

High-precision calibration of binocular camera with super-resolution technology

SUN Ting¹, MA Chao^{1,2,*}, SUN Tian³, PEI Shanshan⁴, and LONG Qian⁵

1. Faculty of Innovation Engineering, Macau University of Science and Technology, Macau 999078, China;

2. Zhuhai Macau University of Science and Technology Science and Technology Research Institute, Zhuhai 519000, China;

3. School of Mathematics and Statistics, Hubei University, Wuhan 430000, China;

4. Beijing Smarter Eye Technology Company Limited, Beijing 100191, China;

5. College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300000, China

Abstract: The calibration of parameters for onboard stereo cameras is crucial for achieving efficient visual-assisted driving. However, in practical scenarios, low-resolution images often result in inaccuracies in feature point extraction, thereby affecting the accuracy of camera parameter calibration. To address this issue, this paper proposes a self-calibration method for stereo cameras based on joint de-noising, de-mosaic, de-ringing, and super-resolution network (JDDDSN) super-resolution reconstruction. By reconstructing images into higher-resolution images with richer details, feature points are extracted for extrinsic calibration of stereo cameras. For real-world driving scenarios, the reconstructed images achieve noise and ringing artifact reduction while obtaining clearer high-resolution images. This study further investigates the impact of the super-resolution reconstruction network on target area calibration at various distances. Additionally, it highlights the significant role of super-resolution in enhancing stereo camera calibration accuracy by removing dynamic points and focusing on static regions. Through a series of experiments, this paper validates the effectiveness and practicality of the JDDDSN super-resolution reconstruction network in improving stereo camera calibration accuracy, demonstrating its application value in the field of stereoscopic vision.

Keywords: super-resolution reconstruction, camera calibration, self-calibration, binocular camera.

DOI: 10.23919/JSEE.2026.000105

1. Introduction

The stereo vision-based autonomous driving system relies on cameras to perceive the environment, enabling functions such as obstacle avoidance and path planning [1,2]. Accurate camera calibration is crucial for determining the three-dimensional (3D) coordinates of objects, which is essential for tasks like distance prediction and obstacle

avoidance [3–6]. However, factors such as lighting conditions and noise can degrade image quality, leading to reduced precision in feature point detection and, consequently, affecting the accuracy of camera calibration. Super-resolution (SR) reconstruction technology addresses these challenges by denoising and deblurring low-resolution images, transforming them into clearer, high-resolution versions, which may help resolve the issue. This approach can improve feature point detection accuracy, enhance camera calibration precision, and, in turn, boost the overall reliability of the stereo vision-based system.

Single image super-resolution (SISR) aims to reconstruct high-resolution images from low-resolution inputs [7–9]. Deep learning techniques, particularly convolutional neural networks (CNNs), have shown significant advancements over traditional image reconstruction methods in this domain [10–13]. To further enhance edge and texture details, Xu et al. [14] proposed the joint demosaicing and super-resolution (JDSR) network. More recent studies have increasingly focused on integrating super-resolution with tasks such as denoising and demosaicing [15–18]. However, high-frequency quantization during reconstruction often leads to ringing artifacts, especially along sharp edges [19–21]. To address this issue, Sun et al. [22] introduced the joint denoising, demosaicing, de-ringing, and super-resolution network (JDDDSN), which effectively mitigates ringing artifacts and improves overall image quality.

Traditional high-precision camera calibration methods typically rely on reference objects [23–25]. These techniques extract feature points from target images, establish equations that relate pixel coordinates to real-world spatial coordinates, and compute calibration parameters using optimization algorithms. However, these methods impose

strict requirements on the calibration environment and reference object configuration, limiting their application to offline calibration and posing challenges for large-scale production and routine maintenance [26–28]. Self-calibration methods, in contrast, can compute calibration parameters without relying on specific reference objects, thus overcoming these limitations. By leveraging geometric consistency constraints, such as parallel lines and feature points across multiple images, these methods enable effective computation of camera parameters, broadening the scope of camera calibration applications [29–33].

During vehicle operation, activating a self-calibration system enables real-time updates of camera parameters, even as objects in the driving scene continuously change. Different object categories provide varying levels of texture information in images, which can affect the accuracy of image feature extraction [34]. In practical scenarios, dynamic objects are commonly encountered, and their associated feature points may degrade localization accuracy, posing challenges for calibration. To address this, semantic segmentation techniques can be employed for pixel-level scene classification, enabling the identification and segmentation of static objects in road environments [35–39]. Among these, the object-contextual representations network (OCRNet) leverages global contextual information to enhance the semantic understanding of various objects. It effectively aggregates information from regions with similar pixel properties, improving the capture of semantic relationships and, in turn, boosting segmentation accuracy. This approach has demonstrated significant results on datasets such as Cityscapes and ADE20K [40]. However, in complex scenes, semantic segmentation can suffer from under-segmentation issues. To mitigate this, Xie et al. [41] proposed a dynamic mask repair method that uses clustering algorithms to process disparity maps, effectively addressing these challenges.

To improve the accuracy and practicality of binocular camera calibration in dynamic driving environments, it is crucial to ensure effective feature point extraction while simultaneously enhancing image quality. This paper presents a camera calibration system that integrates super-resolution reconstruction with self-calibration, specifically designed for dynamic environments. The main contributions of this work are as follows:

(i) A super-resolution reconstruction network, JDDDSN, combined with distance information, is employed for binocular camera self-calibration in road scenes. Initially, the original left and right images are processed by the super-resolution network, resulting in super-resolution left image and right image. Next, feature points are extracted from these super-resolution images. Camera calibration is then performed by filtering the points

based on their distances from the camera, which enables the determination of the external camera parameters.

(ii) The influence of static object feature points derived from basic super-resolution reconstruction on camera calibration accuracy is investigated. Building upon super-resolution reconstruction, the OCRNet semantic segmentation network is employed to classify different object types and extract feature points from static objects. These static object feature points are then utilized for the extrinsic parameter calibration of binocular cameras.

(iii) Semantic segmentation often encounters under-segmentation issues. To mitigate this, we refine the filtering of dynamic points by integrating disparity clustering with multi-view geometric methods, leading to more accurate static feature points for camera extrinsic parameter calibration.

(iv) Experimental results confirm that, in driving scenarios, super-resolution reconstruction enables the extraction of more accurate feature point positions, thereby significantly improving the accuracy of camera calibration.

2. Methods and details

Image super-resolution reconstruction aims to enhance the resolution of low-quality images, restoring them to high-resolution images with rich texture details [42]. In driving scenarios, images are often affected by noise and poor lighting conditions, which degrade their quality and negatively impact feature point extraction, leading to significant localization errors. Therefore, applying super-resolution techniques to reconstruct high-resolution images is essential for ensuring the accurate extraction of feature points in subsequent processing steps.

In driving scenarios, the significant differences between distant and near feature points can impact the accuracy of camera parameter estimation. Additionally, moving objects in the scene can introduce errors in feature point localization, and the uncertainty associated with dynamic points during driving can compromise the accuracy of camera calibration. To address these challenges, we first apply stereo vision algorithms to the reconstructed, high-quality super-resolution images to estimate the distances of feature points. This allows us to filter feature points based on their distances and evaluate the effect of near and far feature points on camera calibration accuracy. Next, we employ a semantic segmentation network to generate probabilistic masks for dynamic objects, effectively removing dynamic feature points to ensure that only static features are used for camera parameter estimation. To further mitigate potential under-segmentation issues in complex environments, we integrate disparity clustering and multi-view geometry techniques, which enhance the accuracy of dynamic point exclusion.

In summary, this section describes the super-resolution reconstruction algorithm applied prior to feature point detection, the principles of stereo camera calibration, and the clustering and multi-view geometry techniques employed to ensure the thorough removal of dynamic feature points.

2.1 Super-resolution reconstruction algorithm

Image acquisition can be compromised by various factors, resulting in the degradation of the original high-resolution image.

lution image (I_{HR}) into a low-resolution image (I_{LR}). This phenomenon is referred to as image degradation [43]. Fig. 1 illustrates the image degradation process, which can be mathematically modeled as follows:

$$I_{LR} = D(I_{HR}, \alpha) \quad (1)$$

where I_{LR} is the degraded low-resolution image, I_{HR} is the original high-resolution image, $D(\cdot)$ is the nonlinear degradation process, α is the nonlinear parameter of degradation.

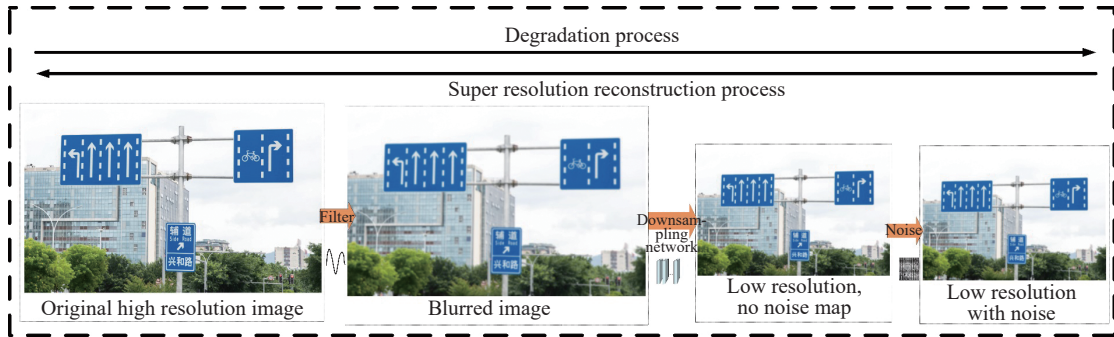


Fig. 1 Image degradation and super-resolution reconstruction

The image super-resolution reconstruction process is essentially the inverse of the image degradation process described above. In this approach, a high-resolution image is reconstructed from a low-resolution image, which can be mathematically represented as

$$I_{SR} = F(I_{LR}; \theta) \quad (2)$$

where I_{SR} is the reconstructed super-resolution image, $F(\cdot)$ is the mapping function between I_{LR} and I_{SR} , and θ is the parameter of the mapping function.

