Online pedestrian tracking with multi-stage re-identification

Yi-Fan Jiang Hyunhak Shin Jaeyong Ju Hanseok Ko
School of Electrical Engineering, Korea University
Anam-dong, Seongbuk-gu, Seoul, Korea
{yfjiang, hhshin, jyju}@ispl.korea.ac.kr hsko@korea.ac.kr

Abstract

Nowadays the task of tracking pedestrians is often addressed within a tracking-by-detection framework, which in most cases entails that the position of each target has been detected before tracking begins. However in some cases, a pedestrian who is being tracked may be obscured by other targets or obstacles, and during this period they may change their trajectory or speed (track drift), and sometimes such a target may leave the FOV (Field of View) [10] but appear again later. These temporary disappearances and absence of detections disrupt the work of the detectors to such an extent that there is a significant decline in performance.

In this paper, we propose a novel approach to pedestrian tracking based on multi-stage re-identification. To deal with the problems discussed above, the proposed framework is comprised of a two-stage re-identification algorithm dealing with cases of track drift and re-entry into the FOV individually, in order to match the identities of lost and re-appeared targets through a comparison of the affinities between their appearance, size and position, and also to update the status of re-identified targets through this assessment. The experimental results demonstrate that this framework can effectively handle complex temporary lost and re-entry situations with robustness, and that its performance is state of the art.

1. Introduction

Multi-pedestrian tracking is the operation of estimating the current status of multiple pedestrians, including the identification of their trajectories based on their appearances and changes in position over time [13]. Recently it has been widely carried out through the use of several kinds of practical technologies such as cctv in public areas, driverless vehicles, unmanned aerial vehicles focused on intelligence-gathering, and so on [13, 12, 24, 26, 25]. However, due to the variety and complexity of the locations in which multi-pedestrian tracking is carried out, it is often challenging to maintain a high level of performance in the face of the frequent occlusion of targets and their unpredictable re-appearances.

Due to the recent rapid development of object detectors, tracking-by-detection methods [3, 4, 18] have increasingly arisen. With these methods, there is an additional target detection stage before tracking begins, during which targets are individually detected on a frame-by-frame basis by an object detector. Then these methods generally build multiple trajectories by making associations between the detections provided by the detector. Such tracking-by-detection approaches can be roughly categorized into batch and online methods.

Batch approaches [14, 7, 27, 28, 16, 15, 19, 8] generally use all detections from video footage in its entirety to build multiple target trajectories called tracklets. To be more specific, short tracklets can be generated by linking the detections of adjacent frames, and then longer tracklets are built with associated global detections. This approach often displays superior performance than online approaches due to its use of the future information of frames since it is an offline method. This is especially clear when applied to complex occlusion scenes [14, 27]. However, the limitation of the global information that determines this approach is difficult to overcome in real-time applications.

Online methods [6, 23, 3, 21, 22, 20] can be applied to real-time applications because they only perform detection association among past frames and the current frame, and do not make use of offline information. On the other hand, because online methods rely to a significant degree on detection quality, partly obscured or inaccurate detections can severely disrupt tracking procedures, produce fragmented trajectories and cause track drift problems. However, recently many advanced approaches based on online methods have been adopted. For example, Ju *et al.* [11] handles the track drift problem through multi-stage data association, but it gives no consideration to the problem of target re-appearance. Solera et al. *et al.* [22] incorporates some selective features to improve tracking performance but performs badly in addressing the track drift problem, and it also gives no consideration to the target re-

