This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2967956, IEEE Transactions on Instrumentation and Measurement

An Automatic Analog Instrument Reading System Using Computer Vision and Inspection Robot

Jian Huang[1], *Senior Member, IEEE,* Junzhe Wang[1], *Student Member, IEEE,* Yihua Tan[1,*], *Member, IEEE,* Dongrui Wu[1], *Senior Member, IEEE,* and Yu Cao[1], *Student Member, IEEE.*

***Abstract*—Since manual inspection of analog instruments is inefficient, many computer vision-based automatic reading systems have been proposed recently. However, most of them use fixed cameras, which are costly due to the large number of used cameras. Although some other systems adopting the pan-tilt-zoom camera and the movable inspection robot can avoid using plenty of cameras, they have to overcome high computational cost in aligning the camera to the tested instrument. Meanwhile, most existing systems are instrument type dependent, and hence cannot handle multiple types of instruments simultaneously. In this paper, first, based on an inspection robot, an automatic reading system equipped with a pan-tilt-zoom camera is designed for different types of round-shape analog instruments. Then, a fast camera alignment algorithm based on visual servo is proposed, in which YOLOv3 is applied and improved to locate the instrument, and guide the camera to iteratively align to the instrument. Finally, a monocular-vision pointer reconstruction algorithm is proposed to accurately read the instrument. Experimental results demonstrated that our proposed system is fast and reliable in the camera alignment process, and is effective in reading different types of analog instruments during the robot-based inspection.**

***Index Terms*—Automatic Reading System, Analog Instrument, Robot-based Inspection, Fast Camera Alignment, Monocular-vision Pointer Reconstruction.**

I. INTRODUCTION

ANALOG pointer instruments of devices are widely used in monitoring the working status of low-voltage substations, water plants, chemical plants, *etc.* They are generally installed at different locations, at height between 0.5m to 1.5m. Manual inspection is the prevalent way to read these instruments nowadays [1]. It is inefficient, costly, and error-prone. So, the vision-based automatic analog instrument reading system, which is more efficient, cheaper, and more reliable, has been gaining the interest of researchers. Many such systems have been proposed recently [2]–[16].

Most earlier systems work in condition that the camera is fixed in front of the tested instruments [2]–[11]. Belan *et al.* proposed a monocular system, in which radial sampling projection [17], [18] and Bresenham line drawing [19], [20] were used to locate the instrument pointer. Jaffery *et al.* [8] proposed a monocular system, in which a Dynamic Sliding Window Algorithm (DSWA) was used to locate the instrument pointer. Zheng *et al.* [9] also proposed a monocular system, in which a novel Multi-Scale Retinex with Color Restoration (MSRCR) algorithm was used to overcome light variations in analog instrument reading. And Yang *et al.* [10] proposed a binocular system using Tsai's calibration [21] and Scale-Invariant Feature Transform (SIFT) [22], in which the information of two camera views was used to restore the pointer information of the real world. However, these automatic reading systems would be costly if there were a large number of analog instruments to be read, because a set of system can only read one instrument.

With the developments of robot technology [23], [24] and Simultaneous Localization and Mapping (SLAM) technology [1], the above issue can be solved by some other systems [12]–[16] using the PTZ camera and combining with the movable inspection robot. Powered by SLAM, an inspection robot can navigate to specified locations so that those systems can read all the analog instruments during the robot-based inspection. However, due to the localization errors of SLAM, the PTZ cameras of those systems need to be aligned with the analog instrument before the automatic reading. Fang *et al.* [12] proposed a visual servo system using SIFT [22] or Speeded-Up Robust Feature (SURF) [25], in which the posture of the PTZ camera was adjusted by repeatedly calculating the horizontal and vertical camera deviation angles between the image center and the instrument center and iteratively controlling the PTZ motors. Similar systems and approaches were also proposed by Li *et al.* [13], Mai *et al.* [14], and Liu *et al.* [15]. Inspired by a face detection work [26], Song *et al.* [16] proposed a visual servo system using Adaboost and Haar-like features, in which the camera deviation angles were first calculated from the result of Adaboost instrument detection and were then corrected by iterative controls of the PTZ motors. However, there are still some challenges in the camera alignment process to be overcome, *e.g.*, SIFT and SURF are time-consuming [12]–[15], and the performance of Adaboost instrument detector is not always reliable [16].

Furthermore, nearly all existing automatic reading systems (both those using fixed cameras and those combining with the inspection robot) only considered one type of analog instruments [2]–[16]. They cannot adequately handle the situation where there are different types of analog instruments in low-voltage substations, water plants, *etc.*

In this paper, an automatic analog instrument reading system using computer vision and inspection robot is constructed. This system is designed for different types of round-shape

*Corresponding Author, e-mail yhtan@hust.edu.cn. This work was supported by the International Science and Technology Cooperation Program of China (No. 2017YFE0128300), the Fundamental Research Funds for the Central Universities (HUST: 2019kfyRCPY014), and the Research Fund of PLA of China (BWS17J024).

[1]Key Laboratory of Ministry of Education for Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, Hubei, 430074 P.R. China.

analog instruments. Then, a Fast Camera Alignment Algorithm (FCAA) based on space information and visual servo instrument detection is proposed in our system. Benefiting from the proposed YOLOv3-tiny-dw detector, which improves the state-of-the-art YOLOv3 [27] by combining it with depthwise separable convolutions in MobileNet [28], the FCAA performs fast and reliable camera alignment. Finally, a Monocular-vision Pointer Reconstruction Algorithm (MPRA) based on SURF [25] and Random Sample Consensus (RANSAC) [29] is proposed in our system to accurately read the readings of different types of round-shape analog instruments.

Our main contributions are:

1) We provide a scheme of building an automatic analog instrument reading system using computer vision and inspection robot.
2) We propose FCAA based on the improved YOLOv3-tiny-dw detector, which performs fast and precise camera alignment.
3) We propose MPRA, which is effective in reading different types of round-shape analog instruments.
4) We validate the above system and algorithms on a real inspection robot platform. In addition, we confirm the robustness to illumination changes and different instrument heights of this system.

The remainder of this paper is organized as follows. Sections II∼IV describe the details of our automatic reading system, the FCAA, and the MPRA, respectively. Section V presents the experimental results of the FCAA and the MPRA. Finally, Section VI draws conclusions.

II. Design of the Automatic Analog Instrument Reading System

Our automatic analog instrument reading system is designed on a SLAM inspection robot introduced in our previous work [3]. Independently, this robot based on laser rangefinder theory can draw a 2-D obstacle map relevant to the horizontal plane during its movement. Thus, its location and posture can be estimated by applying a Monte Carlo method [30] to matching the present laser data and the obstacle map during real-time inspections. Additionally, with the motion control, the robot can navigate to specified locations of the space.

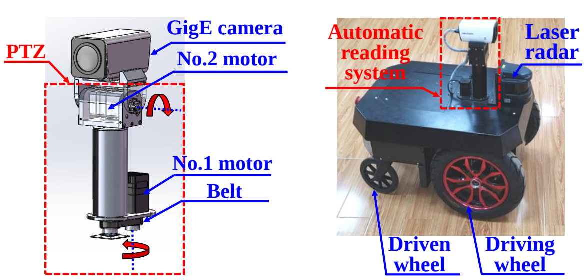


Fig. 1. The appearance of our automatic reading system. (a) The 3-D appearance and the constitution of our automatic reading system. (b) The inspection robot equipped with our automatic reading system.

The 3-D appearance model of our automatic analog instrument reading system is shown in Fig. 1(a). According to this appearance model, our automatic reading system mainly constitutes of a gigabit ethernet (GigE) camera which supports image acquisition and optical zoom, a two-degrees-of-freedom PTZ which supports horizontal and vertical posture adjustments, and other mechanical components which are used to attach the above two parts. In this system, the GigE camera is driven by a 2.50GHz Nvidia® Denver™ CPU. Meanwhile, the PTZ is driven by a 168MHz ARM® Cortex™ – M4 MCU. The mentioned CPU and MCU are both equipped on the robot we used, while they were originally the controllers of robot navigation and robot movement respectively. The inspection robot equipped with this automatic reading system is shown in Fig. 1(b).

As the two-degrees-of-freedom PTZ shown in Fig. 1(a), its postures in the horizontal and vertical Degrees Of Freedom (DOF) can be respectively adjusted by No.1 and No.2 stepper motors, where a equal-speed transmission belt is attached to the axis of No.1 motor to drive the horizontal PTZ adjustment, and magnetic rotary encoders are respectively used to acquire real-time absolute angular positions of No.1 and No.2 motors. More details are described below.

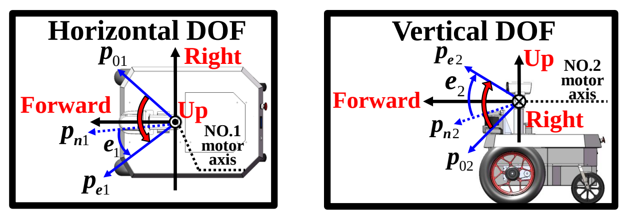


Fig. 2. The diagrams of PTZ adjustment. (a) The illustration of horizontal PTZ adjustment. (b) The illustration of vertical PTZ adjustment.

