Tracking Multiple Players in Beach Volleyball Videos

Xiaokang Jiang, Zheng Liu, and Yunhong Wang(✉)

The Laboratory of Intelligent Recognition and Image Processing,
School of Computer Science and Engineering,
Beihang University, Beijing 100191, China
yhwang@buaa.edu.cn

Abstract. Multi-object tracking has been a difficult problem in recent years, especially in complex scenes such as player tracking in sports videos. Player movements are often complex and abrupt. In this paper, we focus on the problem of tracking multiple players in beach volleyball videos. To handle the difficulties of player tracking, we follow the popular tracking-by-detection framework in multi-object tracking and adopt the multiple hypotheses tracking (MHT) algorithm to solve the data association problem. To improve the efficiency of the MHT, we use motion information from Kalman filter and train an online appearance model of each track hypothesis. An auxiliary particle filter method is adopted to handle the missing detection problem. Furthermore, we obtain the significant performance on our beach volleyball datasets, which demonstrate the effectiveness and efficiency of the proposed method.

1 Introduction

With the explosive growth of various video data, automatic video processing has become more and more important in order to reduce the manual effort for video analysis. Among all kinds of video data, sports videos captured from different kinds of matches, such as football, basketball and volleyball has attracted a lot of research interests, due to their huge popularity and tremendous commercial value.

Compared with multiple object tracking (MOT) in other scenes, multiple player tracking in sports video is much more difficult due to the following reasons: (1) players in the same team are always visually similar, making the appearance information less discriminative and unreliable; (2) sports players often interact with others in complex ways and (3) the occlusions are much more frequent and severe. All of these issues together have posed quite a great challenge to the tracking system, which requires not only reliable observations but also a sophisticated tracking strategy to make the system robust.

Recent progress on Multi-Object Tracking (MOT) has focused on the tracking-by-detection strategy, where object detections from a category detector are linked to form trajectories of the targets. In this work, we aim to track multiple highly dynamic and interactive players in beach volleyball videos. Firstly, we follow the popular tracking-by-detection framework in MOT. Concretely, we employ the DPM [1] detector to get the detection results. Next, to solve the data association problem, the classical multiple hypotheses tracking (MHT) algorithm [2] is adopted.

Z. Zhang and K. Huang (Eds.): IVS 2016, CCIS 664, pp. 65–71, 2016.
DOI: 10.1007/978-981-10-3476-3_8

The contributions of this paper lie in the following three-fold: (1) We train multiple online appearance models to improve the efficiency and accuracy of the MHT tacking method; (2) An auxiliary color-based particle filter is applied to handle missing detections; (3) We analyze the robustness of the method, in particular the influence of each part of the tracking system.

The paper is structured as follows: After discussing related work in the following section, Sect. 3 describes the details of MHT. Section 4 talks about how the online appearance model and the particle filter help improving the tracking result. Section 5 presents experiments and analyzes the performance of our method.

2 Related Work

As object detection has made impressive improvements in recent years, MOT has focused on the tracking-by-detection strategy. Different from the data association-based tracking approaches, there are many methods use the detection results from a probabilistic inference perspective. A Kalman filter [3] is an early representative method. Then based on sequential Monte Carlo sampling, particle filters [4] gained much attention because of their simplicity, generality, and extensibility in a wide range of challenging applications.

There are many papers about multiple player detection and tracking in sports video. In [5], Huang et al. first detected the players and ball based on extracted foreground and then performed shape analysis to remove false alarms. To conquer the problem of complex multiple object interactions with overlaps and ambiguities, Okuma et al. [6] used an offline boosted particle filter to track each player in a mixture representation.

However, probabilistic inference methods cannot solve the highly dynamic and interactive players in sports videos, due to the severe occlusions and abrupt movements of players. So recently many methods have posed multi-object tracking as data association. Majority of the batch methods formulates MOT as a global optimization problem in graph-based representation, due to their computational efficiency and optimality. The problem of associating tracklets has been investigated using a variety of method, such as the Hungarian algorithm [7], k-shortest paths [8], cost-flow networks [9] and discrete-continuous optimization [10]. Kim et al. [11] follow the classical formulation in [2] and corporate the appearance modeling with MHT. We also apply the MHT in multiple player tracking of beach volleyball.

