

Received 14 December 2024, accepted 31 December 2024, date of publication 3 January 2025, date of current version 9 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3525181



RESEARCH ARTICLE

Real-Time Stitching Algorithm of Vehicle Side View Image Based on Multi-Region Fast Phase Correlation

MUBIN XU^{ID}, XIAOYONG LIU^{ID}, AND CHUANHAI WAN^{ID}

School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Xiaoyong Liu (lxy@xjtu.edu.cn)

ABSTRACT Aiming at the problem that the camera field of view is too small to obtain the complete side image of the vehicle in the traffic management scene, a real-time stitching algorithm of vehicle side view image based on Multi-Region Fast Phase Correlation (MFPC) is proposed. Firstly, the background subtraction method based on Gaussian mixture model is used to obtain vehicle foreground image sequence, and the image is downsampled by Gaussian pyramid to reduce the running time of the program. Subsequently, multi-region phase correlation and registration check based on normalized cross-correlation are used to improve the accuracy of registration, and a local inverse discrete Fourier transform method is proposed to improve the computational efficiency. To mitigate background interference in registration, a peak filtering algorithm is proposed, combined with a sub-pixel refinement algorithm to enhance accuracy. Experimental results indicate that the proposed algorithm has better registration performance than the traditional phase correlation and methods based on feature point detection. It requires only 26% of the time taken by the traditional phase correlation, with an average processing time per frame of 3.81 ms, which meets the real-time requirements. In practical applications, the stitching accuracy reached 99.33%, demonstrating high precision and robustness.

INDEX TERMS Image registration, image stitching, phase correlation, road vehicles, vehicle detection.

I. INTRODUCTION

In the intelligent transportation system, vehicle information recognition is a critical technology, which plays an important role in the fields of traffic supervision, expressway toll and truck overload control. In order to obtain vehicle type information, cameras are usually set up around the road to obtain vehicle side images. However, a single camera often has the problem of limited viewing angle and incomplete image coverage. The use of multiple cameras will not only greatly increase the cost, but also cause the image to be unclear when dealing with ultra-long vehicles because the front or rear of the vehicle is too far away from the camera. To solve the above problems, it is necessary to stitch multiple images with partial body information collected by a single camera into a complete side view image of the vehicle.

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar^{ID}.

Image registration is the most critical step in image stitching, which directly affects the final effect of stitching. At present, the mainstream image registration methods can be broadly divided into feature-based method [1], [2], [3] and frequency domain method [4], [5], [6]. The registration method based on spatial features detects and matches key feature points in images for registration. Commonly used feature point detection methods include SIFT [7], SURF [8], and ORB [9]. The SIFT algorithm extracts and describes feature points in scale space, and is robust to scaling, rotation, and illumination changes, but it requires a lot of computation. The SURF algorithm is an improvement of SIFT, which uses the method of integral image and box filter to improve the running speed while maintaining the detection accuracy. The ORB algorithm combines FAST and Harris corner detector to detect feature points, and uses BRIEF descriptors for feature description, which greatly improves the calculation speed, but is less accurate than SIFT and SURF. The registration method

based on spatial features relies on the matching of feature points between images, but there may be a background with rich features such as trees in the images obtained by the camera, resulting in a large number of feature points being detected and matched to the background, which will cause serious interference to vehicle registration. Additionally, the body part of some vehicles is smooth and the feature points are few, which also brings great difficulties to the feature-based registration method.

The image registration method based on frequency domain converts the image to frequency domain and performs registration through frequency domain information. The phase correlation [10] is the most widely used method in frequency domain. It computes image translation parameters by applying discrete Fourier transforms and calculating the normalized cross-power spectrum. Because the phase correlation alone uses phase information, it reduces the dependence on image intensity and content, making it robust to frequency noise, intensity differences, and illumination disturbances [11]. Chen et al. [12] improved the phase correlation by using Fourier-Mellin transform to solve the registration problem with rotation. Tzimiropoulos et al. [13] proposed the normalized gradient correlation, which showed high accuracy and robustness for large-scale transform images.

For the vehicle image stitching scenario, Avizzano et al. [14] proposed a novel stitching algorithm based on Kalman filter and Gaussian mixture model, which is used for the stitching and digital twin construction of train side images in dynamic environments. By introducing Gaussian mixture model to model the input noise, this method can effectively separate a variety of interference sources, such as reflected noise, highlight-induced pseudo-matching, and repetitive texture. In addition, combined with the dynamic prediction function of Kalman filter, the robustness of feature point matching is significantly improved by selecting the most likely real motion data. However, this study focuses primarily on train side data and lacks validation of the algorithm's performance on other vehicle types or in more complex scenarios, such as those involving environmental background information. Zhang et al. [15] proposed a subway train bottom monitoring system based on multi-camera fusion, which uses linear array cameras for image acquisition and employs a 2-D Gamma correction algorithm to address the issue of uneven lighting. The SIFT algorithm and RANSAC algorithm are used for feature matching and mismatching elimination, and finally, image stitching is completed through affine transformation to generate a seamless panoramic image of the train bottom. However, this method has insufficient adaptability to parallax issues in stitched images, resulting in feature matching errors and blurred ghosting problems. El Saer et al. [16] proposed a vehicle appearance detection and stitching framework that combines the deep learning models YOLOv8 and SAM with the image matching algorithm RoMa. A filtering module based on motion mode

is introduced to eliminate background points and mismatched points in high-reflection areas through the dominant parallax vector, thereby improving the accuracy of stitching. Under challenging lighting and reflection conditions, the frame is able to generate seamless panoramic images of vehicle sides. However, the methods mentioned in this paper are computationally expensive, especially when processing high-resolution images, which require a large amount of computing resources, and lack of datasets of actual road vehicles to comprehensively evaluate the effect of the methods, which is difficult to apply in toll road vehicle detection systems with requirements for running time.

For dynamic vehicle side image stitching scenes with complex backgrounds, feature point detection methods are susceptible to interference from complex background features, and it is difficult to apply to vehicles with smooth body surfaces. Moreover, outdoor camera images are often disturbed by various noises, which further complicates feature point detection. Although many image denoising methods have been proposed in recent years [17], [18], [19], [20], the denoising algorithm will bring a certain amount of time consumption, which is unacceptable in applications with high real-time requirements. The phase correlation does not rely on specific image features, performs well in images with few feature points or inconspicuous textures, and is robust to noise and lighting changes, making it more applicable. However, it also has some shortcomings. First of all, if the phase correlation method is directly applied to the whole image, the computational amount is large and cannot meet the real-time requirements. Secondly, the phase correlation requires a large overlapping area between the images to be registered [21], [22], and an excessively large background area will interfere with the phase correlation results. Meanwhile, the stitching of continuous multi-frame images has high requirements for registration accuracy [23], [24].

Based on the above problems, this paper proposes a real-time stitching algorithm of vehicle side view image based on Multi-Region Fast Phase Correlation (MFPC), which can stitch successive frames of images containing partial body information into a complete side view image of vehicle. The algorithm first uses the background subtraction method based on Gaussian mixture model to obtain the vehicle foreground image sequence, and then uses Gaussian pyramid downsampling to reduce the image size to speed up the algorithm. The registration accuracy is improved by multi-region phase correlation and registration check based on normalized cross-correlation. The local inverse discrete Fourier transform effectively improves the computing efficiency. Peak filtering and sub-pixel refinement algorithm reduce the background interference and improve the registration accuracy. Finally, the Sigmoid weighted fusion algorithm is used to make the transition natural and complete the panoramic image stitching.

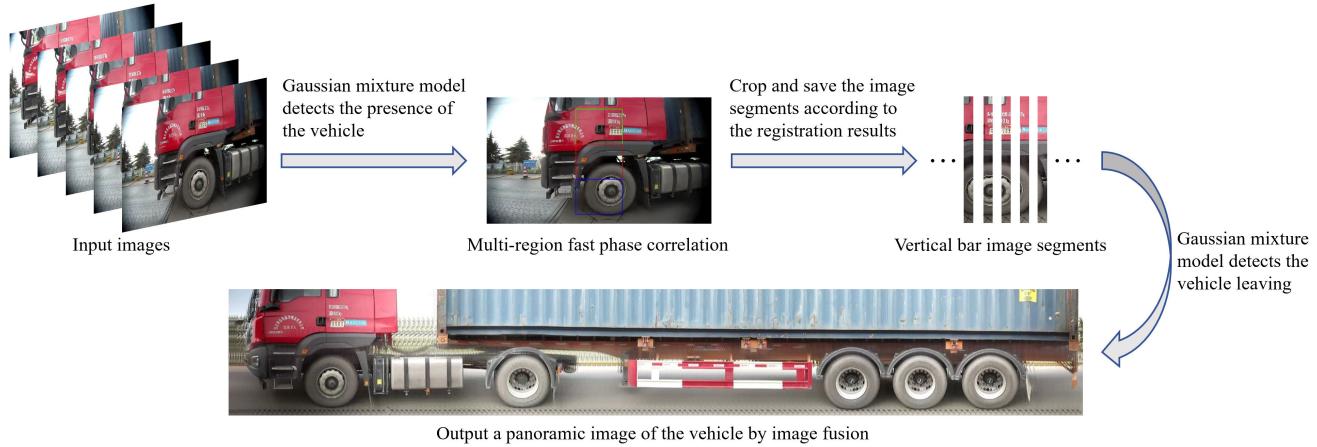


FIGURE 1. The proposed vehicle side view image stitching framework.

The contributions of this paper are as follows:

- We propose a vehicle side view image stitching framework, as shown in Figure 1. The framework uses the background subtraction method based on Gaussian mixture model to detect the entry and exit of vehicles, performs image registration between consecutive frames to obtain the pixel displacement when vehicles are within the detection area, and crops and saves vertical bar image segments based on the pixel displacement. When the vehicle leaves, these image segments are stitched together and fused into a panoramic image of the vehicle.
- We propose an image registration method, named MFPC, which can stitch moving objects in images with complex still backgrounds and has strong adaptability to object size. In this paper, it can stitch various types and sizes of moving vehicles. Additionally, it has a strong tolerance for noise.
- We have made improvements to the operational efficiency and accuracy of the phase correlation. In terms of operational efficiency, we use Gaussian pyramid downsampling to reduce the image size to improve the processing speed, and propose a local inverse discrete Fourier transform to simplify the process of inverse Fourier transform for the scene where there is only horizontal displacement between the images to be registered. In terms of accuracy, we propose peak filtering and sub-pixel algorithms, which reduce the requirement for the overlap percentage between images to be registered, effectively improving registration accuracy.

