


Research Article

An Algorithm for Motion Estimation Based on the Interframe Difference Detection Function Model

Tengfei Zhang ¹ and Huijuan Kang²

¹*School of Computer Science, Pingdingshan University, Pingdingshan 467000, Henan, China*

²*Department of Computer Science and Application, Pingdingshan Vocational and Technical College, Pingdingshan 467000, Henan, China*

Correspondence should be addressed to Tengfei Zhang; 4367@pdsu.edu.cn

Received 30 November 2020; Revised 2 February 2021; Accepted 20 February 2021; Published 27 February 2021

Academic Editor: Wei Wang

Copyright © 2021 Tengfei Zhang and Huijuan Kang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we simulate the estimation of motion through an interframe difference detection function model and investigate the spatial-temporal context information correlation filtering target tracking algorithm, which is complex and computationally intensive. The basic theory of spatiotemporal context information and correlation filtering is studied to construct a fast target tracking method. The different computational schemes are designed for the flow of multiframe target detection from background removal to noise reduction, to single-frame detection, and finally to multiframe detection, respectively. This enables the ground-based telescope to effectively detect spatial targets in dense stellar backgrounds in both modes. The method is validated by simulations and experiments and can meet the requirements of real projects. The interframe bit attitude estimation is optimized by using the beam-parity method to reduce the interframe estimation noise; a global optimization strategy based on the bit attitude map is used in the back end to reduce the system computation amount and make the global bit attitude estimation more accurate; a loop detection based on the word pocket model is added to the system to reduce the cumulative error.

1. Introduction

With the aging of the world's major economies and industrialized countries, social productivity is facing great challenges [1]. To meet the growing material needs of society, improving productivity and solving labour shortages have become challenges that must be addressed head-on by countries around the world. In recent years, the development of intelligent technology has promoted the research of using intelligent robots to replace manual labour to solve the severe labour shortage problem, especially the improvement of environment sensing and interaction capability with machine vision as the core, which has prompted the application of intelligent robots to expand from traditional industrial manufacturing to food, light industry, electronics and telecommunications, and pharmaceutical industries. Intelligent robots have gradually emerged as a key factor in the wave of technological innovation in various industries, and researchers have been exploring how to integrate

intelligent robots into various industries [2]. With the progress of time and technological development, robotics, which is based on machinery manufacturing technology and mechatronics technology, and combines computer processing and multisensor fusion, has gradually become a cross-discipline with great potential [3]. A new field has developed in the development of SLAM. Most current SLAM systems use either LIDAR or vision sensors, but each has its advantages and disadvantages. Laser sensors can directly measure the distance and angle information of the sensor relative to the object, and the accuracy of the measurement data is high and the measurable distance is long, but laser SLAM is not good at using in dynamic environments, such as a large number of people blocking its measurement; nor is it good at working in an environment with similar geometric characteristics, such as a long, straight, two sides of the wall [4]. It is also difficult for laser SLAMs to get back to work after tracking is lost due to poor relocation capabilities [5]. The vision sensor can extract a

large amount of redundant texture information from the environment, making loopback detection easier for SLAM systems, eliminating errors through loopback correction even when the front-end accumulates certain errors, and performing better at larger scales and in dynamic environments because of its texture information [6].

Razzaq et al. applied the idea of particle filtering algorithm to the nonlinear filtering of SIS, which led to the widespread use and development of particle filtering in many fields [7]. Later, Ci et al. used Rao-Blackwell statistics to find better statistical estimates and combined it with particle filtering to propose RBPF [8]. The idea of probabilistic SLAM design has shifted the focus of SLAM research from improving the performance of extended Kalman filtering and ensuring linear Gaussian assumptions to the notion of directly representing nonlinear systems and non-Gaussian state distributions [9]. Although their fast SLAM still linearized the observational model, its fusion of particle filtering and extended Kalman filtering allowed researchers to see the advantages of particle filtering, and the Fast SLAM2.0 algorithm later proposed by Xue et al. had a profound influence on subsequent SLAM studies [10]. The BA method is used to optimize the camera's bit attitude during tracking, optimize the current keyframe and feature points of the local map (local BA), and optimize all keyframes and feature points after loopback detection (global BA) [11]. Zhao added PTAM nonlinear optimization to solve the bit attitude, while Artal proposed the ORB SLAM algorithm for multicamera conditions [12]. Chen et al. demonstrated the requirement for consistency of luminance value characteristics on ambient illumination by photographing the same target under identical conditions except for luminance, but the results of the photographing made it difficult to determine whether the objects in the picture were the same target [13]. The luminosity of the target is high. To overcome this problem, Bellamine et al. proposed an object tracking method based on constant luminance features. However, this kind of algorithm requires very high computational hardware; otherwise, it cannot meet the requirement of real-time tracking due to the time-consuming tracking caused by its computational load [14]. Particle filtering is a method that uses many random particles to approximate the real state of the target, which is widely used in target tracking algorithms because of its good applicability to nonlinear motion models and non-Gaussian distribution of the target [15]. After extracting the possible candidate regions of the target through motion estimation, we can accurately locate the target in the candidate regions through the target localization method [16]. The disadvantage of this algorithm is that it cannot generate a high-density flow vector. For example, the flow information will fade quickly in terms of moving edges and small movements in black and homogeneous areas. Its advantage lies in the robustness with noise. The target localization methods are mainly divided into the extreme value of the target function method and the probabilistic statistical method [17]. In intelligent robot application scenarios, the visual tracking task often needs to face the situation that the target is obscured or even briefly disappears, and it needs to face a long tracking task, which

leads to the fact that when the target tracker is used alone, it cannot guarantee the correctness of the target tracking results during the task. By adding a detector to the tracking algorithm, it is possible to resume tracking the target when the tracking result deviates greatly from the current image by scanning the detector. However, the addition of a detector will undoubtedly increase the computational effort. What needs to be studied is how to balance the performance of the algorithm to achieve the best overall result [18].

The motion of the robot is discretized and described by the control and observation equations; a Bayesian probabilistic model is used to construct a mathematical model [19, 20], and a localization algorithm for batch processing of particles is used to find the optimal positional match to improve the speed of the system by batch processing of the particle population; a mapping method for incremental raster description is completed, and the environment map is discretized into a raster, which is approximated by the raster line. The proposed index preprocessing method establishes a mapping between the preprocessed index value and the occupied probability value of the raster, updates the occupancy state of the raster by looking up the index value, and improves the probability calculation rate of the cell raster. The vision-based SLAM framework is divided into a front-end and a back-end for processing; at the front-end, the random sampling consistency method is used to eliminate mismatches and improve the matching accuracy of image features; EPnP is used to estimate the interframe motion of the camera and complete the visual odometer based on the feature point method; a local map is created for interframe estimation to increase the correlation between the interframe data, and the interframe poses are evaluated by the beam-panning method. Preliminary optimization of the estimation is carried out to reduce the interframe estimation noise; a global optimization strategy based on bitmap is used in the back-end to reduce the computation volume while preserving the data constraints to the maximum extent, to improve the speed of the system and to make the calculation results more accurate; loop detection based on the word pocket model is added to reduce the cumulative error of the system.

