Retraction

Retracted: Technical Research on Moving Target Monitoring and Intelligent Tracking Algorithm Based on Machine Vision

Wireless Communications and Mobile Computing

Wireless Communications and Mobile Computing has retracted the article titled "Technical Research on Moving Target Monitoring and Intelligent Tracking Algorithm Based on Machine Vision" [1] due to concerns that the peer review process has been compromised.

Following an investigation conducted by the Hindawi Research Integrity team [2], significant concerns were identified with the peer reviewers assigned to this article; the investigation has concluded that the peer review process was compromised. We therefore can no longer trust the peer review process and the article is being retracted with the agreement of the Chief Editor.

The authors do not agree to the retraction.

References

[1] P. Xu, Z. Zhou, and Z. Geng, "Technical Research on Moving Target Monitoring and Intelligent Tracking Algorithm Based on Machine Vision," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 7277926, 13 pages, 2022.

[2] L. Ferguson, "Advancing Research Integrity Collaboratively and with Vigour," 2022, https://www.hindawi.com/post/advancing-research-integrity-collaboratively-and-vigour/.

WILEY | Hindawi

Research Article

Technical Research on Moving Target Monitoring and Intelligent Tracking Algorithm Based on Machine Vision

Pengfei Xu [iD], Zhiqing Zhou, and Zexun Geng

Department of Information Engineering, Pingdingshan University, Pingdingshan, 467000 Henan, China

Correspondence should be addressed to Pengfei Xu; 3221@pdsu.edu.cn

Machine vision is an important branch of the rapid development of modern artificial intelligence, and it is a key technology to convert the image information of monitoring targets into digital signals. However, due to the wide range of machine vision applications, this research focuses on its application in video surveillance. In the era of artificial intelligence, the detection and tracking of moving objects have always been a key issue in video surveillance. The simulation of human vision is realized by combining the relevant functions of the computer and the image acquisition device, which enables the computer to have the ability to recognize the surrounding environment through images. The intelligent video analysis technology can automatically analyze and extract the key useful information from the video source with the powerful data processing ability of the computer, so as to realize the computer's "understanding" of the video. It allows the computer to "understand" what is shown in the video or what kind of "event" happened and provides a new method and reliable basis for accident detection and accident analysis. Therefore, after a brief introduction to machine vision, moving target monitoring methods, and intelligent tracking algorithms, this paper will focus on moving target monitoring and intelligent tracking strategies for video surveillance. In addition, this paper will focus on introducing the principle of intelligent tracking algorithm through formulas and compare the accuracy and success rate of target monitoring and intelligent tracking between the machine vision-based algorithm and other algorithms during the experiment. Finally, experiments show that the monitoring and tracking effect of machine vision combined with "cloud" is the best, and the overall average can reach 85.7%. Based on this, this paper fully confirms the feasibility of the moving target monitoring and intelligent tracking algorithm based on machine vision.

1. Introduction

With the continuous development of science and technology, video surveillance has been widely used in banking, electricity, transportation, and other fields. The global demand for video surveillance is also growing rapidly, and the expectations for its personalization and intelligence are getting higher and higher. However, the basic operation of surveillance video is still mainly processed manually. Moreover, due to many limitations of hardware, the monitored images are sometimes not clear, especially when there are multiple moving targets at the same time; there may be cases where effective information cannot be obtained. At this time, only relying on video surveillance staff to identify surveillance images or videos and make corresponding judgments will bring huge challenges to the real time and accuracy of information. In the era of artificial intelligence, moving target monitoring and intelligent tracking are not only an important part of computer vision research but also the focus of cloud storage and cloud computing. The application of video surveillance technology based on machine vision can transform traditional manual passive monitoring into active machine monitoring, thereby freeing the on-duty personnel in the monitoring room from the heavy work of "staring" at the screen. Machine vision can quickly detect and track video information with the same characteristics from massive data, greatly improving the efficiency of information and resource utilization and improving the overall performance of video surveillance systems. Cloud-based machine vision technology enables object detection and

identification of moving objects with little or no human intervention. Therefore, in today's environment, the research of machine vision is very necessary.

The innovation of this study is the following:

(1) The machine vision tracking algorithm combined with "cloud" is used to realize the capture of video surveillance targets, and at the same time, the robustness of the algorithm is also well reflected in the aspect of intelligent tracking

(2) This research takes video surveillance as a specific direction and proposes a cloud-based machine vision technology that integrates identification, monitoring, and tracking, which will provide new ideas for future research in this field

(3) This study also studies the specific movements of the moving target, which is novel to a certain extent, and there is no research that is the same as this part in the existing research materials

2. Related Work

Many experts and scholars at home and abroad have put forward a lot of opinions on the research related to machine vision, moving target monitoring, and tracking algorithms.

Robie et al. have mainly studied several methods based on machine vision: methods for recording high-quality videos for automated analysis, video-based tracking algorithms for estimating the position of interacting animals, and machine learning methods for identifying interaction patterns [1].

Zhao et al. first proposed an IoT and digital twin tracking solution framework for security management and then developed an indoor security tracking mechanism for detecting stationary behavior and possessing automatic learning genetic localization technology. It is mainly used to identify abnormal situations and obtain accurate location information in real time [2].

Liang and Zhi locate the region of interest and select it by automatically adjusting the rotation of the camera to achieve the purpose of tracking and identifying moving targets, and on the basis of further analysis, the target monitoring is achieved [3].

Wang et al. proposed a system for remotely monitoring the displacement of moving objects. It uses quadrature self-injection-locked (QSIL) radars and additional phase shifters for conventional SIL radars to repeatedly switch the phase delay of the injected signal between 0 and 90 degrees. Therefore, it can be used to determine the displacement of moving targets without distortion due to large injection angles. In addition, they also proposed a measurement-and-difference-based method for calibrating the DC and phase offsets before performing arctangent demodulation to extract displacement information [4].

Danion and Randall point out that the ability to track moving objects with smooth eye movements is crucial in both perception and action tasks. In the study, they had participants observe a visual target. The target follows a smooth but unpredictable trajectory in the horizontal plane and is instructed to track the target with their gaze or a joystick-controlled cursor [5].

Wu et al. studied the tracking of moving targets by single-photon lidar (SPL) in the marine aerosol environment and introduced a tracking algorithm to deal with the ion-scattered electrical signals of SPL. In addition, they also proposed a two-stage (TS) dynamic model with an adaptive strong Kalman filter based on the SPL tracking technique [6].

