Journal of Multimedia Processing and Technologies

DLINE JOURNALS

Print ISSN: 0976-4127
Online ISSN: 0976-4135

JMPT 2025: 16 (1)
https://doi.org/10.6025/jmpt/2025/16/1/20-27

Moving Object Detection and Tracking Technology Based on Hybrid Algorithm

Cheng Zhou
School of Physical Education
Hunan University of Arts and Science
415000, Changde, Hunan, China
53145721@qq.com

ABSTRACT

Object tracking is a hot topic in visual technology and is widely applied in scenarios such as intelligent monitoring, autonomous driving, and robot visual perception. In recent years, with the sports industry's rapid development, tracking targets (balls and players) in complex sports scenes represented by basketball and football has attracted increasing attention. This paper focuses on tracking targets (balls as single targets and players as multiple targets) in competitive sports scenes like basketball and football. A small target detection network based on multi-scale features and a triangulation algorithm is employed to fuse the two-dimensional coordinates of the ball into three-dimensional coordinates. Additionally, a simplified motion model is proposed for the non-linear motion of the ball, and a Kalman filter is used to obtain accurate and smooth three-dimensional tracking trajectories. The proposed method achieves 2D and 3D tracking accuracies of 0.81 and 0.92 on the basketball public dataset, respectively.

Keywords: Motion Object Detection, Competitive Sports, Multi-Scale Features, Triangulation Algorithm

Received: 4 September 2024, Revised 30 November 2024, Accepted 11 December 2024

1. Introduction

With the advancement of technology, artificial intelligence technologies, represented by visual technology, have been widely integrated into various industries, such as intelligent monitoring, smart transportation, and robot visual perception. In recent years, the sports industry, with promising market prospects, has gradually become an experimental field for artificial intelligence technology. With the improvement in the quantity and quality of sports data, as well as the enhancement of computer computing power and continuous algorithm improvements, artificial intelligence technology has brought about many changes in the traditional sports industry, revolutionizing athletic performance, sports commercialization, and national fitness. For example, training-

assistant software based on visual technology effectively enhances athletes' training levels, and automatic sports data statistics systems based on visual technology save a lot of manual annotation costs [1].

Video object tracking mainly involves extracting features of the target and locating the target based on these features. It can generally be divided into three steps: target detection, target feature extraction, and target tracking. In target detection, motion target detection algorithms detect moving targets in the scene, remove interfering target areas, and retain the target areas to be tracked. Motion targets mainly refer to objects with specific semantics in sports videos, such as balls and players. Tracking balls and players is primarily to determine their positions and ultimately obtain their motion trajectories. This work not only helps achieve tracking of specific players, statistical analysis of player running data, 3D virtual scene reconstruction, and assistant referee decisions but also serves as the foundational work for advanced semantic tasks such as action recognition, event detection, exciting video editing, tactical analysis, and player and team performance scoring [2]. Therefore, tracking motion targets has become a significant research focus in sports video analysis, receiving widespread attention from researchers in visual technology.

Due to the complexity of team sports competitions, represented by basketball and football, compared to individual sports events like badminton and tennis, there are more participants, more intense movements, more complex video content, and more incredible difficulty in target tracking. Therefore, this paper focuses on applyingtarget-tracking methods in complex, competitive sports scenes, specifically team sports videos. It employs a small target detection network based on multi-scale features and a triangulation algorithm to fuse the two-dimensional coordinates of the ball into three-dimensional coordinates. Additionally, a simplified motion model is proposed for the non-linear motion of the ball, and a Kalman filter is used to obtain accurate and smooth three-dimensional tracking trajectories.

2. Related Work

Today's sports events increasingly rely on advanced computer technology, whether from the perspective of sports competition and appreciation or in terms of tactical analysis, physiological monitoring of athletes, scientific training, and player selection. Currently, research in the academic field on sports video analysis mainly focuses on the detection and tracking of moving targets (balls and players) on the field, while research on the content analysis of sports videos is relatively limited. The main reason is that the detection and tracking of balls and players in sports videos still face many challenges, such as severe occlusion and similarity in player appearance. The subsequent analysis at the advanced semantic level heavily relies on the detection and tracking results of balls and players [3].

