Tracking of multiple objects based on the histogram of oriented gradients for pedestrian detection

Amari Dakhlia
LATIS- Laboratory of Advanced Technology and Intelligent Systems,
ENISo, Sousse University - Tunisia
amari_dakhlia@yahoo.com

Sami GAZZAH
LATIS- Laboratory of Advanced Technology and Intelligent Systems,
ENISo, Sousse University - Tunisia
sami.gazzah@gmail.com

***Abstract*—Thiswork presents our detection system and tracking pedestrian-based tracklets in a complex scene. The first task of our system is the detection of pedestrians using the HOG detector, this system is able to recognize all the pedestrians present in the scene. The second task is the tracking of all detected pedestrians to get complete trajectories for a long time while preserving the identity of each pedestrian, for this we used the notion of tracklets. The experiments show that our approach outperforms the detector to find the undetected objects and the developed method eliminates the false positives and shows the effectiveness of tracking.**

***Keywords*— Multi-objects tracking, Tracklets, HOG, Occlusion.**

I. INTRODUCTION:

Tracking is a computer vision It aims to provide over time the positions of mobile targets. Studying of this issue can be carried out by approaches which can be divided into two major categories: one for mono-object and one for multi-object.

In this work we focus on in multi-object tracking due to its importance for many computer vision applications. Multi-object tracking partially simple subject when objects are easily distinguished from each other and from the background. However, in complex scenes the production of complete trajectories remains a difficult issue due to the frequent occlusion by clutter or other objects, the similar appearance of the different objects and the variation of lighting. In this context, we propose a method that allows robust tracking of multiple objects under harsh conditions. Several multi-object tracking methods exist in the literature. But most of them used the detection-tracking method because they are more robust in complex environments. The basis of the tracking algorithms is to find the similarities between the detected objects. For this, some clues are assembled to calculate the resemblance like motion, location and appearance.

In this paper, we propose a multi-object tracking method based on tracklets. The tracklet is a partial trajectory extracted by a follower during a short period of time and therefore less subject to the drift and occlusion of a long path. In the first step, we use the HOG detector [Dalal et al., 2005] for pedestrian detection and the linear SVM classifier [Hassan et al., 2006] for the classification of objects. In the second step, we use a method [Bae et al., 2014] based on local and global association, in which the reliable tracklets generated in the local phase and the potential tracklets generated by the temporal sliding window are associated by the Hungarian algorithm in order to obtain the optimal association optimal overall.

The paper is organized as follows. In Section 2, we start with a presentation of the related work. Then, our approach is described in Section 3. Our experiments are detailed in Section 4. Finally, we conclude and we present our perspectives in Section 5.

II. RELATED WORK:

As mentioned in the introduction, we focused on the method of tracking by detection. By tracking, we mean the set of algorithms enabling the detection and the tracking of objects in a video sequence. We give a brief review of the main detectors et the main tracking methods.

Detectors: In the following we cite several important detectors. Quite a lot of researchers have used the pseudo-haar [Viola et Jones, 2001] for the detection of objects. As an example, [Shanshan and al., 2014] used the HAAR for pedestrian detection. Another approach was used for the detection of objects, which is based on the detector YOLO [Joseph et al., 2016]. Several studies [Mao et al., 2015] [Haythem et al., 2016] have shown that the histograms of oriented gradients (HOG) method are very responsive for the extraction of characteristics of the human form in particular.

Tracking: The multi-object tracking domain is very large. The tracking methods can be divided into two categories: online and offline. In the case of real-time applications, online approaches are more useful. In fact, they give results fairly comparable to those offline. It uses online information and based solely on past observations to build the complete trajectories of each detected object. In this context we focus on multi-object tracking based on tracklets. Several works considered this subject such as [Bing Wang et al., 2015]. The authors use two steps for the association of tracklets: a local step and a global step. [Junliang Xing et al., 2009] have used in the local stage; a specific filter with observer selection that could deal with partial object occlusion: filter is used to generate a set of reliable tracklets. In the global stage, the detection responses are collected from a temporal sliding window. Another paper [Qian Yu et al., 2007] used the local and global association, at the local level, the moving image blobs, provided from the motion detection, are associated into tracklets by a MCMC (Markov Chain Monte Carlo) Data Association algorithm. Both motion and appearance likelihood are considered when local data association is performed. Then, at the global level, trackletsare linked by their appearance and spatio-temporal consistence on the global map.