From the perspective of control system design, Li et al. studied the two guidance laws of multiple autonomous UAVs for coordinated tracking of static and moving targets outside the defense zone. It adopts quantitative analysis method, fixes relevant parameters, only measures one neighbor in the communication topology network, and uses guidance laws to track static or moving targets [7].

3. Machine Vision and Moving Target Monitoring and Intelligent Tracking Algorithms

3.1. Machine Vision. Machine vision is a comprehensive technology in artificial intelligence, which mainly deals with target information including images and videos [8]. After that, the target image information is output as digital information through machine vision processing, so as to achieve the purpose of processing target object information and performing other analysis operations. The application areas of machine vision are shown in Figure 1.

Machine vision is generally used in working environments that are unbearable for manual operations or occasions where artificial vision is difficult to meet the needs. At the same time, machine vision monitoring and tracking can greatly improve the automation of production. Moreover, machine vision can easily realize the integration of information, which is one of the basic technologies of modern computer integrated manufacturing [9]. The working principle of machine vision is shown in Figure 2.

3.2. Moving Target Monitoring in Video Surveillance. In order to realize real-time monitoring of moving objects in video surveillance, the first thing to do is to detect the moving objects from the image sequence. The background model is obtained by the statistical method, and the model is updated in real time to adapt to the changes of light and scene, and then, the morphological method and the area of the connected domain are used for postprocessing. The detection effect directly depends on the accuracy of the image preprocessing operation. The specific implementation process of moving target monitoring is shown in Figure 3.

3.2.1. Image Filtering. Images or videos collected by machines are usually easily interfered by external factors such as illumination transformation or object motion [10, 11], thus affecting the useful information in detection. Interference is usually called noise, which is mainly generated under the influence of the external environment. Due to the interference of noise, several isolated pixels will
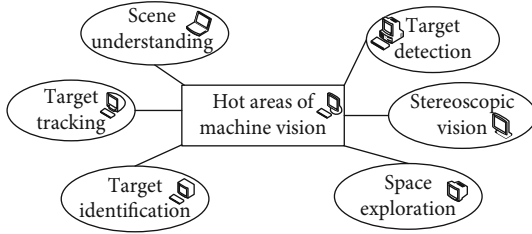
Figure 1: Application areas of machine vision.

appear on the image or video. If these pixels are not filtered in advance, it will have a huge impact on our subsequent image monitoring, judgment, and tracking. Different filters can be used to denoise according to the characteristics of the noise. Several commonly used filtering methods are introduced.

(1) Mean Filter. Mean filtering, as the name suggests, is to filter the video image by taking the average value. The working principle of the filter is to first select an area composed of several pixels for the pixel to be processed (m, n) and then randomly take the average value of the pixel values in the area to replace the original pixel value. The formula can be expressed as

$$f(m, n) = \frac{1}{M} \sum_{i,j \in N} g(i, j). \tag{1}$$

In the formula, $f(m, n)$ represents the final result value of filtering, N is the target area, and M is the number of pixels in the target area N. The degree of filtering is related to the size of N.

(2) Median Filter. Median filtering and mean filtering have certain commonality in method. The difference is that the median filter replaces the value of each pixel in the region with the median value.

The specific method is shown in

$$f = \text{med}\{a1, a2\} = \left\{ m(n + 1), \frac{1}{2m(m + 1)} \right\}. \tag{2}$$

In the formula, $a1$ and $a2$ are ordered number sequences and f represents the median value in the number sequence.

In order to objectively evaluate the effect of these two filtering methods, this paper introduces the peak signal-to-noise ratio (PSNR) as the standard of image evaluation. PSNR is one of the most widely used image quality evaluation methods at present, and its definition is shown in

$$\text{PSNR} = 10 \times \lg \left(\frac{x_{\max}^2}{\text{MSE}} \right). \tag{3}$$

In the formula, $x_{\max}$ is the maximum value of the original image noise, MSE is the mean square error, and its definition formula is as follows:

$$\text{MSE} = \frac{1}{AB} \sum_{m=1}^{M} \sum_{n-1}^{N} (y_k(a, b) - y_{k-1}(a, b))^2. \tag{4}$$

In the formula, A and B represent the length and width of the image, respectively, and $y_k(a, b)$ and $y_{k-1}(a, b)$ represent the current image noise value and the reference noise value, respectively.

By calculating the PSNR value of the filtering method, we found that the PSNR value of the median filter is the highest, which means that the median filter has the best filtering effect, and the noise suppression effect is the best choice in the filtering technology. Moreover, since the median filter does not produce image edge blur, it can better preserve the edge characteristics of the image to facilitate subsequent edge extraction. Therefore, in this study, median filtering is selected as the image filtering processing technique in image preprocessing.

3.2.2. Moving Target Monitoring Method. After preprocessing the image, we believe that we can easily find and detect the region containing moving objects in the video or image sequence. At the same time, we use methods such as image segmentation to detect and optimize the best description function of the moving target area to avoid the dynamic environment and the impact on the detection. The following are some commonly used moving target monitoring methods.

(1) Time Difference Method. The time difference method, also known as the interframe difference method, is one of the simplest and most commonly used moving target monitoring methods. The working principle of the method is to calculate the time difference between the images of two adjacent frames or adjacent multiple frames in the image and perform thresholding processing on the difference image. The time difference method is relatively simple to implement in principle and operation method and can realize fast positioning and detection of moving targets. It has a good self-adaptive ability to the real-time changes of the monitoring scene and is suitable for monitoring occasions with high real-time requirements. However, this method is limited by the moving speed of the target. On the one hand, if the target moves too fast, the intersection of two consecutive frames may be empty, and it will be very difficult to precisely locate the moving target. On the other hand, if the target stops moving, there will be cases where the presence of a moving target cannot be detected.

(2) Background Subtraction. In order to better obtain the complete information of the moving target area, the background subtraction method came into being. It first analyzes the sequence of the acquired video or image on the time scale and then uses the sequence to analyze the pattern changes in the video surveillance scene and then establishes
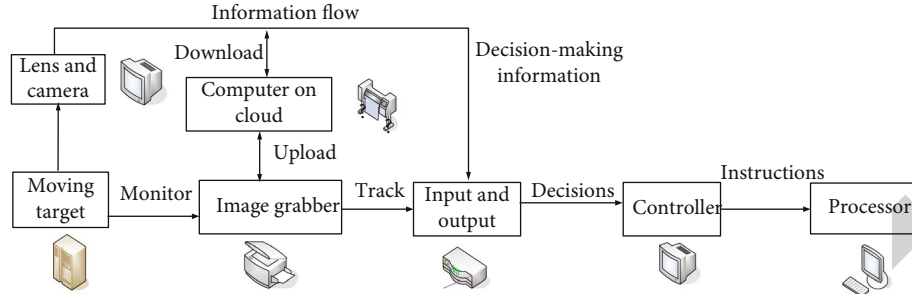
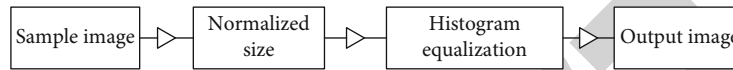Figure 2: The working principle of machine vision.