The goal of super-resolution reconstruction is to minimize the difference between the reconstructed super-resolution image (I_{SR}) and the original high-resolution image, ensuring that they are as similar as possible.

This objective can be expressed as

$$\bar{\theta} = \arg_{\theta} \min L(I_{SR}, I_{HR}) + \lambda \Phi(\theta) \quad (3)$$

where I_{SR} is super-resolution image, L is the loss function, λ is the balance factor, and $\Phi(\theta)$ is the regularization term.

Deep learning-based image super-resolution algorithms leverage CNNs to extract features from low-resolution images, while learning both linear and nonlinear mappings between low- and high-resolution images. The JDDDSN algorithm simultaneously performs four tasks: denoising, demosaicking, deblurring, and super-resolution, providing an end-to-end deep learning solution. This algorithm incorporates a balanced attention mechanism (BAM) and an adversarial loss function, effectively recovering image tex-

ture details while preserving high-frequency information that is often lost during the denoising process due to smoothing operations. Specifically, the network is built upon a U-shaped network (U-Net) architecture, primarily consisting of convolutional (Conv) layers for feature extraction and rectified linear unit (ReLU) as activation functions. The architecture of this network is shown in Fig. 2.

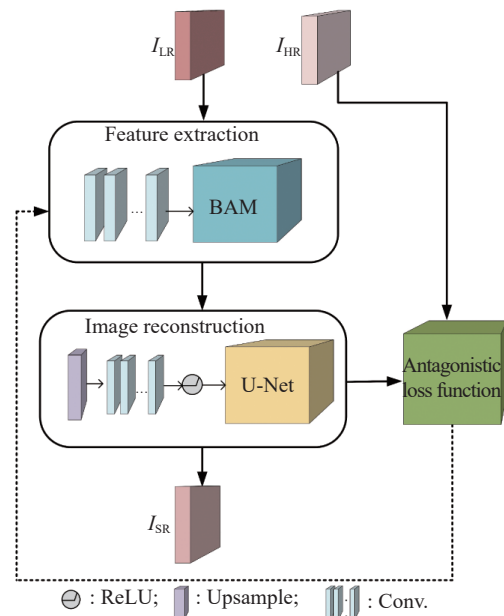


Fig. 2 JDDDSN network architecture

The network model consists of three main components: feature extraction, image reconstruction, and overall loss calculation. In the feature extraction module, we assume the noise level is known in advance and define it as a hyperparameter. The feature extraction process begins with preprocessing steps, including a convolutional layer and a transposed convolutional layer, to enhance image resolution and improve the quality of feature mapping. Subsequently, the extracted features pass through several basic units, with the BAM serving as the core unit of the feature extraction module to enhance focus on key features.

In the reconstruction stage, upsampling convolutions are applied to transform the extracted features into full-resolution ones. To mitigate the loss of high-frequency information caused by smoothing during forward propagation, a UNet module is incorporated at the final stage of the network, ensuring the fine preservation of these details and enhancing the overall reconstruction quality.

In the overall loss calculation component, the generated high-resolution images are compared with the ground-truth images using an adversarial loss module to measure the perceptual differences between the generated and real images. The final loss value is computed by combining this adversarial loss with the losses calculated between the ground-truth images and the blurred images, and the resulting loss is backpropagated to optimize the entire network. This design facilitates effective collaboration among different tasks during image restoration, balancing the reconstruction of texture details with denoising performance, and significantly enhancing the overall performance of the model.

The total loss function L_{total} of the network is defined as follows:

$$\begin{cases} L_{\text{total}} = \alpha \text{MSE}(I_{\text{LR}}, I_{\text{SR}}) + \beta L_{\text{ant_loss}} \\ L_{\text{ant_loss}} = F_3(F_2(F_1(I_{\text{LR}})), I_{\text{HR}}) \end{cases} \quad (4)$$

where F_3 , F_2 and F_1 represent the feature extraction module, image reconstruction module, and adversarial loss function module, respectively. α and β are weighting coefficients. The adversarial loss $L_{\text{ant_loss}}$ is calculated by the adversarial loss function module using the generated image and the original high-resolution image.

Currently, the mainstream objective evaluation methods for SISR reconstruction primarily include peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM).

PSNR is defined as the ratio of the maximum possible signal power to the power of noise that affects the accuracy of its representation. It focuses primarily on the differences between pixel values and is typically expressed in logarithmic decibel units. In the context of image super-

resolution reconstruction, reconstruction errors are treated as noise, and the maximum possible power corresponds to the highest possible gray value of the image pixels. Typically, the reconstruction error is quantified using the mean squared error (MSE). For the high-resolution image and the super-resolution image obtained after reconstruction, the MSE between the I_{SR} and I_{HR} can be calculated as follows:

$$\text{MSE} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (I_{\text{HR}}(i, j) - I_{\text{SR}}(i, j))^2 \quad (5)$$

where M, N denotes the width and height of the image, respectively. Therefore, the PSNR of I_{SR} can be expressed as follows:

$$\text{PSNR} = 10 \lg \left(\frac{L^2}{\text{MSE}} \right). \quad (6)$$

For 8-bit images, L is usually 255. The unit of PSNR is dB, and the larger the PSNR value, the higher the quality of the reconstructed image.

SSIM is used to evaluate the structural similarity between the reconstructed image I_{SR} and the high-resolution image I_{HR} [44]. The SSIM index assesses the relationship between two images by considering three types of statistical information: structural similarity, luminance, and contrast. The calculation of SSIM can be specifically represented as follows:

$$\begin{cases} l(I_{\text{SR}}, I_{\text{HR}}) = \frac{2\mu_{\text{SR}}\mu_{\text{HR}} + C_1}{\mu_{\text{SR}}^2 + \mu_{\text{HR}}^2 + C_1} \\ c(I_{\text{SR}}, I_{\text{HR}}) = \frac{2\sigma_{\text{SR}}\sigma_{\text{HR}} + C_2}{\sigma_{\text{SR}}^2 + \sigma_{\text{HR}}^2 + C_2} \\ s(I_{\text{SR}}, I_{\text{HR}}) = \frac{2\sigma_{\text{HRSR}} + C_3}{\sigma_{\text{SR}}\sigma_{\text{HR}} + C_3} \\ \text{SSIM} = [l(I_{\text{SR}}, I_{\text{HR}})]^\alpha \cdot [c(I_{\text{SR}}, I_{\text{HR}})]^\beta [s(I_{\text{SR}}, I_{\text{HR}})]^\gamma \end{cases} \quad (7)$$

where μ_{HR} and μ_{SR} denote the mean of I_{HR} and I_{SR} , respectively; σ_{HR} and σ_{SR} denotes the standard deviation of I_{HR} and I_{SR} , respectively, and σ_{HRSR} denotes the covariance between them. $l(I_{\text{SR}}, I_{\text{HR}})$ is the brightness comparison between images, $c(I_{\text{SR}}, I_{\text{HR}})$ is the contrast comparison between images; $s(I_{\text{SR}}, I_{\text{HR}})$ is the structure comparison between images. In general setting, $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$. SSIM values range from $[0, 1]$, the larger the SSIM value, the better the effect of the generated image. This index evaluates the quality of the reconstructed image from the perspective of the human visual system, so it can better meet the requirements of perceptual assessment.

2.2 Principle of binocular camera calibration

A binocular camera system captures the spatial information of objects in the 3D world by leveraging the princi-

ples of stereo vision [45]. The epipolar geometry between the two images forms the foundation of stereo vision. For a point X in the world coordinate system, the corresponding image point in the left camera image I is x and the corresponding image point in the right camera image I' is x' . The line connecting the camera centers C and C' is

the baseline, and the plane defined by this baseline is the epipolar plane, as illustrated in Fig. 3. The point where the baseline intersects the image planes of the two cameras are called the epipoles e and e' , respectively. The intersection of the epipolar plane with the image planes is called the epipolar line.

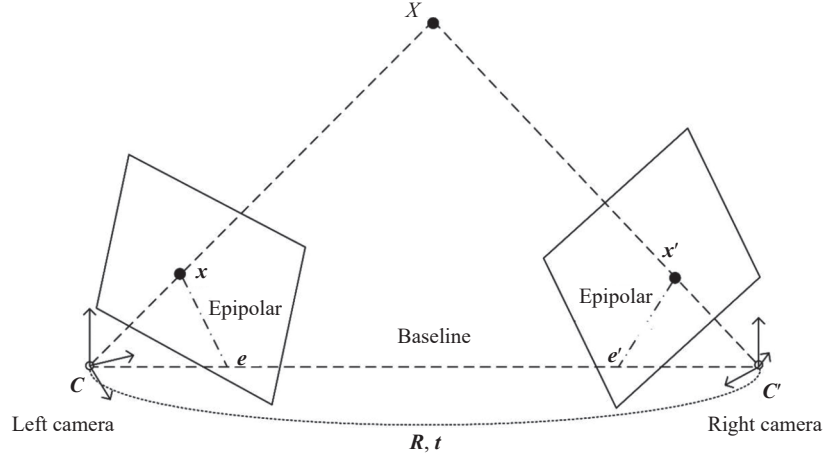


Fig. 3 Epipolar geometry

The epipolar geometry establishes a constraint on the corresponding points in the left and right images:

$\forall x \in I$, in I' there exists an epipolar line l'_x corresponding to it, and the corresponding point lies on l'_x . Similarly, $\forall x' \in I'$, in I there exists an epipolar line l_x corresponding to it, and the corresponding point x lies on l_x .

