Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080210
ISSN 2371-8838 Vol. 8 Num. 2

Tracking algorithm based on spatial progressive matching strategy and optimized correlation

Yunhua Jia[1,a], Siyue Cha[1,b], Chengjiang Zhou[1,c,*]

[1]Information Institute, Yunnan Normal University, 768 Juxian Street, Kunming, China
[a]jia_yunhua@163.com, [b]3124736080@qq.com, [c]chengjiangzhou@foxmail.com
[*]Corresponding author

Keywords: Target tracking, Computer vision, Correlation optimization

Abstract: The task of multi-target tracking is to correctly associate the identity of the same target in the two scene scenes. How to improve the accuracy of target tracking is still full of challenges. In this article, we propose a tracking algorithm based on spatial progressive matching strategy and optimized correlation. In it, we divide the targets in the scene according to the area of the target detection frame in the scene. By Prioritizing matching of target groups with a larger target frame area, and then matching target groups with a smaller target frame area is the spatial progressive matching strategy we propose. We noticed that in certain scenes where the target moves too quickly, the traditional intersection-to-union method becomes limited to some extent. Therefore, we substituted it with a circular intersection-to-union ratio method, which is more effective in accurately associating the targets in those scenes.

1. Introduction

Multi-target tracking is one of the important research fields of computer vision. With the development of the field of artificial intelligence in recent years, multi-target tracking technology has become increasingly mature and gradually integrated into our daily lives. Currently, target tracking technology is effectively used in areas such as overdrives cars, technologically advanced medical scenarios, sports competition result analysis[1][2], and intelligent robots[3]. The application of target tracking technology in these fields will continue to have a positive impact on our future lives.

Since the neural network was proposed, many computer disciplines have developed rapidly, among which target detection technology has a greater impact. Among them, the most influential one is undoubtedly the YOLO series of trackers. After several years of development, the YOLO detector currently covers multiple versions. After the qualitative development of the target detector, it has also played a very active role in the development of multiple disciplines. The impact, the greater impact is on the field of target tracking. Target detection can be said to be the eye of multi-target tracking. The detector can provide the tracker with a lot of useful target information, such as the location of the target, the target category, and the confidence of the target. According to the effective information about the target provided by the detector, the target Researchers in tracking-related fields can use this basis to track targets more effectively. Sort tracker[4] is one of the most representative multi-target trackers. It was proposed by Alex Bewley and others in 2016. The tracker uses the Kalman filter to predict the position information of the target in upcoming frames. In addition, the team used the Hungarian algorithm to determine the correlation between the detection target and the orbital target. This approach greatly enhanced the operating efficiency in comparison to the multi-target tracking algorithm of previous years. At the same time, its accuracy is also greatly improved compared to other tracking methods. The proposal of this method brought the development of the field of multi-target tracking into the era of two-stage trackers and promoted the development of the field of multi-target tracking. However, due to the impact of its matching strategy, its tracking accuracy was also restricted at the same time. In order to improve the accuracy of multi-target tracking, we propose a spatial progressive matching strategy based on the characteristics of the target in the scene. This strategy integrates the characteristics of near-large and far-small targets, and detects targets based on the size of their detection frame area. Matching the larger target groups first, and then matching the smaller target groups, makes the target matching have certain spatial logic, thus improving the accuracy of target tracking. In addition, in some scenes where the target moves too fast, the traditional intersection and union method cannot effectively correlate the target. In view of this problem, we replace the traditional target rectangular frame with a circular frame with a larger search range. Through this method, the tracking accuracy of multiple targets in scenes that move too fast is improved. By combining the two methods mentioned above, the tracking accuracy of our proposed tracking algorithm is improved compared to the baseline method. The method we propose mainly includes the following three points:

(1) A spatial progressive matching strategy is proposed to make target matching more spatially logical information and more conducive to the correct matching of targets.

(2) Combined with the circular intersection ratio method, in some scenes where the target moves too fast, it is more conducive to maintaining the identity information of the target than the traditional intersection ratio method.

(3) By combining the spatial progressive matching strategy and the circular intersection and union ratio method, the tracking accuracy of our proposed tracking method is improved compared to the baseline method.

2. Related Work

In this part, we will introduce some related work of our research, including target detectors and consistency measurement methods.

Object detector. YOLO (You Only Look Once) is an advanced target detection technology series. Its unique feature is that it can achieve accurate positioning and classification of targets in the image through a single forward pass, while maintaining a high detection speed. From the earliest YOLOv1[5] to the latest YOLOv8[6], this series has continued to evolve, introducing key technologies such as multi-scale feature maps, anchor boxes, and residual structures, allowing the model to have stronger generalization capabilities and better performance while maintaining high performance. Fast training speed has become one of the important technologies in the field of target detection.