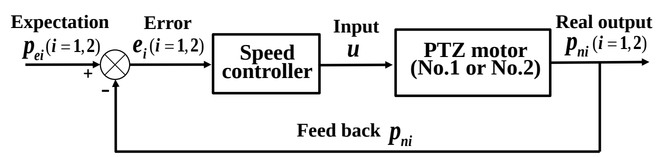


Fig. 3. The control diagram of NO.1 and No.2 motors.

With real-time position feedbacks of encoders, a normal PID speed controller is applied to control NO.1 and NO.2 motors, so that the horizontal and vertical postures of PTZ can be adjusted. As show in Fig. 2(a) and Fig. 2(b), assume the motor positions p_{0i} $(i = 1, 2)$ to be the horizontal and vertical initial conditions respectively, the motor positions p_{ni} the horizontal and vertical real outputs respectively, the motor positions p_{ei} the horizontal and vertical expectations respectively, and e_i the errors between p_{ei} and p_{ni} in different DOFs. The speed controller will respectively calculate pulse control signals u to control NO.1 and NO.2 motors rotating from p_{0i} to p_{ei} in different DOFs as the control diagram shown in Fig. 3.

Combining with the characteristics of the constructed system and the inspection robot we used, the automatic analog instrument reading approach is elementarily designed. First, the robot will start navigating to the parking point relevant to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2967956, IEEE Transactions on Instrumentation and Measurement

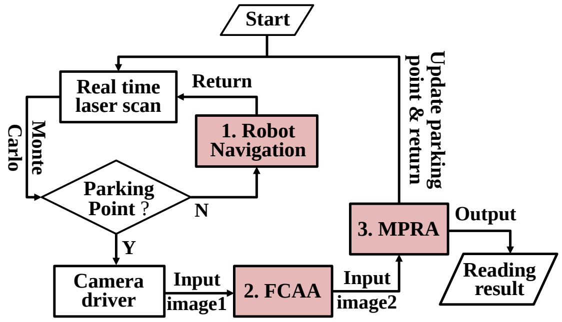


Fig. 4. The flowchart of our automatic reading system.

the first instrument to be read. Second, the proposed FCAA will be performed to quickly align the PTZ camera with the instrument to be read and acquire instrument images that meet the requirements of the automatic reading (see details in Section III). Third, the proposed MPRA will be performed to accurately read the instrument (see details in Section IV). Forth, the robot will navigate to the next parking point, while the second and third steps will be repeated until all analog instruments are read. Finally, all the analog instruments can be read. This process of automatic analog instrument reading using the movable inspection robot is illustrated in Fig. 4.

(a) (b) (c)

Fig. 5. The appearances of different types of analog instruments to be read. (a) A Type-1 instrument. (b) A Type-2 instrument. (c) A Type-3 instrument.

To build our automatic reading system, three types of analog instruments widely used in low-voltage substations are selected as the reading targets. Their appearances are respectively shown in Fig. 5(a)∼Fig. 5(c), in which they are named as type-1∼type-3 instrument respectively. They are round-shape, with thin pointers, and of similar size. Meanwhile, a configuration file is created. For each instrument, some spatial information (*e.g.,* the expected robot parking point coordinate) is written into this file, as well as the instrument type information.

III. Fast Camera Alignment Algorithm (FCAA)

It is known that the robot parking points are usually set far forward (1-5m) from the relevant analog instruments to reduce the complexity of robot navigation [1], [3], [12]–[16] during the robot-based analog instrument inspections. In this case, the general camera without ultra high resolution must zoom in to acquire the instrument image with clear scale lines for automatic reading. Meanwhile, it is known that the camera always zooms in centering on the image center. And the longer time the camera zooms in, the smaller vision field will be. In this case, the PTZ must position the instrument center to the image center initially so that the instrument to be read will not be lost from the vision field after zooming in the camera.

The FCAA is then proposed based on these two ideas. At the very beginning, a YOLOv3-tiny-dw instrument detector based on YOLOv3 [27] and depthwise separable convolutions [28] is proposed and trained. In the execution of FCAA, the PTZ is adjusted first according to some spatial information to align the camera with the instrument elementarily. Then, with a wide camera vision field, the PTZ is adjusted again based on the constructed YOLOv3-tiny-dw instrument detector. Finally, with a small camera vision field, the image of an enlarged and clear instrument can be acquired.

A. YOLOv3-tiny-dw Instrument Detector

1) Production of dataset: To train and test a YOLOv3-tiny-dw detector for wide-vision-field instrument detection, a challenging instrument dataset is constructed for type-1∼type-3 analog instruments introduced in Section II.

In this paper, each robot parking point is set 2.5m right in front of the relevant instrument. In this way, the analog instruments are similar in size in the wide-vision-field inspection images acquired by the PTZ camera of our system. Thus, a 64×64 window is used to label ground truth instrument bounding boxes in different 768×432 inspection images, and form an instrument dataset. And it is divided into a training dataset including 2016 samples and a testing dataset including 1220 samples. Some samples are shown in Fig. 6.

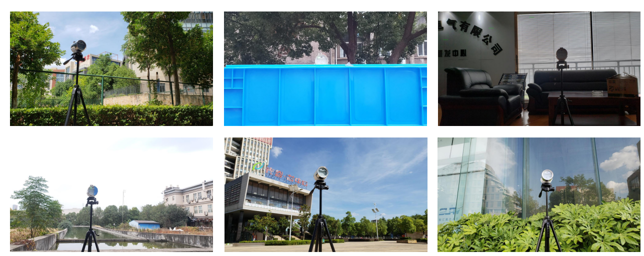


Fig. 6. Some training and testing samples for YOLOv3-tiny-dw.

This challenging dataset fully takes the influencing factors of instrument detection into account, such as indoor and outdoor illumination changes, different camera shooting angles, different environment complexity, different degrees of occlusion, *etc.* Since the instruments are generally set far away from each other in actual operating environment, there is only one certain type analog instrument in each 768×432 sample.

2) Construction of YOLOv3-tiny-dw detector: You Only Look Once version-3 (YOLOv3) [27] is well known as the state-of-the-art object detector. It uses a very deep and complex convolutional neural network (DarkNet-53) to extract stronger features than most other deep learning object detectors [31]–[35]. However, tremendous feature computations in DarkNet-53 made it extremely difficult to run YOLOv3 on most embedded computing platforms, such as embedded robot systems. To alleviate this problem, YOLOv3-tiny was proposed. The network architecture was simplified based on DarkNet-53 in YOLOv3, making it more efficient in feature extraction [36].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2967956, IEEE Transactions on Instrumentation and Measurement

TABLE I
NETWORK ARCHITECTURE OF OUR YOLOV3-TINY-DW INSTRUMENT DETECTOR

Filter type/Stride	Filter shape	Output size
	Network input size	
	$416 \times 416 \times 3$	
Conv dw/1	$3 \times 3 \times 3$	$416 \times 416 \times 3$
Conv pw/1	$1 \times 1 \times 3 \times 16$	$416 \times 416 \times 16$
Maxpool/2	$2 \times 2 \times 16$	$208 \times 208 \times 16$
Conv dw/1	$3 \times 3 \times 16$	$208 \times 208 \times 16$
Conv pw/1	$1 \times 1 \times 16 \times 32$	$208 \times 208 \times 32$
Maxpool/2	$2 \times 2 \times 32$	$104 \times 104 \times 32$
Conv dw/1	$3 \times 3 \times 32$	$104 \times 104 \times 32$
Conv pw/1	$1 \times 1 \times 32 \times 64$	$104 \times 104 \times 64$
Maxpool/2	$2 \times 2 \times 64$	$52 \times 52 \times 64$
Conv dw/1	$3 \times 3 \times 64$	$52 \times 52 \times 64$
Conv pw/1	$1 \times 1 \times 64 \times 128$	$52 \times 52 \times 128$
Maxpool/2	$2 \times 2 \times 128$	$26 \times 26 \times 128$
Conv dw/1	$3 \times 3 \times 128$	$26 \times 26 \times 128$
Conv pw/1	$1 \times 1 \times 128 \times 256$	$26 \times 26 \times 256$
Maxpool/2	$2 \times 2 \times 256$	$13 \times 13 \times 256$
Conv dw/1	$3 \times 3 \times 256$	$13 \times 13 \times 256$
Conv pw/1	$1 \times 1 \times 256 \times 512$	$13 \times 13 \times 512$
Maxpool/1	$2 \times 2 \times 512$	$13 \times 13 \times 512$
Conv dw/1	$3 \times 3 \times 512$	$13 \times 13 \times 512$
Conv pw/1	$1 \times 1 \times 512 \times 1024$	$13 \times 13 \times 1024$
Conv/1	$1 \times 1 \times 1024 \times 256$	$13 \times 13 \times 256$
Conv/1	$3 \times 3 \times 256 \times 512$	$13 \times 13 \times 512$
Conv/1	$1 \times 1 \times 512 \times 255$	$13 \times 13 \times 255$
AvgPool	Global	
Connected	1024×1000	
Softmax		

Based on the network architecture of YOLOv3-tiny, a more efficient YOLOv3-tiny-dw detector is proposed to better adapt to embedded robot system in this paper. To further decrease parameter computations, certain convolutional layers in the network of YOLOv3-tiny are replaced with their corresponding depthwise separable convolution units [28], forming the network of our YOLOv3-tiny-dw detector. As shown in Fig. 7, a depthwise separable convolution unit consists of a depthwise convolutional layer (Conv dw) and a pointwise convolutional layer (Conv pw). In the condition of fixed output, the general convolutional layer (Conv) and its corresponding depthwise separable convolution unit have equivalent feature description ability, but the latter needs much less parameter computations [28]. The architecture of this novel convolutional neural network is given in Table I.

