Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Implementation of panoramic image acquisition system based on stitching in microscopic imaging

Yuhang + Liu
Nanjing University Suzhou Campus
Yu Tao
Nanjing University of Aeronautics and Astronautics
Zhu Kai
Nanjing University Suzhou Campus
Ding Yao
ammdistin@nju.edu.cn

Nanjing University of Aeronautics and Astronautics

Article

Keywords: Visual Servoing, Micro Imaging, Image Stitching

Posted Date: August 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10508065/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Implementation of panoramic image acquisition system based on stitching in microscopic imaging

Liu Yuhang[1+]， Yu Tao[2, +],Zhu Kai[3], Ding Yao[2,3, *]

[1] School of Integrated circuits, Nanjing University Suzhou Campus, Suzhou, Jiangsu, 215163,P.R.China

2 College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu,210016, P.R.China

[3] School of Robotics and Automation, Nanjing University Suzhou Campus, Suzhou, Jiangsu, 215163, P.R.China

+ These two authors contributed equally to this work

* Corresponding author's e-mail:ammdistin@nju.edu.cn

Abstract:

Microscopic imaging has a wide range of applications in biology, medicine, and microscopic visual inspection. This imaging method has the characteristics of high resolution but small field of view. Traditional microscopy imaging is based on manual observation, where key information is manually photographed and stored. The preservation of image information for the entire scene is still in the stage of video preservation. How to effectively generate panoramic images is an urgent challenge that needs to be addressed. The purpose of this study is to use visual servoing technology to drive the movement of the pan tilt in microscopic imaging based on the prior features of pan tilt motion, and to achieve image feature matching, segmentation, and stitching, thereby obtaining panoramic information of the scene in microscopy. After testing, it was found that panoramic information with a resolution of 15488 * 15456 can be saved with a camera resolution of 640 * 480.

Keywords: Visual Servoing; Micro Imaging; Image Stitching;

Microscopic imaging and analysis play an important role in fields such as biology, medicine, and chips, and there are currently two common methods. One is an image acquisition system composed of a large field of view lens and a high-resolution camera. This structure has good imaging consistency, but the pixel density is proportional to the hardware cost. To achieve good imaging resolution, the hardware cost is relatively high; Another method is to complete the imaging of the target through image stitching. The resolution of this method is mainly affected by the lens. On the basis of a certain camera resolution, selecting a high magnification lens can achieve better imaging results. Then, single images are stitched together to obtain a complete field of view image. However, the stitching process is cumbersome, and if the coordination between motion and photography is not high, information may be lost or distorted.

This paper optimizes the hardware and software structure to achieve optimal hardware cost, builds a two-dimensional gimbal mobile test target, and selects a universal CMU device (640)* 480@30fps USB drive free industrial camera, paired with a 100x lens, completes scene image information acquisition, selects suitable ROI areas to extract image features, drives the gimbal movement to stop based on feature matching results, and completes image acquisition, storage, segmentation, and stitching, ultimately forming a panoramic image.

Compared to existing similar products, the structure involved in this project is simple and the cost is significantly reduced. At the same time, based on the camera calibration, the system displays the scale on the imaging results, providing reference for displaying the actual physical size of the

target. In addition, the system is designed with a manual mode, allowing users to manually select suitable areas for image acquisition. However, the system has the problem of not being able to automatically collect multiple samples simultaneously, and can only process one sample at a time.

Related works

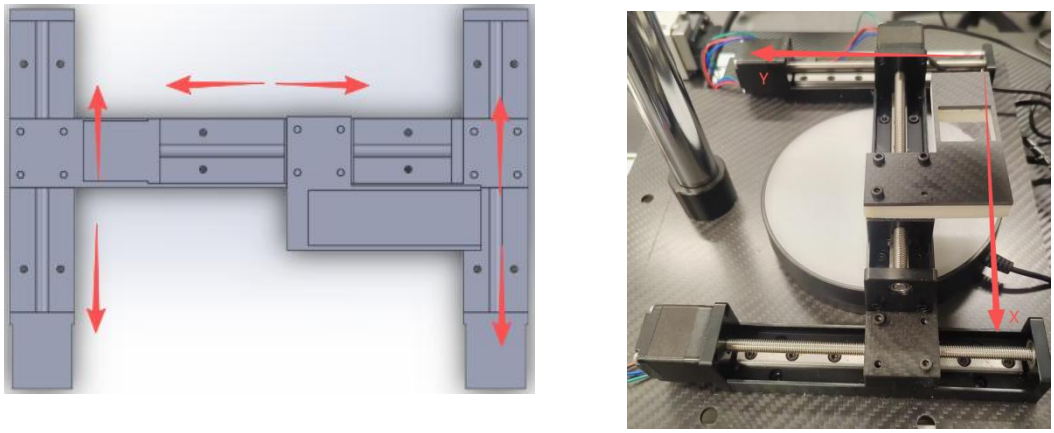
In the fields of biomedicine, material science and nanotechnology, microscopic imaging can carry out high-resolution observation of microstructure. Combined with digital image processing algorithms, it can further improve the contrast, sharpness and image reconstruction ability, and provide data support for image reconstruction in related fields. In electrocatalysis, due to the time-space resolution of traditional electrochemical imaging methods, it is impossible to effectively obtain the dynamic electrocatalytic current distribution. Ma et al[1]. used an interference electro-optic microscope to study the HER activity of double-layer MoS_2, observe the spatial evolution of HER current, reveal the sequential activation of HER sites along the trajectory aligned with the MoS_2 zigzag or handrail direction, facilitate researchers to intuitively observe the structural relationship of dynamic evolution on the nano electrochemical interface, which is of great significance for promoting the rational design of electrocatalysts. Benjamin J Vakoc et al[2]. observed tumor angiogenesis, lymphangiogenesis, tissue survival rate and vascular and cell responses to treatment using optical frequency domain imaging (OFDI) method and completely intrinsic contrast mechanism, demonstrating the potential of OFDI to promote the exploration of physiological and pathological processes and evaluate treatment strategies.Combining artificial intelligence with automatic control, microscopic imaging has realized data-driven optimization. Through adaptive adjustment of image sampling mode, and even real-time feedback control of acquisition process, the trade-off between imaging speed and resolution can be improved without significantly increasing hardware complexity[3,4].Hill and Park put forward the concept of visual servo for the first time[5], and used the visual data information obtained by the computer to guide and control the robot movement. On this basis, some scholars developed visual calibration servo and non visual calibration servo[6,7]. They judged the relative position between the camera and the target to be measured through the real-time characteristics of the image, and generated a motion strategy based on this to drive the movement between the two to achieve the desired relative position and pose.

