are implemented using the NVIDIA Jetson TX2 + FPGA hardware framework, enabling real-time ranging and display.

2. System Framework

In this paper, a wide-field-of-view binocular stereo sensing system based on infrared and visible light is established; the system block diagram is shown in Figure 1. This system consists of two sets of uncooled long-wave infrared-imaging modules with a field-of-view greater than 120°, visible-high-definition-imaging modules, and digital-video-processing modules. The system's operating temperature ranges from $-40\text{ }^{\circ}\text{C}$ to $70\text{ }^{\circ}\text{C}$. The digital-video-processing module utilizes hardware circuits featuring the NVIDIA Jetson TX2 + FPGA combination and receives digital video signals from the visible light-imaging component and the infrared-imaging component, enabling the module to perform image processing. A hardware diagram of the digital-video-processing module is shown in Figure 2.

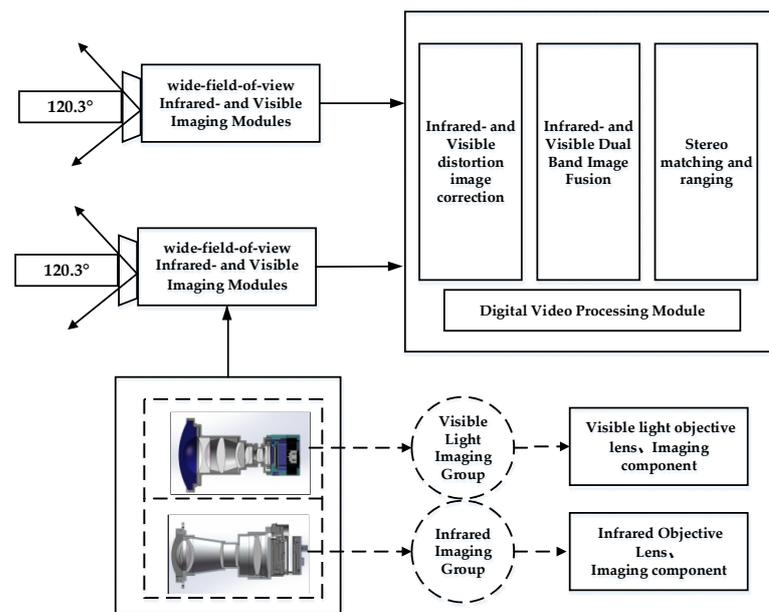


Figure 1. Schematic diagram of the wide-field binocular stereo perception system with both infrared- and visible light components.

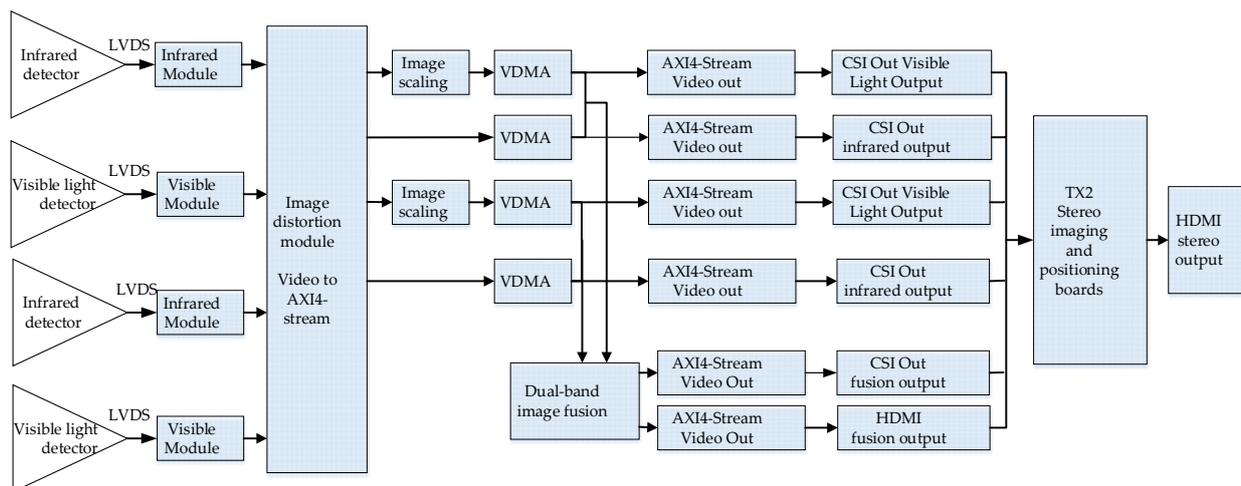


Figure 2. The digital-video-image-processing module.

2.1. Design and Simulation of the Wide-Field Binocular Stereo Perception-Dual-Band-Imaging System

2.1.1. The Principle of Binocular Stereo Imaging

In response to the specific requirements of night-time vehicle driving and rapidly changing work environments, we designed a wide-field-binocular stereo perception system with dual-band-electro-optical imaging. The visiblelight-imaging component incorporates a large-array CMOS device from Rockchip Electronics Co., Ltd. (Jiangsu, China) that has high sensitivity and definition. This system boasts a pixel resolution of 1920 (H) \times 1080 (V), with each pixel measuring 13 μm (H) \times 13 μm (V), and it operates at a frame rate of 50 Hz. This component can function effectively under low-light conditions down to 10^{-3} lx. Furthermore, the infrared-imaging component employs a non-cooled infrared focal plane detector from Yantai IRay Technology Co., Ltd. (Yantai, China) that offers a pixel resolution of 1024 (H) \times 768 (V), with each pixel measuring 14 μm . The frame rate of the infrared component is also 50 Hz.

Binocular stereo sensing systems are mainly divided into two types according to their placement [27]: a parallel model and a convergent model. The parallel model has two

camera optical axes parallel to each other. Moreover, the structure is simple and easy to calculate. The advantage of this structure is the presence of only a negative horizontal parallax, with no vertical parallax. Disadvantages include a small common area and a lack of stereoscopic information in the left and right sides of the single viewing area, which will cause a waste of information. The convergence model can adjust the angle between the two optical axes to obtain a larger effective field-of-view, with positive-, negative-, or zero-horizontal parallax; however, the camera body will produce vertical parallax, which causes a certain gradient distortion. In this study, considering the advantages and disadvantages of the parallel model and the convergence model, the visible and infrared components were fused. For this purpose, we selected an optical design based on the parallel model, as shown in Figure 3. To realize binocular stereo vision, the binocular stereo sensing system consisted of two visible light-objective lens groups and two infrared-objective lens groups.

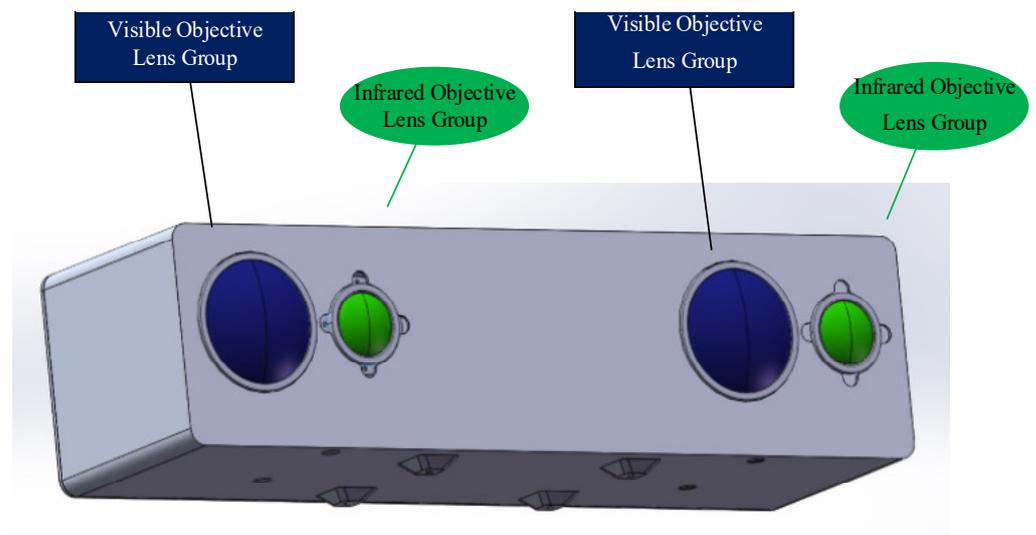


Figure 3. Schematic diagram of the infrared- and visible light-lens structure in the binocular-stereo-perception system.

Binocular-imaging-distance-measurement technology relies on the binocular disparity to establish an ideal model for binocular ranging. In this model, both cameras have identical specifications and parameters, including matching camera models, consistent focal lengths, and the parallel alignment of optical axes. The model is shown in Figure 4.

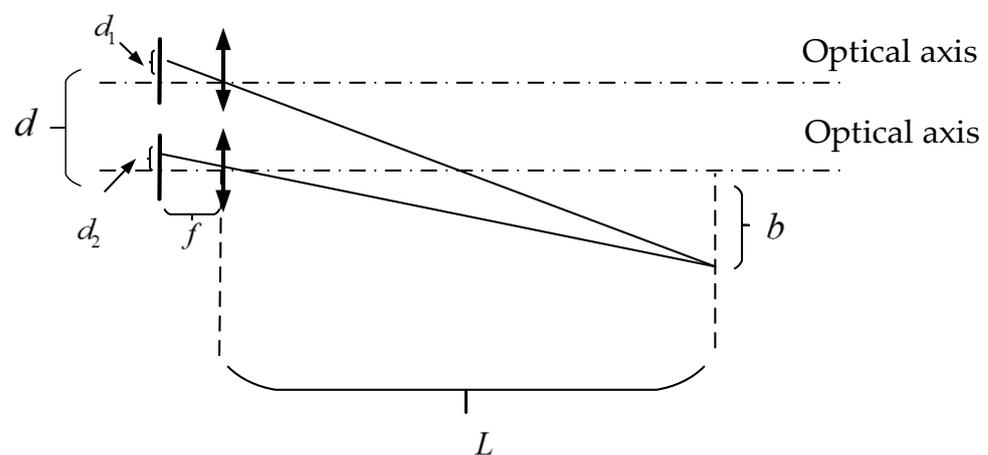


Figure 4. The parallel dual-axis-ranging model with the same focal length.