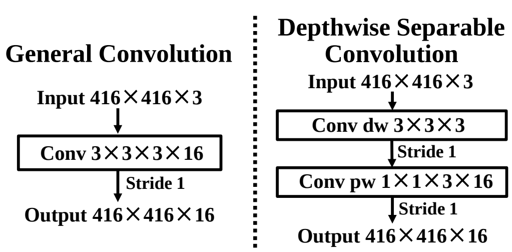
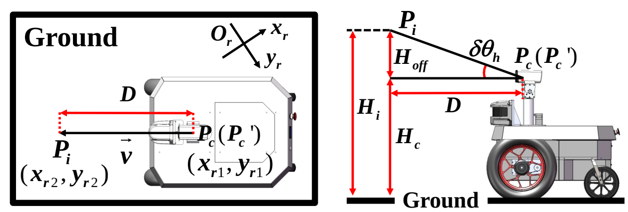


Fig. 7. The schematic of depthwise separable convolution.

Referenced on what YOLOv3-tiny did in detector construction [27], [36], the above network is pre-trained on the standard ImageNet 1000 classes dataset first. Then, the YOLOv3-tiny-dw instrument detector is trained using our training dataset. For an input image in YOLOv3-tiny-dw instrument detection, $13 \times 13 \times 255$ and $26 \times 26 \times 255$ feature maps will be output and used for multi-scale (13×13, 26×26) instrument prediction. The instrument bounding boxes will be directly predicted from the whole input image by using dimension clusters as anchor boxes. The logistic regression will be used to predict an objectness score for each instrument bounding box. By setting a threshold of the above score, the best instrument bounding box can be selected [27], [36]. This detector turned out to be fast and reliable in instrument detection experiments.

B. PTZ Adjustment Based on Spatial Information

This is the first step in the execution of FCAA, which aims to align the camera with the instrument elementarily by using the spatial information previously recorded in the configuration file (see Section II). During the robot-based inspection, it works under the settings that the real-time robot posture $\overrightarrow{v'}$ is estimated centering on the camera image center P'_c, and the default camera posture is set to be the same as $\overrightarrow{v'}$.



(a) (b)

Fig. 8. The top view and side view of an example robot navigation result. (a) The top view. (b) The side view.

As shown in the top view of a robot navigation result in Fig. 8(a), the camera image center P'_c is theoretically coincident with the expected parking point P_c while P'_c is already aligned to the instrument plane center P_i in the horizontal direction, where (x_{r1}, y_{r1}) and (x_{r2}, y_{r2}) respectively represent the coordinate of $P_c(P'_c)$ and P_i in the 2-D SLAM coordinate system $O_r x_r y_r$, $\overrightarrow{v}$ represents the expected robot posture, and D represents the $L2$ distance between (x_{r1}, y_{r1}) and (x_{r2}, y_{r2}), which can be calculated by

$$D = \sqrt{(x_{r1} - x_{r2})^2 + (y_{r1} - y_{r2})^2}. \quad (1)$$

As shown in the side view of a robot navigation result in Fig. 8(b), there is an deviation angle $\delta\theta_h$ to be corrected to align P'_c to P_i in the vertical direction, where H_c and H_i are respectively the heights of P'_c and P_i. That is, we only need to adjust the PTZ in its vertical DOF.

For each instrument installed at different heights, (x_{r1}, y_{r1}), (x_{r2}, y_{r2}), H_c, and H_i are known. In this case, the vertical deviation angle $\delta\theta_h$ can be calculated first by

$$\delta\theta_h = \arctan\frac{H_{off}}{D}, \ H_{off} = H_i - H_c. \quad (2)$$

The control strategy shown in Fig. 3 is applied to adjust the vertical PTZ posture, and then $\delta\theta_h$ can be corrected. Finally, 768×432 images are acquired respectively with a short camera focal length and a long camera focal length, shown in Fig. 9(a) and Fig. 9(b).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2967956, IEEE Transactions on Instrumentation and Measurement

(a) (b)

Fig. 9. Result images of PTZ adjustment based on spatial information. (a) The image with wide vision field. (b) The image with small vision field.

Obviously, the PTZ camera is not aligned to the instrument precisely enough by executing this PTZ adjustment. The instrument in the wide-vision-field image is lost from the camera vision field by zooming in the camera focal length. There are two reasons. First, the estimation errors of Monte Carlo method [30] always exist during the robot navigation. Second, it is hard to find the camera image center and the instrument plane center in the real world so that P_c', P_i, H_c and H_i are not precise by manual calibration.

C. PTZ Adjustment Based on Visual Servo Instrument Detection

This is the second step in the execution of FCAA. According to Fig. 9(a), the instrument to be read is already positioned in wide camera vision field by PTZ adjustment based on spatial information. However, the further PTZ adjustment is needed yet to align the camera with the instrument precisely.

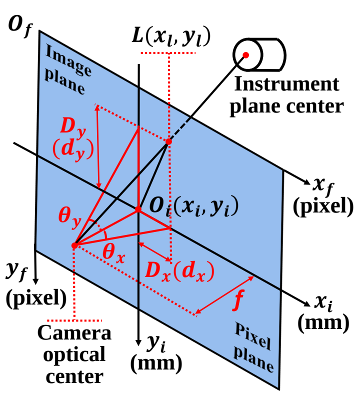
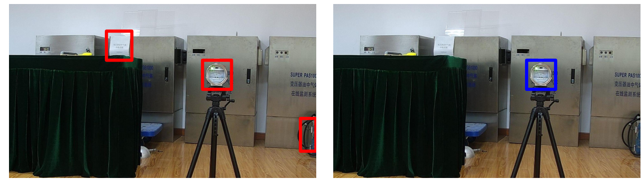


Fig. 10. The pinhole camera system during the robot-based inspection.

Assume $O_i x_i y_i$ is the image plane coordinate system, and $O_f x_f y_f$ the pixel plane coordinate system. A pinhole camera system is built in Fig. 10 based on [21]. There are camera deviation angles θ_x, θ_y between the camera image center O_i and the imaging point L of the instrument center, which respectively resulted by the horizontal and vertical image deviations D_x, D_y in the image plane $O_i x_i y_i$. Meanwhile, the pixel deviations d_x and d_y in the pixel plane $O_f x_f y_f$ are respectively corresponding to D_x and D_y, while the shortest camera focal length f and the pixel coordinate (x_i, y_i) of O_i are already calibrated based on [21]. In this case, when the pixel coordinate (x_l, y_l) of L is known in the pixel plane $O_f x_f y_f$, the camera can be aligned to the instrument by calculating θ_x and θ_y, and correcting them.



(a) (b)

Fig. 11. Results of the YOLOv3-tiny-dw instrument detection. (a) The detection result without threshold selection. (b) The detection result selected by an objectness score threshold of 0.5.

First, the YOLOv3-tiny-dw instrument detector proposed previously is applied to the original image. Take an example as Fig. 9(a), the results are shown in Fig. 11(a) and Fig. 11(b). They are respectively the detection result without threshold selection and the ultimate predicted instrument bounding box selected by an objectness score threshold of 0.5, in which (x_l, y_l) is represented by the pixel coordinate of the center of that bounding box. Next, the pixel deviations d_x, d_y and the image deviations D_x, D_y are calculated by

$$d_x = x_l - x_i, \; D_x = d_x \cdot R_x, \tag{3}$$

$$d_y = y_l - y_i, \; D_y = d_y \cdot R_y, \tag{4}$$

where R_x and R_y are respectively the physical height and physical width of one pixel, and they are determined by the camera used. Then, the camera deviation angles θ_x and θ_y can be calculated by

$$\theta_x = \arctan(D_x/f), \; \theta_y = \arctan(D_y/f). \tag{5}$$

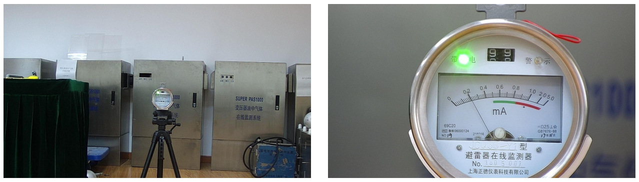
As shown in Fig. 2(a) and Fig. 2(b), the settings of positive directions in PTZ's horizontal and vertical DOFs are the same as that in the image plane $O_i x_i y_i$ and the pixel plane $O_f x_f y_f$. In this way, the control strategy shown in Fig. 3 is applied to adjust the horizontal and vertical PTZ postures respectively, and the camera deviation angles θ_x and θ_y can be corrected in the end.

Theoretically, the camera will be aligned with the instrument by applying the above method, in which the instrument detection is performed only once. However, the performance of this PTZ adjustment based on one-time instrument detection can be unreliable in practice. First, motion errors of the motors and position errors of the encoders will both influence the motion precision of the PTZ. Second, some unintended wrong classifications of the instrument detection will lead to a wrong posture of the PTZ.

