Composite Vision Method for Measurement of 3D Bead Profile, Part 1 — Principles

A novel as-welded bead three-dimensional profile measurement method based on passive and active vision coordination for multi-layer and multi-pass welding

BY S. WU, C. LI, F. CHENG, Z. WANG, AND J. JIAO

Abstract

Identifying the as-welded bead 3D profile during multi-layer and multi-pass (MLMP) welding is challenging since the as-welded bead is easily disturbed and similar to the surrounding weld beads. However, the profile of the as-welded bead determines whether the preset path planning and the welding parameters need to be adjusted to avoid incomplete fusion. In MLMP welding, passive vision sensors can distinguish weld beads but struggle to extract height information, while active vision sensors excel at obtaining height details. Thus, in this study, an innovative composite vision method integrating a passive vision sensor and an active vision sensor was proposed for the as-welded bead profile measurement in MLMP welding. A deep-learning model based on U-Net was used to extract the weld bead features in the passive vision image, and coordinate transformation was used to correspond the coordinates of passive vision images to those of active vision images. The experiment's results showed that the method can precisely measure the profile dimensions of the as-welded bead during MLMP welding, with an error in the width direction of less than 5% and an error in the height direction of less than 7%, meeting the actual MLMP welding requirements.

Keywords

- Multi-Layer and Multi-Pass Welding
- Composite Vision Method
- Weld Bead Profile
- Image Processing
- Deep Learning

Introduction

Multi-layer and multi-pass (MLMP) welding, which is usually used to weld medium-thick plates, is an essential process for manufacturing industries, especially shipbuilding, pipeline construction, and heavy machinery (Refs. 1–3). In most situations, MLMP welding is mainly manual welding or semi-automated welding, in which the operators are responsible for planning the position of the torch and welding parameters according to in-situ bead performance and workpiece distortion, while the mechanical equipment is responsible for executing movement and welding instructions. Therefore, skilled welders still play a significant role in MLMP welding.

However, fully automated MLMP welding has been initially realized in some special situations, such as fillet welding without full penetration requirements and MLMP welding below three layers (Refs. 4, 5). The above situations can be automated mainly due to the small demand for real-time planning and the simply achieved high-precision assembly. However, for large structures, such as ship decks and offshore platform jackets, the distortion during welding is serious and cannot be ignored, and high-precision assembly becomes much harder. Therefore, offline programming planning will not be suitable. In these cases, robotic MLMP welding is mostly used in teaching and playback (Ref. 6).

https://doi.org/10.29391/2025.104.026

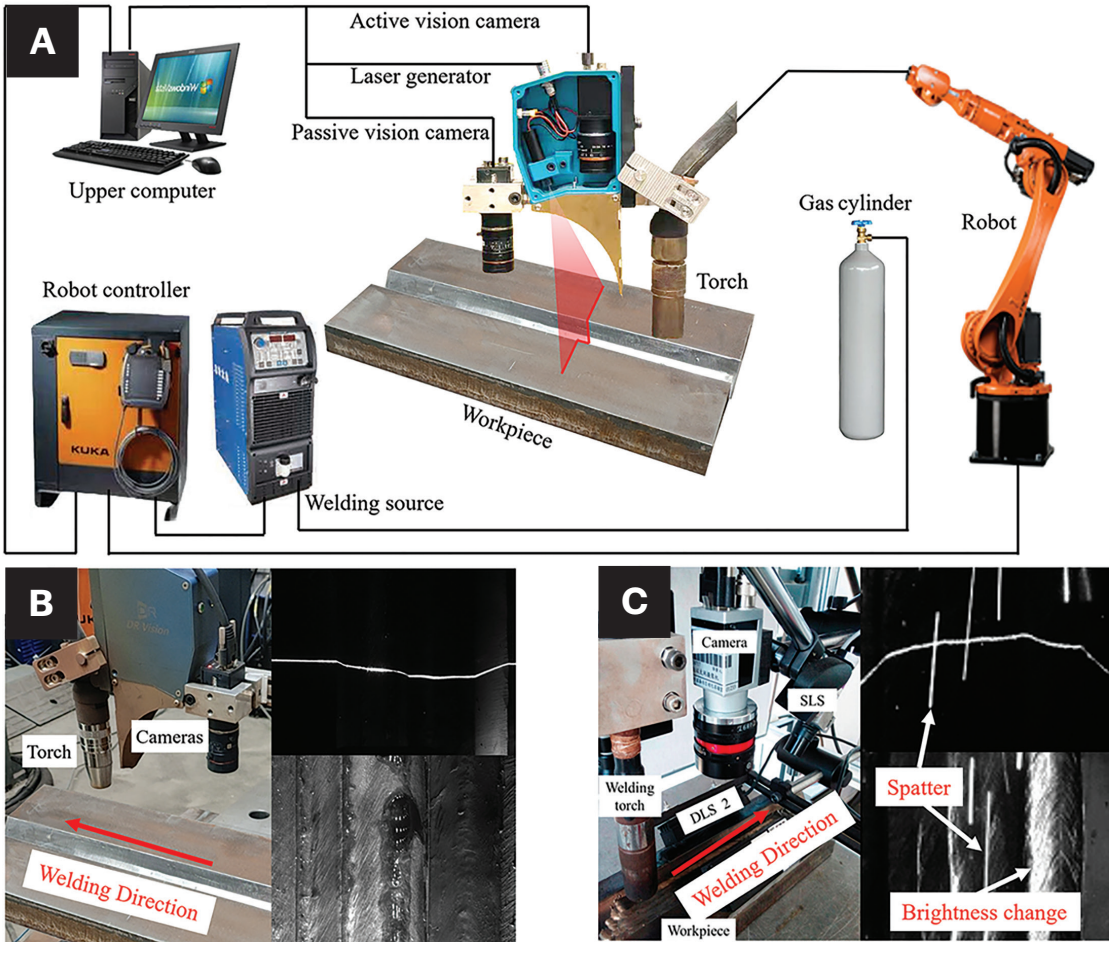

Fig. 1 — Experiment platform setup and vision camera positions: A — Experiment platform setup; B — vision camera positions and captured images in this study; C — vision camera position and captured images (Ref. 19).

Table 1 — Result of Laser Strip Calibration

Coefficient	Image 1	Image 2	Image 3	Average Value
k	1044.2	1038.5	1044.8	1042.5
j	0.0537	0.0536	0.0535	0.0536

By studying the operational behavior of skilled welders, it can be found that they will adjust the position of the torch and the welding parameters according to their experience after observing the as-welded bead profile. The relationship between the as-welded bead profile and the subsequent weld parameters built intuitively by skilled welders is the key to perfect MLMP welding beads and the key reason recent robots cannot be fully automated. Hence, accurately measuring the profile dimensions of the as-welded bead in MLMP welding is the most important problem to be solved.

Integrating sensors with robots has become a highly viable approach to achieving precise weld bead profile measurements. Common sensing methods in welding include arc sensing (Ref. 7), acoustic sensing (Ref. 8), thermal sensing (Ref. 9), and vision sensing (Refs. 10, 11). Vision information represents over 80% of the data gathered during welding (Ref. 12). Therefore, vision sensing is the most advantageous method for the observation of the weld bead profile. Vision sensing is divided into passive and active types. For instance, Ge (Ref. 13) and Xu (Ref. 14) have utilized diagonal differential operator processing and improved Canny algorithms to detect the edges of weld beads and weld pools. However, passive vision has shown significant shortcomings in practical applications, as it struggles to obtain height information and is highly susceptible to environmental interference during welding. Given the limitations of passive vision, active vision sensors have emerged. Active vision sensors, aided by auxiliary light sources, such as lasers, are widely used in various welding applications (Ref. 15). Laser line features are

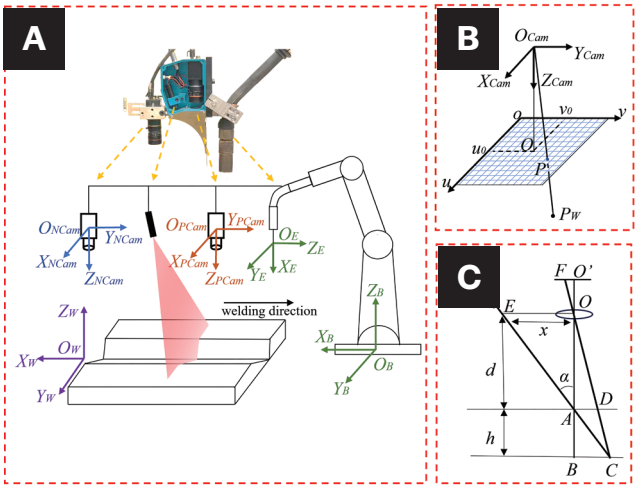

Fig. 2 — Schematic diagram of the coordinate systems in the experiment platform: A — The coordinate systems of the experiment platform; B — projection model of the camera; C — triangular projection model of the laser line.

extracted and processed using various algorithms. Zhang (Ref. 16) proposed the adaptive dynamic thinning algorithm to determine the geometric parameters of the molten pool depression. Zou (Ref. 17) successfully minimized noise interference and achieved high tracking accuracy by combining morphological image processing with a continuous convolution operator-tracking algorithm. Jin (Ref. 18) extracted weld bead width and height dimensions from active vision images through edge detection and smoothing processing. In summary, advanced image processing can obtain more-accurate dimensional results. However, these methods are limited if the edge features are unclear, such as the partly covered adjacent weld beads in MLMP welding.

A composite vision method was recently proposed to solve the MLMP weld bead recognition problem. Zeng (Ref. 19) proposed a recognition method to identify the edges of weld beads using a directional light source and a line laser, in which the sensors and the torch are fixed, and the plate is moved. The directional light source images and the line laser images are alternately collected and separately processed. Then the processed results are fused to accurately obtain the edges of weld beads. However, the method uses a large number of devices that occupy more space, resulting in poor system accessibility; the control logic for alternately capturing images by sensors is relatively complex, which imposes high-performance requirements on the devices.