Here, d represents the camera's inter-image plane spacing (baseline width), d_1 and d_2 are the distance from the image point to the image plane center, f is the camera's focal length, and L is distance to the object. Using the principles of similar triangles, the following relationships can be derived:

$$L = \frac{fd}{d_1 - d_2} = \frac{fd}{c} \quad (1)$$

where c represents the difference in the imaging positions of the object point in the two fields of view, which is commonly referred to as the parallax value.

Under the conditions of parameter determination in a binocular-stereo-perception system (such as focal length and baseline), ranging accuracy is determined by the parameter c . Presently, binocular stereo-matching algorithms can achieve sub-pixel-level matching accuracy, resulting in superior disparity precision. Results can be obtained by differentiating Equation (1):

$$dL = \frac{fd}{c^2}dc = \frac{L^2}{fd}dc. \quad (2)$$

Hence, as long as the parameters of the binocular-stereo-perception system are determined, the distance to the target can be calculated by measuring the disparity. System parameters typically consist of intrinsic and extrinsic parameters. Intrinsic parameters include the focal length, principal point coordinates, and distortion coefficients of the left and right cameras, among others. Extrinsic parameters encompass the relative transformation between the left and right cameras, involving rotation and translation matrices. Due to potential errors during camera installation, such as non-parallel alignment of the lens and imaging plane, it is necessary to recalibrate the camera to obtain updated focal lengths, intrinsic parameters, and extrinsic parameters.

2.1.2. The Simulation and Design of Visible- and Infrared Objective Lenses

A. Simulation design for the visible light objective lens

In the wide-field binocular stereo perception-optical-imaging system, the visible light component employs a high-resolution and low-light CMOS-imaging module from Rockchip Electronics Co., Ltd., which enables imaging in both day and night scenes. This component features large pixels and sensor-imaging areas, imposing stringent requirements on the optical system. We utilized the CODE V(10.2) software for optical system design, implementing a "telephoto-type" optical path structure with 12 lenses. By introducing appropriate non-spherical elements while keeping the total number of lenses, glass thickness, and imaging quality constant, we enhanced the light-gathering capabilities of the visible light objective lens without compromising its transmittance. The system design is depicted in Figure 5. Table 1 presents the optical-design specifications for the visible light objective lens.

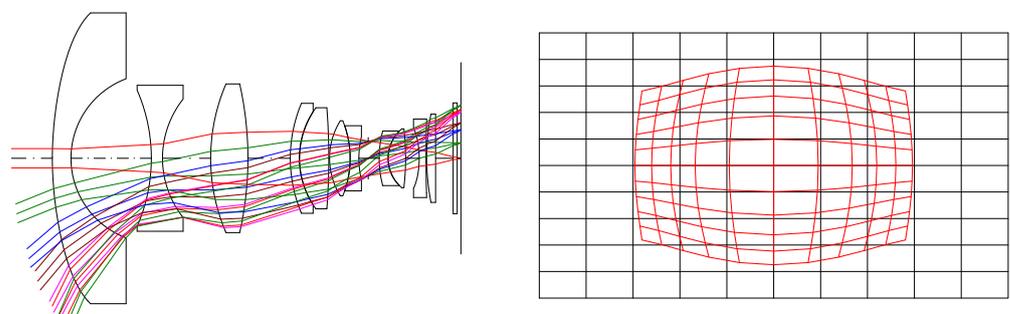


Figure 5. Visible light objective lens-optical system and distortion diagram.

Table 1. Optical-design specifications for the visible light lens.

Parameters	Value	Parameters	Value
Objective focal length (f)	10.96 mm	Diagonal field	133°
F-number	2.1	Distortion (maximum on the diagonal)	−45%
Horizontal field	124°	Maximum working distance	97.16 m
Vertical field	74.4°		

The maximum effective range L of the visible light-optical-imaging system is as follows:

$$L = \frac{f\sqrt{wh}}{8d_{pix}} = 97.16 \text{ m} \quad (3)$$

where w represents human height, h represents shoulder width, and d_{pix} represents pixel size.

At a distance of $L = 15$ m, there is a depth-calculation deviation of $dL \leq 20$ cm. Under the condition of a matching algorithm precision at the 0.1 pixel level, the baseline d can be determined as follows:

$$d \geq \frac{L^2 dC}{fdL} = 13.52 \text{ cm} \quad (4)$$

where L represents the distance from the target to the system, which is referred to as the test distance. dC is set at a 0.1 pixel level, and dL stands for the resolvable distance.

At a distance of $L = 30$ m, the depth-calculation deviation of $dL \leq 1$ m. When these values are incorporated into Equation (4), the following result is obtained:

$$d \geq \frac{L^2 dC}{fdL} = 10.82 \text{ cm}. \quad (5)$$

It can be seen clearly that $d \geq 13.52$ cm. The formula for visible light stereo acuity $d\gamma$ is

$$d\gamma = \frac{\alpha d}{L^2} dL \quad (6)$$

where α is a constant of 206,265 when converting from radians to arcseconds:

$$d\gamma = \frac{\alpha}{f} dc = 24.47''. \quad (7)$$

Stereopsis is the ability to resolve the smallest horizontal disparity between retinal images from both eyes. The normal value for stereopsis should be less than 60''. A smaller value of stereopsis indicates better stereo vision.

Utilizing CONE V software, we simulated environmental temperature variations and obtained the transfer functions for the visible light-optical system at 20 °C, 50 °C, and −40 °C, as well as the diffuse spots, as shown in Figure 6, Figure 7, and Figure 8, respectively.

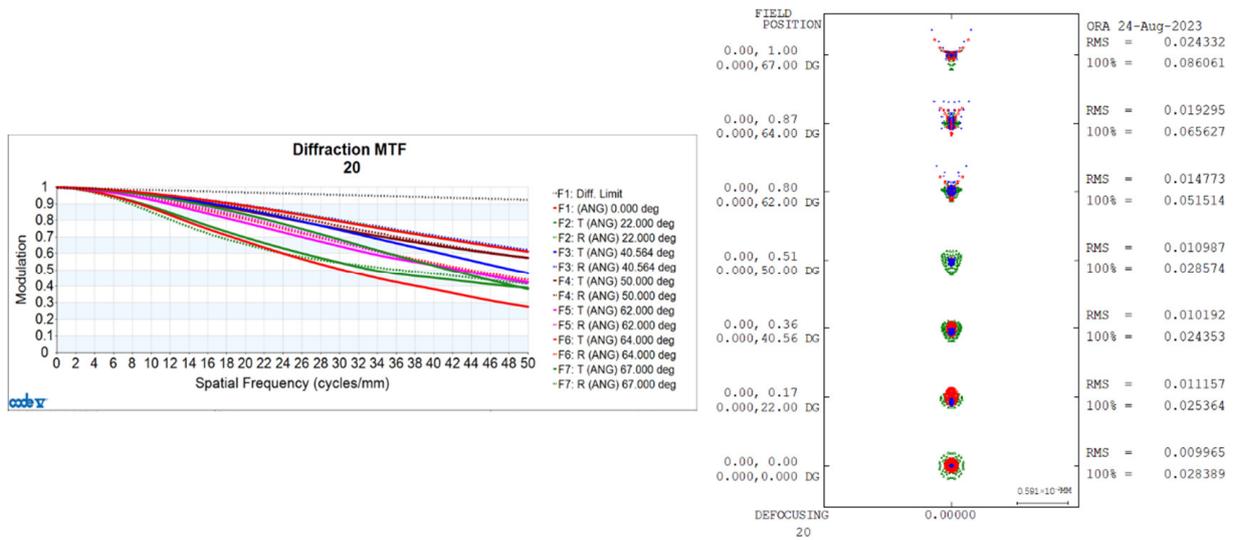


Figure 6. Transfer functions and diffuse spots of visible light-optical systems at 20 °C.

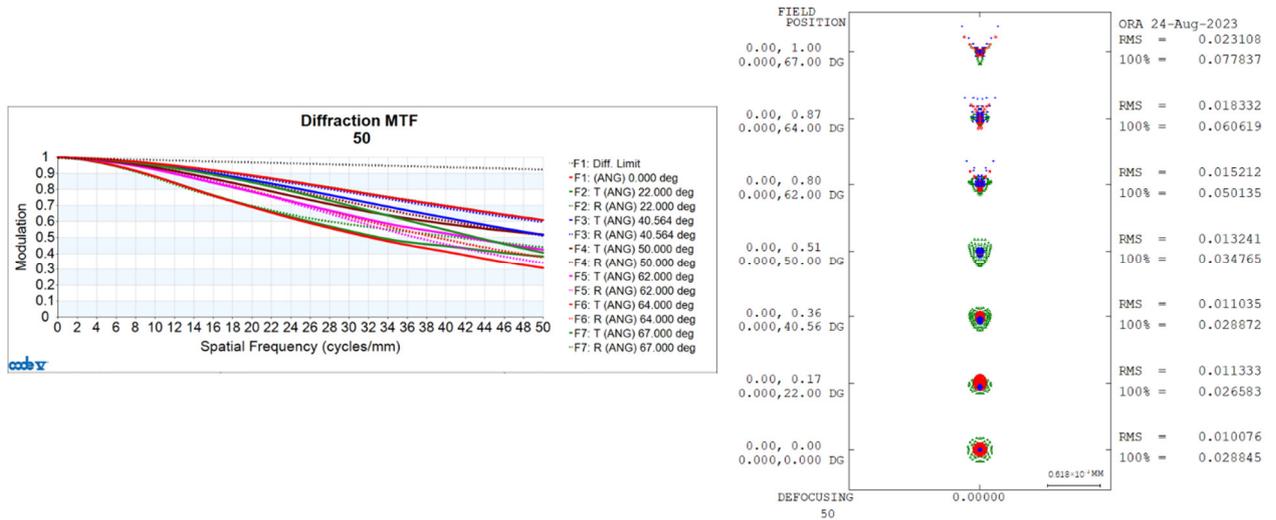


Figure 7. Transfer functions and diffuse spots of visible light-optical systems at 50 °C.