In ball-centric sports, obtaining the ball's motion trajectory quickly and accurately is important for referee decisions, match analysis, and live displays. Initially, people usually separated ball detection from tracking, first detecting the ball and then connecting it to form trajectories. Early ball detection methods mostly used motion object screening based on appearance features or template matching techniques based on color statistics, while tracking methods used template region matching or Kalman filtering. For example, Tolic D et al. first used inter-frame differencing to extract moving objects, then used template matching to detect the football among candidate targets, and tracking was achieved through template search in the neighborhood of the previous football position [4]. Li R. et al. calculated the image gradients first and then used circular detection based on the Hough transform to detect the ball. Due to the football's small scale, fast motion, and frequent occlusions, the detection effect

could have been better [5]. Since balls in team games are often occluded, many missed, and false detections occur during the detection process. Therefore, researchers tend to consider ball detection and tracking together. For example, Ma H T et al. proposed a ball detection method based on trajectory analysis, which first detects multiple candidate objects and then analyzes the motion trajectories of these candidates to determine the ball's motion trajectory [6]. Zhao J W S proposed a ball-tracking algorithm that combines detection and tracking [7]. It first calculates the detection of candidate objects. Then, it obtains the ball in the video through the Viterbi algorithm's optimal path calculation, followed by a Kalman filter to track the ball. The detection and tracking process is repeated by identifying and judging whether the tracker loses the ball based on the tracked area.

The detection and tracking of players aim to obtain the motion trajectories of each player on the field at different moments. In sports competitions, coaches can understand each player's running situation through the player's motion trajectory, thus grasping the player's physical condition or competitive status during training or matches. Early player detection methods usually used image differencing or foreground-background separation. These methods have fast running speeds and good real-time performance but often suffer from severe false positives and false negatives due to factors such as other moving targets, lighting changes, and occlusions in sports videos. With many pedestrian detection methods widely adopting machine learning approaches, some player detection tasks also use machine learning methods. For example, Yu H. et al. used the Ada Boost algorithm to detect players in soccer and basketball scenes, and the player features were represented by handcrafted Histograms of Oriented Gradients (HOG) and Haar features [8]. Deep learning-based object detection methods have become main stream in recent years, and player detection tends to adopt such methods. For example, Ding Y et al. proposed a player detection method based on convolutional neural networks, which can be applied to various complex sports scenes, such as basketball, soccer, and ice hockey, considering player body changes and scale changes caused by camera movements [9].

In summary, research on sports video analysis, including the detection and tracking of moving targets, action recognition, event detection, and content understanding, is gradually advancing at the semantic level, interdependent, and mainly relies on low-level semantic object detection and tracking. Therefore, to better understand the high-level semantic content of sports matches, detecting and tracking moving targets in sports videos is one of the most fundamental problems to be urgently solved at this stage [10]. Since object detection technology is relatively mature, this paper focuses on researching target-tracking methods in sports scenes.

3. Motion Object Detection Algorithms

The main task of motion object detection is to analyze video sequences, extract the regions of moving objects in the images, and remove the regions with interference. Motion object detection can be divided into two categories based on the background: object detection in dynamic backgrounds and object detection in static backgrounds. Object detection in dynamic backgrounds extracts targets from captured video sequences when the camera rotates. Since the background in video sequences is constantly changing, compensation for the background and extracting and matching features for objects in different sequences is required.