Figure 3: The specific implementation process of moving target monitoring.

a mathematical model of the video or image. Finally, it realizes the monitoring of moving objects by comparing the background of the image and the video surveillance scene. The advantage of this method is that the principle and calculation are very simple, and the operation speed is relatively fast.

(3) Optical Flow Method. The optical flow method refers to the speed of pattern motion in time-varying images, which was first proposed in the 1940s [12]. In moving object detection, the optical flow method estimates the motion field according to the image sequence and then distinguishes the background and foreground of the image according to the change of the motion field. The optical flow method contains the motion information of the monitored target, so it not only can be used for moving target detection but also can be used for target tracking directly. At the same time, the optical flow method can realize the monitoring and tracking of moving objects in dynamic and static backgrounds.

(4) Feature Matching Method. When the camera moves, the foreground and background points of the video image will become indistinguishable, and background modeling is difficult to establish. Considering the interference of system noise and environmental noise in the real monitoring environment, the idea of using the relevant features of moving objects to realize moving object detection comes into being. The core content of the so-called feature matching method is to use different methods for target modeling for different types of moving targets and then use the target classification method of pattern recognition to achieve real-time detection of moving targets. The process generally includes the following:

Learning process: use the foreground and background information of the target as positive samples and negative samples to train the classifier, and the classifier uses a large amount of training data to achieve certain results through operations.

Judgment process: calculate the similar features of the current image, use the trained classifier to judge the category, and update the corresponding parameters of the classifier in time according to the judgment result.

3.3. Target Intelligent Tracking in Video Surveillance. Intelligent tracking of moving objects is achieved on the basis of successful detection in videos or images [13]. The key to the real-time and effective tracking of moving objects is how to quickly and effectively realize the correlation matching of each moving object in the video image sequence. The basic realization process is shown in Figure 4. Its operation process mainly includes the following:

(1) Use the moving object detection method to realize the effective detection of motion, extract the characteristics of the moving object on the basis of the effective segmentation of the moving object, and construct a representation model of the moving object feature

(2) Detect the newly entered moving object in the scene, extract its effectively expressed motion features, and define the state of the newly detected moving object through the analysis and matching of the motion features

(3) Utilize the tracking result of the target to update the feature information of the moving target and its tracking model, and estimate the possible motion state information of the tracking target

3.3.1. Target Matching. Target matching is to calculate the similarity and relevance of the target to be detected and use data to measure. The association matching of moving objects will change due to changes in the video surveillance scene. The plane elastic constraint is an effective target constraint method, which can match the points at the same position in the two video images through the plane homography matrix to obtain the matching target. At the same
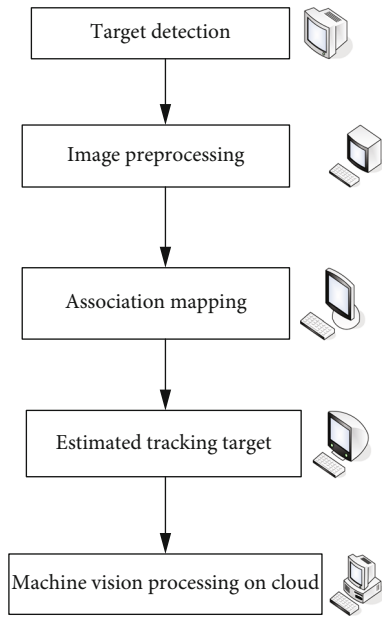
Figure 4: Target intelligent tracking process realization.

time, for the points in different video images, it can also reestablish the matching between the target and the target by means of 3D reconstruction. There are generally the following methods to obtain the association matching between moving objects.

(1) SURF Feature Detection Method. In 1999, some scholars proposed the SIFT algorithm to sample and reconstruct images to obtain multi-image data, but its computational complexity was too high and was discarded later. In 2006, on the basis of SIFT, a SURF (Speed-up Robust Feature) method based on acceleration robustness was proposed. Its basic idea is to use the median filter to construct the scale space and use the integral in the scale space to quickly complete the image matching.

(2) Feature Point Recognition Method. After the features of the two video images are basically generated and determined, the SURF algorithm describes the Euclidean distance of the vector based on this feature as a measure of the similarity between the images. But this method has certain limitations. The specific method of feature point-based recognition is to randomly acquire a feature point in the image A at first. Then, it looks for the feature point with the smallest Euclidean distance in the image B and finds two points with the smallest distance from it, namely, the optimal matching point and the second optimal matching point. At this time, if the ratio of the closest distance to the next closest distance is less than a fixed value, this pair of matching points is selected. But if the two are very close, it might be wrong to choose one of them, thus rejecting both matching points. By reducing the preset ratio threshold, although the number of matching feature points obtained will be less, the matching result will be more accurate and stable.

(3) Local Feature Method. The three-dimensional matching method based on local features includes several steps such as key point search, key point quality calculation, feature extraction, and target matching. The initialization of the 3D model can be described as follows: first, the local features of the scene image are extracted, and the scene is matched with a series of local depth images from different perspectives used in the reconstruction of the 3D model to find the depth image with the largest matching point with the scene. This shows that in this series of depth images from different perspectives, this image is the most similar to the pose of the target in the scene. After initializing the 3D model offline, the scene image is retrieved with the described algorithm to retrieve the target to be matched in the cluttered scene, that is, to find some key points that roughly match the model. All the point clouds outside the continuous surface where these key points are located are removed, which is background segmentation.

(4) ICP Method of Fusion Texture. Using two point sets, Euclidean distance sum formula and singular value decomposition method to calculate the value, iteratively update until the preset error threshold is met. The algorithm only considers the geometric information of the image, which is prone to more mismatch points. It also expands the idea of the traditional ICP algorithm and changes the traditional ICP algorithm using only the Euclidean distance of geometric coordinates as the basis for finding the closest point to a six-dimensional distance formula that combines geometric information and texture information. It uses the geometric information and texture information separately. Compared with the Color_ICP algorithm, since only the geometric information is used in the search, it improves the speed of the nearest point search; that is, the matching efficiency can be improved.