If the projection matrices of the left and right cameras are P and P' , and the intrinsic matrices are K and K' , the parametric equation for the back-projected ray $l_x^{(b)}$ of any point $\forall x \in I$ is

$$X(s) = P^+x + sC, \quad s \in (-\infty, \infty) \quad (8)$$

where P^+ is the pseudo-inverse of P , meaning $P^+P = I$. C is the optical center of the first camera, so $PC = 0$.

Thus, it can be obtained l'_x :

$$\begin{aligned} l'_x &= e' \times x' = (P'C) \times (P'X(s)) = \\ &= (P'C) \times (P'P^+x + sP'C) = \\ &= (P'C) \times (P'P^+x) = [e']_{\times} P'P^+x. \end{aligned} \quad (9)$$

The matching points of the left and right images are on the same plane

$$x'^T l'_x = x'^T F x = 0, \quad (10)$$

$$F = [e']_{\times} P'P^+. \quad (11)$$

By normalizing the feature points using the camera intrinsic matrices K and K' , the following constraint equation can be obtained:

$$\begin{cases} \hat{x} = K^{-1}x \\ \hat{x}' = K'^{-1}x' \\ \hat{x}'^T K'^T F K \hat{x} = \hat{x}'^T E \hat{x} = 0 \end{cases} \quad (12)$$

where the relationship between the essential matrix E , the rotation matrix R and the translation vector t is given by

$$E = [t]_{\times} R \quad (13)$$

where $[t]_{\times}$ is the skew-symmetric matrix of t .

Once the matching point pairs are obtained from the left and right images, the essential matrix can be decomposed into matrices R and t according to their singular values.

The binocular camera system has the advantage of accurate 3D spatial resolution of x, x' through imaging points in the left and right camera views. Using the parallax principle, the precise position (X, Y, Z) of each imaging point in the world coordinate system can be calculated. Usually, the camera coordinate system is set to the world coordinate $\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j$ system, so as to obtain the 3D space point (X_c, Y_c, Z_c) under the camera coordinate system, which accurately reflects its position in the world coordinate system. It is shown as follows:

$$\begin{cases} Z_c = \frac{bf}{disp} \\ X_c = \frac{x - c_x}{f} \cdot Z_c \\ Y_c = \frac{y - c_y}{f} \cdot Z_c \end{cases} \quad (14)$$

where b is the baseline distance of the binocular camera, f is the focal length of the left and right cameras, and $disp$ is the disparity of the imaging point pair.

2.3 Disparity clustering

In complex environments, semantic segmentation networks may suffer from under-segmentation, leading to category masks that do not fully cover the objects in the image. To ensure comprehensive extraction of feature points from static objects, clustering algorithms can be employed to supplement object category information in the image. Additionally, it has been observed that using disparity map information yields better clustering results than relying solely on RGB images [46]. Therefore, we utilize the K-means clustering algorithm to cluster objects based on disparity information.

The K-means algorithm is a classical partition-based clustering algorithm, known for its efficiency and simplicity in implementation, without the need for pre-labeled data [47]. The algorithm operates through an iterative process, which involves the following steps. First, k data objects are randomly selected as the initial cluster centers. Next, each of the remaining objects is assigned to the cluster with the nearest center, based on the calculated distance to each cluster center. The cluster centers are then updated by recalculating the mean value of all objects within each cluster. This process is repeated until convergence, using the following formula:

$$\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j \quad (15)$$

where μ_i is the center of the cluster, $|C_i|$ is the number of data points in the i th cluster, and x_j is the point belonging to the i th cluster.

The process is repeated until convergence is achieved, which occurs when the changes in cluster centers become negligible. Typically, the MSE is used as the objective function, aiming to minimize the sum of squared distances between each data point and its nearest cluster center. The objective function can be represented as follows:

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (16)$$

where J is the objective function, k is the number of clusters, x_j is the data point, and μ_i is the center of cluster C_i .

To account for the possibility that different objects may share the same disparity value, we also incorporate the pixel coordinates of the feature points into the cluster-

ing criteria. Assuming that (d, x, y) represents both the disparity value and the pixel coordinates of the feature points, the clustering criteria can be expressed as follows.

To better state this method, we calculate the cluster center of the i th cluster Mid_i :

$$\text{Mid}_i = \left(\frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}, \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}, \frac{1}{N_i} \sum_{j=1}^{N_i} d_{ij} \right) \quad (17)$$

where N_i is the number of feature points in the i th cluster; x_{ij} , y_{ij} and d_{ij} are the coordinates and visual deviation of the j th feature point in the i th cluster, respectively. By iterative adjustment, we ensure that the distance of points within the same cluster to the cluster center is minimized, while the distances to other clusters are maximized.

3. Experiment and details

This study aims to evaluate the performance of the JDDDSN super-resolution reconstruction network using images captured by onboard stereo cameras in real-world driving scenarios, with a specific focus on its ability to enhance the accuracy of stereo camera parameter self-calibration. In Subsection 3.1, we assess the effectiveness of the JDDDSN model using both public datasets and images captured in actual driving environments. Subsequently, Subsection 3.2 investigates the impact of feature points at varying distances, within images reconstructed by the JDDDSN, on the self-calibration results of stereo camera parameters in driving scenarios. Finally, Subsection 3.3 examines the impact of static feature points, derived from super-resolution reconstruction, on the self-calibration of stereo camera parameters. After applying the JDDDSN super-resolution reconstruction to the images, dynamic feature points are filtered using the OCRNet semantic segmentation network, allowing for the extraction of feature points from static object regions. Additionally, disparity clustering and multi-view geometry techniques are utilized to address under-segmentation in dynamic regions. The analysis then evaluates how these extracted feature points influence the self-calibration results of the stereo camera parameters.

The overall three-stage experimental pipeline is illustrated in Fig. 4, where Stage 1 corresponds to JDDDSN super-resolution reconstruction, Stage 2 to distance-based feature point filtering, and Stage 3 to dynamic point removal via semantic segmentation and disparity-based clustering.

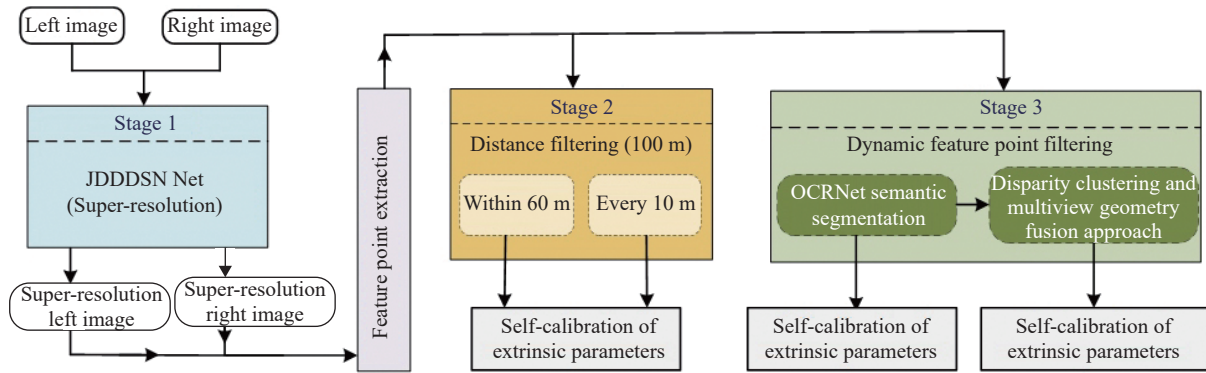


Fig. 4 Overall three-stage experimental pipeline

In the experiments, we utilized a binocular camera system equipped with a field-programmable gate array (FPGA) and a dual-core processor based on the reduced instruction set computer (RISC) architecture, namely an advanced RISC machine (ARM) processor, supporting controller area network (CAN) interfaces. As detailed in Table 1, this system comprises a left and a right camera with identical focal lengths and optical centers. The stereo camera was employed to capture image data across various scenes. Furthermore, the pseudo-code for the proposed algorithm is presented in Algorithm 1.

Table 1 Binocular camera parameters

Name	Specific parameter
Resolution	1 280×720 pixels
Baseline/mm	120
(c_x, c_y)	(649.5382, 340.9208) pixels
Focal length of lens/mm	8
Dynamic range/dB	120
Processor	FPGA, Dual-Core ARM
Support interface	CAN, RS485, network port
Dimensions/cm ³	25×15×8

Algorithm1 Binocular image processing and camera parameter calculation based on super-resolution reconstruction

%This pseudocode describes how to achieve self-calibration of stereo camera extrinsic parameters using JDDDSN reconstruction, feature extraction and matching, semantic segmentation, and dynamic point clustering and filtering.

Input Images ‘Original_LeftImage’ & ‘Original_RightImage’

Output Rotation matrix ‘ R ’, translation Vector ‘ t ’

Procedure

Step 1: JDDDSN super-resolution reconstruction (%Subsection 3.1-Use the original left and right images to perform JDDDSN super-resolution reconstruction to obtain higher-quality input)

LeftImg_SR, RightImg_SR = JDDDSNReconstruction (Original_LeftImage, Original_RightImage)

Step 2: feature extraction and matching (%Subsection 3.2-Extract and match feature points from the two super-resolution images in Step 1.)

