

## Retraction

# Retracted: Sports Target Tracking and Speed Measurement Method Based on Time Domain Vision Sensor

### Security and Communication Networks

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Security and Communication Networks. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] M. Wang and W. Duan, "Sports Target Tracking and Speed Measurement Method Based on Time Domain Vision Sensor," *Security and Communication Networks*, vol. 2022, Article ID 8553888, 12 pages, 2022.

## Research Article

# Sports Target Tracking and Speed Measurement Method Based on Time Domain Vision Sensor

Minting Wang<sup>1</sup> and Wei Duan<sup>2</sup>

<sup>1</sup>Weinan Normal University, Weinan 714099, Shaanxi, China

<sup>2</sup>College of Physical Education, Xi'an Siyuan University, Xi'an 710038, Shaanxi, China

Correspondence should be addressed to Wei Duan; xasycollege@163.com

Received 8 March 2022; Revised 20 April 2022; Accepted 30 April 2022; Published 7 June 2022

Academic Editor: Muhammad Arif

Copyright © 2022 Minting Wang and Wei Duan. 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.

Moving target tracking technology is an important topic in the field of time domain sensors, which is widely used in medical diagnosis, intelligent robots, human-computer interaction, education, and entertainment. This article is based on the time domain vision sensor to explore the methods of sports target tracking and speed measurement. The article introduces the regional gray-scale correlation matching method in the vision sensor algorithm, the noise model of the time domain vision sensor, and the method of eliminating noise. At the same time, the tracking effects of SIFT algorithm, Mean Shift algorithm, and MSO algorithm on moving targets are studied, analyzing the ANF algorithm, TSML algorithm, and IHTLS based on the time domain visual sensor to compare the TUD-Crossing test set, the ETHBAHNHOF test set, and the SMOT data set; then the image matching was purified and analyzed based on the time domain visual sensor. The research results show that, in the method of exploring moving target tracking based on time domain vision sensor, when the improved SIFT algorithm uses feature adaptive fusion to describe the target, it can automatically allocate the proportion of color features and texture features, which improves the accuracy and reliability of moving target tracking. Among them, the DNP algorithm can control the final accuracy rate above 80%, and its recall rate is basically greater than 80%. In the case that the accuracy of rough matching is less than 50%, the final matching accuracy after processing by multiple purification algorithms is above 98%.

## 1. Introduction

With the improvement of sports competitive level, how to more easily and quickly master the training level of the players and provide the corresponding sports parameters to provide support for the players' daily practice has become an urgent problem for sports workers. In recent years, local governments have increased their research in the field of sports scientific research. In the past, the sports training model that relied solely on the observation records of the coaches and the intuitive judgment of the audience could no longer meet the requirements of the ever-increasing level of human competition. With the development of technology, sports target measurement and tracking technology is becoming more and more perfect. However, in sports videos, because the colors of the players' clothes and the background field may be the same, there is still cross-blocking between the players.

Since the visual sensor has better accuracy and memory, it can quickly capture the moving target. By collecting a large amount of video image information of high-level athletes in regular training and large-scale competitions, and effectively analyzing this information, it can change the inadequacy of coaches who only rely on manual observation and experience to guide athletes' technical movements in the past, greatly improving the training effect of athletes. At the same time, in the field of motion video analysis, moving target detection and tracking technology has played an indispensable role. Through real-time detection and tracking of athletes, the athlete's trajectory can be analyzed, and it is convenient to correct the subtle movement differences that cannot be detected by the athletes in training or competition, thereby improving the athlete's training effect and competition performance.

Combined with the research progress at home and abroad, different scholars have made certain researches on time domain sensors and target motion tracking: Feng and Feng research the performance in the time domain and frequency domain through the shaking table test of the three-story frame structure; the displacement of all floors is measured by using a camera to track high-contrast man-made targets or low-contrast natural targets (such as bolts and nuts) on the surface of the structure [1]. Lenero-Bardallo et al. proposed a new type of event-based vision sensor with two operating modes; the operation and readout mode can be switched through two control bits. The sensor has low latency and low power consumption when detecting spatial contrast [2]. Zhang and Chen proposed a feature-level data fusion method that automatically evaluates the welding quality of aluminum alloy tungsten gas shielded welding in real time. Continuously extract and select multiple characteristic parameters from sound and voltage signals, and carefully analyze the spectral distribution of argon atoms related to weld penetration before selecting the characteristic parameters [3]. Prasov and Khalil focused on measuring the effect of noise on tracking error. The purpose of Prasov's research is to provide readers with the relationship between the parameters of the high-gain observer and the tracking error and its subsequent derivatives [4]. Lyu and Smith compare three wireless swing sensors with high-speed video. The swing speed of the high-speed video is measured by distinguishing the position coordinates of the tracking mark on the bat. On average, the speed reported by the wireless bat swing sensor is 8% slower than the speed found by video tracking [5]. Lauzon-Gauthier et al. have developed a machine vision sensor that uses paste structure analysis to predict deviations from the optimal amount of pitch in the anode formulation. It can help operators mitigate the impact of changing anode raw materials (coke and pitch) [6]. Zhou et al. developed a three-dimensional vision sensor that uses 20 to 100 lasers arranged in a circular array, called omnidirectional dot matrix projection (ODMP). Based on the imaging characteristics of the sensor, ODMP can image areas with high image resolution [7]. Yang et al. research and implement the AER vision sensor bus arbitration simulator, which is used to simulate the process of event sampling, AER allocation, and object tracking. The experimental results show that the visual information generated by the AER vision sensor only accounts for 5%–10% of the total number of pixels of the image sensor. The target tracking accuracy of the event aggregation arbitration is higher than that of the specific area arbitration and the arbitration in the high-speed object tracking application of the round-robin method is reduced by 10%–20% [8]. However, these scholars did not effectively combine the time domain sensor with moving target tracking, but only talked about its unilateral significance.

The main content of the research and innovation of this article is reflected in the following: (1) Article introduces the algorithm based on time domain sensor and analyzes the SIFT algorithm of moving target tracking based on time domain sensor. (2) And it compared the algorithms based on time domain sensors and then based on the time domain sensors to track and detect the moving targets in the image, and at the same time, the image matching was purified and analyzed.

## 2. Method of Moving Target Detection Based on Time Domain Vision Sensor

*2.1. Algorithm Based on Time Domain Vision Sensor.* Based on the time domain vision sensor algorithm is the most important and most difficult step in the vision sensor algorithm, and it is also a research hotspot and difficulty in recent decades. It looks for one or more transformations in the transformation space. Due to changes in shooting time, shooting angle, and natural environment, the use of various sensors, and sensor defects, the captured images not only are affected by noise, but also have serious grays distortion and geometric distortion [9]. In this situation, how to achieve accuracy, high matching accuracy, fast speed, good robustness, and strong anti-interference ability of the matching algorithm will become the goal pursued by human beings. According to mathematical theories and methods, it provides humans with a new way of matching [10].

*2.1.1. Regional Gray-Scale Correlation Matching Method.* The matching algorithm based on regional gray scale is the most common and mature matching method used in stereo vision. The elements to be matched have a fixed-size pixel window, and the similarity criterion is the maximum similarity between two pixels windows [11]. The function of the pixel window is to make the similarity criterion in the region gray-based matching algorithm have a judging criterion. The matching algorithm is as follows.