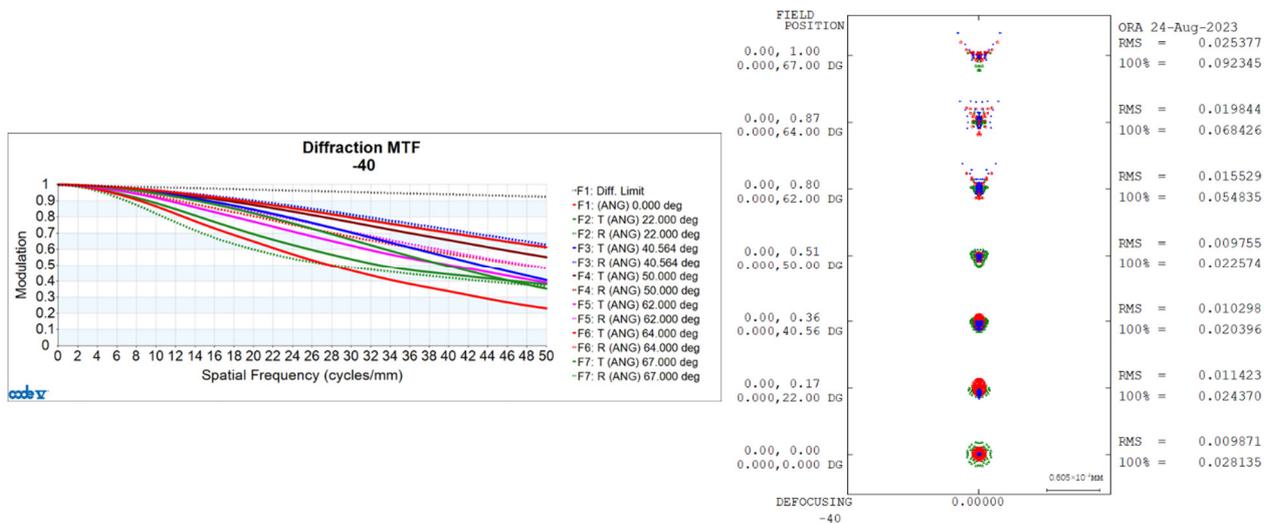


Figure 8. Transfer Functions and Diffuse Spots of Visible Light-Optical Systems at -40 °C.

B. The Simulation Design of an Infrared objective lens

This paper addresses the design of a wide-field, relative aperture long-wave infrared optical system for the 1024 (H) × 768 (V) long-wave infrared detector from Yantai IRay Technology Co., Ltd. The infrared objective lens system, as depicted in Figure 9, operates without active cooling over a wide temperature range. The front surface of the first lens is non-spherical, enabling passive temperature compensation at different temperatures by adjusting the system's back focal length. Table 2 provides the optical-design specifications for the infrared objective lens.

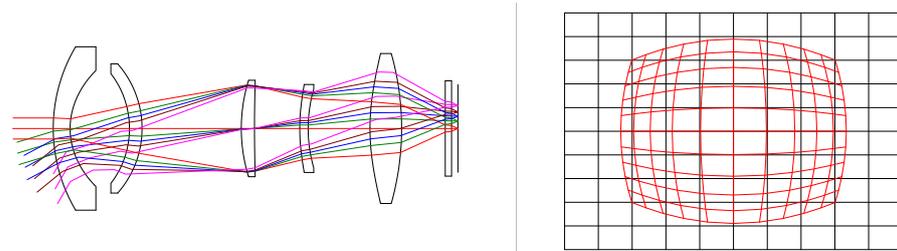


Figure 9. The optical system of the infrared objective lens and a distortion map of the infrared objective lens at 20 °C.

Table 2. The optical-design specifications for the infrared objective lens.

Parameters	Value
objective focal length (f)	6 mm
F-number	2.1
horizontal field	121.6°
vertical field	107°
distortion (maximum on the diagonal)	−47%

The maximum effective range L of the infrared objective lens-optical system is

$$L = \frac{f\sqrt{wh}}{8d_{pix}} = 40.67 \text{ m.} \quad (8)$$

At a distance of $L = 15 \text{ m}$, with a depth-calculation deviation $dL \leq 30 \text{ cm}$, the baseline d can be calculated as follows:

$$d \geq \frac{L^2 dC}{fdL} = 21.26 \text{ cm.} \quad (9)$$

At a distance of $L = 30 \text{ m}$, with a depth-calculation deviation $dL \leq 2 \text{ m}$, the baseline d can be calculated as follows:

$$d \geq \frac{L^2 dC}{fdL} = 12.75 \text{ cm.} \quad (10)$$

According to Formulas (9) and (10),

$$d \geq 21.26 \text{ cm.} \quad (11)$$

The formula for the stereo acuity $d\gamma$ is as follows:

$$d\gamma = \frac{\alpha d}{L^2} dL \quad (12)$$

where d is the baseline length, dL is the resolvable distance, L is the distance from the target to the system, and dL is the binocular ranging accuracy. The formula is as follows:

$$d\gamma = \frac{a}{f}dc = 58.44'' \tag{13}$$

With a decrease in temperature, the infrared optical system experiences changes in inter-lens spacing, lens thickness, refractive index, and curvature radius. Therefore, temperature variations inevitably lead to defocusing of the system's focal plane, resulting in a degradation of image quality. The transfer functions and diffuse spots of the infrared optical system based on system simulation analysis are shown in Figure 10, Figure 11, and Figure 12, respectively, at temperatures of 20 °C, 50 °C, and −40 °C.

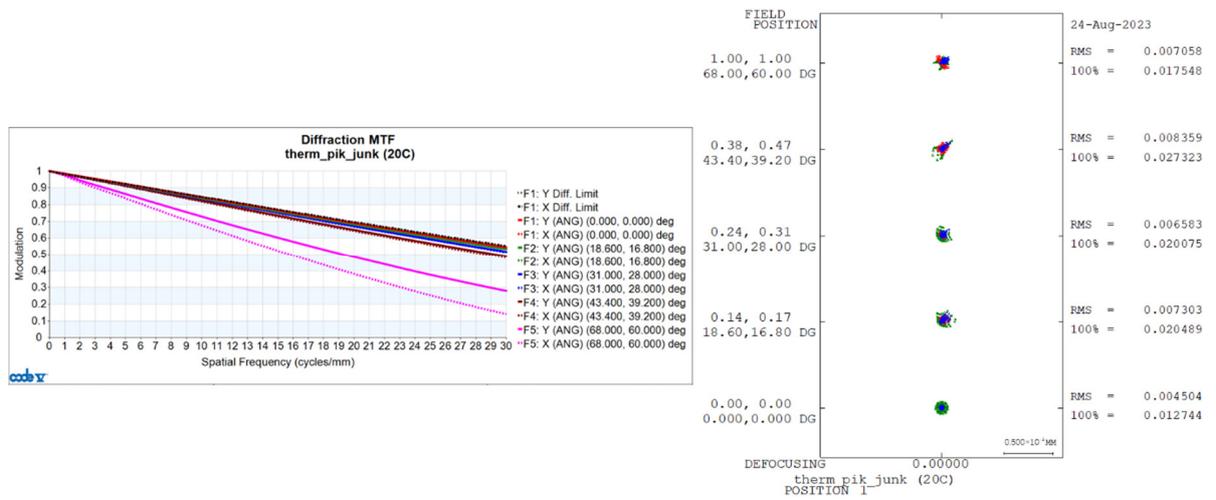


Figure 10. Transfer Functions and Diffuse Spots of Infrared Optical Systems at 20 °C.

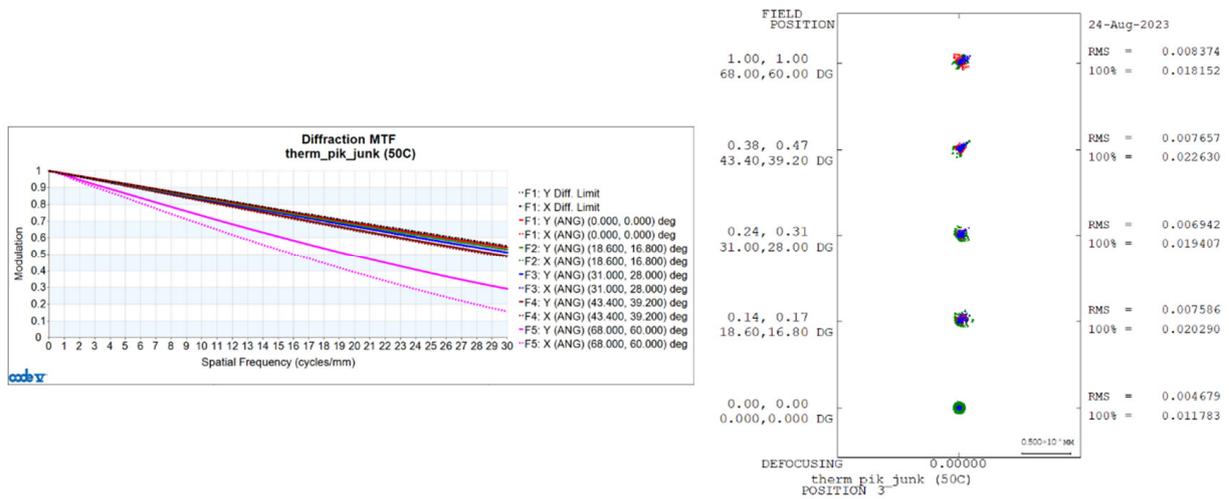


Figure 11. Transfer Functions and Diffuse Spots of Infrared Optical Systems at 50 °C.