To reduce the influence of the above factors, a PTZ adjustment based on visual servo instrument detection is adopted in practice. Its main idea is to apply the PTZ adjustment based on one-time instrument detection iteratively, in which the result image similar to Fig. 9(a) will be acquired with a wide vision field at the end of each iteration, the result image of the former iteration will be input to the next iteration, and those iterations will be stopped when the calculated camera deviation angles θ_x and θ_y satisfy

$$|\theta_x| \le T_{\theta x}, \; |\theta_y| \le T_{\theta y}, \tag{6}$$

where $T_{\theta x}$ and $T_{\theta y}$ are respectively the maximum tolerable limits of θ_x and θ_y summarized from plenty of experiments.



(a) (b)

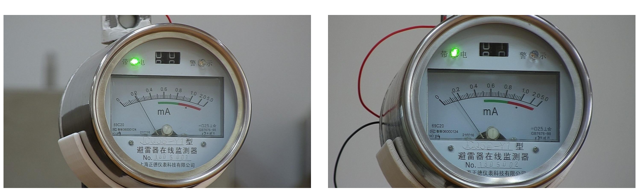
Fig. 12. Result images of PTZ adjustment based on visual servo instrument detection. (a) The image with wide vision field. (b) The image with small vision field.

768×432 result images are acquired respectively with a short camera focal length and a long camera focal length, shown in Fig. 12(a) and Fig. 12(b). It can be seen that the instrument to be read is basically in the middle of the wide camera vision field, and the main body of that instrument is still in the camera vision field after zooming in the camera focal length. Meanwhile, the image shown in Fig. 12(b) contains sufficient details of the scale lines and the pointer, so that it can satisfy the requirements of automatic analog instrument reading in MPRA.

In experiments of the camera alignment using our FCAA, the FCAA was shown to be effective for type-1∼type-3 analog instruments, while the FCAA using visual servo instrument detection was shown to perform better than the FCAA using one-time instrument detection, according to the success rate and precision of the camera alignment.

IV. Monocular-vision Pointer Reconstruction Algorithm (MPRA)

According to previous sections, the round-shape analog instruments to be read may be installed at different heights. Meanwhile, a series of errors always exist during the camera alignment using FCAA. These factors will lead to different degrees of instrument deformation during the FCAA process, as shown in Fig. 13, which may further influence the accuracy of the instrument reading.



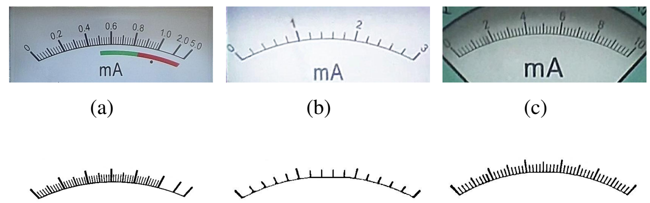
(a) (b)

Fig. 13. The instrument deformation caused in FCAA process. (a) The first example. (b) The second example.

To this end, an universal and precise Monocular-vision Pointer Reconstruction Algorithm (MPRA) is proposed for the automatic reading of different types of analog instruments. Its main idea is to remove the influence of that deformation by reconstructing the pointer pixels onto the strict forward-looking scale region of the instrument to be read, and precisely calculating the reading based on polar transform [6], [9].

A. Preparation

To realize the MPRA for monocular vision, template images of the forward-looking instrument scale region are prepared for type-1∼type-3 analog instruments, in which a color template image and a black-white template image with scale lines only are prepared for each type of instrument, as shown in Fig. 14(a)∼Fig. 14(f).



(a) (b) (c)

(d) (e) (f)

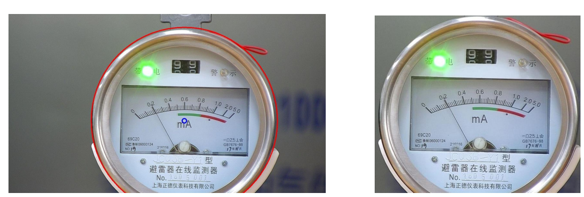
Fig. 14. Template images used in MPRA. (a) Type-1 color template. (b) Type-2 color template. (c) Type-3 color template. (d) Type-1 black-white template. (e) Type-2 black-white template. (f) Type-3 black-white template.

Since the type information of different analog instruments to be read is already known in the configuration file, the relevant template images can be selected according to this information at the beginning when a certain type of instrument is supposed to be read, and then applied to the MPRA process.

B. Determination of Actual Scale Region

This is the first step of MPRA. By using a template matching-based search method, it aims to determine the actual scale region with different degrees of deformation in the result image of FCAA, so that there will be less computations in the further steps of MPRA.

First, Circle Hough Transform (CHT) [37] is applied to locate the round-shape instrument in the 768×432 image shown as Fig. 12(b), so that the range of searching the actual scale region can be narrowed. The result is shown in Fig. 15(a), in which the detected circle edge is indicated in red, and the detected circle center in blue. Based on this result, a square instrument region shown in Fig. 15(b) can be cropped out, and prepared for scale region search.



(a) (b)

Fig. 15. Determination of the instrument region. (a) The result of CHT. (b) The instrument region cropped based on (a).

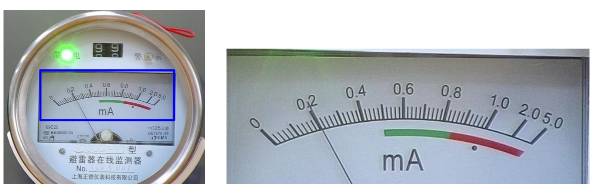
Second, a searching window of the same size as the relevant color template image is applied to scan across the prepared instrument region at different locations, while the Normalized Cross Correlation (NCC) is applied to evaluate the matching degrees between each scanned image region and the above

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2967956, IEEE Transactions on Instrumentation and Measurement

template image, in which a NCC coefficient σ_{ncc} can be calculated over a scanned image region by

$$\sigma_{ncc} = \frac{\sum_i \sum_j [I(i,j) - \overline{I}][T(i,j) - \overline{T}]}{\sqrt{\sum_i \sum_j [I(i,j) - \overline{I}]^2 \cdot \sum_i \sum_j [T(i,j) - \overline{T}]^2}}, \quad (7)$$

where $I(i,j)$ and $\overline{I}$ are respectively the gray-scale pixel intensity relevant to the ith row and the jth column in this image region and the mean gray-scale pixel intensity in this region, while the meanings of $T(i,j)$ and $\overline{T}$ are similar for the relevant template image. According to the properties of the NCC coefficient, the scanned image region relevant to the σ_{ncc} closest to 1 is considered to be the most similar to the relevant template image and is regarded as the actual scale region in the end. It should be noted that the searching window will be scanned at multiple scales with a sliding stride of four pixels to ensure good performances on both the precision and computational cost of the above template matching-based search, in which the rescaling factors of the original template image are respectively 0.8, 0.9, 1.0, 1.1, 1.2. The results are shown in Fig. 16, where Fig. 16(a) shows the result of the determined actual scale region, and Fig. 16(b) shows the cropped actual scale region which is prepared for the further steps.



(a) (b)

Fig. 16. Determination of the actual scale region. (a) The result of the template matching-based search. (b) The actual scale region cropped based on (a).

C. Reconstruction of Pointer Pixels

This is the second step of MPRA. By using SURF [25], it aims to reconstruct the pixel pointers in the actual scale region plane onto its relevant template image plane which represent the strict forward-looking scale region, so that the influence of the instrument deformation can be removed.

First, the coordinate mapping relationship between the actual scale region plane shown in Fig. 16(b) and its relevant color template image plane is calculated to prepare for the reconstruction of pointer pixels, in which SURF [25] is applied to detect and match the feature points respectively belonged to the actual scale region and its relevant color template image, RANSAC is applied to further exclude those mismatched pairs of feature points, and the coordinate mapping relationship represented by a homography matrix

$$\mathbf{T}_{3\times3} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (8)$$

can be calculated by plugging the homogeneous pixel coordinates of multiple matched pairs of feature points into

$$[x_s \ y_s \ 1]^{\mathrm{T}} = \mathbf{T}_{3\times3} \cdot [x'_s \ y'_s \ 1]^{\mathrm{T}} \quad (9)$$

where $[x_s \ y_s \ 1]^{\mathrm{T}}$ and $[x'_s \ y'_s \ 1]^{\mathrm{T}}$ are the homogeneous pixel coordinates of a matched pair of feature points respectively belonging to the color template image plane and the actual scale region plane, and then performing the least square fitting. As shown in Fig. 17, these are the best matched pairs of SURF points between the actual scale region and its relevant color template image.

Fig. 17. The matched pairs of SURF points in the MPRA process.

(a) (b)

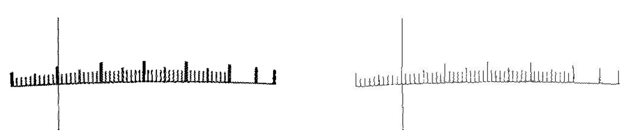
Fig. 18. Reconstruction of the pointer pixels. (a) The pointer pixels determined by HT. (b) The result of the pointer reconstruction.

Next, the pointer pixels of the actual scale region can be determined by Hough Transform (HT) [3], [7], [9]. As shown in Fig. 18(a), we determine the pointer pixels by locating the pointer pixel line (indicated in red) which is the mean of the pixel lines (indicated in blue) detected by HT.

The fact is that the coordinate system of a color template image is in coincidence with that of its relevant black-white template image. In this case, finally, the pointer pixels of the actual scale region plane can be reconstructed onto its relevant black-white template image plane by the matrix $\mathbf{T}_{3\times3}$, in which the reconstructed pointer pixels are assigned black on this template image plane. Thus, the influence of the deformation is removed, and the computations in further steps will be reduced. The result is shown in Fig. 18(b).