As shown in Figure 1, on the  $i$ -th scan line, there is a feature point  $x_1(a, b)$  in the left image of the image pair. To find the corresponding feature point  $x_2$  in the right image, use a correlation window of area  $Z = (2m + 1) \times (2n + 1)$ , and the center of the window is located at feature point  $x_1$ . At the same horizontal scan line along  $x_1$  in the right picture, look for the feature point  $x_2$  within a certain parallax range  $(-r, +r)$  [12]. If  $H_1(a, b)$ ,  $H_2(a, b)$  represent the gray level at the feature point of  $(a, b)$  in the image pair,  $r$  represents the disparity, and  $\overline{H_1(a, b)}$ ,  $\overline{H_2(a, b + r)}$  represent the average gray level of the neighborhood where  $x_1$  and  $x_2$  are located, then the normalized covariance  $W(x_1, x_2)$  of  $x_1, x_2$  is defined as

$$W(x_1, x_2) = \frac{\sum_{x=-m}^m \sum_{y=-n}^n [H_1(x+a, y+b) - \overline{H_1(a, b)}] [H_2(x+a), H_2(y+b+r) - \overline{H_2(a, b+r)}]}{\left\{ \sum_{x=-m}^m \sum_{y=-n}^n [H_1(x+a, y+b) - \overline{H_1(a, b)}]^2 \right\}^{1/2} \left\{ \sum_{x=-m}^m \sum_{y=-n}^n [H_2(x+a), H_2(y+b+r) - \overline{H_2(a, b+r)}]^2 \right\}^{1/2}} \quad (1)$$

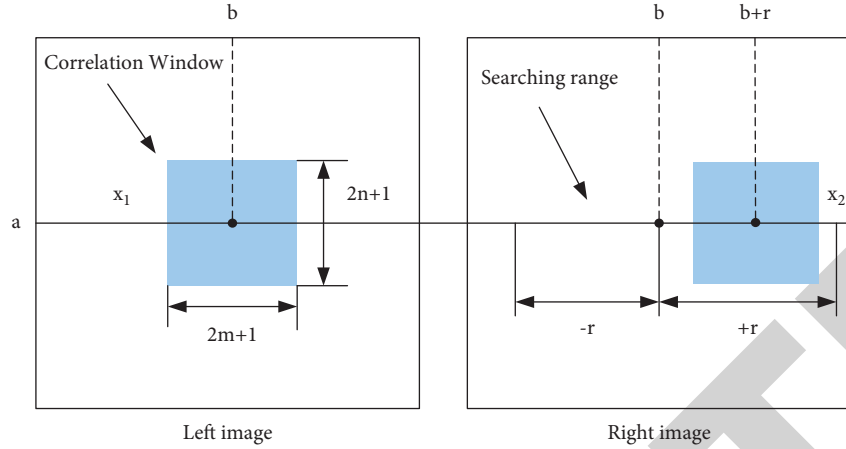


FIGURE 1: Schematic diagram of ZNCC matching algorithm.

The correlation coefficients between  $b$  feature points and  $T0$  neighborhood  $T-2, T-1, T0, T+1$ , and  $T+2$  are  $W-2, W-1, W0, W+1$ , and  $W+2$ . According to the subpixel positioning formula, the parallax  $r_{bt}$  between the characteristic point  $b$  and its corresponding point  $T$  is

$$r_{bt} = t_0 + \left(\frac{7}{20}\right) \frac{(2W_{-2} + W_{-1} - W_{+1} - 2W_{+2})}{(2W_{-2} - W_{-1} - 2W_0 - W_{+1} + 2W_{+2})} - b. \quad (2)$$

In order to reduce the amount of calculation and speed up the matching speed [13], the algorithm can be optimized from the following two aspects:

- (1) The parallax range is limited to reduce the amount of calculation.

Observe the parallax range of the entire image pair from the stereo image pair, set the parallax range to  $(-r, +r)$ , and then for a certain feature point in the left image, in the picture on the right, you only need to search along the range of  $(a-r, b)$  to  $(a+r, b)$  of the same horizontal scan line at that point, eliminating many unnecessary calculations [14].

- (2) Box filtering technology method speeds up the calculation.

Simplify the equation to

$$W(x_1, x_2) = \frac{\sum_{x=-m}^m \sum_{y=-n}^n [H_1(x+a, y+b)H_2(x+a, y+b+r) - ZH_1(a, b)\overline{H_2(a, b+r)}]}{\left\{ \sum_{x=-m}^m \sum_{y=-n}^n [H_1^2(x+a, y+b) - Z\overline{H_1(a, b)}]^2 \right\}^{1/2} \left\{ \sum_{x=-m}^m \sum_{y=-n}^n [H_2^2(x+a, y+b+r) - Z\overline{H_2(a, b+r)}]^2 \right\}^{1/2}} \quad (3)$$

Among them,  $\overline{H(a, b)}$  and  $\sum_{x=-m}^m \sum_{y=-n}^n H^2(a, b)$  are calculated according to the box filter method shown in Figure 2. Let  $H(b) = \sum_{y=-n}^n H(a+y, b)$ ; then,

$$\begin{aligned} \overline{H(a, b+1)} &= \overline{H(a, b)} + Z^{-1}[H(b+m+1) - H(b-m)], \\ \sum_{x=-m}^m \sum_{y=-n}^n H^2(a, b+1) &= \sum_{x=-m}^m \sum_{y=-n}^n H^2(a, b) + [H(b+m+1) \\ &\quad + H(b-m)][H(b+m+1) - H(b-m)]. \end{aligned} \quad (4)$$

Equations (4) and (5) are recursive formulas, the result of the  $b$ th operation can be applied to the  $(b+1)$ th operation, and only the  $(b+m+1)$ th column pixels and the  $(b-m)$ th column pixels need to be added and subtracted [15].

**2.1.2. The Source of Noise and the Noise Model of Domain Vision Sensor.** Image preprocessing is to separate each text image and hand it over to the recognition module for

recognition. This process is called image preprocessing. In image analysis, the processing performed before feature extraction, segmentation, and matching are performed on an input image [16]. In view of the uncontrollable factors such as noise and illumination changes in the visual sensor, this paper takes some measures to weaken their influence on the detection results and improve the reliability of the image matching algorithm [17]. This is because in the process of image acquisition and transmission, it is limited by the external environment of the vision sensor, lens, sensor chip, circuit design, and other factors, which will affect the quality of the image, resulting in the situation of adaptation and false detection. Therefore, noise is the primary problem to be solved in image preprocessing in visual sensor detection.