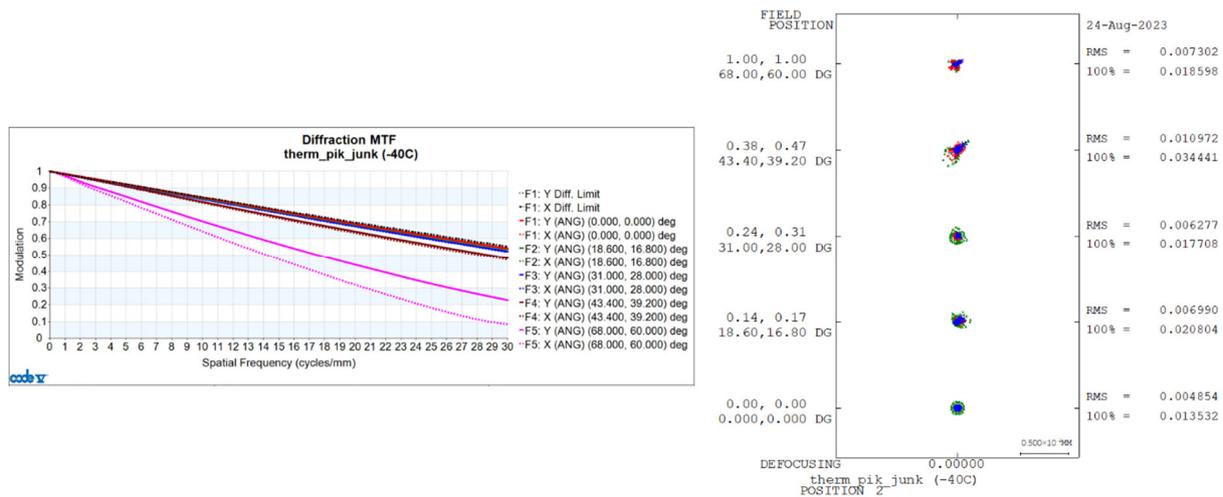


Figure 12. Transfer Functions and Diffuse Spots of Infrared Optical Systems at $-40\text{ }^{\circ}\text{C}$.

In summary, without moving any optical elements, we achieved alignment of the image plane with the detector target surface during changes in environmental temperature. With temperature variations, the lens assembly, employing optical passive thermal compensation, can maintain consistent magnification and requires no active optical components. The relative positions of the optical axes remain relatively unchanged, resulting in high image registration accuracy. Large-field optical systems often exhibit significant distortion to improve the field-of-view. Image distortion is corrected to obtain high-quality fused images. Based on the imaging quality of the fused lens system and image registration effectiveness, further improvements in the system's ranging accuracy were achieved.

Baseline distance $d \geq 21.26\text{cm}$; if baseline distance $d = 35\text{cm}$, then the ranging accuracy is as follows:

$$dL \geq \frac{L^2 dC}{fd}. \quad (14)$$

Table 3 presents the calculated accuracy of visible and infrared ranging at test distances of 15 m and 30 m.

Table 3. Infrared- and visible light-ranging accuracy.

Serial Number	Measuring Distance (L)	Visible Light-Ranging Accuracy (L)	Infrared-Ranging Accuracy (L)
1	15 m	$dL \geq 7.63\text{cm}$	$dL \geq 15.00\text{cm}$
2	30 m	$dL \geq 30.50\text{cm}$	$dL \geq 60.00\text{cm}$

2.2. Real-Time Image Distortion Correction and Simulation

The optical system design model produces aberrations that are unavoidable due to the system's non-coaxial nature, field-of-view, focal length, and other auxiliary factors. In this study, a binocular stereo sensing optical imaging system with a large field-of-view is designed, in which the aberration rate produced by the visible objective lens is -45% , and the aberration rate of the infrared objective lens is -47% , as shown in Section 2.1. Here, we propose a shareable elliptical aberration for real-time correction.

Image distortion is mainly the result of geometric distortion of the pixel positions of the image after imaging. Geometric distortion is further categorized into linear and nonlinear distortion, which refer to a mixture of several distortions that work together [28]. The causes of nonlinear aberrations are mainly categorized as radial aberrations, centrifugal aberrations, and thin prismatic aberrations [29,30]:

$$x_d = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \quad (15)$$

$$y_d = y(1 + k_1r^2 + k_2r^4 + k_3r^6) + 2p_1xy + p_2(r^2 + 2y^2) \quad (16)$$

$$x^2 + y^2 = r^2 \quad (17)$$

where x and x_d represent, respectively, the reference image and distortion image in the X-direction; y and y_d are, respectively, the reference image and distortion image in the Y-direction; k_1, k_2, k_3, p_1 , and p_2 are distortion correction parameters. When the aberration coefficient is greater, the aberration correction is less effective. Aberration correction usually considers radial aberrations and ignores the effect of tangential aberrations, which can describe the nonlinear aberrations of the lens.

Because the resolution ratio of the image $H:V \neq 1:1$, there is a poor edge effect for the image edge distortion correction edge when using the standard concentric circle distortion model. Therefore, an elliptical distortion correction model is proposed using the improved standard concentric circle distortion model:

$$x = x_d \frac{1}{(1 + k_1r^2)}, y = y_d \frac{1}{1 + k_1r^2} \quad (18)$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = r^2 \quad (19)$$

where H is the height of the image, V represents the image's width, $a = H/2$ and $b = V/2$ are shown in Figure 13.

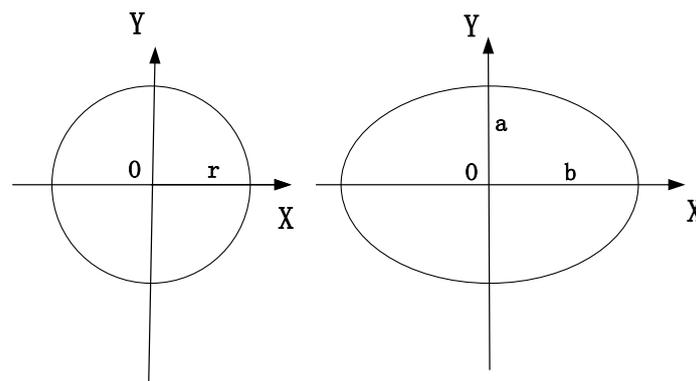


Figure 13. Standard-concentric-aberration model and elliptical-aberration-correction model.

The elliptical distortion correction model cannot cover the image completely ($H \times V$). We further improve the elliptical distortion correction model using an approximate elliptical distortion correction model:

$$x = k_2 \frac{x_d}{1 + k_1(r - R)^2}, y = k_2 \frac{y_d}{1 + k_1(r - R)^2} \quad (20)$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = r^2 \quad (21)$$

where k_1 and k_2 represent the distortion coefficients; x and x_d represent the corrected image and distorted image in the X-direction; and y and y_d represent the corrected image and the distorted image in the Y-direction, respectively.

In this paper, real-time corrections are performed based on FPGA. A common approach in FPGA hardware circuits is the correction-mapping table, which first calculates the corresponding correction result of the image and stores it in the hardware circuit and determines the corresponding corrected video output from the correction-mapping table based on the input real-time video. The correction-mapping table is a straightforward computation but suffers from the disadvantage of depositing an approximate elliptic-distortion-correction-mapping table, which requires a large amount of hardware-circuit

resources. Thus, we proposed elliptical distortion correction model only needs to store one quarter of the data in the hardware circuitry, as shown in Figure 14.

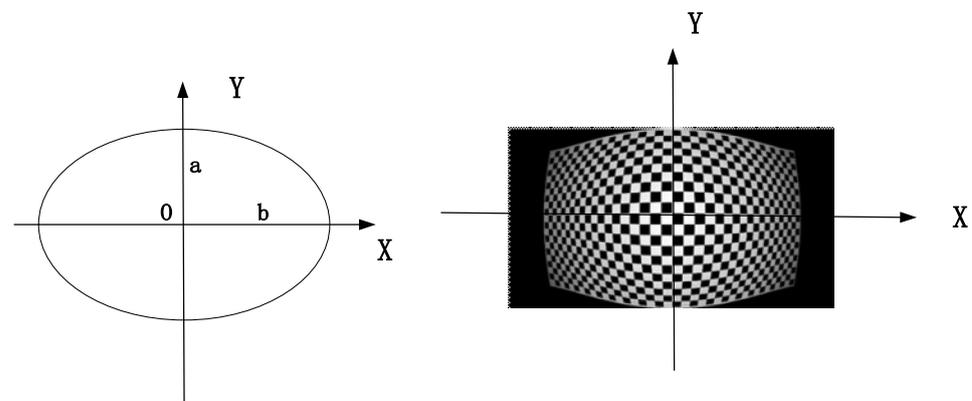


Figure 14. Correction-mapping table for approximate elliptical-distortion correction.

The generated checkerboard grid $H \times V$ images are shown in Figures 15a and 16a and are standard checkerboard grids of 1920×1080 and 1024×768 , respectively. Figures 15b and 16b present, respectively, the aberration simulation images derived using the standard checkerboard grid as the input image through the visible optical system model and those generated using the infrared optical system. Through the myopic elliptic-aberration model proposed in this paper for aberration correction, the aberration-correction map of the visible image and the aberration-correction map of the infrared image are obtained, as shown in Figures 15c and 16c, respectively. The edges of the image are still aberrated, but the field-of-view is increased.