978-1-5386-2939-0/17/$31.00 ©2017 IEEE

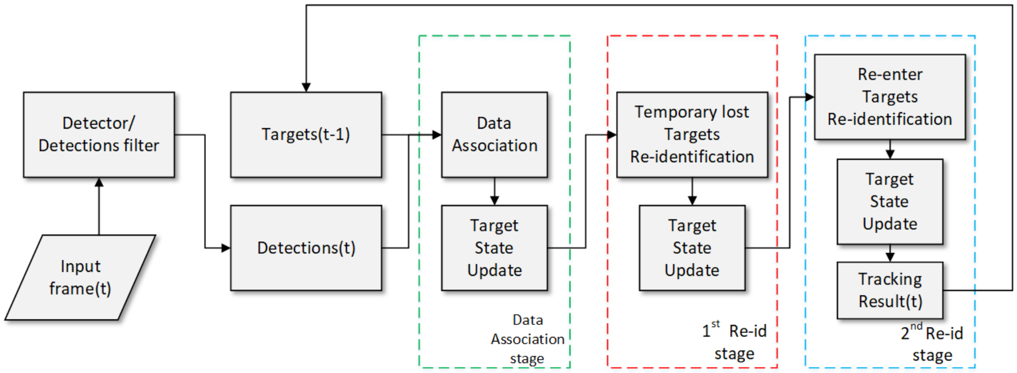


Figure 1. Structure diagram of proposed method.

appearance problem. Therefore, these online approaches are still not adequately addressing the track drift problem, which is caused by pedestrians changing their trajectories or pace while being obscured for lengthy periods, and then re-appearing [23]. Another problem it is necessary to address is that pedestrians exit the FOV but then re-enter it [6, 3], during which time data association can be difficult to perform. Our proposed method is an enhancement of the approach of Ju *et al.* [11], as we add the new affinity term of size and also improve the algorithm to make data association more reliable. The performance of the track correction stage is improved through the first re-identification stage of our proposed method. Furthermore, an additional stage of re-identification has been incorporated in order to handle the target re-entry problem discussed above.

To tackle the various problems encountered by existing methods, we propose a pedestrian tracking framework with two-stage re-identification, which focuses on dealing with track drift and re-identification in cases of target re-entry. To handle the track drift problem, we carry out the first stage of re-identification based on the establishment of associations between re-appearance detections and lost targets contained in past frames, through the use of a target appearance model and target size information. The second stage of re-identification is used to associate re-appearance detections with targets who had exited the FOV, and this association is determined according to a target appearance model.

2. Proposed method

2.1. Overview

The structure diagram of our proposed method is shown in Figure 1. Input video is transmitted to the detector frame by frame in order to capture detections of individual pedestrians. Then output detections are sent through a detection filter, which makes assessments based on the size of the tracks and the overlap between detections and tracks, because a large number of false positive detections caused by objects such as poles and dustbins tend to be generated by the detectors. The filtered detections are associated with the tracks contained in the last frame and the state of the tracks are updated. Then the data association stage takes place, with the goal of matching the detections from the current frame with the targets in the previous frame. Next, the two-stage re-identification process is carried out in order to deal with temporarily lost targets and their re-entry. In the first stage, we find a match between temporarily lost tracks and their re-appearance detections through analysis of appearance and shape affinities. In the second stage, tracks which exited the FOV are matched with unmatched detections from previous stage through the use of only shape and appearance affinities. A greedy algorithm [6] is used to resolve the problem of assigning an identity to a target. After concluding this process, tracking results for the current frame are generated and are exploited in the subsequent frame.

2.2. Data association stage

Before the data association stage commences, data preparation is carried out. If object detection i is detected in frame t, we denote it as d_t^i, then group such detections as D_t. There are two groups of detections D_t^{u1} and D_t^{u2}, which leads to $D_t^{u2} \subset D_t^{u1} \subset D_t$, where D_t^{u1} is the group of unmatched detections produced by the data association stage and D_t^{u2} is the group of unmatched detections produced by the first re-identification stage. Then the track i of frame t is denoted as v_t^i, and grouped as V_t.

Given targets produced by the previous frame V_{t-1}, in order to associate them with the detections of the current frame D_t, we use the product result of the three criteria of appearance, shape and position to represent the level of affinity of the target-detection pairs of ith target and jth detection:

$$\begin{aligned}\mathcal{A}_{ds}(v_{t-1}^i, d_t^j) = \mathcal{A}_a(v_{t-1}^i, d_t^j)\mathcal{A}_s(v_{t-1}^i, d_t^j)\\ \mathcal{A}_p(v_{t-1}^i, d_t^j)\end{aligned} \quad (1)$$

where $\mathcal{A}_a$, $\mathcal{A}_s$ and $\mathcal{A}_p$ are the appearance affinity, size affinity and position affinity respectively. Each of them round from 0 to 1, the higher number representing a higher affinity between the target-detection pair. After we assess the affinity of the input pairs, a greedy algorithm [6] is used to conclude the assignment process.