The rest of this paper is organized as follows. Section II briefly introduces the theoretical basis of the traditional phase correlation, and then introduces our vehicle side image stitching process in detail, including the vehicle foreground image sequence acquisition, the MFPC algorithm and the image fusion. Section III presents performance comparison experiments of various image registration methods and ablation experiments of the proposed MFPC algorithm, along with evaluations of image quality and algorithm robustness

for different stitching methods. Subsequently, the effectiveness of the proposed stitching method is demonstrated through its application at a toll station. Finally, Section IV summarizes this paper.

II. PROPOSED METHOD

A. TRADITIONAL PHASE CORRELATION

Phase correlation is based on the translation property of the two-dimensional discrete Fourier transform: when an image is translated in the spatial domain, its Fourier transform produces a corresponding phase change. When there is a translation relationship between two images, the translation parameters can be determined by calculating their phase difference in the frequency domain. Consider two grayscale images of size $M \times N$, denoted as $f_1(x, y)$ and $f_2(x, y)$, which are related by a pure translation transformation:

$$f_2(x, y) = f_1(x - \Delta x, y - \Delta y), \quad (1)$$

where Δx and Δy are offsets on the x and y axes respectively.

The two-dimensional discrete Fourier transforms of $f_1(x, y)$ and $f_2(x, y)$ are respectively $F_1(u, v)$ and $F_2(u, v)$, by formula (1) their relationship can be obtained as follows:

$$F_2(u, v) = F_1(u, v)e^{-j2\pi(\frac{u\Delta x}{M} + \frac{v\Delta y}{N})}. \quad (2)$$

By calculating the normalized cross-power spectrum of $F_1(u, v)$ and $F_2(u, v)$, expressed as $R(u, v)$, we can obtain:

$$R(u, v) = \frac{F_1(u, v) \cdot F_2^*(u, v)}{|F_1(u, v) \cdot F_2^*(u, v)|} = e^{j2\pi(\frac{u\Delta x}{M} + \frac{v\Delta y}{N})}, \quad (3)$$

where $F_2^*(u, v)$ is the complex conjugate of $F_2(u, v)$. Then the inverse discrete Fourier transform of the normalized cross-power spectrum is performed to obtain the correlation function $r(x, y)$:

$$r(x, y) = \mathcal{F}^{-1}\{R(u, v)\} = \delta(x + \Delta x, y + \Delta y). \quad (4)$$

This results in a Dirac function in two-dimensional space, with the translation parameters Δx and Δy determined by locating its peak.

B. VEHICLE FOREGROUND IMAGE SEQUENCE ACQUISITION BASED ON GAUSSIAN MIXTURE MODEL

The video captured by the camera includes background frames without vehicles and foreground frames with vehicles. To obtain a panoramic side view of each vehicle through stitching, detecting vehicle entry and exit is necessary. The image sequence from entry to exit frames contains the complete side view of the vehicle. In this paper, the background subtraction method based on Gaussian mixture model [25], [26], [27] is used to detect the entry and exit of vehicles, and the vehicle image sequence of each vehicle is obtained. The stitching starts when the vehicle is detected to enter, ends when the vehicle leaves, and then outputs a panoramic side view image of the vehicle. Gaussian mixture model is a background modeling method based on image pixels. For each pixel in a image, the probability distribution $p(f)$ of its gray value $f(x, y)$ can be described by multiple Gaussian components with different means $u_i(x, y)$ and variances $\sigma_i^2(x, y)$:

$$p(f) = \sum_{i=1}^K w_i \frac{1}{\sqrt{2\pi}\sigma_i(x, y)} \exp \left\{ -\frac{(f(x, y) - u_i(x, y))^2}{2\sigma_i^2(x, y)} \right\}, \quad (5)$$

where w_i is the weight of each Gaussian component, which satisfies $\sum_{i=1}^K w_i = 1$.

Each pixel in the current frame is compared to the corresponding Gaussian background model, and according to the results, each pixel belongs to the foreground or the background. If the number of foreground pixels in the current frame is greater than the threshold, the current frame is classified as a foreground frame, and the vehicle is regarded as entering and existing in the detection area. When the vehicle exists in the detection area, the Gaussian background model is not updated until the number of foreground pixels in a certain frame is less than the threshold, and it is determined that the vehicle has left the detection area. With this, we have completed the detection of vehicle entry and exit. Specifically, we set the threshold that determines whether the current frame is a foreground frame to 1/20 of the total number of pixels in the selected region of interest. After practical test, this setting can be sensitive enough to vehicle detection and will not be affected by changes in the lighting environment.

C. PROPOSED MFPC ALGORITHM

1) GAUSSIAN PYRAMID DOWNSAMPLING

For the stitching of vehicle images, it is worth paying attention to ensure the rapidity of image registration. The time complexity of two-dimensional fast Fourier transform for an image with size $M \times N$ is $O(MN\log(MN))$, so the operation time of phase correlation is closely related to the image size, and reducing the image size can speed up the operation. Using the downsampling can effectively accelerate the speed of image registration, but it will lead to the loss of image information. To minimize the impact of image information

loss, Gaussian pyramid [28] is used for downsampling. Gaussian pyramid takes the original image as the bottom layer, and each upper layer carries out a Gaussian convolution operation followed by downsampling. The downsampling is achieved by removing even rows and columns. Gaussian convolution is realized by convolution of the original image with Gaussian kernel, which can preserve the features of the image in the scale space transformation. Gaussian kernel is defined as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp \left(-\frac{x^2 + y^2}{2\sigma^2} \right), \quad (6)$$

where σ is the standard deviation of the Gaussian distribution.

After the Gaussian convolution operation is performed, even rows and even columns of the image are removed to achieve downsampling. With each Gaussian pyramid downsampling, the image size changes to 25% of the previous layer image. When using the image after k times Gaussian pyramid downsampling for phase correlation processing, if the resulting displacement is $(\Delta x, \Delta y)$, then the original image's displacement is $((k+1)\Delta x, (k+1)\Delta y)$. It can be seen that Gaussian pyramid downsampling will reduce the accuracy of the displacement parameters. To balance the speed and accuracy, only one Gaussian pyramid downsampling is carried out in this paper.

2) MULTI-REGION PHASE CORRELATION AND REGISTRATION CHECK BASED ON NORMALIZED CROSS-CORRELATION

The phase correlation is suitable for images in which all pixels have translation, but the vehicle side images captured by the camera usually only contain a small portion of the moving vehicle, with the majority being static background. When the whole side image is used for phase correlation operation, there may be multiple peak points after the inverse Fourier transform of the mutual power spectrum of the two consecutive frames. The position of each peak point corresponds to the translation of different overlapping regions, and the value of the peak point corresponds to the percentage of the overlapping region in the whole image [29]. Since most of the overlapping regions are static background images, the maximum peak point is located at the position where the phase frequency difference is 0, so that the detected translation is 0. If a single Region of Interest (ROI) is selected on the image for detection, the detection result will be improved compared with using the whole image. However, due to different vehicle body heights, different distances from the vehicle to the camera, and vacant areas in the vehicle body, as shown in Figure 2, most of the images in the single ROI will be background, resulting in registration failure of phase correlation.

In order to improve the reliability and accuracy of the phase correlation, this paper selects multiple ROIs on the image to perform phase correlation operations on the two consecutive frames. To enhance the accuracy of displacement detection

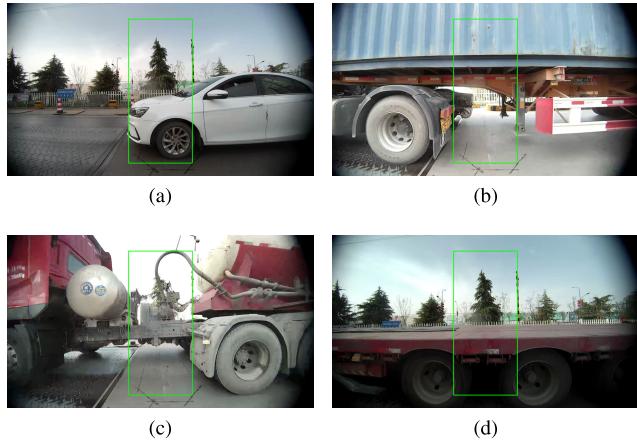


FIGURE 2. Various vehicle body position situations.

without consuming excessive computational resources, three regions are selected for calculation.

Performing phase correlation operations on multiple ROIs separately will obtain multiple translation parameters Δx , from which the most suitable one need to be selected. Normalized Cross-Correlation (NCC) [30] is applied for registration check to determine the optimal translation parameter. The NCC calculation formula is as follows:

$$\text{NCC}(f, g) = \frac{\sum_{x,y} (f(x, y) - \bar{f})(g(x, y) - \bar{g})}{\sqrt{\sum_{x,y} (f(x, y) - \bar{f})^2 \sum_{x,y} (g(x, y) - \bar{g})^2}}, \quad (7)$$

where f is the image to be registered, g is the reference image, \bar{f} is the mean pixel value of the image to be registered, \bar{g} is the mean pixel value of the reference image, x and y represent the pixel coordinates.

To improve the speed of NCC calculation and fairness to each ROI, and reduce the interference of background region to registration check, a rectangular region containing all ROIs in the previous frame is selected as a reference image, and a region after the rectangular region is shifted Δx in the next frame is selected as an image to be registered. According to Δx obtained by phase correlation operation of each ROI, the corresponding NCC is derived, and Δx with the maximum NCC value is selected as the translation parameter on the x axis.