In visual calibration servo, the large amount of image data makes the real-time performance of the system low. To solve this problem, Yan Y et al[8] combined breadth first search algorithm, Canny algorithm, Harris algorithm, and parallel processing strategy to design an image edge detection algorithm to optimize the image data, and established an eye in hand visual servo system. The results show that with the continuous increase of data volume, the performance of the proposed algorithm is still better than other image edge detection algorithms throughout the entire process, with accuracy, recall, and F1 score indicators exceeding 95%, 86%, and 90%, respectively. At an image size of 4096 * 2160, the computational efficiency of the research algorithm can reach 110FPS, and the average running time in the test dataset is at least 30.28ms.To improve the operational accuracy of industrial robots, Zhao, Z et al[9] has designed an intelligent 3D collaborative manufacturing platform for non integral mobile industrial robots based on improved binocular vision. The platform uses fuzzy function estimation and variational function recognition to identify object pose. The method has been tested to have better trajectory control effect and robustness, and can effectively control non integral mobile industrial robots.

Visual servoing is mainly divided into three methods: geometric feature-ba[12], and hybrid visual

servoing control (HVS)[13]. IBVS compares real-time measured image features with reference features and calculates feature errors to form closed-loop feedback. The advantage of this method is that it has strong robustness to camera model bias and requires less real-time computation compared to PBVS, but lacks depth information. Sharma et al[14]. improved the observability and controllability of the 2-DOF robotic arm using two different image parameters, particularly in adjusting the position and motion trajectory of the active camera. Moreno et al[15]. proposed a machine vision based system that controls a robotic arm by analyzing the operator's arm movements. Sekkat et al[16]. proposed a controller based on deep reinforcement learning (DRL) for determining the three-dimensional position of an object to optimize the control of a robotic arm. Oliva et al[17]. combined visual feedback with neural networks to enhance the learning ability of control strategies for 2-dOF planar robots. Sahu et al[19]. integrated key components such as object recognition and trajectory tracking, and used a deep learning based precision object detection framework to detect and locate targets in real time. They generated a control strategy to control the 2-dOF robotic arm, and verified the effectiveness of the control strategy through simulation and experimental verification.

Construction of Microscopic Imaging System



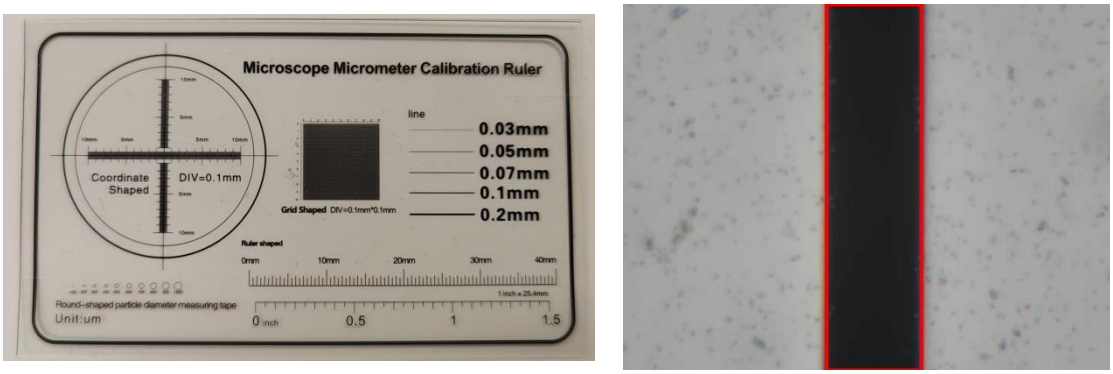
(a) the design of the 2-D PTZ (b) the actual object of the 2-D PTZ

Fig 1. microscopic imaging system

The microscopic imaging system includes camera, lens, bracket, and 2-D PTZ, the design of the 2-D PTZ is shown in Figure 1a, and the actual object is shown in Figure 1b. The PTZ consists of three slide stepper motors and matching drivers, using ATmega328P as the PTZ driver main control. ATmega328P communicates with the upper computer through a serial port, and there is a stage on the PTZ to place the template to be observed. Based on the scene information collected by the camera, the upper computer provides corresponding driving instructions to drive the PTZ to move along the X-axis or Y-axis in conjunction with image acquisition, achieving collaborative control of motion image acquisition.

To facilitate later use, we need to calibrate the imaging system before use. The first is coordinate system calibration. There are pixel coordinate system of the camera and motion coordinate system of the 2-D PTZ in the system. In order to achieve better imaging effect, it is hoped that the X axis and Y axis of the pixel coordinate system and the motion coordinate system of the 2-D PTZ are parallel and in the same direction. Since the sliding table is fixed, the calibration is completed by rotating the camera during calibration. When calibrating, first obtain the initial image, extract an

image feature at the upper left corner, record the coordinates of the current image feature pixel as (u_0, v_0), and then drive the slide motion along the *X* axis. When the feature moves to another edge of the image, perform feature matching. The coordinates after matching are (u_x, v_x). For a pair of images with a size of (m, n), if v_x-v_0<n * 1%, the calibration is considered to be completed when the *X* axis is matched. The *Y*-axis is registered in the same way, and the registration standard is u_y-u_0<m * 1%. In actual operation, because the *X* axis and *Y* axis are orthogonal, when the *X* axis registration is completed, the *Y* axis also meets the registration requirements.



(a) Calibration board (b) Calibration image

Fig. 2 Calibration of pixel and PTZ motion step pulse

The second is the calibration of pixel and PTZ motion step pulse, the relationship between motion pulse and pixel is determined through calibration. The motion of the PTZ is realized by pulse counting. In order to facilitate the later motion control, it is necessary to analyze the image changes generated by the PTZ motion process through images, so as to provide a basis for the generation of motion control strategies. As shown in Figure 2, place a calibration plate on the stage, use the point features on the calibration plate as the moving image features, record the pixel coordinates (u_0, v_0) of the current feature point, and then move the stage horizontally to the next position. At this moment, the feature point coordinates are (u_1, v_1), and the corresponding horizontal direction PTZ motion pulse is N_{xstep1}, so there is a relationship as shown in Equ. 1.

$$\begin{aligned} N_{xstep1} &= a(u_1 - u_0) + b(v_1 - v_0) \\ &= a\Delta u_1 + b\Delta v_1 \end{aligned} \tag{1}$$

From any two groups (N_{xstep1}, Δu, Δv), a and b can be calculated as Equ. 2.

$$\begin{aligned} a &= (n_1\Delta v_2 - n_2\Delta v_1) \Big/ (\Delta u_1\Delta v_2 - \Delta u_2\Delta v_1) \\ b &= (n_1\Delta u_2 - n_2\Delta u_1) \Big/ (\Delta u_2\Delta v_1 - \Delta u_1\Delta v_2) \end{aligned} \tag{2}$$

In actual calibration, we move several positions respectively, record the corresponding pulse number N_{xstep} and pixel coordinates (u_n, v_n), and then combine them in pairs to calculate a group (a, b). Take the arithmetic mean of the calculated groups of (a, b) data as the calibration coefficient.



