Research and Implementation of Vehicle Tracking Algorithm Based on Multi-Feature Fusion

Heng Guan[1(✉)], Huimin Liu[1], Minghe Yu[2], and Zhibin Zhao[1]

[1] School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China
guanheng@stumail.neu.edu.cn
[2] School of Software, Northeastern University, Shenyang 110819, China

Abstract. The main task of multi-object tracking is to associate targets in diverse images by detected information from each frame of a given image sequence. For the scenario of highway video surveillance, the equivalent research issue is vehicles tracking, which is necessary and fundamental for traffic statistics, abnormal events detection, traffic control et al. In this paper, a simplified and efficient multi-object tracking strategy is proposed. Based on the position and intersection-over-union (IOU) of the moving object, the color feature is derived, and unscented Kalman filter is involved to revise targets' positions. This innovative tracking method can effectively solve the problem of target occlusion and loss. The simplicity and efficiency make this algorithm applicable for the perspective of real-time system. In this paper, highway video recordings are explored as data repository for experiments. The results show that our method outperforms on the issue of vehicle tracking.

Keywords: Video surveillance · Multiple object tracking · Data association

1 Introduction

Multi-object tracking is an important research issue in computer vision. It has a wide range of applications, such as video surveillance, human-computer interaction, and driverless driving. The main task of multi-object tracking is to correlate the moving objects detected in the video sequence and plot the trajectory of each moving object, as shown in Fig. 1. These trajectories will contribute to abnormal event report, traffic control and other applications.

In the practical application of high-speed scenes, most of the surveillance cameras are installed on the roadside, so in the process of video surveillance, vehicles often occlude each other. When the target vehicle reappears in the video after occultation, it is likely recognized as new, which results in the loss of the original target. A simplified and efficient correlation algorithm is proposed in this paper. After calculating the correlation degree of each object in different images by its position, motion and color information, we match and merge those objects with a high degree as an identified target.

W. Ni et al. (Eds.): WISA 2019, LNCS 11817, pp. 72–78, 2019.
https://doi.org/10.1007/978-3-030-30952-7_8

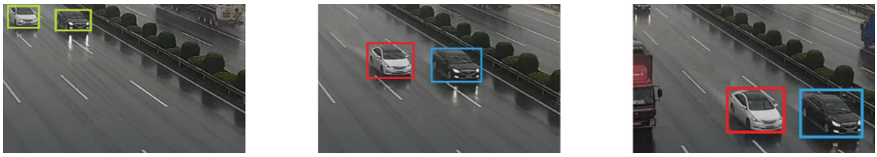

Fig. 1. Vehicle detection and tracking on a highway

The main contributions of this paper are listed below:

- A vehicle location prediction model is proposed to accurately predict the location when occluded.
- A data association method based on multi-feature fusion is proposed for data association for simplification and efficiency.
- Tracking algorithm is improved on real data sets as experiments.

2 Related Work

From RCNN [1] series to the SSD [2] and YOLO [3] series, many achievements appear in object detection. As YOLOv3 is an end-to-end object detection method, it runs very fast [4, 5].

For video analysis, multi-object tracking has been a research focus which consists of Detection-Free Tracking (DFT) and Detection-Based Tracking (DBT) [6]. Because the vehicle in the video is tracked continuously, DBT is considered in this paper.

Based on DBT, many researchers propose different solutions [7–9]. However, these methods do not involve image features, the results are not acceptable for occlusion.

CNN is also introduced for multi-object tracking [10–13]. Based on SORT [8, 14, 15] uses CNN to extract the image features of the tracking target. Although CNN provides more precision, it shows powerlessness in real-time computation. Because the performances on single object tracking in complex scenarios have been improved [16–18], [19] transforms the multi-object tracking problem into multiple single object tracking problems, and also achieves good performance.

For location revision, Kalman filter [20] can filter the noise and interference and optimize the target state. However, Kalman filter is only applicable to Gauss function and deals with linear model.

Based on the above works, this paper proposes a method on multi-object tracking by correlating the color and the location features. It predicts the location when the target is lost in an image, and then applies the unscented Kalman Filter for calibration.

