

مجمع کنفرانس ملی دانشجویی
مهندسی برق ایران



Object tracking using particle filters and deep convolutional network

Ali safari*

Department of Computer Engineering, Faculty of Engineering, Yazd University, Yazd, Iran
ali.safari@stu.yazd.ac.ir

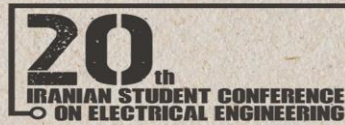
Ali bashiri*

Department of Computer Engineering, Faculty of Engineering, Yazd University, Yazd, Iran
alibashiri@stu.yazd.ac.ir

Abstract

There is a useful method for quick and efficient tracking of multiple objects called simple online and real-time tracking (SORT). By adding visual information, SORT algorithm performance can be improved. The number of identity switches can be minimized by this. A deep network that is offline on a wide data set of qualified pedestrians has been used since the main structure of the algorithm has a lot of computational complexity. In order to extract more and higher quality visual information that can assist the object recognition algorithm, the focus of this article is on the design of this deep network. To enhance data association efficiency, the paper also used a particle filter instead of a Kalman filter. On two standard datasets, MOT16 and MOT17, we checked our proposed method and compared its performance with other available methods. The results indicate that, relative to the current methods in this area, the tracking accuracy (52.2) on the MOT17 dataset is increased. Experimental assessment demonstrates that in dynamic settings, our proposed architecture increases the number of identity switches and preferably tracks goals.

Keywords: Computer Vision, Multiple Object Tracking, Detection, Data Association, Particle filter



مجمع کنفرانس ملی دانشجویی
مهندسی برق ایران



Introduction

In the accuracy and efficiency of object detectors, especially ordinary detectors, recent developments have inspired and provoked multi-target tracking methods by noticing them [1, 2]. These techniques operate in such a way that, by means of a first-rate object detector, they detect the targets for each frame and these detections are compatible with the application of online and offline trackers [4, 5]. These methods are acceptable in cases such as pedestrian tracking, where the goal feature is discriminatory and the target shows a simple motion pattern [6,5]. There are trackers, too, that rely less on appearance than others. In order to get used to the algorithms in these potential situations, they also need a large number of variables and skills to be tuned [8, 9,11].

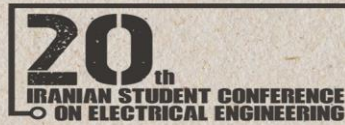
Tracking-by detection has recently evolved into the most important model of compound object tracking, as described earlier, in object recognition. In this way, in a global optimization problem, we usually find object trajectories, which advance the entire batches of video immediately. Flow network formulations [5, 14] and probabilistic graphical models [15, 8, 17] are the most common frameworks of this kind. These methods may not be acceptable in online circumstances in which the target should be addressed individually, step by step. For each frame base, the Joint Probabilistic Data Association Filter (JPDAF)[19] and Multiple Hypothesis Tracking (MHT)[18] present the data association. Such approaches are among the common and traditional techniques that exist in this area.

Since the speed of a large number of accurate trackers is considered too slow for real-time applications, the fairly obvious trade-off between accuracy and speed is evident. Simple online and real-time tracking (SORT) [20] is one of the simple frameworks that aims to improve both speed and precision. By applying the Hungarian approach with an association metric that tests bounding box overlap, which is Kalman filtering, the output of this method is in image space and data relation for each frame. At high frame rates, this undemanding technique achieves positive results.

SORT carries out quite a large number of identity transformations. The explanation for this is that when state evaluation uncertainty is sluggish, the employed association metric is accurate. As a result, SORT has a lack of blockage tracking since they are usually visible in the front section of camera views. We alter the place of the association metric with a more informed metric that incorporates motion and shape data as a solution to this problem. In comparison to deep sorting architectures, we used the deep neural network to integrate motion and appearance information [21]. We also substituted a particle filter for the Kalman filter. The value of the algorithm for particle filters is its simplicity and flexibility. Also, working with the model of the Gaussian multimodality system is undemanding. A great deal of related literature is given in [42]. Unlike measurement supplies, facts and data can be used in a particle filter system, which has greatly enhanced the efficiency of tracking. By a combination of this network, we build toughness against misses and occlusions at the same time as we keep the framework easy to implement and well-organized and applicable to online scenarios. The following sections are arranged in this paper: Section 2 provides a brief evaluation of related literature in the field of multiple object tracking. Until a demonstration of the expected framework functionality on standard benchmark series in Section 4 is completed, Section 3 proposes our approach and deep network architecture. Finally, Section 5 provides us with a summary of the findings learned and addresses future developments.