Matches_pts = MatchFeatures (LeftImg_SR, RightImg_SR)

Step 3: filter matches based on distance (%Subsection 3.2-Use the matched feature points and filter out unreliable matches based on distance)

Matches_pts = FilterMatches (Matches_pts, Threshold)

Initial calculation of rotation and translation matrix (%with the filtered matching points to calculate the initial $[R, t]$)

$[R, t]$ = calculate essential matrix and decompose (matches_pts)

Step 4: dynamic point removal based on semantic segmentation (%Subsection 3.3-with the OCRNet semantic segmentation model to identify static and dynamic regions in the super-resolution images, filtering out dynamic feature points)

Perform semantic segmentation and remove dynamic points

Static_matches_pts = remove dynamic points (matches_pts, LeftImg_SR, RightImg_SR)

Recalculate rotation and translation matrix after static point filtering

$[R, t]$ = Calculate essential matrix and decompose (static_matches_pt)

Step 5: Dynamic feature point clustering and filtering (%with K-means clustering applied to the disparity map from super-resolution images to further identify and filter dynamic feature points in the matches)

% Calculate the final rotation and translation matrix with filtered matches

cluster_disp = Perform K-means clustering (disparity_map)

%with multi-view geometry to detect possible dynamic points in the clusters to further optimize the set of feature points

```

cluster_left_pts, cluster_right_pts = Detect clusters using
multi view geometry (LeftImg_SR, RightImg_SR, cluster_disp)
Matches_pts = filter matches by clusters (matches_pts,
cluster_left_pts, cluster_right_pts)
# Final calculation of rotation and translation matrix
[R, t] = Calculate essential matrix and decompose
(matches_pts)
End of procedure

```

3.1 Super-resolution reconstruction image experiment

During the driving process, various road scenes are

encountered, and factors such as vehicle speed and lighting conditions can significantly impact the quality of images captured by the stereo camera system [48,49]. Vision-based algorithms for binocular camera calibration require high-resolution images to accurately extract feature point positions. To achieve effective reconstruction for urban road scenes, this study utilizes the widely adopted DIV2K dataset, commonly used in image super-resolution tasks, along with 800 high-resolution images of different scenes captured using a high-definition camera, to train the JDDDSN network. The examples of the captured scenes are shown in Fig. 5.

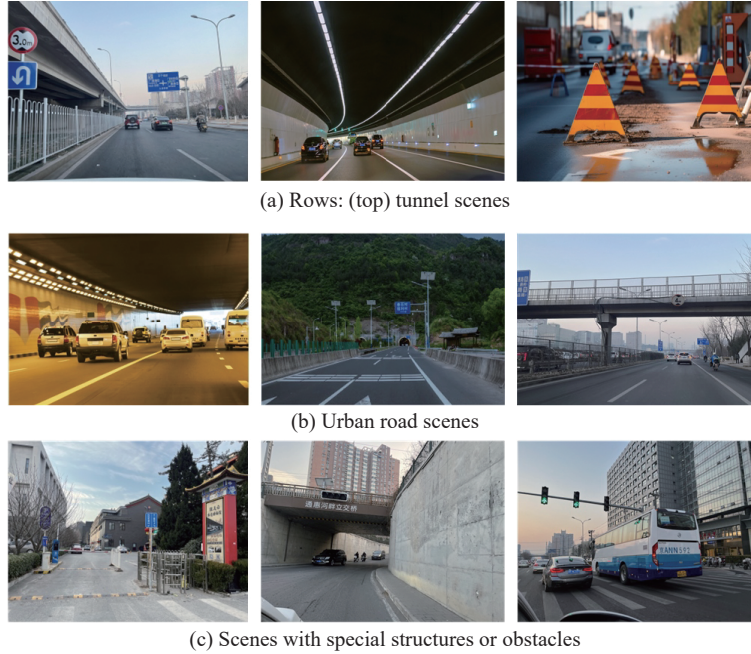


Fig. 5 Dataset captured by the camera (part)

The JDDDSN network was trained using a dataset of 1 500 high-resolution images, comprising 800 images from the DIV2K dataset and 700 high-resolution road scene images collected in-house. All images underwent preprocessing, including resizing to a uniform resolution and standard normalization. The AdamW optimizer was employed with an initial learning rate of $1e-4$, and a step decay strategy halved the learning rate every 50 iterations to facilitate model convergence. The training was conducted in the PyTorch 1.12.1 framework, with an NVIDIA GeForce RTX 3080 GPU. The total training time was approximately 35 h.

To evaluate the performance of the JDDDSN network, we compared it with classical algorithms and lightweight image super-resolution methods, including bicubic, super-resolution convolutional neural network (SRCNN), super-resolution generative adversarial network (SRGAN), and very deep super resolution (VDSR). We conducted a quantitative comparison using PSNR and SSIM indices across four benchmark datasets, as detailed in Table 2. Additionally, we performed a qualitative assessment of the JDDDSN model against other classical methods through visual comparison, as illustrated in Fig. 6.

Table 2 Super-resolution reconstruction results under different scaling factors

Sampling rate	Method	Set5(PSNR/dB)/(SSIM)	Set14(PSNR/dB)/(SSIM)	BSD100(PSNR/dB)/(SSIM)	Urban100(PSNR/dB)/(SSIM)
2×	Bicubic	33.66/0.929	30.24/0.869	29.56/0.843	26.88/0.840
	SRCNN	36.66/0.954	33.45/0.907	31.36/0.888	29.50/0.895
	SRGAN	—	—	30.48/0.863	29.40/0.847
	VDSR	37.53/0.959	33.03/0.912	31.90/0.896	30.76/0.914
	JDDDSN	37.88/0.960	33.65/0.923	32.01/0.902	31.12/0.931

Continued

Sampling rate	Method	Set5(PSNR/dB)/(SSIM)	Set14(PSNR/dB)/(SSIM)	BSD100(PSNR/dB)/(SSIM)	Urban100(PSNR/dB)/(SSIM)
3×	Bicubic	30.39/0.868	27.55/0.774	27.21/0.739	24.46/0.735
	SRCNN	32.75/0.909	29.30/0.822	28.41/0.786	26.24/0.799
	SRGAN	30.87/0.896	28.03/0.781	26.90/0.710	25.23/0.759
	VDSR	33.66/0.921	29.77/0.831	28.83/0.797	27.14/0.827
	JDDDSN	33.64/0.916	29.81/0.827	28.79/0.791	27.56/0.843
4×	Bicubic	28.42/0.810	26.00/0.703	25.96/0.668	23.14/0.658
	SRCNN	30.48/0.863	27.50/0.751	26.90/0.710	24.52/0.722
	SRGAN	29.40/0.847	26.02/0.739	25.16/0.668	24.38/0.737
	VDSR	31.35/0.883	28.01/0.768	27.29/0.725	25.18/0.752
	JDDDSN	30.99/0.871	28.11/0.771	27.70/0.741	25.89/0.786

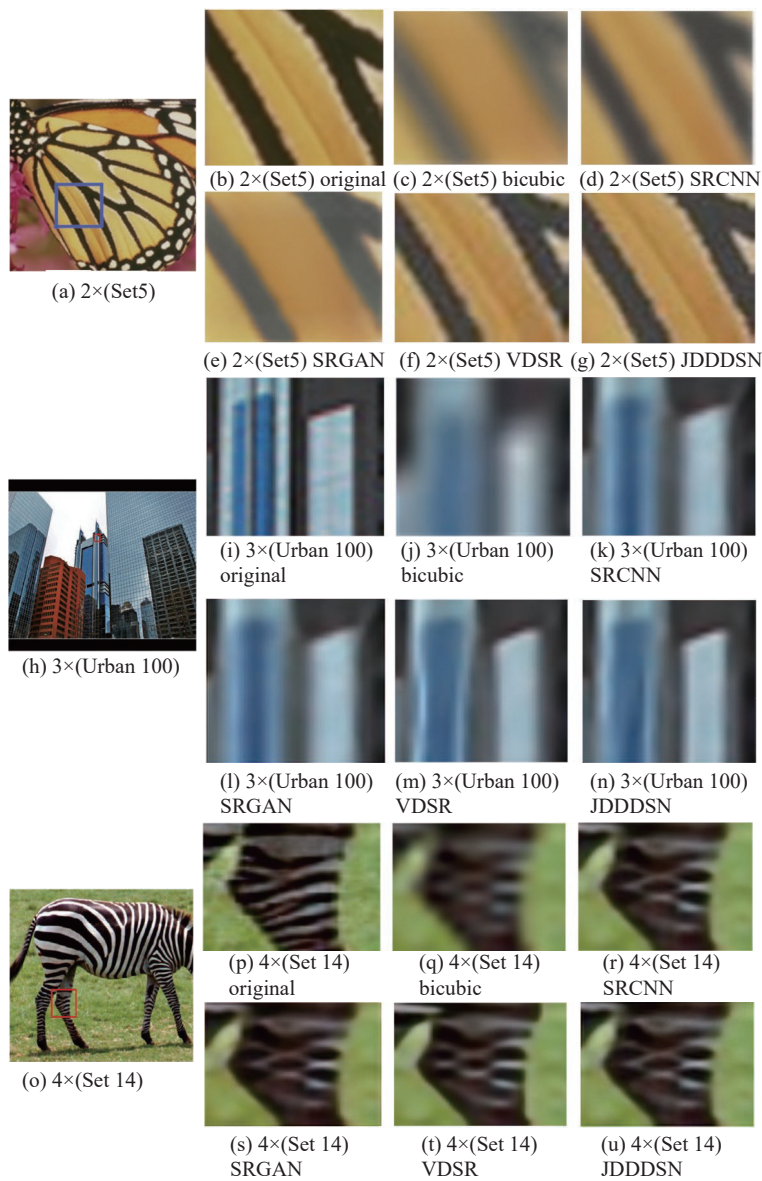


Fig. 6 Reconstruction results of different algorithms with different scaling factors

As seen from the evaluation metrics in Table 2, JDDDSN achieves the best performance on the Set14 and Urban100 datasets, and it also exhibits the best performance on most of the metrics for the Set5 and BSD100 datasets. Fig. 6 demonstrates the performance of images reconstructed by the JDDDSN model in terms of texture details, particularly in the clear and detailed reconstruction of the texture edges on the butterfly wings. Moreover, the texture details of building edges in the distant scene are also more sharply defined.

