Hindawi
Complexity
Volume 2021, Article ID 6853809, 12 pages
https://doi.org/10.1155/2021/6853809

Research Article

Privacy-Oriented Successive Approximation Image Position Follower Processing

Ying Miao,[1] **Danyang Shao**,[2,3] **and Zhimin Yan**[2,3]

[1]*School of Economics and Management, Shenyang Aerospace University, Shenyang, Liaoning 110136, China*
[2]*SLZY (Shenyang) Hi-Tech Co., Ltd., Shenfu Reform and Innovation Demonstration Zone, Shenyang, Liaoning 110172, China*
[3]*UAV Division of Liaoning ITRI, Shenfu Reform and Innovation Demonstration Zone, Shenyang, Liaoning 110172, China*

Correspondence should be addressed to Ying Miao; miaoying@email.sau.edu.cn

Received 26 April 2021; Revised 24 May 2021; Accepted 25 May 2021; Published 7 June 2021

Academic Editor: Zhihan Lv

In this paper, we analyze the location-following processing of the image by successive approximation with the need for directed privacy. To solve the detection problem of moving the human body in the dynamic background, the motion target detection module integrates the two ideas of feature information detection and human body model segmentation detection and combines the deep learning framework to complete the detection of the human body by detecting the feature points of key parts of the human body. The detection of human key points depends on the human pose estimation algorithm, so the research in this paper is based on the bottom-up model in the multiperson pose estimation method; firstly, all the human key points in the image are detected by feature extraction through the convolutional neural network, and then the accurate labelling of human key points is achieved by using the heat map and offset fusion optimization method in the feature point confidence map prediction, and finally, the human body detection results are obtained. In the study of the correlation algorithm, this paper combines the HOG feature extraction of the KCF algorithm and the scale filter of the DSST algorithm to form a fusion correlation filter based on the principle study of the MOSSE correlation filter. The algorithm solves the problems of lack of scale estimation of KCF algorithm and low real-time rate of DSST algorithm and improves the tracking accuracy while ensuring the real-time performance of the algorithm.

1. Introduction

Along with the continuous development of information acquisition devices and the rapid progress of science and technology, personal identification technology has attracted more and more widespread attention. Whether it is national public security service and supervision, military reconnaissance, or personal health monitoring, family monitoring, etc., it has important value and significance. Traditional identification technologies based on ID cards, passwords, etc. have been difficult to meet the rapidly developing needs of modern society. Iris, fingerprint, face, gait, footstep signal, and voice pattern belong to the category of biometric identification [1]. Biometric features belong to the inherent characteristics of the human body, which are neither easy to lose nor forgotten, and, at the same time, have the feature of being portable. Biometric features of the human body are mainly divided into physiological features (iris, fingerprint, face, voice print, etc.), which are relatively stable. Behavioral features (gait, footsteps signals, etc.), which have a certain physiological basis, can reflect the psychological changes of a person. Biometric technologies based on iris, fingerprint, face, gait, and voice pattern have been applied and even promoted at present [2]. Iris recognition is a kind of personal identification by capturing human iris images and analyzing iris features through a collection instrument. The iris structure, as well as the characteristics, will not change from the birth of humans and is unique. Iris recognition is also used in a wide range of applications, such as access control, safes, and other security devices, but iris recognition requires

human subjective involvement. Covert iris recognition does not require the cooperation of the target person, but the quality of the obtained iris image is low. Fingerprint recognition technology can identify individuals based on unique information such as the pattern and detailed features of human fingerprints.

Fingerprints are easy to obtain and use, reliable, miniaturized, inexpensive, and cost-effective. However, fingerprints are affected by skin and humidity, and fingerprints can be easily copied, for example, by candles or playday [3]. Face recognition is now widely used and relatively mature, through the captured facial image or video extraction and analysis of the characteristics of the entire face and its contours to achieve personal recognition. However, face image data needs to be acquired through cameras or cams, which are not easily accessible on some specific occasions, such as perimeters and private places. Also, when the wrongdoers carry out illegal and criminal activities, they will intentionally avoid the camera. Setting the motion target as the human body, the detection and tracking of moving human body have important application value in the monitoring of static backgrounds such as shopping malls and roads. Similarly, when the application environment is no longer a single, simple static background, the previous detection method is no longer applicable, and when the vision sensor is installed on the mobile robot platform, the background of the real-time screen becomes dynamic with the movement of the mobile robot [4]. The detection and tracking of moving human bodies need to be studied, so that the detection and tracking of moving targets are not limited by the change of scenes. For example, in the field of indoor and outdoor navigation, mobile robots can use this technology to track and monitor targets of interest; in the field of service-oriented robots, this technology can be used to allow robots to follow targets set by users and provide services or monitor the status of the targets at any time, reducing the human labour. In the field of service-oriented robots, the technology can be used to allow the robot to follow the target set by the user and provide services or monitor the status of the target at any time, reducing human labor while improving the quality of service and better realizing the change of life through technological development.

Vision technology occupies an extremely important position in the artificial intelligence of mobile robots. The research in this paper is based on the support of a horizontal scientific research project, and vision-based motion target detection and tracking are used as one of the many functional modules of mobile robots to realize the tracking of motion targets by mobile robots in motion. After the motion target is determined to be a human body, the first step is to detect the moving human body from the real-time image received by the host computer. In this paper, we start from the idea of feature detection and variable part detection, based on the bottom-up model, to detect the feature points of key parts of the human body, and finally output the detection results after the feature point confidence map prediction, to detect the moving human body. Based on the understanding of several typical correlation target tracking algorithms, we analyze and compare the advantages and disadvantages of each, fuse the position filter of the KCF algorithm and the scale filter of DSST algorithm based on the theoretical basis of the MOSSE correlation filter, and conduct an in-depth study of the fusion algorithm with practical application scenarios, and realize the tracking of the dynamic human body based on this algorithm.

