Research Article

Research on Video Target Detection and Tracking in Football Matches

Xiaoju Zou,1 Yuanling Huang,1 Nairun Zhou,1 and Zuoming Fang2

1Department of Physical Education, Guangzhou City University of Technology, Guangzhou 510800, Guangdong, China
2College of Physical Education, Guangdong University of Education, Guangzhou 510800, Guangdong, China

Correspondence should be addressed to Zuoming Fang; fangzuoming@gdei.edu.cn

Received 11 May 2022; Revised 19 June 2022; Accepted 6 July 2022; Published 31 July 2022

Copyright © 2022 Xiaoju Zou et al. 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.

Computer vision is an interesting branch of artificial intelligence which is dedicated to how electronic devices can achieve the level of capabilities to perceive things just like ordinary human beings do. Usually for implementation of this, electronic devices must be equipped with additional sensors which has the capability to enable these devices to view things just like an eye of the human being does. In addition, it is an interested research domain, because if these devices, both computers and robots, can achieve this capability, then it would be easier to perform various tasks which are reserved for human beings only. In addition, as per our understanding, AI based devices and systems are utilized to automate various activities, however, this is subjected to the feasibility of these activities in the problem domain. Moreover, in certain scenarios, these systems, i.e., AI based smart systems, are utilized for the regular assistance of the field experts in various domains of life such as in medical sector, a decision support system could help a doctor in the diagnosis process of the disease. For this purpose, computer vision uses cameras and computers to acquire and process visual information. The purpose of computer is to achieve people’s understanding of vision. Computer vision is the intersection and combination of disciplines, including mathematics and image processing, as well as computer science. Video for the detection of target and tracking is a hot topic in the field of computer vision and has received much attention at home and abroad. Video for the detection of target and tracking is a process of inferring the position of moving target in video by computer. The main task is to accurately locate the target in each frame of the video, and then generate the motion track of the target, giving the complete area of the moving target in video at any time.

1. Introduction

Computer vision is an interesting branch of the artificial intelligence which is dedicated to how electronic devices can achieve the level of capabilities to perceive things just like ordinary human beings do. Usually for implementation of this, electronic devices must be equipped with additional sensors which has the capability to enable these devices to view things just like an eye of the human being does. In addition, it is an interested research domain, because if these devices, both computers and robots, can achieve this capability, then it would be easier to perform various tasks which are reserved for human beings only. In addition, as per our understanding, AI based devices and systems are utilized to automate various activities, however, this is subjected to the feasibility of these activities in the problem domain. Moreover, in certain scenarios, these systems, i.e., AI based smart systems, are utilized for the regular assistance of the field experts in various domains of life such as in medical sector, a decision support system could help a doctor in the diagnosis process of the disease. For this purpose, computer vision uses cameras and computers to acquire and process visual information. The purpose of computer is to achieve people’s understanding of vision. Computer vision is the intersection and combination of disciplines, including mathematics and image processing, as well as computer science. Video for the detection of target and tracking is a hot topic in the field of computer vision and has received much attention at home and abroad. Video for the detection of target and tracking is a process of inferring the position of moving target in video by computer. The main task is to accurately locate the target in each frame of the video, and then generate the motion track of the target, giving the complete area of the moving target in video at any time.
Sports videos are a highly frequent form of video in entertainment videos in everyday life, occupying a substantial share of daily entertainment life, and football-related videos in sports videos are extensively watched by sports video viewers. Football coaches may use the video to study the detailed data of the necessary movement, and players can use the video to record the associated data of their own movement, which can be retrieved from the target movement’s track in the football match film. These requirements can be satisfied and met with the use of high-cost labour and material resources for football matches. As a result, this study investigates football video target recognition and tracking systems in depth.

The following are the paper’s innovations: (1) detecting the video target of a competition using a background model and visual detection technology, then integrating for the detection of target findings with the basic meanshift video target tracking algorithm to track the identified target. (2) When compared to traditional teaching methods, the Intelligent Teaching System for Chinese as a foreign language proposed in this paper has a high level of overall stability and can significantly improve students’ learning efficiency and teaching quality.