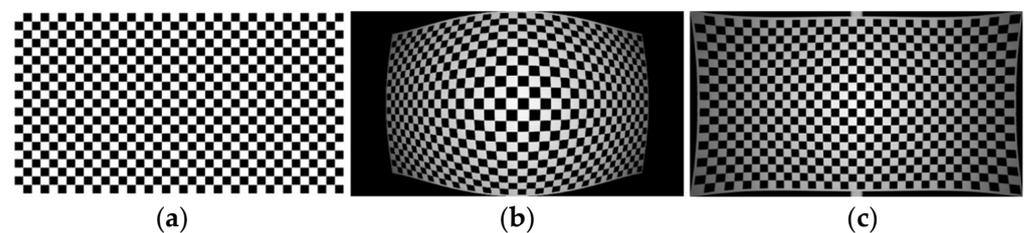


Figure 15. The 1920×1080 Visible light distortion effect: (a) standard checkerboard grid; (b) optical-system-aberration map (aberration rate -45%); and (c) aberration-correction-result map.

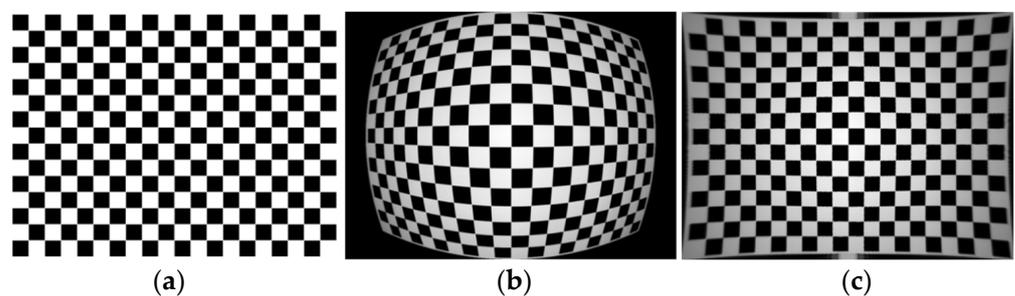


Figure 16. The 1024×768 infrared-distortion effect: (a) standard checkerboard grid; (b) optical-system-aberration map (aberration rate -45%); and (c) aberration-correction-result map.

2.3. Infrared- and Visible Light-Fusion Algorithm

The infrared- and visible light images are horizontally aligned after image alignment and limit correction to realize the left and right images, and the common adaptive image-enhancement method [31] is used to enhance the infrared- and visible light images to improve the bright and dark regions in the image to improve the contrast of the image.

Infrared- and visible light-image matching is a type of heterogenous split matching, and it is difficult to find the same type of corresponding feature points for the two types of images. In this paper, we use the SURF + RANSAC algorithms to realize stereo matching [32,33]. The SURF algorithm has the characteristics of rotation, scale transformation, and brightness invariance; compared to the SIFT algorithm, SURF reduces the complexity of the algorithm, reduces the dimensionality of the feature descriptor from 128 to 65 dimensions, and reduces the computational amount by double [34]. The anomalous data in the matching process is filtered out by using the RANSAC algorithm.

Image fusion involves the use of image information from multiple imaging sensors in a unified scene to increase the perception of the scene and the ability to recognize targets and other objectives. Unlike traditional multi-scale fusion methods, this study adopts a multi-scale infrared- and visible light-fusion method based on the work in [35], which has the unique characteristics of retaining scale-specific information and reducing the edge halo, taking into account the different characteristics of the infrared image and the visible image. This study also adopts the traditional “maximum–absolute” fusion rule. With this optimization, the useful visual details can be better transferred to the fused image while suppressing the noise in the infrared image:

$$V(p) = |I_p - I_1| + |I_p - I_2| + \dots + |I_p - I_N| \quad (22)$$

where I_p is the intensity value at pixel p in image I , $V(p)$ is the significant value of pixel p , and N is the total number of pixels in image I .

If the two-pixel intensity values are equal, then

$$V(p) = \sum_{j=0}^{L-1} M_j |I_p - I_j| \quad (23)$$

where j is the pixel point, M_j is the number of pixels with pixel intensity the same as that of point j , and L is the number of gray levels in the image. The infrared image I_{ir} and visible image I_{vi} are the input images. Then, the base layer B_F of the fused image is obtained as

$$B_F = W_b B_1 + (1 - W_b) B_2 \quad (24)$$

$$W_b = 0.5 + \frac{I_{ir} - I_{vi}}{2} \quad (25)$$

Infrared images usually contain coarse-scale structural information and noise and lack visual details compared with visible light images. The fused detail layer obtained by the “max–absolute” rule is improved to make the image look more natural and more suitable for human visual perception. The weighting factor W^j of “max–absolute” is

$$W^j = \begin{cases} 1, & |d_1^j| > |d_2^j| (j = 1, 2, \dots, N) \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

Next, we apply a Gaussian filtering for denoising:

$$W_d^j = \text{Gaussian}(W^j, \delta_s) \quad (27)$$

where setting $\delta_s = 2$, M^j for the fusion details in the j_{th} layer is accomplished according to the “max–absolute” rule, as follows:

$$M^j = W_d^j d_1^j + (1 - W_d^j) d_2^j. \quad (28)$$

The fusion detail layer D^j of the j_{th} layer can be obtained according to the weighted least squares method:

$$D^j = \sum_P \left((D_p^j - M_p^j)^2 + \lambda a_p^j (D_p^j - (d_2^j)_p)^2 \right) \quad (29)$$

where $a_p^j = \left(\left| \sum_{q \in w_p} (d_1^j)_q \right| + \varepsilon \right)^{-1}$ represents the spatial variation weight; p represents the spatial position of the pixel; ε is a constant equal to 0.0001; and w_p represents a rectangular window centered on pixel point p .

The fused image I_F after combining the base layer B_F and the detail layer D^1, D^2, \dots, D^N is

$$I_F = B_F + D^1 + D^2 + \dots + D^N. \quad (30)$$

2.4. Binocular Stereo Ranging Algorithm

Unlike the traditional binocular stereo-ranging method [36], we seek to acquire binocular stereo-ranging information from the image obtained via aberration correction and image fusion. As shown in Figure 17, parallax information is used to restore the depth. Here, Q_L and Q_R are the coordinate origins of the two camera coordinate systems, and A is the baseline length. To find the corresponding point of the world coordinate system in the stereomage pair for point Q , the corresponding direction vectors in the virtual coordinate system are $\overrightarrow{Q_L Q}$, $\overrightarrow{Q_R Q}$. Q' is the projection of Q in the $XO_L Z$ plane. The angles of the vector $\overrightarrow{O_L Q}$, $\overrightarrow{O_R Q}$ with the plane $YO_L Z$, $YO_R Z$ are β_1, β_2 . If the angles with the $O_L Z$ -axis and $O_R Z$ -axis are both α ; then, we apply the following:

$$\theta_1 = \frac{\pi}{2} - \beta_1, \theta_2 = \frac{\pi}{2} - \beta_2. \quad (31)$$

$$\frac{B}{\sin(\pi - \theta_2)} = \frac{A}{\sin \theta} = \frac{A}{\sin(\theta_2 - \theta_1)} \Rightarrow B = \frac{A \sin \theta_2}{\sin(\theta_2 - \theta_1)}. \quad (32)$$

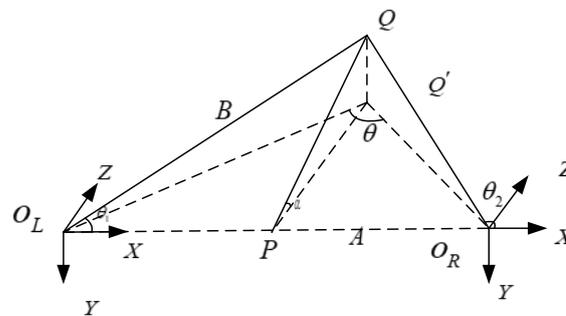


Figure 17. Schematic diagram of binocular stereo-ranging information.

Then, the depth information of Q is

$$Q_{\text{depth}} = Q'P = B \sin \theta_1 \cos \alpha. \quad (33)$$

3. Experiments and Results

3.1. Test Platform

The experiment adopted a wide-field-of-view binocular stereo perception system and a calibration checkerboard grid (ambient and heated state), as shown in Figure 18. The calibration checkerboard grid was based on the Boltier principle design for an active infrared-radiation-calibration checkerboard grid. We used a JY-260 microcomputer temperature controller device from Jiangyin Jinyu Electric Heating Appliance Co., Ltd. (Jiangsu, China) to control the temperature of the calibration checkerboard grid in a range from

−50 °C to −260 °C, the calibration checkerboard grid was 10 × 7, the length of the squares was 100 mm, and the overall size was 1040 mm × 740 mm. Infrared-component calibration was carried out using the power supply. After the infrared component was calibrated, the power supply was used to heat the calibration checkerboard grid, and the temperature was controlled with the JY-260 microcomputer temperature controller at 50 °C, which heated and cooled the white box to produce corners with a large grayscale gradient on the infrared image, which was convenient for corner identification.



Figure 18. Wide-field-of-view binocular stereo sensing system and calibrated checkerboard grid.

3.2. Calibration Test Results

The binocular stereo sensing system was calibrated using sensors for tessellated grids in a three-dimensional scene. Under the optical imaging model, the relationship between the spatial coordinate system of the object points in the scene and the corresponding image points in the image plane was established, which, in turn, determined the internal and external parameters of the camera. In the experiment, calibration of the internal and external parameters of infrared and visible cameras for binocular stereo sensing systems was based on Taylor model-calibration. The images used in this experiment were acquired by visible light and infrared left and right cameras in different orientations for calibrating the checkerboard grid, and then the coordinates of the corner points were extracted.