2. Related Work

Gao et al. built a wilderness search and rescue robot system, in which the operator locally controls a remote mechanical cart for search and rescue tasks [5]. Augmented reality is a new technology developed based on virtual reality technology. Virtual reality technology requires a computer to generate a fully virtual space, which is extremely demanding for computers [6]. Augmented reality technology is based on the superimposition of virtual information or models on real scenes, which not only reduces the complex modelling process, but also enhances human involvement and sense of presence [7]. Nowadays, augmented reality has a wide range of applications in industrial maintenance, medical care, entertainment, etc. van der Helm et al. point out that footstep vibration signal detection systems are extremely important in national security and military applications, but the most popular commercial seismic sensors and geophones are difficult to achieve satisfactory results in footstep vibration signal detection applications, because of the poor low-frequency signal response of current instrumentation, resulting in a reduced target detection range, unsatisfactory sensitivity thresholds, and reduced target detection range [8], low damping and accuracy response, unsatisfactory detection of distant targets in high-noise environments, low noise immunity leading to unreliable operation of instrumentation, large size and high price, hindering the widespread use and promotion of seismic sensors. Through the principle of binocular stereo vision measurement, according to the three markers in the static image, obvious features can be obtained from the image as a reference to analyze the image in real-time to calculate the Jacobi matrix of the target image combined with the formula of visual localization measurement to predict the three-dimensional coordinates of the target, the distance, and the next direction of movement to achieve adaptive tracking of moving targets [9]. Since this algorithm does not require the internal and external parameters of the camera, as well as the initial position of the target, the measurement error is large and only approximately gives information about the position of the target in the next frame, but the algorithm is more robust [10].

The study in [11] discussed the problem of target-oriented control of submersible vehicles to follow the target by controlling the motion of the visual sensors. The study in [12] considered target following control methods for targets in the case of constant acceleration and variable acceleration rotation. In [13], a motion target detection and tracking system were constructed using an industrial camera with adaptive appearance model and vision algorithms, and a Kalman prediction-based tracker is used to achieve target tracking. The study in [14] proposed a vision-based SLAM

method PTAM to estimate UAV attitude, using multiview geometric constraints to detect moving objects in UAV images, making full use of the unrestricted and continuous attitude changes of UAVs, and detecting dynamic objects by extracting feature points, which was tested using a quadrotor UAV platform and successfully detected targets for emotionally unstable small UAVs. The study in [15] proposes a method for safe flight in areas, where navigation information is weak or even inaccessible, first using UAV monocular vision for real-time localization and map building, and ground robots using the ORB_SLAM2 (ORiented Brief SLAM2) system to build a global map and reestablish the 3D environment, with cooperative ground-air cooperation for autonomous navigation.

Intelligent robot detection and tracking target direction is a relatively popular research direction, by carrying vision sensors to obtain the location of the target, the use of tracker to achieve the tracking of the target. And there is relatively little research on UAVs using vision methods to achieve positioning and achieve flight, tracking, and traversal indoors, which can be combined with these directions to synthesize for GPS-free device UAVs using vision sensors to achieve autonomous positioning, detection, tracking, and traversal of obstacles indoors. The door frame target tracking method with an unknown motion state is proposed. In the whole system, the motion position of the target is obtained by autonomous positioning of the UAV platform and spatial positioning of the target, and the Kalman filter method is used to establish the target model, predict and estimate the position of the random motion target, complete the motion tracking of the target, and ensure the accurate traversal of the UAV on the moving door frame obstacle. The feature-based matching method uses some feature points that reflect the characteristics of the image itself to perform stereo matching and solve the problem by emphasizing the structural information in the image. The feature point-based matching algorithm first extracts the feature point information of the left and right views, and after matching the left and right view feature points, the coordinate differences of these feature points are used to obtain the parallax information. The feature-based matching method has fast matching speed and robustness to noise and illumination changes, but the parallax map obtained by the feature matching method is sparse due to the sparsity and discontinuity of the image feature points themselves. In this paper, the feature-based matching method is selected according to the requirements. In the target depth measurement application, only a pair of matching points is theoretically needed to obtain the depth of the target, and a dense parallax map is not required. The feature-based matching method has fast matching speed and good real-time performance, so it is suitable for the application of target depth calculation in this paper. First, use a larger initial window to do the mean filtering (the integration graph realizes the mean filtering, not much introduction, and you can refer to our previous blog) and assign the holes in the large area. Then, the next time you filter, reduce the window size to half of the original one, filter again using the original integral map, assign smaller holes (overwrite the original value), and so on, until the window size becomes 3×3, then stop filtering at this time, and get the final result.

3. Analysis of Privacy Successive Approximation Type Image Position Follower Processing

3.1. Approximate Image Position Follower Processing Design. Background subtraction is extremely similar to the interframe difference method, in which each frame is differenced from the background frame image to obtain the motion target. Background subtraction detects the motion target by modelling and updating the background model. Like interframe differencing, background subtraction requires a series of processing of the obtained target area after the differencing operation to obtain an accurate motion target [16]. The background subtraction method is used to detect the target in four parts, background modelling, background updating, differential detection of the target, and subsequent processing, in which the background model needs to be mathematically analyzed to build a mathematical model that can characterize the background, and the background model plays a key role in the whole detection process and will directly affect the effect of target detection.

The principle of background subtraction is shown in equations (1) and (2), where $f_n(x, y)$ is the pixel value at the point with coordinates (x, y) in the current frame $B_n(x, y)$ pixel value of the corresponding point in the established background image, D_n is the difference image, R_n' is the binarized image when the grey value of the pixel point in the binarized image is 255, which means that this point belongs to the target object, and the grey value of the background point is 0, the same for R_n'. By performing the connected domain analysis, we can obtain an image containing the complete motion target R_n'.

$$D_n(x, y) = \left| f_n(x, y) + B_n(x, y) \right|, \tag{1}$$

$$R_n'(x, y) = \begin{cases} 255, & D_n(x, y) \le T, \\ 0, & \text{others.} \end{cases} \tag{2}$$

In background subtraction, the background is updated in real-time to obtain the background image at the current moment, because the difference operation between the current frame and the background image is required, and equation (3) represents the update of the background model, where $B_{n-1}(x, y)$ represents the background image at frame $n-1$, and $\alpha \in [0, 1]$ is the model update rate.

$$B_n(x, y) = B_{n-1}(x, y)(1 + \alpha) - f_n(x, y) \cdot \alpha. \tag{3}$$

In the actual application process of the background subtraction method, the algorithm is simple and easy to implement; in a static background, due to the difference between this method and the background model, static objects can be detected, and due to the update of the background model, this method has strong resistance. Interference ability: at the same time, it is possible to conduct in-depth research on the establishment and update of the background model, to realize the detection of the target in the complex environment including the dynamic background.