Methods of measuring consistency. The intersection ratio is a consistency measurement method often used in the field of multi-target tracking. It is also called the overlap degree. It is mainly used to measure the degree of overlap between the detection frame and the tracking frame. It is usually used to match the detected detection target with the previous one. The IOU method is widely used in the field of target tracking, and plays a role as a tool in major trackers. Through the IOU method, problems such as overlap and occlusion between targets can be solved to a certain extent. It has been solved to a certain extent, which plays a great role in improving the accuracy of the target tracking system. Currently, the IOU method is still a popular method for measuring tracker consistency.

3. Method

In this section, we mainly introduce the proposed tracking framework and the circular intersection and union ratio method.

3.1. Tracking Framework

As shown in Figure 1, including our proposed matching strategy, we describe our proposed spatial progressive matching strategy according to the process shown in the figure. First, we divide all target detection frames into detection frames with larger areas and detection frames with smaller areas according to their area size. After the division is completed, we divide the detection frame set with a larger area into a detection frame set with a larger confidence level and a detection frame set with a smaller confidence level according to its confidence level. Next, we first match and associate the detection frame set with larger area and higher confidence with the tracking frame. Then, the tracking frames that are not successfully associated are matched and associated with detection frames that are larger in area and have lower confidence. Then, the tracking frames that are still not successfully associated in the second matching are matched and associated with the smaller detection frame set. After the three matching associations are completed, we match all the unmatched detection frames and all the unmatched tracking frames again. After that, we will update the successfully matched orbits using Kalman filtering. If the life of the unmatched orbits reaches a certain length, we will directly delete them. For unmatched detection frames, a new identity is given and a new track is established.

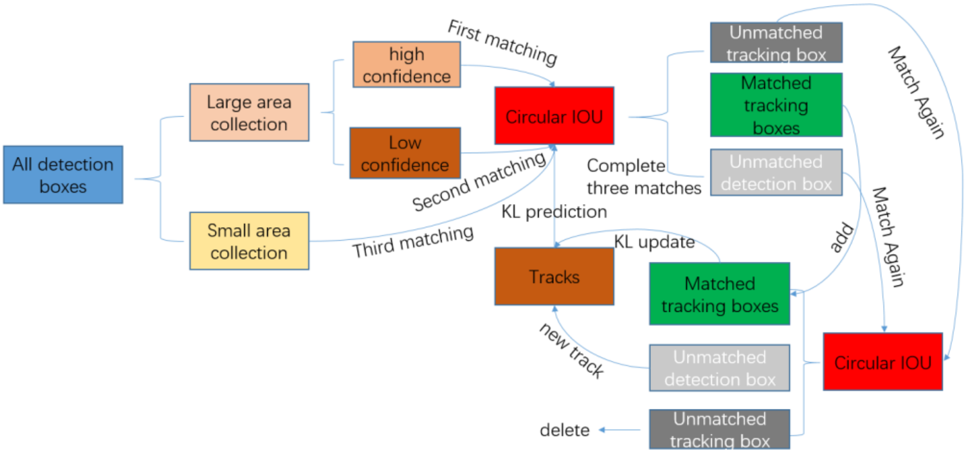

Figure 1: Tracking framework flow chart

3.2. Circular Intersection Ratio

As shown in Figure 2, assume that A is the tracking frame we predict, and B is the target detection frame we use the detector to detect in this frame. When the target moves too fast, the prediction may be inaccurate when using the predictor. In this case, the same situation as the target orbit box and detection box in Figure 2 may occur. It is impossible to use the original IOU method at all. The target can be correctly associated. However, at this time, after we use the circular intersection ratio method, we can find that the detected target overlaps with the orbital target. The overlap value calculated in this way can serve as an important basis for target association.

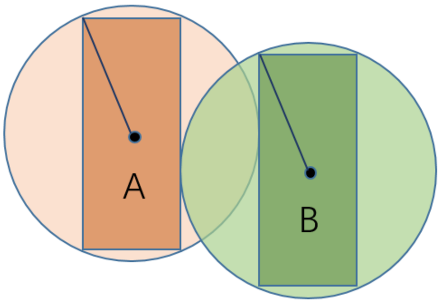

Figure 2: Overview of the circular intersection and union ratio method

$$\text{part1} = r_1^2 * \frac{\arccos(d^2 + r_1^2 - r_2^2)}{2 * d * r_1} \tag{1}$$

$$\text{part2} = r_2^2 * \frac{\arccos(d^2 + r_2^2 - r_1^2)}{2 * d * r_2} \tag{2}$$

$$\text{part3} = \frac{\sqrt{(-d + r_1 + r_2)*(\text{d}+r_1 - r_2)*(\text{d-}r_1 + r_2)*(\text{d}+r_1 + r_2)}}{2} \tag{3}$$