In this paper, we used the HOG descriptor and the linear SVM classifier for detection and classification. For the tracking we used the local and global association of tracklets.

III. APPROACH OVERVIEW:

The idea of our approach is inspired by [Bae and Yoon, 2014]. The authors proposed a method based on tracklet confidence, they used the local association to associate the tracklets with high confidence and detections provided online. For the tracklets with a low confidence and other tracklets and detections they used the global association. Therefore, our contribution is represented in the detection phase because Bae et al. have used predefined detection.

A. *STEP1:DETECTION BY HOG DESCRIPTOR*

For the detectionwe used the supervised learning method because the latter produces better performance for object detection.Such process consists of two phases, a first stage of learning or training which consists of grouping an image number of the positive and negative examples with their labels then it tries to construct a predictive system is able to determine the class of a new unknown sample in the online decision-making phase.

B. *STEP2: TRACKING USING TRACKLETS*

As mentioned in the previous section, the tracking phase is based on the method of Bae and Yoon with which we construct the complete trajectory of each detected pedestrian.

1. TRACKLETS:

A tracklet is a mini-trajectory that consists of a noued$O_i$chain that represents a single detected pedestrian that appears in N image with the same identity (ID) at the start timet_{is} and at the final time t_{if}. A reliable tracklet with high reliabilities based on: length, affinity and occlusion.

The tracklet confidence $conf(T^i)$ can be calculated on the basis of the following criteria:

$$conf(T^i) = (\frac{1}{L}\sum\nolimits_{k\in\left[t_s^i,t_f^i\right],v^{i(k)=1}} \Lambda(T^i, z_k^i)) * \max((1 + \beta.\log((L-w)/L)),0)$$

Where:

- $w$is the number of images during which an object i is missing due to occlusion ($w = t_f^i - t_s^i + 1 - L$).
- L is the length of a tracklet ($L + \left|T^i\right|$).
- $\beta$is a control parameter relying on the performance of a detector. When a detector shows high accuracy, βshould be set to a large value.

The tracklet confidence is based on two steps [Bae and Yoon] :

$$\hat{\mathrm{T}}_{1:t}^{\mathrm{MAP}} = \underset{\mathrm{T}_{1:t}}{\mathrm{argmax}} \iint \mathrm{P}\left(\mathrm{T}_{1:t}\middle|\mathrm{T}_{1:t}^{(\mathrm{hi})}, \mathrm{T}_{1:t}^{(\mathrm{lo})}\right) \times \underbrace{\mathrm{p}\left(\mathrm{T}_{1:t}^{(\mathrm{lo})}\middle|\mathrm{T}_{1:t}^{(\mathrm{hi})}, \mathrm{Z}_{1:t}\right)}_{\text{global phase}} \underbrace{\mathrm{p}\left(\mathrm{T}_{1:t}^{(\mathrm{hi})}\middle|\mathrm{Z}_{1:t}\right)}_{\text{local phase}} d\mathrm{T}_{1:t}^{(\mathrm{hi})} d\mathrm{T}_{1:t}^{(\mathrm{lo})}.$$

Here, $T_{1:t}^{hi}$ and $T_{1:t}^{lo}$represent a set of triplets with high confidence and a set of triplets with low confidence, respectively.$P$is the position of