To calculate the level of appearance affinity, we used a template matching-based method [11]. The development of the template was based on a 24 bin RGI histogram which extracted data from each patch of detection. Each patch was resized 30 × 70. It includes a template update mechanism which revises the latest target template on the basis of the matched detection template after every data association stage. Then the old template is inserted into a historical template pool and the oldest template is discarded (there is a limit of 30 historical templates that are retained). Appearance affinity of ith target and jth detection is defined below:

$$\mathcal{A}_a(v^i_{t-1}, d^j_t) = \begin{cases} A_a(v^i_{t-1}, d^j_t) & if\, A_a(v^i_{t-1}, d^j_t) > \tau_a \\ 0, & otherwise \end{cases} \tag{2}$$

$$\begin{aligned} A_a(v^i_{t-1}, d^j_t) &= \rho_L \cdot f(\mathcal{T}^{d^j}, \mathcal{T}^{v^i}_L) \\ +(1-\rho_L) &\max_{\mathcal{T}^{v^i}_{H(k)} \in \mathcal{H}^{v^i}} f(\mathcal{T}^{d^j}, \mathcal{T}^{v^i}_{H(k)}) \end{aligned} \tag{3}$$

where $\mathcal{T}^{d^j}$ is the template of detection d^j and $\mathcal{T}^{v^i}_L$ is the latest template of target v^i, $\mathcal{T}^{v^i}_{H(k)}$ where $k \in [1, N_h]$ is the historical template pool of v^j. Here we use the Bhattacharyya coefficient $f(\cdot,\cdot)$ to calculate the affinity of two template. Finally, ρ_L is the weight of latest target affinity. τ_a is a threshold selected in order to contain the appearance affinity within a limited range.

The size affinity is define as follows:

$$\mathcal{A}_s(v^i_{t-1}, d^j_t) = \exp\{-\{\frac{|h^{v^i}_{t-1} - h^{d^j}_t|}{h^{v^i}_{t-1}} + \frac{|w^{v^i}_{t-1} - w^{d^j}_t|}{w^{v^i}_{t-1}}\}\} \tag{4}$$

where h and w are the height and width of detections and targets respectively.

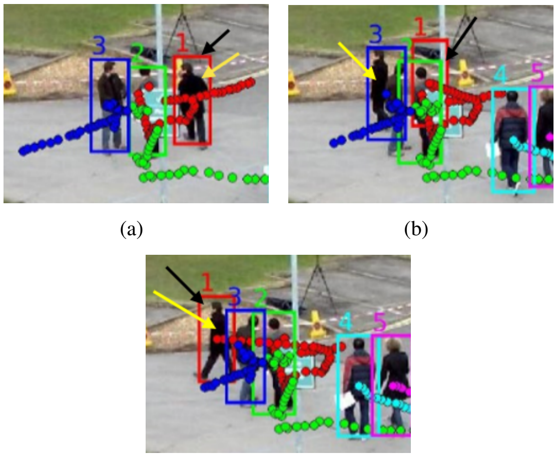
The position affinity is defined as follows:

$$\mathcal{A}_p(v^i_{t-1}, d^j_t) = \begin{cases} A_p(v^i_{t-1}, d^j_t) & if\, A_p(v^i_{t-1}, d^j_t) > \tau_p \\ 0, & otherwise \end{cases} \tag{5}$$

$$\begin{aligned} A_p(v^i_{t-1}, d^j_t) = n_p \exp\{-(\mathbf{p}^{v^i}_{t-1} + \mathbf{s}^{v^i}_{t-1} - \mathbf{p}^{d^j}_t)^T \\ \mathbf{C_p}^{-1}(\mathbf{p}^{v^i}_{t-1} + \mathbf{s}^{v^i}_{t-1} - \mathbf{p}^{d^j}_t)\} \end{aligned} \tag{6}$$

where $\mathbf{p}^{v^j}_t = (x^{v^i}_t, y^{v^i}_t)$ and $\mathbf{s}^{v^i}_t = (s^{x,v^i}_t, s^{y,v^i}_t)$ represent the target's position and speed. $\mathbf{C_p}$ is the covariance of position and n_p is normalising factor.