D. Determination of Analog Instrument Reading

This is the final step of MPRA, in which the reading of a certain type (type-1$\sim$type-3) analog instrument can be determined by the method proposed in our previous work [3].



(a) (b)

Fig. 19. Result images of the polar transform and the image thinning. (a) The result of the polar transform. (b) The result of the image thinning.

First, the polar transform [6], [9] is applied to the result of the pointer reconstruction shown in Fig 18(b). As shown in Fig. 19(a), the arc belonged to the scale lines is converted to a straight line, so that the issue of determining the reading is

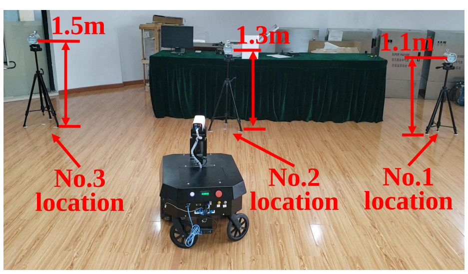
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2967956, IEEE Transactions on Instrumentation and Measurement

simplified. Then, an image thinning algorithm [38] is applied to Fig. 19(a). As shown in Fig. 19(b), the one-pixel-width skeleton is extracted, so that it is easier to locate each scale line and the pointer. Next, each scale line and the pointer are located by counting black pixels over each column in Fig. 19(b), in which the column have the largest number of black pixels is regarded the pointer location. There are different designs of the scale marks of type-1∼type-3 analog instruments. To read a certain type instrument, the configuration file is used again here to further acquire some other type information, *e. g.*, the unit scale value, the number of scale lines, and the maximum scale value. In this case, finally, the reading of the instrument can be calculated based on the relative location relationship between the pointer line and the scale lines, just like the method in [3].

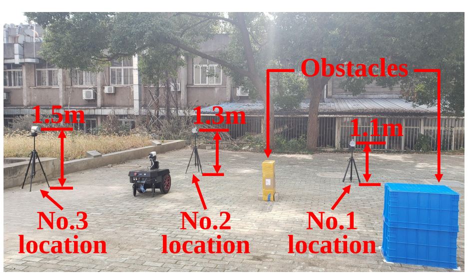
In experiments of the MPRA, the MPRA was shown to be effective and precise for the automatic reading of type-1∼type-3 analog instruments during the robot-based inspection.

V. Performance of Our Automatic Recognition System

To evaluate the performance of the automatic analog instrument reading system during complex real-world inspections, similar indoor and outdoor experimental scenes were set up to simulate the actual operating environments of low-voltage substations, water plants, chemical plants, *etc*. They are respectively shown in Fig. 20(a) and Fig. 20(b).



(a)



(b)

Fig. 20. The experimental scenes. (a) Indoor scene. (b) Outdoor scene.

In both scenes of Fig. 20, three locations named NO.1∼3 were set to put the instruments to be read, where each robot parking point was set 2.5m in front of the corresponding instrument location. As mentioned in Section I, the analog instruments are generally installed at 0.5∼1.5m height in actual operating environments, so that workers can read during manual inspections. Hence, different heights were selected for instrument installations at NO.1∼3 locations, which were 1.1m, 1.3m, and 1.5m respectively. In this case, the evaluation experiments of the YOLOv3-tiny-dw detector, the FCAA, and the MPRA were designed and carried out respectively.

A. *Performance of YOLOv3-tiny-dw Instrument Detector*

In the camera alignment process using FCAA, positioning the instrument plane center to the image center relies on the instrument detection, in which the instrument detection is carried out by our YOLOv3-tiny-dw detector. Therefore, the detection performance of our YOLOv3-tiny-dw detector largely determines the performance of the camera alignment process during robot-based inspections. To evaluate the instrument detection performance of the proposed detector, first, several common-used object detectors were tested on the challenging testing dataset we produced and compared with the proposed detector. Then, the proposed detector was also evaluated in the condition of instrument occlusion.

1) Comparison of instrument detectors: First, several classic object detectors including the Cascade Adaboost [26], [39], the Dual Cascade Adaboost [40], the HOG-SVM [41] were constructed and tested as comparative detectors. Meanwhile, a Faster HOG-SVM detector was proposed in the experiments, and also tested as a comparative detector. The default configurations of the detectors are described below.

Comparative detector I. Cascade Adaboost: The local feature classifier contained 20 levels of Haar-like features. The most discriminative features were selected by Adaboost to build strong classifiers. And the detector detected at all scales from 48×48 to 82×82 with a rescaling factor of 1.2 and a sliding stride of 4 [39].

Comparative detector II. Dual Cascade Adaboost: The local feature classifier contained 8 levels of Haar-like features and 12 levels of MS-LBP blocks. The most discriminative features were selected by Adaboost to build strong classifiers. And the detector detected at all scales from 48×48 to 82×82 with a rescaling factor of 1.2 and a sliding stride of 4 [40].

Comparative detector III. HOG-SVM: The local feature classifier was trained using the HOG feature and the linear soft-margin SVM. The detector was divided into 8×8 grids to calculate the HOG feature. And the detector detected at scales 48×48, 64×64, 80×80 with a sliding stride of 8 [41].

Comparative detector IV. Faster HOG-SVM: Only the differences between the standard HOG-SVM are listed below. It is known that, an instrument to be detected usually occupies a small Region Of Interest (ROI) in the wide-vision-field inspection image. It would be a waste of time if global detection is performed. In this case, Selective Search [42], [43] were applied to suggest small instrument region proposals first, and then the detector merely detected in region proposals to avoid redundant computations [39]–[41].

It should be noted that, the above four detectors are supposed to train local feature classifiers first, and then detect objects by using sliding window strategy. Hence, based on ground truth labels in our training dataset, patches were cropped for their classifier training. Since the limitation of Nvidia® acceleration, they are tested by only the Nvidia® Denver™ CPU.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2967956, IEEE Transactions on Instrumentation and Measurement

TABLE II
INTERPOLATED TPR VALUES FOR DIFFERENT FPR VALUES ON THE ROC CURVES

TPR (%) / Detector	False Positive Rate (FPR)							CPU time/s	GPU time/s
	0.005	0.010	0.015	0.020	0.025	0.030	0.035		
Comparative I. Cascade Adaboost	77.89	81.50	83.80	86.58	87.89	88.39	88.55	0.240(≈4fps)	–
Comparative II. Dual Cascade Adaboost	83.98	87.59	89.39	90.70	91.03	91.52	91.52	0.605(<1fps)	–
Comparative III. HOG-SVM	91.80	93.28	94.26	94.75	95.57	96.07	96.07	1.439(<1fps)	–
Comparative IV. Our Faster HOG-SVM	90.28	92.41	94.05	94.54	95.20	95.20	95.52	0.173(≈6fps)	–
Comparative V. YOLOv3	99.84	100.00	100.00	100.00	100.00	100.00	100.00	2.958(<1fps)	0.096(≈10fps)
Comparative VI. YOLOv3-tiny	99.34	99.34	99.34	99.34	99.34	99.34	99.34	0.672(≈2fps)	0.082(≈12fps)
Our YOLOv3-tiny-dw	99.84	100.00	100.00	100.00	100.00	100.00	100.00	0.463(≈2fps)	0.066(≈16fps)

Then, the latest and the state-of-the-art YOLOv3 [27] and YOLOv3-tiny [36] were constructed on our training dataset and tested as comparative detectors. The default configurations of the detectors are described below.

Comparative detector V. YOLOv3: A 53-layer convolutional neural network was used to extract features. The instrument bounding boxes were predicted at three scales (13×13, 26×26, 52×52) by adding two passthrough layers into DarkNet-53, as well as using K-means clusters and logistic regression [27]. Meanwhile, the objectness score threshold was set to 0.5.

Comparative detector VI. YOLOv3-tiny: A 16-layer convolutional neural network was used to extract features. The instrument bounding boxes were predicted at two scales (13×13, 26×26) by adding a passthrough layer into its network, as well as using K-means clusters and logistic regression [36]. Meanwhile, the objectness score threshold was set to 0.5.

Since the Nvidia® acceleration is available to YOLOv3 and YOLOv3-tiny, they (plus our YOLOv3-tiny-dw) were tested by both the Nvidia® Denver™ CPU and the Nvidia® Kepler™ GPU in our robot system.

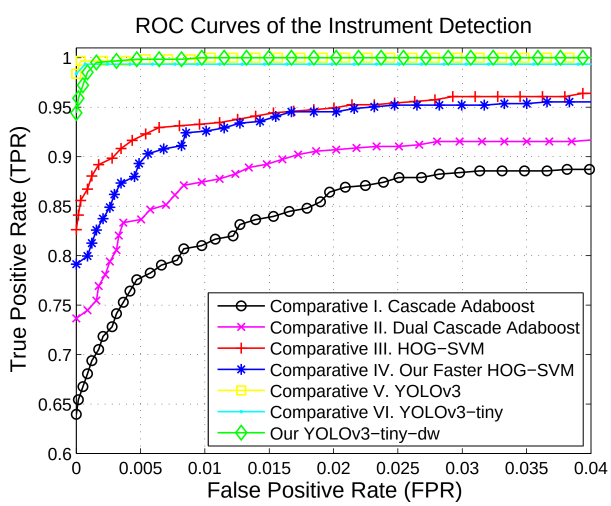


Fig. 21. The ROC curves which show the performances of seven different instrument detectors on our testing dataset.