Fig. 3 Calibration of pixel and physical size (Binary image)

The third calibration is the calibration of pixels and physical dimensions. In order to facilitate the user to analyze the scene image in the later stage, when the camera lens and other hardware platforms are determined, the pixel and physical dimensions can be calibrated in advance, stored and drawn on the saved image. When the user analyzes the scene image in the later stage, the physical dimensions of the target can be obtained through pixels. The specific calibration method is to collect the image on the calibration plate, as shown in Figure 3, then binarize it, use image segmentation to segment the scale bar, respectively calculate the number of pixels in the connected area of each row of pixels, and then average it to get the ratio of pixels to physical size.

In the calibration process, the calibration of pixel and PTZ motion step pulse and the calibration of pixel and physical size are integrated. Measure the horizontal width and vertical width of 0.03mm, 0.05mm, 0.07mm, 0.1mm and 0.2mm calibration respectively to determine the corresponding relationship between the physical size and the pixel x-y in the pixel coordinate system, the number of step pulses required during the movement of the PTZ and the pixel coordinates of the upper left corner of the calibration line. The calibration process is shown in the figure3. Taking the 0.2mm measurement as an example, measure and calculate the position coordinates of the upper left corner and the width of the calibration line when the calibration line is placed horizontally and vertically, and record the number of pulses required before and after the movement, and the results are listed in Table 1.

Table 1 Pixel measurement results

	mid_h Width /Pixel	*up_h width /Pixel*	*down_h Width /Pixel*	*h_avg_width (pixel)*	*h_Pixel/Size (Pixel/mm)*
0.03mm	13.32	13.32	13.12	13.25	441.76
0.05mm	28.30	28.44	28.26	28.33	566.70
0.07mm	40.53	40.44	40.44	40.46	578.13
0.1mm	60.85	61.63	61.54	61.34	613.41
0.2mm	121.25	122.09	122.11	121.81	609.09

	mid_v Width /Pixel	*left_v width /Pixel*	*right_v Width /Pixel*	*v_avg_width (pixel)*	*v_Pixel/Size (Pixel/mm)*
0.03mm	14.83	15.86	14.14	14.94	498.22
0.05mm	27.17	28.21	28.33	27.91	558.22
0.07mm	38.51	39.58	39.88	39.32	561.84
0.1mm	56.85	59.26	58.84	58.32	583.21
0.2mm	113.46	116.00	116.16	115.20	576.04

In table 1,*h_width* represents the width of the calibration line in the horizontal direction, and *v_width* represents the width of the calibration line in the vertical direction, *mid up down left mid right* indicates the position of the calibration line, *h_avg_width* is the average pixel width of the calibration line in the horizontal direction, and *v_avg_width* is the average pixel width of the calibration line in the vertical direction. The columns *h_pixel/size* and *v_pixel/size* are the relationship between pixels and physical dimensions calculated according to the average pixel width and the physical dimensions of the calibration line. The unit is pixels/mm, which represents the number of pixels occupied per mm.

According to the values of *h_pixel/size* and *v_pixel/size* in Table 1, the corresponding relationship between pixels and physical dimensions is 561.83 pixels/mm for the horizontal calibration line, and 555.51 pixels/mm for the vertical calibration line.

At the same time, record the coordinates of the upper left corner of the calibration line before and after moving and the number of pulses required for moving in the relationship between the calibration pixel and the physical size, as shown in Table 2,and the number of pulses required for calibration line movement is shown in Table 3.

Table 2 Coordinate of anchor point of calibration line

	mid_h	***up_h***	***down_h***	***mid_v***	***left_v***	***right_v***
0.03mm	(0,194)	(0,14)	(0,277)	(287,0)	(49,0)	(608,0)
0.05mm	(0,212)	(0,47)	(0,401)	(298,0)	(17,0)	(589,0)
0.07mm	(0,215)	(0,46)	(0,377)	(300,0)	(54,0)	(533,0)
0.1mm	(0,173)	(0,50)	(0,373)	(271,0)	(27,0)	(547,0)
0.2mm	(0,155)	(0,20)	(0,325)	(217,0)	(15,0)	(495,0)

In table 2, *up* indicates that the calibration line is at the top of the image, *mid* indicates that the calibration line is at the middle of the image, *dow*n indicates that the calibration line is at the bottom of the image, *h* indicates that the calibration line is placed horizontally in the image, and *v* indicates that the calibration line is placed vertically in the image.

Table 3 List of calibration line movement pulse records

Movement	***Number of pixels to move***	***Number of mobile pulses***	***Pulse/Pixel***
0.03down_up_h	263	1300	4.94
0.03mid_up_h	180	800	4.44
0.05down_up_h	354	1800	5.08
0.05mid_up_h	165	800	4.84
0.07down_mid_h	162	800	4.93
0.07mid_up_h	169	800	4.73
0.1down_mid_h	200	1000	5
0.1mid_up_h	123	600	4.87
0.2down_mid_h	170	900	5.29

0.2mid_up_h	135	700	5.18
0.03left_mid_v	238	600	2.52
0.03mid_right_v	321	800	2.49
0.05left_mid_v	281	700	2.49
0.05mid_right_v	291	700	2.40
0.07left_mid_v	246	600	2.43
0.07mid_right_v	233	600	2.57
0.1left_mid_v	244	600	2.45
0.1mid_right_v	276	700	2.53
0.2left_mid_v	202	500	2.47
0.2mid_right_v	278	700	2.51

The meaning of *movement* in Table 3 is shown in Figure 4.