After the data association stage, the status of the target is updated according to the assignment result of the greedy algorithm. If there are still some un-associated tracks, they are inserted into set V^{u1}_t.



(a) (b)

(c)

Figure 2. Example of first re-identification stage in sequence PETS'09 S2L1 (the black arrow indicates the predict result, the yellow arrow indicates the true position of the target): (a) In the 45th frame target No.1 (red bounding box) is turning to top-left, but from the subsequent frame until the 61st frame, the target is occluded. (b) In the 61st frame, the target re-appears, but during the occlusion period our algorithm could only predict the target status by its velocity in the 45th frame since no detection could take place during the following frames. Therefore, there was a disparity between the real trajectory and the predicted trajectory. (c) In the 62nd frame, the target was re-identified and its historical status was corrected with equation (9).

2.3. First re-identification stage

With this stage of analysis, the method aims to deal with the target drift problem when the target is temporarily lost. This kind of problem is usually caused by the target's sudden change in motion when it is occluded for a long period. To handle this, if a target which is continuously maintaining un-associated status after the data association stage for T_c frames, we denote it as an 'uncertain target'(here we use $T_c = 6$). If there are uncertain targets and unassociated detections exist after the data association stage, the first re-identification algorithm, based on appearance and size affinity, carries out the re-assignment of identification to them. The basis on which the first re-identification algorithm determines affinity can be calculated as follows:

$$\begin{aligned} &\mathcal{A}_{r1}(v^{u1,i}_t, d^{u1,j}_t) = \\ &\begin{cases} A_a(v^{u1,i}_t, d^{u1,j}_t) A_s(v^{u1,i}_t, d^{u1,j}_t) & if\ dist(v^{u1,i}_t, d^{u1,j}_t) \\ & \le R^{v^{u1,i}}_t \\ 0, & otherwise \end{cases} \end{aligned} \tag{7}$$

$$R^{v^{u1,i}}_t = \rho w^{v^{u1,i}}_t \min(n^{v^{u1,i}}_t, n_{max}) \tag{8}$$

where $dist(\cdot,\cdot)$ is the distance between target and detection. R limits the re-identification activation range according to the uncertain degree of a target. We used the number of continuously un-associated frame of a target $n_t^{v^{u1,i}}$ to represent the uncertain degree. Here $n_t^{v^{u1,i}}$ increase only when $n_t^{v^{u1,i}} > T_c$. $w_t^{v^{u1,i}}$ is the width of target and ρ, n_{max} are positive constants. The re-identification activation range has a positive correlation to the degree of uncertainty of a target, it also takes into consideration the fact that a target is larger when it is closer to a camera. After the re-identification between $v_{t'}^{u1,i}$ and $d_{t'}^{u1,i}$ is concluded, the historical state of $v_{t'}^{u1,i}$ is determined through a calculation of its average speed:

$$\mathbf{s}_m^{v^{u1,i}}(new) = \frac{\mathbf{p}_t^{d^{u1,i}} - \mathbf{p}_{t'}^{v^{u1,i}}}{t - t'},\ k \in [t^{'} + 1, t] \tag{9}$$

after average speed is determined we can also predict the position $\mathbf{p}_m^{v^{u1,i}}(new)$. An example of first re-identification stage is shown in Figure 2.

When the first re-identification stage is concluded, the remaining unassociated detections are inserted into detection set D^{u2}, in preparation for the second re-identification stage. If a target maintains unassociated status continuously for more than 20 frames, it is inserted into target set V^{u2}. This means it has already exited the FOV, and the target's status is updated with a Kalman filter. In addition, if a target has exited the FOV, it is inserted into V^{u2}.