The remainder of the paper is divided into the sections listed below. A detailed review of the existing methodologies is reported in the section two of the paper where feasibility of every scheme along with expected issues are described. In section three, computer vision-based football video for the detection of target method based on the background model is reported where how the vision capability of the devices can be used. Meanshift-enabled target tracking algorithm is described in detail and how it is effective in resolving the scenarios. Results of the experimental observations were described in section five of the paper which is followed by the summary of the whole manuscript.

2. Related Work

In football video, the detection of target and tracking is of great importance to the analysis of football video, so it has received wide attention and obtained corresponding research results. Gao [1], aiming at the problems for the detection of target in football video, adopts the method of football video for the detection of target and tracking based on improved Gauss model. First, it deeply analyses the drawbacks of the model itself, converts color image to gray image, and dynamically expands in areas with small differences in video similarity, so as to improve the efficiency and accuracy for the detection of target and tracking in football video with Gauss model. The experimental results show that this method has a strong ability to deal with video interference in football matches and a large detection range, but it has the problem of low accuracy of detection [1]. Tian [2] puts forward a robust algorithm for football video for the detection of target and tracking in order to achieve the stable detection and tracking of football video targets. The image of the match video is matched and calibrated, and the moving target in the video is detected by the image cumulative difference map. To achieve efficiency for the detection of target, the matching complexity of video feature points is reduced through the corresponding strategies of network sampling, thus solving the problem of nonuniform concentration of image feature points in video. Intensity filter algorithm is used to improve the significance of the target in the match video, to filter out the unwanted false targets, to propose a scale calculation and area detection method based on the detected targets, and then to use Kalman filter to track the detected targets. The experimental results show that this method can detect the moving target accurately in football video, but it does not track the target steadily [2]. Zhou et al. [3] proposed a color space-based detection and tracking method which is robust to light due to the problem of target loss during the staggered motion of football video targets. Edge detection is performed on the relatively accurate moving target’s binary image after converting the moving target’s binary image extracted from the color space, using brightness detection to judge whether the filtered binary image is morphologically operated, and then converting the moving target’s binary image extracted from the color space. After obtaining the minimal enclosing rectangle frame, the associated motion tracks of different color targets are produced, and the multi-target tracking of the football match video is accomplished. It can address the problem of video tracking target staggered motion loss target better than previous methods, but it has poorer detection accuracy [3]. Zou et al. [4] is widely used in many fields such as video surveillance and artificial intelligence for target tracking. In order to detect and track the moving target of football video quickly, a for the detection of target and tracking method combining Kalman filter and particle filter is proposed. Under the framework of Kalman filter, combined with particle filter to predict the location of moving objects, the method proposed in this paper is compared with a single Kalman filter method. The experimental results show that the proposed method has higher accuracy and better real-time performance, but due to the complexity of the process, the tracking effect is poor [4].

3. Football Video for the Detection of Target Method Based on Background Model

Computer vision is among the potential candidate solutions for the problems in hand. It is used to enable a system to perceive various things in a professional manner which preserve expected levels of the precision and accuracy.

3.1. Background Extraction of Optical Flow Field. Figure 1 shows the video for the detection of target process for a football match. Figure 1 shows that the image optical flow field in football match video mainly refers to the change of image brightness mode position in adjacent frames. The function of video target time is defined as \((x(t), y(t))\). The corresponding gray level of image point in video is \(E(x(t), y(t), t)\). Based on the video brightness and smoothing constraints, the formula (1) and formula (2) error functions are expressed as follows:
Soccer match video target detection

Extract background image model

Video background update policy

Get video differential image

Ghost target determination

Classification of image background high-frequency disturbance degree

Pixel count statistics

The end of the

\[ D_t(i) = \left\{ (x, y) \mid \sqrt{(x-x_t(i))^2 + (y-y_t(i))^2} \leq r \right\}. \] (3)