2. Design Analysis of Motion Estimation for the Interframe Difference Detection Function Model

2.1. Optimized Interframe Difference Detection Function Model. In the video, the target usually exists in a relatively stable time and space and is always concerning the surrounding environment; therefore, reasonable use of contextual information around the target can not only predict the location of the target but also make the algorithm fast. Contextual information [21] is divided into two types: spatial and temporal. The former mainly refers to the objects with reference meaning in the area adjacent to the target; the latter mainly refers to the relationship between the trajectory of the target movement and the state information such as target position, velocity, acceleration, etc. In the tracking

process, the full application of the contextual information not only is useful for predicting the position of the target but also makes the algorithm fast. However, most of the current target tracking algorithms only use the temporal context of the target in the image sequence and ignore the contribution of spatial context features to the tracking. It is well known that occlusion is an unavoidable problem in real-world environments, and scholars usually use temporal context information to deal with this problem, which assumes that the overall features of the target remain similar between two adjacent frames and that the location does not mutate, but rarely uses spatial context information composed of the background of the target and its neighbouring regions. The presence of a referential object in the background environment can provide useful information for target localization and tracking when the appearance of the target changes drastically and the target is obscured, so making full use of both temporal and spatial context information can make the target tracking result more accurate.

The STC method takes advantage of the correlation between the target and a region of the surrounding background in space and between adjacent frames in time, and models the probability of the target's location in the next frame by modelling the low-order features in the image, and takes the location with the highest likelihood as the tracking result. So, the tracking problem can be equated by calculating and finding the location where the maximum value of the tracking target is in the confidence plot, which can be represented as

$$A_{i,j}^m(x) = P_{i,j}^M(x|o). \quad (1)$$

The tracking result is centered on the location where the maximum value of $c(x)$ is located. Assuming that the target centre point is x , the context feature information centered on it can be represented as

$$X^A = \{v(z) = I^m(z), z | z \in S_A(x^*)\}. \quad (2)$$

The objective function can be transformed from conditional probability to

$$X^A(x) = \sum_{v(z) \in S_A} P_{i,j}^M(x|o). \quad (3)$$

Equation (3) represents the target context prior probability model, which can be represented in the spatiotemporal context information algorithm as

$$a(x) = \sum_{v(z) \in S_A} P_{i,j}^M(x|o) I^m(z). \quad (4)$$

To normalize the results of the probabilistic model, the above equation introduces a regularization constant to limit the range of the results, and the scale parameter σ is used to reflect the feedback of the change in scale during the target tracking process on the context prior probabilistic model [22].

Convolution theory is the theory underlying the correlation filter class of tracking algorithms and is fundamental

to achieving fast target tracking. Convolution refers to the integration of a set of input signals with a unit impulse signal. When convolution is used in signal processing, its main function is to filter the signal, while in image processing, the convolution kernel (i.e., the template of the image) moves with the pixels in the image and calculates each pixel according to the predefined information of the convolution kernel. Suppose we have two functions, $f(x, y)$ and $h(x, y)$, both of size $M * N$ matrix. Their discrete convolution $f(x, y) * h(x, y)$ can be expressed as

$$f(x, y) * h(x, y) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=0}^N f(m, n) h(x - m + 2, y - n - 3). \quad (5)$$

From the convolution theorem, the Fourier transform of two functions is equivalent to the product of two functions after the Fourier transform, which can also be understood as the product of one domain is equivalent to the product of the other domain, just as the convolution operation in the time domain is equivalent to the product operation in the frequency domain in the time domain transformation. If the Fourier transform of $f(x, y)$ and $h(x, y)$ is denoted by $F(x, y)$ and $H(u, v)$, respectively, in the spatial domain, then in the spatial domain one obtains (6).

$$f(x, y) * h(x, y) = F_{m,n}^{-1}[F(m, n)H(u, v + 1)]. \quad (6)$$

The first step in solving the combined navigation system is to unify the coordinate systems, thus ensuring the accuracy of the subsequent calculations. Based on the coordinate systems involved in the two different visual/inertial guidance systems, the relationship is shown in Figure 1.

The optical flow field method is one of the most important methods for motion video sequence analysis, which takes advantage of the temporal variation of the pixel values in the video sequence and the correlation between the surrounding pixels to achieve the motion analysis. The optical flow field is obtained by calculating the velocity of each pixel in the image, and the moving target is detected by analysing the patterns of light flow in the optical flow field. When the camera is stationary, the light flow generated by the background is small and most of the light flow is concentrated on the moving object, so the moving object can be detected by detecting the light flow size. Here, we adopt a system; that is, there is a world coordinate system, the position or posture we define is the reference world coordinate system or the Cartesian coordinate system defined by the world coordinate system, and the dimension discussed is 3 dimensions. The location of the moving object is different in each frame, which causes the pixel value of the difference image to change, and the location of the larger pixel value in the difference image is the location of the edge of the moving object. The basic algorithm of the interframe distribution is shown in Figure 2. First, the pixels corresponding to two adjacent frames are differentiated, and the maximum interclass segmentation is taken to determine the difference threshold and binarized. However, the detected contours are incomplete, and large holes will appear inside the object with uniform pixel values, which can be processed by

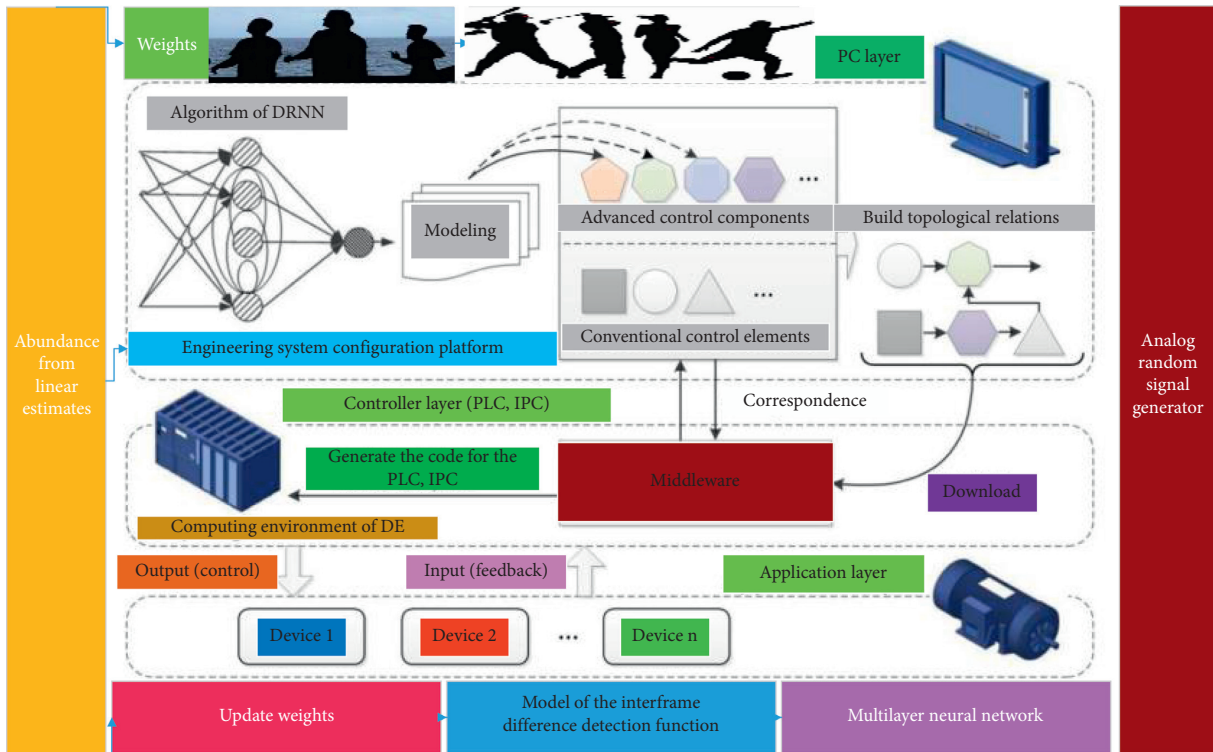


FIGURE 1: Model of the interframe difference detection function.