3.1 Faster R-CNN Object Detection Algorithm

Faster R-CNN uses the ZF model (Zeiler and Fergus model) and the VGG16 model (Simonyan and Zisserman model) as the backbone networks. The entire network first uses a set of essential convolutional neural networks to extract feature maps of the input images. These feature maps are used in the subsequent RPN and RoI Pooling

(region of interest pooling) layers. The RPN layer first uses the Softmax function to classify the generated anchor boxes as positive or negative samples. Then, it uses bounding box regression to refine the positions of the anchor boxes, obtaining accurate region proposals based on the original image. The loss function is a commonly used metric to evaluate the performance of models in object detection tasks, and the IoU Loss used in Faster R-CNN is the most widely used loss function. Compared to the Smooth L1 Loss used in Fast R-CNN, it addresses the problem of 4-point coordinates not having certain correlations and the problem of significant discrepancies in IoU for detection boxes with the same size. However, it also brings corresponding issues: although it reduces the discrepancies in IoU for detection boxes with the same size, the discrepancies still exist. The IoU is defined as follows:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

In Equation (1): A represents the area of the predicted box, and B represents the area of the ground-truth box.

3.2 ECO Object Tracking Algorithm

The ECO algorithm is a tracking algorithm with efficient convolutional features. This algorithm mainly uses deep and handcrafted features, such as HOG and CN features, to achieve object tracking for the correlation filtering algorithm. The ECO algorithm inherits the C-COT algorithm, transforming the feature maps into the continuous spatial domain through interpolation, as shown in Formula (2):

$$J_d\left\{x^d\right\}(t) = \sum_{n=0}^{N_d-1} x^d[n] b_d\left(t - \frac{T}{N_d} n\right) \tag{2}$$

Among them, b_d represents the interpolation kernel for the period T>0. Φ is the entire interpolated feature. Ψ represents the interpolation function for channels 1 to D. In the C-COT algorithm; continuous multi-channel convolution filters are used for score prediction. ECO improves it using principal component analysis (PCA), as shown in Formula (3):

$$S_{pf}\{x\} = Pf * J\{x\} = f * P^T J\{x\} \tag{3}$$

where f is the filter for channel D, * represents the convolution operation, P represents the projection matrix with D rows and C columns obtained through PCA, and $S_{pf}\{x\}$ represents the response score value. First, a Gaussian mixture model (GMM) method is used to compress the training set, and then the L2 norm of the convolution response score and the Gaussian label error is taken.

3.3 Target Tracking and Detection in Basketball Motion

The Faster R-CNN-based object detection method is used for 2D ball detection in various camera views. Considering the ball's small size, a small object detection method proposed in the literature is adopted here, and the VGG16 backbone network is modified to extract multi-scale features. Although the 2D detection of the ball has improved compared to traditional methods, there are still a few missed detections and false alarms caused by occlusion and noise interference.

The ECO-based object tracking method is used for 2D ball tracking in various camera views. Compared to the original ECO version, two improvements are made: first, cross-view information based on *epipolar* constraint is used to remove false alarms and compensate for missed detections; second, detection information is fully utilized

to update the tracking model, thus alleviating tracking drift caused by occlusion during the ball tracking process.

The triangulation algorithm is used to fuse multiple 2D coordinates effectively into a single 3D coordinate. Given the ball's coordinates in two different camera views, i and j, the output of the triangulation algorithm generally includes the ball's 3D coordinates *ptij* of the ball and the corresponding reprojection error *etijt*. The reprojection error is actually the distance between the obtained 3D coordinates and the true 2D coordinates of the ball's reprojected point. The average value of the 3D coordinates corresponding to errors less than a certain threshold is taken as the final fused 3D coordinates.

Kalman filtering is used to obtain a smooth 3D trajectory of the ball. Considering the ball's nonlinear motion characteristics, a simplified motion model is proposed to apply the Kalman filtering method and obtain a smooth 3D trajectory. Figure 1 shows this section's overall framework of the proposed 3D ball tracking method. The left side of the figure represents a multi-camera scene composed of three fixed cameras, where solid circles represent balls appearing in different cameras, and rectangular dashed boxes represent occlusion obstacles.