According to the CMOS noise model, it is convenient to study the influence of noise in the CMOS image sensor acquisition system on the image matching and detection results [18]. Here simply compare the normalized inverse mean absolute difference sum (NISAD), normalized inverse

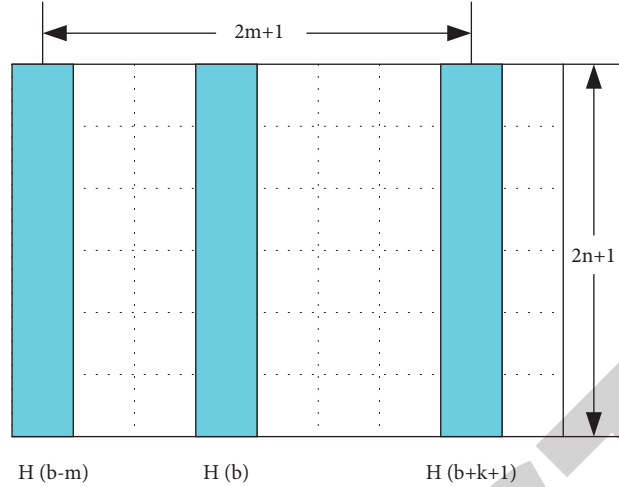


FIGURE 2: Schematic diagram of box filter filtering technology.

mean variance sum (NISSD), normalized inverse maximum absolute difference (NIMD), normalized direct correlation (NDC), and normalized cross-correlation (NCC); these similarity measures change under the influence of noise.

First, give the definition of the above measures:

$$\begin{aligned} \text{NIASAD}(a, b) &= 1 - \frac{1}{MN} \sum_p \sum_q |D(p, q) - A(a + p, b + q)|, \\ \text{NIASAD}(a, b) &= 1 - \frac{1}{MN} \sum_p \sum_q (D(p, q) - A(a + p, b + q))^2, \\ \text{NDC}(a, b) &= \frac{\sum_p \sum_q D(p, q) A(a + p, b + q)}{\sqrt{\sum_p \sum_q D^2(p, q) \sum_p \sum_q A^2(a + p, b + q)}}, \\ \text{NDC}(a, b) &= \frac{\sum_p \sum_q (D(p, q) - \bar{D})(A(a + p, b + q) - \bar{A}(a, b))}{\sqrt{\sum_x \sum_y D^2(p, q) \sum_p \sum_q A^2(a + p, b + q)}}. \end{aligned} \quad (5)$$

Figure 3 shows the effect of noise on different similarity measures. The left picture is the influence of different noises on the similarity measure. The abscissa is the code of the parametric noise model. The ordinate is the normalized similarity measure. It can be seen that the similarity measure is the same; the template becomes worse as the noise increases, but the impact is not great [19]. The figure on the right is the influence of different sizes of templates on the similarity measurement. The abscissa is the code name of the template. The larger the number, the larger the template (where 1 means the template is  $5 \times 5$ , 2 means the template is  $15 \times 15$ , and so on, 10 means the template is  $95 \times 95$ ); it can be seen that the NDC and NCC measurements are less affected by the size of the template, while the NIADD and NIASD measurements are greatly affected by noise under the small template, which seriously affects the matching results. The experimental results show that the size of the noise and the size of the template affect the value of the similarity measure to varying degrees, and the use of a suitable similarity measure can resist these effects to a certain extent [20].

The method to eliminate the influence of noise is mainly to reduce or eliminate the noise in the image by detecting noise, filtering, and other means. At present, the commonly used noise elimination methods mainly include the mean value method.

Mean value method is a linear processing method to eliminate image noise. The basic idea is to replace the gray level of the center pixel with the average gray level of the neighboring pixels, calculated as follows:

$$\bar{A}(x_0, y_0) = \frac{1}{\text{num}} \sum_{(x, y) \in N_0} A(x, y). \quad (6)$$

Num is the number of pixels in the neighbor. This method is simple and fast and can effectively eliminate Gaussian noise, so it is still a commonly used method [21]. There are still some defects in the mean method, which causes serious damage to the edge of the image and makes the image blurred, and its improvement is mainly to optimize its weight, also known as weighted mean.

$$\bar{A}(x_0, y_0) = \frac{1}{\sum_{(x, y) \in N_0} u(x, y)} \sum_{(x, y) \in N_0} u(x, y) A(x, y), \quad (7)$$

where  $u(x, y)$  is the weight.

Regarding the selection and improvement of weights, it mainly focuses on spatial correlation and gray-scale correlation. For example, considering the huge spatial correlation between real pixels in an image, weights related to distance can be used, such as

$$u(x, y) = \frac{1}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} \quad (8)$$

$$u(x, y) = \frac{1}{2\pi\epsilon^2} e^{-(x-x_0)^2 + (y-y_0)^2 / 2\epsilon^2}.$$

**2.2. SIFT Algorithm of Moving Target Tracking.** The core of the SIFT algorithm is to construct a scale space for the collected images, then find feature points in different scale



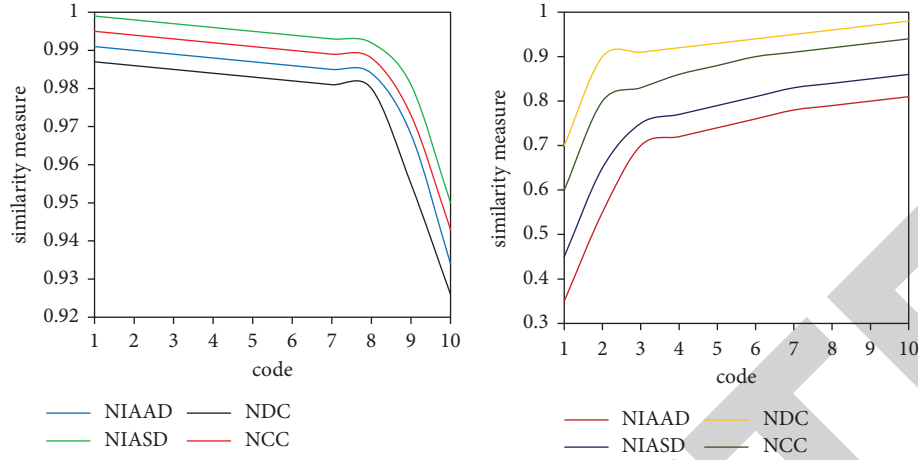


FIGURE 3: The effect of noise on the similarity measure.

spaces, and calculate the direction of the feature points. Through these feature points, it is found that it is not sensitive to scale scaling, image rotation, noise, and illumination but has an anti-interference effect on the change of angle [22].

The SIFT algorithm randomly selects samples to update the background model in a memoryless manner. The selection of samples has nothing to do with the length of time the samples exist and is not selected in the order of time. If 6 samples are randomly selected to update the background model, Figure 4 shows three different results of selecting samples.

The implementation of the SIFT algorithm is as follows.

**2.2.1. Detection of Extreme Points.** Before performing extreme point detection, the scale space must be constructed first [23]. The difference of Gaussian is used here, which is the DOG scale space. The DOG function is defined as

$$\begin{aligned} D(a, b, c) &= [H(a, b, kc) - H(a, b, c)] * P(a, b) \\ &= W(a, b, kc) - W(a, b, c). \end{aligned} \quad (9)$$

After constructing the DOG scale space, judge whether the sampling point is the maximum value of the scale space where the point of itself and its neighborhood is located. The basis for judging whether a sampling point is an extreme point is whether the sampling point is the maximum value of its domain scale space. If so, it is said that the sampling point is a local extreme point of the image; otherwise it is not.