These six comparative instrument detectors together with our YOLOv3-tiny-dw were tested on our challenging testing dataset, which included 1220 images. Referenced on the PASCAL VOC criteria [44], seven Receiver Operating Characteristic (ROC) curves are drawn in Fig. 21 to show the overall performances of these instrument detectors, in which the x-axis and y-axis of the ROC coordinate system are respectively corresponding to the False Positive Rate of detection (FPR) and True Positive Rate of detection (TPR). Meanwhile, a group of FPR values are sampled evenly from the x-axis of the ROC coordinate system to intuitively show their corresponding TPR values for each ROC curve, as presented in Table II.

Combining the results shown in Fig. 21 and Table II, conclusions can be made as below. From the perspective of the overall detection performance, the YOLOv3-tiny-dw instrument detector proposed in our robot system was much better than the Cascade Adaboost, the Dual Cascade Adaboost, the HOG-SVM, and the Faster HOG-SVM. And it performed almost the same as the state-of-the-art YOLOv3 and YOLOv3-tiny. From the perspective of the mean detection time, the proposed YOLOv3-tiny-dw was the fastest (16fps) when GPU was available. And it was always the fastest compared with YOLOv3 and YOLOv3-tiny, both on CPU and on GPU. That is to say, the application of depthwise separable convolutions [28] does promote the efficiency of YOLOv3 and YOLOv3-tiny. Meanwhile, the Faster HOG-SVM proposed in our experiments was found the fastest (5fps) when only CPU was available. In summary, with the GPU acceleration, the proposed YOLOv3-tiny-dw is fast and accurate. It can meet the application requirements of indoor and outdoor robot-based inspections, and is also the reason why the FCAA can work efficiently.

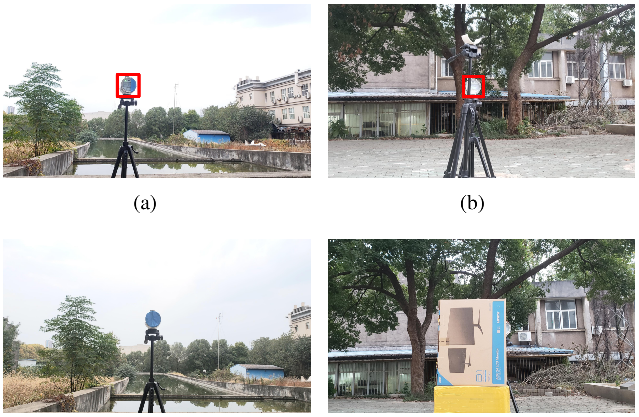
2) Discussion on instrument occlusion: In the indoor and outdoor inspection scenes, some unexpected occlusion (*e.g.*, leaves, dirt accumulated on instrument plane, the occlusion caused by different visual angles, *etc.*) would affect the instrument detection performances. Hence, it is meaningful to discuss the performance of our YOLOv3-tiny-dw instrument detector on the instrument occlusion.

TABLE III
HIT RATES ON DATASETS WITH DIFFERENT DEGREES OF INSTRUMENT OCCLUSION

Occlusion degree	Hit	Miss	Total	Hit rate (%)
Slight	197	3	200	98.5
Serious	19	111	200	9.5

Additionally, a special dataset of 200 images with slight instrument occlusion and a special dataset of 200 images with significant instrument occlusion were constructed in the same way as the previous testing dataset. The hit rates of type 1∼3 analog instruments in YOLOv3-tiny-dw detection are listed in

Table III. YOLOv3-tiny-dw demonstrated excellent robustness to slight instrument occlusion (a hit rate of 98.5%), but it is vulnerable to significant instrument occlusion (only a hit rate of 9.5%). Some detection examples are shown in Fig. 22.



(a) (b) (c) (d)

Fig. 22. Some example detection results on special datasets with different degrees of instrument occlusion. (a), (b) Success detections in slight occlusion condition. (c), (d) Failed detections in serious occlusion condition.

In fact, in actual low-voltage substations, water plants, chemical plants, *etc.*, cleaning different kinds of instrument occlusion is also an important work to be performed regularly. The reason is that, it is unavailable and meaningless to read an analog instrument whose main indication region is invisible in the inspection. In this case, the proposed YOLOv3-tiny-dw instrument detector satisfies indoor and outdoor application requirements on occlusion robustness, and can handle slight occlusion with excellent performances.

B. Comparison of FCAA Based on Visual Servo Instrument Detection and FCAA Based on One-time Instrument Detection

In Section III, one-time instrument detection and visual servo instrument detection were proposed as two execution methods of the FCAA. Via the error analysis of our robot system, we believed that the latter would present higher success rate and better precision on camera alignment.

TABLE IV
CONFIGURATIONS OF FCAA COMPARISON EXPERIMENTS FOR A CERTAIN-TYPE ANALOG INSTRUMENT

Object: type-1∼3 instrument		Instrument height: 1.3m	
Total inspection	FCAA type	Experimental scene	Inspection time
40	One-time detection	Indoor	10
		Outdoor 9:00 am	10
		Outdoor 12:00 am	10
		Outdoor 4:00 pm	10
40	Visual servo detection	Indoor	10
		Outdoor 9:00 am	10
		Outdoor 12:00 am	10
		Outdoor 4:00 pm	10

In this part, comparison experiments were designed for type-1∼type-3 analog instrument to confirm the above argument. For each type of instrument, the FCAA-based camera alignment was executed during 80 different inspections, in which the one-time way was adopted during 40 inspections, and the visual servo way was adopted during the other 40 inspections. To make our experiment results more persuasive and general, we also considered disturbations possibly caused by variable experiment scenes (*e.g.,* illumination changes, environmental changes, *etc.*). Hence, the above experiments were carried out in four different scenes. Meanwhile, to follow the variable-controlling principle, 1.3m height (*i.e.* the NO.2 locations in both Fig. 20(a) and Fig. 20(b)) was adopted for each type of instrument. The detailed experimental configurations are listed in Table IV.

For easy to compare, some definitions were made first. Refer to a result image of FCAA shown in Fig. 12(b), the FCAA process was considered successful when the actual scale region of a certain type instrument (see Fig. 16(b)) was totally positioned in this small-vision-field image, and the camera alignment precision was represented by the Euclidean distance between instrument center and image center. In this case, the experimental results respectively corresponded to type-1 instrument, type-2 instrument, and type-3 instrument were recorded and the performances of two types of FCAA were compared.

TABLE V
SUCCESS RATES OF FCAA (ONE-TIME DETECTION)

Instrument type	Success	Failure	Total	Success rate (%)
Type-1	28	12	40	70
Type-2	29	11	40	72.5
Type-3	31	9	40	77.5

TABLE VI
SUCCESS RATES OF FCAA (VISUAL SERVO DETECTION)

Instrument type	Success	Failure	Total	Success rate (%)
Type-1	38	2	40	95
Type-2	37	3	40	92.5
Type-3	40	0	40	100

As the results presented in Table V and Table VI, the overall camera alignment success rates of type-1∼type-3 instruments were found respectively 70%, 72.5%, 77.5% by using the F-CAA based on one-time detection, and those rates dramatically increased to 95%, 92.5%, 100% by using the FCAA based on visual servo detection. As the results shown in Fig. 23(a), Fig. 23(c), and Fig. 23(e), the instrument centers positioned by the FCAA based on visual servo detection generally converged much closer to the image center than those positioned by the FCAA based on one-time detection. Meanwhile, the overall camera alignment precision distributions of type-1∼type-3 instruments shown in Fig. 23(b), Fig. 23(d), and Fig. 23(f) indicated that, the median precision values acquired by using the FCAA based on one-time detection were respectively 194, 239, 186 pixels, and those values acquired by using the FCAA based on visual servo detection dramatically improved to 69, 82, 96 pixels. Also, the distribution section of the precision values was generally much smaller in the latter situation.

In summary, the FCAA with visual servo instrument detection did achieve higher success rate and higher precision than the FCAA with one-time instrument detection during the