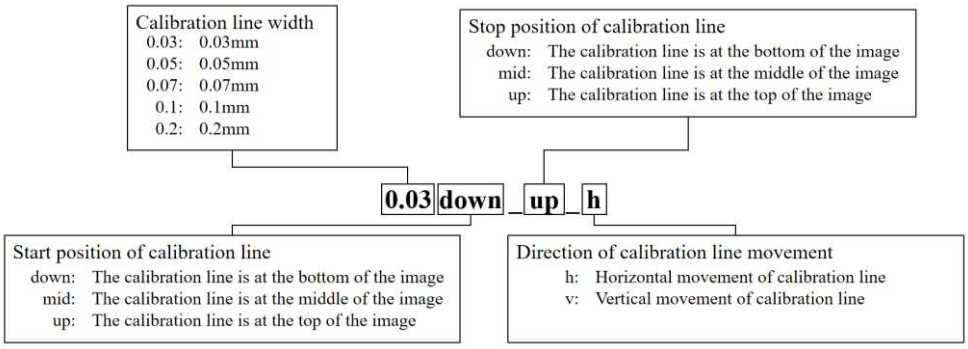
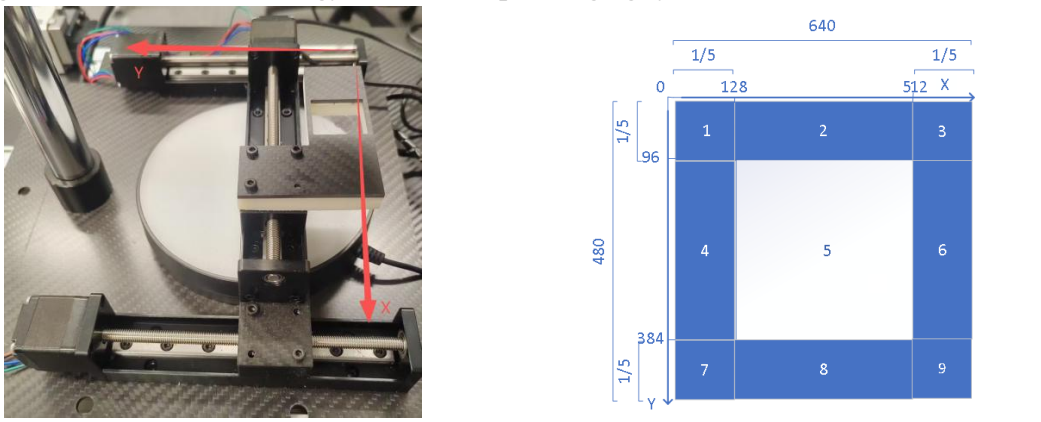


Fig.4 The meaning of *movement*

According to the data calculation in Table 3, for the calibration line along the horizontal direction, the corresponding relationship between pixels and pulses is 4.934999186 pulses moving one pixel; For the calibration line along the vertical direction, the corresponding relationship between pixels and pulses is 2.491243483 pulses moving one pixel. From this, it can be calculated that if the camera's viewing angle moves horizontally, the left 1/5 area will move to the right 1/5 image area that needs to move 4/5, 512 pixels will need to be moved, and approximately 1895 pulses will be needed; When the camera angle is vertically moved, moving the upper 1/5 area to the lower 1/5 requires 384 pixels, or about 1276 pulses.

Design of motion control strategy for microscopic imaging system



(a) carrier platform (b) Schematic diagram of ROI area for single frame image acquisition

Fig. 4 Load Platform Design

To facilitate microscopic imaging, a sample platform was constructed as shown in Figure 4. The target under test was placed on the platform. Considering the limited field of view, a black matte cover glass was added to the platform to obscure areas outside the target, ensuring uniform imaging results around the target and aiding visual localization of the test area.

When collecting images of this scene, there will be obvious boundary features. As shown in Figure 4b, 1/5 of the boundary of a single frame image is selected as the ROI area. By analyzing the ROI area, the positional relationship between the camera and the object under test is determined, and a driving strategy is generated based on the positional relationship.

A) Position determination and motion strategy generation

Transfer the color image to HSV space and use the image information from *H* or *S* channels. For grayscale images, calculate directly by selecting the upper 1/5, lower 1/5, left 1/5, and right 1/5 of the image and calculating the pixel average values top_{mean}, $bottom_{mean}$, $left_{mean}$, and $right_{mean}$ in the above areas. Select a threshold thd_{mean}. When $top_{mean}<thd_{mean}$ and $left_{mean}<thd_{mean}$, it means that the current camera is located in the upper left corner of the target to be measured, and the relative position is denoted as pos_1. When only $top_{mean}<thd_{mean}$ is present, it means that the current camera is located at the upper edge of the target to be measured, and the relative position is denoted as pos_2. When $top_{mean}<thd_{mean}$ and $right_{mean}<thd_{mean}$, it indicates that the current camera is located in the upper right corner of the target being measured, and the relative position is denoted as pos_3. When $right_{mean}<thd_{mean}$, the current camera is located at the right edge of the target being tested, the relative position is denoted as pos_6. When the top_{mean}, $bottom_{mean}$, $left_{mean}$, and $right_{mean}$ are all greater than the thd_{mean}, the current camera is located in the middle area of the target being tested, and the relative position is denoted as pos_5. When only the $left_{mean}$ is less than the thd_{mean}, the current camera is located at the left edge of the target being tested, and the relative position is denoted as pos_4. When the $bottom_{mean}$ and the $left_{mean}$ are less than the thd_{mean}, the current camera is located at the bottom left corner of the target being tested, and the relative position is denoted as pos_7. When only the $bottom_{mean}$ is less than the thd_{mean}, the current camera is located at the bottom of the target being tested, and the relative position is denoted as pos_8. When $bottom_{mean}<thd_{mean}$ and $right_{mean}<thd_{mean}$, it indicates that the current camera is located in the lower right corner of the target to be tested, and the relative position is denoted as pos_9. When the top_{mean}, $bottom_{mean}$, $left_{mean}$, and $right_{mean}$ are all less than the thd_{mean}, it indicates that the current camera does not intersect with the target being measured and no image information of the target has been collected. At this time, the relative position is recorded as pos_0.