To further assess the robustness of the JDDDSN model in real-world driving scenarios, we conducted a qualitative comparison of images across various scenes, as

shown in Fig. 7. In challenging environments, such as night tunnel scenes, the JDDDSN model effectively recovers fine details of road markings and vehicle contours that are often lost under low-light conditions. Furthermore, in daytime scenarios, JDDDSN successfully mitigates artifacts, such as ringing effects beneath guardrails, which are more pronounced in images processed by other methods. These results clearly demonstrate that the JDDDSN model significantly enhances image sharpness and clarity. The improvements in noise reduction and preservation of fine details directly contribute to greater stability in subsequent feature point extraction, reducing matching errors, and further improving calibration accuracy.

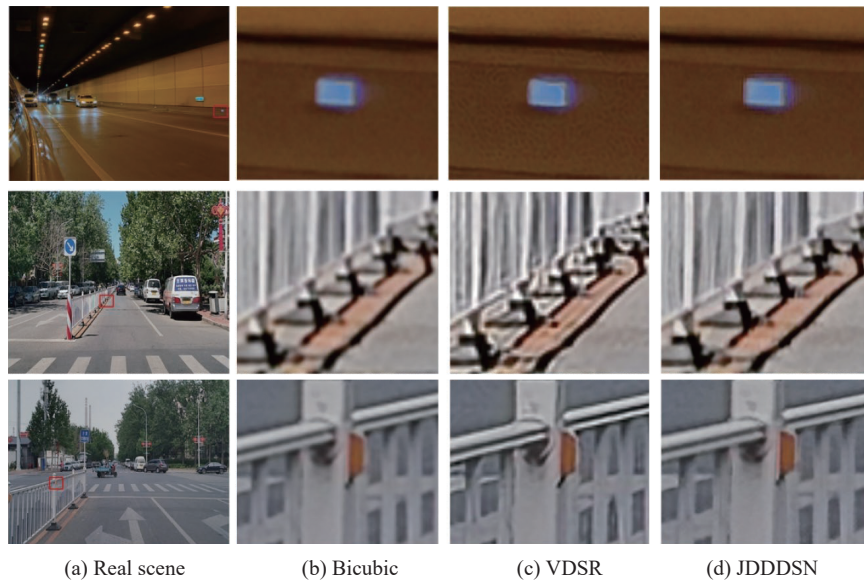


Fig. 7 Reconstruction effects of different algorithms in real scenes when the scaling factor is 2

3.2 Distance calibration experiment based on super-resolution reconstruction

Vehicle driving is a dynamic process in which the relative distance between the camera and the objects it captures continuously changes, influencing the image resolution. This fluctuation in resolution leads to varying confidence levels in the extracted feature points, which can adversely affect the self-calibration of the stereo camera's external parameters [50–52]. To address this issue and improve the reliability of feature point extraction at different distances, this study investigates the impact of super-resolution reconstruction of object feature points at varying distances on the accuracy of camera calibration.

We used road scenes captured by stereo cameras, as shown in Fig. 8. The left and right images of the original road scene were reconstructed using the VDSR and

JDDDSN networks to generate super-resolution images. Feature points were then extracted from these images using the scale-invariant feature transform (SIFT) algorithm, and the fast library for approximate nearest neighbors (FLANN) matching method was applied to obtain corresponding point pairs for camera calibration. The rotation (R) and translation (T) matrices of the camera were then calculated. Following this, the original images were rectified using the R and T matrices to produce newly corrected stereo images. The positions of the feature points in the rectified images were obtained, and the relative distances between the feature points and the camera were determined using the disparity formula (14) derived in Subsection 2.2. Finally, the disparity errors between feature points in the corrected images were evaluated using the designated formula $\text{error}_y = |y_{\text{right}} - y_{\text{left}}|$.

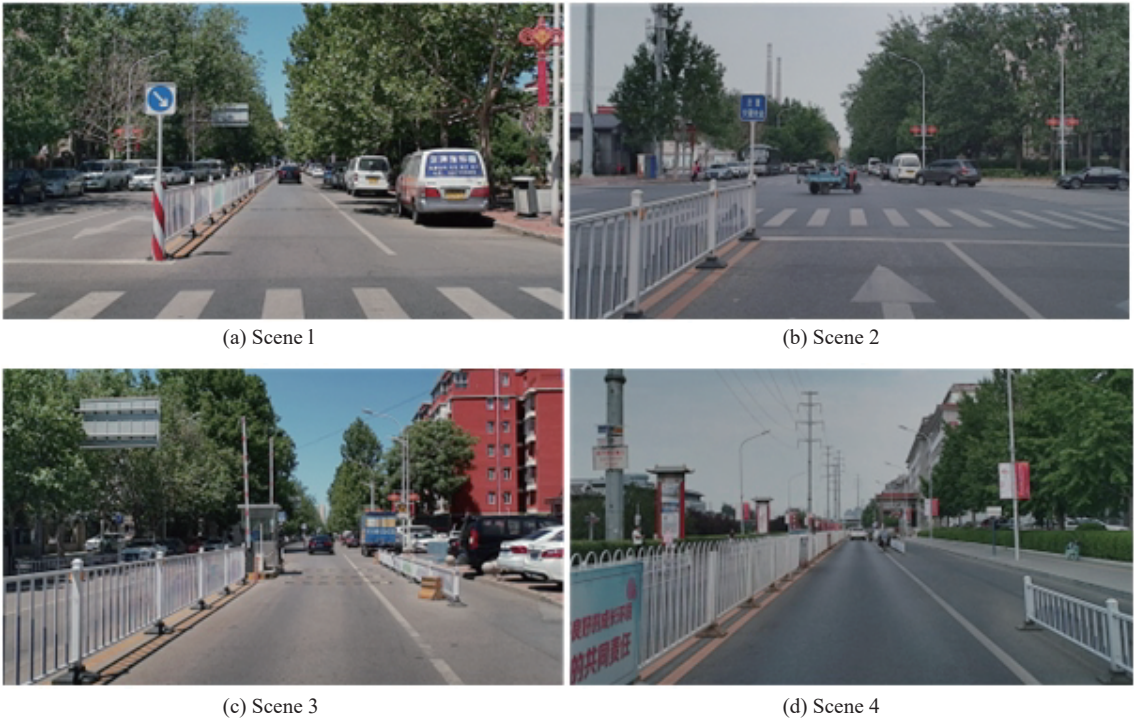
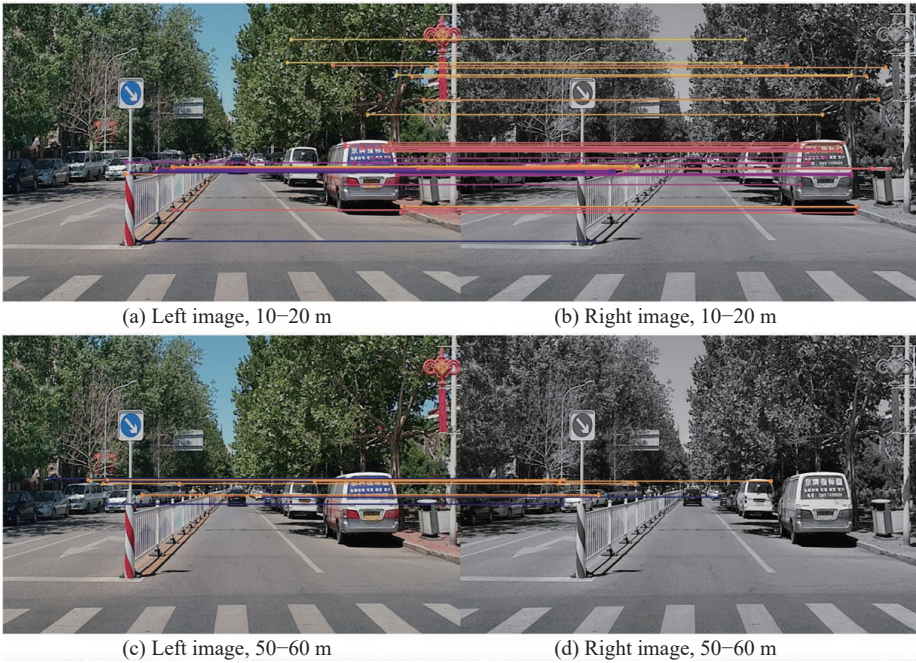


Fig. 8 Different road scenes

Fig. 9 illustrates the distribution of matching points from near to far distances in Scenario 1. Initially, we constrained the distance between the camera and objects to within 60 m for preliminary exploration. Feature points were extracted from the original left and right images, as well as the super-resolved images recon-

structed by the VDSR and JDDDSN networks, across four sets of scenes. Matching points within 60 m were selected based on stereo matching principles, and these matched points were used to calibrate the camera. The experimental data obtained are shown in Table 3 and Fig. 10.



