Int. J. Nonlinear Anal. Appl.
Volume 11, Special Issue, Winter and Spring 2020, 369-378
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2020.4630

Online target tracking via deep convolutional network approach

Mahbubeh Nazarloo[a], Meisam Yadollahzadeh Tabari [b*] ,Homayoon Motameni[c]

[a] *Department of Computer Engineering, Babol Branch, Islamic Azad University,Babol, Iran.*
[b] *Department of Computer Engineering,Babol Branch, Islamic Azad University, Babol, Iran.*
[c] *Department of Computer Engineering,Sari Branch, Islamic Azad University, Sari, Iran.*

Abstract

There is a useful approach for multiple objects tracking easy and efficient that is called simple online and real time tracking(SORT). SORT algorithm performance can be improved by adding visual information. This can reduce the number of identity switches. Because the main framework of the algorithm has a lot of computational complexity, a deep network has been used that is offline on a large data set of trained pedestrians. the focus of this article is on the architecture of this deep network in order to extract more and higher quality visual information that can help the object recognition algorithm. The paper also used a particle filter instead of a Kalman filter to improve data association performance. We tested our proposed method on two standard datasets, MOT16 and MOT17, and compared its performance with other available methods. The results show that the tracking accuracy(52.2) on the MOT17 dataset is improved compared to the existing methods in this field. Experimental evaluation shows that our proposed architecture improves the number of identity switches and ideally tracks goals in complex environments.

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

1. Introduction

Recent developments have inspired and provoked multi-target tracking methods in the accuracy and efficiency of object detectors, especially ordinary detectors by noticing them [1, 2]. These techniques work in a way, that they detect the targets for each frame by the means of a first-rate object detector and these detections are associated with applying online and offline trackers [4, 5]. In cases like pedestrian tracking, that the target aspect is discriminative and the target shows a plain motion pattern, these approaches are suitable [6,5]. also, there are trackers, that depend on the appearance

[*] Corresponding Author: Meisam Yadollahzadeh Tabari

Received: August 2019 *Revised:* August 2020

less than others. They frequently need tuning of an enormous amount of factors and skills in order to get used to the algorithms into these possible situations [8, 9,11].

Recently development in object recognition, tracking-by detection has turned into the most important model in compound object tracking, as mentioned earlier. In this way, we find object trajectories typically in a worldwide optimization issue, which progress the whole video batches straight away. The most popular frameworks of this type are flow network formulations [5, 14] and probabilistic graphical models [15, 8, 17]. During online situations, that the target individually should be presented step by step, these approaches can not be appropriate. Joint Probabilistic Data Association Filter (JPDAF) [19] as well as Multiple Hypothesis Tracking (MHT) [18] present data association for each frame basis. These methods are among the popular and conventional methods that exist in this field.

As the speed of a large number of precise trackers is considered too slow for real-time applications, the trade-off between accuracy and speed come into sight relatively obvious. One of the simple frameworks that attempts to develop both speed and accuracy is simple online and real-time tracking (SORT) [20]. the performance of this method is in image space and data connection for each frame by applying the Hungarian approach with an association metric that evaluates bounding box overlap, which is Kalman filtering. This undemanding method accomplishes positive performance at high frame rates.

SORT proceeds quite a large amount of identity switches. The reason for it is that the employed association metric is precise once state assessment ambiguity is low. As a result, SORT has a shortage of tracking through blockage because they become visible normally in the front part of camera views. we change the place of association metric with a metric that is more informed, which unites motion and form data as a solution for this issue. To combine motion and appearance information, we used the deep neural network with unlike deep sort architecture [21]. We also replaced the Kalman filter with a particle filter. The advantage of the particle filter algorithm is its plainness and flexibility. also, this is undemanding to deals with the Gaussian multimodality system model. A lot of associated literature are provided in [42]. Facts and details from unlike measurement supplies are able to be used in a particle filter framework, which has improved the tracking performance to a great extent. We develop toughness against misses and occlusions through a combination of this network at the same time as we keep the system simple to apply and well-organized and relevant to online scenarios.

This paper is organized as the following sections: Section 2 presents a short assessment of associated literature in multiple object tracking area. Section 3 proposes our approach and architecture of deep network before a demonstration of the anticipated framework usefulness on regular benchmark series in Section 4 is done. Eventually, Section 5 gives us a review of the learned results and argues potential developments.