In the coordinate system shown in Fig.4b, when located at pos_1, the PTZ needs to move in the negative *X-axis* direction to facilitate the camera's capture of the image on the right side of the current position. When located at pos_2, the PTZ needs to continue moving in the negative *X-axis* direction to facilitate the camera's capture of the image on the right side of the current position.When located at pos_3, the PTZ needs to move in the negative *y-axis* direction to facilitate the camera's capture of images from the lower side of the current position.When located at pos_6, the PTZ needs to move in the positive *X-axis* direction to facilitate the camera's capture of the image on the left side of the current position. When located at pos_5, the PTZ needs to continue the previous motion strategy and move in the negative or positive direction along the *X-axis*, making it easier for the camera to capture images of the next position. When located at pos_4, the PTZ needs to move in the negative *Y-axis* direction to facilitate the camera to capture images of the lower side of the current position. There are two situations when the camera is at the bottom of the

target being tested, one is that when located at pos_7, the PTZ needs to move along the negative *X-axis* direction to facilitate the camera to capture the image on the right side of the current position, when located at pos_8, the PTZ needs to move along the negative *X-axis* direction to facilitate the camera to capture the image on the right side of the current position, when located at pos_9, it indicates that the movement has reached the end of the test area and the image acquisition is complete. Another scenario is that when located at pos_9, the PTZ needs to move along the positive *X-axis* direction to facilitate the camera's capture of the image on the left side of the current position, when located at pos_8, the PTZ needs to move along the positive *X-axis* direction to facilitate the camera's capture of the image on the left side of the current position, when located at pos_7, it indicates that the movement has reached the end of the test area and the image acquisition is complete.

B) Image feature extraction, matching, and image stitching design based on location judgment

Feature extraction and matching of images are key technologies in visual servoing,J. Gao et al[19] describes a method to construct seamless image，approach blends the homographies in the alignment procedure to perform a nonlinear warping, once the images are geometrically stitched, they are further processed to blend seams and reduce curvilinear visual artifacts due to the nonlinear warping. Liao T et al[20]propose a novel image stitching method using multi-homography warping guided by image segmentation, for each segmented content in the overlapping region, select its best-fitting homography with the lowest photometric error, for each segmented content in the non-overlapping region, calculate a weighted combination of the linearized homographies, finally, the target image is warped via the best-fitting homographies to align with the reference image, and the final panorama is generated via linear blending. Comprehensive experimental results on the public datasets demonstrate that our method provides the best alignment accuracy by a large margin, compared with the state-of-the-art methods.In order to solve the problem of increasing difficulty in image feature extraction caused by lighting changes, noise interference and geometric distortion, Liu, Y[21] et al introduced StitchEval's comprehensive benchmark test framework from four aspects of structural similarity (SSIM), mean square error (MSE), peak signal to noise ratio (PSNR) and human rating subjective score (SS), and systematically analyzed the robustness of different stitching algorithms by applying light transformation, noise interference and geometric changes to the original dataset, providing valuable reference for optimizing future stitching methods and improving the adaptability of algorithms in real world scenes.

In a microscopic imaging system, the target plane to be measured is parallel to the camera plane and moves horizontally along the *X axis* or *Y axis*. During the movement, the relative distance between the two planes does not change. Therefore, there is no rotation or scaling change in the image features during relative motion, which greatly reduces the difficulty of image feature matching. Therefore, template matching is used in the system to achieve feature extraction and matching of images.

When the camera and the target being tested are in different relative positions, there will be different image feature extraction and matching strategies.When located at pos_1, select the 1/5 area on the right side of the current image (img_{start}) as the ROI, obtain a single channel image of this area, and use the Canny operator to calculate the image edge information. Calculate the bounding rectangle with the largest connected edge area, and use the single channel image framed by this bounding rectangle as the matching template for matching in the subsequent process. The ROI to

be matched in the image is the left 1/5 area of the camera. When the generated template is used to complete the matching in the left 1/5 area, stop the pan tilt motion and save the current image as img_{end}. Due to the absence of rotational scaling during the image acquisition process, there is no need to correct the pre stitched image through affine transformation.In this system, img_{star} and img_{end} have the same size (m,n). In img_{star}, the pixel coordinates of the upper left corner of the obtained feature template are (x_0, y_0), then crop img_{star} $[0{:}x0, 0{:}n]$ as the image to be stitched $img_{stitch1}$. In img_{end}, the coordinates of the upper left corner of the matching area are (x_1, y_1), at this time, the cropped img_{end} $[x_1{:}m, 0{:}n]$ is the image to be stitched $img_{stitch2}$.The stitching process is based on establishing an empty image matrix, first placing $img_{stitch1}$ at position $(0,0)$, and then placing$img_{stitch2}$ at position $(0, x_0+1)$ for stacked stitching. When the PTZ drives the template to be photographed to move to the left, that is, at pos_1, pos_2, pos_5, and pos_8, this image feature extraction matching and stitching method can be used.

When located at pos_3, select the ROI area on the lower 1/5 side of the current image(img_{star}), obtain a single channel image of this area, and use the Canny operator to calculate the image edge information. Calculate the bounding rectangle with the largest connected edge area, and use the single channel image framed by this bounding rectangle as the matching template for matching in the subsequent process. The ROI area to be matched in the image is the top 1/5 area of the camera. When the generated template is used to complete the matching in the top 1/5 area, stop the pan tilt motion and save the current image as img_{end}. In img_{star}, if the pixel coordinates in the upper left corner of the obtained feature template are (x_0, y_0), then crop img_{star} $[0{:}m, 0{:}y_0]$ as the image to be stitched $img_{stitch1}$. In img_{end}, the coordinates of the upper left corner of the matching area are (x_1, y_1). At this time, img_{end} $[0{:}m, y_1{:}n]$ is cropped as the image to be spliced $img_{stitch2}$. The splicing adopts a stacking mode, that is, img_{stitch1} is stacked below img_{stitch2}.When the gimbal drives the target to move upwards, that is, at pos_3, pos_4, and pos_6, this image feature extraction and matching method can be used.However, in practical applications, due to the need to horizontally stitch multiple low resolution images into one high-resolution image, for ease of operation, the above improvement is to record coordinate points (x_1, y_1) after matching, use reference points (x_1, y_1) to uniformly crop the high-resolution horizontally stitched image, and then stack and stitch it with the previous image.