In formula (3), both \( x \) and \( y \) represent the target mark points in the football game video. The reasonable selection of radius \( r \) can ensure the coverage of all target areas and effectively reduce the background points of the target area in the video. The estimated \( D_t \) of the football game video target area in frame \( t \) in the game video is expressed by the following formula:

\[ D_t = \bigcup_{i=1}^{M_t} D_t(i). \] (4)

In formula (4), \( M_t \) is the number of mark points in the \( t \) frame of the competition video. After obtaining the area \( D_t \) of the moving target in each frame of the video, the background corresponding to the point in the video is estimated according to the gray value of the points outside the target area. Assuming that \( F_t(x, y) \) is a function for determining whether each point in frame \( t \) in the game video is the target area, according to the solution of obtaining the optical flow field of frame \( N \), the fixed background area \( USB \) of the target area in the video image is eliminated. When the area of the image background in the football game video is extracted, the background model of the football game video target image is calculated by formula (5), expressed as

\[ CB(x, y) = \frac{\sum_{t=1}^{N} f_t(x, y) \left(1 - F_t(x, y)\right)}{\sum_{t=1}^{N} \left(1 - F_t(x, y)\right)}. \] (5)

In formula (5), \( N \) is the number of frames of the input football game video image, and \( f_t(x, y) \) is the gray value of the \( t \) frame image in the video at the \( (x, y) \) coordinate. After obtaining the fixed background area and background model of the video image, a frame image \( I \), if it belongs to the pixel of the image background area, the difference between the background model of the football game video image and the image \( I \) is calculated and expressed as

\[ I - CB = \bigcup_{(x, y) \in USB} |I(x, y) - CB(x, y)|. \] (6)

In formula (6), \( I(x, y) \) is all pixels in image \( I \). If the difference of the sum is greater than the set threshold, it indicates that the background of the game video changes greatly and the convective light field needs to be extracted again [7, 8]. The model of the video image color difference is expressed as

\[ I_D(x, y) = \max \left| R_b(x, y) - R_k(x, y) \right|, \left| G_b(x, y) - G_k(x, y) \right|, \left| B_b(x, y) - B_k(x, y) \right|. \] (7)

In formula (7), \( R_b(x, y) \) is the color background image’s value of pixel \( R \) in the football game video, \( G_b(x, y) \) is the color background image’s value of pixel \( G \) in the football game video, \( B_b(x, y) \) is the color background image’s value of pixel \( B \) in the football game video, \( R_k(x, y) \) and \( G_k(x, y) \) are the three primary color values of the current frame image of the football game video, and the difference image formation can improve the integrity for the detection.
of target area of the football game video by using color information [9, 10].

3.2. Ghost Elimination of Video Target in Football Match. In the process of detecting the moving target of the game video, the pixel \( x \) of any image has a corresponding counter \( f_{\text{count}}(x) \). If the moving target image pixel waiting to be detected is a foreground pixel, it is necessary to judge whether the image pixel in the background model is restrictive. If it has significant characteristics, the pixel \( f_{\text{count}}(x) \) of the video frame waiting to be detected needs to be increased by 1, which is expressed as

\[
f_{\text{count}}(x) = \begin{cases} f_{\text{count}}(x), & f(x) = 1 \cap S(x) = 1, \\ 0, & \text{else}. \end{cases}
\] (8)

In formula (8), \( f(x) \) is the function of judging whether the image pixel \( x \) is a foreground image. \( S(x) \) represents the detection result of the significance of the background model saved in the detection target image relative to the pixel \( x \), and the fuzzy set theory is introduced when the background of the football game video image is updated. \( Z \) represents the element set of the game video image, \( z \) is a certain kind of element \( Z = \{z\} \), and the fuzzy set \( A \) in \( Z \) is represented by the membership function \( \mu_A(z) \). The fuzzy set of football game video is an ordered pair composed of \( z \) value and corresponding membership function.

\[
A = \{z, \mu_A(z) \mid z \in Z\}. \tag{9}
\]