3) LOCAL INVERSE DISCRETE FOURIER TRANSFORM

In the phase correlation operation, after obtaining the correlation function $r(x, y)$ by performing inverse discrete Fourier transform on the normalized cross-power spectrum $R(u, v)$, in order to make the energy more concentrated and easy to analyze, it is necessary to perform fftshift operation on $r(x, y)$. Firstly, divide $r(x, y)$ into four parts: upper left, lower left, upper right, and lower right. Then, perform a diagonal swap operation, which involves swapping the upper left part with the lower right part and the upper right part with the lower left part, resulting in $r_{shift}(x, y)$. At this

point, the displacement amount ($\Delta x, \Delta y$) can be obtained by subtracting the peak coordinate (x^*, y^*) from the center coordinate $(M/2, N/2)$ of $r_{shift}(x, y)$, that is:

$$(\Delta x, \Delta y) = (\frac{M}{2} - x^*, \frac{N}{2} - y^*). \quad (8)$$

In the case where the camera is directly facing the vehicle for shooting, the vehicle's movement in the image is only horizontal, with very small vertical displacement. Therefore, only the horizontal displacement Δx needs to be calculated. Based on this horizontal displacement, the image can be stitched together. The vertical displacement Δy has been tested to be in the range $[-3, 3]$, with the majority of cases being 0. So given $\varepsilon = 3$, find the peak within the region of $y \in [N/2 - \varepsilon, N/2 + \varepsilon]$ in $r_{shift}(x, y)$. This will lead to redundant calculations in the process of inverse discrete Fourier transform for normalized cross-power spectrum $R(u, v)$. To avoid unnecessary calculations, the local inverse discrete Fourier transform is performed in this paper.

The two-dimensional inverse discrete Fourier transform can be decomposed into a one-dimensional inverse discrete Fourier transform in two directions:

$$\begin{aligned} r(x, y) &= \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} R(u, v) e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})} \\ &= \frac{1}{M} \sum_{u=0}^{M-1} \left(\frac{1}{N} \sum_{v=0}^{N-1} R(u, v) e^{j2\pi \frac{vy}{N}} \right) e^{j2\pi \frac{ux}{M}} \\ &= \frac{1}{M} \sum_{u=0}^{M-1} (R_y(u, y)) e^{j2\pi \frac{ux}{M}}, \end{aligned} \quad (9)$$

where:

$$R_y(u, y) = \frac{1}{N} \sum_{v=0}^{N-1} R(u, v) e^{j2\pi \frac{vy}{N}}. \quad (10)$$

Formula (9) and (10) show that $R_y(u, y)$ is obtained by performing one-dimensional inverse Fourier transform on each column of $R(u, v)$, and then $r(x, y)$ is obtained by performing one-dimensional inverse Fourier transform on each row of $R_y(u, y)$.

We aim to find the peak in the region where $y \in [N/2 - \varepsilon, N/2 + \varepsilon]$ in $r_{shift}(x, y)$, so only need to compute the inverse Fourier transform of $R_y(u, y)$ over the range $y \in [0, \varepsilon] \cup [N - \varepsilon, N - 1]$, which is:

$$\begin{aligned} r(x, y) &= \begin{cases} \frac{1}{M} \sum_{u=0}^{M-1} (R_y(u, y)) e^{j2\pi \frac{ux}{M}}, & y \in [0, \varepsilon] \cup [N - \varepsilon, N - 1] \\ 0, & y \in [\varepsilon + 1, N - \varepsilon - 1]. \end{cases} \end{aligned} \quad (11)$$

The time complexity of the two-dimensional fast Fourier inverse transform is $O(MN \log(MN))$, and the time complexity of the local inverse discrete Fourier transform proposed in this paper is $O(MN \log N + M(2\varepsilon + 1) \log M)$, which reduces

$O(M(N-2\varepsilon-1)\log M)$ operations, where ε is generally much less than N , and $\varepsilon = 3$ in this paper.

The result of inverse Fourier transform of $R_y(u, y)$ in the range of $y \in [0, \varepsilon] \cup [N-\varepsilon, N-1]$ constitutes a new image $r^*(x, y)$ with $(2\varepsilon+1)$ rows and M columns. Then the fftshift operation mentioned above is applied to obtain $r_{shift}^*(x, y)$, and subsequent peak search operation will be performed in $r_{shift}^*(x, y)$.

4) PEAK FILTERING AND SUB-PIXEL REFINEMENT

To minimize interference from the background, phase correlation in each area requires peak filtering. When both the background region and the vehicle region exist in two images for phase correlation, there will be a peak located at coordinate $(M/2, \varepsilon)$ in $r_{shift}^*(x, y)$ corresponding to the background area and a peak corresponding to the vehicle's displacement. When the area of the background region is larger than that of the vehicle region in the image, the peak value of the background region may be larger than that of the vehicle region, resulting in incorrect displacement detection.

To avoid this situation, this paper first performs global peak detection on $r_{shift}^*(x, y)$. The final peak coordinate (x^*, y^*) is (x_0, y_0) when the peak coordinate (x_0, y_0) do not locate at $(M/2, \varepsilon)$. If it is located at $(M/2, \varepsilon)$, given $m = 3$, set the function values of $r_{shift}^*(x, y)$ to 0 in an $m \times m$ area centered at $(M/2, \varepsilon)$, and perform global peak detection again. Record the coordinate of the peak found this time as (x_1, y_1) , and the peak value as p . Given threshold $T = 0.1$, if the peak p is greater than T , the final peak coordinate (x^*, y^*) is taken as (x_1, y_1) , otherwise as (x_0, y_0) . The filtering process is shown in Algorithm 1, and the comparison of the results of $r_{shift}^*(x, y)$ before and after filtering is shown in Figure 3.

Algorithm 1 Peak Filtering Algorithm

Input: Image $r_{shift}^*(x, y)$, width M , height parameter ε , window size m , threshold T

Output: Final peak coordinate (x^*, y^*)

- 1: Perform global peak detection on $r_{shift}^*(x, y)$
- 2: Let the detected peak coordinate be (x_0, y_0)
- 3: **if** $(x_0, y_0) = (M/2, \varepsilon)$ **then**
- 4: Set the function values of $r_{shift}^*(x, y)$ to 0 in an $m \times m$ area centered at $(M/2, \varepsilon)$
- 5: Perform global peak detection on the modified $r_{shift}^*(x, y)$
- 6: Let the new peak coordinate be (x_1, y_1) and the peak value be p
- 7: **if** $p > T$ **then**
- 8: $(x^*, y^*) \leftarrow (x_1, y_1)$
- 9: **else**
- 10: $(x^*, y^*) \leftarrow (x_0, y_0)$
- 11: **end if**
- 12: **else**
- 13: $(x^*, y^*) \leftarrow (x_0, y_0)$
- 14: **end if**
- 15: **return** (x^*, y^*)

Since images are quantized into pixels, the peak coordinates produced by the inverse Fourier transform of the normalized cross-power spectrum are restricted to integer values. However, vehicle movement may have sub-pixel displacement, leading to some deviation in detection [31], and the accuracy is further reduced after downsampling. To enhance the accuracy of displacement measurement, sub-pixel refinement of the displacement is required. When the vehicle has sub-pixel displacement, subsidiary peaks will appear in the area adjacent to the peak point, and the amount of displacement can be refined by combining these subsidiary peaks. In this paper, weighted centroid method is used to solve the sub-pixel displacement of image. Sub-pixel accuracy is achieved by calculating the weighted centroid of $r_{shift}^*(x, y)$ within a $(2k+1) \times (2k+1)$ region centered at (x^*, y^*) , where k is taken as 1. Let \hat{x} be the abscissa of the peak point after sub-pixel refinement, and its calculation formula is as follows:

$$\hat{x} = \frac{\sum_{x=x^*-k}^{x^*+k} \sum_{y=y^*-k}^{y^*+k} x \cdot r_{shift}^*(x, y)}{\sum_{x=x^*-k}^{x^*+k} \sum_{y=y^*-k}^{y^*+k} r_{shift}^*(x, y)}. \quad (12)$$

Considering only the displacement in the horizontal direction, the final Δx is:

$$\Delta x = \frac{M}{2} - \hat{x}. \quad (13)$$

D. IMAGE STITCHING AND FUSION

After obtaining the displacement amount Δx for each frame of the vehicle, Δx can be used as the width to crop the vertical bar image for stitching. Since each vertical bar image comes from different frames, the brightness of different frames may be different. If these vertical bar image sequences are directly stitched, there will be obvious gaps at the splicing position. Therefore, image fusion method should be used to smooth the transition at the splicing position. Considering the aspect of real-time, the weighted fusion algorithm is a good choice. Compared with the weighted fusion algorithm with linear weights, the Sigmoid weighted fusion algorithm improves the edge detail retention ability through nonlinear transition [32]. Therefore, the Sigmoid weighted fusion algorithm is used for image fusion in this paper. This algorithm requires that there is an overlapping area between the two images to be stitched, so for each frame of the image, a vertical bar image with a width of $\Delta x + l$ is cropped for stitching, where l is the width of the given overlapping area, which is set to 20 in this paper. Suppose that the two images to be stitched are $f_1(x, y)$ and $f_2(x, y)$ respectively, and $f(x, y)$ is the stitched image, then the Sigmoid weighted fusion algorithm is calculated as follows:

$$f(x, y) = \begin{cases} f_1(x, y), & (x, y) \in f_1 \\ (1-w(x))f_1(x, y) + w(x)f_2(x, y), & (x, y) \in (f_1 \cap f_2) \\ f_2(x, y), & (x, y) \in f_2 \end{cases} \quad (14)$$