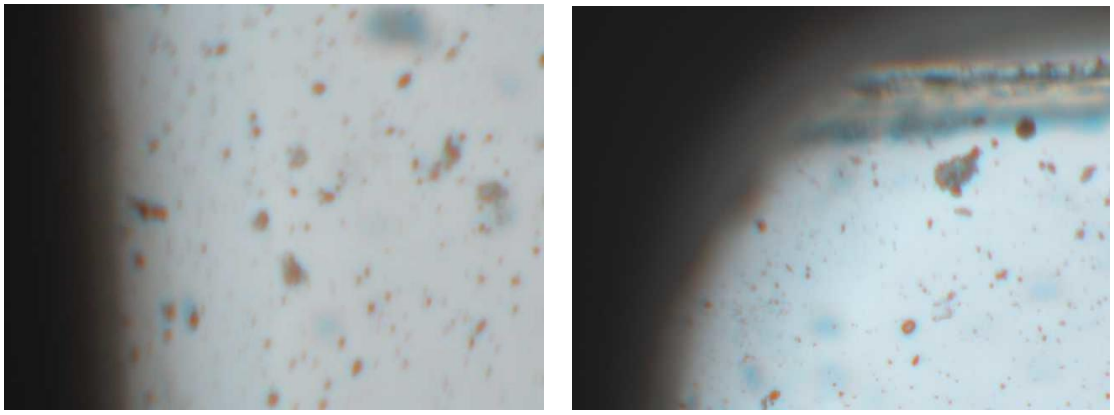
When located at pos_4 or pos_6, it means that the gimbal needs to move downwards with the target to be tested first, and then move laterally. Lateral motion is divided into left translation and right translation. The strategy of left translation has been described in the previous text, while right translation and left translation are similar, except that the ROI area is swapped. The extraction, matching, and stitching strategies of image features remain unchanged.

When the PTZ drives the target to move upwards, there are two situations when approaching the bottom. Firstly, matching is achieved in the ROI area at the top 1/5, and in this case, the matching is still carried out according to the previous strategy. Secondly, when the average gray value of the ROI area at the bottom 1/5 is less than the threshold, it is judged that the relative position between the camera and the target is pos_7 or pos_9. At this time, the PTZ stops moving upwards, and then image feature matching is performed in the current field of view. The anchor coordinates of the upper left corner of the template matching are (x_1, y_1), and based on this, horizontal cutting is performed for later image stitching.

Image stitching experiment

A) Position determination

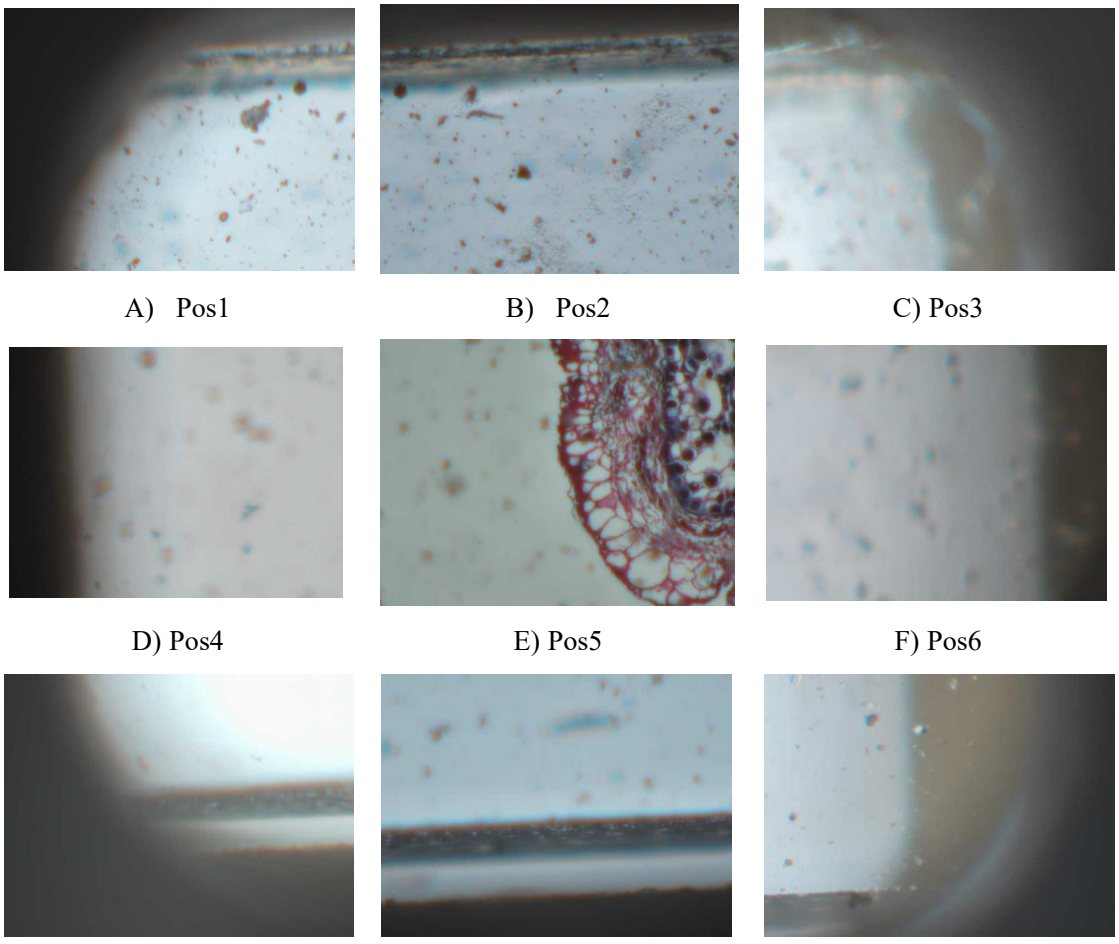
The relative position determination between the camera and the target being tested is the key to driving the movement of the pan tilt zoom. To facilitate labeling, we encode different positions separately. When the image information meets the criteria for position determination, the corresponding position encoding is given through the serial port.



(a) Random starting position (b) Automatic photography initialization position

Fig. 5 System initialization scenario

Firstly, manually adjust the gimbal to the position shown in Figure 5a, and then start the program. The system will automatically move to the starting position shown in Figure 5b.



