Multiple Object Tracking in Video Sequences: A Review Article

Krit Rojanarungruengporn
Faculty of Engineering
Chulalongkorn University Bangkok, Thailand
krit.rojana@gmail.com

***Abstract*— Multiple object tracking is one of important research area in computer vision. Despite numerous researches proposed to solve the problem, it remains critical due to appearance variation, change in illumination and object occlusions. In this paper, recent approaches to tackle this problem are reviewed. Problem formulation, categorization, key components of multiple object tracking are presented. Evaluation and result comparisons are highlighted. Finally, the interesting issues which can become research topic in the future are presented.**

***Keywords*— Multiple object tracking, Video, Review, MOT, MTT**

1. INTRODUCTION

Because of the rapid technology development and economic of scale in manufacturing, nowadays, both the cost per computational unit and the cost of camera are getting lower which encourage the rapid advances of computer vision technology. Multiple object tracking is one of challenging problem in computer vision. It concerns about locating multiple targets, finding each target in the video sequences to maintain its identity and linking all the targets over the sequences to make its individual trajectory. Multiple object tracking is now a very interesting topic because of its academic and commercial use cases. The real-world applications can range from tracking the group of animals (such as fish[1],[2] , birds [3]), tracking leaves of plant for calculation of the growth rate [4],[5] , tracking cars on the road [6] and tracking pedestrians on the street[7], visual surveillance[8], VR(Virtual reality)[9].

Multiple object tracking is a challenging problem. If we compare multiple object tracking with the single one, we will find that the same principles and procedure are still valid except that we do it with multiple objects. The same problems for single target such as change in appearance, illumination, scale and out of plane rotation are still there. Additionally, there are issues that existed for multiple targets: Counting number of objects to be tracked which can be varied over time, maintain each object identity over different frames, occlusion, same appearance between multiple object and the interaction between them.

In this review paper, we will provide fundamental aspects of multiple object tracking which consist of understanding what is the problem of multiple object tracking in Section 2, multiple object tracking categorization in Section 3, multiple object tracking components and recent approaches in Section 4, evaluation in Section 5, conclusion in section 6, lastly, interesting research topic for future work in section 7.

2. Understanding the problem

In general, multiple object tracking problem is very similar to multiple variable estimation problem. According to [10], the aims of multiple object tracking is to estimate the optimum sequential states of all the tracking objects in the video sequence. Given the sequence, the i-th object in t-th frame can be denoted as s_t^i, $S_t = (s_t^1, s_t^2, \dots, s_t^{M_t})$ as the states of all the M_t objects in t-th frame and $S_{i_s,i_e}^i = \{s_{i_s}^i, \dots, s_{i_e}^i\}$ as the sequence of states of i-th object, where i_s, i_e are the first and last frames which i-th object exists, and $S_{1:t} = \{S_1, S_2, \dots, S_t\}$ is all the sequential states of all the objects from frame 1-th to t-th.

The same convention is applied for the observation which is the result from detection-based tracking. o_t^i denotes observation properties of object i-th in frames t-th, and $O_t = (o_t^1, o_t^2, \dots, o_t^{M_t})$ denotes the observations of all the M_t objects in t-th frame, and $O_{1:t} = \{O_1, O_2, \dots, O_t\}$ denotes all the sequential observations from first frame to t-th frame.

The estimation can be done by using MAP (Maximal a posteriori) estimation from the conditional distribution of sequential states given all the observations:

$$\bar{S}_{1:t} = \underset{S_{1:t}}{\operatorname{argmax}}\, P(S_{1:t}|O_{1:t}) \qquad (1)$$

The different approaches to solve the problem can be thought as different ways to solve MAP problem.

3. Multiple object tracking categorization

There are so many ways to group multiple object tracking algorithms based on different criteria, and it is not an easy task to pick one and use it as a general criteria. [10] proposed three criteria which follow the standard procedure of execute the task that are initialization method, processing mode and output type.

3.1 Initialization Method

Most of the works on multiple object tracking can be categorized two groups depend on the initialization process [11]: Detection-Based Tracking (DBT) and Detection-Free Tracking (DFT).

Detection-Based Tracking DBT consists of two main tasks which are object detection and connecting the object to trajectories. The pretrained detector is needed in this initialization method. Its task is to detect the objects or motion in each frame of the video and obtain object hypotheses [12]. After that, in linking step, the detected hypotheses will be linked together to form trajectories. There are two issues associated with DFT. First, Since the detector is pretrained, it can only detect specific object used in the training phase. Second, the performance of the DFT is heavily depend on the detector, not on the tracking phase.

Detection-Free Tracking (DFT), on the other hands, is not required object detection, however, it need manually initialize number of objects in the first frame. Compare to DBT, nowadays, DBT is much more popular than DFT, because DBT can add and remove the objects that entering and leaving the scene automatically given a lot of benefits to many applications. Table 1 lists pros and cons of DBT and DFT.