3 Data Association

Given an image sequence, we employ object detection module to get n detected objects in the t-th frame as $S^t = \{S_1^t, S_2^t, \ldots, S_n^t\}$. The set of all the detected objects from time t_s to t_e is $S^{t_s:t_e} = \{S_1^{t_s:t_e}, S_2^{t_s:t_e}, \ldots, S_n^{t_s:t_e}\}$, where $S_i^{t_s:t_e} = \langle S_i^{t_s}, S_i^{t_s+1}, \ldots, S_i^{t_e} \rangle (i \in n)$ is the

trajectory of the object i-th from t_s to t_e. The purpose of multi-object tracking is to find the best sequence of states of all objects, which can be modeled by using MAP (maximum a posteriori) according to the conditional distribution of the states of all observed sequences, defined as:

$$\bar{S}^{t_s:t_e} = \arg\max_{S^{t_s:t_e}} \mathrm{P}(S^{t_s:t_e}|O^{t_s:t_e}) \tag{1}$$

For more accuracy, prediction on the object position according to the historical trajectory is necessary. Equations (2) and (3) show the prediction method.

$$S_i^t = \left\langle Loc_{S_i^t}, MotionInfo_{S_i^t}, Features_{S_i^t} \right\rangle \tag{2}$$

$$\begin{cases} Loc_{S_i^t} = \left[\left(x_{S_i^t}, y_{S_i^t} \right), \left((x+w)_{S_i^t}, (y+h)_{S_i^t} \right) \right] \\ MotionInfo_{S_i^t} = \left[\bar{v}_{S_i^t}, \bar{a}_{S_i^t}, k \right] \\ Features_{S_i^t} = \left[color_{S_i^t}, type_{S_i^t} \right] \end{cases} \tag{3}$$

$Loc_{S_i^t}$ is the location of i-th object in t-th frame, including upper left coordinates $\left(x_{S_i^t}, y_{S_i^t} \right)$ and lower right coordinates $\left((x+w)_{S_i^t}, (y+h)_{S_i^t} \right)$. $MotionInfo_{S_i^t}$ is motion information of i-th object in t-th frame, including average velocity $\bar{v}_{S_i^t}$, average acceleration $\bar{a}_{S_i^t}$ and motion direction k; $Features_{S_i^t}$ is characteristic information of i-th object in t-th frame, including color features $color_{S_i^t}$ and object class $type_{S_i^t}$.

Kalman filter is used to calibrate the position. However, Kalman filter algorithm is applicable to linear model and does not support multi-object tracking. Therefore, UKF is used for statistical linearization called nondestructive transformation. UKF first collects n points in the prior distribution, and uses linear regression for non-linear function of random variables for higher accuracy.

In this paper, three data are used for data matching – location information, IOU (Intersection over Union) and color feature. When calculating the position information of each S_i^{t-1} and S_i^t, we use the adjusted cosine similarity to calculate the position similarity of S_i^{t-1} and S_i^t. Formula for calculating position similarity $L_{confidence}$ between S_i^{t-1} and S_i^t is defined as (4).

$$L_{confidence} = \frac{\sum_{k=1}^{n} \left(Loc_{S_i^t} - \overline{Loc}_{S_i^t} \right)_k \left(Loc_{S_i^{t-1}} - \overline{Loc}_{S_i^{t-1}} \right)_k}{\sqrt{\sum_{k=1}^{n} \left(Loc_{S_i^t} - \overline{Loc}_{S_i^t} \right)_k^2} \sqrt{\sum_{k=1}^{n} \left(Loc_{S_i^{t-1}} - \overline{Loc}_{S_i^{t-1}} \right)_k^2}} \tag{4}$$

The object association confidence between S_i^{t-1} and S_i^t is shown as:

$$\text{Confidence}\left(S_i^{t-1}, S^t\right) = \left\{\text{c}_1\left(S_i^{t-1}, S_1^t\right), \text{c}_2\left(S_i^{t-1}, S_2^t\right), \ldots, \text{c}_n\left(S_i^{t-1}, S_n^t\right)\right\} \tag{5}$$

Thereby $Loc_{S_i^t}$ is:

$$Loc_{S_i^t} = Loc_{S^t}{}_{\text{maxIndex}\left(\text{Confidence}\left(S_i^{t-1}, S^t\right)\right)} \tag{6}$$

4 Experiments

The real high surveillance videos are used as the experimental data set, and three different scenes are intercepted: ordinary road section, frequent occlusion and congestion. Each video lasts about 2 min with a total of 11338 pictures annotated according to the MOT Challenge standard data set format.