A) Pos1 B) Pos2 C) Pos3

D) Pos4 E) Pos5 F) Pos6

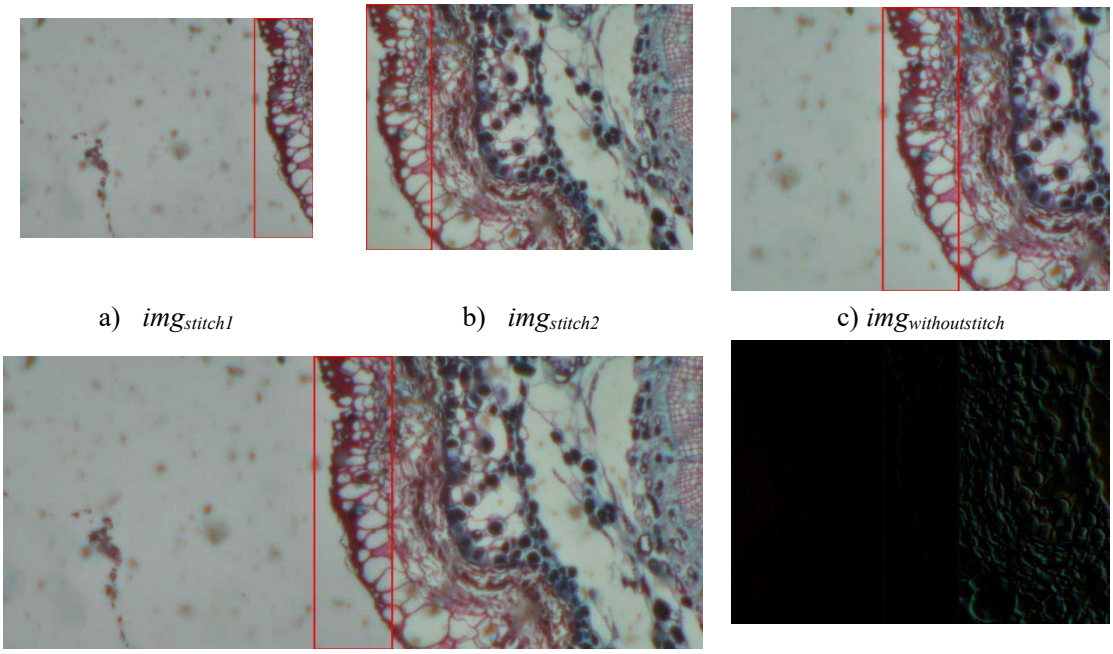
G) Pos7 H) Pos8 I) Pos9

Fig. 6 Location Diagram

For the convenience of observation, mark on the displayed image when the motion reaches the corresponding position. Each position is marked as shown in Figure 6.

B) Two consecutive image stitching tests

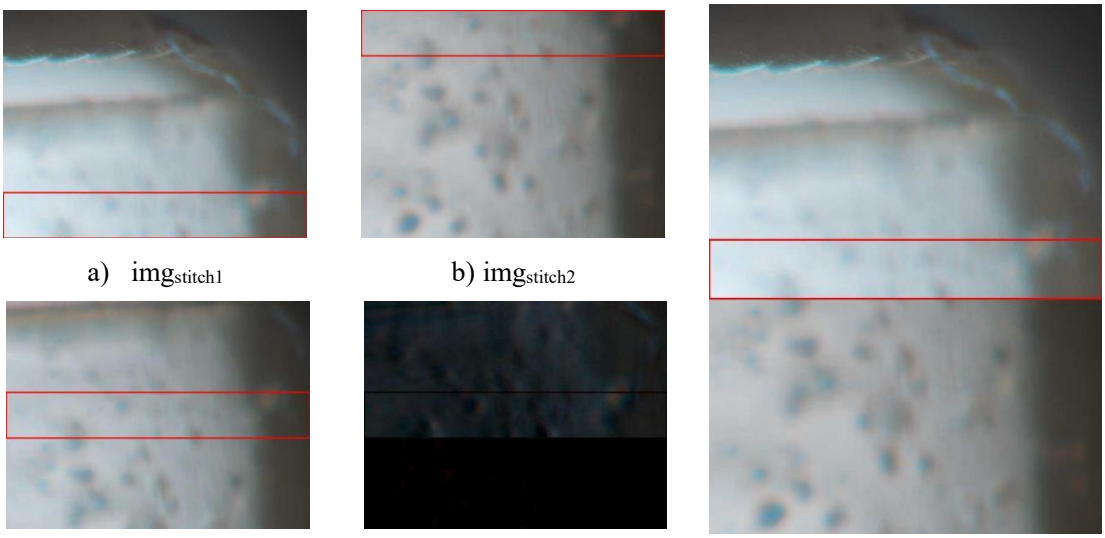
Image stitching is the foundation for constructing high-resolution images from low resolution images, and an important indicator to measure the stitching effect is whether information is lost or mutated. Due to the stacking method used for stitching, there is no information variation caused by image processing techniques such as affine transformation. The main focus is on observing whether there is information loss during image stitching. The comparison of the stitched images is shown in Figure 7 and Figure 8.



a) $img_{stitch1}$ b) $img_{stitch2}$ c) $img_{withoutstitch}$

d) $img_{withstitch}$ e) comparison results

Fig. 7 Horizontal image stitching



a) $img_{stitch1}$ b) $img_{stitch2}$

c) $img_{withoutstitch}$ d) comparison results e) $img_{withstitch}$

Fig. 8 Vertical image stitching

C) Panoramic high-resolution imaging results

The result of combining multiple low resolution images in the same row into a high-resolution

image is shown in Figure 9. Then, this high-resolution, high aspect ratio image is combined into a high-resolution panoramic image, and the final effect is shown in Figure 10.



Fig. 9 Single line high-resolution imaging effect

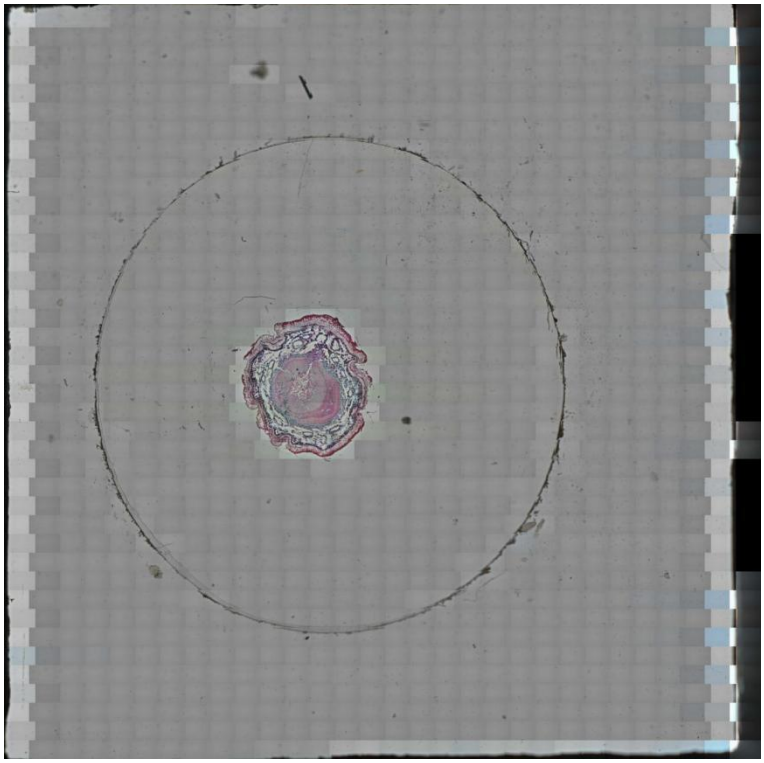


Fig. 10 Full scene high-resolution imaging effect

From Figure 10, it can be seen that this stitching method exhibits a noticeable mosaic effect when generating panoramic photos, while the jagged edges of the scene are more pronounced. In response to the above situation, we further optimized the imaging method.