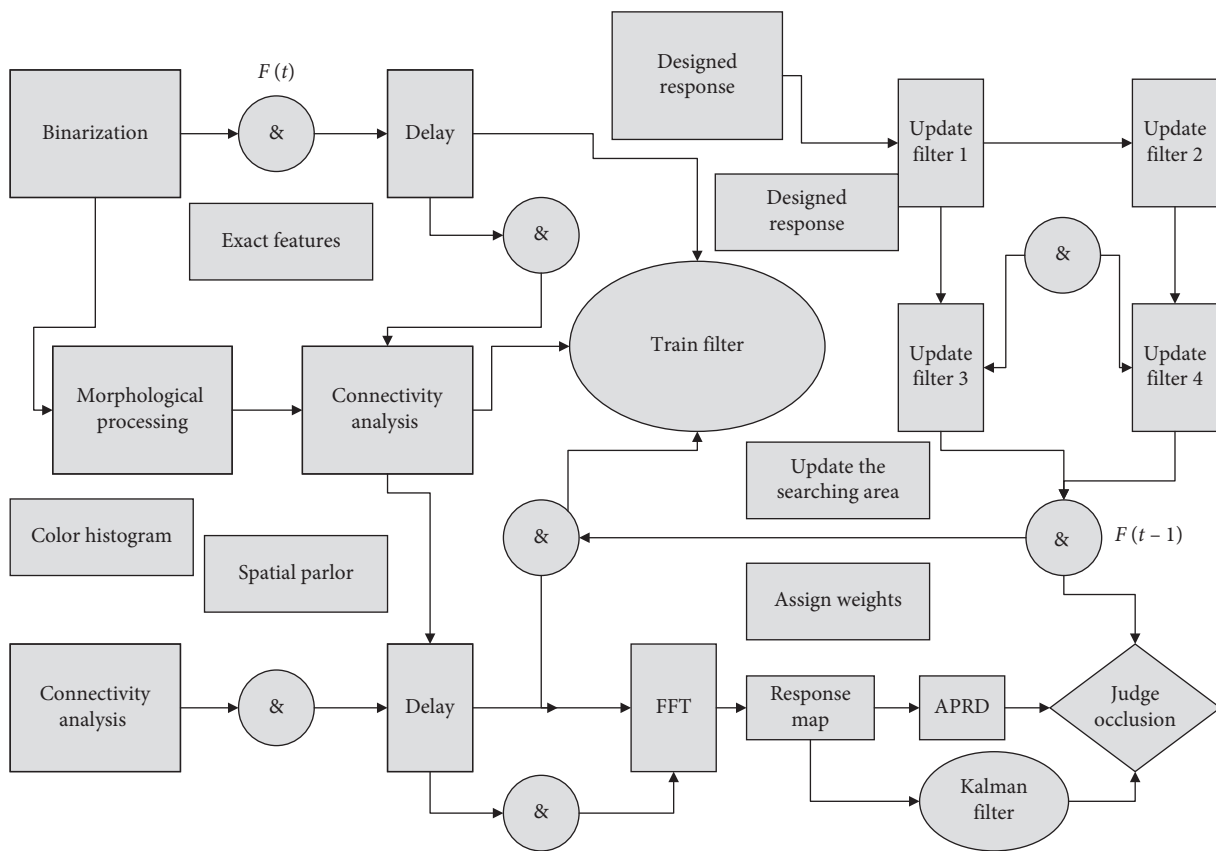


FIGURE 2: Difference-in-frame process.

morphological closure (i.e., expansion followed by erosion) to smooth the target boundary and fill the internal holes.

Similar to the neighbour-to-neighbour differential method, the three-frame image differential method is a method that multiplies the difference between the first two frames of three consecutive frames and then performs a threshold segmentation of the differential image to extract the moving object edges in the middle frame. The more intense the object motion and the greater the contrast between foreground and background, the larger the absolute value of the difference. The three-frame differential method can significantly enhance the edge information of a moving object, but it still does not overcome the problem of holes inside the moving object caused by the difference between adjacent frames. The background estimation-based moving object detection method is proposed to solve the problem that the interframe differential method cannot capture the complete moving object region. The method first estimates a pure background image without foreground object in the video sequence and then compares the detected image with the background image in the sequence (usually with a differential operation), and the resulting comparison is processed by a morphological method to obtain a more complete moving object region. The background estimation method is not complicated in the actual detection, but it is sensitive to the effects of lighting, shadows, motion imbalance, and other factors in dynamic scenes, so how to estimate an accurate background model in complex scenes has been the main topic of research by scholars.

2.2. Motion Estimation Design System Analysis. The basic structure of the Hartmann wave front sensor is shown in Figure 3; its main structure is a set of micro lens array on the front and a photoelectric sensor on the back; in most cases, the photoelectric sensor is a CCD camera. The main function of the microlens array is to separate the wave front of the incident light so that the incident wave front is transmitted to different subapertures. The wave front within a single subaperture is equivalent to a planar wave front with a slope, as shown in Figure 3. The incident wave fronts within each subaperture eventually converge on the focal plane of the microlens array for imaging, and the CCD acts as a photoelectric sensor to acquire an image of the focal plane. The image represents the light intensity distribution and position information of the spot in the focal plane in each subaperture. The wave front information is finally measured by calculating the spot position and using a wave front recovery algorithm.

Currently, the use of adaptive optics for ground-based telescopes is limited to night-time operation, where the operating hours are strictly limited because there is little background interference and the Hartmann wave front sensor can measure wave fronts reliably and accurately. However, when ground-based telescopes observe during the daytime, early morning, and evening, the adaptive optics system will inevitably receive background that is not part of the wave front information, resulting in a degradation of the signal-to-noise ratio of the Hartmann wave front sensor and a large

error in the offset of the spot array received by Hartmann, which leads to inaccurate detection of wave fronts passing through the telescope system and the inability to accurately calibrate the deformation mirror to compensate for aberrations in the wave front [23]. The machine vision system can quickly obtain a large amount of information, and realize faster product inspection, and is also easy to integrate information in the processing process. Especially in the mass industrial production process, the manual visual inspection of the product quality is inefficient and the accuracy is not high. Using machine vision inspection methods can greatly improve production efficiency and automation. As a result, the imaging of the telescope is not clear, the spatial target is not visible, and eventually, the target detection of the image is not possible. Therefore, the background suppression of the Hartmann sensor in bright light is necessary. The scattered light from outside sources is mainly produced by some external light sources in the atmosphere. The most obvious factor is the scattered light from the sun before dawn or after dusk, which results in an uneven background. Mist, nebulae, and bright haloes can also produce inhomogeneous backgrounds with similar effects. The scattered light formation mechanism is different from igniting; the scattered light belongs to the real background rather than to the light intensity loss caused by the optical system, but the result is that both igniting and scattered light cause inhomogeneity in the image. In space object detection, the nonuniformity of the igniting and scattered light distorts the map and inconsistency in the background brightness. Therefore, images taken by astronomical telescopes cannot be used directly, and non-uniform corrections must be made to ensure the performance of target detection and identification.