Related Works

Multi-object tracking has been overcome generally and conventionally by the Joint Probabilistic Data Association (JPDA) filters [22,23] and Multiple Hypothesis Tracking (MHT) [19]. Despite the fact that there is considerable uncertainty in object assignments, they postpone making complex decisions. The number of tracked objects raises the complexity of integrating these techniques, making them inefficient for simultaneous purposes in highly active environments. In visual multiple object tracking (MOT), the JPDA formulation [22] is considered by Reza tofighi et al [23]. His aim was to resolve the



مجمع کنفرانس ملی دانشجویی
مهندسی برق ایران



problems of combinational complexity in the resolution of integer programs. Kim et al. [24] applied the exterior model for every target in order to reduce the MHT chart. she wanted to accomplish state-of-the-art presentation. Such methods, however, put off the conclusion that online tracking is unacceptable.

Typical multi-target trackings are known as a network flow problem [5] or its variation [14]. Most of these methods rely on a large number of object shapes and assume that their goals are basic motion models. They work well in environments and pedestrians or means of transportation are the goals. One of the disadvantages of implementing the formulation of a network flow is that it has to plan starting and finishing positions and/or goal times that might be difficult to define in the future. Brendel et al. [6], while considering only two-frame relations, applied an extremely serious set formulation. Zamir et al [26] propose the implementation of a General Maximum Clique partitioning formulation. This technique selects the largest candidate from each tracklet with the goal of achieving global association. The formulation of the linear assignment [11,14,15] is comparable with the generalized linear assignment (GLA). The positive thing about it, however, is that it allows tracks to start and end anywhere in time and place.

The method of MOT solver is suggested by Braso et al [3]. This approach is based on networks exchanging messages. To present both function learning and final result estimation, it utilizes the natural graph problem structure. They propose a novel time-aware neural message passing update stage that is inspired by classic graph MOT formulations.

The strategy for several tracking techniques is to shape online education models of both objects [27,29] and the global model [30,33]. Motion is also incorporated to assist related tracklet detection [34,30], as well as patterns of appearance. When one-to-one correspondence models are considered, uniformly ideal solutions such as the Hungarian algorithm [35] can be applied [37].

The system of Geiger et al. [37] applies the Hungarian algorithm [35] in a couple of steps process. The earliest stage is when tracklets are formed with connecting detections from corner to corner of closest frames, where geometry and type signals are exchanged to shape the similarity matrix. Subsequently, by using geometry and appearance signals, the tracklets are linked to another to reconnect broken trajectories caused by occlusion. The union technique of these pair measures restricts this approach to batch computation.

To develop SORT efficiency, the Nicolai Wojke et al. [21] method combines appearance information. Throughout longer phases of occlusions, they will follow targets and successfully drop the amount of identity switched down induced by this Extension. However, as discussed in the following section, we use new deep network architectures for appearance descriptors.

Methodology

This strategy is expressed through the methodology of standard assumption tracking with recursive particle filtering and data link for each frame. In the next sections, the various parts of this system and our anticipated technique will be developed.

Detection and state estimation

The results of identification in data association based on MOT have a strong influence on the presentation of tracking. The original formulation in [22] is accompanied by both the Detection and Kalman filtering framework.

Instead of using the Kalman filter, we used the Particle filter to track targets according to the concept set out in [43], and we were inspired by their concept. In the paper, a target tracking algorithm is designed that integrates the particle filter and convolution network. In the particle filter system, the extracted function from convolutional networks is represented. The target local and spatial data are fully applied in order to indicate the state shift of the entity. The local shape alteration and partial occlusion problem of the target are best figured out as the global information sections of the particle

filter are integrated in order to find out the condition of the current targets. This is based on the target state that is dealt with various details.