In formula (9), \( \mu_A(z) \) describes the ghost degree of image pixels in the background model, and \( z \) describes the image pixels to be detected in the video image. When the pixel \( f_{\text{num}}(x) \) of the video image is determined to be a ghost pixel, the corresponding image pixel needs to add \( N_0 \) expansion operation. When the video pixel is determined not to be a ghost pixel during detection, \( f_{\text{num}}(x) \) needs to subtract \( N_1 \) expansion operation, which is expressed as

\[
f_{\text{num}}(x) = \begin{cases} f_{\text{num}}(x) + N_0, & g(x) = 1, \\ f_{\text{num}}(x) - N_1, & g(x) = 0. \end{cases} \tag{10}
\]

In formula (10), \( g(x) \) represents whether the football game video pixel \( x \) is a function of ghost pixel determination, \( N_0 \) and \( N_1 \) represent fixed constants, and the range of \( f_{\text{num}}(x) \) value is set in the interval \([0, \varphi]\), \( \varphi \) represents the factor of time sampling. For this reason, the factor of secondary time sampling is expressed as

\[
\phi = \varphi^* F(x). \tag{11}
\]

In formula (11), \( F(x) \) is a fuzzy function. By constructing the flicker matrix pixel of the target pixel as the judgment principle, the degree of high-frequency disturbance of the video background is graded, and the matching threshold is set to \( R' \), which is expressed as

\[
R' = \begin{cases} R(1 + \lambda_{\text{inc}}), & m \geq R_b, \\ R(1 - \lambda_{\text{dec}}), & m < R_b. \end{cases} \tag{12}
\]

In formula (12), \( m \) represents the data of the flicker matrix pixel of the target pixel of the football game video, and \( R_b \) represents the adaptive update boundary threshold of \( R' \), and \( \lambda_{\text{inc}} \) and \( \lambda_{\text{dec}} \) represents the given parameters [11, 12].

4. Soccer Game Video Target Tracking Algorithm Based on Meanshift

Meanshift, frequently known as a mode-seeking methodology, is a nonparametric feature-space technique, i.e., it is used for the analysis of mathematics, for defining the maxima of a density function. Cluster analysis in computer vision and image processing are examples of application fields. Combined with the above football game video for the detection of target results based on the background model, set the space \( R^d \), \( d \) represents the dimension of the game video space, select \( n \) video sampling points \( x_i, i = 1, 2, \ldots, n \) in this space, and the basic form of meanshift vector at the reference video point \( x \) is expressed as

\[
M_h(x) = \frac{1}{k} \sum_{x_j \in x} (x_i - x). \tag{13}
\]

In formula (13), \( x_i - x \) represents the offset of the game video target sample point \( x_i \) from the target reference point \( x \), and \( M_h(x) \) is the basic vector of meanshift algorithm [13, 14]. The sample point \( x_i \) in the target area of the game video is obtained in the density function \( f(x) \). When the density gradient value is greater than 0 or less than 0, the direction of the function gradient and the direction of density promotion are the same, and the direction of the sample sampled by the target in the \( s_j \) area is consistent with the direction of the density gradient. Therefore, the direction of the vector \( M_h(x) \) of the meanshift algorithm is the same as the direction of the density gradient [15, 16].

Figure 2 shows the schematic diagram of mean value and Figure 3 shows the schematic diagram of mean value offset. In Figures 2 and 3, the square area is \( s_j \), and the direction pointed by the arrow is the vector to determine the offset of the target reference point \( x \). The hollow circle represents \( n \).
target sampling points in the football video, which is satisfied with \( x_i \in s_h \), and the solid circle represents the reference point \( x \) of the target of the meanshift algorithm. The result of the basic vector \( M_h(x) \) is offset to the maximum direction of the probability density of the sampling point of the moving target.

It can be seen from formula (13) that no matter how far the given target reference point \( x \) is from the target sampling point \( x_i \) in the football game video image area \( s_h(x) \), the calculation result of the vector \( M_h(x) \) of meanshift has the same impact. Therefore, the convergence speed of the video target reference point will become slow [17, 18]. In most cases, the closer the video target sampling point is to the target reference point, the greater the impact on the vector result will be. Therefore, different values will be given to the target sampling points with different distances from the moving target reference point. Formula (14) is the expansion formula of meanshift algorithm vector \( M_h(x) \), which is expressed as

\[
M_h(x) = \sum_{i=1}^{n} G(x_i - x/h)w(x_i)(x_i - x) / \sum_{i=1}^{n} G(x_i - x/h)w(x_i).
\]  

In formula (14), \( G \) represents the extended coefficient and \( w \) represents any constant.