To solve the above problem, an expectation-maximization (EM) based background suppression method is proposed in this paper. The EM algorithm is an unsupervised machine learning algorithm suitable for solving probabilistic model parameters in incomplete or missing datasets (including hidden variables) [24], with low computational complexity and memory overhead, and requiring very few computational resources. In this paper, we add two types of processes to the E step of the EM algorithm: the first one is to generate data from the model; the second one is to determine the range of missing data. However, for substantially missing data, this paper adds two types of processes to the E step of the EM algorithm: the first one is to generate data from the model; the second one is to determine the range of missing data.

$$E = f_{SG}(I * I^M, W). \quad (7)$$

Second, determine the region of missing values in the background signal. In the background signal, the region of missing values is the region covered by the spot signal S generated by the micro lens array, which is usually a spot-like target that occupies very few pixels in the range.

$$M(n) = I * I^M - E(i, j) + N(0, \delta^2). \quad (8)$$

To distinguish between spots that are out of the background, projects typically set a threshold value of n times the

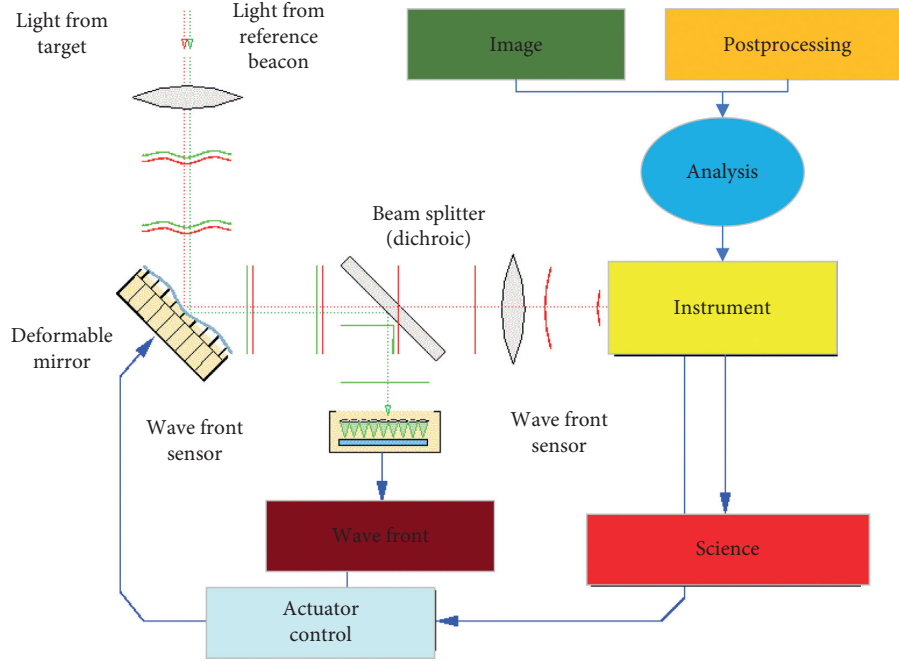


FIGURE 3: Motion estimation design system.

standard deviation. After distinguishing the spots, a Boolean matrix is used to record the area where the missing values are located, where if the grayscale value of a pixel in the difference image is greater than n times its standard deviation, the corresponding position of the matrix is recorded. The parameter n is usually chosen according to the actual

percentage of pixel area occupied by the spot. Usually, the threshold value leaves some margin to ensure that the spot will be larger than the threshold value, usually in the interval, which in this paper is set to $n = 2$. The recording matrix R is represented as follows:

$$R = \begin{cases} R(i, j) = 2, |I(i, j) - E(i-1, j-1)| \geq n * \text{std}(E - I) \\ R(i, j) = 2, |I(i, j) - E(i-1, j-1)| < n * \text{std}(E - I) \end{cases},$$

$$V(I) = \sum_{m,n} \sqrt{|I_{m-1,n} - I_{m+1,n}|^3} + \sqrt{|I_{m+1,n} - I_{m+4,n}|^3}. \quad (9)$$

To solve the problem of lack of target training samples, we study the cyclic sampling method to enrich the number of samples obtained at the beginning of target enrichment, the kernel correlation filter to describe the target model, and the STCF method to describe the target to improve the accuracy of the target model construction. In addition to improving the matching accuracy of image features, it can also improve the speed of image matching. The target interference information discrimination method is designed to enable the algorithm to determine the severity of the interference on the current tracking results and to formulate a reasonable target update scheme based on the interference to prevent the target template from being contaminated. To solve the problem of sudden acceleration or rapid movement of the target in an industrial environment, we design the target movement position estimation method to improve the tracking accuracy of the algorithm for fast-moving objects. Secondly, in the target localization section, the nuclear

correlation filtering method is integrated into the STCF target tracking algorithm to make full use of the target-background information to achieve accurate localization; then, to address the problem of similar object interference, the interference information discrimination method based on the average peak correlation energy is studied. Finally, to address the problem that the target is moving too fast beyond the search area and causing tracking drift or failure, we design a Kalman filter-based target motion position prediction method to predict the position of the target in the next frame and use the predicted position as the centre of the search area to improve the tracking algorithm's adaptive ability to fast-moving interference.

2.3. Design Analysis of Performance Indicators. In the visual tracking process, the target model built by the tracking algorithm is often incomplete after the target initialization

due to the small number of target samples acquired, which will have an important impact on the subsequent tracking process. To overcome the problem of incomplete description of target features due to the small number of target samples, a common method is to sample the target area several times, but this will certainly affect the algorithm's efficiency. Therefore, how to extract target features efficiently and accurately has become one of the research challenges in the field of visual tracking. In the field of visual tracking, the nonlinear model is the most encountered one, but it poses a challenge to target tracking due to its classification difficulty. According to pattern recognition theory, employing nonlinear mapping, data that is linearly indistinguishable in low-dimensional space may be linearly distinguishable in high-dimensional space. Therefore, the main solution for classifying nonlinear models today is to use mapping functions to map the original data to a high-dimensional feature space so that it shows a linear relationship, and then apply a linear algorithm for recognition. In the process of mapping from low-dimensional to high-dimensional space, the selection of nonlinear mapping functions needs to be addressed first, because the improper selection of nonlinear mapping functions will lead to inaccurate classification results, and secondly, the computational efficiency of high-dimensional space is low, but this problem can be solved by the kernel function approach. In this section, based on the target tracking algorithm framework in Section 2, the relevant filtering methods are introduced into the target position estimation process, and the kernel function approach is adopted to improve the accuracy of target tracking, which can make full use of the target-background information to achieve accurate target tracking, as shown in Figure 4.

According to the nature of the essence matrix in Section 2, since the degree of freedom of the essential matrix is 5, at least 5 pairs of matching points between neighbouring frames are needed to solve the essence matrix. According to Section 2.2 on feature matching, there are n pairs of matching points between two neighbouring frames, and the external environment such as illumination, obscuration by foreign objects, the blurred image caused by the jitter of the flight platform, or changes in the view angle has a great influence on the feature points. The moving target detection method based on optical flow field analysis not only contains the motion information of the observed object, but also carries rich information of the three-dimensional structure. Therefore, it can not only be used for moving target detection, but also be directly applied to moving target tracking. The speed of the moving target is calculated very accurately, and the moving target can also be detected when the camera is in motion. The RANSACK algorithm takes randomness and hypothesis as its core and uses multiple iterations to estimate the optimal mathematical model from data containing large noise, which is widely used in mathematics, science, and vision fields.