We selected 18 images from the collected data for calibration, as shown in the binocular stereo left lens's visible light-camera-calibration diagram in Figure 19, and in the right lens's visible light camera-calibration diagram in Figure 20, and in the right lens's infrared camera-calibration diagram in Figure 21, and in the right lens's infrared camera-calibration diagram in Figure 22. The binocular stereo sensing system's visible camera and its infrared camera inside and outside the parameters of the calibration results are shown in Tables 4 and 5. We perform parameter calibration based on a calibrated chessboard, which absorbs heat uniformly after heating, overcoming the problem of difficult to recognize corner points in infrared cameras, and at the same time solving the calibration problems arising from the different imaging characteristics of infrared and visible light.

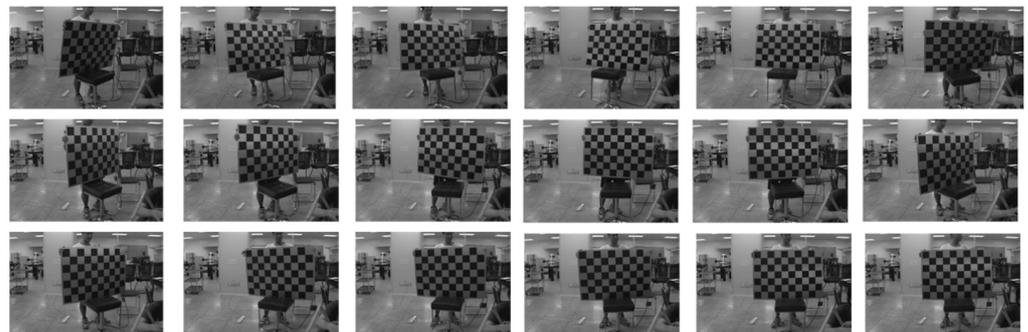


Figure 19. Calibration chart for the visible light camera (left lens).

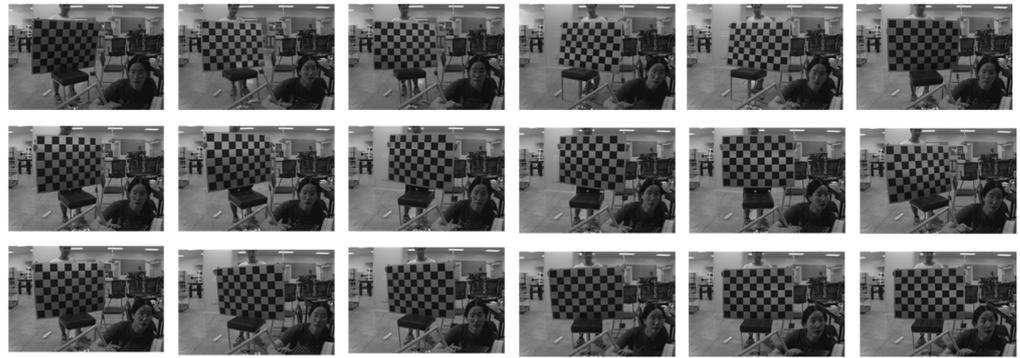


Figure 20. Calibration chart for the visible light camera (right lens).

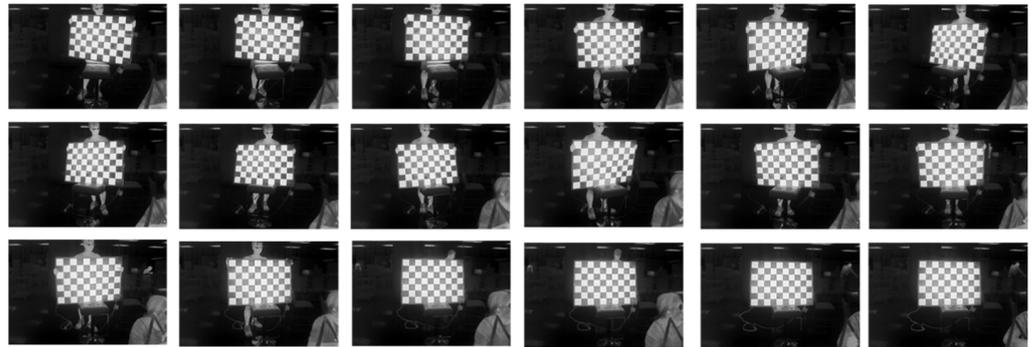


Figure 21. Calibration chart for the infrared camera (left lens).

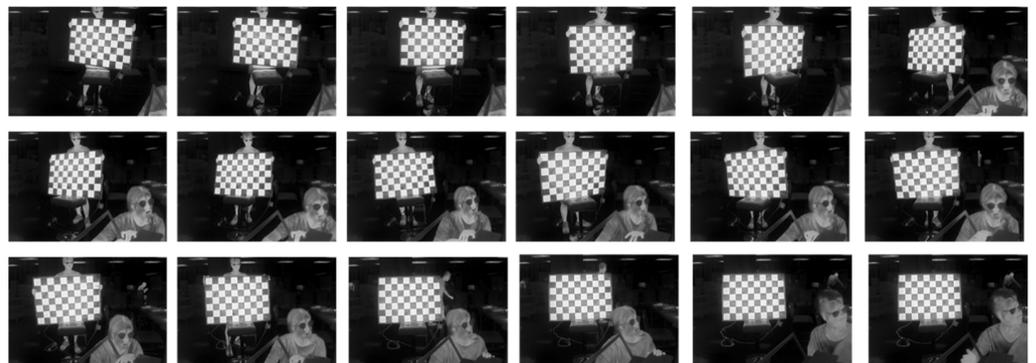


Figure 22. Calibration chart for the infrared camera (right lens).

Table 4. Internal and external parameters of the visible light camera of the binocular stereo sensing system.

Parameter	Calibration Results (Left)	Calibration Results (Right)
(c, d, e)	$(1.0001, -1.2 \times 10^{-5}, 1.7 \times 10^{-5})$	$(1.0001, -2.2 \times 10^{-5}, -2.12 \times 10^{-4})$
(u_0, v_0)	$(487.0978, 964.3319)$	$(464.1441, 949.0153)$
$(a_0, a_1, a_2, a_3, a_4)$	$(-835.384, 0, 3.296 \times 10^{-4}, 3.290 \times 10^{-7}, -3.189 \times 10^{-10})$	$(-819.290, 0, 3.642 \times 10^{-4}, 2.727 \times 10^{-7}, -2.909 \times 10^{-10})$
R	$\begin{pmatrix} 0.9976 & -0.0009 & 0.0684 \\ 0.0007 & 0.9999 & 0.0019 \\ -0.0684 & -0.0018 & 0.9976 \end{pmatrix}$	
T	$[-335.5285 \quad 9.9994 \quad -16.8602]$	

Table 5. Internal and external parameters for the infrared camera of the binocular stereo sensing system.

Parameter	Calibration Results (Left)	Calibration Results (Right)
(c, d, e)	$(0.992, -1.47 \times 10^{-3}, 6.2 \times 10^{-5})$	$(0.9998, -9.5 \times 10^{-5}, 1.9 \times 10^{-5})$
(u_0, v_0)	$(479.4945, 959.2869)$	$(449.2642, 932.4781)$
$(a_0, a_1, a_2, a_3, a_4)$	$(-780.385, 0, 2.472 \times 10^{-3}, 1.522 \times 10^{-8}, -7.214 \times 10^{-12})$	$(-796.383, 0, 8.938 \times 10^{-5}, 3.160 \times 10^{-8}, 1.7333 \times 10^{-10})$
R	$\begin{pmatrix} 0.9989 & 0.0371 & 0.0291 \\ -0.0375 & 0.9992 & 0.0133 \\ -0.0286 & -0.0144 & 0.9995 \end{pmatrix}$	
T	$[-355.90 \quad 11.0089 \quad 35.05714]$	

3.3. Aberration-Correction Results

The aberration maps for this experiment were acquired indoors on a calibrated checkerboard grid using a binocular stereo sensing system. The aberration-correction algorithm in this paper was used to obtain the aberration-correction map (shown in Figures 23 and 24). The aberration rate of the visible image was 45%, indicating that the image aberration caused by the large field-of-view was greatly improved at the edges and that the field-of-view was not lost.

**Figure 23.** Aberration and correction maps for visible light camera (left lens).**Figure 24.** Aberration and correction maps for visible light camera (right lens).

3.4. Ranging Test Results

As shown in Figures 25 and 26, a person was selected as the recognition target and could be recognized by both visible light and infrared cameras. Several ranging measurements were taken at 5 m, 10 m, 15 m, 20 m, 25 m, 30 m, 40 m, and 60 m. By recognizing the target, the target point was framed out and the depth value of the feature point inside the frame was calculated. Sometimes the feature point contained both the foreground and the background feature points. In this study, the analysis was carried out using Equation (34), eliminating the unwanted data. The results of the ranging tests are shown in Table 6.

$$\bar{d}_V \pm \frac{\sum |d_i - \bar{d}_V|}{n} \quad (34)$$

where d_i is the distance value of the feature point, and \bar{d}_V is the average value of the distance value of the feature point.