In recent years, deep-learning technology has made significant progress. With its unique network structure, U-Net has performed exceptionally well in image segmentation tasks and has gradually been applied to the welding field. Zhang (Refs. 20–23) measured and predicted the profile of the weld pool by using active vision and deep-learning models such as CNN and U-Net. Feng (Ref. 24) proposed an end-to-end deep-learning-enhanced framework for multi-source sensing of weld pool images, processing images with different deep-learning models, including U-Net, and jointly predicting the weld depth. These indicated that it is feasible to use multi-sensor and deep learning to accurately measure MLMP weld bead profiles.

In this study, a composite vision method with a passive vision sensor and an active vision sensor was proposed for measuring intermediate as-welded bead profiles, for which the time difference of sensor acquisition was compensated. This paper is organized as follows: the experiment platform setup, calibration methods, and results are described in the "Experiment Platform and Calibration" section; a series of image-processing algorithms and feature extraction methods for intermediate as-welded beads of each layer are proposed in the "Image Processing and Feature Extraction section; finally, the feasibility of the method and system is verified by comparing the fitted dimensions with measured dimensions.

Experiment Platform and Calibration

Experiment Platform

The overall experiment platform setup is shown in Fig. 1A. It was equipped with a passive vision camera (Hikrobot MV-CS016-10GM) for capturing images of the as-welded bead surface and an active vision camera (Hikrobot MV-CA130-A0GM) for capturing line laser images (generated by a laser generator). A welding source (Aotai MAG-500PRO) and its accompanying gas cylinder were responsible for the welding work. A robot (KUKA KR 10) and its accompanying robot controller were used to control the motion of the torch. The upper computer was an industrial computer, which was responsible for receiving and processing the images collected by cameras and coordinating calculations.

Most importantly, the vision cameras of the platform were placed behind the torch. Figures 1B and C show the cameras behind and in front of the torch and their corresponding images, respectively. Compared with being placed in front of the torch, the cameras placed behind the torch could be less affected by spatter and changes in brightness. In addition, placing the cameras behind the torch could also avoid re-scanning and improve the efficiency of the overall MLMP welding.

Coordinate System Calibration

There were two cameras, a laser generator, and a robot in the proposed experiment platform; therefore, establishing a method to convert relationships between the coordinate systems of so many devices was the most important problem to be solved for the novel proposed method. The calibration of the platform included the calibration of the two cameras' intrinsic parameters, hand-eye calibration (Ref. 25) between the cameras and the robot, and the calibration of the line laser, as shown in Fig. 2. The principles and derivation processes of various calibration methods are detailed in the appendix; the main text focuses on introducing the calibration results and their applications.

The calibration of the two cameras' intrinsic parameters was the process of using two-dimensional plane information and a small amount of three-dimensional spatial information

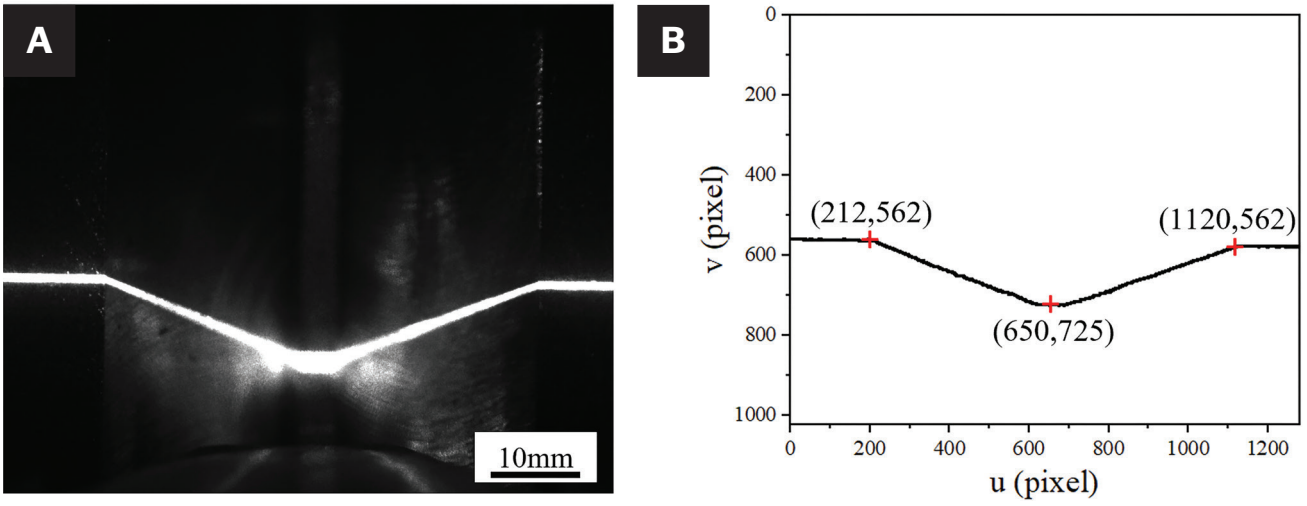

Fig. 3 — The coordinates of the groove in the laser line image: A — Laser line image of the groove; B — the coordinates of the groove.

to analyze the geometric relationship between the object points and their corresponding image points. Currently, the classic method for camera calibration is Zhang's calibration method (Ref. 26), which calculates the camera's intrinsic matrix by capturing images of a checkerboard from different angles. The camera intrinsic parameters were used to correct images and map image coordinates to camera coordinates and from camera coordinates to image coordinates.

According to Zhang's calibration method, the two-camera intrinsic parameters are shown as follows:

$$A_{PCam} = \begin{bmatrix} 3489.66491208 & 0 & 513.37436298 \\ 0 & 3495.77851983 & 525.40219628 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A_{NCam} = \begin{bmatrix} 3588.81355492 & 0 & 662.84378143 \\ 0 & 3597.11205395 & 644.22898891 \\ 0 & 0 & 1 \end{bmatrix}$$

Hand-eye calibration was carried out to obtain the transformation matrices between the camera and robot. The camera extrinsic parameters were utilized for matrix calculations with the robot tool coordinate, thus acquiring the transformation matrix of the camera to the robot coordinate system. Once these transformation matrices were obtained through hand-eye calibration, the relationship matrix between the two cameras could be determined via simple matrix operations. Subsequently, the pixel positions on the images of the two cameras were mapped to each other using the camera intrinsic parameters and the relationship matrix between the two cameras.

By calibrating and calculating the camera extrinsic matrices and the corresponding robot position matrices (Ref. 27), the hand-eye calibration matrices for the two cameras were obtained as follows:

$$T_E^{PC} = \begin{bmatrix} 0.06173209 & -0.05903626 & 0.99634526 & -198.50645812 \\ -0.99807261 & 0.00269187 & 0.06199862 & -3.92231431 \\ -0.0063422 & -0.99825221 & -0.0587563 & 63.01801909 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$T_E^{NC} = \begin{bmatrix} 0.03704426 & -0.06942367 & 0.99689923 & -101.93328551 \\ -0.99931362 & -0.00251013 & 0.03695917 & 3.07329363 \\ -0.0000635 & -0.99758411 & -0.06946901 & 143.10624454 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The transformation matrix between the two camera coordinate systems was as follows:

$$T_{NC}^{PC} = \begin{bmatrix} 0.99964116 & -0.02471744 & -0.0103254 & -96.57317261 \\ 0.02464664 & 0.99967221 & -0.00692897 & -6.995607946 \\ 0.01049328 & 0.00667201 & 0.99992269 & -80.08822545 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

In addition, it was necessary to establish a time correspondence between the passive and active vision images of the same location in the weld bead based on welding parameters and the physical relationship between the two cameras, since the two cameras simultaneously captured images from different positions. The time difference between the passive and active vision images at the same position was $t = l/v_w$, where l is the distance from the line laser in the active vision image to the center of the passive vision image and v_w is the welding speed.

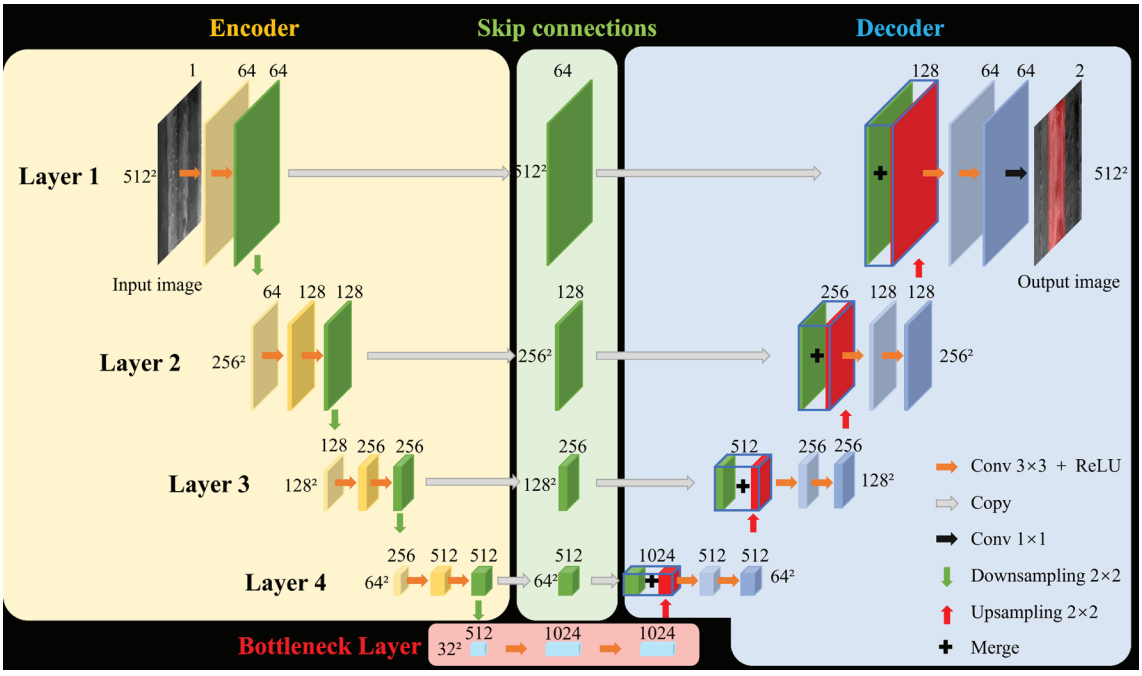

Fig. 4 — The overall structure model of the proposed U-Net.