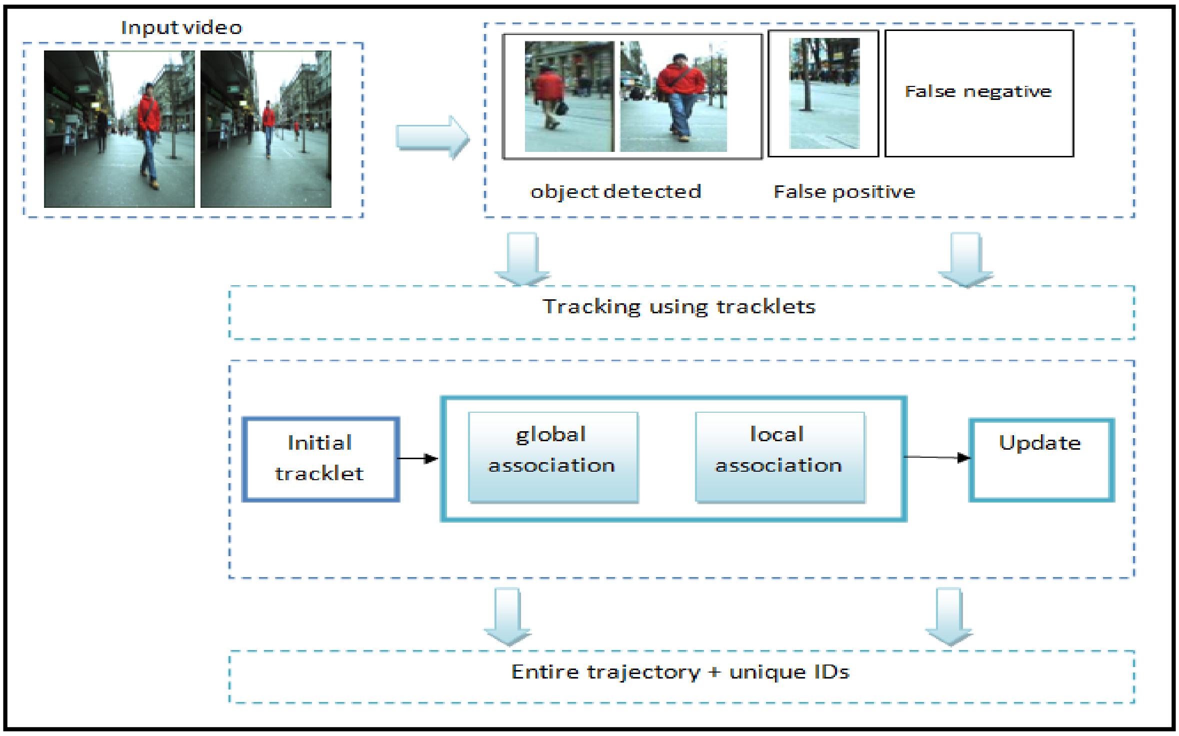

Figure.1: Block diagram of the proposed approach.

tracklet and $Z_{1:t}$is a set of all detections up to frame t.

As mentioned in the previous equation, the problem of association is solved in two phases: tracklets with high confidence and detections provided on-line are locally associated, while the tracklets with a low confidence are globally associated with other tracklets and detections.

2. LOCAL ASSOCIATION:

In the local the tracklets with high confidence $T^{i(hi)}$ are associated with a set of detections. When a h set of tracklets with a high confidence and n detections find in the image t, a score matrix S_{h*n} is computed [Bae et Yoon, 2014].

3. GLOBAL ASSOCIATION:

For the global association, tracklets with low confidence $T^{i(low)}$are globally associated with other tracklets and detections, these tracklets are likely to be fragmented [Baeet Yoon, 2014].

4. DISCRIMINATIVE APPEARANCE LEARNING:

The modeling of the appearance is very important both in the local and the global association to associate tracklets and detections of the same object distinguishing different objects, its aim is the collection of samples discriminating and learning

online using ILDA. The main cause of the use of ILDA is that the appearances of several objects can be distinguished with a single LDA projection matrix [Bae and Yoon, 2014].

IV. EXPERIMENTS:

In this section, we present the used metrics and the obtained results.

A. Metrics:

For the ETHMS (Sunny and Bahnhof) sequences, we use the CLEAR MOT metrics (MOTP, MOTA, precision, recall and false positives (FP)) in order to compare with related works. The MOTP metric is multiple-object tracking precision to evaluate the tracking results with bounding boxes of ground truth. The MOTA metric is the multiple-object tracking accuracy to measure the ID switch, the false positives and the false negatives.

C. DATA BASE:

For the test we used two databases: ETH_ Bahnhof which contains 400 images and 696 objects detected and in ETH_Sunny database, a static camera was used to record a set of contained 434 images and 655 objects detected.

D. Results:

Our result of detection by the HOG detector is illustrated in Figure.2.

Figure.2: Detection result

The use of our detection in the follow-up program provides an acceptable result in terms of false positives and the identity of each pedestrian (see Figure 3).

