International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC 2014)

Multiple Object Tracking Based on Motion Estimation and Structural Constraints

Wan Qi
Laboratory of Intelligent Vision Based Monitoring for Hydropower Engineering
College of Computer and Information,China Three Gorges University
Yichang, Hubei, China
ivy_wanqi@163.com

Chen Peng
Laboratory of Intelligent Vision Based Monitoring for Hydropower Engineering
College of Computer and Information,China Three Gorges University
Yichang, Hubei, China
chenpeng@ctgu.edu.cn

Liu Jun-qing*
Laboratory of Intelligent Vision Based Monitoring for Hydropower Engineering
College of Computer and Information,China Three Gorges University
Yichang, Hubei, China
junqingliu@ctgu.edu.cn

Lei Bang-jun
Laboratory of Intelligent Vision Based Monitoring for Hydropower Engineering
College of Computer and Information,China Three Gorges University
Yichang, Hubei, China
Bangjun.lei@ieee.org

Abstract—To solve the time-consuming problem and the low efficiency of the global exhaustive searching in the object tracking, this paper propose a new search strategy based on motion estimation and structural constraints. First, the motion vector of one object is calculated, associating with the location of the object in the previous frame, its moving direction and scope are predicted in the current frame. Then, with the combination of structural constraints between objects, the accurate search direction and scope of the other targets can be determined. We choose five videos for the experiment to confirm the superiority of the search algorithm in this paper. For each video, all these measurements are averaged over all objects, over all frames, and over five separate runs of the tracker. Experimental results show that the new search method can narrow the search range and enhance the searching efficiency under the condition of no affect on the tracking accuracy, thus the complexity of the multi-object tracking algorithm will be reduced.

Keywords-multiple object tracking; motion estimation; motion vector; structural constraints; online structured SVM algorithm

I. INTRODUCTION

Object tracking is a well-studied aspect in computer vision, and has been widely used in many practical applications (e.g., military guidance[1], robot[2], intelligent transportation[3], pedestrian detection[4]). The key factors to realize the object tracking are correct object segmentation, reasonable object representation and accurate object identifycation. The object feature extraction is prerequisite of object tracking and detection. Generally the foreground detection[5] is adopted to obtain the tracking object. In recent years, approaches for model-free tracking[6] became popular. In model-free tracking, the object of interest is manually annotated in the first frame of a video sequence (using a rectangular bounding box). The object feature is extracted in the rectangular box area, and we train a classifier with the object and background characteristics as inputs to get positive and negative samples. Then, samples are selected in the area that object may occur and the positive and negative samples are updated throughout the rest of the video. Combining the classify-cation results with some appropriate method, we can determine the position of the object. Little object information and dramatic changes in object appearance make model-free tracking become a challenging task.

Object feature representation methods mainly contain gray feature[7], geometric characteristics[8], subspace learning[9], sparse representation[9], color characteristics (e.g., camshift algorithm[10], meanshift algorithm[11]) and local binary pattern[12,13]. Learning approaches commonly include adaboost[14], neural networks[15], multiple instance learning[6] and structured output learning to predict object transformations[16]. Although model-free tracking has signifycantly improved in recent years, it's still very difficult to track multiple objects look similar at the same time. Zhang Lu et al.[17,18] successfully exploit such spatial constraints between objects in model-free trackers by developing a structure-preserving object tracker(**SPOT**) that incorporates spatial constraints between objects via a pictorial-structures framework (e.g., star model or the minimum spanning tree model) to avoid confusion between objects When all objects move in the same direction. Histogram-of-gradient (**HOG**) features[4] are sensitive to the spatial location of the object, in this paper we extract HOG features to represent the object. The search strategy is a sliding-window exhaustive search in the region of objects may occur. Dalal-Triggs detector[4] is capitalized to track objects. We train the individual object classifiers and the structural constraints jointly using an online structured **SVM**, which greatly improves the accuracy of multi-objects tracking. Due to the global search method consuming too much time, its real-time is poor. Taking into account the fact that sudden dramatic change in the object position is impossible in video sequence, this paper presents a search method based on motion estimation and structural constraints for multi-object tracking. In the first frame we select one object for its position, but its moving trend in the next frame is unknown, the search range of the object can only be roughly determined in the second frame, after finding the object, we combine the object's location information in two frames to calculate the motion vector. Because of the structural constraints between the objects, we can accurately determine the search range of the remaining objects by means of the known motion vector, and thus find all the objects. The motion vector and the structural constraints are constantly updated in the left frames. In summary, our main contribution are narrowing the search area, reducing search time and increasing search efficiency without affecting the accuracy of tracking.