The advantage of the optical flow method for target detection is that it is not limited by the scene information to

get the location of the moving target, so it is also applicable in the dynamic background, where the camera is moving at the same time [17]. But, at the same time, the method also has disadvantages; firstly, it is sensitive to light and cannot be applied to scenes with changing light; secondly, the accurate detection of the target and the real-time of the detection cannot be satisfied at the same time, because of the large amount of computation in the optical flow method.

The camera device in this study is a monocular fisheye camera with an imaging pixel resolution of 640 ∗ 360. The monocular fisheye camera has a large field of view and can acquire a large field of view, while the pinhole model is different, where the light reflected from the object passes through the camera aperture to form an image at the back-end sensor, and the imaging model and imaging schematic of the pinhole are shown in Figure 1.

Camera planar imaging is a three-dimensional mapping process of the real scene, and the quality of the camera depends on the performance of the camera, postprocessing of the image, such as aberration correction, and the target world coordinates to solve the need for accurate camera parameters. Camera parameters are mainly internal and external parameters. The internal parameters of the camera are the physical parameters inherent to the camera itself, including the focal length, image centroid, and aberration. The external parameters are the positions of the camera body coordinate system relative to a real-world coordinate system [18]. In this paper, the camera internal and external parameters are solved by the function of calibrated fisheye camera in vision software, and the camera internal and external parameters are applied to the target localization part. As the camera in the process of production cannot ensure that the plane of light incidence for absolute compliance and natural imaging changes, called image aberrations, image aberrations can be divided into radial aberrations and tangential aberrations. Radial aberration refers to the parallel direction of the imaging radius, the image position in the imaging position shift; and tangential aberration refers to the direction parallel to the plane of light incidence tangent, the image position in the imaging position of the pixel point shift.

The expression for the image height r of the model is r, where the parameters on the right side of the expression include the focal length f of the fisheye camera and the angle of incidence rag of the light. Assuming that the pixel coordinates of the fisheye image are (x, y), the plane of the target image is the plane where $P\ (X, Y, Z)$ is located in Figure 1, which is regarded as $Z = R$, and the expression R is the radius of the fisheye image in this study. From the schematic diagram in Figure 1, it is obtained that

$$\begin{aligned} x &= r\ \cos\ \beta, \\ y &= r\ \sin\ \beta, \\ r &= f\beta. \end{aligned} \tag{4}$$

After the 3D coordinates of the spatial point are obtained, we need to transform this spatial point through the spherical perspective, and the spatial target point will be projected through the sphere to the imaging plane, which will intersect with the virtual sphere in the process of light incidence at the point P_1, and then mapped to the imaging plane. According to the relationship of similar triangles, considering the triangle in the figure OPP_3' with OP_3P_3', it is known that

$$\begin{aligned} \frac{X_1}{X} &= \frac{Y_1}{Y} = \frac{Z_1}{Z}, \\ X &= \frac{ARXX_1}{Z_1}, \\ Y &= \frac{ARYY_1}{Z_1}, \\ Z &= R. \end{aligned} \tag{5}$$

The plane of the corrected target image in this paper is the $Z = R$ plane, and this study first assumes that the coordinates of the target's position on the image are (u, v) after the image distortion correction. The expression of the coordinate's parameter is shown in

$$u = R\ \tan\left(\frac{\sqrt{x^2 + y^2 + z^2}}{f}\right)\sin\left(\arctan\left(\frac{y}{x}\right)\right), \tag{6}$$

$$v = R\ \tan\left(\frac{\sqrt{x^2 + y^2 + z^2}}{f}\right)\cos\left(\arctan\left(\frac{x}{y}\right)\right). \tag{7}$$

According to the above study, the radius of the fisheye image will be adjusted according to the good or bad image condition to achieve the best result of image aberration correction. The whole aberration correction process has more calculation steps, and to reduce the number of loop iterations, the matrix form operation can be used to improve the operation efficiency of the algorithm and finally complete the aberration correction of the fisheye camera.

The problems of the binary image obtained after the color feature extraction are completed, mainly including noise, disconnected image regions, and the existence of more pretzel noise points. To address the existing problems, this paper proposes solving them by using image operations of expansion and erosion, which connect adjacent elements in the image, then find out the obvious maximal value region or minimal value region in the image, and also get the gradient and other information of the image. The expansion method in morphological processing is to perform a specific structural element operation on the input image based on the maximum value of the image related neighborhood and assign this maximum value to the pixel specified at the reference point, which will cause the highlighted area in the image to gradually grow and fill the target internal void. The erosion operation is a specific structure element operation on the input image based on the minimum value of the image's associated neighborhood, and the erosion process gradually makes the highlighted area smaller, and the amount of reduction is determined by the structure element,

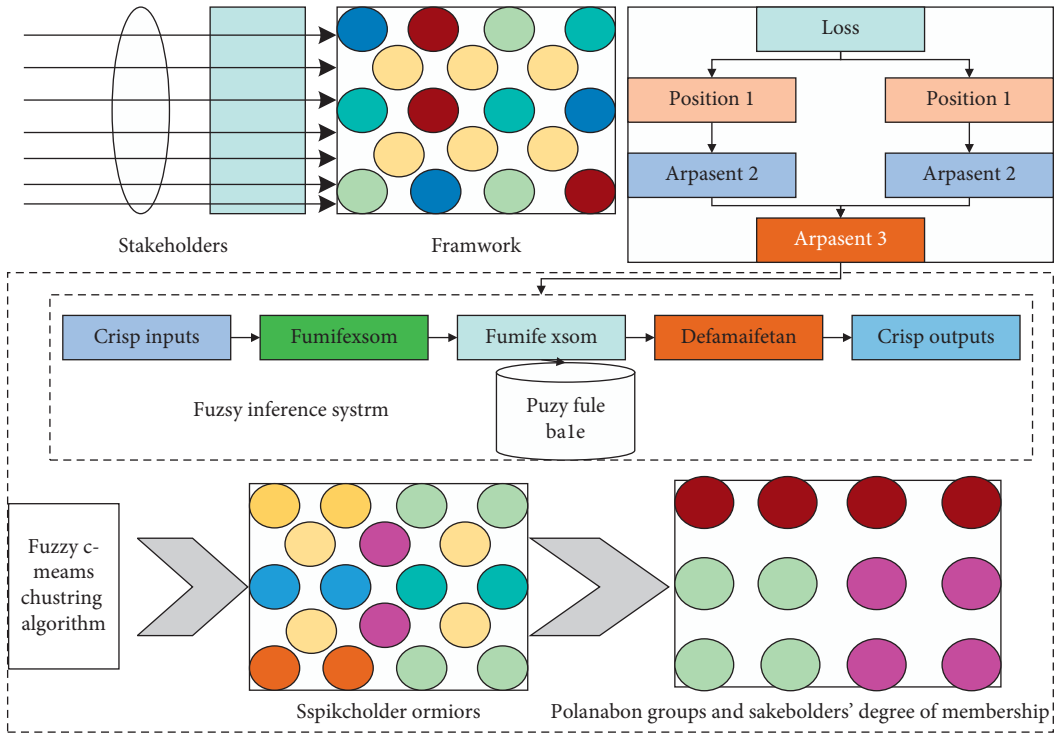

Figure 1: Imaging principle.