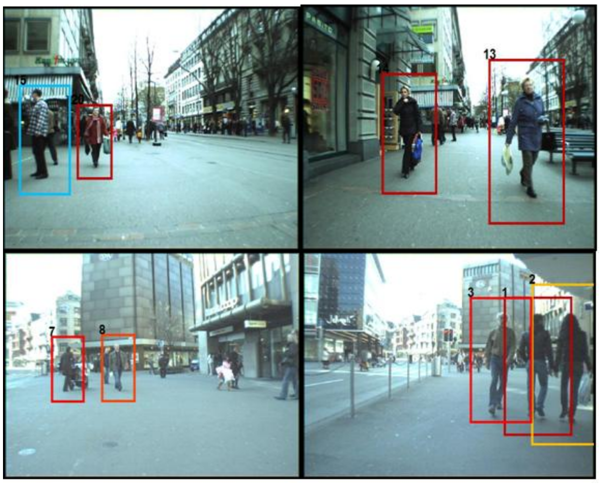

Figure.3: Tracking results with the database ETHMS (Sunny and Bahnhof).

Table 1 gives the results of our tracking method and a selection of the works in the state of the art. Our tracker is evaluated on the ETHMS (Bahnhof and Sunny) database. For this dataset, we also use the metric (MOTA, MOTP, presion, recall and false positives (FP)). Figure 3 illustrate the efficiency of our approach.

V. Conclusion:

In this paper we proposed a new tracking approach based on the tracklet with the use of HOG descriptor for the detection of pedestrians. Our method is evaluated on the public ETHMS database (Sunny ET Bahnhof). We obtained a system capable of reading different databases and detected the pedestrians present in each sequence.

As future work, we plan to improve the algorithm at the level of the local or global association. The study of other methods of pedestrian detection is more effective than the method used. We also plan to generalize our system to detect other objects (animals,cars...)

Base de données	Caméra	Méthode	Métriques	
			MOTA	MOTP
ETHMS (Sunny and Bahnof)	Caméra mobile	Bae and Yoon [33]	0.72	0.64
		Faster R-CNN[28] (Yosra DORAI 2017)	0.74	0.66
		Notre approche	0.71	0.62

Table 1 : CLEAR MOT metrics tracking results on the mono-camera database ETHMS (Sunny and Bahnhof).

VI. References:

[11] N. DALAL , et B. TRIGGS. "Histograms of oriented gradients for human detection". In: Computer Vision and Pattern Recognition, CVPR. IEEE Computer Society Conference on. IEEE, pages 886-893, 2005.

[1] M. HASAN, et F. BORIS. "SVM : Machines vectors de support our separators vastest merges". Rapport technique, pages 64, 2006.

[2] S-H. BAE, et K-J. YOO. "Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning". In : Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1218-1225, 2014.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features, in IEEE Proc. Of the conférence on Computer Vision and Pattern Recognition", pp. 511–518, 2001.

[4] ZHANG, Shanshan, BAUCKHAGE, Christian, et CREMERS, Armin B. "Informed haar-like features improve pedestrian detection". In : Proceedings of the IEEE conference on computer vision and pattern recognition. pages (947-954). 2014.

[5] REDMON, Joseph, DIVVALA, Santosh, GIRSHICK, Ross, et al. "You only look once: Unified, real-time object detection". In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pages (779-788). 2016.

[6] AMEUR, Haythem, HELALI, Abdelhamid, MAAREF, Hassen, et al. Improved vehicle detection system based on customized HOG. 2016.

[7] WANG, B., WANG, G., CHAN, K. L., et al. Tracklet Association by Online Target-Specific Metric Learning and Coherent Dynamics Estimation. 2015.

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

[9] X. Junling et al. ."Multi-Object Tracking through Occlusions by Local Tracklets Filtering and Global Tracklets Association with Detection Responses". 2009.

[10] Q. YU et G. Medioni. "Map-enhanced detection and tracking from a moving platform with local and global data association". Article, pages (3-3), 2007.

[11]Y. DORAI, F. CHAUSSE, S. GAZZAH, et al. "Multi Target Tracking by Linking Tracklets with a Convolutional Neural Network". In : VISIGRAPP (6: VISAPP). pages 492-498, 2017.