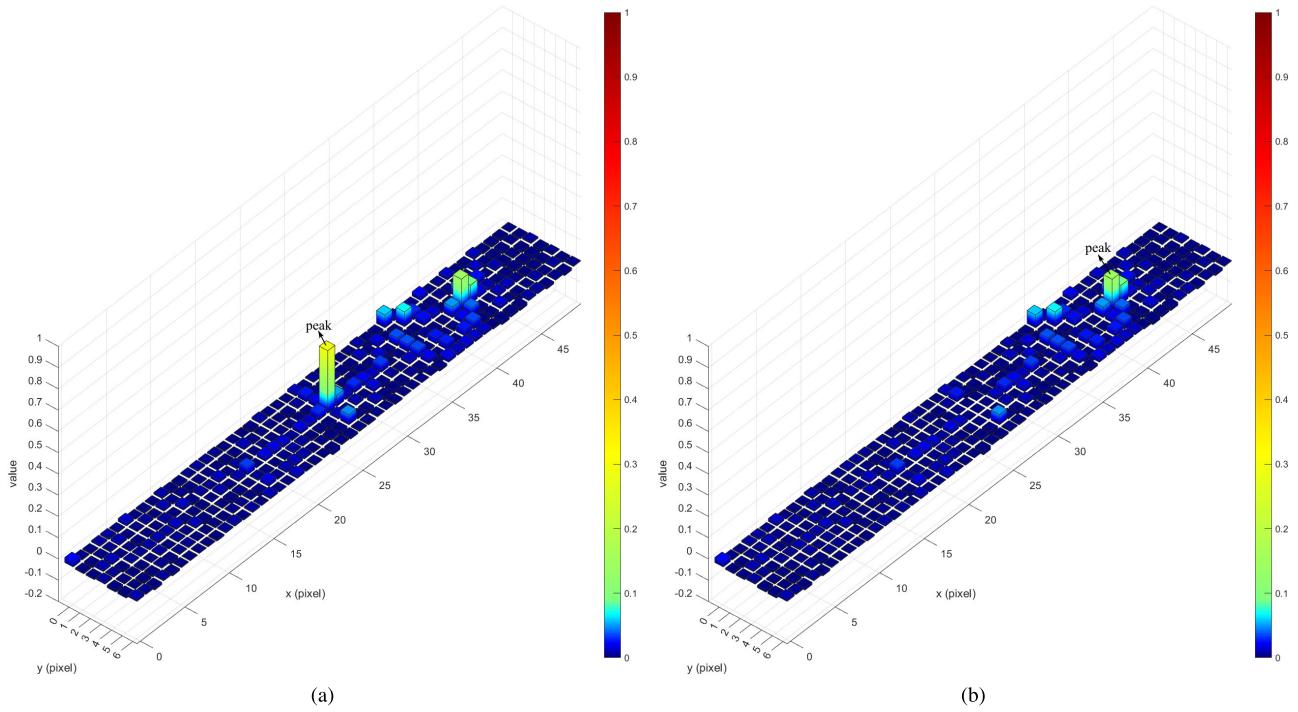


FIGURE 3. $r_{shift}^*(x, y)$ before and after filtering: (a) $r_{shift}^*(x, y)$ before filtering; (b) $r_{shift}^*(x, y)$ after filtering.

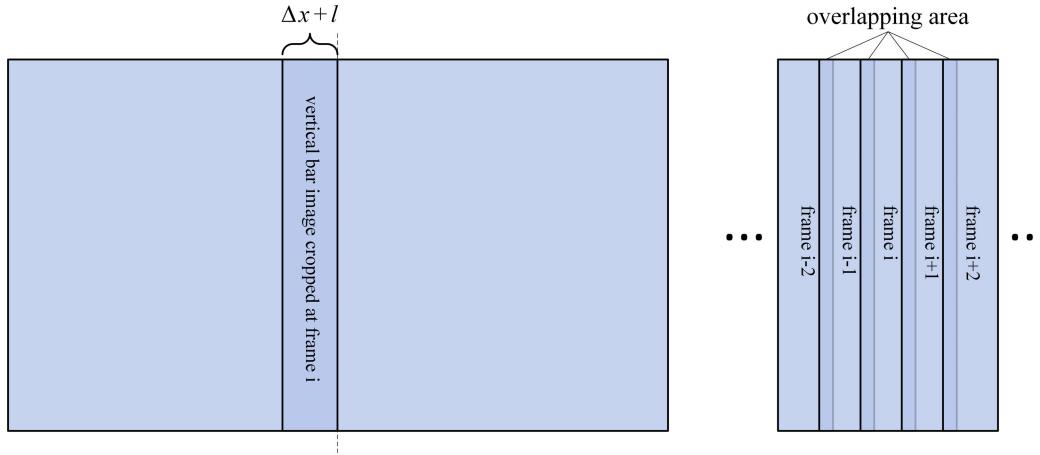


FIGURE 4. Image cropping and fusion process.

where $w(x)$ is the Sigmoid weight function, and its calculation formula is as follows:

$$w(x) = \frac{1}{1 + e^{-\frac{k}{l}(x-c)}}, \quad (15)$$

where, k is the parameter that controls the steepness of the curve, and k is set to 5 in this paper. c is the center of the horizontal coordinate of the overlapping area.

The process of image cropping and fusion is shown in Figure 4.

III. EXPERIMENTS

A. EXPERIMENT EQUIPMENT

In this paper, the camera deployed on the side of the toll station road is used to collect images. The camera model is SONY IMX327, the collected image size is 1920×1080 , and the collection frame rate is 50 fps. The captured images are processed in real time using NVIDIA Jetson AGX Xavier, an edge computing device with an 8-core ARM v8.2 64-bit CPU @ 2.26 GHz and 32 GB LPDDR4 memory. The program is based on Ubuntu 20.04, OpenCV 4.5.4



FIGURE 5. Equipment installation diagram.



FIGURE 6. ROI distribution: (a) Single ROI distribution; (b) Three ROIs distribution.

and C++. The installation diagram of the equipment is shown in Figure 5.

B. COMPARISON OF REGISTRATION PERFORMANCE

To demonstrate the effectiveness of the proposed algorithm, a comparative experiment is conducted, comparing the SIFT, SURF, ORB, Phase Correlation (PC), and the MFPC proposed in this paper. Except for the different registration methods used, the stitching process of each group follows the stitching framework mentioned in this paper. The code implementations of SIFT, SURF, and ORB algorithms all come from OpenCV. All parameters are set to their default values, and the FLANN algorithm [33] is used for feature point matching. The PC algorithm is also implemented based on OpenCV. The experimental data for the comparative experiment is a video containing a moving truck collected on site. The single ROI selected by SIFT, SURF, ORB, and PC is shown in Figure 6(a), and its size is 400×900 . In addition, the background subtraction method based on Gaussian mixture model selects the same ROI region. And the three ROIs selected by MFPC, as shown in Figure 6(b), are obtained by dividing the single ROI in Figure 6(a) into three areas from top to bottom on average.

Record the displacement for each frame obtained from the five methods respectively, as shown in Figure 7, and the stitching results obtained by each method are shown in Figure 8. Record frames with displacement not equal to 0 as registered effective frames. The registration ratio is defined as the ratio of the number of registered effective frames

TABLE 1. Comparison of registration ratio and running time of five methods.

Method	Registration ratio	Time (ms)
SIFT	99.43%	212.38
SURF	100.00%	109.22
ORB	98.86%	25.41
PC	57.14%	14.48
MFPC	100.00%	3.81

to the total number of frames, and the registration ratio of each of the five methods is recorded. Moreover, in order to illustrate the computational efficiency of the proposed algorithm, the running time of five methods is compared, and the average time of each method for single frame processing is recorded. It should be noted that for image registration of consecutive frames, since the results of feature point detection or discrete Fourier transform from the previous frame can be retained to assist in the registration of the next frame, various methods only need to perform feature point detection or discrete Fourier transform on the current frame, thus avoiding unnecessary time consuming. The registration ratio and running time results of the five methods are shown in Table 1.

As can be seen from Figure 7, 8, and Table 1, the registration ratio of the SIFT algorithm is high, but the resulting displacement curve fluctuates greatly and takes too long. The smoothness of displacement curve obtained by the SURF algorithm is improved and the running time is reduced, but it still cannot meet the real-time requirement. The ORB algorithm significantly reduces the processing time, but the displacement curve fluctuates greatly. The above three methods based on feature point detection are not accurate enough in displacement detection, resulting in ghosting and vacancy in stitched images, as shown in the areas outlined by the ellipse in Figure 8(a), (b), and (c). Although the PC algorithm takes less time, the registration ratio is very low, resulting in a large number of vacant areas in the stitched image. The MFPC algorithm proposed in this paper effectively solves the problems existing in PC algorithm, achieving a 100% registration ratio. The resulting displacement curve is notably smoother, and there is basically no ghosting and vacancy in the stitched image. Moreover, although the total area of the three ROIs selected by MFPC algorithm is the same as that of the single ROI selected by the previous four methods, the processing time of single frame is 3.81 ms due to the Gaussian pyramid downsampling and local inverse discrete Fourier transform, which consumes only 26% of the PC algorithm. This remarkably low processing time not only ensures that image processing is completed within the 20 ms interval dictated by the camera frame rate but also makes the algorithm well-suited for deployment on low-speed CPUs, effectively meeting real-time requirements.