**2.2.2. Feature Point Positioning.** The collected images will inevitably have noise, so that the extreme points extracted by the above steps may not necessarily be the feature points we want, so the second step is required: precise positioning of the feature points. Here, the Taylor expansion of the DOG function is as follows:

$$D(A) = D + \frac{\alpha D^R}{\alpha A} A + \frac{1}{2} A^R \frac{\alpha^2 D}{\alpha A^2} A. \quad (10)$$

Suppose the extreme point is  $\bar{A} = (a, b, c)^R$ , where  $a$  represents the row of the image,  $b$  represents the column of the image, and  $c$  represents the scale of the image. In the process of calculation,  $a$ ,  $b$ , and  $c$  need to be correct, as in the formula:

$$\bar{A} = -\frac{\alpha D^R}{\alpha A} \left( \frac{\alpha^2 D}{\alpha A^2} \right)^2. \quad (11)$$

After correction, get the formula:

$$D(\bar{A}) = D + \frac{1}{2} \frac{\alpha D^R}{\alpha A} \bar{A}. \quad (12)$$

If  $|D(\bar{A})| \geq 0.03$ , we think that this point is the feature point we want; otherwise, we remove it.

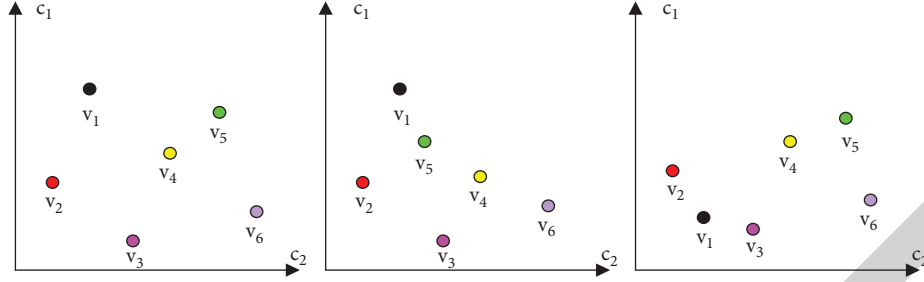
The above steps can only eliminate points with large differences. Among the extracted feature points, in addition to the large differences, there are some unstable points that need to be corrected. Such unstable points are called pseudo-feature points. The gray value of pseudo-feature points mainly changes in the edge direction, while the gray value changes in the direction perpendicular to the edge are small [24]. For this feature point, you can use the Hessian matrix to estimate the derivative method. The Hessian matrix is expressed as

$$J = \begin{pmatrix} D_{aa} & D_{ab} \\ D_{ab} & D_{bb} \end{pmatrix}, \quad (13)$$

where  $D_{aa}$  is the second-order partial derivative of the Gaussian difference pyramid in the  $A$  direction,  $D_{bb}$  is the second-order partial derivative of the Gaussian difference pyramid in the  $B$  direction, and  $D_{ab}$  is the mixed partial derivative of the Gaussian difference pyramid in the  $A$  direction and the  $B$  direction.

Assuming that the eigenvalue of the Hessian matrix has the largest value  $\alpha$ , the smallest value is  $\beta$ , there exists  $\alpha = \gamma * \beta$ , and the ratio can be expressed as

$$\frac{R_n(J)^2}{\det(J)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(n+1)^2}{n}. \quad (14)$$

FIGURE 4: Three different selection results when  $N=6$ .

Among them  $R_n(J) = D_{aa} + D_{bb}$ ,  $\det(J) = D_{aa} \times D_{ab} - D_{ab} \times D_{bb}$ . When  $\gamma$  is equal to 1, that is,  $\alpha = \beta$ , the minimum value can be taken, and the threshold is defined as the formula:

$$\frac{R_n(J)^2}{\det(J)} < \frac{(n+1)^2}{n}, \quad (15)$$

where  $\gamma = 10$ . If the Hessian matrix satisfies this formula, it can be considered that this extreme point is the characteristic point we want; otherwise this point is discarded [25].

The background modeling of the abovementioned SIFT algorithm is accurate, and the specific flowchart of using the SIFT algorithm for moving target detection is shown in Figure 5. In the figure,  $\#min$  represents the minimum threshold,  $P$  is the current number of frames, and  $frames$  is the total number of frames of the video sequence.

### 3. Experimental Results of Sports Target Tracking and Speed Measurement Based on Time Domain Vision Sensors

**3.1. Tracking Algorithm and Comparison Based on Time Domain Vision Sensor.** When studying the SIFT algorithm, Mean Shift algorithm, and MSO algorithm, the same group of moving targets has been tracked. The tracking effect can probably be felt from the figure and the experimental effect. In order to more rigorously explain the tracking effect of each algorithm. Let us make quantitative analysis through multiple sets of experiments, as shown in Figure 6.

MSO is an improved algorithm of Mean Shift, so here the MSO algorithm and Mean Shift algorithm are used to track the moving target during the movement process shown in Figure 6, and the effect diagram is shown in Figure 7. The red is the tracking process of the Mean Shift algorithm, and the blue is the tracking process of the MSO algorithm. From the effect in the figure, it can be seen that the effect of the Mean Shift is not ideal. When the scale of the moving target changes, the tracking window of the Mean Shift algorithm remains unchanged, resulting in a less ideal tracking process, while the tracking effect of the MSO algorithm is better. Because the MSO algorithm uses the ORB algorithm to update the tracking window, and when the Mean Shift algorithm tracking effect is not good, it will use the ORB algorithm to rematch it to locate the moving target, so the effect is better. The ORB algorithm can be used to quickly create feature vectors for key points in an image, which can be used to identify objects in an image.

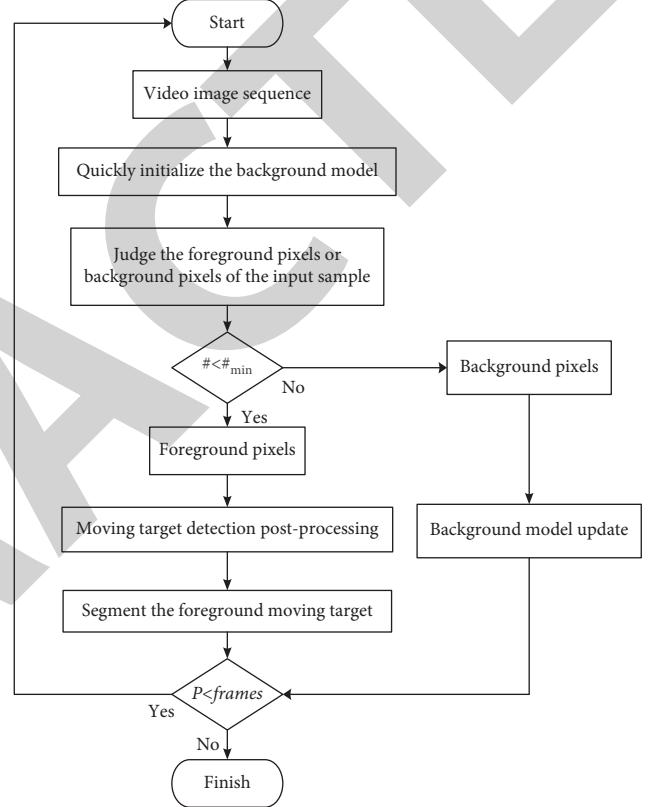


FIGURE 5: SIFT algorithm target detection flowchart.