The performance indicators of multi-object tracking show the accuracy in predicting the location and the consistency of the tracking algorithm in time. The evaluation indicators include: MOTA, combines false negatives, false positives and mismatch rate; MOTP, overlap between the estimated positions and the ground-truth averaged over the matches; MT, percentage of ground-truth trajectories which are covered by the tracker output for more than 80% of their length; ML, percentage of ground-truth trajectories which are covered by the tracker output for less than 20% of their length; IDS, times that a tracked trajectory changes its matched ground-truth identity [6].

Firstly, we implement the basic IOU algorithm to calculate the association of targets. The comparison is between IOU, SORT and IOU17. In the experiment, our method sets $T_{minhits}$ (shortest life length of the generated object) as 8, T_{maxdp} as 30; the same to SORT. IOU17 sets $T_{minhits}$ as 8, σ_{iou} as 0.5; the objective detecting accuracy is set as 0.5 and the results are shown in Tables 1, 2 and 3.

Table 1. Accuracy for ordinary section

Method	MOTA(↑)	MOTP(↑)	MT(↑)	ML(↓)	IDS(↓)
IOU	43.1	76.1	22.8%	30.7%	238
IOU+ position prediction	52.5	77.2	24.6%	26.8%	178
IOU+ color feature	53.2	76.7	23.2%	27.8%	184
IOU+ position prediction + color feature	55.9	78.6	26.6%	24.5%	147
SORT [8]	48.6	75.5	21.2%	29.4%	266
IOU17 [9]	51.4	76.8	23.1%	27.6%	192

It is illustrated from Tables 1, 2 and 3 that the method with predicted position and color features we proposed in this paper outperforms other methods. On frequently occlusion scenario, predicted position method greatly improves the accuracy. In congestion sections, color features are more helpful in accuracy than location prediction.

Table 2. Accuracy for frequently occlusion section

Method	MOTA(↑)	MOTP(↑)	MT(↑)	ML(↓)	IDS(↓)
IOU	32.1	73.3	13.1%	48.3%	482
IOU+ position prediction	39.7	76.6	16.3%	40.8%	314
IOU+ color feature	37.6	75.4	15.2%	42.5%	357
IOU+ position prediction + color feature	42.8	78.6	18.6%	38.7%	286
SORT [8]	34.1	76.0	14.6%	46.7%	477
IOU17 [9]	36.9	78.6	16.2%	41.3%	376

Table 3. Accuracy for congestion sections

Method	MOTA(↑)	MOTP(↑)	MT(↑)	ML(↓)	IDS(↓)
IOU	38.2	75.3	18.6%	36.3%	369
IOU+ position prediction	43.3	76.5	19.8%	32.4%	287
IOU+ color feature	45.5	77.3	21.6%	30.2%	245
IOU+ position prediction + color feature	47.9	79.3	23.2%	29.5%	212
SORT [8]	39.2	76.8	17.3%	34.2%	385
IOU17 [9]	42.8	77.8	19.6%	33.6%	302

The method in this paper does not predict the position in the case of amble because of the large deviation. However, the color feature doesn't change in amble. For amble, the color feature will make the accuracy significantly increase.

SORT [8] and IOU17 [9] only use the position information of the target to correlate the data over image feature and position prediction. In the case of frequent congestion and occlusion, the vehicle can't be tracked effectively because of the decrease of the accuracy of the object detection and the occlusion of the tracking target.