C. ABLATION EXPERIMENT

The MFPC algorithm proposed in this paper reduces processing time through Gaussian pyramid downsampling

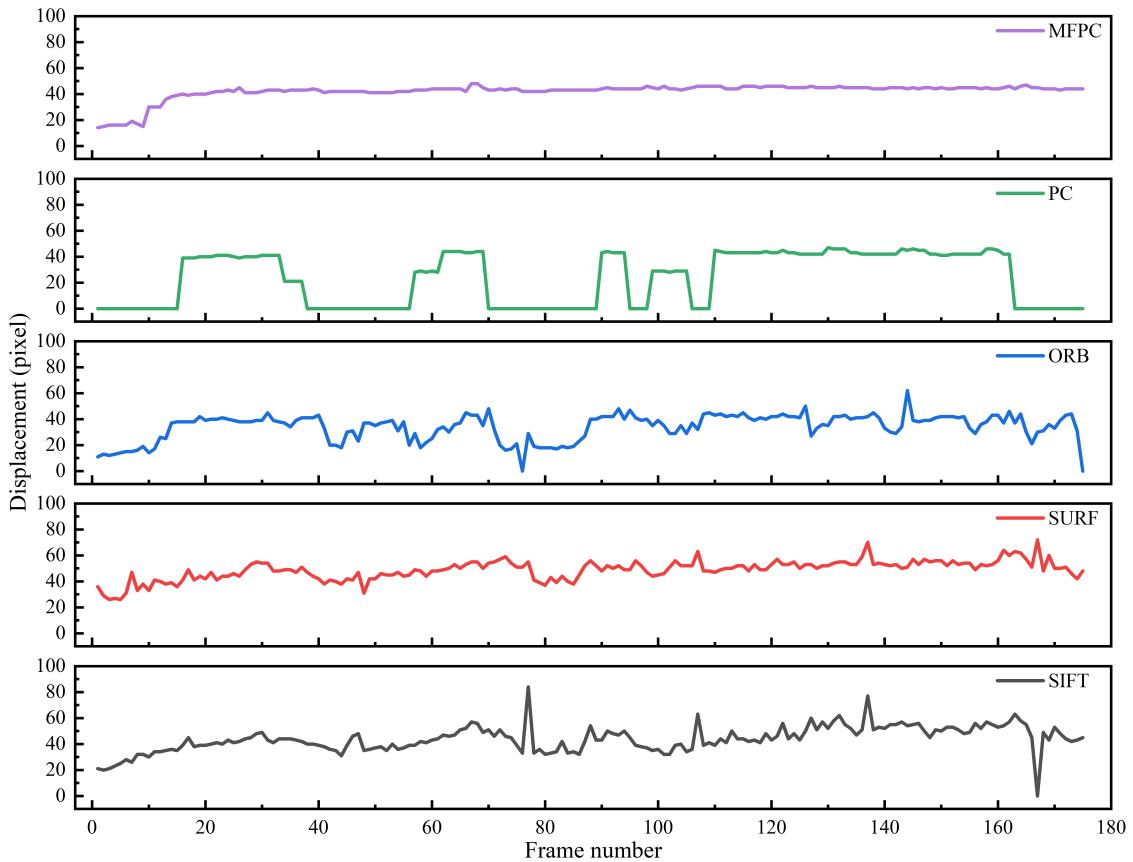


FIGURE 7. Comparison of displacement results obtained by five methods.

TABLE 2. The groups of ablation experiment.

Group	Downsampling ¹	Local IDFT ²	Peak filtering
A	No	No	No
B	Yes	No	No
C	No	Yes	No
D	Yes	Yes	No
E	No	No	Yes
F	Yes	Yes	Yes

¹ Downsampling = Gaussian pyramid downsampling.

² Local IDFT = Local inverse discrete Fourier transform.

and local inverse discrete Fourier transform, and improves registration accuracy through peak filtering. To demonstrate the role these methods play in reducing processing time and improving registration accuracy, different groups are set up for ablation experiments on the MFPC algorithm based on whether these methods are used, as shown in Table 2.

Since it is impossible to accurately know the pixel displacement of each frame of the vehicle captured by the camera, Adobe Photoshop 2023 is used to extract the complete vehicle from the original image. Then place the vehicle image into the background image, simulate the scene of a vehicle driving from right to left and make a video. The speed of the vehicle is set to 30 pixels per frame. According to the above method, three test videos are generated based on

three vehicle images. Figure 9 shows several frames extracted from video 1.

To evaluate the impact of the algorithm on registration accuracy, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used to evaluate the displacement error, with the following formulas:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (17)$$

where y_i is the true displacement, which is set to 30 in this paper, \hat{y}_i is the displacement obtained through registration, and n is the total number of frames.

The displacement results obtained by image registration of each frame in each group are shown in Figure 10. The MAE, RMSE, and average processing time per frame are presented in Table 3, and the stitching results are shown in Figure 11.

From groups A to D in Figure 10(a) and (b), it can be seen that due to the possibility of the background area in the ROI being larger than the vehicle body area during vehicle movement, some frames in the middle of the curve may fail to register. This results in missing areas in the stitched

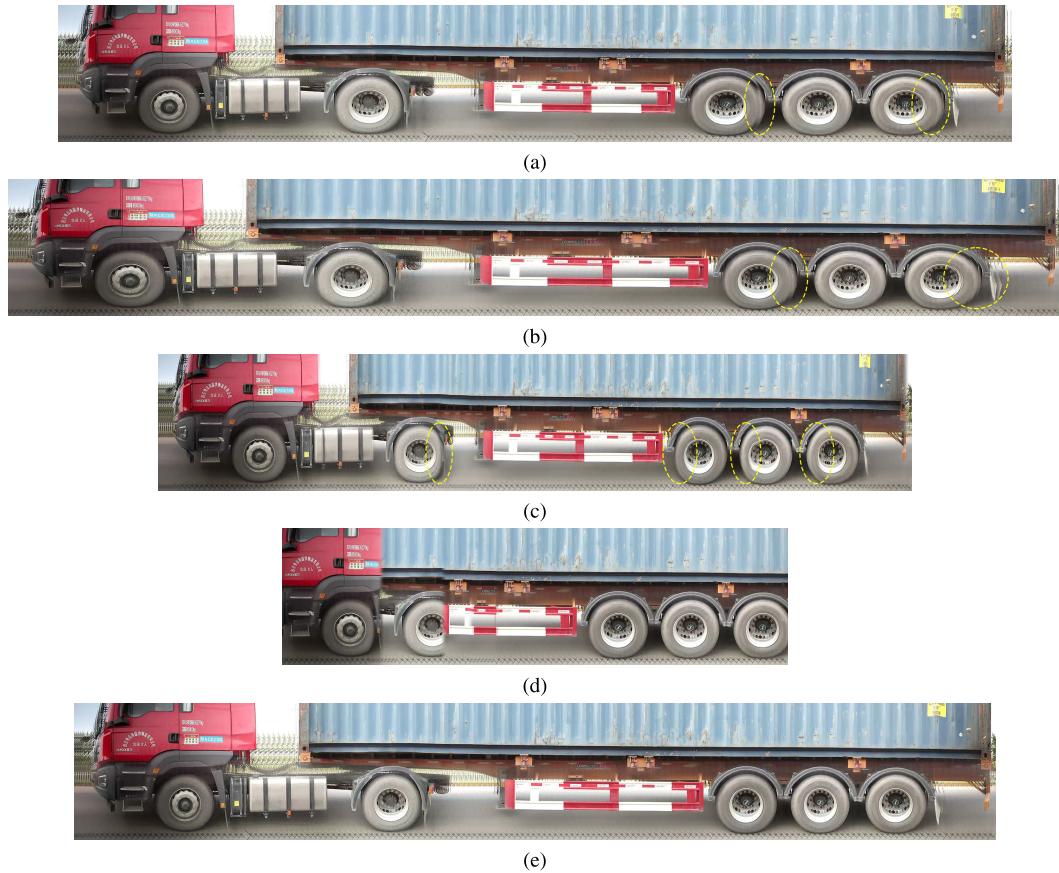


FIGURE 8. Comparison of stitching results obtained by five methods: (a) The result of SIFT; (b) The result of SURF; (c) The result of ORB; (d) The result of PC; (e) The result of MFPC.



FIGURE 9. Several frames extracted from video 1.

TABLE 3. The results of ablation experiment.

Group	MAE				RMSE				Time (ms)			
	Video1	Video2	Video3	Avg	Video1	Video2	Video3	Avg	Video1	Video2	Video3	Avg
A	5.88	3.53	3.46	4.29	13.28	10.29	10.19	11.25	14.30	14.13	14.48	14.30
B	5.88	4.73	3.46	4.69	13.28	11.88	10.19	11.78	4.31	4.40	4.53	4.41
C	5.88	3.53	3.46	4.29	13.28	10.29	10.19	11.25	11.39	11.28	11.39	11.35
D	5.90	4.12	3.48	4.50	13.28	11.11	10.19	11.53	3.68	3.69	3.63	3.67
E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	14.50	14.17	14.35	14.34
F	0.04	0.02	0.02	<u>0.03</u>	0.20	0.14	0.14	<u>0.16</u>	3.75	3.71	3.84	<u>3.77</u>

images of the first and second vehicles from (a) to (d) in Figure 11. The peak filtering algorithm effectively addresses this issue, as evidenced by the successful registration of the middle parts of the curves in groups E and F of Figures 10(a) and (b). Additionally, the peak filtering

algorithm reduces the requirement for the overlap area ratio between images to be registered in the phase correlation, enabling successful registration even when only a small portion of the vehicle's front enters or the majority of its rear exits the ROI. This is demonstrated by displacements

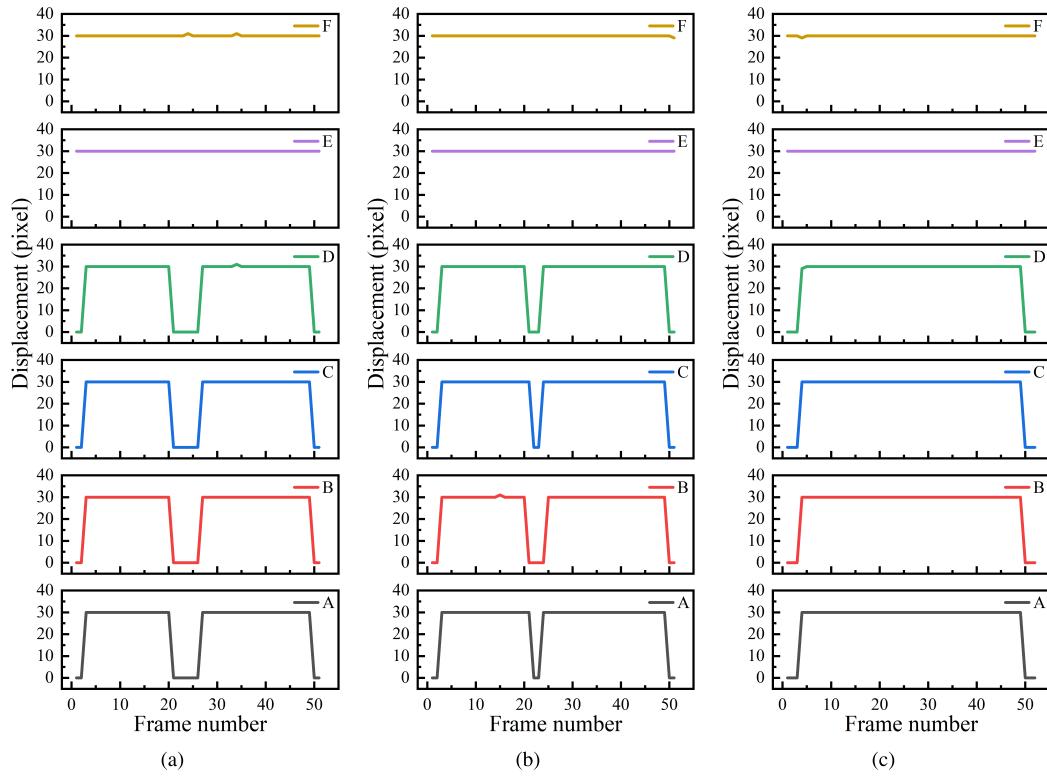


FIGURE 10. The displacement graphs obtained by the registration of each group on three videos: (a) The displacement graphs based on video 1; (b) The displacement graphs based on video 2; (c) The displacement graphs based on video 3.