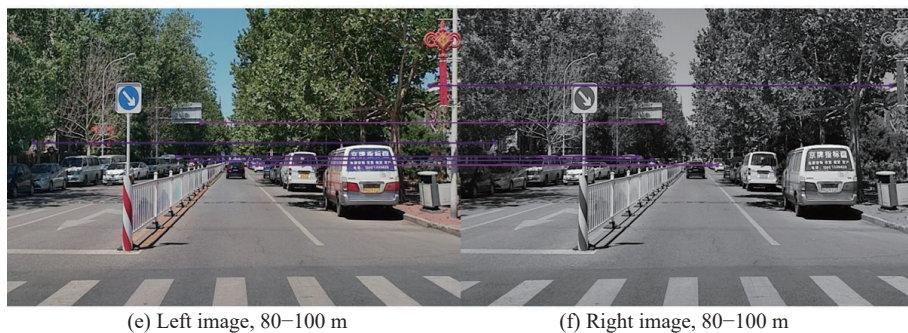


Fig. 9 Distribution of matching points across different distances in Scene 1

Table 3 Calibration error within 60 m of the relative camera

Image	Scene 1	Scene 2	Scene 3	Scene 4	Pixel
Original image	0.161 1	0.127 5	0.130 5	0.142 6	0.140 4
VDSR image	0.160 6	0.127 7	0.126 8	0.135 2	0.137 5
JDDDSN image	0.149 2	0.127 6	0.127 6	0.104 8	0.127 3

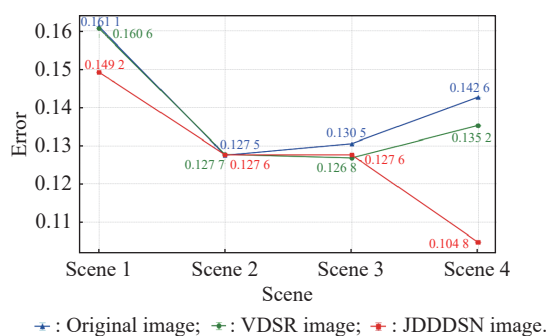


Fig. 10 Comparison of camera calibration error within 60 m

The preliminary findings indicate that when the distance between the camera and the object is limited to 60 m, the calibration accuracy of images reconstructed by the VDSR and JDDDSN super-resolution networks is superior to that of the original images. The overall mean error was reduced from 0.1404 for the original images to 0.1375 and 0.1273 for the VDSR and JDDDSN reconstructions, respectively. Furthermore, the calibration effect of the JDDDSN network reconstruction is significantly better than that of the VDSR reconstruction, with consistently lower errors across all scenes. In particular, in Scene 4, the calibration error of the image reconstructed by VDSR was reduced from 0.1352 to 0.1048. These results demonstrate that the JDDDSN network's

super-resolution reconstruction offers clear advantages in enhancing the calibration accuracy of the stereo camera's extrinsic parameters.

In driving scenarios, accurate depth estimation within the closer range of up to 60 m is crucial for real-time obstacle detection. For longer distances, ranging from 60 m to 100 m, static features such as distant buildings can typically be detected. Super-resolution technology significantly enhances the reconstruction and retention of image details, thereby improving the precision of feature extraction. This, in turn, contributes to a notable increase in the overall accuracy of the calibration process. Therefore, building on our findings within the 60 m range, we further explored the impact of the JDDDSN super-resolution reconstruction algorithm on calibration errors across different distances within a 100 m range. To further explore the impact of the JDDDSN super-resolution reconstruction algorithm on calibration errors across different distance segments within 100 m, we calibrated stereo cameras in four sets of scenes after excluding the influence of highly similar elements such as trees and sky. The calibration was conducted for object regions at intervals of 10 m, resulting in experimental results for 10 distance segments, as summarized in Table 4, with a visual representation of the outcomes provided in Fig. 11.

Table 4 Errors at different distances

Image	Distance/m										Pixel
	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100	
Original image	0.0124	0.1033	0.1170	0.1174	0.1086	0.0719	0.0759	0.3347	0.3333	0.5738	
SR image	0.0113	0.0382	0.1154	0.0630	0.0845	0.0701	0.0650	0.1127	0.2988	0.2038	

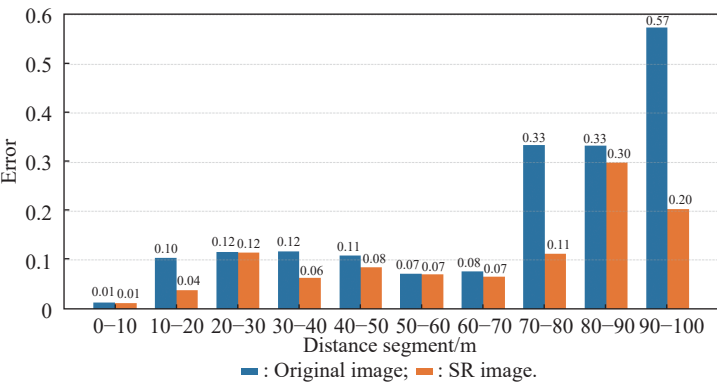


Fig. 11 Comparison of parallel allelic errors at different distance segments

Based on the data analysis presented in Table 4, it is evident that images reconstructed using the JDDDSN super-resolution algorithm exhibit smaller calibration errors compared to the original images across all distance segments within a 100 m range. Notably, the calibration error in the distance range of 90 to 100 m is reduced from 0.5738 in the original image to 0.2038 after applying super-resolution reconstruction. This reduction in parallel isoposition error clearly demonstrates the enhanced detail performance and improved feature recognition accuracy achieved by utilizing JDDDSN super-resolution reconstruction. The data presented in Table 3 and Table 4 provide strong evidence of the significant effectiveness of using images reconstructed with the JDDDSN super-resolution algorithm for feature point extraction, greatly enhancing the accuracy of camera calibration.

3.3 Static object region calibration experiment based on super-resolution reconstruction

In driving scenarios, the rapid movement of dynamic

objects, such as pedestrians and moving vehicles, poses challenges for feature point extraction. Conversely, static objects in the scene, such as fences, traffic signs, utility poles, and buildings, facilitate the extraction of stable feature points. Super-resolution reconstruction provides a more robust foundation for filtering dynamic feature points. Therefore, this section will investigate the impact of super-resolution reconstruction on the calibration of stereo camera extrinsic parameters using feature points derived from static objects.

To screen out dynamic points, the OCRNet semantic segmentation model is initially employed to segment the image into different object categories, generating segmentation results for the four scenes, as shown in Fig. 12. Based on these results, dynamic objects are removed, while masks for four types of static objects—buildings, fences, traffic lights, and traffic signs—are retained. These static object masks are then fused with the original image. Finally, feature points are extracted from the fused image to complete the self-calibration of the stereo camera’s extrinsic parameters.

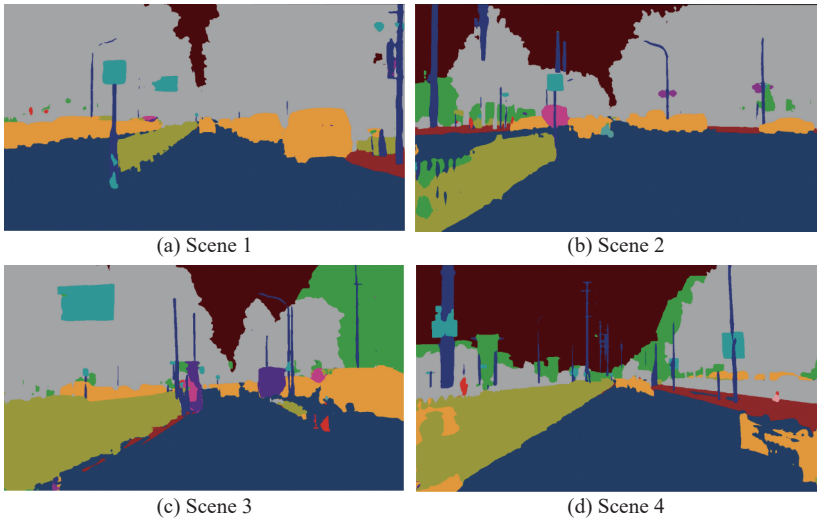


Fig. 12 OCRNet semantic segmentation results for different scenes

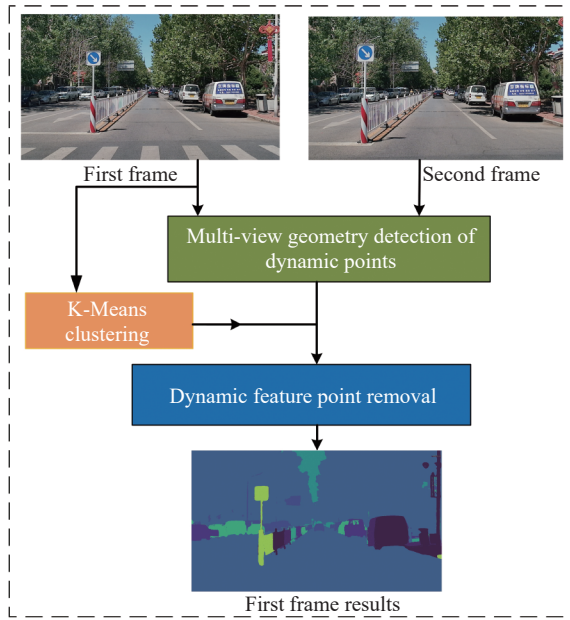
We applied the aforementioned steps to both the original images and the super-resolution reconstructed images in the same scene to complete the stereo camera calibration for static object areas, calculating the parallel isoposition errors for three groups of calibrations. The experimental results are presented in Table 5. The data indicate that calibration accuracy is significantly improved after reconstruction with the JDDDSN network when only static objects are retained. Specifically, the mean calibration error decreases from 0.1237 in the original image to 0.1019 with VDSR network reconstruction, and further reduces to 0.0992 with JDDDSN network

reconstruction. Additionally, the parallel isoposition error of the JDDDSN super-resolution reconstructed image is markedly reduced across different scenarios. For instance, in Scene 4, the error is reduced from 0.1108 in the original image to 0.0427 with JDDDSN. In comparison, the VDSR method achieves an error of 0.0492 in the same scenario, which, although improved, does not match the effectiveness of the JDDDSN network. These results demonstrate that the JDDDSN network outperforms the VDSR method in enhancing the accuracy of feature point detection and stereo camera parameter calibration.