II. MULTI-OBJECT TRACKING SEARCH ALGORITHMS IN THIS PAPER

The flow chart of multi-object tracking search algorithms in this paper is shown in Fig .1. We assume that the whole video frame number is F_N, n_i represents the object, and i is a positive integer. We choose n_1 to calcu-late the motion vector. A rough estimate of the search range of n_1 in the second frame can be got according to the position of n_1 in the first frame, then we find the position of n_1, combining the position of n_1 in the first frame, the motion vector of n_1 can be calculated. Because of the structural constraints among objects, the remaining objects $n_2, \ldots, n_i$ share the motion vector of n_1, thereby, their search area will be determined narrowly, which is useful for finding them successfully. Structural constraints up-date based on the location of objects in current frame. We apply the motion vector that is not updated and updated structural constraints to determine the motion trends and search range of all objects. After we find all of them, the motion vector and structural constraints will be updated for tracking in the next frame. From the third frame, the updated motion vector and structural constraints are utilized for tracking, the search algorithm will circulate until the last frame.

Owing to no extreme mutation in the position of objects in video sequence, we take the area around the object in previous frame as the current search scope, define search scope of each object below:

$$S_i = kS_{n_i} \qquad k \in R^+, i, n \in N^+ \tag{1}$$

Where S_{n_i} is object area of each target, S_i is search scope of each object, and S_i is a multiple of S_{n_i} . In the search process we use two points of a diagonal line in a rectangular box to represent the object n_i and its search scope in the video sequence, so (1) convert into coordinate calculation:

$$\{P_i, Q_i\} = kB_i = k\{X_i, w_i, h_i\} \tag{2}$$

where

$$S_i = \{P_i, Q_i\} \tag{3}$$

$$S_{n_i} = B_i = \{X_i, w_i, h_i\} \tag{4}$$

We represent the bounding box that indicates object n_i

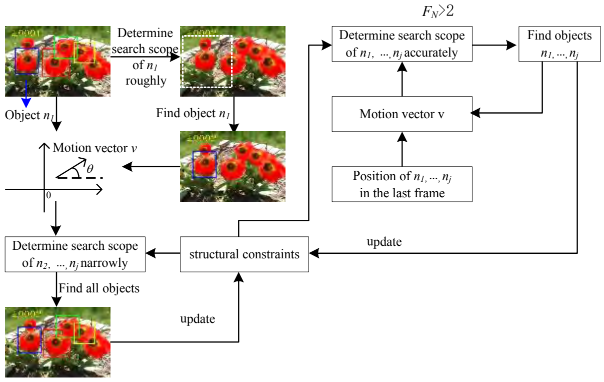

Figure 1.Flow chart of multi-object tracking search algorithms

by B_i={X_i,w_i,h_i} with center location X_i=(x_i,y_i), width w_i, and height h_i; both w_i and h_i are fixed. We choose two points to represent the search scope S_i={P_i,Q_i} of n_i , point $P_i = (x_{p_i}, y_{p_i})$ and point $Q_i = (x_{Q_i}, y_{Q_i})$ have a relationship with B_i. We put these parameters into (4):

$$\{x_{p_i}, y_{p_i}, x_{Q_i}, y_{Q_i}\} = k\{x_i, y_i, w_i, h_i\} \tag{5}$$

where

$$x_{p_i} = x_i - \eta h_i \tag{6}$$

$$y_{p_i} = y_i - \eta w_i \tag{7}$$

$$x_{Q_i} = x_i + (k - \eta) h_i \tag{8}$$

$$y_{Q_i} = y_i + (k - \eta) w_i \qquad 0 < \eta < k \tag{9}$$

k is decided by the object size and the object's proportion in the image frame, typically 4-8 times of the object. η is decided by the motion vector, the direction angle of the motion vector is continuous, considering the complexity of the algorithm, we generally use a limited direction to approximate it, in this paper, we select eight directions as the direction angle of the motion vector based on existing research results, each direction 45°, the angle of the motion vector is θ, then

$$\eta = \delta \arctan \theta \qquad \delta \in R \tag{10}$$

When we first determine the search range of the object in the second frame, due to the motion vector is unknown, a rough range around the object is determined with object as the center, then η take 1/2 of k. After finding the object, combining with the object position in the prior frame, the motion vector of the object can be calculated.