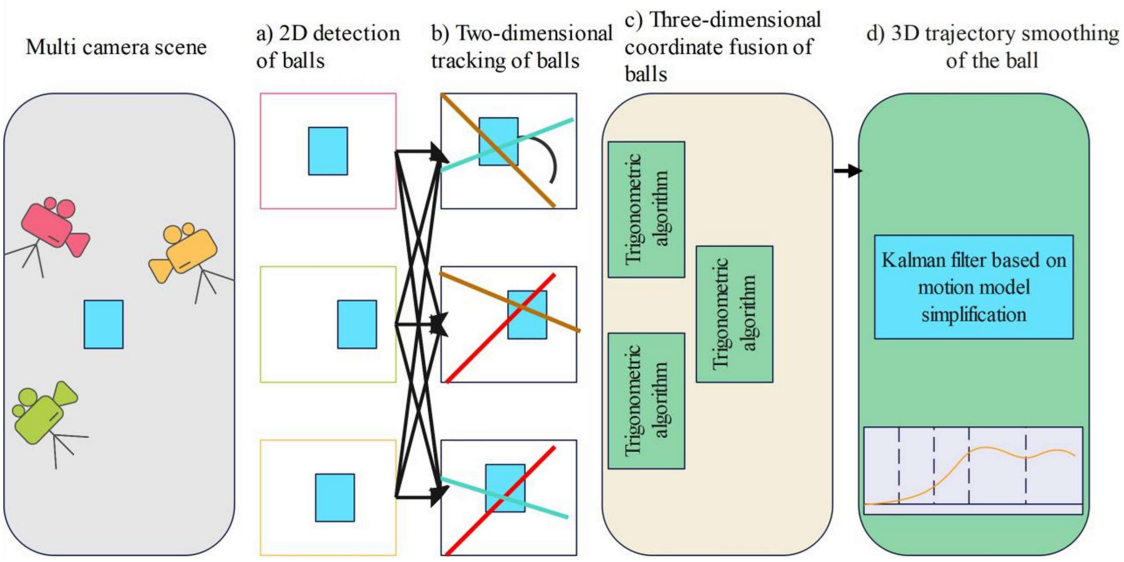

Figure 1. Shows the overall schematic diagram of the 3D ball tracking framework based on small object detection and multi-view fusion

4. Experimental Design and Results Analysis

4.1 Experimental Dataset

APIDIS Basketball Dataset: This dataset is captured using 7 synchronized cameras, with 5 cameras installed on the ground and two fisheye cameras installed at the top of the basketball court. Each camera has a resolution of 800*600 and a frame rate of 25fps. Due to the influence of camera positions and lighting conditions, the ball tracking in this dataset is challenging. Despite this difficulty, this dataset is one of the few public datasets involving team sports videos. In addition to 1500 frames of synchronized image sequences used for testing and comparison, the APIDIS dataset also includes image sequences from other periods of the basketball game. For the Faster R-CNN detection method, the image frames from different periods are used for annotation and model training. In

contrast, the mentioned 1500 frames of synchronized image sequences are used as the test dataset.

Self-collected Dataset: This dataset consists of two sets, one collected during a basketball training match at a university in China and the other collected during a youth football match at a primary school in China. Both datasets were captured using 8 ordinary surveillance cameras installed around the sports fields. Compared to the APIDIS basketball dataset, these two datasets have better lighting conditions, higher resolution, and are more conducive to 3D ball tracking.

The primary evaluation metric for target tracking is tracking accuracy, which describes the proportion of frames in an image sequence where the distance deviation between the estimated position and the true position of the tracking is less than a given threshold. The position refers to the center coordinate of the ball, and the distance is represented by the Euclidean distance between two center coordinates.