After conducting multiple sets of experiments, the data of the SIFT algorithm, Mean Shift algorithm, and MSO algorithm when tracking the moving target are, respectively, counted, as shown in Table 1.

It can be seen from the table that SIFT has the highest accuracy rate and the best tracking effect, but it also takes the longest time; although the Mean Shift algorithm takes a short time, its accuracy is also reduced, and the effect is not ideal. The accuracy of MSO is higher than that of Mean Shift algorithm, and it takes less time than Mean Shift algorithm, which can better balance real time and accuracy. Its accuracy rankings from highest to lowest are SIFT, MSO, and Mean Shift, and its time-consuming rankings from highest to lowest are SIFT, Mean Shift, and MSO. It can be seen that the MSO algorithm is better than the Mean Shift algorithm in terms of time consumption and accuracy.

Combining the experimental results and the comparison results of the tracking performance of the three algorithms in



FIGURE 6: Movement process of moving target.

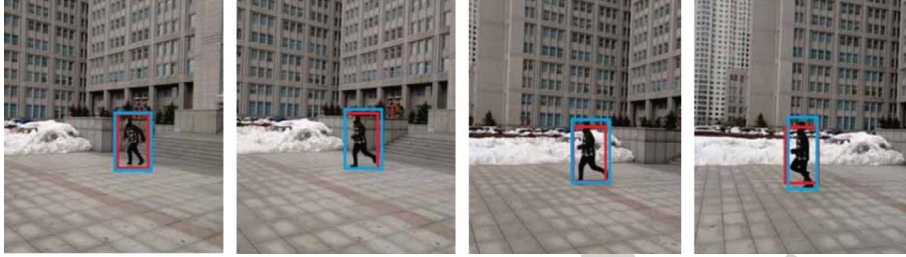


FIGURE 7: Tracking effect diagram.

TABLE 1: The comparison image of algorithm.

Group	Algorithm	Picture size	Feature points	Number of matches	Algorithm time-consuming
First group	SIFT	340*340	458012	261	1232
	Mean Shift		369865	179	361
	MSO		417965	198	301
Second group	SIFT	320*240	201356	169	1271
	Mean Shift		198638	119	361
	MSO		158634	151	271
Third group	SIFT	800*640	298689	909	6098
	Mean Shift		224613	659	1762
	MSO		259687	791	1309
Fourth group	SIFT	180*140	309698	99	841
	Mean Shift		296873	82	253
	MSO		278635	89	181

TABLE 2: Comparison of tracking performance of each algorithm.

	Average number of iterations	Single frame running time (s/frame)	Average running time (s)	Tracking results
Mean Shift algorithm	6.0231	0.1269	0.2198	Fail
MSO algorithm	5.3685	0.3589	0.4391	Success
SIFT algorithm	4.5826	0.4092	0.5593	Success

Table 2, the tracking results of the SIFT algorithm are accurate, which reduces the tracking error while also reducing the number of iterations and also meets the requirements of the algorithm's real-time performance.

Comprehensive analysis shows that, compared with the other two algorithms, the improved SIFT algorithm based on feature fusion, when the improved SIFT algorithm uses feature adaptive fusion to describe the target, it can automatically allocate the proportion of color features and texture features, which improves the accuracy and reliability of moving target tracking. Even in video scenes with similar background colors and light changes, the improved SIFT tracking algorithm can still maintain an ideal tracking effect.

*3.2. Detection and Tracking of Image Moving Targets Based on Time Domain Sensors.* At the same time, this article compares experiments on the TUD-Crossing test set, the ETHBAHNHOF test set, and the SMOT data set. The TUD-Crossing data set is a video sequence in the TUD pedestrian database used to evaluate the performance of the tracker. ETHBAHNHOF is a video sequence used to evaluate the performance of multitarget tracking. SMOT is a test video sequence containing a variety of scenes, including slalom, acrobatics, crowds, and other scenes.

Table 3 compares the performance of the comparison method and the method in this paper on the TUD-Crossing, ETHBAHNHOF, and SMOT data sets. The comparison index is the number of target ID exchanges. The data in the table is



TABLE 3: Comparison of the number of exchanges of tracking target IDs (unit: times).

Algorithm name data set	TUD-crossing	ETH	SMOT(campus)
DP	36	39	41
TSML	12	13	13
IHTLS	9	6	6
ANF	6	4	5

the number of target ID exchanges. The total number of targets detected in TUD-Crossing is 846, the total number of targets detected in the ETH data set is 1768, and the total number of targets detected in the campus scene in the SMOT data set is 922. It can be seen from the table that the number of ID exchanges using the DP algorithm is the highest. This is because this method does not use metric learning to distinguish the targets, and the tracked fragments obtained are not necessarily reliable. Moreover, simple appearance features are used to estimate the transition probability when tracking segment associations, so it is easy to exchange target IDs. Although the TSML algorithm uses the target appearance feature metric learning method to refine the tracking segments, it still uses the appearance features for association when the tracking segments are associated, so it is easy to exchange the target ID when the target appearance is very similar. The target ID exchange frequency of the IHTLS algorithm is relatively low; however, by calculating the distance between target detections and combining target detections to form a tracking segment, it directly performs tracking segment association without refining the tracking segment. This tracking segment is not necessarily reliable and easily affects the subsequent tracking segment association. The improved target tracking method in this paper makes full use of the characteristic information of the target to associate the target when the target detection is used to form the tracking segment. In the tracking segment association stage, the target's motion information is used to achieve continuous tracking of the target when the target appearance is similar and to reduce the number of target ID exchanges.

In this paper, the performance index comparison of five scenarios is obtained by testing on the SMOT data set. The balls scene refers to nearly 50 randomly rolling table tennis balls. The scene of this crowd is a densely populated surveillance image, and the detection target is a human head. Due to the large number of people, there are frequent occlusions in the scene. Acrobat is a scene of acrobatic performances by actors wearing the same costumes. Slalom is three skiers skiing on the slalom arena. The target in the scene frequently disappears from the screen for a long time and then reappears. In the juggling scene, the performer uses 3 balls to perform. The ball moves quickly and the size and color of the ball are the same. The test result is obtained by running the corresponding algorithm ten times in each scenario and averaging.