so that the target edges are simplified, and the pretzel noise points are removed. The morphological operation logically transforms each pixel of the image by the structural element and the neighborhood, and finally, the pixel coordinate transformed image is obtained.

During human walking, each person has a unique walking style based on individual biometric characteristics, including the position of the center of gravity, the size of the stride, the way the foot contacts the ground, the difference in shoes, and mood changes. As shown in Figure 2, four sets of signals originated from two testers with leather and sneakers on their feet, respectively, and each segment was selected from the step of foot vibration signal closest to the G5 detector. (a) is the two-step foot vibration signal generated by tester F1 with leather shoes; (b) is the two-step foot vibration signal generated by tester M1 with leather shoes; it can be seen that the amplitude of the foot vibration signal generated by M1 is larger than that of F1, because its weight is larger than that of F1: (c) is the two-step foot vibration signal generated by tester F1 with sneakers; it can be seen that the amplitude of the signal generated by sneakers is much smaller than that of leather shoes; (d) is the two-step foot vibration signal generated by tester F1 with sneakers; it can be seen that the amplitude of the signal generated by sneakers is much smaller than that of leather shoes. From the four sets of signals, it can be seen that there are differences in the foot vibration signals of the same person with two shoe types, different people with the same shoe type and different shoe types, such as amplitude and duration; there is a similarity between the two-foot vibration signals generated by the same person with leather shoes or sports shoes. There is a similarity between two footsteps of the same person wearing leather shoes or sports shoes. Therefore, it is possible to identify individuals by using the foot vibration signals generated during human walking.

First, the result of the detection module is used as the input of the tracking module to initialize the tracking algorithm, then the position of the target in the next frame is predicted by particle filtering or sliding window motion model, and candidate regions are generated; then, traditional features or depth features are extracted from these candidate regions; whether the predicted region is the target is verified according to feature matching and discriminative or generative appearance model, and if the region is a target, the target with the best state in the current frame is selected as the result for target localization. In the process of tracking, the appearance model is updated according to the current tracking result in each frame, so that the tracking algorithm can adapt to the changes since the shape and angle of the moving target will change.

3.2. Privacy Successive Approximation Image System Design. Based on the above discussion, this paper can realize the dynamic mapping relationship between the camera coordinate system H, the robot global coordinate system (robot arm base coordinate system) R, and the tool coordinate system E of the end-effector hand claw, as shown in Figure 3.

A path does not take time into account and is defined only as a specific sequence of robot configurations. The robot's end moves from the initial point to the middle point and then to the endpoint. The intermediate process of its movement is a path, which is different from a trajectory in that it is independent of time and has no relationship with it

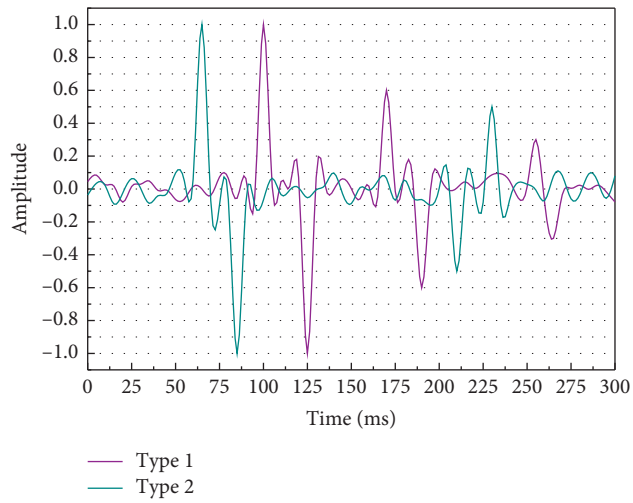

Figure 2: Vibration events.

when each part of the path is reached, so the path of the robot's hand claw from point A to point B and then to point C is always the same. However, if the speed through each part of the path is not the same, the trajectory is not the same. So, even if the robot passes through the same points, but the elapsed time is different, the points on the path and the trajectory of the robot's hand claw may be different. Trajectory planning focuses on the relationship between variables such as displacement, velocity, and acceleration of the robot trajectory in the joint space, which is to find out the intended trajectory profile of the robot according to the current demand. Trajectory planning generally involves "interpolation" of polynomial functions to slowly move the paths of the robot's end-effectors at their initial and end positions closer to a given path and to generate a series of points over time that can be used and controlled by the PC [19]. Currently, the main methods for trajectory planning are cubic polynomial interpolation, quintuple polynomial interpolation, and spline curves. Since the robot end-effector morphology is easier to observe in the Cartesian coordinate system, the path endpoints are generally given in the Cartesian coordinate system.

Through the above analysis of the time domain and frequency domain of the foot vibration signal, it can be seen that the features in the time domain and frequency domain are the reflection and manifestation of the global characteristics of the foot vibration signal, but since the foot vibration signal is a nonstationary signal, it is not enough to consider its global characteristics only. Therefore, this section analyzes the local characteristics of the footstep vibration signal, extracts its time-frequency characteristics, observes its local refinement information in the time and frequency domains, finds the time distribution of different frequency components in different footstep vibration signals, and distinguishes in detail the footstep vibration signals generated by different individual people walking. Due to the influence of ground friction resistance, etc., in the process of actual operation, the actual wheel speed of the left and right wheels of the trolley is not the same as the speed originally intended to be controlled, so when it is needed to move forward in a straight line, there is always a deviation, so a control method is needed to make it achieve the desired control effect. In the controlling idea, PID is to compare the current real speed collection result of the wheel with the desired control speed in real-time and make up for less and subtract more, so that the ideal control can be achieved. After the above analysis, after determining the control strategy of the trolley, the control method of the trolley is studied. The control of the trolley is mainly to achieve the control of the angular velocity and linear velocity of the trolley wheels. The forward and backward movements of the mobile robot can be accomplished by controlling the linear velocity, while the rotating and turning around movements of the mobile robot can be accomplished by controlling the angular velocity. From the hardware point of view, when controlling the linear speed and angular speed of the trolley, the actual speed of the left and right wheels is not the same due to external factors such as ground friction, so when you want to control the trolley to go straight, it will not go straight, so there will be errors; to avoid this situation, you need to join the PID control; the idea of PID control is to collect the real speed of the wheels back in real-time. The idea of PID control is to collect the real speed of the wheel in real-time and compare it with the speed you want to control, and more is reduced to make up for it, as shown in Figure 4.