4.2 Experimental Results Analysis

In the 2D detection stage of the ball, this paper adopts multi-scale features from small object detection methods, which effectively alleviates the missed detection and false alarm phenomena in the 2D detection process. In the 2D tracking stage of the ball, this paper solves the tracking drift problem by introducing cross-view information based on *epipolar* constraint and a detection-based model update strategy. The comparison of the results of various 2D benchmark methods is shown in Figure 2. Figure 2 is a curve graph between accuracy and 2D-pixel threshold. The horizontal axis represents the threshold from 1px to 50px, and the vertical axis represents the proportion of frames with tracking deviation less than the corresponding threshold (i.e., accuracy). From Figure 2, it can be observed that the introduction of multi-scale features improves the 2D detection accuracy of the ball from 0.68 to 0.74. The improved ECO tracking algorithm based on multi-view fusion further enhances the 2D tracking accuracy of the ball, increasing it from 0.67 to 0.77 compared to the improved KCF tracking algorithm. After fusing the 3D coordinates and conducting 3D tracking, the 3D coordinates are reprojected back to the 2D camera plane. The corresponding accuracies of the 2D results are 0.76 (iVGG16 + iECO + Triangulation) and 0.81 (iVGG16 + iECO + Triangulation + iKalman).

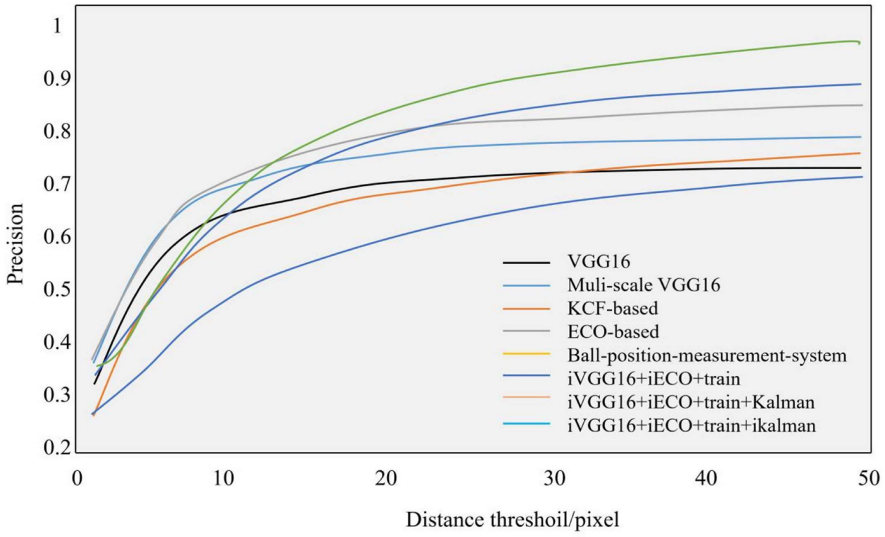

Figure 2. Shows the precision-threshold relationship curves corresponding to various methods at the 2D level

At the 3D level, the triangulation-based 3D coordinate fusion method and the Kalman filter-based 3D ball tracking method can obtain the 3D coordinates of the ball. Similar to the 2D level, the precision-threshold relationship curves are used to represent the 3D results, as shown in Figure 3. The horizontal axis represents the threshold range set in 3D space (from 1 cm to 1000 cm), and the vertical axis represents the tracking accuracy. From Figure 3, it can be seen that compared to the 3D results obtained from 3D coordinate fusion, the accuracy corresponding to the 3D trajectory after Kalman filtering is improved from 0.79 (iVGG16 + iECO + Triangulation) to 0.81 (iVGG16 + iECO + Triangulation + Kalman). The improved Kalman filter based on simplified motion models further increases the accuracy to 0.92 (iVGG16 + iECO + Triangulation + iKalman). This indicates that the proposed method in this paper performs better than other relevant methods at various thresholds.