The minimum spanning tree model is constructed based on the objects marked in the first frame, it is obtained by searching the set of all possible completely-connected paths for the tree with minimum total, this pictorial-structures framework put multiple objects close together, as a entirety. With these two prerequisites (1) all objects move in the same direction, (2)they constitute a group, a conclusion is made that the motion vector of n_i are roughly the same. We calculate the range of all objects, finally find all tracking objects, and update structural constraints and the motion vector. We repeat the process above to find all the tracking objects until the last frame.

III. EXPERIMENTAL RESULTS AND ANALYSIS

Setup. This algorithm is implemented on MATLAB and Visual c + + platform, and tested on desktop computer with Intel processors (CoreTM i5-3470, 3.20 GHz). We made five groups contrast experiment to prove that the algorithm is effect, videos are respectively Red Flowers, Hunting, Parade， Students and Vehicle, the average length of the videos is 957 frames.

SPOT show the latest achievements of multi-object tracking algorithms based on structural constraints, it's also the most representative results, we evaluate the performance of the trackers by measuring (1) average pixel distance error (ALE) : the average pixel distance of the center of the identified bounding box to the center of the ground-truth bounding box, (2) tracking accuracy rate (CDR) : the average percentage of frames for which the overlap between the identified bounding box and the ground-truth bounding box is at least 50 percent to make a right tracking, we define CDR as the average percent-tage of frames that objects are tracking correctly and total frames of the video, and (3) time ratio (TR): the percentage of time that the algorithm cost to finish entire tracking in this paper and **SPOT**.

Results. Table 1 presents comparison of experimental data of search algorithm in this paper and **SPOT**, we believe that the following three conditions is excellent: (1) average pixel distance error (ALE) is as small as possible; (2) higher tracking accuracy rate (CDR) is better; (3) less time ratio (TR) is better. For each video, these three measurements are averaged over all objects, over all frames, and over five separate runs of the tracker. The first two items are the average data of all the objects in each video, the last one is the running time of tracking for the entire video.

Table I.Data comparison of SPOT and our method

	SPOT		Our Method		TR
	ALE	CDR	ALE	CDR	
Red Flowers	9.5	0.99	7.9	0.99	0.49
Hunting	19.4	0.87	17.8	0.87	0.20
Parade	9.2	0.68	4.9	0.68	0.61
Students	9.4	1.00	7.2	1.00	0.41
Vehicle	3.7	1.00	2.7	1.00	0.68

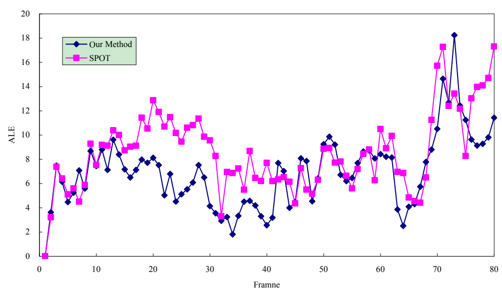

Figure 2. Comparison chart of Hunting between two algorithms

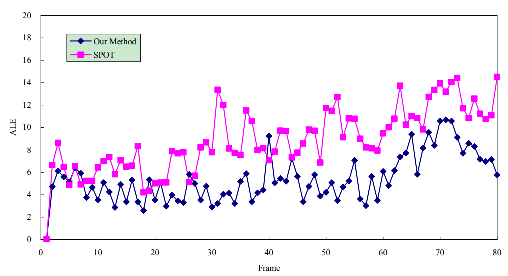

Figure 3. Comparison chart of Students between two algorithms

We select respectively 80 frames from video Hunting and Students to make comparison in position pixel differ-rence. Fig .2 is the comparison chart for Hunting between our algorithm and **SPOT**, and Fig .3 is for Students. From these two figures, a conclusion come out that our algorithm is superior to **SPOT** for its overall position pixel difference is less.

The pictures below is contrast of tracking effect between our algorithm and the exhaustive search algori-thm in **SPOT**, the first column of the pictures is for **SPOT**, and the second column is for algorithm in this paper.