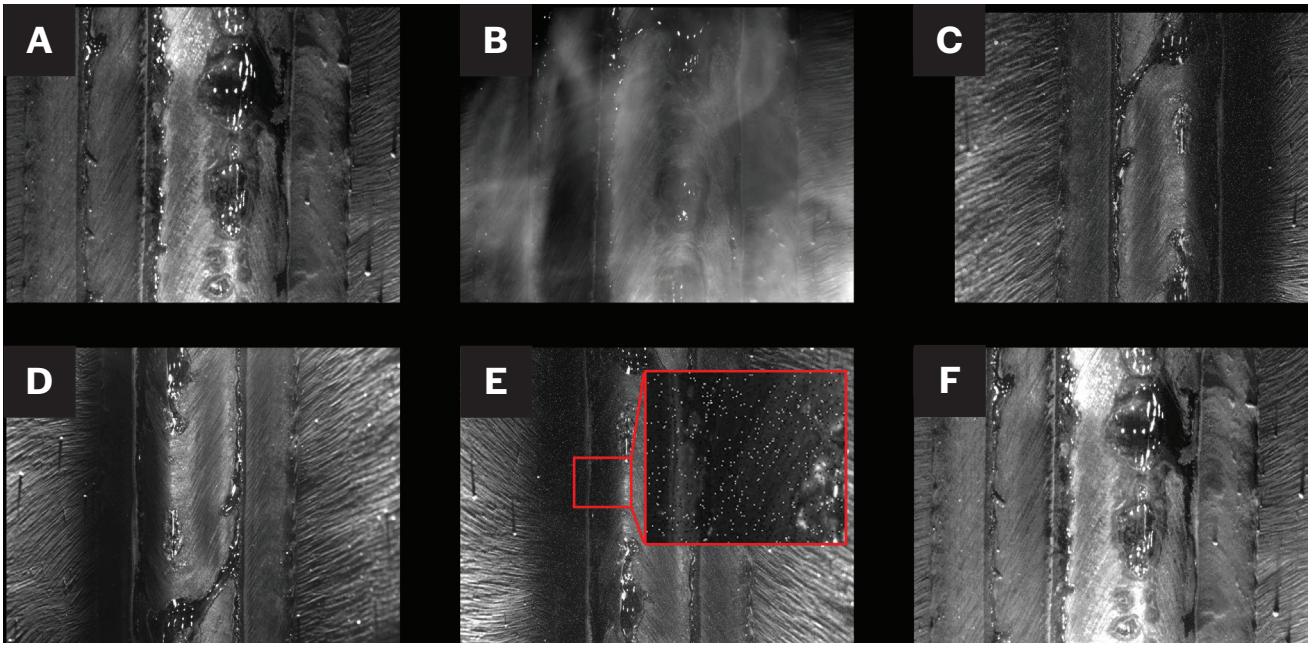

Fig. 5 — Image distribution of the dataset: A — Normal image; B — disturbed image; C — translated image; D — mirrored image; E — image with addition of Gaussian noise; F — image with brightness change.

The calibration of the line laser was carried out to obtain the relationship between the actual width, w, and height, h, of the laser line and the pixel width Δv and height Δu in the active vision image, which was as follows:

$$h = \frac{\Delta v \cdot d}{k - \Delta v} \tag{1}$$

$$w = j \cdot \Delta u \tag{2}$$

The k and j values were calculated by comparing the actual measured dimensions with the image pixel dimensions of the groove (the original joint preparation). The laser line image of the groove was processed to obtain the pixel dimensions of the groove, as shown in Fig. 3. A caliper was used to measure the precision-machined groove; the measured thickness of

the groove was 36.7 mm ($h = 36.7$ mm) and the width was 48.8 mm ($w = 48.8$ mm), where the distance from the laser generator to the workpiece surface was 198.5 mm ($d = 198.5$ mm). The coordinate of the line laser at the bottom of the groove was (650,725), and the coordinates at the top of the groove were (212,562) and (1120,562), resulting in $\Delta v = 163$ and $\Delta u = 908$. The calculated k and j values were 1044.2 and 0.0537. To minimize measurement errors, the k and j values for three images were calculated, as shown in Table 1. It can be observed that the variation in k and j values across several images remained within 1%, thereby satisfying the accuracy requirements of this method. Consequently, the relationships between the coordinates of the line laser in the active vision image and the actual dimensions were as follows:

$$h = \frac{198.5 \cdot \Delta v}{1042.5 - \Delta v} \quad (3)$$

$$w = 0.0536 \cdot \Delta u \quad (4)$$

Image Processing and Feature Extraction

The image processing algorithms of the passive and active vision images were used to extract the features of the intermediate MLMP as-welded beads, and the compositing methods of these features are also detailed in this section. The experiment's conditions are listed in Table 2, and the typical passive and active vision images captured by the proposed composite vision method are shown in Fig. 1B.

Passive Vision Image Processing

The challenge of extracting the as-welded bead features from a passive vision image involves avoiding the interference of the features of the groove and other weld beads. To address this issue, a U-Net neural network, which is a highly efficient convolutional neural network architecture specifically designed for image segmentation (Ref. 28), was

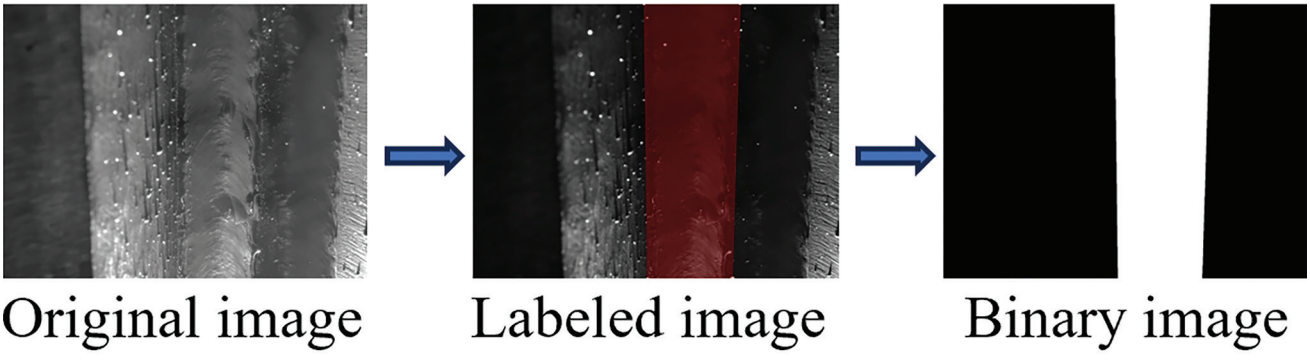

Fig. 6 — Annotation process of the dataset.

Table 2 — Welding Parameters and Acquisition Parameters Setting

Parameters	Values
Workpiece (S355) Dimensions (mm^3)	375 × 100 × 40
Gas (80%Ar+20%CO_2) Flow Rate (L/min)	20
Average Current (A)	220
Voltage (V)	24
Welding Speed (m/min)	0.36
Frame Rate of Active Vision Camera (fps)	30
Frame Rate of Passive Vision Camera (fps)	30
Exposure Time of Active Vision Camera (μs)	500
Exposure Time of Passive Vision Camera (μs)	6000

Table 3 — The Sample Distribution of the Dataset

Image Dataset	Tra. Set	Val. Set	Total
Normal	270	30	300
Disturbed	189	21	210
Translated	459	51	510
Mirrored	459	51	510
Addition of Gaussian Noise	459	51	510
Brightness Change	459	51	510
Total	2295	255	2550

proposed to segment the as-welded bead features from the passive vision images.

The overall structure model of the proposed U-Net is shown in Fig. 4. This model comprised four components: encoder, skip connections, bottleneck layer, and decoder. Images are represented as blocks in Fig. 4, and their sizes included resolution (left or right side of each layer) and channel number (above each block). Depending on the resolution of the input image and the difficulty of feature extraction, the proposed U-Net contained four layers in the encoder part. At the beginning of the encoder, an input image with a resolution of 512 × 512 pixels was processed through two convolutions (orange arrow in Fig. 4) with 64 convolution kernels in Layer 1, obtaining a green block with a size of 512 × 512 × 64. Two convolutions were used to enhance the ability of feature extraction; the number of convolution kernels was equal to the channel number of the obtained block. The green block obtained from each layer had two functions. First, it was copied (gray arrow in Fig. 4) to the corresponding layer of the decoder through skip connections. Second, it was used for downsampling (green arrow in Fig. 4), which helped extract as-welded bead features of different scales by reducing resolution. As the layers went deeper, the number of convolution kernels gradually increased to 128, 256, and 512 to learn more-complex and comprehensive as-welded bead features. The bottleneck layer received the down-sampling result of Layer 4 of the encoder part, and the channel number was adjusted by two convolutions to facilitate upsampling (red arrow in Fig. 4) in the decoding part. The decoder part also contained four layers corresponding to the encoder part, and the processing sequence of blocks was carried out from Layer 4 upward. Upsampling restored the resolution and changed the channel number to a value that matched the corresponding layer, to merge with the block copied through skip connections. In each layer, the merged blocks underwent two convolutions. At the end of the decoder, a convolution (black arrow in Fig. 4) with a kernel size of 1 × 1 and kernel number of 2 was applied to obtain an output image with a size of 512 × 512 × 2, whose two channels represented the as-welded bead area and background area respectively, thus extracting the as-welded bead feature from the passive vision image.

By analyzing about 10,000 passive vision images of ten intermediate as-welded beads, these images could be divided into normal images (as shown in Fig. 5A) and disturbed images with interference such as welding fume (as shown in Fig. 5B), with an approximate ratio of 86:14. The proportion of disturbed images in the dataset was appropriately increased to ensure the recognition accuracy of the model for disturbed images. To better expand the dataset, translation, mirroring, addition of Gaussian noise, and brightness change were carried out on the dataset images, and the results are shown in Figs. 5C–F. Finally, a dataset containing 2550 images was obtained, and the sample distribution is shown in Table 3. In order to facilitate the model training, the dataset's images were uniformly adjusted to a resolution of 512 × 512 pixels without distortion and annotated according to the process shown in Fig. 6. In addition, the width comparison of the measured values and fitted values is shown in Table 4.