2. RELATEDWORKS

The Joint Probabilistic Data Association (JPDA) filters [22, 23] and Multiple Hypothesis Tracking (MHT) [19] has solved multi-object tracking commonly and conventionally. they postpone making complicated decisions despite the fact that there is a great ambiguity in object assignments. The combination of these methods difficulty is increasing in the amount of tracked objects, which make them not practical for concurrent purposes in extremely active surroundings. the JPDA formulation [22] in visual Multiple ObjectTracking (MOT) is considered by Rezatofighi et al.[23]. His purpose was addressing the combinational complexity problems in solving 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. though, such approaches put off the conclusion, that are inappropriate for online tracking.

Typical multi-target trackings are known as a network flow problem [5] or its variation [14]. most of these approaches depend on a great amount of object form and presume motion models that are simple as their priorities. they function great in settings and the targets are pedestrians or means of transportation . one of the drawbacks of applying a network flow formulation is that it needs arranging starting and finishing locations and/or times of the target that possibly will be hard to identify ahead. Brendel et al. [6] applied an extremely serious set formulation, though they regard just two-frames connections.

Applying a General Maximum Clique partitioning formulation is suggested by Zamir et al [26]. This method chooses on from every tracklet the greatest candidate with the intention of attaining global association. The linear Assignment formulation [11,14,15] is comparable with the Generalized Linear Assignment (GLA). however, the good thing about it is that it lets tracks start and end anyplace in place and time.

MOT solver method is proposed by Braso et al .[3]. This method is based on message passing networks. it uses the natural graph problem structure to present both feature learning and final result calculation. A novel time-aware neural message passing update step which is inspired by classic graph MOT formulations is proposed by them.

Many tracking techniques plan is forming form models of both the objects [27,29] and the global model [30,33] by online education. Motion is frequently integrated to support associating detection to tracklets [34,30], as well as appearance models. Universally ideal solutions like the Hungerian algorithm [35] can be applied [37] when one-to-one correspondence models are considered.

The method of Geiger et al. [37] applies the Hungarian algorithm [35] in a couple of steps procedure. The earliest step is when tracklets are shaped with connecting detections from corner to corner of closest frames, where geometry and form signals are shared to shape the similarity matrix. afterward, the tracklets are connected with another by using geometry and appearance cues in order to link broken trajectories that are caused by occlusion. These pair steps union technique limits this approach to batch computation.

Nicolai Wojke et al. [21] method combines appearance information to develop SORT performance. They can follow objects throughout longer stages of occlusions and successfully dropping the amount of identity switched down caused by this Extention. however, we use new deep network architecture for appearance descriptor as explained in the subsequent part.

3. METHODOLOGY

This approach is expressed by regular assumption tracking methodology with recursive Particle filtering and data connection for every frame. the different parts of this system and our anticipatedtechnique will be established in the next sections.

3.1. Detection and state estimation

The detection consequences in data association based on MOT have a strong effect on the tracking presentation. Both the Detection and Kalman filtering framework follow the original formulation in [22].

Instead of using the Kalman filter, we used Particle filter according to the idea set out in [43] to track targets, and we were inspired by their idea. A target tracking algorithm incorporating the Particle filter and convolution network is designed in the paper. The extracted feature from convolutional networks is represented in the particle filter framework. In order to signify the state change of the object, the target local and spatial data are completely applied. since the global information pieces of particle filter are incorporated in the direction of finding out the situation of the present targets, the local form alteration and partial occlusion issue of the aim are better worked out. This is based on the target state that is dealt with different information.

We presume a very common tracking situation in which the camera is uncalibrated also there is no ego-motion data in hand, which is the most typical setup measured in the latest multiple object tracking benchmarks[38], as described in [21]. The inter-frame movements for every object are estimated by the linear steady rapidly model that doesn't rely on further objects or camera movement. the position of every target is planed as:

$$\mathrm{x}=[\mu,\nu,\gamma,h,\dot{x},\dot{y},\dot{\gamma},\dot{h}] \quad (1)$$

Where μ and v signify the horizontal and vertical pixel place of the target center, it coordinates aspect ratio γ, height h,and their relevant velocities in the picture. the detected bounding box is applied to renew the target location, where the velocity mechanism is resolved in the best and the most favorable way by using a Particle filter frame when discovery is connected with an object. the location is basically forecasted with no adjustment by applying the linear velocity form if no recognition is related to the target. We count the frame numbers from the previous successful measurement association for each track in this approach. during Particle filter prediction, this counter is increased and reorganized to 0 when the track has been connected with a measurement. We take them out from the track set if the tracks are more than the prearranged maximum age. A new track hypothesis is arranged for every detection that is not capable of being connected with an access track. These up-to-the-minute tracks are arranged like uncertain throughout their initial three frameworks. We look forward to a flourishing measurement association during each time step.