Table 5 Calibration error of static object feature points

					Pixel
Image	Scene 1	Scene 2	Scene 3	Scene 4	Mean
Original image	0.1367	0.1206	0.1138	0.1108	0.1237
VDSR image	0.1273	0.1183	0.1130	0.0492	0.1019
JDDDSN image	0.1238	0.1194	0.1110	0.0427	0.0992

However, to address potential under-segmentation issues within the OCRNet semantic segmentation model, we employed object clustering information and multi-view geometric methods to refine the filtering of dynamic points. This process is depicted in Fig. 13.

**Fig. 13 Dynamic feature point clustering elimination process**

First, we cluster objects based on the spatial relationships within the disparity map of the stereo image to identify different object categories. The disparity map is calculated as follows:

$$d = x_l - x_r \quad (18)$$

where d is the disparity value, x_l and x_r are the abscissa of the corresponding points in the left and right camera images, respectively.

The identified objects are subsequently analyzed using multi-view geometry to detect and discard dynamically changing points. This process involves keypoint matching and triangulation. If we denote $\mathbf{x}, \mathbf{x}'$ as a matching point in the two images, respectively, the triangulation formula used to calculate the corresponding 3D point is as follows:

$$\mathbf{X} = \lambda \mathbf{P}^{-1} \mathbf{x} = \lambda' \mathbf{P}'^{-1} \mathbf{x}' \quad (19)$$

where $\mathbf{X}$ are the 3D points in the world coordinate system, λ and λ' are the scale factors.

These dynamically changing points are often the primary cause of under-segmentation issues in semantic segmentation models. By filtering out these dynamic points, the image can undergo self-calibration of the stereo camera's extrinsic parameters. The calibration errors observed before and after clustering the scene images are presented in Tables 6–9.

Table 6 Calibration errors of static object feature points in the original image

					Pixel
Image	Scene 1	Scene 2	Scene 3	Scene 4	Mean
Before clustering	0.1367	0.1206	0.1138	0.1108	0.1204
After clustering	0.1233	0.1143	0.1160	0.0933	0.1117

Table 7 Calibration errors of static object feature points after VDSR reconstruction images

Image	Scene 1	Scene 2	Scene 3	Scene 4	Pixel
Before clustering	0.127 3	0.118 3	0.113 0	0.049 2	0.101 9
After clustering	0.109 2	0.110 2	0.113 0	0.041 4	0.093 5

Table 8 Calibration errors of static object feature points after JDDDSN reconstruction image

Image	Scene 1	Scene 2	Scene 3	Scene 4	Pixel
Before clustering	0.123 8	0.119 4	0.111 0	0.042 7	0.099 2
After clustering	0.102 6	0.107 2	0.105 9	0.034 1	0.087 4

Table 9 Calibration error after filtering dynamic points by clustering

Image	Scene 1	Scene 2	Scene 3	Scene 4	Pixel
Original image	0.123 3	0.114 3	0.116 0	0.093 3	0.111 7
VDSR image	0.109 2	0.110 2	0.113 0	0.041 4	0.093 4
JDDDSN image	0.102 6	0.107 2	0.105 9	0.034 1	0.087 4

According to the data in Tables 5–7, the calibration errors for the original images, as well as those for the VDSR and JDDDSN super-resolution reconstructed images, show improvement across all four scenes after clustering. This demonstrates that leveraging object clustering information and employing multi-view geometry

methods effectively filter dynamic points within the scene. This step is crucial for enhancing both the adaptability and accuracy of the model in response to dynamic scene changes.

According to the data on dynamic points filtered by clustering in Table 9, the calibration error of images using super-resolution reconstruction technology in different scenes is significantly lower than that of the original images. The mean calibration error for the original images is 0.1117, whereas the calibration errors after VDSR and JDDDSN super-resolution reconstruction are reduced to 0.0934 and 0.0874, respectively. This further confirms that the JDDDSN super-resolution reconstruction network provides more accurate input data for camera calibration, thereby enhancing the effectiveness and practicality of the stereo camera calibration process.

Additionally, visualization of Fig. 14(a) and Fig. 14(b) demonstrates a significant improvement in feature point matching accuracy for the JDDDSN super-resolution reconstructed images. As shown in the figure, the number of feature point mismatches in the JDDDSN-reconstructed images is substantially reduced, leading to more accurate feature point matching. These findings effectively demonstrate the practical impact of super-resolution techniques in improving the accuracy and stability of feature point detection.

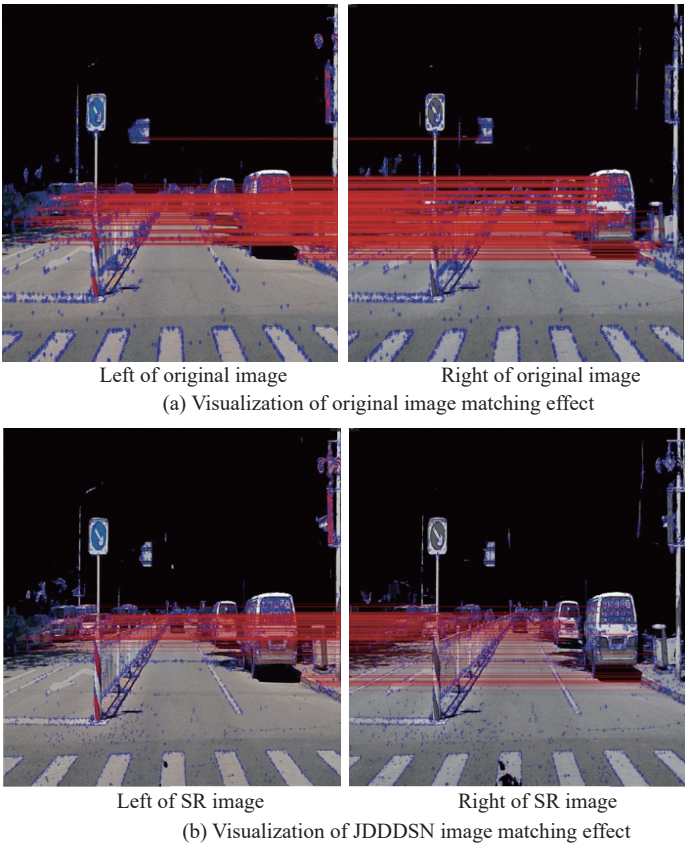


Fig. 14 Comparison of point matching before and after super-resolution in Scene 1

4. Conclusions

This paper proposes a binocular camera self-calibration method based on super-resolution reconstruction, utilizing the JDDDSN super-resolution technique to enhance image resolution while preserving rich image details. Feature points are extracted from the JDDDSN super-resolution reconstructed images for calibrating the extrinsic parameters of the binocular camera. Super-resolution reconstruction image experiment shows that JDDDSN super-resolution reconstruction effectively reduces noise and ringing artifacts in real driving scene images, resulting in clearer, high-resolution images. Distance calibration experiment based on super-resolution reconstruction demonstrates that using JDDDSN super-resolution reconstructed images significantly reduces calibration errors when calibrating object areas at varying distances in road scenes. Static object region calibration experiment based on super-resolution reconstruction indicates that by filtering dynamic regions through semantic segmentation, and removing dynamic points using clustering and geometric methods, the static regions in the binocular camera calibration process are further optimized. With JDDDSN super-resolution reconstruction, the accuracy of feature point detection is enhanced, leading to more precise estimation of binocular camera parameters and reduced calibration errors. This series of experiments validates the effectiveness and practicality of using JDDDSN super-resolution reconstruction to improve calibration accuracy, highlighting its potential value in stereo vision applications.