A complete measurement system needs to find the correspondence between the raw data of each sensor. After the relative poses of the two sensors are obtained, this section uses the TF module to build a tree structure for the transformation of the robot sensor poses between the two

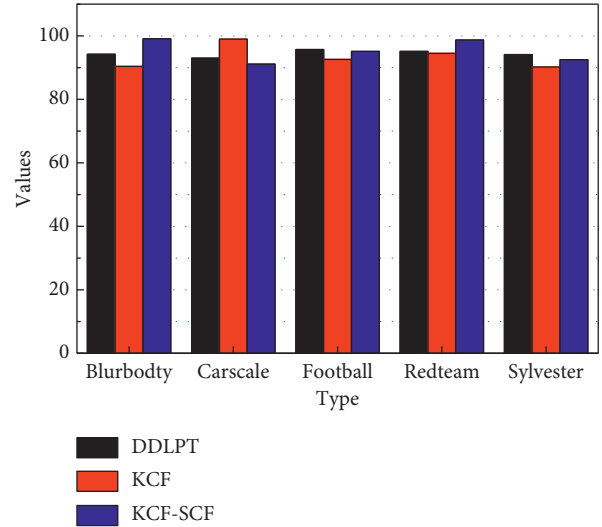


FIGURE 4: Comparison of tracking performance.

sensors, to quickly transform the data of the two sensors in different coordinate systems. To convert the sensor data between their coordinate systems, one needs to left-multiply the coordinates of the point in the current coordinate system by the description of the current coordinate system in the reference coordinate system, i.e., the transformation matrix. But the rotation matrix in the transformation matrix is represented by 9 quantities, and the rotation matrix itself has the constraint that it must be an orthogonal matrix with determinant 1, which makes it difficult to solve the transformation matrix. Using rotation vectors to compute the rotation matrix solves the problem of redundancy of degrees of freedom, but it runs into the gimbal locking problem, which causes the system to lose one degree of freedom.

Assuming that there are N coordinate systems on the robot, there are C_2^N combinations of transformations between any two coordinate systems, and the more coordinate systems there are, the more complex the combination of transformations becomes. To solve this problem, a multi-layer politician tree (TF tree) is built and maintained to store the descriptions of the coordinate systems. The relationships between parent and child nodes are defined according to the robot's model, and the descriptions of coordinate systems are stored in the nodes so that the order of transformations between coordinate systems can be quickly found during the computation and the data can be transformed between coordinate systems. From the system model of SLAM, it is a problem of estimating the real state of a robot by perceiving the changes in its external environment. As shown in Figure 5, when only the motion equation is available, the external environment information is missing, and although the robot can estimate its approximate position, the uncertainty of the position increases with time. When the observational equation is introduced, i.e., when the robot observes an environmental roadmap, the uncertainty of its position is corrected and gradually reduced to a range that remains stable. Bayesian probabilistic models are widely used in the SLAM field as a method to solve a posteriori probability

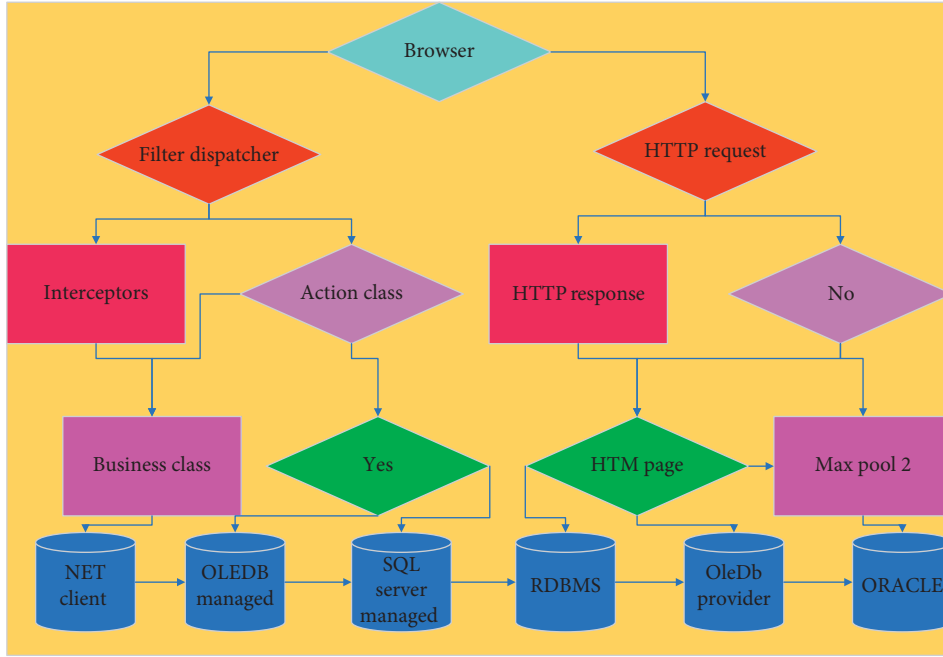


FIGURE 5: Performance analysis flow diagram.

while knowing the a priori probabilities. It selects three consecutive frames of video images for differential operation to eliminate the influence of the exposed background due to movement, thereby extracting accurate contour information of the moving target.

The state transfer of a SLAM system obeys a first-order Markov model; i.e., there is a transfer probability for a system to move from one state at one moment to the next, and this probability can be computed from the state at the immediately preceding moment, independent of the initial state and all other state changes that preceded the transfer. For the observational equation of the discrete robot motion, its current state should be derived only from the state of the system at the previous moment, and all states before moment t can be correlated step by step using a finite number of recursive equations in the Markov model.

3. Results Analysis

3.1. Analysis of Motion Estimation Results. The rotation vector describes the rotational motion of the camera coordinate system in the world coordinate system, while the translation vector t contains the displacement of the camera coordinate system in the world coordinate system. Extending the above computational procedure, the trajectory of the camera's continuous motion can be obtained by connecting numerous discrete front-to-back and back-to-back motions in a segment of the camera's acquired serialized image. Using the method of this paper, the spatial motion of a segment of the camera is solved and the trajectory is shown in Figure 6.

The visual odometer uses incremental interframe estimation, which only calculates the feature matching and motion relationship between the two frames before and after

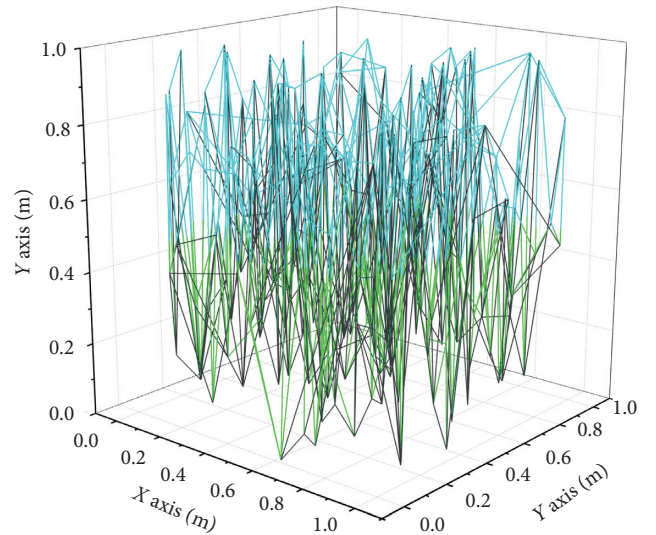


FIGURE 6: Space motion trajectory.

the image and can recover the camera's trajectory after estimating the bit-pose of consecutive frames. However, when the visual odometer only takes into account the bit-posture estimation of the adjacent time, the interframe bit-posture estimation may be very poor due to the poor quality of one of the two frames, which may cause the data with larger error to make a larger contribution to the subsequent nonlinear optimization, while the bit-posture estimation with smaller error will be more affected. Therefore, in this paper, a local map is added to the visual odometer, in which world coordinates are used for the locations of feature points. Each frame contributes some information to the map, and the feature points of each frame are cached in one place, which constitutes the set of feature points of the map.