We assume a very common tracking situation in which the camera is uncalibrated also there is no ego-motion data in hand, which is the most popular setup measured in the latest multiple object tracking benchmarks [38], as stated in [21]. The linear steady fast model, which does not depend on more objects or camera movement, estimates the inter-frame movements for each object. Every target's location is designed as:

$$x = [u, v, \gamma, h, \dot{x}, \dot{y}, \dot{\gamma}, \dot{h}] \quad (1)$$

Where μ and v indicate the horizontal and vertical pixel position of the target center, the aspect ratio γ , height h , and their respective velocities in the image are coordinated. In order to renew the target location, the detected bounding box is applied, where the velocity mechanism is resolved in the best and most favorable way by using a particle filter frame when an object is discovered. Basically, by applying the linear velocity form, the position is predicted without modification if no recognition is connected to the target. For each track in this method, we count the frame numbers from the prior effective measurement association. This counter is increased and reorganized to 0 during the prediction of the particle filter when the track has been linked with a measurement. If the tracks are over the pre-arranged maximum age, we take them out of the track collection. For any detection that is not capable of being linked with an access track, a new track hypothesis is arranged. Throughout their initial three structures, these up-to-the-minute tracks are structured as uncertain. During each time phase, we look forward to a flourishing measurement association.

Data Association

Each goal bounding box geometry is anticipated in the recent framework by forecasting its new status while our intention is to define detections for achievable targets. The cost matrix task will be measured as the intersection-over-union (IOU) space between of recognition and the entire bounding boxes predicted from the objects presented. The assignment is solved optimally by using the Hungarian algorithm. We combine details of motion and appearance by arranging two appropriate metrics for this formulation of the issue.

To formulate a problem, as we discussed earlier, we need to merge motion and appearance data. Between predicted particle states, we use the Mahalanobis space plus recently arrived evaluations. Camera motions that are not counted could lead to rapid disarticulations in the image surface, especially by making the Mahalanobis distance another unconscious metric for tracking through occlusions. It is therefore so much easier to combine the second metric into the assignment problem. For each bounding box detection, a shape descriptor r_j with $\|r_j\| = 1$ is determined. Gallery $R_k = \{r_k^{(i)}\}_{i=1}^{L_k}$ of the last $L_k = 100$ descriptors of the connected form are kept for each k track. After that, the smallest cosine space between the i -th track and j -th identification in shape space is the following metric procedure:

$$d^{(2)}(i, j) = \min \{1 - r_j^T r_k^{(i)} | r_k^{(i)} \in R_i\} \quad (2)$$

The Mahalanobis distance offers details about possible positions of objects based on short-term predictions of motion that are especially realistic. Whereas, when movement is less discriminatory, the cosine space considers type data that are fundamentally realistic in the direction of enhancing identities later than long-term occlusions.

Appearance Feature

The space between shape characteristics is used to compute the significance of similarity in data association. For comparable identity individuals, the importance of affinity must be outsized and broad and undersized for individuals with different identities, based on the optimal and desirable appearance characteristics. The appearance function in our operation is extracted by using a network that is presented in Figure 1 and Table 1 in our network architecture.

On the basis of the results and experiences obtained, it can be said that in deep learning, the deeper the network, the better the network's accuracy, given that there are no problems with the vanishing gradient. Therefore, we tried to improve the precision of detection in the proposed architecture by deepening the grid and also using residual layers to prevent the gradient from disappearing. On MARS data collection, this method has been achieved and contains more than 1,100,000 images of 1,261 pedestrians, which makes this entirely appropriate for deep metric learning in the situation of people monitoring.