Figure 8 shows the MOTA comparison of ANF algorithm, TSML algorithm, and IHTLS. It can be seen from the figure that the tracking accuracy of the IHTLS algorithm and the improved tracking algorithm in this article is relatively high in these five scenarios. Based on the feature metric

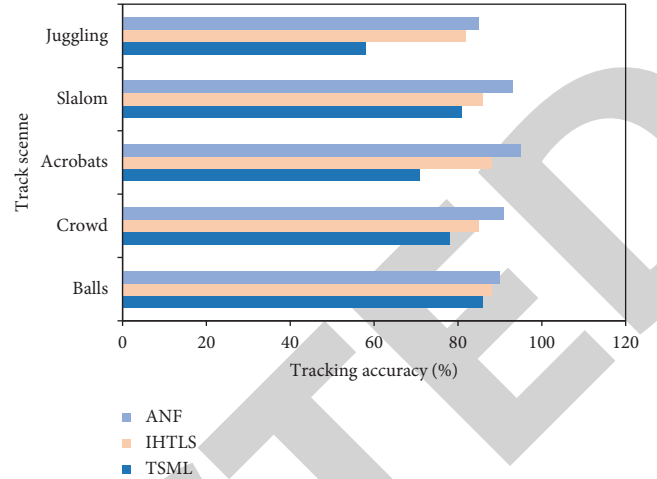


FIGURE 8: Multitarget tracking accuracy.

learning method, in the five scenes, especially the juggling scene, because the appearance of the ball is the same and the target move quickly, it cannot be effectively tracked by the appearance of the target. The Acrobat scene has the highest accuracy among the five scenes, because there is less occlusion between targets in this scene.

Figure 9 shows the comparison of the missing rate and the misjudgment rate of the three tracking algorithms. The missing rate counts the number of missing targets. It can be concluded that the missing rate of the three methods is high when tracking the juggling scene. This is because the three factors of the ball's movement, appearance, and occlusion of the performer increase the difficulty of tracking, resulting in an increase in missing targets, so the missing rate in this scene is relatively high. The false positive rate counts the number of unmatched target hypothetical positions. It can be seen from the figure that the misjudgment rate based on the kinetic model is also relatively high. This is because the IHTLS algorithm based on the kinetic model has relatively high requirements for the detection results. If the detection results contain a large number of false detections, the tracking effect will be greatly affected. The improved method in this paper can eliminate false detections through metric learning, so the false positive rate is reduced. The highest rate of misjudgment in the balls scene is because the athletes are blocked for a long time.

Figure 10 shows the average tracking time of TSML, IHTLS, and the improved algorithm in this paper. The average tracking time is obtained by averaging ten times of tracking the SMOT data set. Although the time performance of this paper is lower than that of the other two methods, this is due to the increased time of feature metric learning. But the time consumed has little effect on the overall performance of target tracking.

3.3. *Image Matching and Purification by Time Domain Vision Sensor.* In order to study the effect of each purification constraint acting alone, rough matching can be performed first, and then each purification algorithm is used to process the rough matching. The experiment is to explore the same

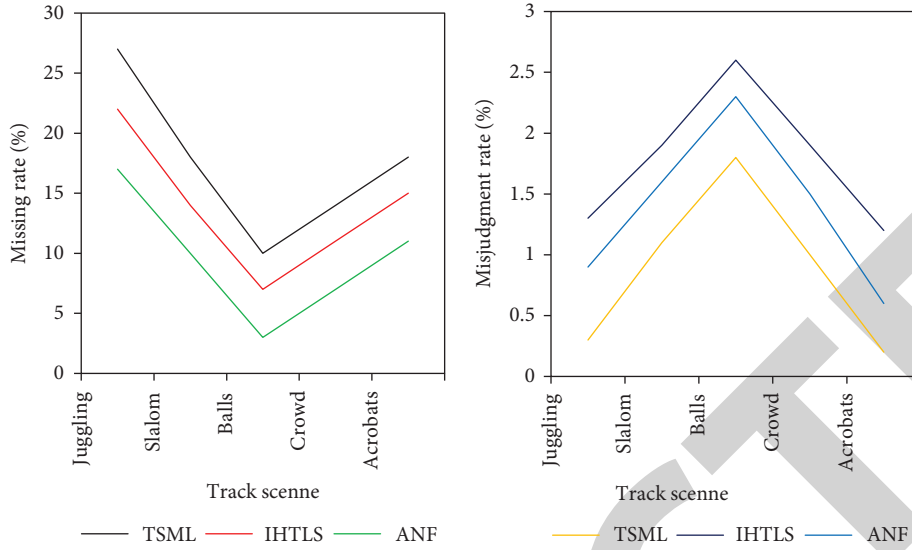


FIGURE 9: Algorithm missing rate and false positive rate.

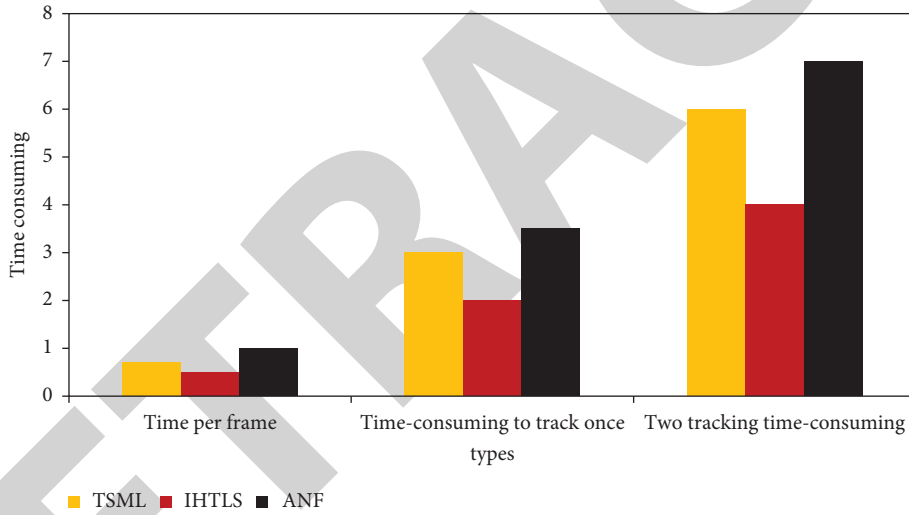


FIGURE 10: Tracking time performance comparison.

TABLE 4: Accuracy of the multipurification algorithm.

	Group 1		Group 2		Group 3	
	Matching logarithm	Accuracy/%	Matching logarithm	Accuracy/%	Matching logarithm	Accuracy/%
Coarse matching	396	46.2	401	40.2	396	50.2
After purification by neighboring ratio	286	62.6	261	62.1	269	69.6
After two-way matching and purification	198	89.2	181	90.9	209	89.7
After purification by angle cosine	202	98.9	156	98.9	195	98.6
Actual exact match	203	—	162	—	203	—

object under the time domain vision sensor. To test the comprehensive effect of the multiple purification algorithm, the following is after rough matching is performed using the nearest neighbor search algorithm; after the nearest neighbor search algorithm is used for rough matching, the nearest neighbor ratio, two-way matching, and angle cosine

purification algorithms are used to eliminate mismatches. The accuracy of the three sets of experiments is shown in Table 4. It can be seen from the table that when the accuracy of rough matching is less than 50%, the final matching accuracy after processing with multiple purification algorithms is above 98%.