5 Conclusions

Target loss happens when occlusion and other events occur in a video surveillance. In this paper, we use linear regression to analyze the vehicle's historical trajectory, and then predict the position of the vehicle when it disappears in a video frame, and recognize the target when the vehicle appears again by the predicted position and color feature. Experiments show that the algorithm also shows it sufficiency for occlusion and congestion.

This method only uses color feature as extracted image feature. Although the computation is light, the deviation for color similarity of targets exits. For the future work, we will research shallow CNN to extract image features to enhance the performances in terms of efficiency and differentiation.

Acknowledgement. This research is supported by the National Key R&D Program of China under Grant No. 2018YFB1003404.

References

1. Girshick, R., Donahue, J., Darrelland, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, pp. 580–587. IEEE (2014)
2. Liu, W., et al.: SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016, Part I. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
3. Redmon, J., Divvala, S., Girshick, R., et al.: You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, pp. 779–788. IEEE (2016)
4. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Hawaii, pp. 6517–6525. IEEE (2017)
5. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. https://arxiv.org/abs/1804.02767. Accessed 15 Mar 2019
6. Luo, W., Xing, J., Milan, A., et al.: Multiple object tracking: a literature review. https://arxiv.org/abs/1409.7618. Accessed 15 Jan 2019
7. Jérôme, B., Fleuret, F., Engin, T., et al.: Multiple object tracking using K-shortest paths optimization. IEEE Trans. Pattern Anal. Mach. Intell. **33**(9), 1806–1819 (2011)
8. Bewley, A., Ge, Z., Ott, L., et al.: Simple online and realtime tracking. https://arxiv.org/abs/1602.00763. Accessed 18 Feb 2019
9. Bochinski, E., Eiselein, V., Sikora, T.: High-speed tracking-by-detection without using image information. In: IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017, Lecce, pp. 1–6. IEEE (2017)
10. Chu, Q., Ouyang, W., Li, H., et al.: Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism. In: IEEE International Conference on Computer Vision, ICCV 2017, Venice, pp. 4846–4855. IEEE (2017)
11. Leal, T., Laura, F.C.C., Schindler, K.: Learning by tracking: Siamese CNN for robust target association. In: Computer Vision and Pattern Recognition Conference Workshops, pp. 33–40 (2016)
12. Son, J., Baek, M., Cho, M., et al.: Multi-object tracking with quadruplet convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Hawaii, pp. 5620–5629. IEEE (2017)
13. Zhao, H., Xia, S., Zhao, J., Zhu, D., Yao, R., Niu, Q.: Pareto-based many-objective convolutional neural networks. In: Meng, X., Li, R., Wang, K., Niu, B., Wang, X., Zhao, G. (eds.) WISA 2018. LNCS, vol. 11242, pp. 3–14. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02934-0_1
14. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. https://arxiv.org/abs/1703.07402. Accessed 22 Feb 2019
15. Chu, Q., Ouyang, W., Li, H., et al.: Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism. In: IEEE International Conference on Computer Vision, ICCV 2017, Venice, pp. 4836–4845. IEEE (2017)
16. Bertinetto, L., Valmadre, J., Golodetz, S., et al.: Staple: complementary learners for real-time tracking. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, pp. 1401–1409. IEEE (2016)

17. Fan, H., Ling, H.: Parallel tracking and verifying: a framework for real-time and high accuracy visual tracking. In: IEEE International Conference on Computer Vision, ICCV 2017, Venice, pp. 5487–5495. IEEE (2017)
18. Li, Y., Zhu, J.: A scale adaptive Kernel correlation filter tracker with feature integration. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014, Part II. LNCS, vol. 8926, pp. 254–265. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16181-5_18
19. Chu, P., Fan, H., Tan, C.C., et al.: Online multi-object tracking with instance-aware tracker and dynamic model refreshment. In: IEEE Winter Conference on Applications of Computer Vision, WACV 2019, Hawaii, pp. 161–170. IEEE (2019)
20. Julier, S.J., Uhlmann, J.K.: New extension of the Kalman filter to nonlinear systems. In: Signal Processing, Sensor Fusion, and Target Recognition VI, vol. 3068, pp. 182–194. International Society for Optics and Photonics (1997)