Table 1. Overview of the CNN architecture

Name	Patch Size/Stride	Output Size
Conv 1	7×7/1	32×128×64
Conv 2	5×5/1	32×128×64
Conv 3	3×3/1	32×128×64
Batch and normalization		32×128×64
Max Pool 4	3×3/2	32×64×32
Residual 5	3×3/1	32×64×32
Residual 6	3×3/1	32×64×32
Residual 7	3×3/2	64×32×16
Residual 8	3×3/1	64×32×16
Residual 9	3×3/2	128×16×8
Residual 10	3×3/1	128×16×8
Residual 11	3×3/2	256×8×4
Residual 12	3×3/1	256×8×4
Conv 13	3×3/1	256×8×4
Conv 14	1×1/1	256×8×4
Batch and normalization		256×8×4
Avg Pool 15	3×3/2	256×4×2
Dense 15		256
Dense 16		256
Batch and normalization		256

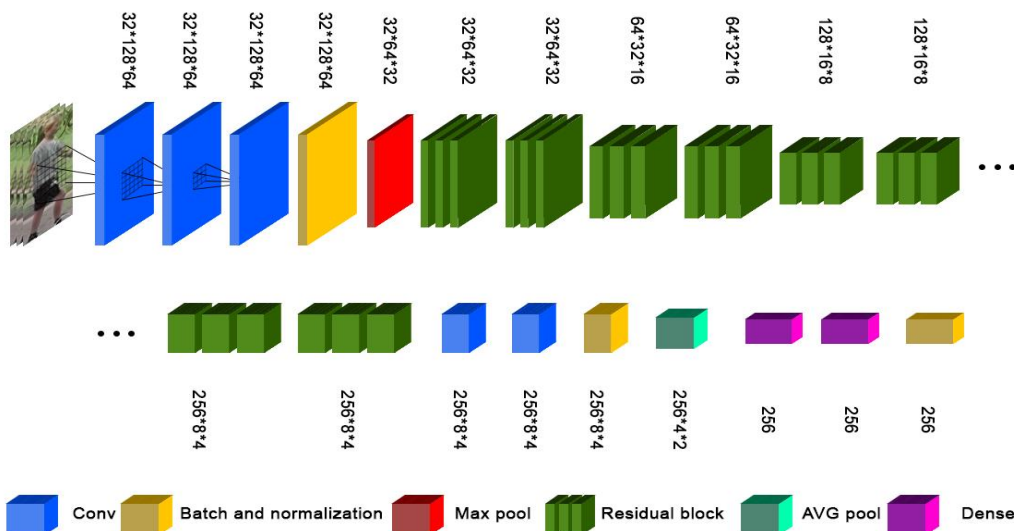
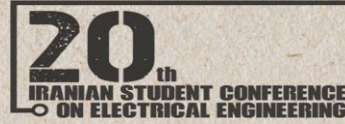
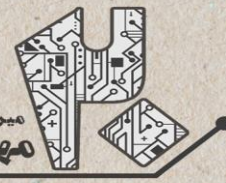


Figure 1. Overview of the CNN architecture



مجمع کنفرانس ملی دانشجویی
مهندسی برق ایران



Usually, our network is made of two elements. With phase 1 and similar padding in the first section, four convolutional layers are added. We have also applied batch normalization between every sheet. Batch normalization reduced the quantity through the movement of the concealed unit values (covariance shift). Batch normalization helps every network layer to learn that it is also more independent of other layers. In the next part, we added the eight remaining blocks of different sizes, one after the other. We increase the depth of the network to increase accuracy and minimize the amount of changes in individuality. We applied two compact layers of 250 as a final point in order to estimate the global feature map. And, right after adding a batch normalization layer, the final network output is achieved.

It is important to learn decent mappings for unintentional initialization of weights from input to output in neural networks. There are some local minimums that may be trapped by back propagation since there is a large search space that includes several weights in training. On the other hand, the randomization function of weight initialization must be selected and carefully defined unless there is a great risk that the production of preparation has decreased to the point of uselessness.

The Adam optimizer algorithm with a $1e-3$ rate of learning was used in the learning process of this network. In a learning process of 200,000 iterations, the latest outcome of our proposed system, shown in Table 2 and Table 3, is obtained.