Set the point set \( \{x_i\}, i = 1, 2, \ldots, n \) of the football game video image, which represents the points in the area where the football game video target is located, and the regional center point is \( x_h \). Take the regional range of the moving target as the basis to normalize the target area. Therefore, for the target template referenced by the center point \( x_h \) of the video image area, the probability distribution of \( x_i \) color at other positions can be assumed to be \( \tilde{q}_u = \{\tilde{q}_u(y)\}, u = 1, 2, \ldots, m \), and the distribution probability of video target color probability is expressed as

\[
\tilde{q}_u = C \sum_{i=1}^{n} k(\|x_i - x_{h}/h\|^2) \delta[b(x_i) - u].
\]  

In formula (15), \( \delta \) for the Crohneck function, \( k(x) \) is called the outline function of the kernel function, which assigns the weights of pixel points to different moving objects. A higher weight is given to moving target pixels that are close to the end point, and the kernel function can increase the overall stability of density estimation. The formula for calculating the nuclear density of a moving target is presented below (16):

\[
C = \frac{1}{\sum_{i=1}^{n} k(\|x_i - x_{h}/h\|^2)} \sum_{i=1}^{m} \tilde{q}_u = 1.
\]  

In formula (16), \( h \) represents the range of the target area of the competition video, and the formula obtains the distribution \( \tilde{p}_u(y) = \{\tilde{p}_u(y), u = 1, 2, \ldots, m \} \) of the target color probability of a location \( y \) in the template of the candidate moving target. The distribution calculation formula of the color probability of a location in the candidate target is expressed by the following formula:

\[
\tilde{p}_u = C_h \sum_{i=1}^{n} \left(\frac{\|x_i - y/h\|^2}{h}\right) \delta[b(x_i) - u],
\]  

\[
C_h = \frac{1}{\sum_{i=1}^{n} k(\|x_i - y/h\|^2)}
\]  

It is understood from formula (17) that the sampled point \( x_i \) of a moving target is expressed as a regularized grid, so that there is no relation between the \( C_h \) value and the \( y \) value. After calculating the range value between the kernel density value and the moving target area, the \( C_h \) value can also be determined.

Therefore, it can be understood that the target tracking process in the competition video is the process of comparing the similarity between the candidate templates and the target templates, and finding the best template for similarity is the best target location [19, 20]. The Bhattacharyya coefficient is calculated to obtain the similarity between the moving target template and the candidate template. The distribution of color probability, \( \tilde{p}_u(y) \) and \( \tilde{q}_u \) similarity, is represented by the following formula:

\[
\rho[p(y), q] = \sum_{u=1}^{m} \sqrt{\rho(p(y), \tilde{q}_u)}.
\]  

The initial position of the video frame image in the current football match is represented by \( y_{h} \). Formula (18) is extended by Taylor as the following equation:

\[
\rho[p(y), q] = \frac{1}{2} \sum_{u=1}^{m} \sqrt{\rho(p_u(y_{h}), \tilde{q}_u) \delta(b(y_{h}) - u)}.
\]  

Formula (19) shows that the size of \( y \)-value will be affected by the right side of the formula after the expansion of Taylor. Therefore, using the Taylor expansion formula in meanshift algorithm to expand the iteration of the similarity between the color probability distribution \( \tilde{p}_u(y) \) and \( \tilde{q}_u \) of the video image, the maximum density estimation of the target region probability of the video is obtained, and the maximum probability estimation is used to determine the latest position \( y_{1} \) of the target motion. After several
iterations, \( y_i \) is the best position for the target tracking in the match video, thus completing the football match video for the detection of target and tracking.