3 Multiple Hypotheses Tracking Algorithm

In the MHT algorithm, observations are localized bounding boxes. Generally, the MHT framework consists of the following five steps.

3.1 Track Tree Construction and Updating

A track tree encapsulates multiple hypotheses starting from a single observation. At each frame, a new track tree is constructed for each observation. Previously existing track trees are also updated with observations from the current frame.

3.2 Gating

To predict the tracking area of each track hypothesis, the motion information is taken into consideration. The Kalman filter is used to predict the location of the target. Then the Mahalanobis distance d^2 between the predict location and a new observation is calculated to decide whether to update a particular trajectory. The distance threshold d_{th} determines the size of the gating area.

3.3 Tracking Score

Each track hypothesis is associated with a track score. The l^{th} track's score at frame k is defined as follows:

$$S^l(k) = \omega_{mot} S^l_{mot}(k) + \omega_{app} S^l_{app}(k) \tag{1}$$

where $S^l_{mot}(k)$ and $S^l_{app}(k)$ are the motion and appearance scores, and ω_{mot}, ω_{app} are the weights that control the contribution of the location measurement and the appearance measurement to the track score, respectively. The motion and the appearance score are calculated like the formulation in [11].

3.4 Global Hypothesis Formation

Given the set of trees that contains all trajectory hypotheses of all targets, we wish to determine the most likely global hypothesis. We follow the formulation in [2]. The global hypothesis formation problem is formulated as a Maximum Weighted Independent Set (MWIS) problem.

3.5 Track Tree Pruning

In order to avoid the exponential growth of the graph size, the pruning step is applied for MHT. We adopt the standard N-scan pruning approach. Figure 1 shows an example.

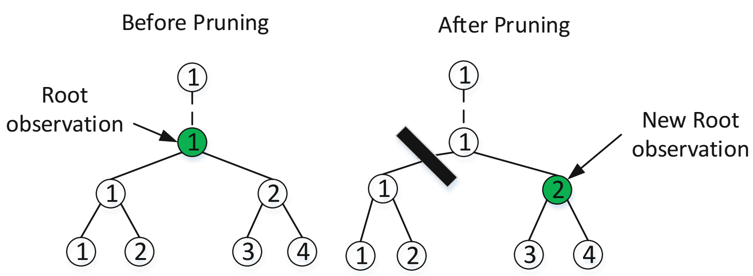

Fig. 1. An example of N-scan pruning ($N = 2$)

We also set a threshold B_{th} to retain a track tree's branches based on its track score. Only the top B_{th} branches are kept. This is because the number of players in beach volleyball is static. In this way, we can make pruning more efficient.

4 Online Appearance Model and Particle Filter Method

4.1 Online Appearance Model

Considering the complexity of beach volleyball player tracking, in addition to the motion estimates method, we build an online appearance model of each target. We utilize the convolutional neural network features trained on the ImageNet+PASCAL VOC dataset in [12]. The 4096-dimensional feature for each observation box is extracted. To save the time and space, a principal component analysis (PCA) is then performed to reduce the dimensionality of the features. In the experiments we take the first 256 principal components. We follow the multi-output regularized least squares framework [13].

4.2 An Auxiliary Particle Filter

In the MHT framework, if there is a sequence of missing observations, the corresponding track hypothesis will be deleted from the hypothesis space. This may cause false tracking results, which is resulted from missing detections of the detector. Figure 2 shows an example.