Figure 4.Tracking results of Red Flowers between two algorithms

Fig .4 is the contrast tracking results of Red Flowers between algorithm in **SPOT** and our method, video Red Flowers is a multi-object tracking in complex environments. Objects account for large proportion in the back-ground, four tracking objects have similar appea-rance and look like non-tracking objects, object cross and occlusion problems exist in tracking process, and there is little change in their relative position. The data of Red Flowers from Table 1 show that (1) tracking accuracy is 0.99 and average pixel distance error is 9.5 in **SPOT**, (2) tracking accuracy is 0.99 and average pixel distance error is 7.9 in this paper, and (3) time ratio is 0.49. Tracking accuracy of two algorithms are the same, in addition, average pixel distance error of our algorithm is lower than **SPOT**. Fig .4 show that the objects in four selected frames can be fully tracked with both two algorithms, and the objects of our algorithm is closer to the center of the rectangle.

Fig .5 show us the results of Hunting between **SPOT** and our method, Hunting is an activity tracking in

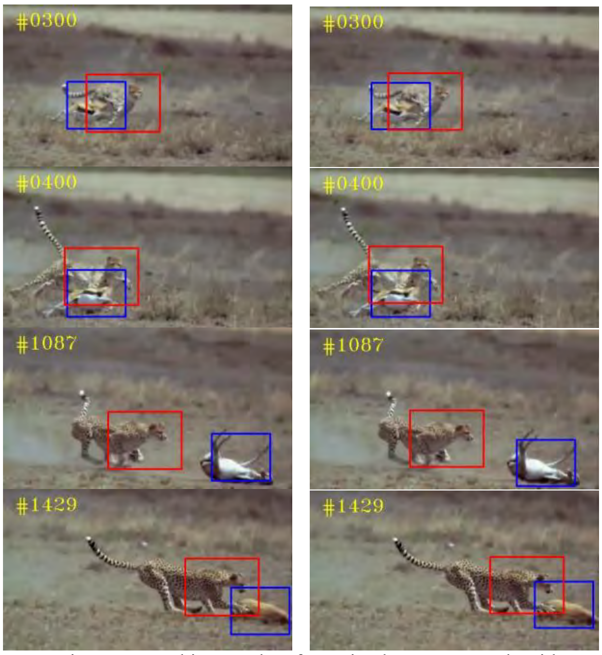

Figure 5. Tracking results of Hunting between two algorithms

simple environment. The difference between the objects and background is little, two tracking objects size vary widely, the cheetah's body account for a significant proportion, but gazelle is opposite. When Cheetah runs after gazelle, appearance and relative position of cheetah and gazelle change dramatically, gazelle also cause partial occlusion of the cheetah.The data of Hunting from Table 1 show that (1) tracking accuracy is 0.87 and average pixel distance error is 19.4 in **SPOT**, (2) tracking accuracy is 0.87 and average pixel distance error is 17.8 in this paper, and (3) the time ratio is 0.20. Tracking accuracy of two algorithms are the same, in addition, average pixel distance error of our algorithm is lower than **SPOT**. Figure 4 show the objects in four selected frames can be

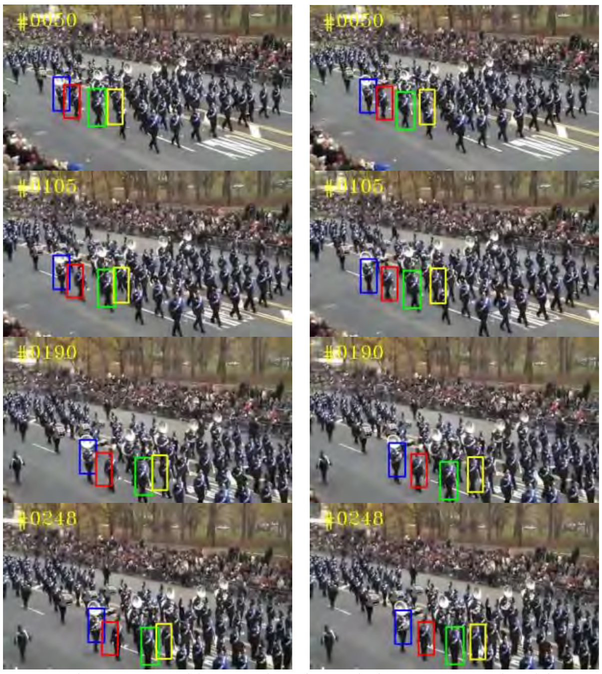
Figure 6. Tracking results of Parade between two algorithms

fully tracked with both algorithm in **SPOT** and in this paper, and the objects of our algorithm is closer to the center of the rectangle.