3.2. Data Association

Each target bounding box geometry is anticipated by forecasting the new status of it in the recent frame when our purpose is to define detections to obtainable targets. the task of cost matrix will be computed as the intersection-over-union (IOU) space among every recognition and the entire the bounding boxes from the presented objects, that was anticipated. By using the Hungarian algorithm, the assignment is solved optimally. We combine motion and appearance detail by arranging two suitable metrics for this issue formulation.

We need to merge motion and appearance information to formulate an issue as we mentioned earlier. we use the Mahalanobis spaceamong anticipated Particle states plus recently arrived assessments. Particularly by making the Mahalanobis distance another unaware metric for tracking through occlusions, camera motions that are not counted couldbring in quick disarticulations in the image surface. Therefore combining the second metric into the assignment issue is so much better. d_j are computed a form descriptor r_j with $\|r_j\| = 1$ for each bounding box detection. gallery $R_k = \{r_k^{(i)}\}_{k=1}^{L_k}$of the last $L_k = 100$ connectedform descriptors for every track k is kept. after that the following metric procedure the slightest cosine space among the $i-$th track and $j-$th recognition in form space:

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

The Mahalanobis distance presents information about potential objects places based on particularly practical for short-term predictions motion. whereas, the cosine space regards formdata that are basically practical in the direction of improving identities later than long-term occlusions when movement is less discriminative.

3.3. Data Association

For computing the similarity importance in data association, the space among form feature is used. The affinity significance has to be outsized and big for similar identity persons and be undersized for persons with unlike identities, based on the perfect and ideal appearance features. By using a network that our network architecture is presented in Figure 1, the appearance feature in our operation is extracted.

Based on the results and experiences obtained, it can be said that in deep learning, the deeper the network, the better the accuracy of the network, provided that the problems of vanishing gradient do not occur. Accordingly, in the proposed architecture, we tried to increase the detection accuracy by deepening the grid and also using residual layers to prevent the gradient from disappearing.

This system has been accomplished on MARS data collection and it includes more than 1,100,000 pictures of 1,261 pedestrians, that makes this completely suitable for deep metric learning in people tracking situation.

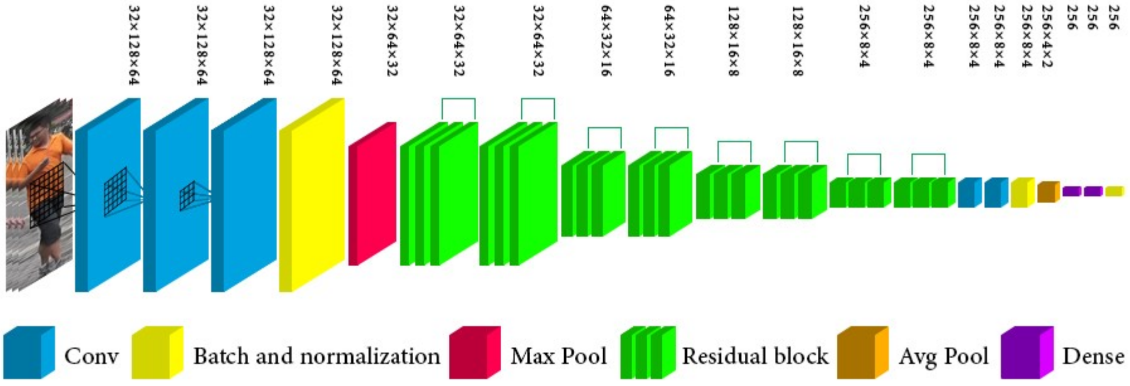

Figure 1: Architecture of our deep convolutional network for feature extraction.

Table 1: Performance of the proposed method on MOT16 benchmark

Method	MOTA	IDF1	MT	ML	FP	FN	ID $_{sw}$
GCRA[44]	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

Generally, our network is made of two parts. Four convolutional layers are applied with step 1 and

Table 2: Performance of the proposed method on MOT17 benchmark

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

similar padding, in the first part. we applied batch normalization between each layer as well. Batch normalization decreased the quantity via what the concealed unit values move around(covariance shift).batch normalization lets every network layer learn itself to be more independent of other layers as well. we applied the eight remaining blocks one after the other of unlike sizes, in the next part. We augment the network depth to expand exactness and decrease the number of individuality changes. As a final point, we applied two compact layers of 250 in order to estimate the global feature map. And the final network output is achieved right after applying a batch normalization layer.