TABLE 2. Performance of the proposed approach on MOT16 benchmark sequences

Method	MOTA	IDF1	MT	ML	FP	FN	ID _{SW}
GCRA[3]	48.2	48.6	12.9%	41.1%	5104	88586	821
oICF [7]	43.2	49.3	11.3%	48.5%	6651	96515	381
MOTDT[10]	47.6	50.9	15.2%	38.3%	9253	85431	792
LMP[12]	48.8	51.3	18.2%	40.1%	6654	86245	481
NOMT[31]	46.4	53.3	18.3%	41.4%	9753	87565	359
MCjoint[13]	47.1	52.3	20.4%	46.9%	6703	89368	370
DMMOT[16]	46.1	54.8	17.4%	42.7%	7909	89874	532
Deep SORT[21]	61.4	-	32.8%	18.2%	12852	56668	781
Our method	57.9	49.7	29.9%	20.5%	5850	70985	330

TABLE 3. Performance of the proposed approach on MOT17 benchmark sequences

Method	MOTA	IDF1	MT	ML	FP	FN	ID _{SW}
MHT_DAM[24]	50.7	47.2	20.8%	36.9%	22875	252889	2314
FWT [33]	51.3	47.6	21.4%	35.2%	24101	247921	2648
HAM_SADF17[41]	48.3	51.1	17.1%	41.7%	20967	269038	1871
EDMT17[32]	50.0	51.3	21.6%	36.3%	32279	247297	2264
MOTDT17[10]	50.9	52.7	17.5%	35.7%	24069	250768	2474
jCC[32]	51.2	54.5	20.9%	37.0%	25937	247822	1802
DMAN[16]	48.2	55.7	19.3%	38.3%	26218	263608	2194
TNT[25]	51.9	58.0	23.5%	35.5%	37311	231658	2294
Our method	52.2	56.1	21.3%	37.1%	26857	237594	1774

We applied the Xavier or variance scaling approach for weight initialization. Compared with the immature weight scaling technique, the Xavier weight initialization technique is a great development. This technique helped us to increase the pace of the field of deep learning in a great way. It thus adapts itself according to the quantity of the weight values. The principle of these methods is that if you can keep the variance constant layer by layer in either the feed-forward or back-propagation direction, your network can learn optimally. Your weight will ultimately saturate your non-linear neurons in

both positive and negative directions as you go through the layers, as if the variance boosts or decreases. This initialization was established to work better with the activation functions of ReLU in general, because in this network we apply the activation function of ReLU:

$$\text{var}(w_i) = \frac{2}{n_{in}} \quad (3)$$

W represents the weights in the above formula and n represents the number of inputs for each node. This network essentially consists of 4,654,764 parameters and one forward pass of 32 bounding boxes that take on Nvidia Titan XP for around 19 ms. So, this network is suitable for online tracking if a modern GPU is available.

Experiments

In order to teach and evaluate our tracking performance, which includes either moving or static camera sequences, we use MOT16 and MOT17 datasets. In addition to top-down observation arrangements, this benchmark estimates the tracking output on seven challenging test sequences that involve front-view scenes with a movable camera. Together with the MOT metrics [41], we use the assessment metrics defined in [40]:

MOTA(↑): Multi-object tracking accuracy.

IDF1(↑): the ratio of correctly identified detections over the average number of ground-truth and computed detections.

MT(↑): the amount of mostly tracked trajectories. i.e. target has a similar label for at least 80% of its life span.

ML(↓): the amount of mostly lost trajectories. i.e. target is not tracked for at least 20% of its life span.

FP(↓): the amount of false detections.

FN(↓): the amount of missed detections.

ID_{sw}(↓): the amount of times ID switches to a dissimilar formerly tracked purpose.

Estimation procedures with (↑), upper scores signify improved performance; while for evaluation procedures with (↓), minor scores denote better performance.



Figure 2: In an ordinary tracking situation with normal occlusion, our system representative performance on the MOT challenge data collection.

In order to apply the MOT benchmark [38] test server, tracking output is calculated where the ground precision is suspended for 11 sequences. In Table 2 and Table 3, several other baseline trackers are assessed with the suggested SORT process. This strategy has effectively eased the number of identity switches.

We understand that a small increase is primarily the number of tracked objects and the decrease in usually missing objects. Generally, due to the combination of appearance details, we effectively

maintain identities through longer occlusions. We can see them by monitoring qualitative performance analysis that we include in additional content. In Figure 2, an excellent tracker performance is shown. This technique is a great opponent of other tracking frameworks online. While we maintain competitive MOTA grades, track fragmentation, and false negatives, the smallest number of online method identity switches are returned by our method. In Figure 3, with other available methods on the MOT17 benchmark, you can see the accuracy of the proposed process.