The main control core board is responsible for the acquisition of binocular camera video data, the direct control of the gimbal servo, and the bidirectional data transmission with the local terminal in this system. The hardware part of the main control board consists of a power management part, microcontroller minimum system, USB2.0 conversion module, and wireless transmission module. The power management part uses a single integrated step-down switching power converter TD1410 with an internal power MOSFET, which has a wide voltage input range and can continuously output a 2 A current. Considering that the components on the main control board are all 3.3 V powered, the microcontroller is responsible for the rudder

control in the system in the two degrees of freedom direction of the gimbal. The STM32F101 series microcontroller with Cortex-M3 core is used in this system [20]. Since the camera data transmission in this paper uses USB protocol, a USB2.0 conversion module is added between the wireless transmission module WRTnode and the camera, which is responsible for the data conversion between the camera and the scoreboard. The wireless transmission module is responsible for the remote communication between the remote end and the local computer, and the real-time acquisition of binocular camera video data. The system chooses WRTnode2P from Beijing Net link Technology Company as the wireless transmission module, which has the advantages of small size, low cost, and strong transmission capability.

4. Analysis of Results

Firstly, for the UAV autonomous localization experiments, the improved PTAM algorithm is adopted as the UAV visual localization algorithm in the method, the SLAM localization function based on monocular vision is realized, and the hovering and dynamic fixed waypoint flight experiments are conducted, respectively. For the hovering experiment, the UAV was used to hover for a long time with the aid of visual positioning, and the indoor positioning system was used to detect the hovering position of the UAV, observe the stability of its hovering, and derive the hovering results. For the dynamic fixed waypoint flight test, the visual SLAM completes the construction of a map of the environment, the UAV flies based on the environmental map, sets the flight waypoints for the UAV, respectively, detects the flight position of the UAV using the motion capture system, compares the set waypoint position, and observes the flight of the UAV, mainly comparing the difference between the real value detected by the motion capture system and the set waypoint position. The results of the two experiments are shown in Figure 5, respectively.

In Figure 5, the UAV is flown in a fixed hover at a certain position to check the hovering and positioning performance of the UAV. In the experiment, the UAV is made to keep the target position for a certain time to verify the stability of single-point hovering, and the trajectory estimated by the visual SLAM algorithm is the red line, and the UAV trajectory captured by the motion capture system is the blue line. According to the results, the mean square errors of the three dimensions are 0.8158 cm, 0.8389 cm, and 0.6904 cm, respectively. The results show that the errors are extremely small, indicating that the visual SLAM algorithm has high accuracy in hovering flight. In Figure 5, the UAV flies according to the set waypoints to check the dynamic flight performance of the UAV. The UAV flies at three different altitudes, and the trajectory estimated by the visual SLAM algorithm is the red line, and the trajectory of the UAV captured by the motion capture system is the blue line. According to the results, the mean square errors of the three dimensions are 8.3615 cm, 7.4998 cm, and 1.5521 cm, which show that the errors are at the centimeter level, indicating that the vision algorithm has high accuracy even under dynamic flight conditions and can meet the UAV's localization and navigation requirements.

According to the detection result output Figure 6, on the image, the position of the four corner points of the inner contour of the door frame P_{uv} can be accurately derived, using the pixel coordinate system door frame vertex imaging position P_{uv} and the internal reference matrix K together to solve $|X_c, Y_c, 1|^T$ the coordinates of the normalized plane, the normalized coordinate data as the constraint equation of the target model known conditions, the relative position of the target is solved, using this method to solve the monocular scale uncertainty problem. Then, the absolute position of the target is solved jointly using the relative position fused with the UAV positional information, and the center of mass of the four points is the final target point, and the absolute position of the center of mass of the door frame is obtained by doing the processing of the absolute position of the four points, and the comparison of the distance solving results of the center of mass of the target is shown in Figure 6, the vertical coordinate represents the distance of the UAV relative to the target, the horizontal coordinate is the time stamp, the blue line in the figure is the motion capture system to get the blue line in the figure as the real value of the motion capture system, and the red line is the measured value calculated by the algorithm.

The comparison between the true and measured values of each distance of the target relative to the UAV is shown in Figure 6. Using the motion capture system, the absolute position of the target is measured and counted as the true value, and the monocular scale solved value is counted as the measured value. This is probably due to the deviation of the higher-order solution method in terms of the iterative cut-off conditions and the position of the target on the image in pixel-level units, but the errors of both the true and measured values are within 3%, which can be further reduced considering the flight stability of the UAV.

For the study of the world position solution of the door frame target in the visual construction map, the target positioning experiment shows that the positioning algorithm can solve the absolute position of the target with high accuracy, and multiple distance measurement experimental results show that it can accurately calculate the position of the target at different distances, which not only solves the problem of monocular scale uncertainty, but also ensures the accuracy of the input measurement value of the UAV at every moment of the tracking link, which provides the basis for it. This not only solves the monocular scale uncertainty problem, but also ensures the accuracy of the UAV input measurements at each moment of the tracking session, which lays the foundation for subsequent stable tracking and traversing.