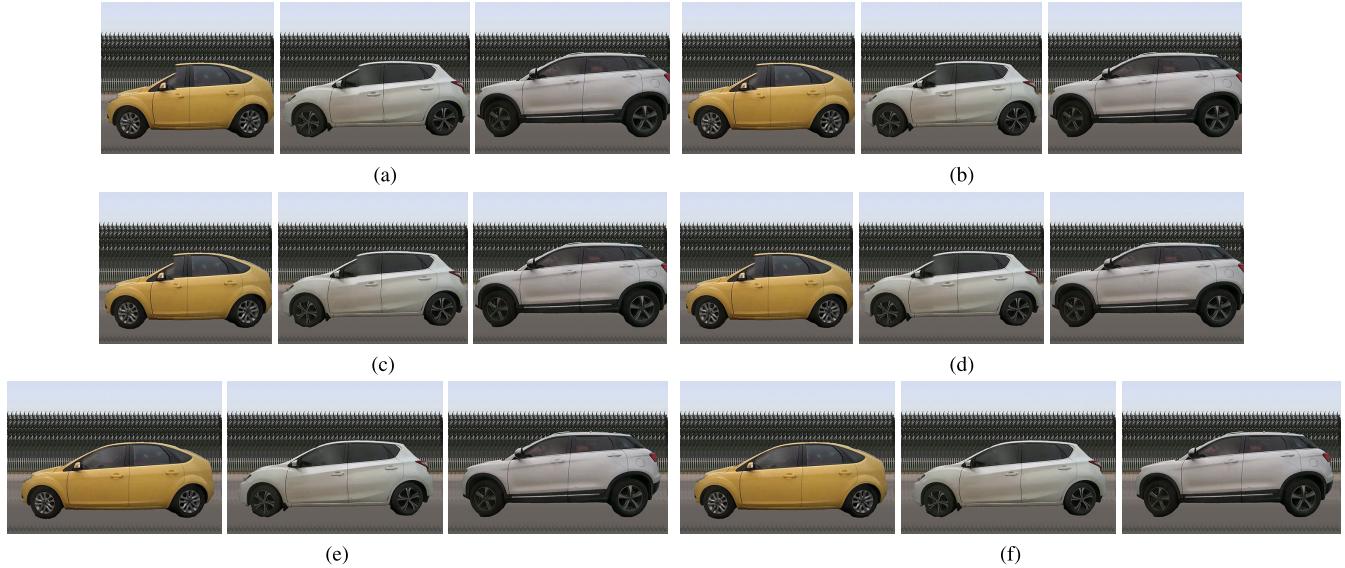


FIGURE 11. The stitching results obtained by each group on three videos: (a) Group A; (b) Group B; (c) Group C; (d) Group D; (e) Group E; (f) Group F.

of 0 detected at the beginning and end of the curves of groups A to D in Figure 10, while groups E and F can register successfully. In Figure 11, this is reflected in the partial absence of the tail of the first and third vehicles in (a) to (d), while the vehicles in groups E and F remain intact, with extra space both in front and back.

As can be seen from Table 3, although the error of group B using Gaussian pyramid downsampling is increased compared to group A, the processing time is significantly reduced to just 30.84% of that of group A. Due to the time cost of the downsampling process, the actual processing time does not reach 25% of the theoretical expectation. By comparing

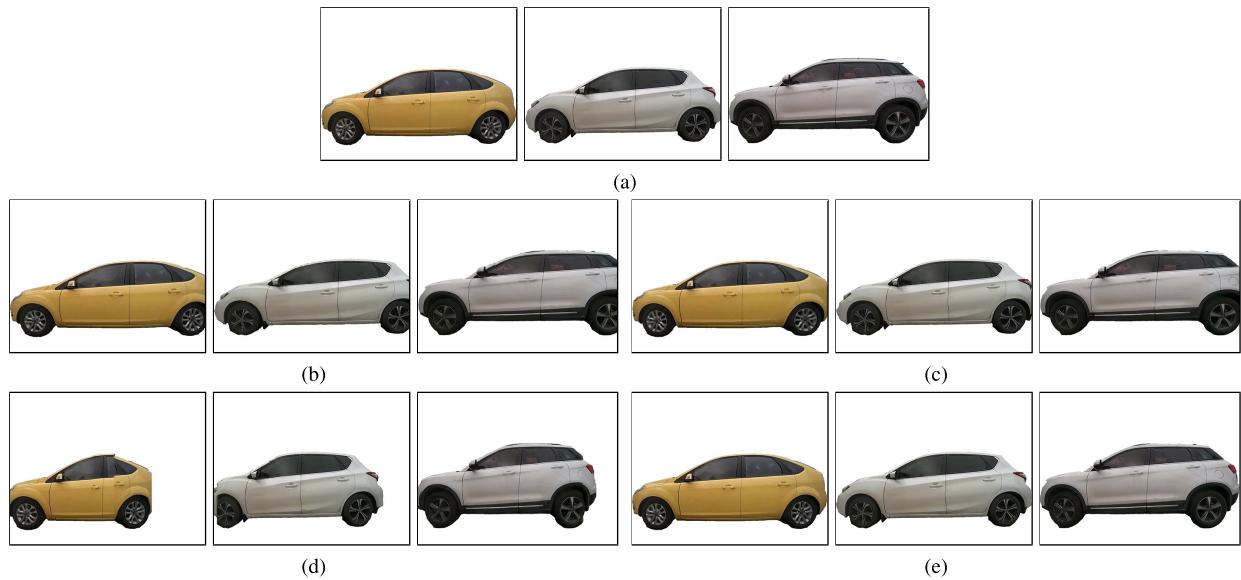


FIGURE 12. The reference images and the cropped stitched images obtained by four methods from three videos: (a) The reference images; (b) The results of SIFT; (c) The results of SURF; (d) The results of ORB; (e) The results of MFPC.

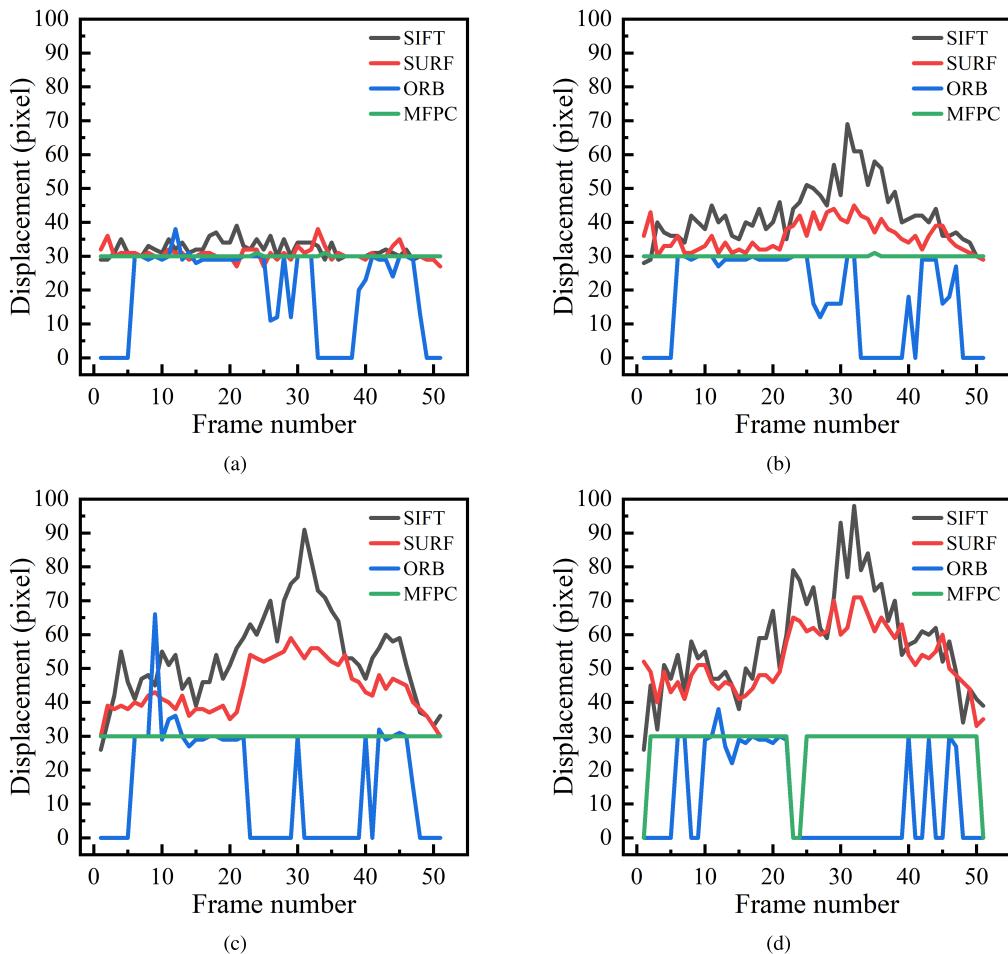


FIGURE 13. The displacement results obtained by four methods of registration under different intensity noise: (a) $\sigma = 0$; (b) $\sigma = 10$; (c) $\sigma = 30$; (d) $\sigma = 50$.