3.3.2. Target Tracking. The tracking problem of moving objects is a very challenging research direction. When the moving target moves in the scene, its motion state and its corresponding motion information are changing in real time, and the objective environment of the monitoring scene is also dynamically adjusted. At the same time, it takes into account the mutual occlusion between moving objects when moving and the reciprocating entry and exit of the moving target in the monitoring scene building; also, the movement of the related monitoring and acquisition equipment and the network transmission rate of the monitoring video data may have a huge impact on the results of target tracking. Therefore, in practical applications, target tracking is carried out in experiments and tests in a hypothetical monitoring environment. For different monitoring environments and moving target characteristics, researchers have proposed many classical tracking algorithms [14, 15].

(1) Kalman Filter. The Kalman filter problem is simply a problem of predicting a target state, that is, using the existing state information to predict and estimate its possible motion state through statistical operations. And it reduces the gap between the estimated value and the

actual measured value as much as possible, and its description formula is as follows:

$$Y_{k+1} = Cx_k + Du_k + v_k. \tag{5}$$

Among them, C and D are both system state matrices, and random signals u_k and v_k represent system process noise and measurement noise, respectively.

(2) Mean Shift Tracking Method. The mean shift algorithm was first proposed in 1975, and its original meaning was the mean vector of the shift. But with the development of the theory, its meaning has undergone many changes. Now usually it refers to an iterative step, which is a hill-climbing algorithm based on kernel density estimation, which can be used for clustering, image segmentation, tracking, etc. The tracking of moving objects using the mean shift method is mainly composed of the following steps:

(1) Construct the feature representation model of the moving target

$$Q_u = C \sum_{i=1}^{n} g(x_0 - x_i h_i) \partial \left[a(x_i)^2 - u \right]. \tag{6}$$

The kernel function is $g(x)$, h represents the bandwidth of the function, and C is the correlation coefficient of the normalized function. When $\sum_{u=1}^{m} Q_u = 1$, it can get

$$C = \frac{1}{\sum_{i=1}^{n} k(x_0 - h_i)}. \tag{7}$$

(2) Establish a description model of the target candidate region

Similar to the steps, if the particle coordinate of the candidate target region is y, then the kernel function can be calculated by the following formula:

$$P_u(y) = C_h \sum_{i=1}^{n} k \partial [b(x_i) - u], \tag{8}$$

where C_h is the correlation coefficient of the normalized function, when $\sum_{u=1}^{m} P_u(y) = 1$:

$$C_h = \frac{1}{\sum_{i=1}^{n} k(x_0 - h_i)}. \tag{9}$$

(3) Similarity measurement between the moving target feature description model and the candidate region model

In the similarity measurement of the color histogram distribution between the moving target feature description

TABLE 1: Result 1.

	Median filter	Kalman filter	Mean shift	Ours
David	80.22	40.13	66.12	70.95
Car	71.21	11.54	45.23	89.21
Car chase	34.19	22.98	43.65	65.32

TABLE 2: Result 2.

	Median filter	Kalman filter	Mean shift	Ours
Volkswagen	11.76	17.78	19.31	20.09
Motocross	22.32	25.31	25.11	30.16
Couple	22.19	23.83	24.91	25.16

model and the candidate region model, the Bhattacharyya coefficient is commonly used in the mean shift tracking method. The formula for calculating the coefficient is as follows:

$$d(y) = \sqrt{1 - q(y)}. \tag{10}$$

Among them, $q(y)$ represents the Bhattacharyya coefficient of the moving target description model and the candidate area model.

$$q(y) = q[q(y), s] = \sum_{u=1}^{m} \sqrt{q_u(y), s}. \tag{11}$$

The mathematical meaning of this coefficient is that the larger the value of $q(y)$, the higher the similarity between the moving target description model and the candidate area model; that is, the moving target candidate area corresponding to $q(y)$ is the current real coordinate position of the moving target we are looking for.

(4) Positioning of moving targets

According to the similarity measurement result between the moving object description model and the candidate area model, when positioning the moving object, the center position y of the moving object in the previous frame of the image is the starting point. It calculates the best matching moving object location area in the current image frame and assigns the centroid of this area to y. Using Taylor's formula for the candidate region model $q(y)$ in the current graphics frame, it can be expanded to get

$$p[q(y), \widehat{q}] \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{q(y_0), q_u} + \frac{C_h}{2} \sum_{i=1}^{n_h} w_i k(y - x_i)^2, \tag{12}$$

$$w_i = \sum_{u=1}^{m} \partial [b(x_i) - u] \sqrt{\frac{q_u}{p_u(y_0)}}. \tag{13}$$

(a)

(b)

FIGURE 5: The tracking test success and accuracy results of the tracking algorithm.

It is observed that formula (12) has nothing to do with y, then let

$$g_{n,k} = \frac{C_h}{2} \sum_{i=1}^{n} w_i k(y - x_i)^2. \tag{14}$$

When formula (14) reaches the maximum value, the matching and positioning of the moving target are completed. Through correlation calculation, the particle y of the current moving target position region can be obtained by iterative calculation from y_0, as shown in

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i g(y - x_i)^2}{\sum_{i=1}^{n_h} w_i g(y - x_i)^2}. \tag{15}$$

Among them, $g(k) = -k(x)$. The target monitoring efficiency of this algorithm is better than that of other algorithms, but this method often needs multiple iterations

to complete the positioning of moving targets during calculation.

Through the analysis and calculation of the Kalman filter and the mean shift method, we can conclude that the Kalman filter realizes the prediction and update of the motion state of the moving target through the ideal linear estimation of the real-time monitoring scene. Its implementation is relatively simple, and it has good real-time application capabilities for monitoring scenarios with relatively simple environments. Unfortunately, the state model and measurement model of the real tracking system are mostly nonlinear. When the nonlinearity is relatively strong, the Kalman filter usually cannot obtain practical results. The mean shift tracking method uses the color histogram distribution of the moving target area as a feature description model. It calculates the optimal value of the similarity through the iterative recursive formula through the similarity measure of the feature model between each moving target, so as to realize the real-time tracking and positioning of the moving target. The algorithm is a parameter-free similarity matching algorithm for kernel density estimation, which has a good adaptability to the nonlinear changing monitoring scene environment. But its disadvantage is that it often ignores the spatial information of the moving target, especially when the tracking target moves too fast or is occluded.

Therefore, according to the environmental characteristics of actual video surveillance, this paper decides to use the basic principle of the mean shift tracking algorithm and organically integrate the median filter image processing technology and proposes a target intelligent tracking strategy based on machine vision.