5. Experimental Result

Moving target pixels that are close to the end point are given more weight, and the kernel function can improve the overall stability of density estimation. Table 1 shows the formula for estimating the nuclear density of a moving object (16). The motion target image of football match video is shown in Figure 4.

Table 2 shows the time comparison between the methods in [1, 2] and the football video for the detection of target based on the background model is proposed in this paper.

By analyzing Table 2, it can be seen that the time for video for the detection of target in football match is 35 s for the methods mentioned in [1] and 46 s for the methods mentioned in [2]. The time for video for the detection of target in football match based on the background model proposed in this paper is 18 s, which is obviously better than the methods mentioned in [1, 2]. Therefore, the methods mentioned in this paper are more suitable for video for the detection of target in football match. Figure 5 shows the comparison of the performance of the background model based for the detection of target method proposed in this paper with that in references [1, 2].

As shown in Figure 5, when the for the detection of target method proposed in document [1] detects the number of video frames, the overall detection effect of the experiment can reach 76 percent at first, but as the number of experimental frames increases, the detection effect deteriorates. When detecting the video image target, the overall detection effect of the for the detection of target method proposed in the document [1] is always worse, whereas the detection effect of the method mentioned in this paper is stable over 90%. The contour of the detected target is complete and the detection effect is good. The tracking times of the meanshift-based video target tracking algorithm presented in this work and those in [1, 2] are shown in Figure 6.
The target tracking technique presented in [2] grows with the growth in the number of video frames in each frame of a football match, as shown in Figure 6, and the tracking time increases gradually. The target tracking technique presented in [1] is essentially the same as that proposed in this study at the start of the experiment, but as the number of video frames rises, so does the tracking time. However, the time required to track a video target using the meanshift method suggested in this study is much less than the time required to follow a video target using the algorithms provided in [1, 2]. This demonstrates that the approach suggested in this study can enhance target tracking efficiency and is better suited to video target tracking in football matches. Table 3 compares the number of iterations of the target tracking algorithm developed in this work to previous approaches.

Table 3: Comparison of tracking iteration times of different methods.

| The evaluation index                  | In this paper, methods | The traditional method |
|---------------------------------------|------------------------|------------------------|
| Frames/frame                          | 120                    | 120                    |
| Iterations/times                      | 96                     | 210                    |
| Number of iterations per frame/time   | 0.8                    | 1.75                   |

By analyzing Figure 7, it can be seen that the video target tracking algorithm based on meanshift proposed in this paper can control the error of video target tracking by less than 10% as a whole. Although the error of the target tracking algorithm proposed in [1] is smaller than that in [2], the error is higher than that of the method proposed in this paper. As the number of video frames increases, the tracking error of the method proposed in [2] increases, so it is explained. The method mentioned in this paper can accurately track the video target of football match, and the tracking accuracy is high.

6. Conclusion

Due to the recent advancement in the field of sensors and other electronic technology, computer vision has become one of the broader and the most challenging research domain where researchers are eager to enable computerized devices to perceive things like humans. With the gradual improvement of image processing technology, the requirements for multi-for the detection of target are also gradually increasing. Football matches are a sport with a large number of viewers. Therefore, the processing of football match videos should also meet the needs of professionals and audiences. In order to improve the effect of football video detection and tracking accuracy, this paper detects and tracks the moving objects in football video, puts forward the methods for the detection of target and tracking, and verifies the effectiveness of the methods through experiments. The experiments prove that the methods can effectively improve the effect of football video for the detection of target and tracking accuracy, which has a certain practical value.