The experiments were conducted for three cases of the flat display, three-dimensional display, and three-dimensional display combined with depth cueing, and five distances were set, 2.4 m, 3.0 m, 3.6 m, 4.2 m, and 4.8 m. The test was repeated five times for each distance, and the actual position of the trolley was recorded with the sign position, and the time for the operator to control the movement of the trolley to the sign position was also recorded for five

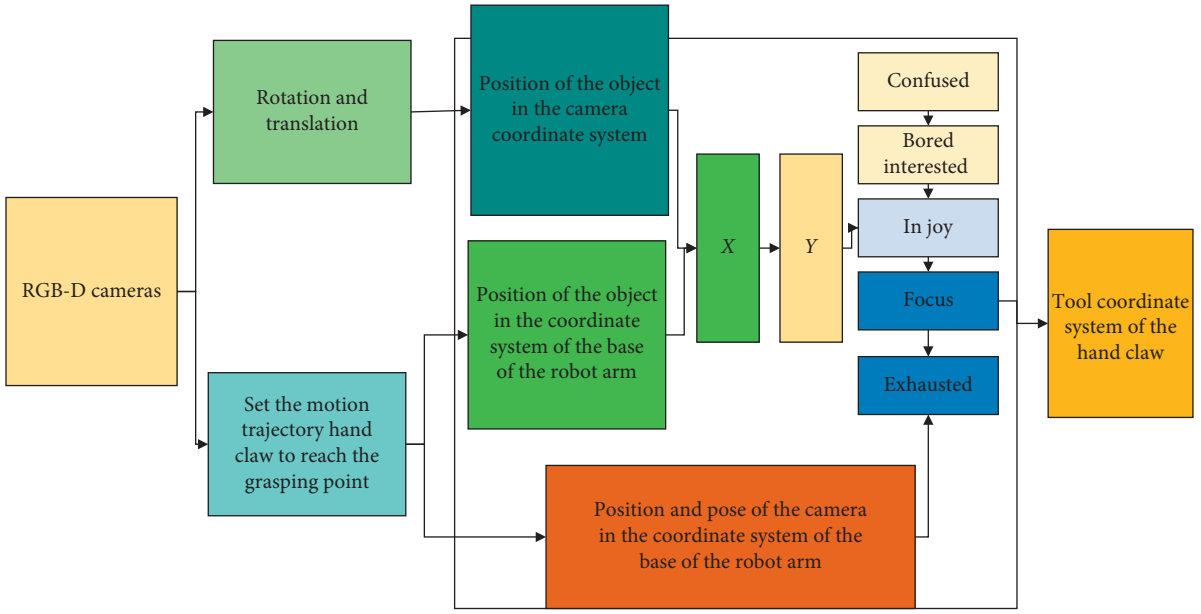

Figure 3: Coordinate conversion flow chart.

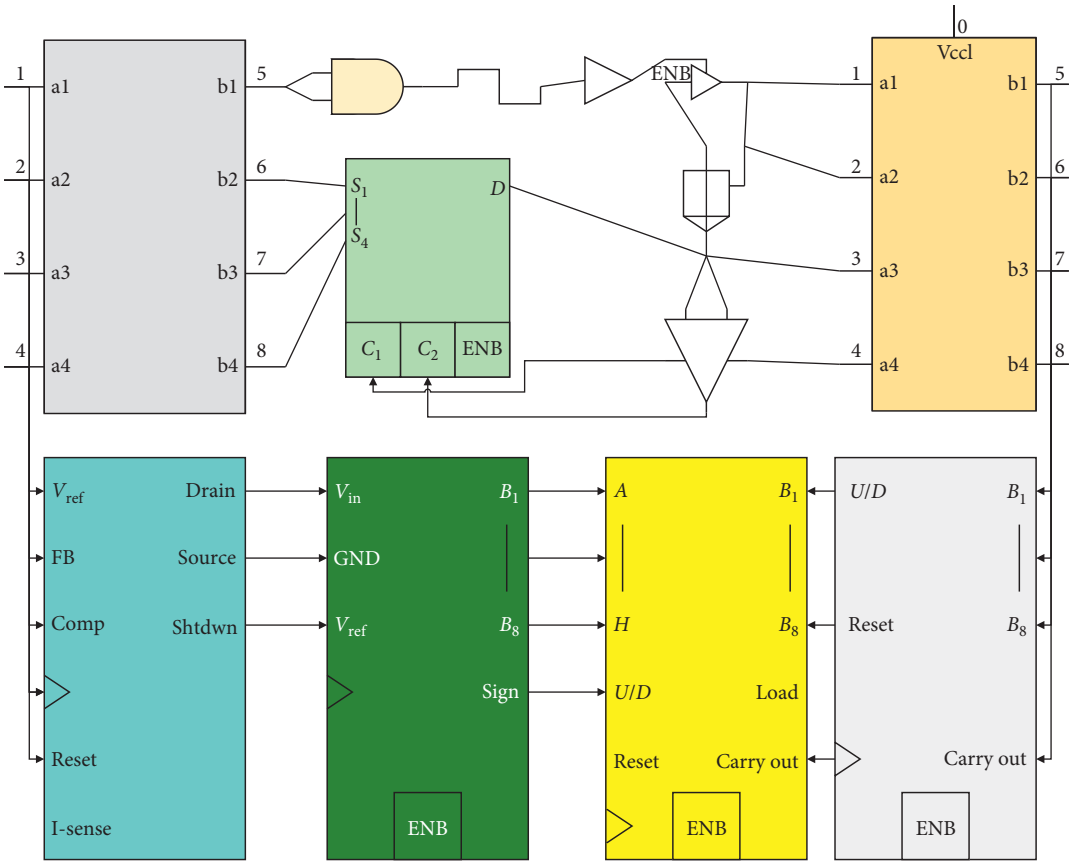

Figure 4: 7 STM32 minimum system schematic.

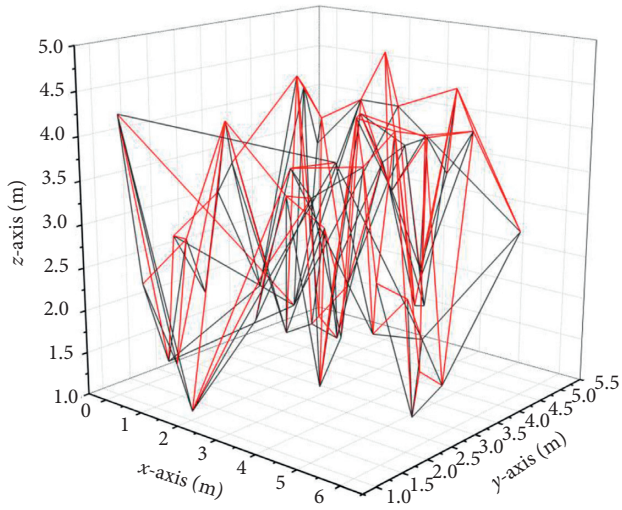

Figure 5: Comparison results between estimated and real values of hovering.