Learning decent mappings from input to output in neural networks for accidental initialization of weights is crucial. There are several local minimums which might be traped by back-propagationbecause the search space that involves numerous weights throughout training is enormous.on the other hand, the weight initialization randomization role has to be chosen and identified cautiously unless there is a great danger that the preparation development decreased to the point of uselessness.

In the learning process of this network, the Adam optimizer algorithm with 1e-3 learning rate has been used. The latest result of our proposed method, presented in Table 1 and Table 2, is obtained in a learning process with 200,000 iterations.

In order to weight initialization, we applied the Xavier or variance scaling method. The Xavier weight initialization method is a great development compared with the immature method of weight scaling. this approach assisted us in a great way with increasing the speed of the deep learning field. thus it adapts itself according to weight values amount. This methods idea is that your network will learn optimally if you could maintain the variance constant layer by layer in either feed-forward or back-propagation direction. As you go through the layers, your weight will ultimately saturate your non-linear neurons in both positive and negative direction as if the variance boosts or reduces. This initialization has been established to work better with ReLU activation functions in general, because we apply ReLU activation function in this network:

$$\mathrm{var}(\mathrm{w}_i) = \frac{2}{n_{in}} \quad (3)$$

In the above formula, w represents the weights and n represents the number of inputs for each node. In the end, this network consists of 4,654,764 parameters and one forward pass of 32 bounding boxes that take about 19 ms on Nvidia Titan XP. so if a modern GPU is accessible, this network is suitable for online tracking.

4. EXPERIMENTS

We apply MOT16 and MOT17 data sets in order to teach and assess our tracking performance that includes either moving or static camera sequences. This benchmark estimates tracking performance on seven demanding test sequences that include front view scenes with a movable camera in addition to top-down observation arrangements. we utilize the evaluation metrics defined in [40], along with the MOT metrics [41]:
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 asimilar 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(↓):theamount of missed detections.
ID$_{sw}$(↓):theamount of times ID switches to a dissimilarformerly tracked purpose.
Estimationprocedures with (↑), upper scores signifyimproved performance; while for evaluation procedures with (↓), minor scores denote better performance.

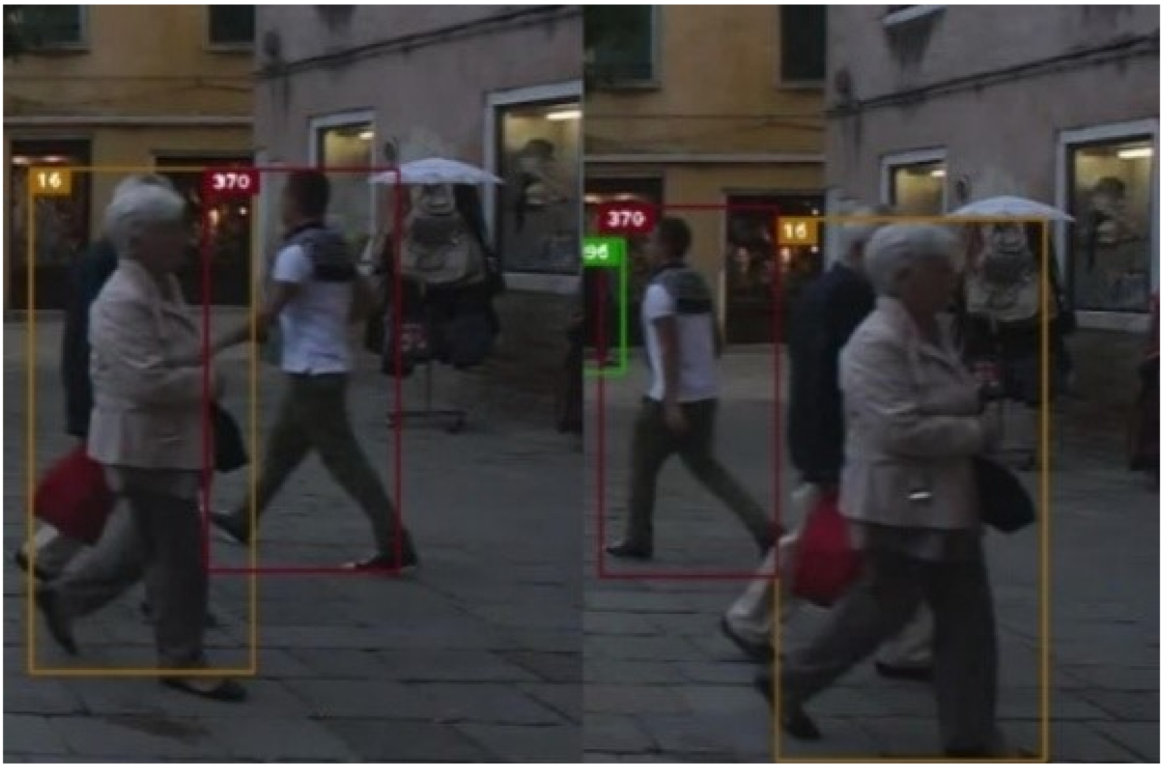

Figure 2: Our method representative output on MOT challenge dataset in an ordinary tracking situation with regular occlusion.