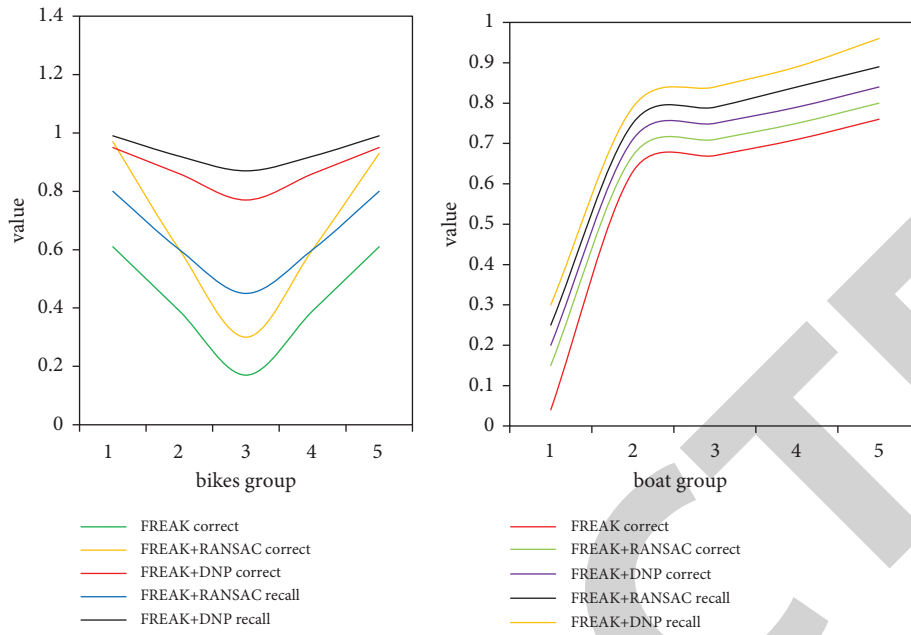


FIGURE 11: Comparison of the purification results of matched pairs between the bikes group and the boat group.

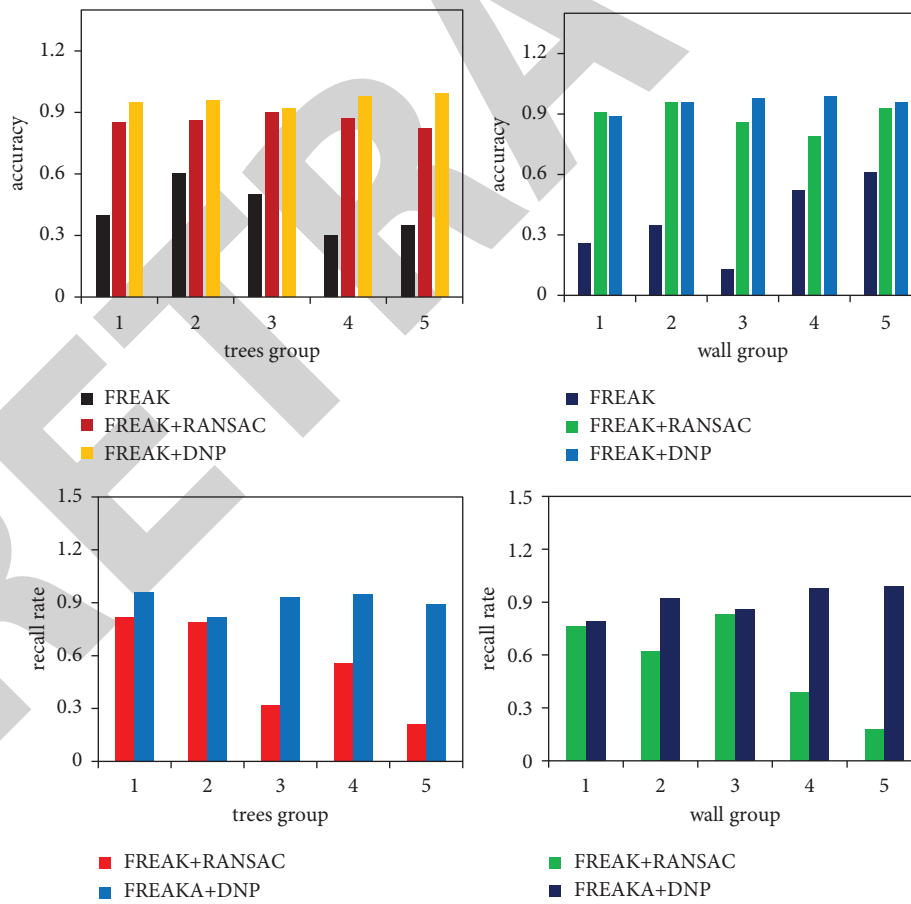


FIGURE 12: Comparison of the purification results of matching pairs between the tree group and the wall group.

Secondly, the FREAK algorithm has good applicability to scale changes, but no new feature point extraction algorithm is proposed. Therefore, this experiment first uses the STAR algorithm to extract feature points of the image, then uses the FREAK algorithm to describe its features, and finally uses the DNP algorithm and the RANSAC algorithm to achieve purification. Select the purification results of four groups of test images, as shown in Figures 11 and 12.

It can be seen from the figure that, for each group of images, the coarse matching accuracy of the original FREAK algorithm decreases as the degree of change increases. The purification method based on RANSAC can obtain better accuracy in some weaker changes, but its accuracy rate fluctuates greatly, and the recall rate is very low. At the same time, the DNP algorithm can control the final accuracy rate above 80%, and its recall rate is basically greater than 80%. Taking the boat group as an example, in the first two experiments, both the RANSAC algorithm and the DNP algorithm can achieve better purification accuracy. However, the recall rate of the RANSAC algorithm has dropped significantly. In the following three experiments, the accuracy of the RANSAC algorithm has been close to zero. At this time, the DNP algorithm can still maintain a fairly high accuracy and recall rate.

#### 4. Discussion

Image sequence moving target tracking technology is widely used in fields such as national defense, military, industrial control, community security, and intelligent transportation. Moving target tracking is to find the moving target of interest in the image sequence and locate each target of interest in the image sequence. By analyzing the sequence of images, the motion parameters, displacement, and velocity of moving objects in consecutive frames can be calculated, which is the purpose of moving object tracking, and can also provide information for image understanding and analysis. The moving target in the image sequence provides more information than the target in a single static image. The purpose of target tracking is to find the accurate position of the target in the next frame of image. The general tracking method is to extract the image of the tracked target, build a template, and then match the entire image in the next frame to search for the target.

Of course, the detection and tracking of moving targets has always been the frontier in the field of time domain vision sensors. It has a wide range of applications in military, security monitoring, traffic control, product testing, medical image analysis, and other fields. It has important application value and broad development prospects. Moving target detection is an important part of video surveillance. It can separate the moving targets in the video stream so that people can analyze the moving targets. When detecting moving objects, it will be affected by the external environment such as light and climate change. Moving target detection and tracking is also an important topic in the field of time domain vision sensors. After long-term development, many results have been obtained. There are many mature algorithms for detecting and tracking moving

targets, but various algorithms are limited to specific application scenarios. This paper studies the previous target detection algorithms and tracking algorithms, applies the multihypothesis tracking algorithm to the tracking of moving targets in the video stream, and has achieved good tracking results.