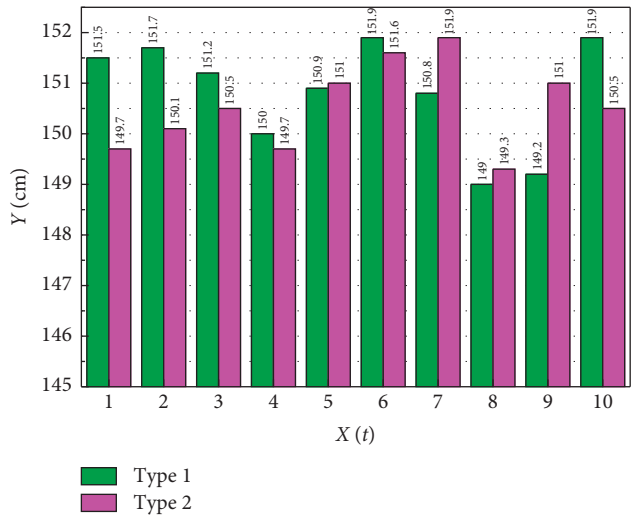

Figure 6: Comparison of real and measured values at different target center-of-mass distances.

experimenters, totaling 125 sets of data. Among them, the original data record of experimental subject 1 is shown in Figure 7.

The relationship between the coordinates of the base and the actual coordinates obtained from the experiment is shown in Figure 8. In the range the robot arm can grasp, the sum of the absolute values of the three-dimensional directional error is positively correlated with the distance of the target object relative to the base, and the farther the distance, the larger the error, but the error can be controlled within 2 cm. There are three main sources of error: the error of the camera placement, the error of the robot arm's kinematics and mechanical dimensions, and the measurement error of the camera itself.

The experimental results show that the coordinate conversion method proposed in this paper makes the grasping success rate high, and the success rate of the hand claw center point reaching the grasping point is almost 100% within the tolerance range of 2 cm, and the highest grasping success rate of 85% is reached at the distance of 130 cm, which can meet the requirements of the explosive detonation robot to grasp explosives. To further verify the effectiveness of the system in this paper, under the same conditions, the traditional fixed vision method is used for grasping experiments, within the acceptable error range, the system in this paper is compared with the fixed vision of the robot arm grasping system, and it not only solves the problem of fixed camera vision obscured by the robot arm and shooting clarity, but also effectively improves the grasping success rate. Firstly, common image preprocessing and contour extraction methods are introduced and selected, and edge detection, i.e., contour extraction method, is analyzed, and then a contour centroid acquisition method based on the snake algorithm for contour extraction is proposed, and the

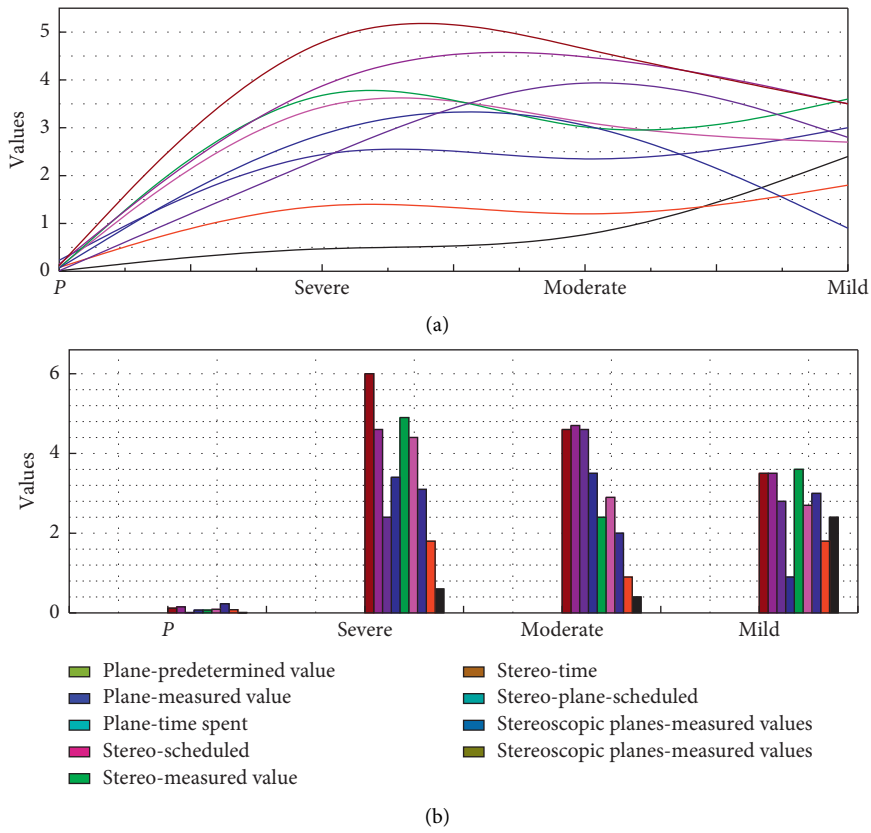

Figure 7: Test subject 1 data.

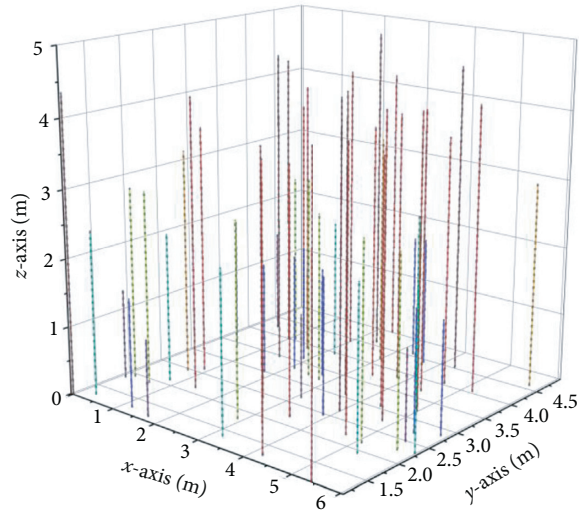

Figure 8: Coordinates and error diagram.

pixel points within the contour are transformed into contour centroids for autonomous grasping experiments. Two target objects, water cups, and square boxes are selected to determine the performance characteristics of the system. The experimental results show that the grasping success rate of the improved grasping system based on the method proposed in this chapter is further improved, and the grasping success rate of the water cup is higher than that of the square box due to the robot's hand claw configuration. This further improves the grasping success rate.