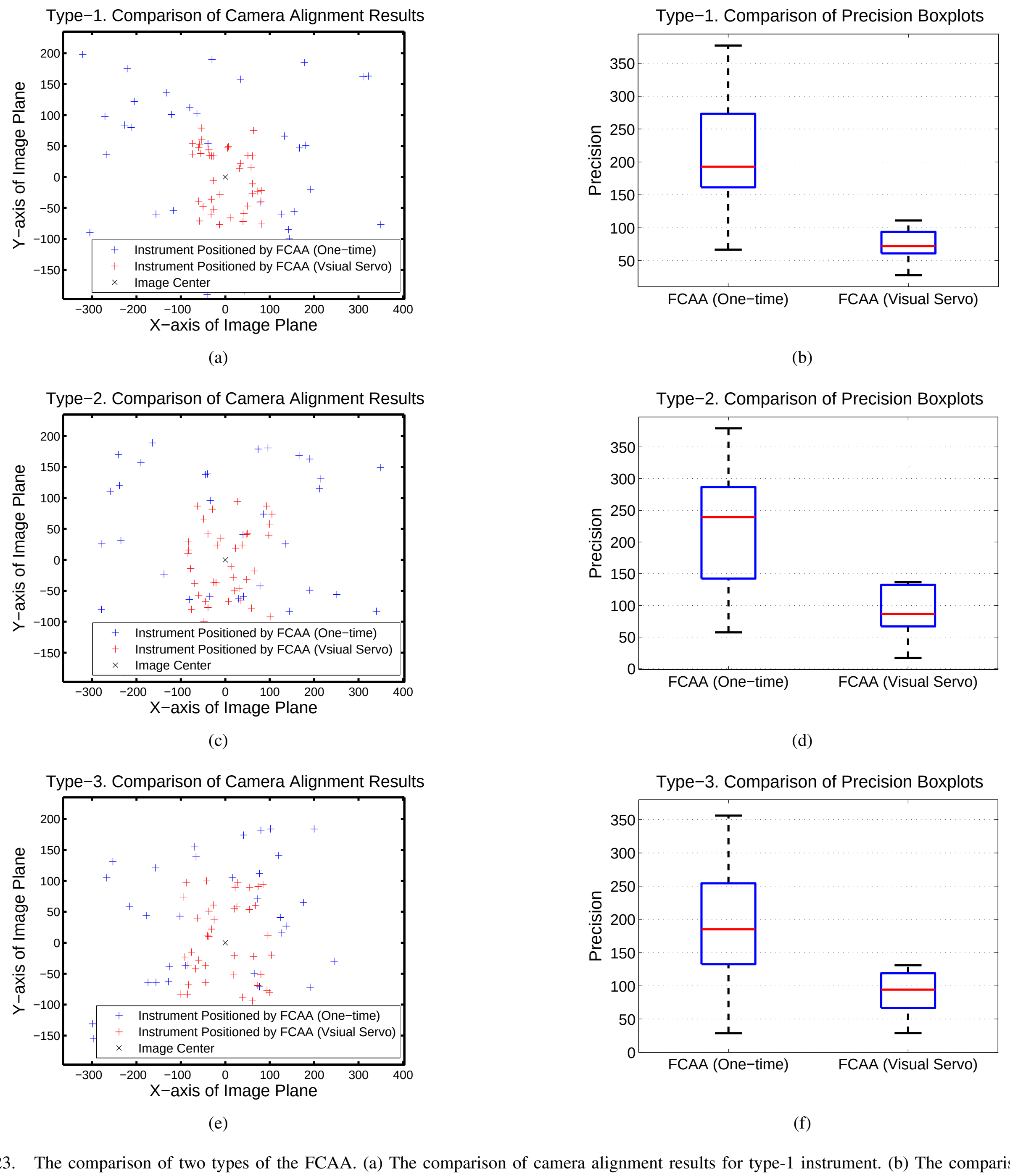


Fig. 23. The comparison of two types of the FCAA. (a) The comparison of camera alignment results for type-1 instrument. (b) The comparison of camera alignment precision box plots for type-1 instrument. (c) The same meaning to (a) but for type-2 instrument. (d) The same meaning to (b) but for type-2 instrument. (e) The same meaning to (a) but for type-3 instrument. (f) The same meaning to (b) but for type-3 instrument.

camera alignment process of the robot-based inspection. This is why we adopted visual servo instrument detection in the FCAA-based camera alignment process rather than one-time instrument detection. Meanwhile, the FCAA was found steady and effective for type-1∼type-3 analog instruments, whether indoor, outdoor or scenes with complex illumination changes.

C. Performance of MPRA

It is known that illumination changes and different instrument heights are two influential factors to the reading accuracy of MPRA. The former always leads to intensity distribution changes in the image, which may affect the SURF-Matching stage in MPRA. The latter brings different degrees of instrument deformation (see Fig. 13), which may affect the pointer reconstruction stage in MPRA. Hence, in our robot system, it is necessary and meaningful to discuss the robustness of MPRA to illumination changes and different instrument heights respectively.

In this case, groups of FCAA-MPRA union experiments were designed and carried out for type-1∼type-3 analog instruments. For the convenience of discussion, some general experimental configurations were specified first. The reading accuracy was represented by the absolute error between the reading acquired by MPRA and the corresponding real reading. The real reading of each instruments to be read was calibrated by adjusting a DC adjustable steady current supply with the help of human eyes. Meanwhile, it should be noted that the unit scale values of type-1∼type-3 analog instruments were respectively 0.020mA, 0.20mA, and 0.20mA.

TABLE VII
EXPERIMENTAL CONFIGURATIONS AND RESULTS OF MPRA IN DISCUSSION ON ILLUMINATION CHANGES

Object: type-1∼3 instrument		Instrument height: 1.3m		reading unit: mA	
Instrument type	Illumination level	Inspection time	Real reading	MPRA reading	Absolute error
Type-1	Indoor	3	0.330	0.3271	0.0029
			0.470	0.4617	0.0083
			0.620	0.6119	0.0081
	Outdoor 9:00 am	3	0.330	0.3361	0.0061
			0.470	0.4657	0.0043
			0.620	0.6109	0.0091
	Outdoor 12:00 am	3	0.330	0.3310	0.0010
			0.470	0.4792	0.0092
			0.620	0.6132	0.0068
	Outdoor 4:00 pm	3	0.330	0.3301	0.0001
			0.470	0.4771	0.0071
			0.620	0.6259	0.0059
Type-2	Indoor	3	0.45	0.439	0.011
			1.50	1.533	0.033
			2.60	2.584	0.016
	Outdoor 9:00 am	3	0.45	0.437	0.013
			1.50	1.442	0.058
			2.60	2.675	0.075
	Outdoor 12:00 am	3	0.45	0.531	0.081
			1.50	1.490	0.010
			2.60	2.554	0.046
	Outdoor 4:00 pm	3	0.45	0.498	0.042
			1.50	1.479	0.021
			2.60	2.595	0.065
Type-3	Indoor	3	1.15	1.163	0.013
			5.00	4.928	0.072
			9.05	9.104	0.054
	Outdoor 9:00 am	3	1.15	1.197	0.047
			5.00	4.961	0.039
			9.05	8.995	0.055
	Outdoor 12:00 am	3	1.15	1.108	0.042
			5.00	4.992	0.008
			9.05	9.101	0.051
	Outdoor 4:00 pm	3	1.15	1.104	0.046
			5.00	5.089	0.089
			9.05	9.092	0.042

1) Discussion on illumination changes: In this part, experiments were designed for type-1∼type-3 analog instruments to study the robustness of MPRA to illumination changes. To obtain comparative experimental results, four different illumination levels were configured based on the indoor and outdoor experimental scenes shown in Fig. 20(a) and Fig. 20(b). These four levels include the indoor case, and the outdoor cases at 9:00 am, 12:00 am, and 4:00 pm respectively. To follow the variable-controlling principle, 1.3m height (*i.e.* the NO.2 locations in both indoor and outdoor experimental scenes) was adopted to install the analog instruments to be read. In the experiments, for each type of instrument, we designed twelve inspection trials for the MPRA-based analog instrument reading. These twelve trials were divided equally into four groups. Each group contains three trials, which were carried out at the same illumination level.

The detailed experimental configurations are presented in Table VII, as well as the experimental results. As can be calculated, for type-1 instrument, the mean absolute reading errors of MPRA in four different illumination levels were respectively 0.0064, 0.0065, 0.0057, and 0.0044 mA (merely 32%, 32.5%, 28.5%, and 22% of the unit scale value of 0.020 mA). For type-2 instrument, those errors were respectively 0.020, 0.049, 0.046, and 0.043 mA (merely 10%, 24.5%, 23%, and 21.5% of the unit scale value of 0.20 mA). And for type-3 instrument, those errors were respectively 0.046, 0.047, 0.033, and 0.059 mA (merely 23%, 23.5%, 16.5%, and 29.5% of the unit scale value of 0.20 mA). The mean absolute reading errors of MPRA had little change when the illumination levels changed randomly. In summary, the MPRA was robust to illumination changes. Meanwhile, it was universal to automatic readings of type-1∼type-3 analog instruments, and with a high reading accuracy.

TABLE VIII
EXPERIMENTAL CONFIGURATIONS AND RESULTS OF MPRA IN DISCUSSION ON DIFFERENT INSTRUMENT HEIGHTS

Object: type-1∼3 instrument		Illumination level: indoor		reading unit: mA	
Instrument type	Instrument height	Inspection number	Real reading	MPRA reading	Absolute error
Type-1	1.1m	1	0.330	0.3251	0.0049
		2	0.470	0.4700	0.0000
		3	0.620	0.6279	0.0079
	1.3m	1	0.330	0.3355	0.0055
		2	0.470	0.4758	0.0058
		3	0.620	0.6134	0.0066
	1.5m	1	0.330	0.3378	0.0078
		2	0.470	0.4798	0.0098
		3	0.620	0.6231	0.0031
Type-2	1.1m	1	0.45	0.461	0.011
		2	1.50	1.579	0.079
		3	2.60	2.633	0.033
	1.3m	1	0.45	0.397	0.053
		2	1.50	1.470	0.030
		3	2.60	2.678	0.078
	1.5m	1	0.45	0.527	0.077
		2	1.50	1.564	0.064
		3	2.60	2.673	0.073
Type-3	1.1m	1	1.15	1.195	0.045
		2	5.00	5.007	0.007
		3	9.05	8.997	0.053
	1.3m	1	1.15	1.203	0.053
		2	5.00	5.072	0.072
		3	9.05	9.081	0.031
	1.5m	1	1.15	1.085	0.065
		2	5.00	5.033	0.033
		3	9.05	9.130	0.080

2) Discussion on different instrument heights: In this part, experiments were designed for type-1∼type-3 analog instruments to discuss the robustness of MPRA to different instrument heights. To obtain comparative experimental results, 1.1m, 1.3m, 1.5m heights (*i.e.* the NO.1, NO.2, and NO.3 locations shown in Fig. 20(a)) were adopted to install analog instruments of the same type which need to be read. To follow the variable-controlling principle, this part of experiments were all developed in the indoor experimental scene shown in Fig 20(a) to maintain constant illumination distributions.