The train loss value and the val (validation) loss value of the proposed U-Net after 200 epochs of training are shown in Fig. 7A; they were used together to evaluate the fitting degree of the model to the data. The loss values were stable at 0.013, indicating that the model's prediction was very close to the real situation. The MIoU (mean intersection over union) value of the proposed U-Net is shown in Fig. 7B; this was used to evaluate the performance of the segmentation model. The MIoU value was stable at 99.05, indicating excellent model performance. The output image of the training U-Net model is shown in Fig. 7C. It can be seen that the proposed U-Net could accurately identify the contour of the as-welded bead from the other weld beads and groove.

Because a fixed position in the as-welded bead could be captured by multiple adjacent passive vision images, the features in multiple adjacent images were used to improve the recognition accuracy of the as-welded bead edge. The capture of the adjacent images was as shown in Fig. 8A. The point $M_i(u_i, v_i)$ was the coordinate of the as-welded bead

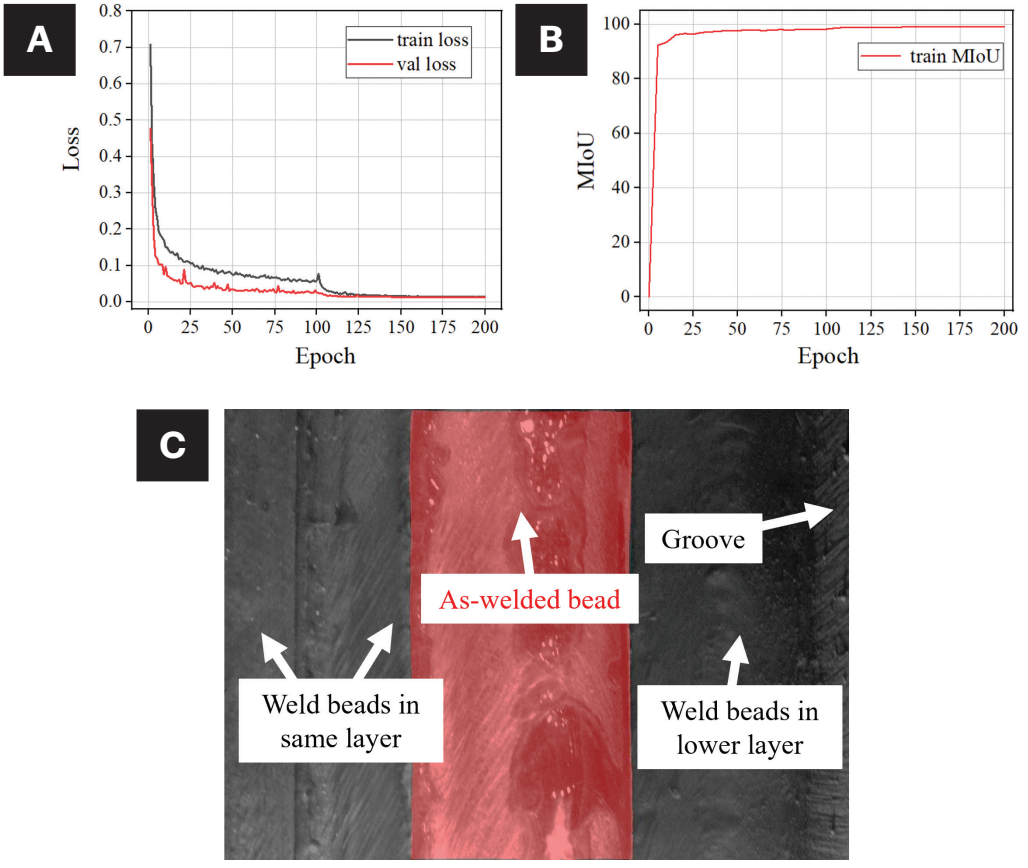

Fig. 7 — Training results and image validation: A — Loss values; B — MIoU value; C — image validation result.

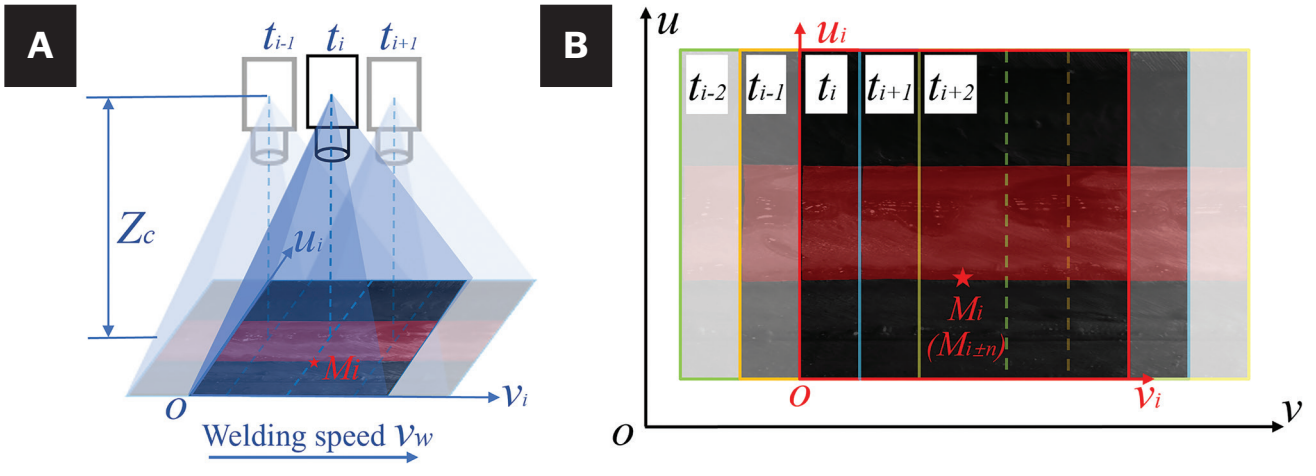

Fig. 8 — The positional relationship of: A — Camera in adjacent moments; B — images captured in adjacent moments.

edge in the image captured at t_i moment. The v_i value was 540 (half of the image height), and the u_i value was determined by identifying the as-welded bead edge. According to the camera parameters and welding speed v_w, five adjacent images could be stably used to calculate the average coordinate of an as-welded bead edge, as shown in Fig. 8B. The image at t_i moment was the main image and the corresponding coordinate in the adjacent image of the point M_i (u_i, v_i) was $M_{i\pm n}$ $(u_{i\pm n}, v_{i\pm n})$, $n \in \{1,2\}$. The $v_{i\pm n}$ value could be obtained as follows:

$$v_{i\pm n} = v_i - (t_{i\pm n} - t_i) \cdot v_w \cdot \frac{Z_c}{f_y} \cdot W_S \qquad (5)$$

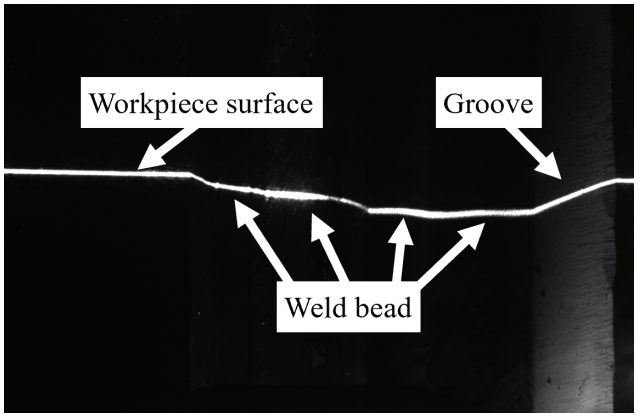

Fig. 9 — Information in active vision image.

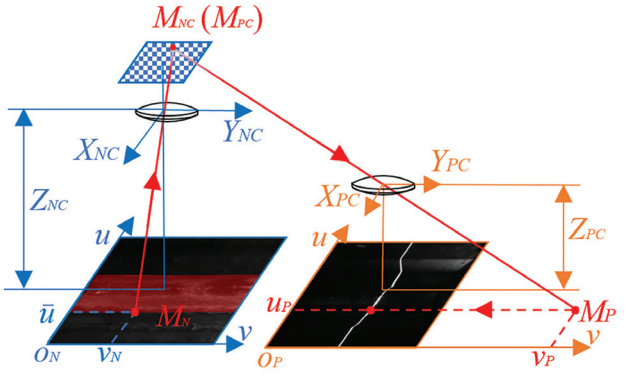

Fig. 10 — Coordinate transformation from passive vision image to active vision image.

where W_s is the width of the camera pixel size available from the camera product specifications. The $u_{i\pm n}$ value was obtained by identifying the as-welded bead edge after determining $v_{i\pm n}$. All $u_{i\pm n}$, $n \in \{1,2\}$ were used to calculate the average value $\bar{u}$, and $(\bar{u}, v_i)$ was taken to be the final recognized coordinate of point M_i.

Active Vision Image Processing

The active vision image, as shown in Fig. 9, contains laser lines of the workpiece surface, groove, and weld beads, so it is difficult to obtain the laser line of the as-welded bead by directly processing the image. According to the relationship between the passive and active vision camera coordinate systems, the obtained coordinates $(\bar{u}, v_i)$ are used to accurately determine the region of interest (ROI) in an active vision image.