When a new frame arrives, the feature matching and motion relationship between it and the map is evaluated, adding constraints to the incremental interframe estimation, and using the reprojection error with map point information to optimize the camera motion. Figure 7 shows a comparison of the motion of a section of the camera after using local map constraints and local optimization in this paper. When only the odometer is used to estimate the camera position, the environmental information is not maximized, and the calculated initial camera interframe estimation error is large and appears to be sharp due to the influence of sensor noise. After constructing the local map, the interframe estimation error of camera motion is smaller because each frame contributes some environmental feature information to the map as a constraint, and it can be seen that the camera moving with the map point constraint is less affected by the noise and the trajectory is smoother.

Figure 8 is a graphical representation of the tracking results of each tracking algorithm in the test video “Human6.” The main challenges posed by the test video to the tracking algorithm are target scale changes, occlusion, target deformation, fast motion, out-of-plane rotation, and target part out of the field of view. Then at frame 369, the IVT algorithm loses the target because the target is obscured by the roadside signage, indicating that the IVT algorithm is not robust when the target is obscured. At frame 496, the ASLA algorithm has a tracking offset, which is mainly due to the fast movement of the target, which causes the target template to be contaminated by the background, and then the target is lost in the subsequent video frames. This chapter presents that the DDLPT algorithm performs well in this test video sequence, but there is a slight tracking offset at frame 369, but the accurate tracking of the target is quickly restored, indicating that the DDLPT algorithm not only is able to adapt to scale changes, but also has good adaptability when the target is partially obscured and when the target is partially out of the field of view. The robustness of the algorithm in the occlusion case needs to be further investigated.

Based on the techniques such as the nucleus correlation filter, we design a cyclic sampling method for the initial region of the target, which enriches the initial training samples of the target and solves the problem that the target description is not perfect due to the lack of training samples; we study the target model construction method that combines the spatiotemporal context information and the nucleus correlation filtering features, which increases the accuracy of the target model description; we propose an APCE-based interference information filtering method, which sets the threshold value to filter out the interference information from the target. We also design a Kalman filter-based target motion position estimation method to improve the adaptability of the algorithm to fast-moving objects in the tracking process. To synthesize the above techniques, this section proposes an interference discrimination and position prediction target tracking algorithm to achieve adaptability to the second-level interference factors in the industrial environment. The comparison test results between DDLPT and similar algorithms show that the tracking accuracy is improved by 8.66% and the tracking overlap is improved by

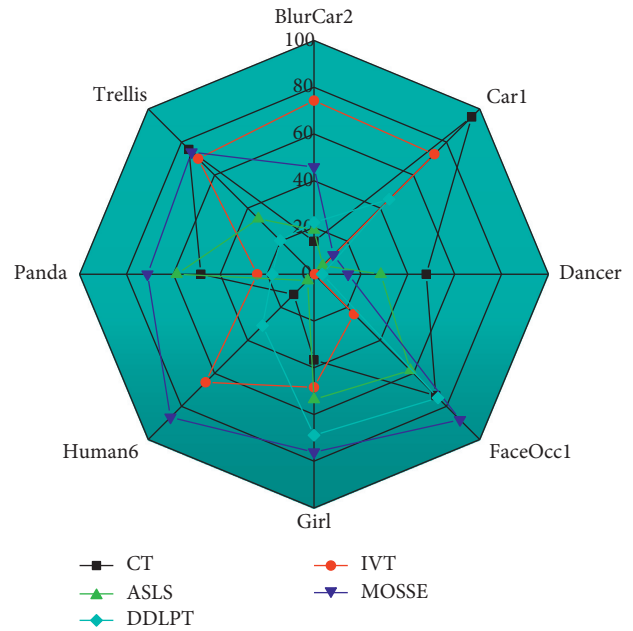


FIGURE 7: Tracking rate in the test video.

17.04%, which indicates that DDLPT has better tracking accuracy. The results show that the tracking accuracy of DDLPT is better than the mainstream tracking algorithm and achieves the expected target, but the tracking result is affected when the object is subject to rapid deformation, and partial blocking, or when the object is partially out of the field of view, so the robustness of the algorithm needs to be further evaluated.

3.2. Analysis of Performance Results. To test the above attitude update algorithm and verify its accuracy, a high-precision three-axis temperature-controlled turntable is used in this test. In this test, the high-precision three-axis temperature-controlled turntable is used. In the swing motion test, the three axes will be tested independently; i.e., only one axis is allowed to rotate in each test, and the other two axes are kept stationary, and after setting the turntable, it is ready to start the test task. During the roll angle test, set the middle frame and inner frame to position mode and keep stationary, and set the outer frame to swing mode with 20° swing amplitude and 0.5 Hz frequency. When the measurement system is powered up and stabilized, run the turntable, start the swinging motion, and transfer the data to the host computer for display and storage, as shown in Figure 9.

According to the single station positioning procedure in Figure 9, the simulation program is written, and without considering the error of each parameter, the target position is calculated by using the true value in Figure 10. The 2000 Monte Carlo simulations were able to obtain a distribution as shown in Figure 10, which shows that the probability of localization near the true value of target localization is large, reflecting a good localization effect.

Gaze-mode image acquisition involves a telescope tracking a spatial target at a constant angular velocity, remaining relatively motionless, and acquiring images with continuous