In the experiments, for each type of instrument, we designed three inspection trials for the MPRA-based analog instrument reading. In each trial, the instruments of same type installed at 1.1m, 1.3m, 1.5m heights were read in order.

The detailed experimental configurations are presented in Table VIII, as well as the experimental results. As can be calculated, for type-1 instruments installed at 1.1m, 1.3m, and 1.5m heights, the mean absolute reading errors of MPRA were respectively 0.0043, 0.0060, and 0.0069 mA (merely 21.5%, 30%, and 34.5% of the unit scale value of 0.020 mA). For type-2 instrument installed at 1.1m, 1.3m, and 1.5m heights, those errors were respectively 0.041, 0.054, and 0.071 mA (merely 20.5%, 27%, and 35.5% of the unit scale value of 0.20 mA). And for type-3 instrument installed at 1.1m, 1.3m, and 1.5m heights, those errors were respectively 0.035, 0.052, 0.059 mA (merely 17.5%, 26%, and 29.5% of the unit scale value of 0.20 mA). The mean absolute reading errors of MPRA increased slightly when the instruments were installed higher. Meanwhile, all the errors in Table VIII never exceeded half of the corresponding unit scale values. As a result, we can conclude that the MPRA was robust to different instrument heights.

In summary, our automatic reading system using the FCAA and MPRA was effective to different types of analog instruments, robust to illumination changes and different instrument heights, and precise in the process of analog instrument reading. During the robot-based inspection, this system performed much better than those discussed on only one type of analog instruments [2]–[16], while it performed much better than human eyes on the reading accuracy.

VI. Conclusion

In this paper, concerning about the automatic reading of different types of analog instruments, an novel automatic analog instrument reading system using computer vision and an inspection robot has been presented. Through integrated application of the proposed FCAA and MPRA, the PTZ camera can be aligned to the instrument to be read, and the instrument reading can be then accurately calculated by removing the influence of the instrument deformation. In the process of FCAA, the location of the instrument to be read is first detected by an improved YOLOv3-tiny-dw instrument detector. Then, the camera deviation angles are calculated. Finally, the camera is aligned to the instrument by visual servo PTZ adjustments. In the process of MPRA, the pointer pixels are first detected by Hough Transform. Then, those pixels are reconstructed onto the front-view instrument scale region to remove the influence of instrument deformation. Finally, the reading is acquired based on the polar transform. Experimental results indicated that our system achieved high speed, high precision, and high success rate during the camera alignment process. Meanwhile, this system was capable of automatic reading of different types of analog instruments, with a reading accuracy higher than that achieved by human eyes. More importantly, the robustness to illumination changes and different instrument heights of our system was also confirmed, both in indoor and outdoor scenes. It can meet the application requirements of both indoor and outdoor robot-based inspections.

References

[1] S. Lu, Y. Zhang, and J. Su, "Mobile robot for power substation inspection: a survey," *IEEE/CAA Journal of Automatica Sinica*, pp. 1–18, 2018.

[2] W. Zheng, H. Yin, A. Wang, P. Fu, and B. Liu, "Development of an automatic reading method and software for pointer instruments," in *2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS)*. IEEE, 2017, pp. 1–6.

[3] J. Wang, J. Huang, and R. Cheng, "Automatic reading system for analog instruments based on computer vision and inspection robot for power plant," in *2018 10th International Conference on Modelling, Identification and Control (ICMIC)*. IEEE, 2018, pp. 1–6.

[4] H. Jiale, L. En, T. Bingjie, and L. Ming, "Reading recognition method of analog measuring instruments based on improved hough transform," in *IEEE 2011 10th International Conference on Electronic Measurement & Instruments*, vol. 3. IEEE, 2011, pp. 337–340.

[5] X. Yue, Z. Min, X. Zhou, and P. Wang, "The research on auto-recognition method for analogy measuring instruments," in *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering*, vol. 2. IEEE, 2010, pp. 207–210.

[6] E. C. Alegria and A. C. Serra, "Automatic calibration of analog and digital measuring instruments using computer vision," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 1, pp. 94–99, 2000.

[7] P. Belan, S. Araujo, and A. Librantz, "Segmentation-free approaches of computer vision for automatic calibration of digital and analog instruments," *Measurement*, vol. 46, no. 1, pp. 177–184, 2013.

[8] Z. A. Jaffery and A. K. Dubey, "Architecture of noninvasive real time visual monitoring system for dial type measuring instrument," *IEEE Sensors Journal*, vol. 13, no. 4, pp. 1236–1244, 2013.

[9] C. Zheng, S. Wang, Y. Zhang, P. Zhang, and Y. Zhao, "A robust and automatic recognition system of analog instruments in power system by using computer vision," *Measurement*, vol. 92, pp. 413–420, 2016.

[10] B. Yang, G. Lin, and W. Zhang, "Auto-recognition method for pointer-type meter based on binocular vision." *JCP*, vol. 9, no. 4, pp. 787–793, 2014.

[11] R. Sablatnig and W. G. Kropatsch, "Automatic reading of analog display instruments," in *Proceedings of 12th International Conference on Pattern Recognition*, vol. 1. IEEE, 1994, pp. 794–797.

[12] H. Fang, Xiaoxiao Cui, Liangkun Cui, and Yuanxin Chen, "An adapted visual servo algorithm for substation equipment inspection robot," in *2016 4th International Conference on Applied Robotics for the Power Industry (CARPI)*, Oct 2016, pp. 1–5.

[13] L. Li, B. Wang, B. Li, P. Xiao, W. Wang, and Y. Li, "The application of image based visual servo control system for smart guard," in *2013 10th IEEE International Conference on Control and Automation (ICCA)*. IEEE, 2013, pp. 1342–1345.

[14] X. Mai, W. Li, Y. Huang, and Y. Yang, "An automatic meter reading method based on one-dimensional measuring curve mapping," in *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*. IEEE, 2018, pp. 69–73.

[15] J. Liu, M. Nie, H. Wu, and X. Mai, "An image-based accurate alignment for substation inspection robot," in *2018 International Conference on Power System Technology (POWERCON)*. IEEE, 2018, pp. 4113–4117.

[16] H. Song and G. Yin, "Accurately localize and recognize instruments with substation inspection robot in complex environments," *Sensors & Transducers*, vol. 174, no. 7, pp. 211–216, 2014.

[17] H. Y. Kim and S. A. De Araújo, "Grayscale template-matching invariant to rotation, scale, translation, brightness and contrast," in *Pacific-Rim Symposium on Image and Video Technology*. Springer, 2007, pp. 100–113.

[18] H. Y. Kim *et al.*, "Ciratefi: An rst-invariant template matching with extension to color images," *Integrated Computer-Aided Engineering*, vol. 18, no. 1, pp. 75–90, 2011.

[19] J. E. Bresenham, "Algorithm for computer control of a digital plotter," *IBM Systems Journal*, vol. 4, no. 1, pp. 25–30, 1965.

[20] J. Bresenham, "A linear algorithm for incremental digital display of circular arcs," *Communications of the ACM*, vol. 20, no. 2, pp. 100–106, 1977.

[21] R. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[23] J. Huang, Y. Wang, and T. Fukuda, "Set-membership-based fault detection and isolation for robotic assembly of electrical connectors," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 160–171, 2016.

[24] J. Huang, M. Zhang, S. Ri, C. Xiong, Z. Li, and Y. Kang, "High-order disturbance observer based sliding mode control for mobile wheeled inverted pendulum systems," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2030–2041, 2020.

[25] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European Conference on Computer Vision*. Springer, 2006, pp. 404–417.

[26] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[27] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[28] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[29] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[30] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte carlo localization for mobile robots," in *ICRA*, vol. 2, 1999, pp. 1322–1328.

[31] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[32] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[35] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263–7271.

[36] J. Redmon, "Darknet: Open source neural networks in c. `http://pjreddie.com/darknet/`," 2013–2016.

[37] H. Yuen, J. Princen, J. Illingworth, and J. Kittler, "Comparative study of hough transform methods for circle finding," *Image and Vision Computing*, vol. 8, no. 1, pp. 71–77, 1990.

[38] L. Siming, "Improved thinning algorithm of engineering drawings input and automatic recognition," *Computer Engineering*, vol. 29, no. 16, pp. 37–38,61, 2003.

[39] P. Viola, M. Jones *et al.*, "Rapid object detection using a boosted cascade of simple features," *CVPR*, vol. 1, pp. 511–518, 2001.

[40] J. Xu, Q. Wu, J. Zhang, and Z. Tang, "Fast and accurate human detection using a cascade of boosted ms-lbp features," *IEEE Signal Processing Letters*, vol. 19, no. 10, pp. 676–679, 2012.

[41] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition (CVPR'05)*, vol. 1. IEEE Computer Society, 2005, pp. 886–893.

[42] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[43] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[44] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2011.