The coordinate transformation from passive vision image to active vision image is shown in Fig. 10. The coordinate of the edge point in the passive vision image coordinate system $o_N - uv$ is $M_N[\bar{u}, v_N]^T$. The point M_N in the passive vision camera coordinate system $M_{NC}[X_{NC}, Y_{NC}, Z_{NC}]^T$ can be calculated by hand-eye calibration. The point $M_{PC}[X_{PC}, Y_{PC}, Z_{PC}]^T$ in the active vision camera coordinate system is obtained by using T_{NC}^{PC}. The point $M_P[u_P, v_P]^T$ in the active vision image coordinate system $o_P - uv$ can be obtained as follows:

$$\begin{bmatrix} u_P \\ v_P \\ 1 \end{bmatrix} = \frac{1}{Z_{PC}} A_{PCam} \begin{bmatrix} X_{PC} \\ Y_{PC} \\ Z_{PC} \end{bmatrix} = \frac{Z_{NC}}{Z_{PC}} T_{NC}^{PC} \cdot A_{NCam}^{-1} \cdot \begin{bmatrix} \bar{u} \\ v_N \\ 1 \end{bmatrix} \tag{6}$$

It is noteworthy that since two cameras were arranged sequentially along the welding direction and their fields of view (FOV) did not overlap, the v_P would exceed the FOV of the active vision camera. However, the u_N remained within the FOV of the active vision camera; therefore, the u_P value was retained for subsequent processing after coordinate transformation.

The u_P value was used as the standard for the u-direction of the ROI, as shown in Fig. 11A. The position of the line laser in the active vision image remained relatively constant because cameras and laser generators were fixed on the torch; therefore, the standard for the u-direction of the ROI was selected based on a fixed value, and the result is shown in Fig. 11B. The extraction result of the ROI is shown in Fig. 11C, which is considered to accurately describe the laser line region of as-welded bead.

A standard image processing strategy including denoising and binarization, contour extraction, center point extraction, and curve fitting was proposed to extract the as-welded bead features in the ROI image, as shown in Fig. 12. Morphological processing using the opening operation was applied within the ROI area for denoising. Then the processed image was binarized by using Equation 7. The results of denoising and binarization are shown in Fig. 12A. The Sobel (Ref. 29) operator was employed to detect the edges of line laser in a binarized image by identifying areas with significant changes in grayscale gradients, and the results of image processing are shown in Fig. 12B. Then the center points of the laser line were extracted using the grayscale centroid method, as described by Equation 8. The extraction results are shown in Fig. 12C. Finally, Equation 9, where $n = 4$, was used to fit the surface profile curve of the as-welded bead through the extracted center points. The surface profile curve of the as-welded bead is shown in Fig. 12D.

$$dst(u,v) = \begin{cases} maxVal, & if\ \mathrm{src}(u,v) > threshold \\ 0, & otherwise \end{cases} \tag{7}$$

$$\bar{v} = \frac{\sum_{v=1}^{L} I(u,v) \cdot v}{\sum_{v=1}^{L} I(u,v)} \tag{8}$$

$$y = a_1 + a_2 x + a_3 x^2 + \cdots + a_{n+1} x^n \tag{9}$$

Feature Extraction

The actual contour curve of the as-welded bead was obtained by using Equations 3 and 4 to transform the fitting curve in the ROI image, as shown in Figs. 13A and B. The maximum coordinate difference along the *u* axis is defined as the width of the as-welded bead since the *u* value of the ROI image was obtained by extracting the accurate contour coordinate. In contrast, the maximum coordinate difference along the *v* axis is defined as the height of the as-welded bead, and this standard was used to verify the fitting accuracy of the as-welded bead profile.

Results

A complete MLMP welding experiment was completed according to the experimental parameters in Table 2, and a total of nine layers and 29 beads were obtained, as shown in Fig. 14A, to verify the accuracy of the composite vision method.

For the verification of fitting results, the surface dimensions of the weld beads were measured by using a caliper after each bead was welded, since the weld bead in the upper layer covered the weld bead in the lower layer during MLMP welding. For example, bead 28 after welding is shown in Fig. 14B, and bead 29 had not yet been welded; the bead

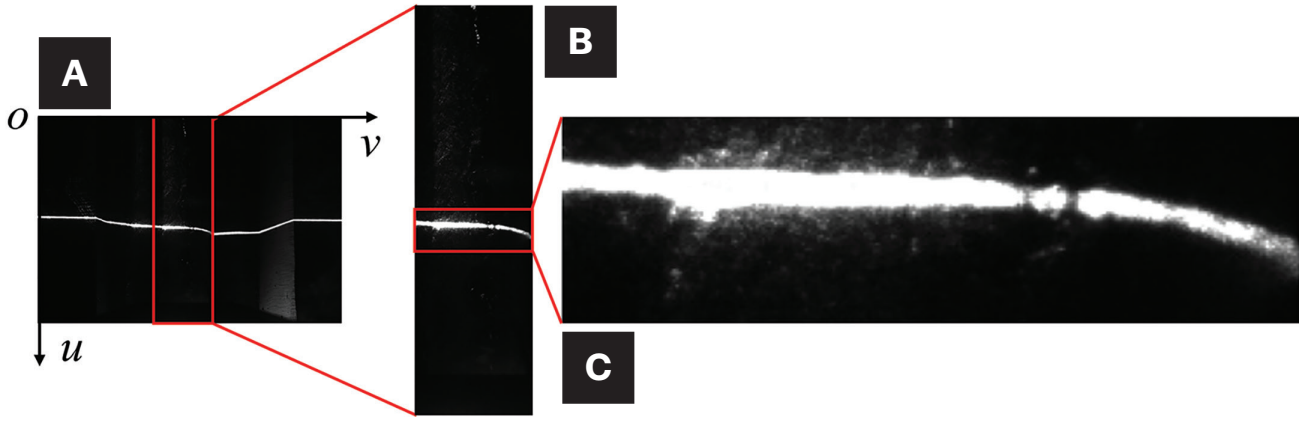

Fig. 11 — Extraction of the ROI for the as-welded bead image: A — The v range of the ROI; B — the u range of the ROI; C — the ROI image.

Table 4 — The Width Comparison of the Measured Values and Fitted Values

	Fitted (mm)	Measured (mm)	Error (mm)	Error rate (%)
1	12.06	11.6	−0.46	3.97
2	12.23	12.0	−0.23	1.92
3	12.18	11.7	−0.48	4.10
4	12.10	11.7	−0.40	3.42
5	12.19	11.9	−0.29	2.44
6	12.33	12.0	−0.33	2.75
7	12.35	12.0	−0.35	2.92
8	12.21	12.0	−0.21	1.75
9	12.40	12.1	−0.30	2.48
Average	12.228	11.89	−0.339	2.860

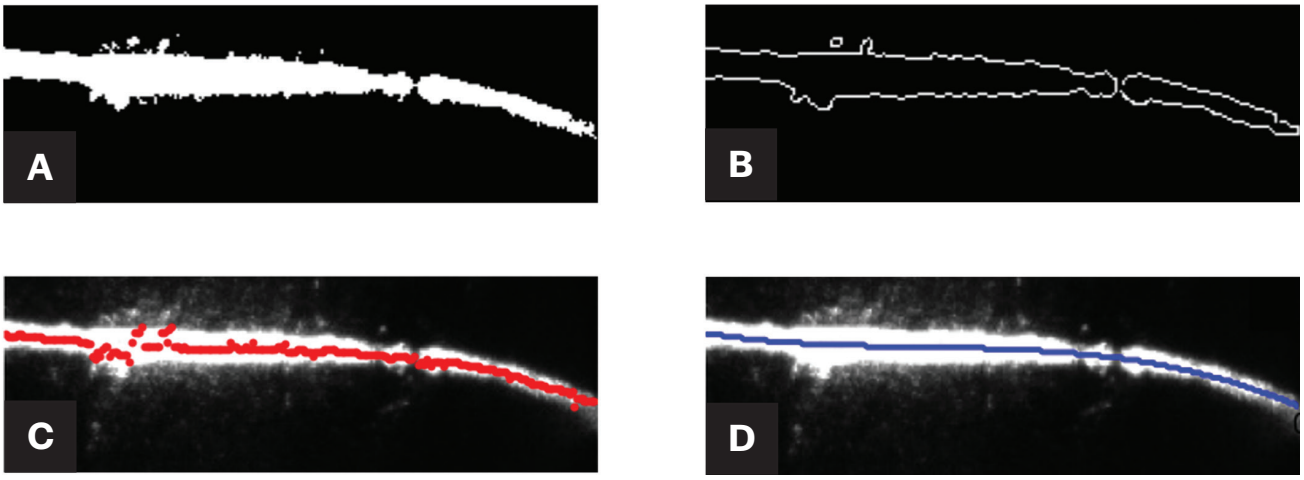

Fig. 12 — The steps of surface profile curve fitting: A — Denoising and binarization; B — contour extraction; C — center of gravity method to extract the center point; D — fitting curve.

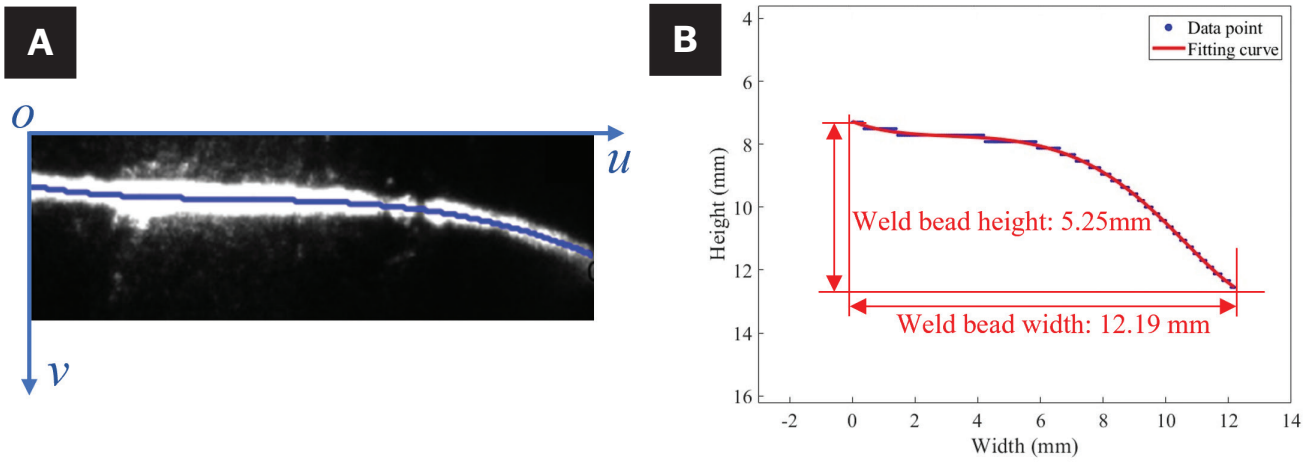

Fig. 13 — The as-welded bead fitting curve: A — Coordinate dimensions; B — actual dimensions.

cross-sections at nine positions, as shown in Fig. 14C, were selected for measurement and fitting curves.