Conclusion

As shown in Figure 10, the system can effectively identify sample features and control the movement of a 2D pan-tilt unit via visual servoing, accurately cutting and stitching sample images to form panoramic images. A comparison with Figure 8 reveals no loss of image information during the stitching process, demonstrating the system's capability to capture panoramic image data of objects through stitching. This approach, which precisely controls the 2D pan-tilt unit movement and image acquisition via IBVS, aims to automate microscopic imaging of biological cell sections, reduce manual operations, and enhance the efficiency of microscopic panoramic imaging.

Data availability

The datasets generated and analysed during the current study are not publicly available due to certain reasons within the organization but are available from the corresponding author on reasonable request.

References

[1] Ma, K., Cui, Y., Ren, Y. et al. Imaging dynamic electrocatalytic processes on nano-strained MoS2 using interferometric electro-optical microscopy. Nat Energy (2026).

https://doi.org/10.1038/s41560-026-02043-4

[2] Vakoc, B., Lanning, R., Tyrrell, J. et al. Three-dimensional microscopy of the tumor microenvironment in vivo using optical frequency domain imaging. Nat Med 15, 1219–1223 (2009). https://doi.org/10.1038/nm.1971

[3] Passmore J B, Rates A, Schröder J, et al. Closed-loop optogenetic control of cell biology enables outcome-driven microscopy[J]. Nature Communications, 2026, 17: 1087.

[4] Sinha A, Lee J, Li S, et al. Lensless computational imaging through deep learning[J]. Optica, 2017, 4(9): 1117-1125.

[5] Current Solutions and Future Trends for Robotic Prosthetic Hands[J]. Mendez Vincent;Iberite Francesco;Shokur Solaiman;Micera Silvestro.Annual Review of Control, Robotics, and Autonomous Systems.2021

[6] Lee D W, Kim S H, Won S C. Visual Servoing Method Using Camera Self-calibration[C]//2006 SICE-ICASE International Joint Conference. IEEE, 2006: 5006-5010.

[7] Gong Z, Tao B, Yang H, et al. An uncalibrated visual servo method based on projective homography[J]. IEEE Transactions on Automation Science and Engineering, 2017, 15(2): 806-817.

[8] Yan, Y., Mokri, S.S. Performance analysis of robotic arm visual servo system based on BFS-canny image edge detection algorithm. Sci Rep 15, 35355 (2025). https://doi.org/10.1038/s41598-025-19221-1

[9] Zhao, Z., Li, W., Li, Y. et al. Embedding high-resolution touch across robotic hands enables adaptive human-like grasping. Nat Mach Intell 7, 889–900 (2025). https://doi.org/10.1038/s42256-025-01053-3

[10] Chu, H. K., Mills, J. K. & Cleghorn, W. L. Image-based visual servoing through micropart reflection for the microassembly process. J. Micromech. Microeng. 21, 065016. https://doi.org/10.1088/0960-1317/21/6/065016 (2011).

[11] Sahu, U. K., Patra, D. & Subudhi, B. Adaptive intelligent vision-based control of a flexible-link manipulator. Electr. Eng. 105, 3263–3281. https://doi.org/10.1007/s00202-023-01875-7 (2023).

[12] Nazari, A. A., Zareinia, K. & Janabi-Sharifi, F. Visual servoing of continuum robots: Methods, challenges, and prospects. Int. J. Med. Robot. Comput. Assist. Surg. 18, 1–26. https://doi.org/10.1002/rcs.2384 (2022).

[13] Wang, Y., Lang, H. & De Silva, C. W. A hybrid visual servo controller for robust grasping by wheeled mobile robots. IEEE/ASME Trans. Mechatron. 15, 757–769. https://doi.org/10.1109/TMECH.2009.2034740 (2010).

[14] Sharma, R. & Hutchineon, S. The observability of robot motion under active camera control. In Proceedings—IEEE International Conference on Robotics and Automation, 162–167 (IEEE, San Diego, CA, USA, 1994). https://doi.org/10.1109/ROBOT.1994.350994.

[15] Moreno, R. J. Tracking of human operator arms oriented to the control of two robotic arms. In Proceedings of 19th Symposium Image, Signal Process and Artificial Vision, STSIVA 2014, 1–4 (IEEE, Armenia, Colombia, 2015). https://doi.org/10.1109/STSIVA.2014.7010125.

[16] Sekkat, H., Tigani, S., Saadane, R. & Chehri, A. Vision-based robotic arm control algorithm using deep reinforcement learning for autonomous objects grasping. Appl. Sci. 11, 1–14. https://doi.org/10.3390/app11177917 (2021).

[17] Oliva, M., Banik, S., Josifovski, J. & Knoll, A. Graph neural networks for relational inductive bias in vision-based deep reinforcement learning of robot control. In Proceedings of the International Joint Conference on Neural Networks, Vol. 2022-July, 1–9 (IEEE, Padua, Italy, 2022). https://doi.org/10.1109/IJCNN55064.2022.9892101. arXiv:2203.05985.

[18] Sahu, U.K., K. S., M., K., A. et al. Autonomous object tracking with vision based control using a 2DOF robotic arm. Sci Rep 15, 13404 (2025). https://doi.org/10.1038/s41598-025-97930-3

[19] J. Gao, S. J. Kim and M. S. Brown, "Constructing image panoramas using dual-homography warping," CVPR 2011, Colorado Springs, CO, USA, 2011, pp. 49-56, doi: 10.1109/CVPR.2011.5995433.

[20] Liao, T., Wang, C., Li, L., Liu, G. & Li, N. Parallax-tolerant image stitching via segmentation-guided multi-homography warping. Signal Process. 230, 109860, DOI: https://doi.org/10.1016/j.sigpro.2024.109860 (2025).

[21] Liu, Y., Wang, Y., Wang, D. et al. A scalable benchmark to evaluate the robustness of image stitching under simulated distortions. Sci Rep 15, 32816 (2025). https://doi.org/10.1038/s41598-025-17730-7