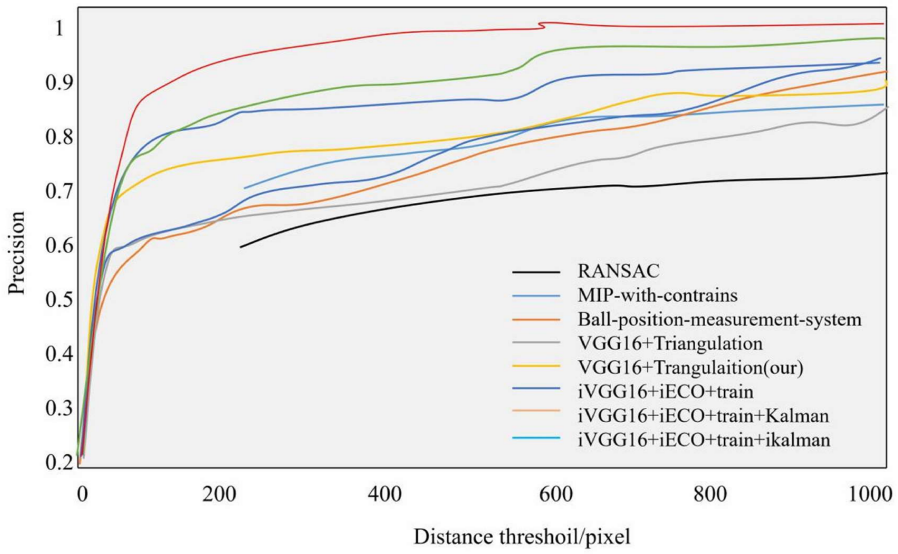

Figure 3. The accuracy-threshold curve displayed at the 3D level

5. Conclusion

Visual technology has been widely applied in sports video analysis in recent years. The detection and tracking of moving objects in sports videos have gradually become hot topics in visual technology research. This paper focuses on tracking balls and players in complex team sports scenarios and proposes a ball's 3D tracking framework based on small target detection and multi-view fusion. Additionally, it introduces cross-view information based on *epipolar* constraints and detection-based model updating to address tracking drift issues. At the 3D level, this paper adopts a triangulation-based 3D coordinate fusion method to merge 2D ball coordinates from multiple views into 3D coordinates. Numerous experiments demonstrate that the designed ball's 3D tracking method can effectively overcome the challenges of small ball size, frequent occlusions, camera calibration errors, and background interference.

References

[1] Katyal, A., Singla, R. (2021). Synchronized Detection of Evoked Potentials to Drive a High Information Transfer Rate Hybrid Brain-Computer Interface Application. *Advanced Biomedical Engineering, 10*, 58-69.

[2] Bhondekar, A. P., Vig, R., Singla, M. L., et al. (2009). Genetic algorithm-based node placement methodol

ogy for wireless sensor networks. *Proceedings of the International Multiconference of Engineers and Computer Scientists, 1*, 18-20.

[3] Jagtap, A. M., Gomathi, N. (2017). Minimizing sensor movement in target coverage problem: A hybrid approach using Voronoi partition and swarm intelligence. *Bulletin of the Polish Academy of Sciences Technical Sciences, 65*(2).

[4] Tolic, D., Fierro, R., Ferrari, S. (2009). Cooperative multi-target tracking via hybrid modeling and geometric optimization. *2009 17th Mediterranean Conference on Control and Automation*. IEEE, 440-445.

[5] Li, R., Xu, H. C., Fu, J. W. (2014). Research on the Detection of Moving Target for Sport Video Object Analysis Using Hybrid Algorithm. *Applied Mechanics & Materials, 687-691*, 3889-3892.

[6] Ma, H. T., He, Y. J., Wang, C. M., et al. (2018). Research on human detection and tracking technology based on UAV vision. *Computer Technology and Development, 28*(10), 115-118.

[7] Zhao, S., Luo, J., Wei, S. (2021). A hybrid eye movement feature recognition of classroom students based on machine learning. *Journal of Intelligent & Fuzzy Systems, 40*(2), 2803-2813.

[8] Yu, H., Sharma, A., Sharma, P. (2021). Adaptive strategy for sports video moving target detection and tracking technology based on mean shift algorithm. *International Journal of System Assurance Engineering and Management*, 1-11.

[9] Ding, Y., Wei, C., Wang, X., et al. (2020). Realization of the Airborne System for Moving Target Detection and Tracking Based on Quadrotor. *IOP Conference Series: Materials Science and Engineering, 782*(5), 052005 (9pp).

[10] Kim, C., Suh, J., Han, J. H. (2020). Development of a Hybrid Path Planning Algorithm and a Bio-Inspired Control for an Omni-Wheel Mobile Robot. *Sensors, 20*(15), 4258.