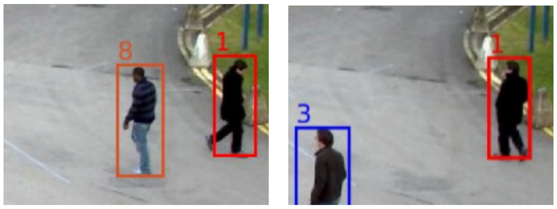
2.4. Second re-identification stage

In second re-identification stage, we focus on dealing with the pedestrian re-enter problem. Sometime pedestrians exit the FOV but re-appear later. In this situation, it is necessary to use a re-identification algorithm to provide the same identities to targets and to update their old target status. To re-identify the re-enter targets we use only appearance affinity to form between targets from V^{u2} and detections from D^{u2}:

$$\mathcal{A}_{r2}(v_t^{u2,i}, d_t^{u2,j}) = \begin{cases} A_a(v_t^{u2,i}, d_t^{u2,j}) & if\ \mathbf{p}_t^{d^{u2,j}} \in boundary\ region \\ 0, & otherwise \end{cases} \tag{10}$$

$$boundary\ region = \{\rho_w w_t^{d^{u2,j}} < x < w_{FOV} - \rho_w w_t^{d^{u2,j}}\}, \{\rho_h h_t^{d^{u2,j}} < y < h_{FOV} - \rho_h h_t^{d^{u2,j}}\} \tag{11}$$

where we only consider the detections which are located within the boundary of region $\rho_w = \rho_h = 5$ in our experiment. An example of the second re-identification stage is shown in Figure 3. After the second re-identification stage, if a target has been re-identified successfully, it is moved from V^{u2} to the current target set. If there are still some unassociated targets, they are kept in the V^{u2} set and their status is updated with blank. The remaining unassociated detections in D^{u2} are removed.



(a) (b)

Figure 3. An example of the second re-identification stage in sequence PETS'09'S2L1 : (a) In the 145th frame, target No.1 is going to leave the FOV. (2) In 225 frame, target No.1 re-appears in the FOV, so re-identification algorithm recovers its identity and update its status.

After this stage is concluded, we obtain and save all tracking results of the current frame, which we later use to make associations with detections in frame $t + 1$.

3. Experimental results

3.1. Dataset and Detections

For performance evaluation, we used following datasets: PETS 2009(PETS) [2] and 2D MOT 2015(MOT) [1]. Two sequences in PETS dataset were used, S2L1 and S2L2. The Town-Centre sequence in MOT dataset was also used. The use of these three sequences enabled a comparison with the results obtained using other state-of-the-art methods.

The detection results of the S2L1, S2L2 and Town-Centre sequences were obtained using an ACF detector [9], and for the remaining MOT sequences we used publicly available detections which can be accessed on the MOT challenge website.

3.2. Evaluation metrics

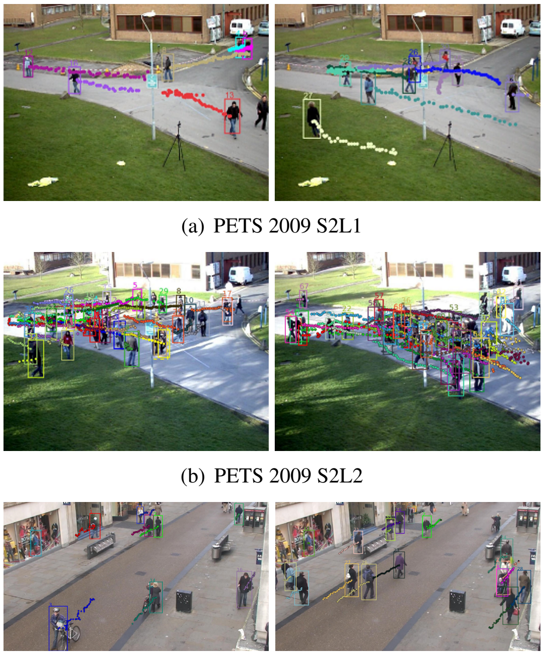
We used CLEAR MOT metrics [5] to evaluate the performance of our method. There are multiple metrics: multiple object tracking precision (MOTP↑) evaluates the alignment between the annotated and the predicted bounding boxes, multiple object tracking accuracy(MOTA↑) together false positives, missed targets and identity switches three factors to calculate the accuracy, the radio of mostly trajectories (MT↑), the ratio of mostly lost trajectories(ML↓) and identity switching(IDs↓). In here ↑ and ↓ mean higher number is better and lower number is better respectively.