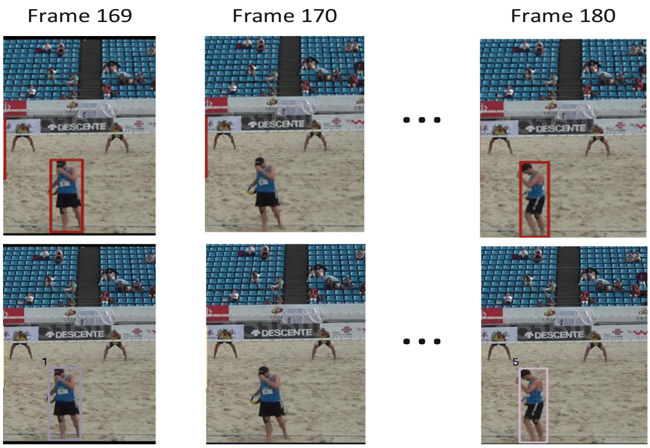

Fig. 2. Illustration of missing detection problem. The first row is the missing detection. The second row is the wrong detection results of MHT (player 1 was identified as player 5)

In our work, color information of each player is used for the observation model in the particle filter framework. We take the HSV color space of each player and create a one dimensional histogram containing $N_H * N_S$ bins and N_V bins appended.

Firstly, the color histogram of each player is calculated on the first frame based on the bounding box with high detection confidential scores. When a bounding box has no adjacent boxes on the next frame, we assume that the missing detection problem happens. Then a similarity matric between the bounding box and players is calculated. If the bounding box is similar to one of the players, a particle filter is initialized to predict the position of the bounding box in the next frame. The particle filter continues until a new bounding box from the detector is found around the predict position. In this way, we can avoid false tracking results in some degrees.

5 Experiments

We evaluate the performance of our framework on a Beach Volleyball Dataset. There are 200 frames labeled for beach volleyball games.

The entire system is implemented in Matlab on the platform of Linux. The popular DPM detector, which is publicly available, is used to get bounding boxes of players. The model trained on the INRIA Person dataset is applied. In Fig. 3, we show some tracking output on our beach volleyball datasets. We can see that our tracking method can handle the occlusion problem well. For comparison, we choose the state-of-the-art MOT method proposed in [11]. The same DPM detector is used in our dataset.

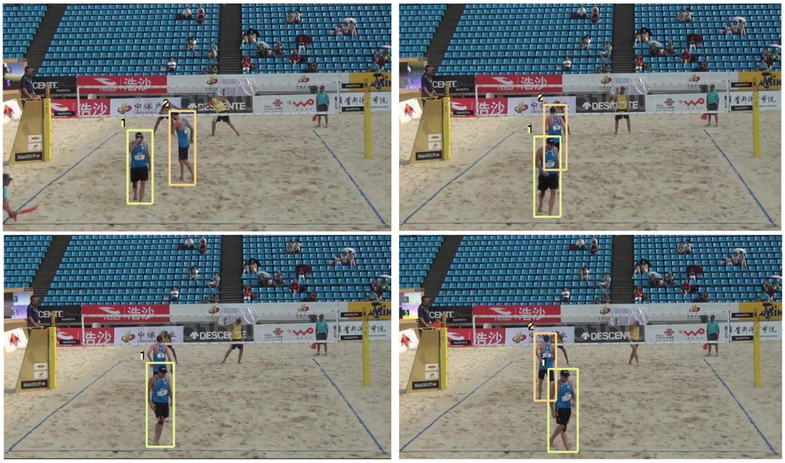

Fig. 3. Tracking output on the beach volleyball dataset.

Algorithm Parameters: In the proposed tracking algorithm, as to MHT, we set $\omega_{mot} = 0.2$ and $\omega_{app} = 0.8$ considering that the appearance model plays a more important role. And the max number of each track tree's branches B_{th} is set to 10. The dummy numbers threshold $N_{miss} = 15$. N about the N-scan is set to 7 in our experiment. And the Mahalanobis distance $d_{th} = 12$. As to the particle filter method, the particle number of each target is set to 50. As to color histogram, we set $N_H = 10$, $N_S = 10$ and

$N_V = 10$. The smaller number of particles means faster speed, while may result in lower accuracy.