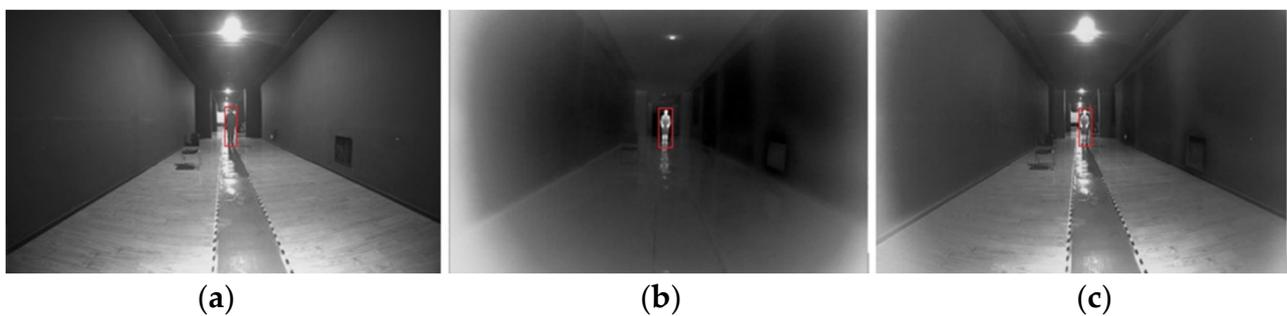


Figure 25. Binocular distance-measurement: (a) visible image; (b) infrared image; (c) and fused image.

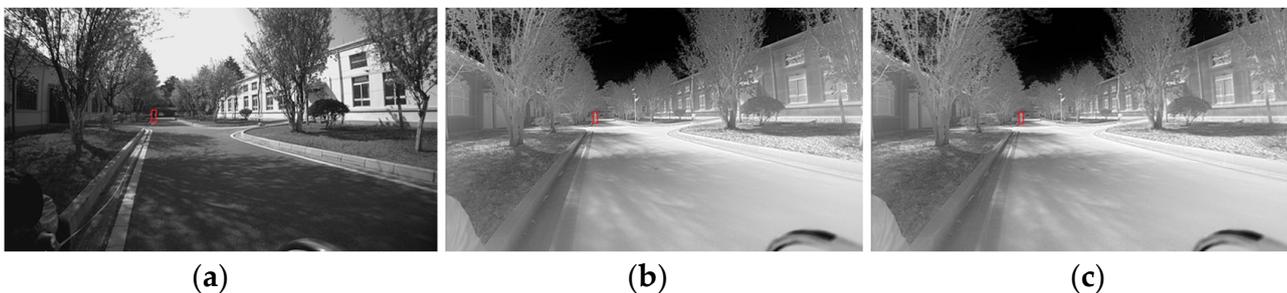


Figure 26. Binocular ranging-scene diagram in the park: (a) visible light camera image; (b) infrared image; (c) and fused image.

Table 6. Binocular distance-measurement results.

Serial Number	Measuring Distance (m)	Visible Light-Ranging		Infrared Ranging		Fusion-Ranging	
		Distance(m)	Difference (m)	Distance (m)	Difference (m)	Distance (m)	Difference (m)
1	5	4.82	0.18	5.05	0.05	5.01	0.01
2	10	10.15	0.15	10.02	0.02	10	0
3	15	14.64	0.36	15.04	0.04	15.02	0.02
4	20	19.01	0.99	19.84	0.16	19.98	0.02
5	25	24.14	0.86	24.53	0.47	24.86	0.14
6	30	31.62	1.62	29.53	0.47	30.37	0.37
7	40	37.99	2.01	39.46	0.54	39.78	0.22
8	60	56.76	3.24	62.55	2.55	60.35	0.35

Table 6 shows the visible ranging error at 0.99 m (distance 20 m), 1.62 m (distance 30 m), and 3.24 m (distance 60 m) and 0.02 m (distance 20 m), 0.47 m (distance 30 m), and 2.55 m (distance 60 m), as well as the fusion-ranging error at 0.16 m (distance 20 m), 0.37 m (distance 30 m), and 0.35 m (distance 60 m). At a close distance, the error of ranging was small, and the error gradually increased with an increase in distance. Comparing the

ranging results for visible and infrared cameras, the ranging accuracy of infrared camera was better than that of visible light camera. The reason for this result is that the reprojection error for the calibration of the infrared camera was less than that of the calibration results for the visible light camera. Our proposed fusion-ranging method has greater accuracy than those using only infrared- or visible light. The reason for this result is that the fusion algorithm combines the advantages of visible- light and infrared, and the extracted feature points are more accurate, so its ranging errors were 0.16 m (distance 20 m), 0.37 m (distance 30 m), and 0.35 m (distance 60 m). In addition, its feature points were more accurate, so its ranging results were higher.

In order to improve the performance of this system, this paper improves the running speed of the algorithms by optimizing the related algorithms, parallel processing, and hardware acceleration, as shown in Figure 27. The major algorithms include: an FPGA-based elliptic-aberration correction, which saves FPGA storage space by storing 1/4 of the data, and the processing time of the algorithm is 6 ms; an FPGA-based adaptive image-enhancement algorithm with a processing time of 0.2 ms, lower computational complexity, and faster processing time; FPGA-based multi-scale infrared and visible fusion with a processing time of 0.01 ms; a TX2-based SURF + RANSAC binocular stereo-matching algorithm which uses feature point extraction, feature point descriptor calculation, matching point search and consistency checking, such that the computational complexity is high and the processing time is 16 ms; and, finally, a binocular stereo ranging method that removes the misjudged points to get the final ranging data with a processing time of about 0.1 ms. To summarize, the algorithms in the binocular stereo sensing system have a total processing time of about 22.31 ms.

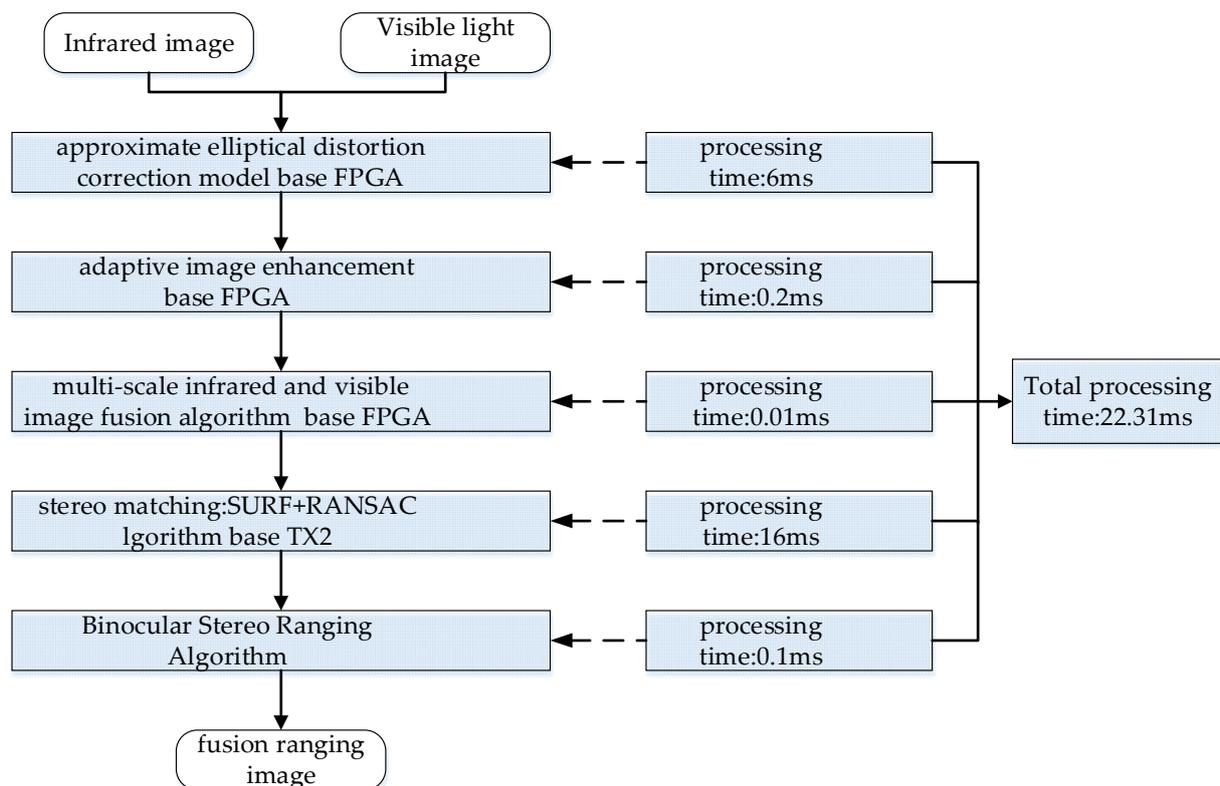


Figure 27. Image-algorithm-processing time for binocular stereo sensing systems.

4. Conclusions

This study proposed a dual-band fusion binocular stereo perception system with a large 120° field-of-view combining infrared- and visible light imaging through two sets of binocular stereo vision. This system enables information acquisition and stereo visual

perception within a wide field-of-view, providing enhanced scene understanding and addressing the limitations of traditional binocular stereo vision systems, such as a narrow field-of-view and inaccuracies in short-distance ranging. Aiming to solve the aberration problems of infrared and visible images with a large field-of-view, this paper adopted the elliptic-aberration-correction model, which was corrected in real-time in the hardware circuit in FPGA, thereby saving the storage resources of FPGA and solving the problem of field-of-view loss. By adopting the multi-scale fusion method, the visible and infrared images compensated for each other's differences, fully utilized the texture information of the visible image and the thermal radiation contour information of the infrared image, and significantly improved the ranging accuracy. The simulation and experimental results demonstrated that the fusion-ranging accuracy outperforms the individual ranging accuracies for infrared- and visible light cameras.

The proposed system can maintain high performance under different light and weather conditions and scenarios. Adding image recognition algorithms for target localization of special targets or obstacles, which has application value in many areas such as driverless vehicles, virtual reality, and robot navigation. With the advancement of hardware and algorithm technology, future research will focus on enhancing the real-time performance of binocular stereo perception algorithms and achieving low-power edge computing, engineering the binocular stereo perception system, and further building the binocular stereo perception system with different resolutions to correct aberrations and accurately recognize objects at different distances. By introducing artificial intelligence algorithms, the system can realize adaptive processing and intelligent decision-making for complex scenes.