3.3. Performance evaluation

The evaluation results of our proposed method are displayed in Figure 4. The results show that the method can effectively perform tracking in challenging locations. In a

Table 1. Performance comparison of our proposed method, the baseline method and other state-of-the-art methods; the baseline method is marked with an asterisk *, while the best performance evaluation metric is in bold.

Dataset(Sequence)	Method	On/Offline	MOTP(%)	MOTA(%)	MT(%)	ML(%)	IDS
PETS(S2L1)	Proposed	Online	75.8	**91.0**	95.1	**0**	**5**
	Ju *et al.*[11]*	Online	75.3	90.4	94.7	**0**	11
	Milan *et al.*[16]	Offline	**76.1**	87.4	**100**	**0**	31
	Milan *et al.*[15]	Offline	74.3	90.3	**100**	**0**	26
	Zamir *et al.*[19]	Offline	69.0	90.3	-	-	11
PETS(S2L2)	Proposed	Online	**73.3**	56.4	**48.9**	**0**	**197**
	Ju *et al.*[11]*	Online	71.8	**56.8**	39.6	**0**	230
	Solera *et al.*[22]	Online	70.8	47.4	14.0	7.0	297
	Pirsiavash *et al.*[17]	Offline	64.1	45.0	39.5	16.3	201
MOT(Town-Centre)	Proposed	Online	**74.7**	**66.4**	55.4	**1.8**	255
	Ju *et al.*[11]*	Online	**74.7**	66.3	**56.1**	8.7	260
	Sanchez-Matilla *et al.*[20]	Online	68.5	17.3	4.9	49.1	201
	Dicle *et al.*[8]	Offline	70.0	15.0	2.2	58.4	**76**


(a) PETS 2009 S2L1

(b) PETS 2009 S2L2

(c) Town-Centre

Figure 4. Example frames from PETS 2009 and MOT datasets.

medium-density crowd sequence, the temporarily lost target problem can be dealt with robustly, and the second re-identification stage works well when faced with cases of target re-entry. However some drawbacks of our method are evident, including a high dependence on the quality of detection, which leads to relatively poor performance during high-density crowd sequences, since no stable detection can be performed in each frame.

In Table 1, our proposed method is quantitatively compared with other state-of-the-art methods. Because our proposed method is an enhancement of Ju *et al.* [11], we set it as a baseline method in order to evaluate performance improvement. As can be seen in the evaluation results of the S2L1 and S2L2 sequences in the PETS dataset, our proposed method displays better performance in terms of most metrics compared to the baseline method. In particular, in terms of MT and IDS metrics, we use additional affinity categories and add a re-entry re-identification stage to deal with the target re-entry problem, with the aim of increasing accuracy and minimizing identity switches. In addition, compared to other state-of-the-art methods using MOTP and MOTA metrics, our proposed method displays relatively good performance. For the Town-Centre sequence of the MOT dataset, almost no target re-entry case occurs, so our method does not perform as well according to the IDS metric as it does in terms of the two sequences above. Although its performance in terms of IDS seems relatively good, its performance in terms of other accuracy metrics is more clearly an improvement over other methods.

4. Conclusion

We have described an online pedestrian tracking framework based on multi-stage re-identification in order to deal with two common tracking problems: (*i*) the track drift problem caused by target motion abruptly changing during a long period of occlusion; (*ii*) the target re-identification problem caused by the target leaving the FOV but re-appearing later. Our experiments have demonstrated our proposed method's high level of effectiveness and robustness compared to other state-of-the-art approaches.

Acknowledgements: This work was supported by the ONRG grant: N62909-16-1-2185. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2017R1A2B4012720). This subject is supported by Korea Ministry of Environment (MOE) as Public Technology Program based on Environmental Policy.

References

[1] 2D MOT 2015 Dataset. `https://motchallenge.net/data/2D_MOT_2015/`.