Data Availability

The datasets used during the present study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Gao, “Detection and tracking of moving object in sports video based on improved Gaussian mixture model,” Modern Electronics Technique, vol. 40, no. 19, pp. 91–93, 2017.
[2] G. Q. Tian, “A novel algorithm for efficient multi-object detection and tracking for infrared dynamic frames,” Infrared Technology, vol. 40, no. 3, pp. 259–263, 2018.
[3] H. P. Zhou, F. Wang, and J. Y. Tian, “Multi-target video tracking algorithm based on HSV color features,” Science Technology and Engineering, vol. 17, no. 20, pp. 184–188, 2017.
[4] H. F. Zou and T. T. Luo, “An improved kalman object real-time detection and tracking algorithm,” Computer Simulation, vol. 39, no. 3, pp. 200–204, 2022.
[5] X. Li, “Design of sports video moving for the detection of target and tracking system,” Electronic Design Engineering, vol. 26, no. 20, pp. 54–58, 2018.
[6] B. Yang, “Grey neural network model for prediction of carbon emissions,” *Computer Measurement & Control*, vol. 25, no. 9, pp. 266–268, 2017.

[7] J. Yang, “Research on for the detection of target and tracking system based on video technology,” *Microcontrollers & Embedded Systems*, vol. 22, no. 2, pp. 46–50, 2022.

[8] S. Wang, J. A. Wang, and D. G. San, “Moving for the detection of target and tracking analysis in video system,” *Guangdong Gongan Keji*, vol. 30, no. 1, pp. 39–42, 2022.

[9] Y. F. Dai, Q. C. Tao, and B. Yang, “Infrared video for the detection of target and tracking based on YOLOV3 and kernel correlation filtering algorithm,” *Computer Applications and Software*, vol. 39, no. 1, pp. 200–205, 2022.

[10] J. G. Li, B. H. Yang, and S. B. Shi, "Multi-object detection locationing and locating algorithm for remote sensing video satellites," *Fire Control and Command Control*, vol. 46, no. 2, pp. 114–119, 2021.

[11] M. Nallasivam and V. Senniappan, “Moving human target detection and tracking in video frames,” *Studies in Informatics and Control*, vol. 30, no. 1, pp. 119–129, 2021.

[12] D. Mulimani and A. Makandar, ”Sports video annotation and multi- target tracking using extended Gaussian mixture model,” *International Journal of Recent Technology and Engineering*, vol. 10, no. 1, pp. 1–6, 2021.

[13] Z. A. Ansari, M. J. Nigam, and A. Kumar, “Accurate tracking of manoeuvring target using scale estimation and detection,” *Defence Science Journal*, vol. 69, no. 5, pp. 495–502, 2019.

[14] P. Martinez and M. Barczyk, “Implementation and optimization of the cascade classifier algorithm for UAV detection and tracking,” *Journal of Unmanned Vehicle Systems*, vol. 7, no. 4, pp. 296–311, 2019.

[15] S. K. Mary, D. Mohanapriya, and K. Mahesh, “Sequential Pixel Text Detector: a cohesive framework for text detection and recognition in video,” *International Journal of Pure and Applied Mathematics*, vol. 118, no. 8, pp. 371–374, 2018.

[16] A. Ekkurthi and G. Sudhavani, ”An advanced H-infinity filtering approach for multiple object tracking in video sequences,” *International Journal of Applied Engineering Research*, vol. 12, no. 24, pp. 15489–15500, 2017.

[17] S. Memon, H. Son, K. H. Memon, and A. Ansari, ”Multi-scan smoothing for tracking manoeuvring target trajectory in heavy cluttered environment,” *IET Radar, Sonar & Navigation*, vol. 11, no. 12, pp. 1815–1821, 2017.

[18] D. Sudha and J. Priyadarshini, ”An intelligent multiple vehicle detection and tracking using modified vibe algorithm and deep learning algorithm,” *Soft Computing*, vol. 24, no. 22, pp. 17417–17429, 2020.

[19] V. Aharonson, V. Y. Coopoo, K. L. Govender, and M. Postema, ”Automatic pupil detection and gaze estimation using the vestibulo-ocular reflex in a low-cost eye-tracking setup,” *SAIEE Africa Research Journal*, vol. 111, no. 3, pp. 120–124, 2020.

[20] K. M. Abughalieh, B. H. Sababha, and N. A. Rawashdeh, ”A video-based object detection and tracking system for weight sensitive UAVs,” *Multimedia Tools and Applications*, vol. 78, no. 7, pp. 9149–9167, 2019.