$$\text{overlap} = part1 + part2 - part3 \tag{4}$$

The formulas we list are only for overlapping situations like Figure 2. Among them, r_1 in the formula represents the radius of target A. We take this radius as half of the diagonal of the rectangular frame. r_2 represents the radius of target B. The radius value is consistent with target A. In the formula, d represents the center distance between target A and target B. In Formula 1, part1 is described as the sector area enclosed from the center of the circle of target A to two intersection points. In Formula 2, part2 is described as the sector area enclosed from the center of the circle of target B to two intersection points. In Formula 3, part3 represents the area of the quadrilateral formed by connecting the intersection points from the center of the circle of target A and the center of the circle of target B respectively. In Formula 4, overlap represents the area of the circular intersection of target A and target B. After we obtain the circular intersection area according to the above formula, we can calculate the intersection ratio of the two targets just like the IOU method.

4. Experiment

4.1. Data Set

The MOT16 data set is a public multi-target tracking related data set. It has 7 video sequences of training data, which are shot by fixed cameras and mobile cameras. This data set covers multiple scenes indoors, outdoors, streets, busy intersections, and shopping malls. Most of the resolutions of the images are 1920*1080. The MOT16-05 data set is special. The resolution of the video frames it contains is 640*480.

4.2. Implementation Details

All our experiments were conducted on an Ubuntu-based system with an NVIDIA GeForce GTX 1050 Ti graphics card. The experiments conducted were all developed using the Python programming language and based on the PyTorch machine learning library. According to the current practice of other trackers, we also intercept the second half of each video sequence in the training data set as our ablation data set.

4.3. Experiments On MOT16 Ablation Data Set

As shown in Table 1, we compared our proposed method with the Baseline method on the MOT16 ablation data set. The Baseline method was re-implemented by us based on the idea of the Sort method. Among them, HOTA, MOTA, ASSA and IDF1 are all representative evaluation indicators in the tracking field. These are related to the accuracy of tracking and the robust performance of tracking. Among them, 'Ours-Circle_IOU' is the experimental data of the matching strategy we proposed on the baseline method, and 'Ours' is the method of circular intersection and union ratio that we added on top of the matching strategy. Note that since we took into account the characteristics of the data set, we only added the circular intersection and union ratio method to the MOT16-13 sub-video sequence. The specific improvement effect is shown in Table 2. The targets in the MOT16-13 data set are characterized by fast movement. Our proposed circular intersection ratio method greatly improves multiple metrics on this dataset. It is easy to see from Table 1 and Table 2 that the effect of our proposed method is significantly improved compared to the baseline method. Especially in scenarios where the target moves quickly, the circular intersection method can better reflect its tracking advantages.

Table 1: Comparison between our proposed method and the Baseline method on the MOT16 ablation data set

	HOTA	MOTA	ASSA	IDF1
Baseline	67.017	77.838	67.531	78.026
Ours-Circle_IOU	67.226	77.581	67.944	78.538
Ours	67.434	77.851	68.246	78.966

Table 2: Detailed comparison between our proposed method and Baseline method on MOT16-13 ablation video sequence

	HOTA	MOTA	ASSA	IDF1
Baseline	57.329	72.79	55.788	72.807
Ours-Circle_IOU	58.437	71.966	57.234	73.69
Ours	62.139	76.715	62.609	80.946

5. Conclusion

In this paper, we propose a tracking method based on spatial progressive matching strategy and circular intersection-union ratio. We divide the target according to the size of the detection frame in the scene, giving priority to the detection target group with a larger detection frame, and then matching the target group with a smaller detection frame. This is the spatial progressive matching strategy we propose. Then based on the movement characteristics of the target, we improved the traditional IOU method, taking the center of the rectangular frame as the center of the circle, and the diagonal of the rectangular frame as the radius, and calculating the intersection ratio in a circular shape. When the target moves relatively quickly. In fast scenarios, we obtain better tracking performance. Overall, our proposed method improves tracking performance to a certain extent.

Acknowledgements

This project was funded by the Yunnan Normal University Graduate Research Innovation Fund (Project No.: YJSJJ23-B172). I also extremely grateful to my friends and tutor in the course of preparing this paper.

References

[1] Kuru A. Exploring experience of runners with sports tracking technology [J]. International Journal of Human–Computer Interaction, 2016, 32(11): 847-860.

[2] Huang Y C, Liao I N, Chen C H, et al. Tracknet: A deep learning network for tracking high-speed and tiny objects in sports applications[C]//2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2019: 1-8.

[3] Senanayake M, Senthooran I, Barca J C, et al. Search and tracking algorithms for swarms of robots: A survey[J]. Robotics and Autonomous Systems, 2016, 75: 422-434.

[4] Bewley A, Ge Z, Ott L, et al. Simple online and realtime tracking[C]//2016 IEEE international conference on image processing (ICIP). IEEE, 2016: 3464-3468.

[5] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[6] Terven J, Córdova-Esparza D M, Romero-González J A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.