Author Contributions: Conceptualization, Y.Z. (Youpan Zhu), D.Z. and Y.Z. (Yongkang Zhou); methodology, Y.Z. (Youpan Zhu), D.Z., Y.Z. (Yongkang Zhou) and W.J.; software, Y.Z. (Youpan Zhu), D.Z., Y.Z. (Yongkang Zhou), L.Z. and G.W.; validation, Y.Z. (Youpan Zhu), D.Z., Y.Z. (Yongkang Zhou), L.Z. and Y.L.; formal analysis, G.W. and Y.L.; investigation, Y.Z. (Youpan Zhu), W.J., and G.W.; resources, W.J.; data curation, Y.Z. (Youpan Zhu), D.Z., and Y.Z. (Yongkang Zhou); writing—original draft preparation, Y.Z. (Youpan Zhu), D.Z., Y.Z. (Yongkang Zhou), W.J., L.Z., G.W. and Y.L.; writing—review and editing, Y.Z. (Youpan Zhu), D.Z., Y.Z. (Yongkang Zhou), W.J., L.Z., G.W. and Y.L.; visualization, G.W. and Y.L.; supervision, W.J. and Y.L.; project administration, W.J.; funding acquisition, W.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Defense Science and Technology Foundation Strengthening Plan (grant number 2021-JCJQ-JJ-1020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: We thank the National Defense Science and Technology Foundation Strengthening Plan for help in identifying collaborators for this work.

Conflicts of Interest: Authors Dan Zhang and Lingling Zhou were employed by the company Yunnan North Optical & Electronic Instrument Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Srivastav, A.; Mandal, S. Radars for Autonomous Driving: A Review of Deep Learning Methods and Challenges. In Proceedings of the Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 7 September 2023.
2. Feng, D.; Haase-Schutz, C.; Rosenbaum, L.; Hertlein, H.; Glaser, C.; Timm, F.; Wiesbeck, W.; Dietmayer, K. Deep Multi-modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges. *IEEE Trans. Intell. Transp. Syst.* **2019**, *22*, 1341–1360. [[CrossRef](#)]
3. Yan, J.; Jiao, H.; Pu, W.; Shi, C.; Dai, J.; Liu, H. Radar Sensor Network Resource Allocation for Fused Target Tracking: A Brief Review. *Inf. Fusion* **2022**, *86–87*, 104–115. [[CrossRef](#)]
4. Fernandez Llorca, D.; Hernandez Martinez, A.; Garcia Daza, I. Vision-based Vehicle Speed Estimation for ITS: A Survey. In Proceedings of the Computer Vision and Pattern Recognition, Nashville, TN, USA, 26 May 2021.

5. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [[CrossRef](#)]
6. Mur-Artal, R.; Tardós, J.D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [[CrossRef](#)]
7. Campos, C.; Elvira, R.; Rodríguez, J.J.G.; Montiel, J.M.; Tardós, J.D. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Trans. Robot.* **2021**, *6*, 37. [[CrossRef](#)]
8. Marr, D. Cooperative computation of stereo disparity. *Science* **1976**, *194*, 283–287. [[CrossRef](#)]
9. Marr, D. A Computational Theory of Human Stereo Vision. *R. Soc. B Biol. Sci.* **1979**, *204*, 301–328.
10. Poggi, M.; Tosi, F.; Batsos, K.; Mordohai, P.; Mattoccia, S. On the Synergies between Machine Learning and Binocular Stereo for Depth Estimation from Images: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5314–5334. [[CrossRef](#)]
11. Jin, Y.; Lee, M. Enhancing Binocular Depth Estimation Based on Proactive Perception and Action Cyclic Learning for an Autonomous Developmental Robot. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *49*, 169–180. [[CrossRef](#)]
12. Gyongy, I.; Dutton, N.A.; Henderson, R.K. Direct Time-of-Flight Single-Photon Imaging. *IEEE Trans. Electron Devices* **2022**, *69*, 2794–2805. [[CrossRef](#)]
13. Zhang, J.; Luo, B.; Su, X.; Wang, Y.; Chen, X.; Wang, Y. Depth range enhancement of binary defocusing technique based on multi-frequency phase merging. *Opt. Express* **2019**, *27*, 36717–36730. [[CrossRef](#)] [[PubMed](#)]
14. Real-Moreno, O.; Rodríguez-Quiñonez, J.C.; Sergiyenko, O.; Flores-Fuentes, W.; Mercorelli, P.; Ramírez-Hernández, L.R. Obtaining Object Information from Stereo Vision System for Autonomous Vehicles. In Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021; pp. 1–6.
15. Yang, F.; Rao, Y. Vision-Based Intelligent Vehicle Road Recognition and Obstacle Detection Method. *Int. J. Pattern Recognit. Artif. Intell.* **2020**, *34*, 2050020. [[CrossRef](#)]
16. Montemurro, N.; Scerrati, A.; Ricciardi, L.; Trevisi, G. The Exoscope in Neurosurgery: An Overview of the Current Literature of Intraoperative Use in Brain and Spine Surgery. *J. Clin. Med.* **2021**, *11*, 223. [[CrossRef](#)]
17. Yin, Y.; Yu, K.; Yu, C.; Bai, X.; Cai, Z.; Meng, X.; Yang, X. 3D Imaging Using Geometric Light Field: A Review. *Chin. J. Lasers* **2021**, *48*, 1209001.
18. Xu, Y.; Liu, K.; Ni, J.; Li, Q. 3D reconstruction method based on second-order semiglobal stereo matching and fast point positioning Delaunay triangulation. *PLoS ONE* **2023**, *17*, e0260466. [[CrossRef](#)]
19. Blake, R.; Wilson, H. Binocular vision. *Vis. Res.* **2011**, *51*, 754–770. [[CrossRef](#)]
20. Pomerleau, D. Efficient Training of Artificial Neural Networks for Autonomous Navigation. *Neural Comput.* **2014**, *3*, 88–97. [[CrossRef](#)]
21. Shao, N.; Li, H.G.; Liu, L.; Zhang, Z.L. Stereo Vision Robot Obstacle Detection Based on the SIFT. In Proceedings of the 2010 Second WRI Global Congress on Intelligent Systems, Wuhan, China, 16–17 December 2010; pp. 274–277.
22. Sivaraman, S.; Trivedi, M.M. Real-time vehicle detection using parts at intersections. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, USA, 16–19 September 2012; pp. 1519–1524.
23. Bontar, J.; Lecun, Y. Computing the Stereo Matching Cost with a Convolutional Neural Network. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: New York, NY, USA, 2014.
24. Li, P.; Chen, X.; Shen, S. Stereo R-CNN based 3D Object Detection for Autonomous Driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7636–7644.
25. Tang, Y.; Zhou, H.; Wang, H.; Zhang, Y. Fruit detection and positioning technology for a Camellia oleifera C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. *Expert Syst. Appl.* **2023**, *211*, 118573. [[CrossRef](#)]
26. Huang, H.; Yuan, E.; Zhang, D.; Sun, D.; Yang, M.; Zheng, Z.; Zhang, Z.; Gao, L.; Panzai, S.; Qiu, K. Free Field of View Infrared Digital Holography for Mineral Crystallization. *Cryst. Growth Des.* **2023**, *23*, 7992–8008. [[CrossRef](#)]
27. Ma, K.; Zhou, H.; Li, J.; Liu, H. Design of Binocular Stereo Vision System with Parallel Optical Axes and Image 3D Reconstruction. In Proceedings of the 2019 China-Qatar International Workshop on Artificial Intelligence and Applications to Intelligent Manufacturing (AIAIM), Doha, Qatar, 1–4 January 2019; pp. 59–62.
28. Hu, Y.; Chen, Q.; Feng, S.; Tao, T.; Asundi, A.; Zuo, C. A new microscopic telecentric stereo vision system—Calibration, rectification, and three-dimensional reconstruction. *Opt. Lasers Eng.* **2019**, *113*, 14–22. [[CrossRef](#)]
29. Peng, Y.T.; Cao, K.; Cosman, P.C. Generalization of the Dark Channel Prior for Single Image Restoration. *IEEE Trans. Image Process* **2018**, *27*, 2856–2868. [[CrossRef](#)]
30. Wu, J.; Liu, G. Noniterative calibration of a camera lens with radial distortion. *Meas. Sci. Technol.* **2012**, *23*, 105013. [[CrossRef](#)]
31. Zhou, Y.; Li, X.; Zhou, J.; Tang, X. A low delay highly dynamic range infrared imaging system for complex scenes based on FPGA. *Measurement* **2023**, *223*, 113745. [[CrossRef](#)]
32. Zhu, Y.; Li, H.; Li, L.; Jin, W.; Song, J.; Zhou, Y. A stereo vision depth estimation method of binocular wide-field infrared camera. In Proceedings of the Third International Computing Imaging Conference, Sydney, Australia, 1–3 June 2023; Volume 2921, pp. 252–264.
33. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded up robust features. *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]
34. Bansal, M.; Kumar, M.; Kumar, M. 2D object recognition: A comparative analysis of SIFT, SURF, and ORB feature descriptors. *Multimed. Tools Appl.* **2021**, *80*, 18839–18857. [[CrossRef](#)]

35. Ma, J.; Zhou, Z.; Wang, B.; Zong, H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys. Technol.* **2017**, *82*, 8–17. [[CrossRef](#)]
36. Mu, Q.; Wei, J.; Yuan, Z.; Yin, Y. Research on Target Ranging Method Based on Binocular Stereo Vision. In Proceedings of the 2021 International Conference on Intelligent Computing, Automation and Applications (ICAA), Nanjing, China, 25–27 June 2021; pp. 81–85.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.