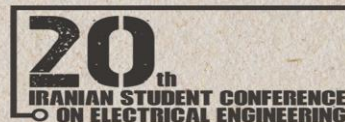
Figure 3: Quantitative comparison plots for the MOT17 benchmark [38] for multi-object tracking accuracy (MOTA).

Conclusions

We suggest promising tracking methods by applying the Particle filter and deep convolutional network. In order to extract effective features for good and stable tracking, we use deep learning methods. The algorithm can efficiently address appearance changes and occlusion problems in a strict manner. The improved approach is superior to conventional tracking approaches in extreme tracking conditions on the basis of experimental effects and has slightly decreased numbers of identity changes compared to the deep sorting system.

References

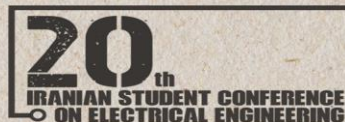
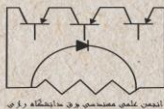
1. N. Dalal and B. Triggs, "Histograms of oriented gradients for humandetection", In CVPR, (2005), 886–893.
2. P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model", In CVPR, (2008), 1–8.
3. C. Ma, C. Yang, F. Yang, Y. Zhuang, Z. Zhang, H. Jia, and X. Xie. "Trajectory factory: Tracklet cleaving and reconnection by deep siamese bi-gru for multiple object tracking" arXiv preprint arXiv:1804.04555, (2018).
4. J. Xing, H. Ai, and S. Lao, "Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses", In CVPR, (2009), 1200–1207.
5. L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows", In CVPR, (2008), 1–8.
6. W. Brendel, M. Amer, and S. Todorovic, "Multiobject tracking as maximum weight independent set", In CVPR, (2011), 1273–1280.
7. H. Kieritz, S. Becker, W. Hubner, and M. Arens. "Online multi-person tracking using integral channel features", 13th IEEE International Conference on, (2016), 122–130.
8. A. Andriyenko, K. Schindler, and S. Roth, "Discretecontinuous optimization for multi-target tracking", In CVPR, (2012), 1926–1933.
9. M. Betke, D. Hirsh, A. Bagchi, N. Hristov, N. Makris, and T. Kunz, "Tracking large variable numbers of objects in clutter", In CVPR, (2007), 1–8.
10. C. Long, A. Haizhou, Z. Zijie, and S. Chong. "Real-time multiple people tracking with deeply learned candidate selection and person re-identification", ICME, (2018).
11. R. Collins, "Multitarget data association with higher-order motion models", In CVPR, (2012), 1744–1751.