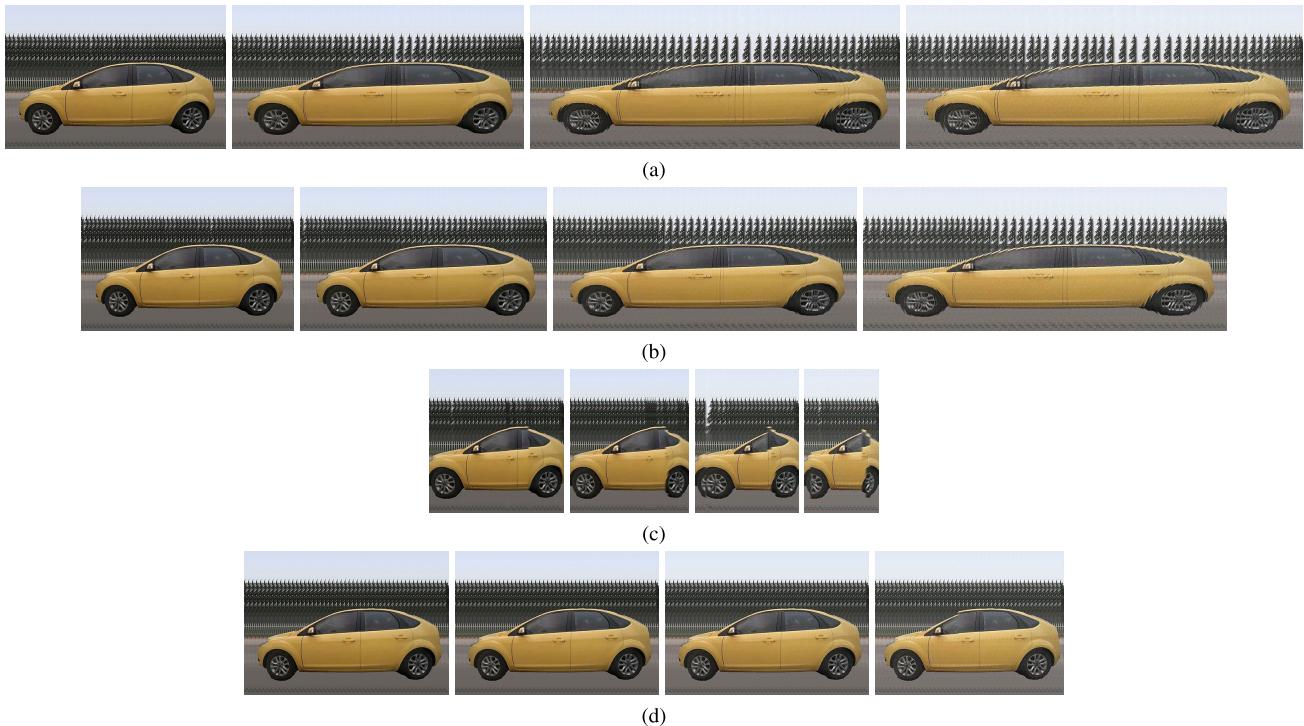


FIGURE 14. The stitching results obtained by four methods under noise of $\sigma = 0, 10, 30, 50$: (a) The results of SIFT; (b) The results of SURF; (c) The results of ORB; (d) The results of MFPC.

groups A and C, it can be observed that the local inverse discrete Fourier transform effectively reduces processing time, bringing the processing time of group C to 79.37% of that of group A, with almost no impact on accuracy. The processing time of group D using both Gaussian pyramid downsampling and local inverse discrete Fourier transform is only 25.66% of that of group A. Comparing the results of groups A and E, it can be seen that the registration accuracy has been significantly improved after using the peak filtering algorithm, and the MAE and RMSE are both 0, showing extremely high accuracy. Compared with group D, the processing time of group F using the complete MFPC algorithm is increased, but it remains at a lower level than that of other groups, with the processing time being 26.36% of group A. Moreover, its accuracy is significantly improved, with the MAE and RMSE both reduced to nearly zero, indicating extremely low error. In terms of speed and accuracy, the complete MFPC algorithm used in group F shows the best performance, with fast processing speed and high registration accuracy, which verifies the advantages of the proposed algorithm.

D. IMAGE QUALITY ASSESSMENT

To evaluate the image quality of the images stitched by various algorithms, the SIFT, SURF, ORB, and MFPC proposed in this paper are used to stitch vehicle images from three videos. It is tested that the traditional phase correlation based on a single ROI has a very low registration rate due to the small proportion of the vehicle body area occupying

the ROI area, making it almost impossible to successfully register each frame, so no record is made. For the images obtained by each algorithm, Peak Signal-to-Noise Ratio (PSNR) [34], Structural Similarity Index (SSIM) [35], and Figure of Merit (FOM) [36], [37] are used to perform image quality evaluation. To avoid the interference of background on the evaluation, vehicle images need to be extracted from images containing background. The reference images are obtained by extracting vehicle images from the original images used to create the videos using Adobe Photoshop 2023. Similarly, the stitched images to be evaluated are extracted using the same method and cropped to match the size of the reference images. The reference images and the cropped stitched images are shown in Figure 12. The quality evaluation program is completed using MATLAB R2023a, and the quality evaluation results are shown in Table 4.

As can be seen from Figure 12, the image stitched by the MFPC algorithm is more complete, and its size is closer to the reference image. As can be seen from Table 4, the three quality evaluation indexes of the images stitched by the MFPC algorithm are all optimal, and compared with the sub-optimal SURF algorithm, the PSNR is increased by 9.37 dB, the SSIM is increased by 0.07, and the FOM is increased by 0.15, showing a strong ability to maintain overall structure and edge details.

E. ALGORITHM ROBUSTNESS TESTING

In practical applications, images are subject to various types of noise in the environment. To demonstrate the robustness

TABLE 4. The quality evaluation results.

Method	PSNR				SSIM				FOM			
	Video1	Video2	Video3	Avg	Video1	Video2	Video3	Avg	Video1	Video2	Video3	Avg
SIFT	16.77	16.33	14.17	15.76	0.88	0.86	0.81	0.85	0.61	0.59	0.57	0.59
SURF	21.28	19.67	15.24	18.73	0.93	0.88	0.83	0.88	0.71	0.65	0.59	0.65
ORB	11.33	11.96	13.90	12.40	0.78	0.80	0.83	0.80	0.43	0.43	0.57	0.48
MFPC	28.57	26.70	29.02	28.10	0.96	0.93	0.96	0.95	0.80	0.76	0.83	0.80

TABLE 5. Displacement error results.

Method	MAE				RMSE			
	$\sigma = 0$	$\sigma = 10$	$\sigma = 30$	$\sigma = 50$	$\sigma = 0$	$\sigma = 10$	$\sigma = 30$	$\sigma = 50$
SIFT	2.31	12.61	24.00	28.08	3.12	15.16	27.21	31.69
SURF	1.39	5.76	14.10	22.80	2.10	7.15	15.98	24.62
ORB	10.55	12.57	16.76	19.45	16.67	18.22	22.14	23.83
MFPC	0.04	0.02	0.00	2.35	0.20	0.14	0.00	8.40

of the proposed algorithm against noise, Gaussian noise distributed as $N(0, \sigma^2)$ is added to video 1, where σ is the standard deviation. In this experiment, σ is successively set to 0, 10, 30, and 50. The robustness testing is carried out using SIFT, SURF, ORB, and MFPC. The traditional phase correlation based on a single ROI is not recorded because the proportion of the vehicle body area to the ROI area is too small, resulting in a 0% registration rate. The displacement results obtained by four methods of registration under different intensity noise are shown in Figure 13. The displacement error is assessed using the MAE and RMSE, and the results are shown in Table 5. The stitching results of each algorithm are shown in Figure 14.

As can be seen from Figure 13 and Figure 14, the SIFT, SURF, and ORB algorithms have relatively low robustness to noise. As the noise increases, the difference between the detected value of the displacement and the true value becomes greater and greater, resulting in severe distortion of the stitched image. In contrast, the proposed MFPC algorithm shows strong robustness against noise, maintaining almost the same registration effect as under noise-free conditions even under Gaussian noise with $\sigma = 30$. With Gaussian noise at $\sigma = 50$, it only results in registration failures in a few frames, having little impact on the stitching result. As can be seen from Table 5, regardless of whether there is noise or not, the MAE and RMSE of the MFPC algorithm are better than those of other algorithms, and show high registration accuracy and robustness.

F. APPLICATION EFFECT

When evaluating the stitching effect for longer vehicles, algorithms that require reference images cannot be used, as no valid reference vehicle image is available for comparison. To ensure the objectivity of evaluation, this paper evaluates the stitching effect by taking the integrity and non-deformation of all wheels in the vehicle images obtained by stitching as the standard, and the non-deformation of

TABLE 6. Stitching result statistics.