[2] PETS 2009 Benchmark Data. `http://www.cvg.reading.ac.uk/PETS2009/a.html`.

[3] S.-H. Bae and K.-J. Yoon. 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.

[4] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. *IEEE transactions on pattern analysis and machine intelligence*, 33(9):1806–1819, 2011.

[5] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008(1):1–10, 2008.

[6] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Online multiperson tracking-by-detection from a single, uncalibrated camera. *IEEE transactions on pattern analysis and machine intelligence*, 33(9):1820–1833, 2011.

[7] W. Brendel, M. Amer, and S. Todorovic. Multiobject tracking as maximum weight independent set. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1273–1280. IEEE, 2011.

[8] C. Dicle, O. I. Camps, and M. Sznaier. The way they move: Tracking multiple targets with similar appearance. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2304–2311, 2013.

[9] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1532–1545, 2014.

[10] L. Esterle, J. Simonjan, G. Nebehay, R. Pflugfelder, G. F. Domínguez, and B. Rinner. Self-aware object tracking in multi-camera networks. In *Self-aware Computing Systems*, pages 261–277. Springer, 2016.

[11] J. Ju, D. Kim, B. Ku, D. K. Han, and H. Ko. Online multi-object tracking based on hierarchical association framework. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–42, 2016.

[12] L. Kratz and K. Nishino. Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes. *IEEE transactions on pattern analysis and machine intelligence*, 34(5):987–1002, 2012.

[13] O. Masoud and N. P. Papanikolopoulos. A novel method for tracking and counting pedestrians in real-time using a single camera. *IEEE transactions on vehicular technology*, 50(5):1267–1278, 2001.

[14] A. Milan, S. Roth, and K. Schindler. Continuous energy minimization for multitarget tracking. *IEEE transactions on pattern analysis and machine intelligence*, 36(1):58–72, 2014.

[15] A. Milan, K. Schindler, and S. Roth. Detection-and trajectory-level exclusion in multiple object tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3682–3689, 2013.

[16] A. Milan, K. Schindler, and S. Roth. Multi-target tracking by discrete-continuous energy minimization. *IEEE transactions on pattern analysis and machine intelligence*, 38(10):2054–2068, 2016.

[17] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1201–1208. IEEE, 2011.

[18] H. Possegger, T. Mauthner, P. M. Roth, and H. Bischof. Occlusion geodesics for online multi-object tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1306–1313, 2014.

[19] A. Roshan Zamir, A. Dehghan, and M. Shah. Gmcp-tracker: Global multi-object tracking using generalized minimum clique graphs. *Computer Vision–ECCV 2012*, pages 343–356, 2012.

[20] R. Sanchez-Matilla, F. Poiesi, and A. Cavallaro. Online multi-target tracking with strong and weak detections. In *European Conference on Computer Vision*, pages 84–99. Springer, 2016.

[21] G. Shu, A. Dehghan, O. Oreifej, E. Hand, and M. Shah. Part-based multiple-person tracking with partial occlusion handling. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1815–1821. IEEE, 2012.

[22] F. Solera, S. Calderara, and R. Cucchiara. Learning to divide and conquer for online multi-target tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4373–4381, 2015.

[23] X. Song, J. Cui, H. Zha, and H. Zhao. Vision-based multiple interacting targets tracking via on-line supervised learning. In *European Conference on Computer Vision*, pages 642–655. Springer, 2008.

[24] X. Song, X. Shao, Q. Zhang, R. Shibasaki, H. Zhao, and H. Zha. A novel dynamic model for multiple pedestrians tracking in extremely crowded scenarios. *Information Fusion*, 14(3):301–310, 2013.

[25] S. Vishwakarma and A. Agrawal. A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, 29(10):983–1009, 2013.

[26] X. Wang. Intelligent multi-camera video surveillance: A review. *Pattern recognition letters*, 34(1):3–19, 2013.

[27] B. Yang and R. Nevatia. Multi-target tracking by online learning of non-linear motion patterns and robust appearance models. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1918–1925. IEEE, 2012.

[28] B. Yang and R. Nevatia. An online learned crf model for multi-target tracking. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2034–2041. IEEE, 2012.