4. Moving Target Monitoring and Intelligent Tracking Strategy Based on Machine Vision

Before the experiment, we first use the median filter image processing technology to preliminarily process the detection results of moving objects, and then, we can use the position function $(q^{\wedge^{(I)}})_{I=1,\cdots L}$ corresponding to multiple subblocks to characterize. In the function, the letter L represents the number of blocks in the moving area, and $q^{\wedge^{(I)}}$ is the subblock position kernel function of the moving target. Then, the position of the moving target can be characterized by the similarity measure with the candidate region model whose position is y:

$$\widehat{q}_y = \partial^{(1)} q^{(1)} + \partial^{(2)} q^{(2)} + \cdots \partial^{(L)} q^{(L)} = \sum_{i=1}^{L} \partial^{(i)} q^{(i)}. \quad (16)$$

Among them, $\partial^{(i)}$ represents the weight of moving target monitoring, and the definitions of $\sum_{i=1}^{L} \partial^{(i)} = 1$ and $\partial^{(i)}$ are shown in

$$\partial^{(i)} = \frac{q\left[\overset{\wedge^{(1)}}{q}(y), p\right]}{\sum_{i=1}^{L} q\left[\overset{\wedge^{(1)}}{q}(y), p\right]}. \quad (17)$$
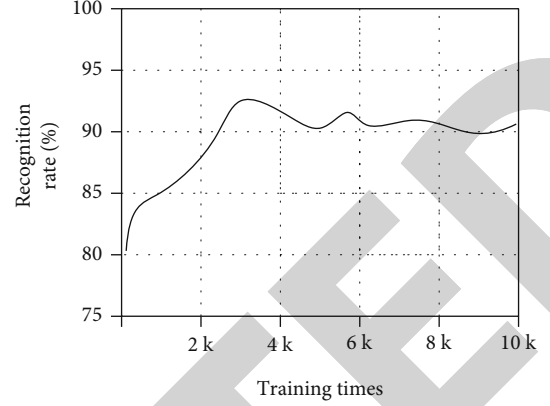

FIGURE 6: Recognition rate and training times.

The larger the value of $\partial^{(i)}$, the higher the similarity. p represents the maximum value of the Bhattacharyya coefficient of the moving target description model and the candidate area model. Therefore, as long as the absolute values of p and $q(y)$ are minimized, the minimum value of the spatial distance can be obtained.

$$q^{(i)} \equiv \frac{p\left[\left(\overset{\wedge^{(1)}}{q}(y)\right), \overset{\wedge^{(L)}}{q}(y)\right]}{\sum_{i=1}^{L}\left[\overset{\wedge^{(1)}}{q}(y)\right]} = \sum_{u=1}^{m} \sqrt{p_u(y)q_u}. \quad (18)$$

After the possible target position of the moving target is calculated, the activity range of the moving area can be roughly obtained. When there is only a single moving object in the scene area, the relative motion state of the moving object can be directly updated.

During the intelligent tracking of moving objects, the motion position of the moving object can be calculated and predicted by using the motion state mimic formula. It achieves the purpose of reducing the calculation iteration times of the mean shift algorithm and narrowing the search range of the moving target and effectively improves the matching accuracy of the moving target [16, 17].

Let the motion state of the moving target at time n be

$$x_n = (x\_\text{center}_n, y\_\text{center}_n, x\_v_n, y\_v_n, \text{width}, \text{height}). \quad (19)$$

Taking the motion state corresponding to formula (19) as the input of the median filter, then the possible position of the moving target at time $n + 1$ is calculated as follows:

$$x\_\text{center}_{t+1} = x\_\text{center}_t + x\_v_t \cdot \theta t, \quad (20)$$

$$y\_\text{center}_{t+1} = y\_\text{center}_t + y\_v_t \cdot \theta t. \quad (21)$$

Considering the limited range of motion displacement of moving objects in adjacent multiframe images, this paper defines an evaluation function as follows:

$$V(i, j) = \alpha \cdot C(i.j) + \beta \cdot D(i, j). \quad (22)$$

TABLE 3: Success rate of specific action.

	Walk	Beat	Stoop	Run	Wander	Squat
Walk	100	3	1	9	15	9
Beat	5	120	9	18	5	3
Stoop	3	2	103	2	3	12
Run	1	2	9	99	0	2
Wander	10	12	0	6	90	5
Squat	4	3	11	8	10	100
Success rate (%)	80.12	91.11	86.32	85.75	86.11	90.54

TABLE 4: The success rate of recognizing moving objects.

	Ours	Median filter	Kalman filter	Mean shift	KCF	TGPR
Walk	0.778	0.716	0.7	0.684	0.654	0.643
Beat	0.7	0.653	0.613	0.6	0.553	0.512
Stoop	0.624	0.579	0.506	0.531	0.476	0.486
Run	0.415	0.426	0.399	0.298	0.324	0.33
Wander	0.141	0.084	0.1	0.063	0.042	0.039
Squat	0.112	0.083	0.2	0.04	0.092	0.396

$C(i, j)$ represents the distance similarity between the ith moving object and the jth object between adjacent frames in the video surveillance scene. $D(i, j)$ represents the displacement position between the ith moving object and the jth object between adjacent frames [18, 19].

The results of median filter, Kalman filter, mean shift, and our method under the DP (%) criterion are shown in Tables 1 and 2.

Tables 1 and 2 show that our machine vision-based target intelligent tracking results perform the best, with an overall average of 60.95%. The closer the tracking samples are to the real target, the better the tracking performance can be guaranteed.

The tracking attribute test results of the target intelligent tracking algorithm based on machine vision on OTB100 are shown in Figure 5.

Figures 5(a) and 5(b) show the average tracking accuracy and success rate of mainstream tracking algorithms including the mean shift algorithm on 11 target tracking attributes. These tracking properties are Ray Change (IV), In-Plane Auto Rotation (IPR), Out of View (OV), Occlusion Change Rate (OCC), Moving Object Deformation (DEF), Out-of-Plane Auto Rotation (OPR), Moving Objects Scale Change (SV), Moving Object Ambiguity (MB), Fast Object Movement (FM), Video Background (BC), and Resolution Change (LR) [19, 20].

After combining the "cloud" technology, we upload this algorithm to the cloud, and let the computer train itself to capture the moving target to complete the training. As shown in Figure 6, when the number of training times reaches about 3k, the recognition rate of target monitoring based on machine vision is stable at 90%. And with the continuous deepening of training, its recognition rate tends to be stable.

In order to better verify the cloud-based machine vision technology, we finally obtained the data in Table 3 by monitoring and judging the different movements of moving objects in the video. The results show that after using the machine vision tracking algorithm combined with "cloud," the success rate of the algorithm in recognizing moving objects has been upgraded to a certain extent.