Fig .6 show us comparision of Parade between **SPOT** and our method, Parade is an orderly human acti-vity tracking in complex environment. Four objects are small, the tracking objects have a high similarity with non-tracking objects. There is only one object covered by non-tracking objects and position of four tracking objects remain unchanged during tracking. The data of Parade show that (1) tracking accuracy is 0.68 and average pixel distance error is 9.2 in **SPOT**, (2) tracking accuracy is 0.68 and average pixel distance error is 4.9 in this paper, and (3) time ratio is 0.61. Tracking accuracy of two algorithms are the same, in addition, average pixel distance error of our algorithm is lower than **SPOT**. Fig .6 show that the objects in four selected frames can be fully tracked with both two algorithms, and the objects of our algorithm is closer to the center of the rectangle.


Figure 7. Tracking results of Students between two algorithms

Fig .7 show us a comparision of Students between algorithm in **SPOT** and our method, video Students is an orderly human activity tracking in simple environment. Three objects are small, the tracking objects will gradually become smaller in the tracking process. There is one object partially covered obscured most of the time, as well as non-object enters its tracking range, causing interference. The data of Students from Table 1 show that (1) tracking accuracy is 1.00 and average pixel distance error is 9.4 in **SPOT**, (2) tracking accuracy is 1.00 and average pixel distance error is 7.2 in this paper, and (3) the time ratio is 0.41. Tracking accuracy of two algorithms are the same, in addition, average pixel distance error of our algorithm is lower than **SPOT**. Fig .4 show that the objects in four selected frames can be fully tracked with both algorithm in **SPOT** and in this paper, and the objects of our algorithm is closer to the center of the rectangle.

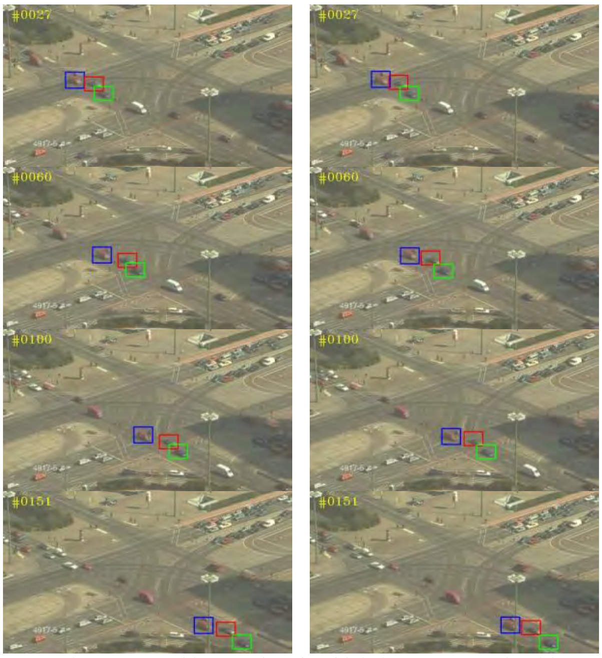

Figure 8. Tracking results of Vehicle between two algorithms

Fig .8 show us a comparision of Vehicle between algorithm in **SPOT** and our method, video Vehicle is vehicle tracking in simple environment. Three objects belongs to a small object tracking. Occlusion problem does not exist, and the relative position substantially unchanged. The data of Vehicle from Table 1 show that (1) tracking accuracy is 0.68 and average pixel distance error is 9.2 in **SPOT**, (2) tracking accuracy is 0.68 and average pixel distance error is 4.9 in this paper, and (3) the time ratio is 0.61. Tracking accuracy of two algorithms are the same, in addition, average pixel distance error of our algorithm is lower than **SPOT**. Fig .4 show that the objects in four selected frames can be fully tracked with both algorithms in **SPOT** and in this paper, and the objects of our algorithm is closer to the center of the rectangle.

