PhD Forum: Tracking multi-object using tracklet and Faster R-CNN

Yosra DORAI
Institut Pascal,Université Clermont Auvergne, France
SAGE ENISo, University of Sousse, Tunisia
yosra.dorai@gmail.com

Sami GAZZAH
SAGE ENISo, University of Sousse, Tunisia
sami.gazzah@gmail.com

Frederic CHAUSSE
Institut Pascal, Université Clermont Auvergne, France
frederic.chausse@univ-bpclermont.fr

Najoua ESSOUKRI BEN AMARA
SAGE ENISo, University of Sousse, Tunisia
najoua.benamara@eniso.rnu.tn

ABSTRACT

In this work, we propose an original approach for tracking using tracklets (mini-trajectories) and Faster R-CNN. In fact, the major tracking problem is how an object can keep the same ID during the entire trajectory. To solve problems of tracking, we proceed on the one hand by the definition of a specific signature for each detected object to build tracklets. On the other hand, we associate these mini-trajectories in order to have complete trajectory.

Keywords

Tracklet, Tracking multi-object, Faster R-CNN

1. INTRODUCTION

The tracking of mono or multi-object in video is a challenge of computer vision. This domain presents various problems (e.g. occlusions, changes of poses, similarity of objects…). The tracking goal is that every object must keep the same and unique ID throughout its appearance in the scene.

Tracking begins by a detection stage. The detection is achieved by the classification of image descriptors (e.g. HOG or Haar [6][2][9]) in previously learned classes using a decision methodology such as SVM[7] or Adaboost[9]. Recently the literature has shown that neural networks and essentially Faster R-CNN has higher performances. Actually no descriptor is needed because the neural network training stage creates its own ones in its first convolutional layers. It determines a list of candidate objects.

In the tracking step, objects detected in the next images of a sequence are associated to the current list. Most of the works use the local, global or both associations [3][10]. Indeed, according to [4] the number of frames is important, in a global association, it can reach the totality of the video.

Beside, usually, the detector presents defects such as missing detection, false positives…In this work, we use tracklets to eliminate the false positives detections and to predict missed one. Tracking is based on detections using Faster R-CNN to build tracklets.

The paper is organized as follows. The description of our approach in section 2, followed by the experiments in section 3. Finally we conclude and we present the future of work in section 4.

ICDSC '16, September 12-15, 2016, Paris, France

DOI: http://dx.doi.org/10.1145/2967413.2974035

2. OVERVIEW OF THE PROPOSED APPROACH

Our approach is composed of two parts(detection and tracking) in order to obtain the entire trajectory of each object appearing on the scene with ID and signature (see fig 1). The first step is detection by Faster R-CNN. Then,in the step of tracking we construct tracklet. We first associate detection to have initial tracklets. In addition, we compare them with detection from the regressor function of Faster R-CNN. We also associate tracklet which have almost the same signatures. Finally, we update tracklets by adding unassociated detections.

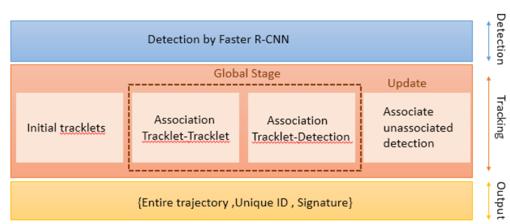

Figure 1: Block diagram of the proposed approach.

2.1 Step 1: Detection by Faster R-CNN

We use a Faster R-CNN, pre-trained initially with Imagenet, and we use the architecture VGG16[8]. It's composed by two modules: RPN (Region Proposal Network) and Fast R-CNN [5].

2.2 Step 2: Tracking using tracklet

A tracklet is a chain of nodes O^i representing one single object detected and appeared in N frames with the same ID i from start time t_s^i to final time t_f^i.

Each node O_t^i is one detection at time t ($t \in [t_s^i, t_f^i]$). It has a unique signature defined by features (e.g. localization, speed, size, appearance…). One object can have multiple chains in one scene because of missed detection or occlusion problems. Indeed when such problem occur the intial trackclet is interrupted. Then when the same object appears again later a new trakcklet is initiated instead of continuing the ancient one. So tracklets have be associated to solve occlusion problems and to predict missed objets.

Initial tracklets are build just after the detection step. First,we choose the number N of frames on which the tracklets are defined. Then we associate the detection from t_s to t_f. The association is done according to overlap and similarity of appearance between successive detections. After initial tracklets constructions, we can predict the following positions using Kalman filtering. In the global stage, we associate tracklets having similar signatures as well as detections provided by the system. Finally, in an update stage, we rectify tracklets to add unassociated detections. As output, we obtain the trajectory of each object with its unique ID and signature.

3. EXPERIMENTS

The evolution framework is with the private Logiraod Traffic dataset (car) and dataset PETS2009 [1](pedestrian)to experiment our approach. In the case of the private Logiraod Traffic dataset, we have successfully traced the trajectories cars despite their resemblance (see Fig. 2).

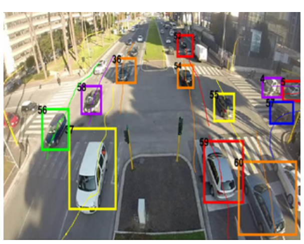

Figure 2: tracking result: Cars are correctly tracked and keep same ID dispite similarity.

For the PETS dataset,we use the S2L1 (walking) extracted from PET sequence S2 (People Tracking). The efficiency of our approach is observed in fig.3, since the two pedestrian (1 yellow and 2 green) have kept their ID from the frame 44 to frame 127, despite the presence of crossing problem

4. CONCLUSION AND FUTURE WORK

In this work, we have proposed a method of tracking multi-object by tracklet using Faster R-CNN in one camera. We have successfully find a way to exceed the occlusion problems and objects have kept the same ID as long as possible. In the future work, we will use a network camera to track the objects.

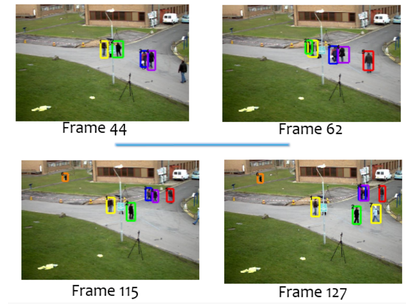

Figure 3: tracking result: IDs(1 yellow,2 green) are correctly kept despite crossing.

5. ACKNOWLEDGMENTS

This work is sponsored by a co-guardianship between the University of Sousse (Tunisia) and Blaise Pascal University (France).

6. REFERENCES

[1] http://www.cvg.reading.ac.uk/PETS2009/a.html.

[2] J. Badie and F. Bremond. Global tracker: an online evaluation framework to improve tracking quality. In *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*, pages 25–30. IEEE, 2014.

[3] S.-H. Bae and K.-J. Yoon. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1218–1225. IEEE, 2014.

[4] A. Dehghan, Y. Tian, P. H. Torr, and M. Shah. Target identity-aware network flow for online multiple target tracking. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1146–1154. IEEE, 2015.

[5] R. Girshick. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1440–1448, 2015.

[6] Y. Mao and Z. Yin. Training a scene-specific pedestrian detector using tracklets. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 170–176. IEEE, 2015.

[7] Y. Pang, Y. Yuan, X. Li, and J. Pan. Efficient hog human detection. *Signal Processing*, 91(4):773–781, 2011.

[8] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[9] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.

[10] Q. Yu and G. Medioni. Map-enhanced detection and tracking from a moving platform with local and global data association. In *Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on*, pages 3–3. IEEE, 2007.