Table 3 shows that after adding cloud technology, the algorithm based on machine vision is very stable, and the overall average success rate of action recognition for a single moving object reaches 86.95%. At the same time, in order to find the shortcomings of the algorithm, we will test the behavior of the characteristics.

It can be seen from Table 4 that our algorithm has certain advantages in identifying characteristic target actions, but for some actions, our algorithm has a relatively low recognition rate.

Figure 7 clearly compares our recognition rate with other algorithms.

As shown in Figure 7(a), in terms of the accuracy of recognizing moving objects, our algorithm can basically complete the goal and has a high success rate. Figure 7(b) shows that the recognition accuracy of our algorithm is above 80% on average.

In order to better characterize the detection accuracy results based on machine vision, this study selects other traditional algorithm transformations to compare with the algorithm used in this study.

Tables 5 and 6 present our detection results for the same video image.

Through statistical analysis of Tables 5 and 6, the algorithm designed in this study has better reproducibility and anti-interference than traditional monitoring and tracking methods [21, 22]. The reason is that, first of all, we divide
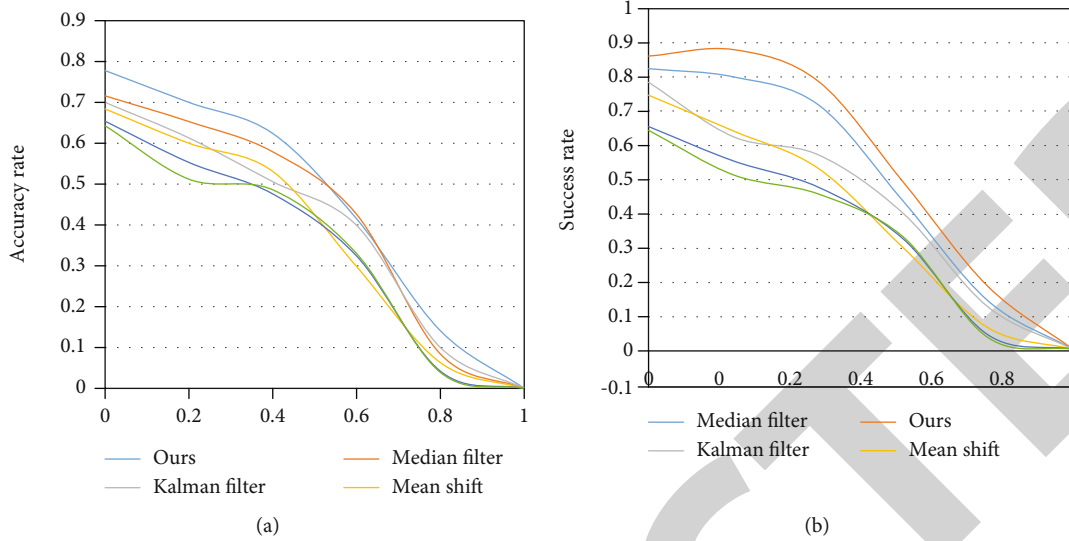
(a)

(b)

Figure 7: Success and accuracy rate of recognizing moving objects.

Table 5: Detection frame comparison.

	Ours	Median filter	Kalman filter	Mean shift	KCF	TGPR
Total frames	800	800	800	800	800	800
False frames	70	121	98	101	80	107
Error frames	50	68	99	72	94	69

Table 6: Detection accuracy comparison.

	Ours	Median filter	Kalman filter	Mean shift	KCF	TGPR
False detection rate (%)	8.75	15.12	12.11	12.5	10.01	13.21
Error detection rate (%)	6.25	8.5	12.17	10.94	11.98	8.32
Accuracy rate (%)	85.7	72.1	71.52	74.2	75.1	74.8

the detection area, which effectively avoids the interference of the video information in the nondetection area. At the same time, the median filter used in this study plays a good filtering and suppressing role in the case of cluttered information and interference noise, so the false detection rate and missed detection rate are greatly reduced [23, 24].

The overall situation of the number of false detections and missed detections in the experiment is shown in Figure 8.

It can be seen from Figure 8(a) that our algorithm maintains more than 80% accuracy, but the error and false detection data are still relatively high. Figure 8(b) shows that the median filter differs from us in the dimension of accuracy; that is, its recognition accuracy is not as good as our algorithm, but its error rate is lower than ours. As can be seen from Figure 8(c), the recognition accuracy of the Kalman filter is significantly different from ours.

The accuracy of intelligent tracking of moving targets is shown in Figure 9.

As can be seen from Figure 9(a), after comparing the error rate and the omission rate horizontally, our algorithm

has the least number of error frames and lost frames. The reason for this can be attributed to the low error rate of our algorithm for specific target actions. As can be seen from Figure 9(b), other algorithms significantly lag behind us in both accuracy and speed.

Although our algorithm is somewhat lower than other algorithms in terms of missed detection and false detection rate, the average success rate of comprehensive tracking monitoring is 85.7%. But in addition, we can also observe that other algorithms have certain advantages in false detection or missed detection. Although they lag slightly behind this algorithm, they also have good performance. At the same time, it also shows that the monitoring and tracking algorithm designed in this paper has considerable room for improvement.