Evaluation: We follow the current evaluation protocols in [14] for visual multi-target tracking. The multiple object tracking accuracy (MOTA), multiple object tracking precision (MOTP). The number of false positives (FP), the number of false negatives (FN), the total number of identity switches (IDS) are also reported.

Table 1 shows the results about our method and the MHT_DAM method on the beach volleyball dataset. We also list the results about the method without the help of the auxiliary particle filter for comparison.

Table 1. Results on the beach volleyball dataset

Method	MOTA	MOTP	FP	FN	IDS
MHT_DAM [11]	64.3%	66.7%	36.3%	2.6%	2
Ours(without PF)	64.5%	67.3%	35.1%	2.6%	2
Ours	67.2%	70.7%	29.7%	0.3%	1

From Table 1, we can see that the MOTA and MOTP of our method are better than the MHT_DAM and method without particle filtering. The FN becomes smaller because the missing detections are alleviated. What's more, the identity switches between players becomes less.

6 Conclusion

We have proposed a novel method based on the tracking-by-detection framework to handle the MOT problem for beach volleyball videos. The MHT method is adopted to overcome the false positive detections in the data association procedure. Online appearance models are trained to improve the efficiency and accuracy of the MHT tacking method. To alleviate the missing detection problem, a color-based particle filter is utilized. Through this way, better observations are provided for multiple player tracking. Experimental results demonstrate that the proposed method achieves better performance, compared with the state-of-the-art approaches.

Acknowledgments. This work was supported in part by the Hong Kong, Macao and Taiwan Science and Technology Cooperation Program of China (No. L2015TGA9004), and the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (No. 61573045, 61421003).

References

1. Felzenszwalb, P.F., et al.: Object detection with discriminatively trained part-based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1627–1645 (2010)

2. Papageorgiou, D.J., Salpukas, M.R.: The maximum weight independent set problem for data association in multiple hypothesis tracking. In: Hirsch, M.J., Commander, C.W., Pardalos, P.M., Murphey, R. (eds.) Optimization and Cooperative Control Strategies. LNCIS, vol. 381, pp. 235–255. Springer, Heidelberg (2009). doi:10.1007/978-3-540-88063-9_15
3. Kalman, R.E.: An introduction to Kalman filter. TR 95-041, University of North Carolina at Chapel Hill, Department of Computer Science (1995)
4. Isard, M., Blake, A.: Condensation—conditional density propagation for visual tracking. Int. J. Comput. Vis. **29**(1), 5–28 (1998)
5. Huang, Yu., Llach, J., Bhagavathy, S.: Players and ball detection in soccer videos based on color segmentation and shape analysis. In: Sebe, N., Liu, Y., Zhuang, Y., Huang, Thomas, S. (eds.) MCAM 2007. LNCS, vol. 4577, pp. 416–425. Springer, Heidelberg (2007). doi:10.1007/978-3-540-73417-8_50
6. Okuma, K., Taleghani, A., Freitas, N., Little, James, J., Lowe, David, G.: A boosted particle filter: multitarget detection and tracking. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 28–39. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24670-1_3
7. Bae, S.-H., Yoon, K.-J.: Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE (2014)
8. Berclaz, J., et al.: Multiple object tracking using k-shortest paths optimization. IEEE Trans. Pattern Anal. Mach. Intell. **33**(9), 1806–1819 (2011)
9. Butt, A.A., Collins R.T.: Multi-target tracking by lagrangian relaxation to min-cost network flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2013)
10. Andriyenko, A., Schindler, K., Roth, S.: Discrete-continuous optimization for multi-target tracking. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2012)
11. Kim, C., et al.: Multiple hypothesis tracking revisited. In: Proceedings of the IEEE International Conference on Computer Vision (2015)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (2012)
13. Li, F., et al.: Video segmentation by tracking many figure-ground segments. In: Proceedings of the IEEE International Conference on Computer Vision (2013)
14. Bernardin, K., Stiefelhagen, R.: Evaluating multiple object tracking performance: the CLEAR MOT metrics. EURASIP J. Image and Video Process. **2008**(1), 1–10 (2008)