The height and width at nine positions of bead 28 are shown in Fig. 15; the height and width were measured using a caliper immediately after bead 28 was welded, and Fig. 15 was taken after bead 29 was welded. For convenience, the closed area enclosed by the white solid and dashed lines in Fig. 15 is used to describe the cross-sectional profile of bead 28 at the corresponding position, where the white solid line is the visible contour and the white dashed line is the reconstructed contour line through the recognition algorithm. The fitting curve and fitting dimensions based on the composite vision method at the selected position are shown in Fig. 16. The comparison of width and height between the measured value and the fitted value are shown in Table 5. The error rates of width and height were calculated by

$$\begin{cases} \varepsilon_W = \dfrac{W_M - W_F}{W_M} \\ \varepsilon_H = \dfrac{H_M - H_F}{H_M} \end{cases}$$

where ε_W and ε_H are the error rates of width and height; W_M and H_M are the measured values of width and height; and W_F and H_F are the fitted values of width and height. It can be seen that the average errors of the as-welded bead width and height were 2.860% and 3.875%, and the maximum errors of the as-welded bead width and height were 4.10% and 6.11%. Other intermediate weld beads were measured in the same way, and the results showed that the errors of width and height were less than 5% and 7%, which is considered to meet the actual requirements of MLMP welding, indicating that the method could effectively extract the profile dimensions of the as-welded bead in MLMP welding and provide a basis for the subsequent adjustment of torch position and welding parameters.

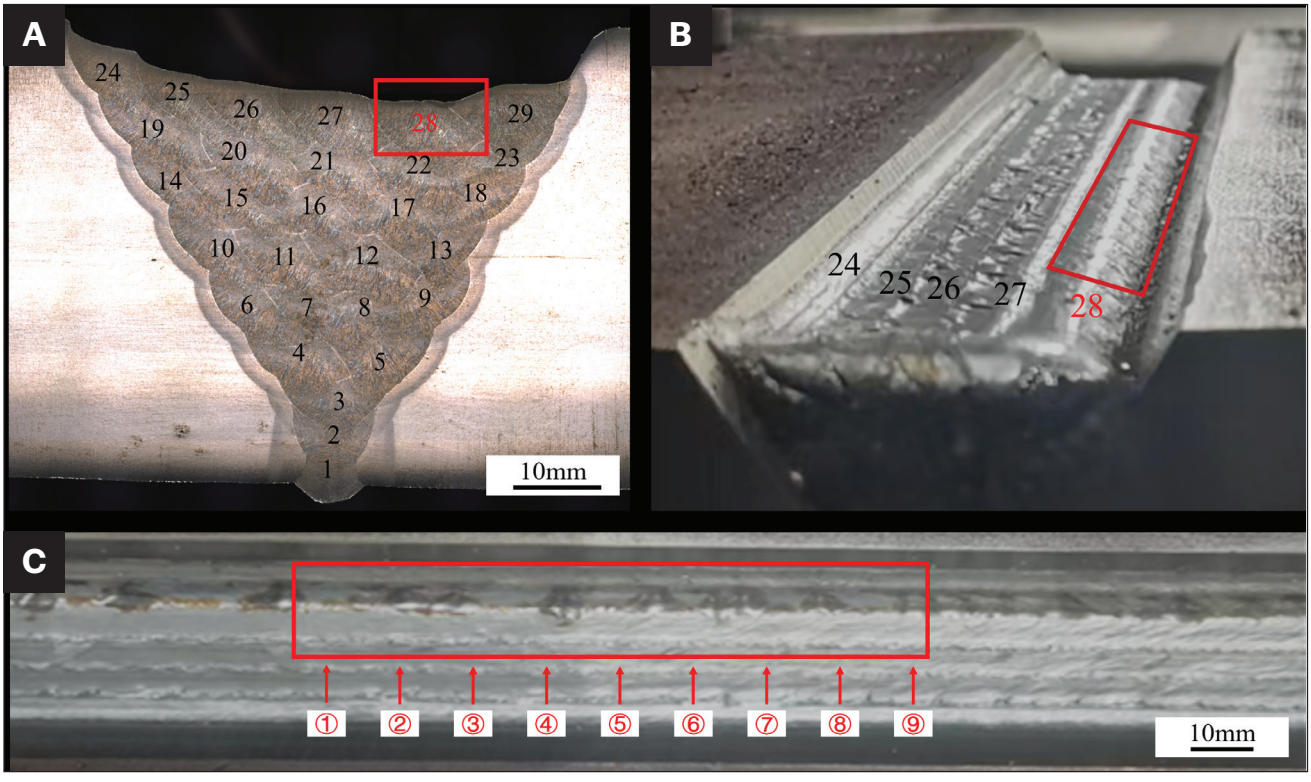

Fig. 14 — The weld beads for verifying accuracy: A — The distribution of weld beads in groove; B — the condition of bead 28; C — the selected position in the as-welded bead.

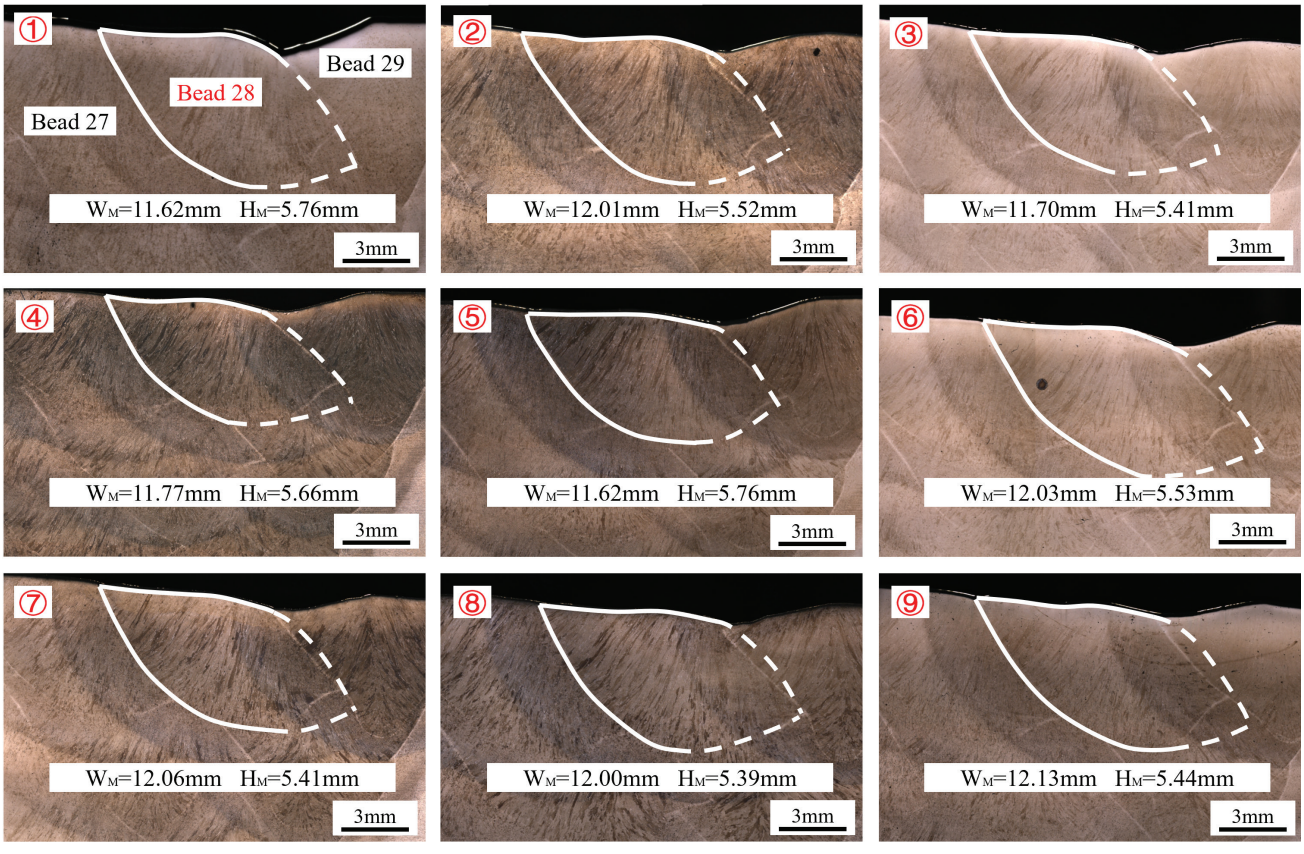

Fig. 15 — The measured value of the selected position.

However, at present, the performance of this method has only been initially verified in an intermediate as-welded bead. The surface profile of the first and last weld beads of each layer will be different from that of the intermediate weld bead due to the influence of the groove sidewall. In Part 2 of our series of papers, the adaptability of this method to the first and last welds of each layer will be highlighted, and corresponding optimization strategies will be proposed.

After the availability of the identification algorithm for all position beads has been verified, we will focus on building the ultimate purpose of this system, which is to achieve overall identification and automatic correction for MLMP welding. This involves researching the algorithms for alignment and correction as well as system automation. The final verification of the overall system in the correction effect is planned to be shown in Part 3.