5. Discussion

Video surveillance technology based on machine vision plays an increasingly important role in modern surveillance systems. It can not only improve the security of our social
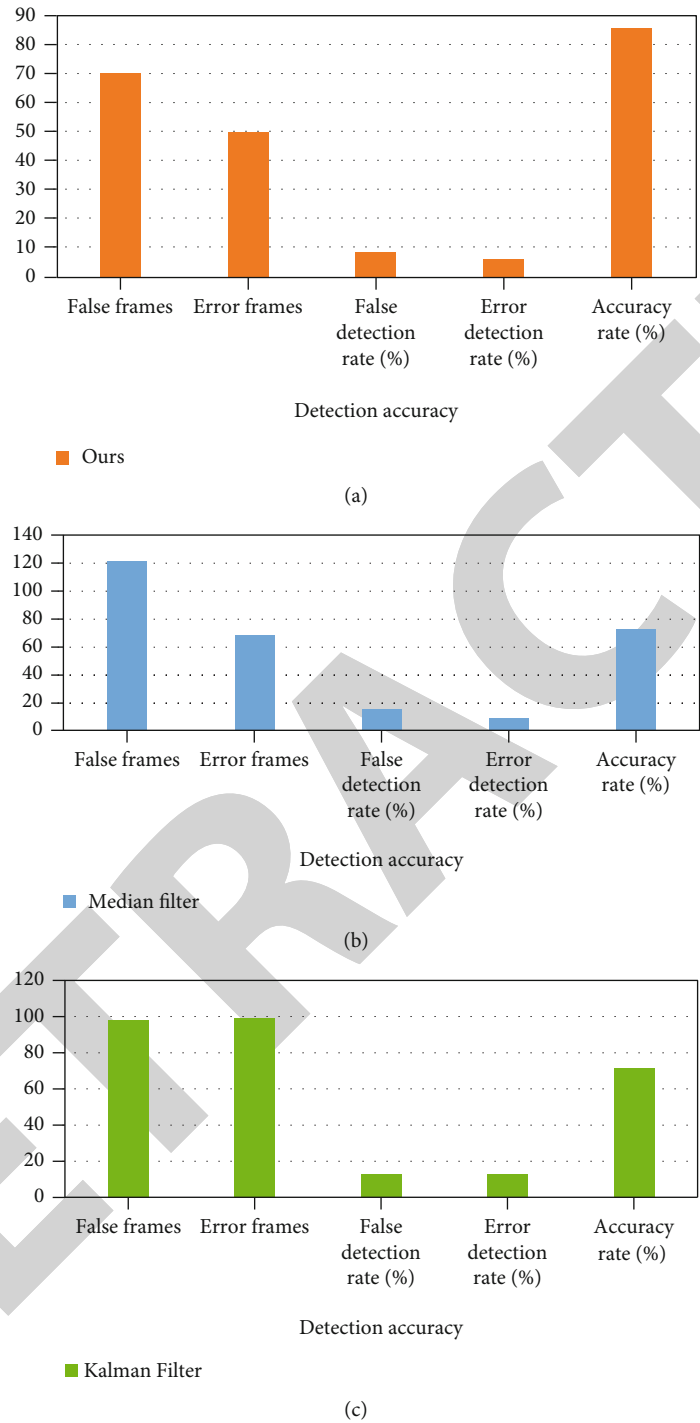
(a)

(b)

Figure 8: Detection accuracy comparison.

governance and effectively avoid accidents but also improve the utilization and processing efficiency of information. At the same time, the combination of machine vision and smart city technologies will become one of the important components of the digital earth, so as to realize the integration of information collection and processing [25, 26]. At present, with the performance improvement and cost reduction of equipment such as digital processors and cameras, machine vision-based video surveillance strategies will become the mainstream. On this basis, the moving target monitoring and intelligent tracking algorithms are constantly being upgraded and adjusted, giving different attention to the different states and actions of moving targets. After acquiring its different states and behaviors, the tracking algorithm will be further iterated to meet the actual needs. Based on this, in the context of artificial intelligence, video surveillance will be automatically captured and analyzed by machine vision by capturing and analyzing moving objects in images or videos
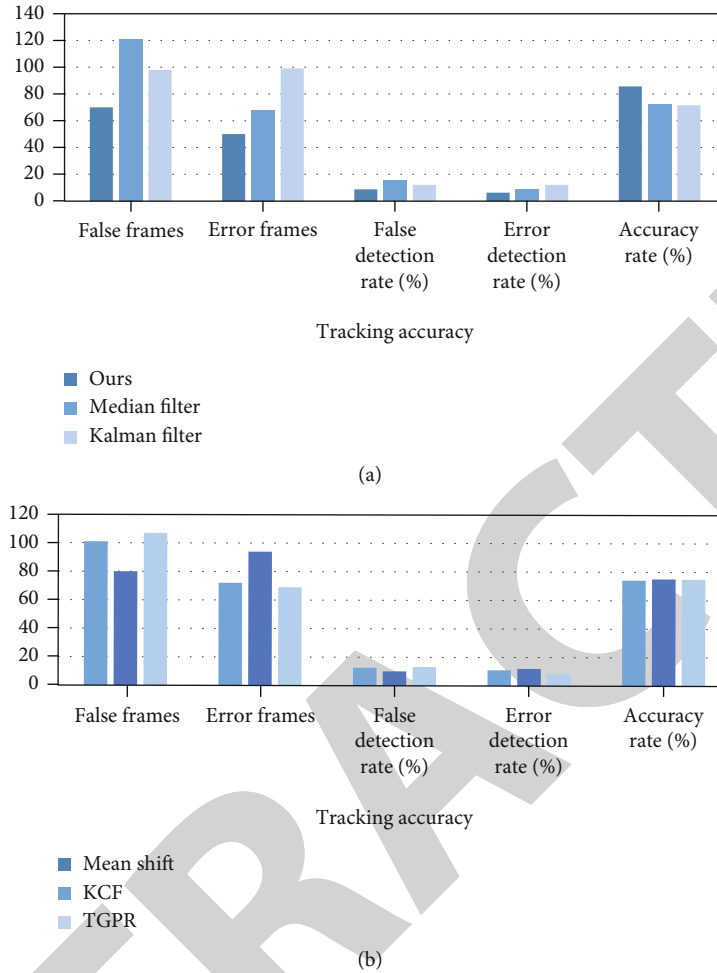
(a)

(b)

FIGURE 9: Tracking accuracy changes.

[27, 28]. With the help of machine vision, the staff can quickly monitor and track moving targets in the scene, so as to give full play to the maximum effect of machine vision. After combining with cloud technology, the abnormal behavior of moving objects will be quickly detected by video surveillance and tracked and alerted in the fastest and best way [29].

6. Conclusion

Video surveillance is a research hotspot in the field of machine vision in recent years, and its related research results have played a huge role in intelligent traffic management, social public safety management, and smart city construction. This paper proposes the research on moving target monitoring and intelligent tracking algorithm based on machine vision technology combined with "cloud." However, there are still many technical difficulties to make the video surveillance system play a greater role in future social governance and security work. First of all, in the intelligent tracking method based on machine vision and moving target monitoring, the theoretical support of the algorithm needs to be further studied. Second, this article only discusses

how to combine machine vision and video surveillance. In real life, if considering the machine vision application of other platforms and other types of sensory data, this kind of combination method has to break through slowly in technology. How to closely link the relevant algorithms and technologies of the monitoring of moving objects and the effective tracking of moving objects in this paper with the actual needs and development direction of the current video surveillance system to provide necessary technical support for building a safe and harmonious social environment will be an urgent problem to be solved in the future.

Data Availability

This article does not cover data research. No data were used to support this study.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

References

[1] A. A. Robie, K. M. Seagraves, S. Egnor, and K. Branson, "Machine vision methods for analyzing social interactions," *Journal of Experimental Biology*, vol. 220, no. 1, pp. 25–34, 2017.