In summary, compared to **SPOT**, the advantages of our method is that we introduce the motion vector, and taking into account the structural constraints to narrow the search range and reduce the error in the object matching process, improve search efficiency, greatly reduce the time-consuming, and enhance the real-time tracking. Meanwhile multi-object occlusion problem can be solved partially, and the precision of the tracking algorithm is improved greatly, which makes a great value.

IV. CONCLUSION

This paper proposes an object search algorithm based on motion estimation and structural constraint, this method incorporates the motion vector and the structure constraint between the objects to narrow the searching scope of the objects on the purpose of reducing the overall complexity of the tracking algorithm. The experimental results show that the algorithm can greatly decrease the time-consuming during the objects detection, under the condition of no effect on the tracking accuracy, so that the whole tracking efficiency will be improved, especially for small objects in large background. Even if the object is occluded, the reduction of the search scope will not lead to objects loss. As a result, the algorithm has great application prospect for intelligent transportation with vehicle tracking in the same direction.

ACKNOWLEDGMENT

The research work was supported by National Natural Science Foundation of China under Grant No. 61272236, Educational Commission of Hubei Province of China under Grant No. D20131306, China Three Gorges Universit ChuTian Scholar Fund under Grant No. KJ2012B001

REFERENCES

[1] Guo L, Song C, Mao Y. H infinity filter in maneuvering target tracking of military guidance field[J]. 2012.

[2] Huang Y H, Wang J L, Jia X M. Research of soccer robot target tracking algorithm based on improved camshift[J]. Advanced Materials Research, 2011, 221: 610-614.

[3] Kanhere N K, Birchfield S T. Real-time incremental segmentation and tracking of vehicles at low camera angles using stable features[J]. Intelligent Transportation Systems, IEEE Transactions on, 2008, 9(1): 148-160.

[4] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886-893.

[5] Ortego D, SanMiguel J C. Stationary foreground detection for video-surveillance based on foreground and motion history images[C]//Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on. IEEE, 2013: 75-80.

[6] B. Babenko, M.-H. Yang, and S. Belongie. Robust object tracking with online multiple instance learning. PAMI, 33(8): 1619–1632, 2011.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[7] Abdel Tawab A M, Abdelhalim M B, Habib S E D. Efficient multi-feature PSO for fast gray level object-tracking[J]. Applied Soft Computing, 2014, 14: 317-337.

[8] Tavanai A, Sridhar M, Gu F, et al. Carried object detection and tracking using geometric shape models and spatio-temporal consistency[M]//Computer Vision Systems. Springer Berlin Heidelberg, 2013: 223-233.

[9] Xie Y, Zhang W, Qu Y, et al. Discriminative subspace learning with sparse representation view-based model for robust visual tracking[J]. Pattern Recognition, 2014, 47(3): 1383-1394.

[10] Zhu L, Hu H. Research of Motion Tracking Based on CamShift Algorithm[J]. Applied Mechanics and Materials, 2013, 263: 2403-2407.

[11] Liu Q, Tang L B, Zhao B J. Algorithm of target tracking based on Mean Shift with adaptive tracking window[J]. Systems Engineering and Electronics, 2012, 34(2): 409-412.

[12] Yan W S. The Target Tracking Algorithm Based on Local Binary Pattern Texture Model[J]. Applied Mechanics and Materials, 2012, 197: 558-563.

[13] SONG X, WANG W, ZHANG W. Vehicle Detection and Tracking Based on the Local Binary Pattern Texture and Improved Camshift Operator[J]. Journal of Hunan University (Natural Sciences), 2013, 8: 009.

[14] Yang G, Liu H. Visual attention & multi-cue fusion based human motion tracking method[C]//Natural Computation (ICNC), 2010 Sixth International Conference on. IEEE, 2010, 4: 2044-2054.

[15] Zhao Q, Wang F, Sun Z. Using neural network technique in vision-based robot curve tracking[C]//Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on. IEEE, 2006: 3817-3822.

[16] Hare S, Saffari A, Torr P H S. Struck: Structured output tracking with kernels[C]//Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 263-270.

[17] [17] Zhang L, van der Maaten L. Structure preserving object tracking[C]//Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013: 1838-1845.

[18] [18] Zhang Lu, van der Maaten L. Preserving Structure in Model-Free Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(4): 756-769

[19] [19] S. Branson, P. Perona, and S. Belongie. Strong supervison from weak annotation: Interactive training of deformable part models. In ICCV, pages 1832–1839, 2011.