مجمع کنفرانس ملی دانشجویی
مهندسی برق ایران



12. S. Tang, M. Andriluka, B. Andres, and B. Schiele. "Multiple people tracking by lifted multicut and person reidentification". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2017), 3539–3548.
13. M. Keuper, S. Tang, Y. Zhongjie, B. Andres, T. Brox, and B. Schiele, "A multi-cut formulation for joint segmentation and tracking of multiple objects", arXiv preprint arXiv:1607.06317, (2016).
14. J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization", IEEE TransPattern Anal. Mach. Intell., vol. 33, no. 9, (2011), 1806–1819.
15. B. Yang and R. Nevatia, "An online learned CRF model for multi-target tracking", in CVPR, (2012), 2034–2041.
16. J. Zhu, H. Yang, N. Liu, M. Kim, W. Zhang, and M.-H. Yang. "Online multi-object tracking with dual matching attention networks". In Proceedings of the European Conference on Computer Vision (ECCV), (2018), 366–382.
17. A. Milan, K. Schindler, and S. Roth, "Detection and trajectory-level exclusion in multiple object tracking", in CVPR, (2013), 3682–3689.
18. D. B. Reid, "An algorithm for tracking multiple targets," IEEE Trans, vol. 24, no. 6, (1979), 843–854.
19. T.E. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association", IEEE J. Ocean. Eng., vol. 8, no. 3, (1983), 173–184.
20. A. Bewley, G. Zongyuan, F. Ramos, and B. Upcroft, "Simple online and realtime tracking", in ICIP, (2016), 3464–3468.
21. N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric", 2017 IEEE International Conference on Image Processing (ICIP), (2017), 3645–3649.
22. Y. Bar-Shalom, "Tracking and data association", Academic Press Professional, Inc, (1987).
23. S. H. Rezatofighi, A. Milan, Z. Zhang, A. Dick, Q. Shi, and I. Reid, "Joint Probabilistic Data Association Revisited", in International Conference on Computer Vision, (2015).
24. C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple Hypothesis Tracking Revisited", in International Conference on Computer Vision, (2015).
25. Gaoang Wang, Yizhou Wang, Haotian Zhang, Renshu Gu, Jenq-Neng Hwang, "Exploit the Connectivity: Multi-Object Tracking with TrackletNet", Computer Vision and Pattern Recognition, arXiv:1811.07258, (2018).
26. A. Zamir, A. Dehghan, and M. Shah, "Gmcp-tracker: Global multi-object tracking using generalized minimum clique graphs", ECCV, (2012).
27. S. H. Bae and K. J. Yoon, "Robust Online Multi-Object Tracking based on Tracklet Confidence and Online Discriminative Appearance Learning", Computer Vision and Pattern Recognition, (2014).
28. R. Henschel, L. Leal-Taixe, D. Cremers, and B. Rosenhahn. "Fusion of head and full-body detectors for multi-object tracking", In Computer Vision and Pattern Recognition Workshops (CVPRW), (2018).
29. Y. Xiang, A. Alahi, and S. Savarese, "Learning to Track : Online Multi-Object Tracking by Decision Making", in International Conference on Computer Vision, (2015).
30. A. Bewley, V. Guizilini, F. Ramos, and B. Upcroft, "Online Self Supervised Multi-Instance Segmentation of Dynamic Objects," in International Conference on Robotics and Automation, (2014).
31. Y.-c. Yoon, A. Boragule, K. Yoon, and M. Jeon. "Online multi-object tracking with historical appearance matching and scene adaptive detection filtering", arXiv preprint arXiv:1805.10916, (2018).
32. J. Chen, H. Sheng, Y. Zhang, and Z. Xiong. Enhancing detection model for multiple hypothesis tracking. In Conf. on Computer Vision and Pattern Recognition Workshops,, (2017), 2143–2152.
33. A. Bewley, L. Ott, F. Ramos, and B. Upcroft, "ALExTRAC: Affinity Learning by Exploring Temporal Reinforcement within Association Chains", in International Conference on Robotics and Automation, (2016).
34. C. Dicle, M. Szaier, and O. Camps, "The way they move: Tracking multiple targets with similar appearance", in International Conference on Computer Vision, (2013).
35. H. W. Kuhn, "The Hungarian method for the assignment problem", Naval Research Logistics Quarterly, vol. 2, (1995), 83–97.
36. M. Keuper, S. Tang, B. Andres, T. Brox, and B. Schiele. "Motion segmentation & multiple object tracking by correlation co-clustering", IEEE transactions on pattern analysis and machine intelligence, (2018).
37. A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, "3D Traffic Scene Understanding from Movable Platforms", Pattern Analysis and Machine Intelligence, (2014).



مجمع کنفرانس ملی دانشجویی
مهندسی برق ایران



38. L. Leal-Taix'e, A. Milan, I. Reid, S. Roth, and K. Schindler, "MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking", arXiv preprint, (2015).
39. L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "MARS: A video benchmark for large-scale person re-identification", in ECCV, (2016).
40. Y. Li, C. Huang, and R. Nevatia, "Learning to associate: HybridBoosted multi-target tracker for crowded scene", in Computer Vision and Pattern Recognition, (2009).
41. K. Bernardin and R. Stiefelhagen, "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics", Image and Video Processing, (2008).
42. M. Lucena, J. M. Fuertes, and N. P. de la Blanca, "Optical flowbased observation models for particle filter tracking," PAA. Pattern Analysis and Applications, vol. 18, no. 1, (2015), 135–143.
43. Hongxia Chu, Kejun Wang, and Xianglei Xing, "Target Tracking via Particle Filter and Convolutional Network", Journal of Electrical and Computer Engineering, (2018).