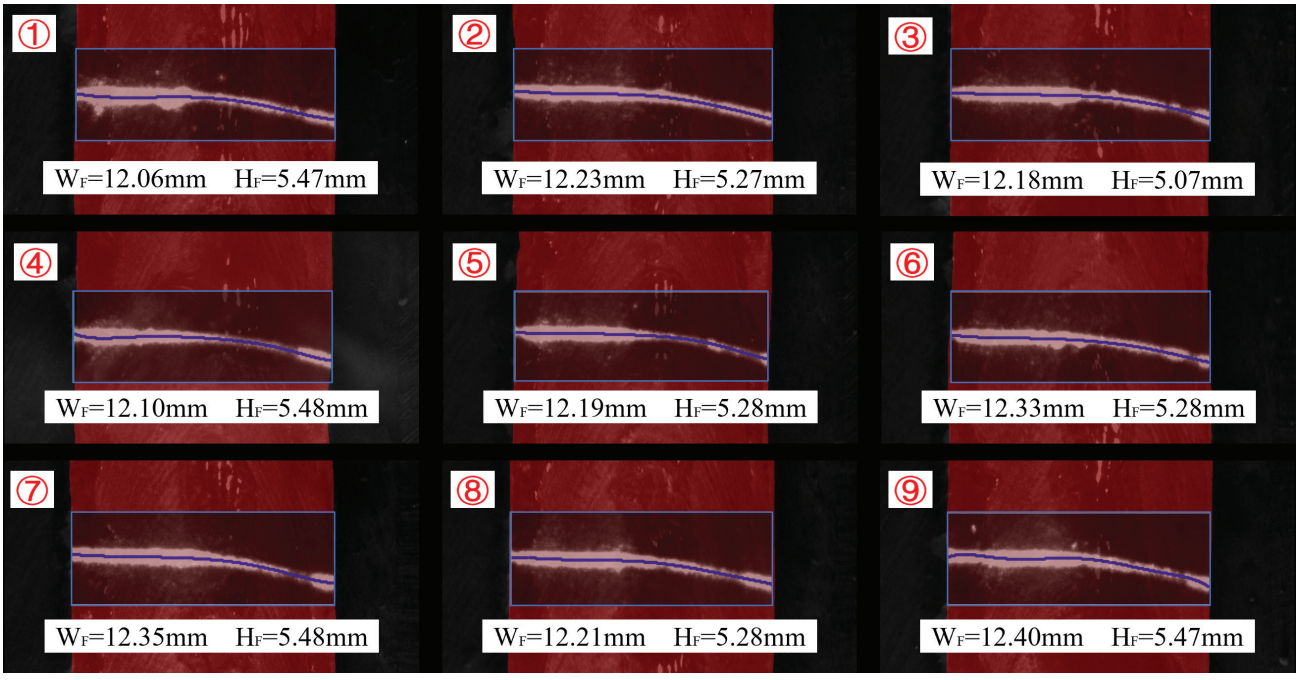

Fig. 16 — The fitted value of the selected position.

Table 5 — The Height Comparison of the Measured Values and Fitted Values

	Fitted (mm)	Measured (mm)	Error (mm)	Error rate (%)
1	5.47	5.7	−0.23	4.04
2	5.27	5.5	−0.23	4.18
3	5.07	5.4	−0.33	6.11
4	5.48	5.6	−0.12	2.14
5	5.28	5.5	−0.22	4.00
6	5.28	5.5	−0.22	4.00
7	5.48	5.4	0.08	1.48
8	5.28	5.4	−0.12	2.22
9	5.47	5.4	0.07	1.30
Average	5.342	5.49	−0.147	3.275

Conclusion

1. A novel composite vision method was proposed for measuring the as-welded bead profile during MLMP welding. The method used passive vision images to identify the as-welded bead edges and, based on that information, used active vision images to obtain a 3D profile of the as-welded bead through coordinate transformation, thereby accurately measuring the as-welded bead profile dimensions. The weld bead height and width fitting errors of this method were less than 5% and 7%, respectively.

2. The active vision images were processed by a segmentation method based on U-Net to solve the extraction challenges due to unobvious as-welded bead features, and the features of multiple adjacent images were used to improve the recognition accuracy of the as-welded bead edge.

3. The active vision images were processed by extracting an ROI with the transformed coordinates from passive vision images and then using a series of image processing algorithms. The intermediate as-welded bead at each moment was fitted out to obtain the real surface profile.

Acknowledgments

This program was supported by the National Natural Science Foundation of China (Grant No. 52005366).

References

1. ASTM International. 2019. ASTM A131/A131M-19, *Standard specification for structural steel for ships*.

2. Alipooramirabad, H., Paradowska, A., Ghomashchi, R., and Reid, M. 2017. Investigating the effects of welding process on residual stresses, microstructure and mechanical properties in HSLA steel welds. *Journal of Manufacturing Processes* 28: 70–81. DOI: 10.1016/j.jmapro.2017.04.030

3. Pesin, A., Drigun, E., Pustovoytov, D. O., and Pesin, I. 2016. Development of the technology of various large bodies manufacturing based on combined methods of deformation. *Key Engineering Materials* 716: 659–666. DOI: 10.4028/www.scientific.net/KEM.716.659

4. Zhang, H., Lu, H., Cai, C., and Chen, S. 2011. "Robot path planning in multi-pass weaving welding for thick plates." In *Robotic Welding, Intelligence and Automation RWIA'2010*, edited by Tarn, T.-J., Chen, S.-B., and Fang, G., 351–359. DOI:10.1007/978-3-642-19959-2_43

5. Zhang, H. J., Zhang, G. J., Cai, C. B., Yin, Z., and Wu, L. 2009. Self-defining path layout strategy for thick plate arc welding robot. *Transactions of the China Welding Institution* 30(3): 61–64.

6. Xu, F., Xiao, R., Hou, Z., Xu, Y., Zhang, H., and Chen, S. 2019. Multi-layer multi-pass welding of medium thickness plate. Technologies, advances and future prospects. *Transactions on Intelligent Welding Manufacturing* 3(4): 3–33. DOI: 10.1007/978-981-33-6502-5_1

7. Yan, Z., Xu, D., Li, Y., Tan, M., and Zhao, Z. 2007. A survey of the sensing and control techniques for robotic arc welding. *Measurement and Control* 40(5): 146–150. DOI: 10.1177/002029400704000503

8. Mahajan, A., and Figueroa, F. 1997. Intelligent seam tracking using ultrasonic sensors for robotic welding. *Robotica* 15(3): 275–281. DOI: 10.1017/S0263574797000313

9. Xu, F., Xu, Y., Zhang, H., and Chen, S. 2022. Application of sensing technology in intelligent robotic arc welding: A review. *Journal of Manufacturing Processes* 79: 854–880. DOI: 10.1016/j.jmapro.2022.05.029

10. Zhang, Y. M., Wu, L., Walcott, B. L., and Chen, D. H. 1993. Determining joint penetration in GTAW with vision sensing of weld-face geometry. *Welding Journal* 72(10): 463-s to 469-s.

11. Zhang, Y. M., Wang, Q. Y., and Liu, Y. K. 2021. Adaptive intelligent welding manufacturing. *Welding Journal* 100(1): 63-s to 83-s. DOI:10.29391/2021.100.006

12. Beattie, R. J., Cheng, S. K., and Logue, P. S. 1988. The use of vision sensors in multipass welding applications. *Welding Journal* 67(11): 28–33.

13. Ge, J., Zhu, Z., He, D., and Chen, L. 2005. A vision-based algorithm for seam detection in a PAW process for large-diameter stainless steel pipes. *International Journal of Advanced Manufacturing Technology* 26: 1006–1011. DOI:10.1007/S00170-004-2070-2

14. Xu, Y., Yu, H., Zhong, J., Lin, T., and Chen, S. 2012. Real-time image capturing and processing of seam and pool during robotic welding process. *Industrial Robot: An International Journal* 39(5): 513–523. DOI: 10.1108/01439911211249805

15. Wang, Z. 2024. The active visual sensing methods for robotic welding: review, tutorial and prospect. *IEEE Transactions on Instrumentation and Measurement* 73. DOI: 10.1109/TIM.2024.3485460

16. Zhang, Y. M., and Kovacevic, R. 1997. Real-time sensing of sag geometry during gta welding. *Journal of Manufacturing Science and Engineering-Transactions of the ASME* 119(2): 151–160. DOI: 10.1115/1.2831090

17. Zou, Y., and Chen, T. 2018. Laser vision seam tracking system based on image processing and continuous convolution operator tracker. *Optics and Lasers in Engineering* 105: 141–149. DOI: 10.1016/j.optlaseng.2018.01.008

18. Jin, Z., Li, H., Wang, Q., and Gao, H. 2017. Online measurement of the GMAW process using composite sensor technology. *Welding Journal* 96(4): 133-s to 142-s.

19. Zeng, J., Chang, B., Du, D., Wang, L., Chang, S., Peng, G., and Wang, W. 2018. A weld position recognition method based on directional and structured light information fusion in multi-layer/multi-pass welding. *Sensors* 18(1): 129. DOI: 10.3390/s18010129

20. Song, H. S., and Zhang, Y. M. 2007. Image processing for measurement of three-dimensional GTA weld pool surface algorithms are explored for processing the image of a dot-matrix laser pattern reflected from the weld pool surface. *Welding Journal* 86(10): 323-s to 330-s.

21. Chao, L. I., Wang, Q., Jiao, W., Johnson, M., and Zhang, Y. M. 2020. Deep learning-based detection of penetration from weld pool reflection images. *Welding Journal* 99(9): 239-s to 245-s. DOI: 10.29391/2020.99.022

22. Yu, R., Kershaw, J., Wang, P. and Zhang, Y. 2021. Real-time recognition of arc weld pool using image segmentation network. *Journal of Manufacturing Processes* 72: 159–167. DOI: 10.1016/j.jmapro.2021.10.019

23. Li, T., Cao, Y., and Zhang, Y. 2024. Analysis of weld pool region constituents in GMAW for dynamic reconstruction through characteristic enhancement and LSTM U-Net networks. *Journal of Manufacturing Processes* 127: 573–588. DOI: 10.1016/j.jmapro.2024.07.084

24. Feng, Y., Chen, Z., Wang, D., Chen, J., and Feng, Z. 2019. DeepWelding: A deep learning enhanced approach to GTAW using multisource sensing images. *IEEE Transactions on Industrial Informatics* 16(1): 465–474. DOI: 10.1109/TII.2019.2937563

25. Horaud, R., and Dornaika, F. 1995. Hand-eye calibration. *The International Journal of Robotics Research* 14(3): 195–210. DOI: 10.1177/027836499501400301

26. Zhang, Z. Y. 1999. Flexible camera calibration by viewing a plane from unknown orientations. *Proceedings of the Seventh IEEE International Conference on Computer Vision* 1: 666–673. DOI: 10.1109/ICCV.1999.791289

27. Tsai, R. Y. and Lenz, R. K. 1989. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation* 5(3): 345–358. DOI:10.1109/70.34770

28. Ronneberger, O., Fischer, P., and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241. DOI: 10.48550/arXiv.1505.04597

29. Marr, D., and Hildreth, E. 1980.Theory of edge detection. *Proceedings of the Royal Society of London* 207: 187–217. DOI: 10.1098/rspb.1980.0020

Appendix

The composition of the experiment platform's coordinate system is shown in Fig. A(A). It includes five coordinate systems: the robot base coordinate system $O_B - X_B Y_B Z_B$; the tool coordinate system at the robot's end effector, specifically the welding torch $O_E - X_E Y_E Z_E$; two camera coordinate systems, the active vision camera coordinate system $O_{PCam} - X_{PCam} Y_{PCam} Z_{PCam}$ and the passive vision camera coordinate system $O_{NCam} - X_{NCam} Y_{NCam} Z_{NCam}$; and the global coordinate system $O_W - X_W Y_W Z_W$.