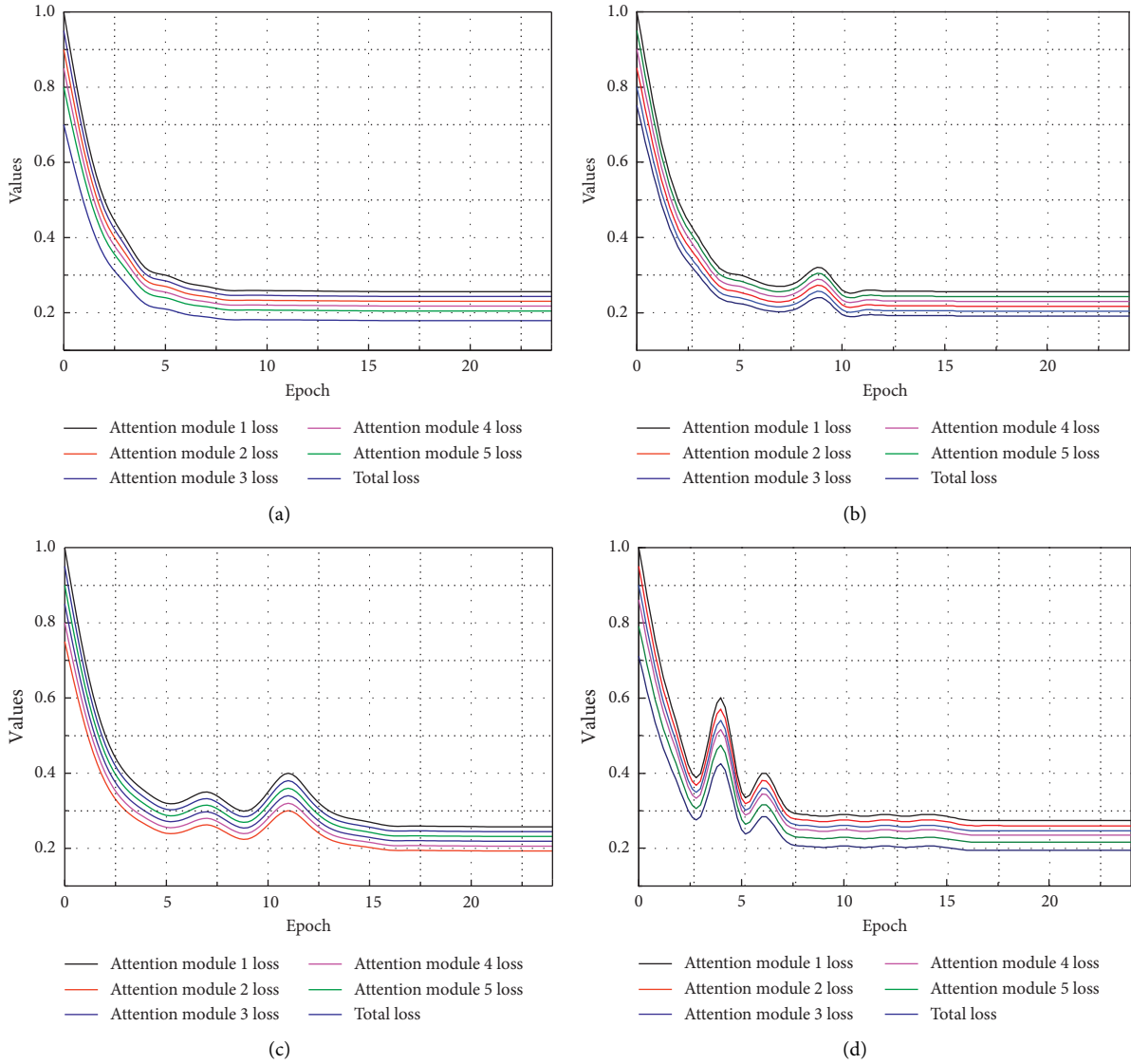


FIGURE 8: Tracking error in the video.

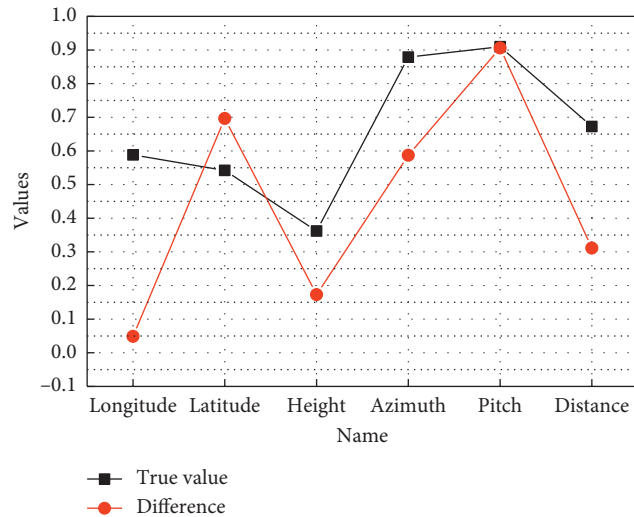


FIGURE 9: Simulation conditions for single station targeting.

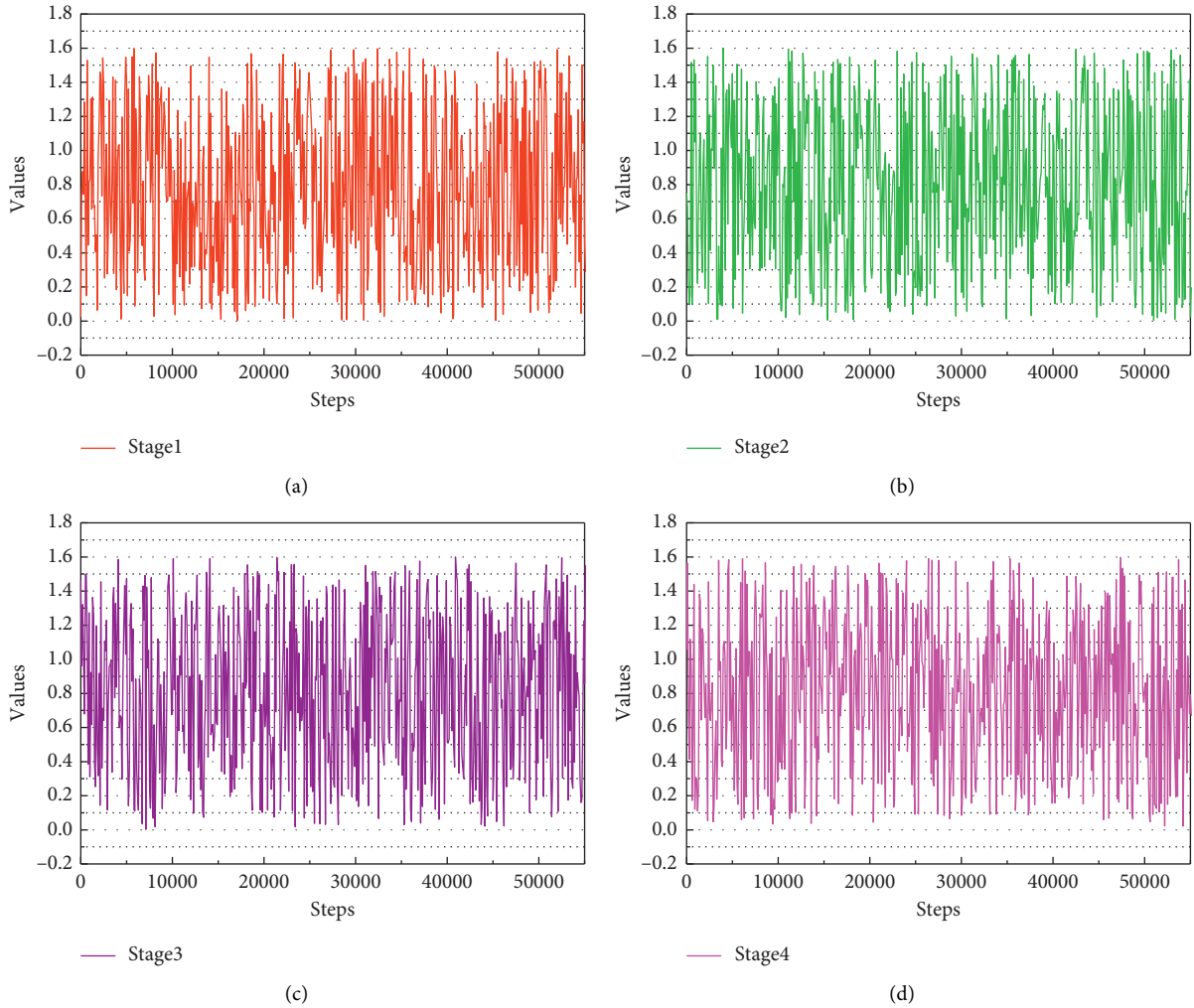


FIGURE 10: Brightness profile of a pixel.

exposure. The spatial target to be observed remains constant over a multiframe image sequence and is shaped as a spot, whereas a stellar object is a long strip of spots, the length of which depends on the exposure time and the relative speed of the spatial target. In gaze mode, the background removal method is the same as in stellar mode, but the target frame and background frame need to be aligned first. Both ignited light and scattered light will cause unevenness in the image and result in low accuracy. First, the stellar image is sparse, so a two-dimensional image is compressed into a one-dimensional array for interarray alignment using projection. The image can be aligned using grayscale projection in both horizontal and vertical directions, and then two one-dimensional arrays on the horizontal and vertical axes. The projection method is very simple to use for registration and greatly reduces the computational cost.