Tracking performance is estimated to apply the MOT benchmark [38] test server where the ground accuracy for 11 sequences is suspended. Many other baseline trackers are evaluated with the proposed SORT method in Table 1 and Table 2. this approach has eased the number of identity switches successfully.

We realize a little augmentation is mostly tracked objects number and reduction in typically lost objects. Generally, we sustain identities through longer occlusions effectively because of appearance information combination.we can see them by tracking output qualitative analysis that we supply in complementary material. An excellent tracker output is revealed in Figure 2.

This approach is a great opponent to other online tracking frameworks. even thoughwe keep competitive MOTA grades, track fragmentation, and false negatives, our method returns the smallest amount of online method identity switches. In Figure 3, you can see the accuracy of the proposed method with other available methods on the MOT17 benchmark.

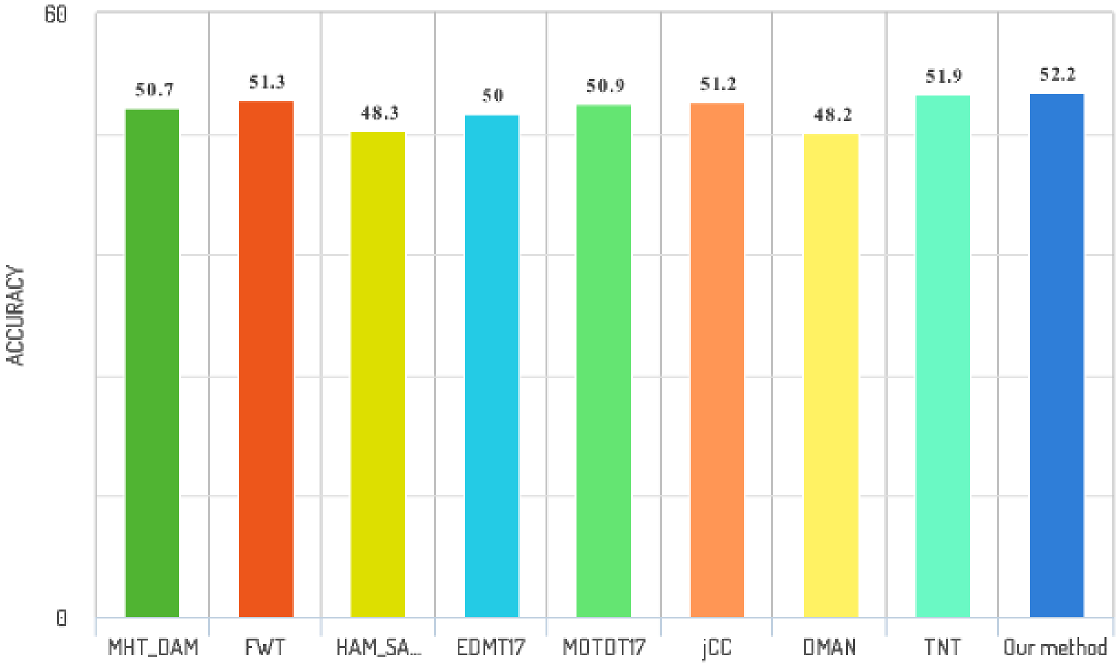

Figure 3: The multi-object tracking accuracy(MOTA) plots of quantitative comparison for MOT17 benchmark [38].

5. CONCLUSION

By applying the Particle filter and deep convolutional network, we propose a promising tracking methods. we use deep learning method in order to pull out effective features for strong and sturdy tracking. Appearance changing and occlusion issues will be solved by the algorithm efficiently in a strict way. Based on experimental consequences, the improved approach is preferable to traditional tracking methods in severe tracking surroundings and it has slightly reduced identity switches numbers compared with the deep sort method.

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] G. Bras´o, LTaix´e. “Learning a Neural Solver for Multiple Object Tracking”In CVPR, (2020),6247–6257.

[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.

[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 objects 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. Sznaier, 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).

[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).

[44] 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).