The Calibration of Camera Intrinsic Parameters

The intrinsic parameters of two cameras were calibrated based on the camera projection model shown in Fig. A(B). The image coordinate system $o - uv$ was a two-dimensional coordinate system. Suppose there was a point $P_W [X_W, Y_W, Z_W]^T$ in the global coordinate system (the subscript represents the coordinate system in which the point or axis coordinates were located, and the superscript T represents the transposed matrix; expressions in this article will follow this convention), and its corresponding projection in the image coordinate system, which is denoted as $P[u,v]^T$. The coordinate of P_W in the camera coordinate system can also be represented as $[X_{Cam}, Y_{Cam}, Z_{Cam}]^T$. The conversion relationship for point P between the image coordinate system and the camera coordinate system was as follows:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_{Cam} \\ Y_{Cam} \\ Z_{Cam} \end{bmatrix} \quad \text{(A1)}$$

where Z_C is the scale factor; f_x and f_y are the scale factors of the u axis and v axis, respectively, and (u_0, v_0) is the principal point coordinate of the camera in the image coordinate system.

The transformation between the camera coordinate system and the global coordinate system belonged to the rigid transformation, which is represented by the rotation matrix R and the translation vector t:

$$\begin{bmatrix} X_{Cam} \\ Y_{Cam} \\ z_{Cam} \end{bmatrix} = R \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} + t \quad \text{(A2)}$$

where R is a 3 × 3 rotation orthogonal matrix and t is a 3 × 1 translation vector. The transformation equation of the global coordinate system to the image coordinate system was obtained as follows:

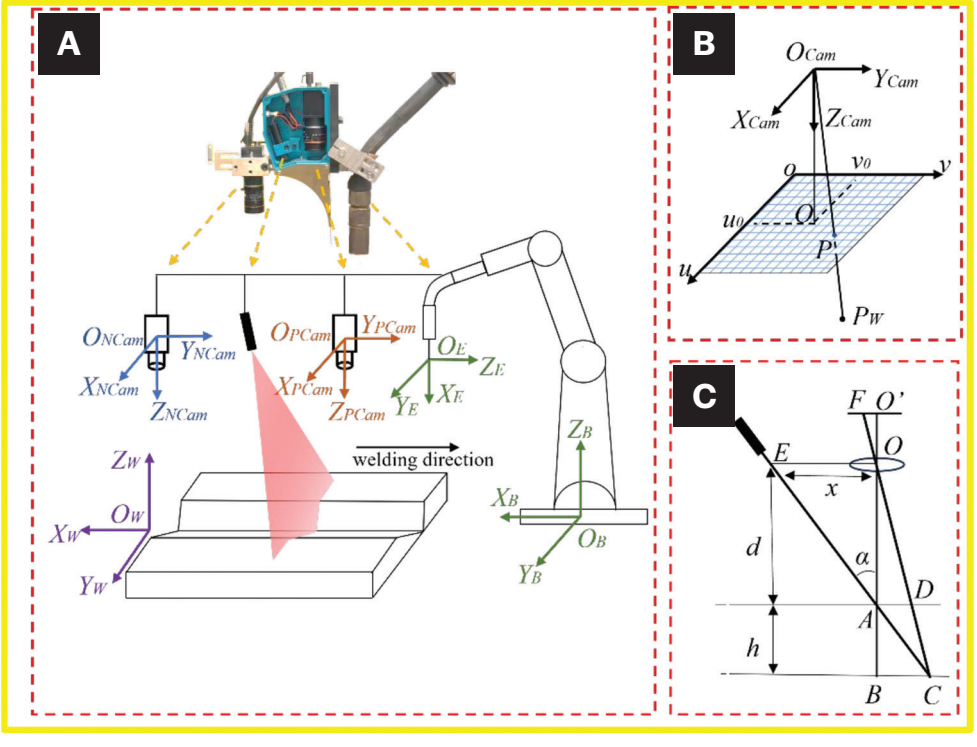

Fig. A — Schematic diagram of the coordinate systems in the experiment platform: A — The coordinate systems of the experiment platform; B — projection model of the camera; C — triangular projection model of the laser line.

$$Z_C\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f_x & 0 & u_0\\0 & f_y & v_0\\0 & 0 & 1\end{bmatrix}[Rt]\begin{bmatrix}X_W\\Yw\\zw\end{bmatrix}$$
$$=AT_W^{Cam}\begin{bmatrix}X_W\\Y_W\\Z_W\end{bmatrix} \quad \text{(A3)}$$

where A is the camera intrinsic parameter matrix and T_W^{Cam} the camera extrinsic parameter matrix that ignores the distortion correction. Since the cameras had a front view of the workpiece, the camera distortion could be ignored in this composite vision method. By calibrating camera intrinsic parameters, image pixel positions could be converted into the camera coordinate system.

Transition Matrix between Cameras

The coordinate transformation matrices between the cameras and the robot were obtained by hand-eye calibration. The robot base coordinate system $O_B-X_BY_BZ_B$ and the global coordinate system $O_W-X_WY_WZ_W$ were always fixed, while the tool coordinate system $O_E-X_WY_WZ_W$ and the camera coordinate system $O_{Cam}-X_{Cam}Y_{Cam}Z_{Cam}$ changed with the robot position. Based on these coordinate transformation relationships, the relationship between point P_W and its corresponding point P_B in the robot base coordinate system was as follows:

$$P_B=\left(T_B^E\right)\left(T_E^C\right)\left(T_C^W\right)P_W \quad \text{(A4)}$$

where T_E^B is the coordinate transformation matrix from the robot base coordinate system to the tool coordinate system, and it can be obtained by robot tool center point calibration; T_E^C is the coordinate transformation matrix from the end-effector coordinate system to the camera coordinate system, which is also expected to be obtained; T_C^W is the coordinate transformation matrix from the camera coordinate system to the global coordinate system, and it can be obtained from camera extrinsic parameters. By changing the robot position, a new set of corresponding relationships was established as follows:

$$\left(T'^E_B\right)\left(T_E^C\right)\left(T'^W_C\right)P_W=P_B \quad \text{(A5)}$$

where T'^E_B and T'^W_C are the coordinate transformation matrices at the other robot position. By combining Equations A4 and A5, the relationship between these coordinate transformation matrices is as follows:

$$\left(T'^E_B\right)^{-1}\left(T_B^E\right)T_E^C=T_E^C\left(T'^W_C\right)\left(T_C^W\right)^{-1} \quad \text{(A6)}$$

Equation A6 can be regarded as an equation of the form $AX = XB$, where $A=(T'^E_B)^{-1}(T_B^E)$; $B=(T'^W_C)(T_C^W)^{-1}$; $X=T_E^C$, which can be obtained by solving the equation $AX = XB$.

The coordinate transformation matrices from the active vision camera coordinate system and the passive vision camera coordinate system to the tool coordinate system are shown as T_E^{PC} and T_E^{NC}, respectively. The coordinate transformation matrix between these two camera coordinate systems T_{NC}^{PC} can be expressed as follows:

$$T_{NC}^{PC}=T_E^{PC}\left(T_E^{NC}\right)^{-1} \quad \text{(A7)}$$

Calibration of the Laser Line

A line laser can be calibrated based on the principle of triangular projection, as shown in Fig. A(C). The relative position between the laser generator and the active vision camera remained constant. The distance from the laser generator and the active vision camera to the weld bead was almost unchanged during welding. By capturing line laser images using active vision camera and assuming a height difference h between two captured images, with positions in Fig. A(C) denoted as O' and F, the angle between laser generator and camera optical axis as α, and the horizontal distance from laser generator to the camera optical center as $OE = x$, the similar triangle relationships are shown as follows:

$$\frac{O'F}{OO'}=\frac{BC}{d+h} \quad \text{(A8)}$$

$$\frac{BC}{x}=\frac{h}{d} \quad \text{(A9)}$$

Combining Equations A8 and A9 and substituting $x = tan\alpha \cdot d$ into the resulting expression, the relationship between h and $O'F$ is as follows:

$$h=\frac{O'F\cdot d}{O'O\cdot\tan\alpha-O'F}=\frac{\Delta v\cdot d}{k-\Delta v} \quad \text{(A10)}$$

where $O'F = \Delta v$, $k = O'O \cdot tan\alpha$. Since the value of k is only related to the intrinsic parameters and the position of active vision camera, it can be assumed that the value of k remains constant during the process of capturing images. The line laser length relationship between the global coordinate system and the active vision image coordinate system is $w = j \cdot \Delta u$, where w is the line laser length value in the global coordinate system and Δu is the line laser length value in the active vision image coordinate system.

SHAOJIE WU (*shaojie@tju.edu.cn*), **CHUANG LI**, **FANGJIE CHENG, ZHIJIANG WANG**, and **JUNFU JIAO** are with the School of Materials Science and Engineering, Tianjin University, Tianjin China. **WU**, **CHENG**, and **WANG** are also with Tianjin Key Laboratory of Advanced Joining Technology, Tianjin, China.