#### 5. Conclusions

In the past sports training, the coaches mainly rely on the intuition and naked eye observation to detect the moving target, which greatly increases the unfairness of sports. To change this situation, time domain vision sensor technology has also been cited in sports many times. It can quickly and effectively capture the moving target and provide more theoretical and data descriptions for the athlete's movements, so as to make a more accurate judgment on the athlete's movement state. We have made additional clarifications in the conclusion section of the article. At the same time, with the development of science and technology, the advancement of society, and the improvement of people's living standards, the safety awareness of groups and individuals has continued to increase, and video surveillance has been more and more widely used. At present, the use of video or image to detect moving targets has been applied in many occasions. For example, a surveillance system is used to detect moving objects in a video to analyze the behavior and characteristics of the moving objects in the image. The use of monitoring methods to detect and automatically track moving targets will greatly reduce the waste of manpower and material resources.

#### Data Availability

No data were used to support this study.

#### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

#### Acknowledgments

This work was supported by the Social Science Fund Project of Shaanxi Province: 1. Research on College Students' Attitude towards Physical Health Standard Policy of Private Universities in Shaanxi Province; 2. Research on Innovative Mode Construction and Ecological Civilization Construction of Sports Tourism in Shaanxi Province under the Strategy of "the Belt and Road."

#### References

- [1] D. Feng and M. Q. Feng, "Vision based multipoint displacement measurement for structural health monitoring," *Structural Control and Health Monitoring*, vol. 23, no. 5, pp. 876–890, 2016.
- [2] J. A. Lenero-Bardallo, P. Hafliker, R. Carmona-Galan, and A. Rodriguez-Vazquez, "A bio-inspired vision sensor with dual operation and readout modes," *IEEE Sensors Journal*, vol. 16, no. 2, pp. 317–330, 2016.

- [3] Z. Zhang and S. Chen, "Real-time seam penetration identification in arc welding based on fusion of sound, voltage and spectrum signals," *Journal of Intelligent Manufacturing*, vol. 28, no. 1, pp. 207–218, 2017.
- [4] A. A. Prasov and H. K. Khalil, "Tracking performance of a high gain observer in the presence of measurement noise," *International Journal of Adaptive Control and Signal Processing*, vol. 30, no. 8–10, pp. 1228–1243, 2016.
- [5] B. Lyu and L. V. Smith, "Evaluation of wireless bat swing speed sensors," *Sports Engineering*, vol. 21, no. 2, pp. 1–6, 2017.
- [6] J. Lauzon-Gauthier, C. Duchesne, and J. Tessier, "A machine vision sensor for quality control of green anode paste material," *Journal of Occupational Medicine*, vol. 72, no. 1, pp. 287–295, 2020.
- [7] F. Zhou, C. Xin, and H. Tan, "Three-dimensional catadioptric vision sensor using omnidirectional dot matrix projection," *Chinese Optics Letters*, vol. 14, no. 11, pp. 80–84, 2016.
- [8] S. Yang, H. U. Yan-Xiang, and X. U. Jiang-Tao, "Modeling and emulation of address-event representation vision sensor bus arbitration," *Journal of Tianjin Polytechnic University*, vol. 35, no. 6, pp. 84–88, 2016.
- [9] S. Engelberg and B. Milgrom, "Tracking using state estimation: a brief introduction," *IEEE Instrumentation and Measurement Magazine*, vol. 22, no. 3, pp. 36–42, 2019.
- [10] C. Hu, Y. Deng, W. Tian, J. Wang, and T. Zeng, "Novel MIMO-SAR system applied for high-speed and high-accuracy deformation measurement," *Journal of Engineering*, vol. 2019, no. 20, pp. 6598–6602, 2019.
- [11] B. Frczek, A. Grzelak, and A. T. Klimek, "Analysis of daily energy expenditure of elite athletes in relation to their sport, the measurement method and energy requirement norms," *Journal of Human Kinetics*, vol. 70, no. 1, pp. 81–92, 2019.
- [12] N. Kwak, G. Lee, J. Kwon, and Y. Eunjung, "Robust measurement validation for radar target tracking using prior information," *IET Radar Sonar & Navigation*, vol. 13, no. 10, pp. 1842–1849, 2019.
- [13] K. Ahmadi and E. Salari, "Social-spider optimised particle filtering for tracking of targets with discontinuous measurement data," *IET Computer Vision*, vol. 11, no. 3, pp. 246–254, 2017.
- [14] A. Wilson, "High-speed cameras enhance automation and measurement applications," *Vision Systems Design*, vol. 24, no. 1, pp. 17–20, 2019.
- [15] D. M. Brown, D. B. Dwyer, S. J. Robertson, and P. B. Gastin, "Metabolic power method: underestimation of energy expenditure in field-sport movements using a global positioning system tracking system," *International Journal of Sports Physiology and Performance*, vol. 11, no. 8, pp. 1067–1073, 2016.
- [16] M. Kobayashi and T. Yamazaki, "A 1,000 [fps]-High-Speed 3D-stacked vision sensor to realize visual feedback control for industrial machinery/robots," *Journal of the Robotics Society of Japan*, vol. 35, no. 8, pp. 579–582, 2017.
- [17] Y. Tang, T. Wu, G. Yuan, and L. Shaohui, "3D Omnidirectional vision sensor for morphology defects detection in pipelines," *Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument*, vol. 38, no. 3, pp. 726–733, 2017.
- [18] D. Henke, E. M. Dominguez, D. Small, and E. S. Michael, "Moving target tracking in SAR data using combined exo- and endo-clutter processing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 251–263, 2017.
- [19] A. Obara, X. Yang, and H. Oku, "Structured light field generated by two projectors for high-speed three dimensional measurement," *Journal of Robotics and Mechatronics*, vol. 28, no. 4, pp. 523–532, 2016.
- [20] I. Castell-Uroz, J. Sole-Pareta, and P. Barlet-Ros, "Network measurements for web tracking analysis and detection: a tutorial," *IEEE Instrumentation and Measurement Magazine*, vol. 23, no. 9, pp. 50–57, 2020.
- [21] M. Beato, G. Devereux, and A. Stiff, "Validity and reliability of global positioning system units (STATSports viper) for measuring distance and peak speed in sports," *The Journal of Strength & Conditioning Research*, vol. 32, no. 10, pp. 2831–2837, 2018.
- [22] T. Hortobágyi, M. Lesinski, M. Gäbler, J. M. VanSwearingen, D. Malatesta, and U. Granacher, "Erratum to: effects of three types of exercise interventions on healthy old adults' gait speed: a systematic review and meta-analysis," *Sports Medicine*, vol. 46, no. 3, p. 453, 2016.
- [23] J. M. Rhodes, B. S. Mason, T. A. Paulson, and V. L. Goosey-Tolfrey, "Altering the speed profiles of wheelchair rugby players with game-simulation drill design," *International Journal of Sports Physiology and Performance*, vol. 13, no. 1, pp. 37–43, 2018.
- [24] Y. Hamada and N. Hayashi, "Effects of eating speed on diet-induced thermogenesis," *Tairyoku Kagaku Japanese Journal of Physical Fitness & Sports Medicine*, vol. 65, no. 3, pp. 287–295, 2016.
- [25] G. Fiorilli, E. Iuliano, M. Mitrotasios et al., "Are change of direction speed and reactive agility useful for determining the optimal field position for young soccer players?" *Journal of Sports Science and Medicine*, vol. 16, no. 2, pp. 247–253, 2017.