References

- [1] XIE Q W, LONG Q, MITA S. Integration of optical flow and Multi-Path-Viterbi algorithm for stereo vision. *International Journal of Wavelets, Multiresolution and Information Processing*, 2017, 15: 1750022.
- [2] MARKUS H, THOMAS W, MICHAEL B, et al. Online extrinsic calibration based on per-sensor ego-motion using dual quaternions. *IEEE Robotics and Automation Letters*, 2021, 6(2): 982–989.
- [3] LI G, LIAO X, HUANG H L, et al. Robust stereo visual SLAM for dynamic environments with moving object. *IEEE Access*, 2021: 32310–32320.
- [4] RAHMAT M A, INDRABAYU A, ANDANI A, et al. Stereo camera calibration for autonomous car applications. *Proc. of the 2nd International Seminar on Machine Learning, Optimization, and Data Science*, 2022: 567–572.
- [5] KANG B, MIAO Y, LIU F, et al. A second-order sliding mode controller of quad-rotor UAV based on PID sliding mode surface with unbalanced load. *Journal of Systems Science and Complexity*, 2021, 34(2): 520–536.
- [6] WEI X H, TANG C Y, WANG B, et al. Three-dimensional cooperative target structure design and location algorithm for vision landing. *Systems Engineering-Theory & Practice*, 2019, 39(11): 2975–2983. (in Chinese)
- [7] LI X Q, DENG C X, LI S S, et al. An improved model for wavelet coefficients estimation and its application to image denoising. *International Journal of Wavelets, Multiresolution and Information Processing*, 2024, 22(4): 2450006.
- [8] CAL T Y, DAN B, HUANG W B. Super-resolution parameter estimation of monopulse radar by wide-narrowband joint processing. *Journal of Systems Engineering and Electronics*, 2023, 34(5): 1158–1170.
- [9] WANG X C, SUN X D, SONG R X. Single image dehazing with V-transform and dark channel prior. *Journal of Systems Science and Information*, 2020, 8(2): 185–194.
- [10] SU H, LI Y, XU Y F, et al. A review of deep-learning-based super-resolution: from methods to applications. *Pattern Recognition*, 2025, 157: 110935.
- [11] QIU Z, ZHUANG C, LIU L, et al. Image super-resolution method based on dual learning. *Journal of Circuits, Systems and Computers*, 2022, 31(16): 2250283.
- [12] KAMBLE T U, MAHAJAN S P. 3D image reconstruction using C-dual attention network from multi-view images. *International Journal of Wavelets, Multiresolution and Information Processing*, 2023, 21(2): 2250044.
- [13] YU H F, HE S W, ZHANG M, et al. DIUN: deeper inception U-network for recovering partial pixelated images. *Journal of Systems Science and Information*, 2022, 10(2): 193–202.
- [14] XU X, YE Y F, LI X. Joint demosaicing and super-resolution (JDSR): network design and perceptual optimization. *IEEE Trans. on Computational Imaging*, 2019, 6: 968–980.
- [15] WEN L X, EGAZARIAN K O. End-to-end learning for joint image demosaicing, denoising and super-resolution. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 3506–3515.
- [16] WU R Y, YANG T, SUN L C, et al. SeeSR: towards semantics-aware real-world image super-resolution. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 25456–25467.
- [17] QIAN G C, WANG Y H, GU J J, et al. Rethinking learning-based demosaicing, denoising, and super-resolution pipeline. *Proc. of the IEEE International Conference on Computational Photography*, 2022: 1–12.
- [18] ARULKUMAR V, ARUNA M, ANANTHA M, et al. Super resolution and demosaicing based self learning adaptive dictionary image denoising framework. *Proc. of the 5th International Conference on Intelligent Computing and Control Systems*, 2021: 1891–1897.
- [19] TAO X, GAO H Y, WANG Y, et al. Scale-recurrent network for deep image deblurring. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 8174–8182.
- [20] WANG X T, XIE L B, DONG C, et al. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data. *Proc. of the IEEE/CVF International Conference on Computer Vision Workshops*, 2021: 1905–1914.
- [21] HUI Z, LI J, WANG X M, et al. Learning the Non-differentiable optimization for blind super-resolution. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 2093–2102.
- [22] SUN T, ZOU B, LONG Q, et al. Deep learning algorithm for image clarification based on end-to-end processing. *Mathematics in Practice and Theory*, 2024, 54(8): 154–167. (in Chinese)
- [23] ZHANG Z Y. Camera calibration with one-dimensional

- objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, 26: 892–899.
- [24] ZHANG B, ZHU D L. Improved camera calibration method and accuracy analysis for binocular vision. *International Journal of Pattern Recognition and Artificial Intelligence*, 2021, 35(9): 2155010.
- [25] XIE Q W, LONG Q, LI J P, et al. Application of intelligence binocular vision sensor: mobility solutions for automotive perception system. *IEEE Sensors Journal*, 2024, 24: 5578–5592.
- [26] WEI L, ZHANG G Y, HUO J, et al. Novel camera calibration method based on invariance of collinear points and pole-polar constraint. *Journal of Systems Engineering and Electronics*, 2023, 34(3): 744–753.
- [27] MONRROY C A, LAMBERT J, EDAHIRO M, et al. Single-shot intrinsic calibration for autonomous driving applications. *Sensors*, 2022, 22(5): 2067.
- [28] XIE Q W, HU X Y, REN L, et al. A binocular vision application in IoT: realtime trustworthy road condition detection system in passable area. *IEEE Trans. on Industrial Informatics*, 2023, 19: 973–983.
- [29] HAGEMANN A, KNORR M, STILLE C. Deep geometry-aware camera self-calibration from video. *Proc. of the IEEE/CVF International Conference on Computer Vision*, 2023: 3415–3425.
- [30] GRUEN A, BEYER H A. System calibration through self-calibration. *Calibration and Orientation of Cameras in Computer Vision*, Berlin: Springer Berlin Heidelberg, 2001: 163–193.
- [31] ZHANG Z Y, TANG Q. Camera self-calibration based on multiple view images. *Proc. of Nicograph International*, 2016: 88–91.
- [32] XIE Q W, LIU R R, SUN Z, et al. A flexible free-space detection system based on stereo vision. *Neurocomputing*, 2021, 485: 252–262.
- [33] MICHITAKA A, SAYAKA M, SATORU K. Accuracy improvement of antenna-gain self-calibration method with electronic calibration module. *Proc. of the International Symposium on Antennas and Propagation*, 2018: 1–2.
- [34] DENG Y F, GUO J J, LI H T, et al. A new self-calibration method of time-sharing measurement based on laser tracker. *Proc. of the 7th International Conference on Mechanical and Aerospace Engineering*, 2016: 228–233.
- [35] YU C Q, WANG J B, GAO C X, et al. Context prior for scene segmentation. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 12413–12422.
- [36] HAO F, WU J Z, ZHANG L, et al. Unsupervised change detection of man-made objects using coherent and incoherent features of multi-temporal SAR images. *Journal of Systems Engineering and Electronics*, 2022, 33(4): 896–906.
- [37] CHOI S, JUNG S, YUN H, et al. RobustNet: improving domain generalization in urban-scene segmentation via instance selective whitening. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 11575–11585.
- [38] CHEN L C, ZHU T K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proc. of the European Conference on Computer Vision*, 2018: 833–851.
- [39] KU T, YANG Q R, ZHANG H. Multilevel feature fusion dilated convolutional network for semantic segmentation. *International Journal of Advanced Robotic Systems*, 2021, 18(2): 665.
- [40] DONG J Y, WANG W W, ZHU L. Object-contextual representations for PointNet. *Proc. of the 34th Chinese Control and Decision Conference*, 2022: 6179–6182.
- [41] XIE W F, LIU X P, ZHENG M H. Moving object segmentation and detection for robust RGBD-SLAM in dynamic environments. *IEEE Trans. on Instrumentation and Measurement*, 2021, 70: 5001008.
- [42] EIRIKUR A, RADU T. NTIRE 2017 challenge on single image super-resolution: dataset and study. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017: 1122–1131.
- [43] FANG J, LIU S Q, XIAO Y, et al. SAR image de-noising based on texture strength and weighted nuclear norm minimization. *Journal of Systems Engineering and Electronics*, 2016, 27(4): 807–814.
- [44] HAO B B, ZHU J G. Wavelet inverse scale space for image restoration. *Journal of Systems Engineering and Electronics*, 2012, 23(6): 929–935.
- [45] HUO J, LI Y H, YANG M. Multi-camera calibration method based on minimizing the difference of reprojection error vectors. *Journal of Systems Engineering and Electronics*, 2018, 29(4): 844–853.
- [46] WANG G L, WU L S, HU Y, et al. Point cloud simplification algorithm based on the feature of adaptive curvature entropy. *Measurement Science and Technology*, 2021, 32(6): 065004.
- [47] JUNG B, SUKHATME G S. Real-time motion tracking from a mobile robot. *International Journal of Social Robotics*, 2010, 2: 63–78.
- [48] RATHI K, GOMATHI V, RAJA S P. A novel approach of visual image reconstruction from fMRI using Kohonen network information maximizing generative adversarial network. *International Journal of Wavelets, Multiresolution and Information Processing*, 2024, 22(6): 2450032.
- [49] ZHANG X D, ZENG H Y, GUO S, et al. Efficient long-range attention network for image Super-resolution. *Proc. of the European Conference on Computer Vision*, 2022: 649–667.
- [50] BARTSCH J, SPERLING Y, BERGMANN R. Efficient vision ray calibration of multi-camera systems. *Optics Express*, 2021, 29(11): 17125–17139.
- [51] SUN X Y, JIANG Y Z, JI Y F, et al. Distance measurement system based on binocular stereo vision. *IOP Conference Series: Earth and Environmental Science*, 2015, 252(2): 052051.
- [52] ZHANG Y Y, WANG X Y, JIANG H S, et al. Binocular camera calibration based on BP neural network optimized by improved particle swarm optimization. *Proc. of the 5th International Conference on Pattern Recognition and Artificial Intelligence*, 2022: 450–454.