5. Conclusion

The autonomous grasping system of the detonation robot based on follow-me vision proposed in this paper can ensure reliable and stable operation, which provides a theoretical basis for the subsequent development of algorithms for target detection, object recognition and deep learning, and real-time feedback in trajectory planning, and has certain theoretical and practical significance. The results show that the coordinate conversion method proposed in this paper makes the grasping success rate high, and the success rate of the center point of the hand claw reaching the grasping point is almost 100% within the tolerance range of 2 cm, and the highest grasping success rate of 85% is reached at the distance of 130 cm, which can meet the requirements of the explosive detonation robot for grasping explosives. To further verify the effectiveness of the system in this paper, under the same conditions, the traditional fixed vision method is used for grasping experiments, within the acceptable error range, the system in this paper is compared with the fixed vision of the robot arm grasping system, and it not only solves the problem of fixed camera vision obscured by the robot arm and shooting clarity, but also effectively improves the grasping success rate. To further improve the grasping success rate, this paper improves the grasping method by using the GVF snake algorithm to make the hand claw grasp the center point of the target object, and the maximum grasping success rate can reach 92%. Also, this paper conducts a comparison experiment by grasping two different shapes of target objects, and the experimental results show that the system has a higher grasping success rate for columnar objects. The dynamics model and observation model investigate the control strategy and control method of follower tracking. The control strategy is designed in the absence of depth information, the tracking target is determined to be the upper body of the human body, and the centerline of the camera is used as the reference to control the following tracking motion of the trolley, and the PID control method based on the ROS system is adopted.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Science Technology Major Project of Liaoning Province China, Development and Application Demonstration of Heavy Duty Industrial Multi Rotor UAV, under No. 2019JH1/10100028.

References

[1] A. Denasi and S. Misra, "Independent and leader–follower control for two magnetic micro-agents," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 218–225, 2017.

[2] P. Paral, A. Chatterjee, and A. Rakshit, "Vision sensor-based shoe detection for human tracking in a human-robot coexisting environment: a photometric invariant approach using DBSCAN algorithm," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4549–4559, 2019.

[3] L.-H. Juang and J.-S. Zhang, "Robust visual line-following navigation system for humanoid robots," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 653–670, 2020.

[4] M. H. Lee, N. P. Nguyen, and J. Moon, "Leader-follower decentralized optimal control for large population hexarotors with tilted propellers: a Stackelberg game approach," *Journal of the Franklin Institute*, vol. 356, no. 12, pp. 6175–6207, 2019.

[5] Z. Gao and G. Guo, "Adaptive formation control of autonomous underwater vehicles with model uncertainties," *International Journal of Adaptive Control and Signal Processing*, vol. 32, no. 7, pp. 1067–1080, 2018.

[6] M. Li and Y. Sun, "General rational approximation of Gaussian wavelet series and continuous-time g_m-C filter implementation," *International Journal of Circuit Theory and Applications*, vol. 48, no. 11, 2020.

[7] I. Saleh and W. M. Y. R. W. Abdul, "Fuzzy logic collision avoidance for autonomous RC car follower utilizing monocular camera as distance approximator," *International Journal of Electrical Engineering and Applied Sciences*, vol. 1, no. 2, pp. 53–60, 2018.

[8] S. van der Helm, M. Coppola, K. N. McGuire, and G. C. de Croon, "On-board range-based relative localization for micro air vehicles in indoor leader–follower flight," *Autonomous Robots*, vol. 44, no. 3, pp. 415–441, 2020.

[9] S. Zhao, "Application of a clustering algorithm in sports video image extraction and processing," *The Journal of Supercomputing*, vol. 75, no. 9, pp. 6070–6084, 2019.

[10] Y. Wang, M. Shan, Y. Yue, and D. Wang, "Vision-based flexible leader–follower formation tracking of multiple nonholonomic mobile robots in unknown obstacle environments," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 3, pp. 1025–1033, 2019.

[11] L. N. Tan, "Omnidirectional-vision-based distributed optimal tracking control for mobile multirobot systems with kinematic and dynamic disturbance rejection," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5693–5703, 2017.

[12] P. Borges, C. Sagastizábal, and M. Solodov, "Decomposition algorithms for some deterministic and two-stage stochastic single-leader multi-follower games," *Computational Optimization and Applications*, vol. 78, no. 3, pp. 675–704, 2021.

[13] C. J. Jermak, M. Rucki, and M. Jakubowicz, "Accuracy of the pneumatic follower for the wooden surface quality assessment," *European Journal of Wood and Wood Products*, vol. 78, no. 6, pp. 1149–1159, 2020.

[14] H. Qin, H. Chen, and Y. Sun, "Distributed finite-time fault-tolerant containment control for multiple ocean bottom flying nodes," *Journal of the Franklin Institute*, vol. 357, no. 16, pp. 11242–11264, 2020.

[15] J. Yang, Y. Zhao, J. Liu et al., "No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions," *IEEE Transactions on Cybernetics*, pp. 1–13, 2020.

[16] S. Hao, L. Yang, and Y. Shi, "Data-driven car-following model based on rough set theory," *IET Intelligent Transport Systems*, vol. 12, no. 1, pp. 49–57, 2017.

[17] C. Nayak, S. K. Saha, R. Kar, and D. Mandal, "Optimal design of zero-phase digital Riesz FIR fractional-order differentiator," *Soft Computing*, vol. 25, no. 6, pp. 4261–4282, 2021.

[18] P. Agarwal, P. Gautam, A. Agarwal, and V. Singh, "Human follower robot using Kinect," *International Research Journal of Engineering and Technology*, vol. 4, no. 4, pp. 1635–1637, 2017.

[19] O. Janbu, R. Johansson, T. Martinussen, and J. Solhusvik, "A 1.17-megapixel CMOS image sensor with 1.5 A/D conversions per digital CDS pixel readout and four in-pixel gain steps," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 9, pp. 2568–2578, 2019.

[20] S. Sukisaki, R. Shimomura, and H. Nobuhara, "Three-dimensional position estimation method via AM pulse light modulation and an application to control multiple UAVs," *Advanced Robotics*, vol. 32, no. 19, pp. 1023–1036, 2018.