Vehicle type	Time of Day	Count	Correct	Accuracy
2-axle car	Daytime	1147	1142	99.56%
	Nighttime	482	477	98.96%
	All Day	1629	1619	99.39%
2-axle truck	Daytime	427	425	99.53%
	Nighttime	201	200	99.50%
	All Day	628	625	99.52%
3-axle truck	Daytime	29	29	100.00%
	Nighttime	26	26	100.00%
	All Day	55	55	100.00%
4-axle truck	Daytime	66	65	98.48%
	Nighttime	50	50	100.00%
	All Day	116	115	99.14%
5-axle truck	Daytime	43	42	97.67%
	Nighttime	15	15	100.00%
	All Day	58	57	98.28%
6-axle truck	Daytime	132	129	97.73%
	Nighttime	57	57	100.00%
	All Day	189	186	98.41%
Total	Daytime	1844	1832	99.35%
	Nighttime	831	825	99.28%
	All Day	2675	2657	99.33%

wheels refers to the aspect ratio of wheels between 0.8 and 1.2. In this paper, the YOLO target detection algorithm [38] is employed to detect the vehicle's wheels. When all the wheels can be identified and the wheels are not deformed, the stitching is considered correct, otherwise the stitching is considered wrong. The training and testing sets for the YOLO target detection algorithm are constructed using a dataset of panoramic vehicle images obtained from the proposed stitching framework. The dataset includes 8000 images for training and 2000 images for testing, ensuring a diverse representation of vehicle types and conditions. These sets are annotated for accurate wheel detection. This process is solely used to evaluate the stitching accuracy and is not the main focus of this study, so the detailed configuration and optimization of the YOLO model is not further explained.

Apply the algorithm proposed in this paper to the toll station for real-time vehicle stitching. Under the condition of turning on the additional lighting at night, it can achieve continuous operation for 24 hours. The stitching results on May 27, 2024 are statistically analyzed. The period from 06:00 to 19:00 is considered as daytime, while the remaining period is considered as nighttime. Vehicles are classified according to the number of axles and whether they carry cargo. The statistical results are shown in Table 6.

Table 6 shows that the total stitching accuracy of the algorithm proposed in this paper reached 99.33%. Moreover, under the assistance of night lighting, the accuracy at nighttime is comparable to that during the daytime. There is no significant difference in accuracy across various vehicle types, indicating the robustness and accuracy of the algorithm presented in this paper.

IV. CONCLUSION

In this paper, a real-time stitching algorithm of vehicle side view image based on multi-region fast phase correlation is proposed, which can stitch multiple consecutive frames containing partial information of vehicle into a complete panoramic image. After obtaining the vehicle foreground image sequence using the background subtraction method based on Gaussian mixture model, the algorithm reduces the image size by Gaussian pyramid downsampling to speed up the operation. In order to solve the problem that the traditional phase correlation is prone to registration failure, phase correlation is carried out in multiple regions at the same time, and the normalized cross-correlation is used to verify the results obtained in each region to obtain the optimal result. For the redundant operations that exist in the inverse discrete Fourier transform process, the local inverse discrete Fourier transform is proposed to simplify the calculation. To avoid interference from the background, the peak corresponding to the vehicle's motion is selected through the peak filtering algorithm. The sub-pixel refinement algorithm applied to the displacement effectively improves the registration accuracy. Finally, the Sigmoid weighted fusion algorithm is used to smooth the transition at the image splicing to obtain a panoramic image of the vehicle.

The experimental results show that the time consumption of the proposed MFPC algorithm is only 26% of the traditional phase correlation, with an average processing time per frame of 3.81 ms, meeting the real-time requirement. Compared with traditional phase correlation and feature point detection methods, it has higher registration accuracy and tolerance to noise. In the application scenario, the stitching accuracy of the proposed algorithm reached 99.33% and was roughly the same during the day and night, demonstrating high accuracy and robustness.

REFERENCES

- [1] Z. Wang, Z. Tang, J. Huang, and J. Li, "A real-time correction and stitching algorithm for underwater fisheye images," *Signal, Image Video Process.*, vol. 16, no. 7, pp. 1783–1791, Oct. 2022.
- [2] Y.-B. Liu, J.-H. Qin, M.-Y. Zhu, and T.-T. Huang, "Fast stitching method for multi-view images of cupping spots," *Signal, Image Video Process.*, vol. 17, no. 5, pp. 1905–1913, Jul. 2023.
- [3] Z. Zhang, L. Wang, W. Zheng, L. Yin, R. Hu, and B. Yang, "Endoscope image mosaic based on pyramid ORB," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103261.
- [4] Y. Han, J. Choi, J. Jung, A. Chang, S. Oh, and J. Yeom, "Automated coregistration of multisensor orthophotos generated from unmanned aerial vehicle platforms," *J. Sensors*, vol. 2019, pp. 1–10, Apr. 2019.
- [5] Z. Hrazdíra, M. Druckmüller, and S. Habbal, "Iterative phase correlation algorithm for high-precision subpixel image registration," *Astrophysical J. Suppl. Ser.*, vol. 247, no. 1, p. 8, Mar. 2020.
- [6] Y. S. Park, Y.-S. Shin, and A. Kim, "PhaRaO: Direct radar odometry using phase correlation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 2617–2623.
- [7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [8] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.
- [9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.
- [10] C. D. Kuglin, "The phase correlation image alignment method," in *Proc. IEEE Int. Conf. Cybern. Soc.*, Jan. 1975, pp. 163–165.
- [11] X. Tong, Z. Ye, Y. Xu, S. Gao, H. Xie, Q. Du, S. Liu, X. Xu, S. Liu, K. Luan, and U. Still, "Image registration with Fourier-based image correlation: A comprehensive review of developments and applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 4062–4081, Oct. 2019.
- [12] Q.-S. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 12, pp. 1156–1168, May 1994.
- [13] G. Tzimiropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki, "Robust FFT-based scale-invariant image registration with image gradients," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1899–1906, Oct. 2010.
- [14] C. A. Avizzano, G. Scivoletto, and P. Tripicchio, "Robust image stitching and reconstruction of rolling stocks using a novel Kalman filter with a multiple-hypothesis measurement model," *IEEE Access*, vol. 9, pp. 154011–154021, 2021.
- [15] Z. Zhang, J. Zhang, and Y. Chen, "An online metro train bottom monitoring system based on multicamera fusion," *IEEE Sensors J.*, vol. 24, no. 17, pp. 27687–27698, Sep. 2024.
- [16] A. El Saer, L. Grammatikopoulos, G. Sfikas, G. Karras, and E. Petsa, "A novel framework for image matching and stitching for moving car inspection under illumination challenges," *Sensors*, vol. 24, no. 4, p. 1083, Feb. 2024.
- [17] A. Khmag, S. A. R. Al Haddad, R. A. Ramlee, N. Kamarudin, and F. L. Malallah, "Natural image noise removal using nonlocal means and hidden Markov models in transform domain," *Vis. Comput.*, vol. 34, no. 12, pp. 1661–1675, Dec. 2018.
- [18] A. Khmag, "Additive Gaussian noise removal based on generative adversarial network model and semi-soft thresholding approach," *Multimedia Tools Appl.*, vol. 82, no. 5, pp. 7757–7777, Feb. 2023.
- [19] J. Yang, M. Ma, J. Zhang, and C. Wang, "Noise removal using an adaptive Euler's elastica-based model," *Vis. Comput.*, vol. 39, no. 11, pp. 5485–5496, Nov. 2023.
- [20] A. Khmag, "Natural digital image mixed noise removal using regularization Perona–Malik model and pulse coupled neural networks," *Soft Comput.*, vol. 27, no. 21, pp. 15523–15532, Nov. 2023.
- [21] S. Samsudin, S. Adwan, H. Arof, N. Mokhtar, and F. Ibrahim, "Development of automated image stitching system for radiographic images," *J. Digit. Imag.*, vol. 26, no. 2, pp. 361–370, Apr. 2013.
- [22] N. T. Pham, S. Park, and C.-S. Park, "Fast and efficient method for large-scale aerial image stitching," *IEEE Access*, vol. 9, pp. 127852–127865, 2021.
- [23] X. Wei, W. Yan, Q. Zheng, M. Gu, K. Su, G. Yue, and Y. Liu, "Image redundancy filtering for panorama stitching," *IEEE Access*, vol. 8, pp. 209113–209126, 2020.

- [24] Y. Liu, M. He, Y. Wang, Y. Sun, and X. Gao, "Farmland aerial images fast-stitching method and application based on improved SIFT algorithm," *IEEE Access*, vol. 10, pp. 95411–95424, 2022.
- [25] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Apr. 1999, pp. 246–252.
- [26] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, pp. 28–31.
- [27] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006.
- [28] T. Li, J. Wang, and K. Yao, "Subpixel image registration algorithm based on pyramid phase correlation and upsampling," *Signal, Image Video Process.*, vol. 16, no. 7, pp. 1973–1979, Oct. 2022.
- [29] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–200, Mar. 2002.
- [30] V. S. Bhadouria, Y.-R. Park, and J. B. Eom, "A sliding template-based approach for position-based whole slide microscopy image stitching for slightly unaligned stages," *Signal, Image Video Process.*, vol. 18, no. 3, pp. 2597–2605, Apr. 2024.
- [31] W. S. Hoge, "A subspace identification extension to the phase correlation method," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 277–280, Feb. 2003.
- [32] C. Xiu, J. Fang, and J. Zhang, "Image stitching method based on adaptive weighted fusion," in *Proc. 33rd Chin. Control Decis. Conf. (CCDC)*, May 2021, pp. 3099–3103.
- [33] M. Muja and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," *VISAPP*, vol. 2, nos. 331–340, p. 2, Jan. 2009.
- [34] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
- [35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [36] N. Tariq, R. A. Hamzah, T. F. Ng, S. L. Wang, and H. Ibrahim, "Quality assessment methods to evaluate the performance of edge detection algorithms for digital image: A systematic literature review," *IEEE Access*, vol. 9, pp. 87763–87776, 2021.
- [37] Y. Meng and J. Zhang, "A novel gray image denoising method using convolutional neural network," *IEEE Access*, vol. 10, pp. 49657–49676, 2022.
- [38] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl. Extraction*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023.



MUBIN XU received the B.S. degree from the School of Electrical Engineering and Automation, Wuhan University, in 2022. He is currently pursuing the master's degree with the School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China. His research interests include image stitching, object detection, and deep learning.



XIAOYONG LIU is currently an Associate Professor with the School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China. His research interests include computer vision and pattern recognition.



CHUANHAI WAN received the B.S. degree from the School of Electrical Engineering and Automation, Wuhan University, in 2022. He is currently pursuing the master's degree with the School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, China. His research interests include computer vision and sensor software driver programming.

• • •