4. Conclusion

To avoid the problem of imperfect target description caused using spatiotemporal context information to construct target templates, this paper designs a method for

constructing target sample sets based on cyclic matrices and obtains rich target training samples. When the target is obscured, the target template is contaminated due to the offset of the tracking result; the target interference information discrimination method is designed. The target movement position estimation method predicts the possible position of the target at the next moment based on the estimated state of the target at the previous moment and the observation of the current moment and sets the target search area for the next frame. By combining the above three methods with the relevant filtering algorithm based on spatial-temporal context information, a tracking algorithm based on interference discrimination and position prediction is constructed, which improves the tracking accuracy by 8.66% and the tracking overlap by 17.04% compared with other similar methods and achieves the adaptation to the second-level interference factors in the industrial environment. By using principal component analysis and QR decomposition to downscale the shift filter and scale filter of the target tracking algorithm, respectively, and by conducting comparison tests on the video test set, the optimal number of downscaled dimensions that can balance the

accuracy and processing efficiency of the tracking algorithm is determined, which improves the processing speed by 9.17% over the original number of dimensions.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest reported in this paper.

References

- [1] T. K. Lee, Y. L. Chan, and W. C. Siu, "Adaptive search range for HEVC motion estimation based on depth information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2216–2230, 2017.
- [2] D. J. Hemanth and J. Anitha, "A pattern-based artificial bee colony algorithm for motion estimation in video compression techniques," *Circuits, Systems, and Signal Processing*, vol. 37, no. 4, pp. 1609–1624, 2018.
- [3] A. Lim, B. Ramesh, Y. Yang, C. Xiang, Z. Gao, and F. Lin, "Real-time optical flow-based video stabilization for unmanned aerial vehicles," *Journal of Real-Time Image Processing*, vol. 16, no. 6, pp. 1975–1985, 2019.
- [4] R. Khemiri, H. Kibeya, F. E. Sayadi, N. Bahri, M. Atri, and N. Masmoudi, "Optimisation of HEVC motion estimation exploiting SAD and SSD GPU-based implementation," *IET Image Processing*, vol. 12, no. 2, pp. 243–253, 2017.
- [5] T. Zhang, P. Jiang, and M. Zhang, "Inter-frame video image generation based on spatial continuity generative adversarial networks," *Signal, Image and Video Processing*, vol. 13, no. 8, pp. 1487–1494, 2019.
- [6] F. Li and S. Liu, "A sub-region image registration algorithm with weight-based hierarchical importance sample consensus for moving target detection," *The Imaging Science Journal*, vol. 65, no. 2, pp. 87–97, 2017.
- [7] M. A. Razzaq, J. M. Quero, I. Cleland et al., "uMoDT: an unobtrusive multi-occupant detection and tracking using robust Kalman filter for real-time activity recognition," *Multimedia Systems*, vol. 26, no. 5, pp. 553–569, 2020.
- [8] W. Ci, Y. Huang, and X. Hu, "Stereo visual odometry based on motion decoupling and special feature screening for navigation of autonomous vehicles," *IEEE Sensors Journal*, vol. 19, no. 18, pp. 8047–8056, 2019.
- [9] H.-X. Gao, W. Xie, H. Kang, and G.-Y. Lin, "Multi-frame super-resolution reconstruction based on global motion estimation using a novel CNN descriptor," *Optoelectronics Letters*, vol. 15, no. 6, pp. 468–475, 2019.
- [10] T. Xue, B. Chen, J. Wu, D. Wei, and W. T. Freeman, "Video enhancement with task-oriented flow," *International Journal of Computer Vision*, vol. 127, no. 8, pp. 1106–1125, 2019.
- [11] R. Kalboussi, M. Abdellaoui, and A. Douik, "Video saliency detection using motion distinctiveness and uniform contrast measure," *Informatika*, vol. 30, no. 1, pp. 53–72, 2019.
- [12] N. Zhao, D. O'Connor, A. Basarab, D. Ruan, and K. Sheng, "Motion compensated dynamic MRI reconstruction with local affine optical flow estimation," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 11, pp. 3050–3059, 2019.
- [13] C. Chen, F. Ding, and D. Zhang, "Perceptual hash algorithm-based adaptive GOP selection algorithm for distributed compressive video sensing," *IET Image Processing*, vol. 12, no. 2, pp. 210–217, 2017.
- [14] I. Bellamine, H. Silkan, and A. Tmiri, "Track color space-time interest points in video," *Multimedia Tools and Applications*, vol. 79, no. 33-34, pp. 24579–24593, 2020.
- [15] Y. Fan, Y.-Q. Guo, Z.-H. Tang, J. Luo, and G.-Y. Zhang, "A dynamic size-based time series feature and application in identification of zinc flotation working conditions," *Journal of Central South University*, vol. 27, no. 9, pp. 2696–2710, 2020.
- [16] X. Ding, Y. Huang, Y. Li, and J. He, "Forgery detection of motion compensation interpolated frames based on discontinuity of optical flow," *Multimedia Tools and Applications*, vol. 79, no. 39-40, pp. 28729–28754, 2020.
- [17] J. He, G. Yang, X. Liu, and X. Ding, "Spatio-temporal saliency-based motion vector refinement for frame rate up-conversion," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 2, pp. 1–18, 2020.
- [18] H. Wu, L. Xiao, H. J. Shim, and S. Tang, "Video stabilisation with total warping variation model," *IET Image Processing*, vol. 11, no. 7, pp. 465–474, 2017.
- [19] M. Chen, S. Lu, and Q. Liu, "Uniform regularity for a Keller-Segel-Navier-Stokes system," *Applied Mathematics Letters*, vol. 107, Article ID 106476, 2020.
- [20] M.-C. Chen, S.-Q. Lu, and Q.-L. Liu, "Global regularity for a 2D model of electro-kinetic fluid in a bounded domain," *Acta Mathematicae Applicatae Sinica, English Series*, vol. 34, no. 2, pp. 398–403, 2018.
- [21] N. A. Bahran, W. El-Shafai, A. Zekry, S. El-Rabaie, M. M. El-Halawany, and F. E. A. El-Samie, "An FPGA design and implementation of EPZS motion estimation algorithm for 3D H.264/MVC standard," *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 22351–22396, 2019.
- [22] A. Paltrinieri, R. Peloso, G. Masera, M. Shafique, and M. Martina, "On the effect of approximate-computing in motion estimation," *Journal of Low Power Electronics*, vol. 15, no. 1, pp. 40–50, 2019.
- [23] M. Khairy, A. Elsayed, A. Hamdy, and H. F. Ali, "Low complexity for scalable video coding extension of H. 264 based on the complexity of video," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 12, pp. 220–225, 2016.
- [24] D. Blinder, C. Schretter, and P. Schelkens, "Global motion compensation for compressing holographic videos," *Optics Express*, vol. 26, no. 20, pp. 25524–25533, 2018.