[2] Z. Zhao, L. Shen, C. Yang, W. Wu, M. Zhang, and G. Q. Huang, "IOT and digital twin enabled smart tracking for safety management," *Computers & Operations Research*, vol. 128, no. 5, pp. 105–183, 2020.

[3] R. Liang, H. Zhi, and M. M. Kamruzzaman, "Methods of moving target detection and behavior recognition in intelligent vision monitoring," *Acta Microscopica*, vol. 28, no. 4, pp. 750–759, 2019.

[4] F. K. Wang, P. H. Juan, S. C. Su, M. C. Tang, and T. S. Horng, "Monitoring displacement by a quadrature self-injection-locked radar with measurement- and differential-based offset calibration methods," *IEEE Sensors Journal*, vol. 19, no. 5, pp. 1905–1916, 2019.

[5] F. R. Danion and F. J. Randall, "Different gaze strategies during eye versus hand tracking of a moving target," *Scientific Reports*, vol. 8, no. 1, pp. 10059-10060, 2018.

[6] C. Wu, W. Xing, Z. Feng, and L. Xia, "Moving target tracking in marine aerosol environment with single photon lidar system," *Optics and Lasers in Engineering*, vol. 127, p. 105967, 2020.

[7] Z. Li, X. Chen, and Z. Zhao, "Design of standoff cooperative target-tracking guidance laws for autonomous unmanned aerial vehicles," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6682160, 14 pages, 2021.

[8] A. Aytekin, Y. Rezaeitabar, S. Dogru, and I. Ulusoy, "Railway fastener inspection by real-time machine vision," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 7, pp. 1101–1107, 2015.

[9] L. Fernandez-Robles, G. Azzopardi, E. Alegre, and N. Petkov, "Machine-vision-based identification of broken inserts in edge profile milling heads," *Robotics and Computer-Integrated Manufacturing*, vol. 44, pp. 276–283, 2017.

[10] T. Dawood, Z. Zhu, and T. Zayed, "Machine vision-based model for spalling detection and quantification in subway networks," *Automation in Construction*, vol. 81, pp. 149–160, 2017.

[11] D. M. Tsai and Y. C. Hsieh, "Machine vision-based positioning and inspection using expectation–maximization technique," *IEEE Transactions on Instrumentation & Measurement*, vol. 66, no. 11, pp. 2858–2868, 2017.

[12] M. Abdollahpour, M. R. Golzarian, A. Rohani, and H. Abootorabi Zarchi, "Development of a machine vision dual-axis solar tracking system," *Solar Energy*, vol. 169, pp. 136–143, 2018.

[13] L. Panahi and V. Ghods, "Human fall detection using machine vision techniques on RGB-D images," *Biomedical Signal Processing and Control*, vol. 44, pp. 146–153, 2018.

[14] W. S. Yue, K. K. Chye, and C. W. Hoy, "Towards smart mobility in urban spaces: bus tracking and information application," *AIP Conference Proceedings*, vol. 1891, no. 1, pp. 1–6, 2017.

[15] O. I. Khalaf, C. A. T. Romero, A. Azhagu Jaisudhan Pazhani, and G. Vinuja, "VLSI implementation of a high-performance nonlinear image scaling algorithm," *Journal of Healthcare Engineering*, vol. 2021, Article ID 6297856, 10 pages, 2021.

[16] M. Elhoseny, G. Ramírez-González, O. M. Abu-Elnasr, S. A. Shawkat, N. Arunkumar, and A. Farouk, "Secure medical data transmission model for IoT-based healthcare systems," *Ieee Access*, vol. 6, pp. 20596–20608, 2018.

[17] O. I. Khalaf and G. M. Abdulsahib, "Energy efficient routing and reliable data transmission protocol in WSN," *International Journal of Advances in Soft Computing and its Application*, vol. 12, no. 3, pp. 45–53, 2020.

[18] O. I. Khalaf, G. M. Abdulsahib, and B. M. Sabbar, "Optimization of wireless sensor network coverage using the bee algorithm," *Journal of Information Science and Engineering*, vol. 36, no. 2, pp. 377–386, 2020.

[19] S. Zhou, M. Ke, and P. Luo, "Multi-camera transfer GAN for person re-identification," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 393–400, 2019.

[20] J. Yang, C. Wang, B. Jiang, H. Song, and Q. Meng, "Visual perception enabled industry intelligence: state of the art, challenges and prospects," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2204–2219, 2020.

[21] S. Ding, S. Qu, Y. Xi, and S. Wan, "Stimulus-driven and concept-driven analysis for image caption generation," *Neurocomputing*, vol. 398, pp. 520–530, 2020.

[22] C. Grange and H. Barki, "The nature and role of user beliefs regarding a website's design quality," *Journal of Organizational and End User Computing (JOEUC)*, vol. 32, no. 1, pp. 75–96, 2020.

[23] H. S. Son, "Smart tracking algorithm for multi-static sonar based on expectation maximisation," *IET Radar, Sonar & Navigation*, vol. 14, no. 10, pp. 1624–1630, 2020.

[24] D. Grzonka, A. Jakóbik, J. Koodziej, and S. Pllana, "Using a multi-agent system and artificial intelligence for monitoring and improving the cloud performance and security," *Future Generation Computer Systems*, vol. 86, pp. 1106–1117, 2018.

[25] E. Erdol and Kohle and Group, "Total to develop artificial intelligence solutions with Google cloud," *Erdol Erdgas Kohle*, vol. 134, no. 6, pp. 61–61, 2018.

[26] C. H. Tseng, C. Lin, H. C. Chang et al., "Cloud-based artificial intelligence system for large-scale arrhythmia screening," *Computer*, vol. 52, no. 11, pp. 40–51, 2019.

[27] T. Needleman, "Cloud accounting just gets better and now it's getting ready to embrace artificial intelligence," *Accounting Today*, vol. 32, no. 1, pp. 22-23, 2018.

[28] A. Retico, M. Avanzo, T. Boccali et al., "Enhancing the impact of artificial intelligence in medicine: a joint AIFM-INFN Italian initiative for a dedicated cloud-based computing infrastructure," *Physica Medica*, vol. 91, p. 140, 2021.

[29] X. Cheng, L. Zhou, and Y. Zhang, "Object tracking algorithm based on multiple loss generative adversary," vol. 48, Tech. Rep. 3, Dongnan Daxue Xuebao (Ziran Kexue Ban)/Journal of Southeast University (Natural Science Edition), 2018.