3.2 Processing Mode

By how the algorithms is processing its data, we can divide them into two groups: Online tracking and offline tracking. For online tracking [14],[15], when the algorithm processes the current frame, it only uses the information from the past up until the current frame. For the offline [16], it collects all observations in every frame first and uses information from all of them to process the current frame hence the name offline tracking. Table 2 lists differences between online and offline tracking.

3.3 Output Type

Depend on the randomness of output, multiple object tracking can be categorized into deterministic and probabilistic tracking. The result of the deterministic method will be constant when running the process multiple times, however, the output for the probabilistic will be different for each run.

The three criteria presented in section 3 is just an easy example for categorize the multiple object tracking. There are other criteria such as Category free tracking (CFT) and Association based tracking (ABT) [17]. In addition, considering the tracking process, [18] proposed to classify the algorithms based on their tracking methods which are Bayesian theory and data association.

TABLE 1: Comparison between Detection-Based Tracking and Detection-Free Tracking [13].

	Factor	Detection-Based Tracking	Detection-Free Tracking
1.	Initialization	Automatically detection, some error depends on detector performance	Manual initialization, exact number of initial objects
2.	Number of objects	Vary	Fixed
3.	Applications	In general, work on specific type of target	Any type of targets
4.	Advantages	Can handle varying number of objects	Do not need object detector
5.	Disadvantages	Performance depends on detector	Manual initialization

TABLE 2: Comparison between online and offline tracking.

	Factor	Online	Offline
1.	Input	Observations from the past up to the current frame	Observations from all the frames
2.	Methodology	Gradually link new observations to the existing trajectories.	Connect all the observations to form trajectories
3.	Advantages	Suitable for online tasks	Can get global optimal result
4.	Disadvantages	Current prediction can suffer from insufficient observations	Heavier computation and memory usage

3.4 Tracking Method

Tracking is the process the connect the observations to the trajectories. For the Bayesian theory, the tracking is done like a prediction process, hence, all the trajectories are estimated according to the Bayesian theory. On the other hands, Data association method focuses on linking the observations to the trajectories by considering the affinity between objects. Moreover, this method can be divided further into two categories which are local optimization-based data association and global optimization-based data association.

4. Multiple object tracking components

When design the multiple object tracking algorithm, there are two main tasks to consider which are how to measure the similarity between objects and how to link the objects based on its similarity across the frames. The first one focus on modeling the observation of the object such as its appearance, motion and position within the frame. The second one involves how to link the identity of the object based on the similarity collected from the observation model to the other identity found in other frames. In short, the first one is how we detect and the second is how we track multiple object. In this section, we start with the review of the observation model which includes appearance model, motion model, exclusion model and integrated model, then we review the tracking method which we categorize them to be following the Bayesian theory and Data association method as mentioned in the section 3.4.

4.1 Observation model

Observation model is the model to describe the properties of the object. In this section will look into each component of the model consist of appearance model, motion model, exclusion model and integrated model

4.1.1 Appearance model

Appearance model is the importance tool for measure the properties of the objects, and, from these properties, we can calculate the similarity between objects to use it in the multiple object tracking. To measure the properties of the object, there some features used in the previous works such as color, shape, gradient, texture, depth, super-pixel, motion or optical flow. In some works, they combine multiple features and use it as the appearance model. The main features used in the previous work are presented in the Table 3, and the example of using combined features appearance model is listed in the Table 4.

TABLE 4: Combined features used for appearance model [10].

Strategy	Combination of features	References
Boosting	Color, HOG, shapes, covariance matrix, etc.	[23]
Concatenating	Color, HOG, optical flow, etc.	[24]
Summation	Color, depth, correlogram, LBP, etc.	[25]
Product	Color, shapes, bags of local features, etc.	[26]
Cascading	Depth, shape, texture, etc.	[27]

According to [18], the appearance model can be grouped into generative model and discriminative model. The first one is model the visual observation of the moving object, and the tracking is only searching all the other observations to find the closed appearance to the appearance of the constructed target model. For the discriminative model, the objects tracking is the binary classification of the object from the background.

4.1.1.1 Generative appearance model

The common features that used to make the generative appearance model are color histogram, HOG, Gaussian mixture model, manifold learning sparse representation and hidden Markov random. [17] used joint sparse representation to select the templates and estimate the coefficients of the them. [28] used hidden Markov random model to represent the joint dependencies of labels and linked trajectories.

TABLE 3: Features used for appearance model [18].

Feature type	Present by	Pros	Cons	References
Color	Color histogram	Efficient, Robust to region rotation, scaling and changing shape.	Sensitive to illumination changes. Not take spatial distribution of the pixel values into account.	[19][20]
Contour	Edge	Less sensitive to illumination changes		[20]
Gradient	HOG, SIFT	Suitable for human detection		[19]
Shape		Low computational time		
Texture	LBP	Accuracy	Computational expensive	[20]
Motion	Optical flow	Handle occlusion	Computational expensive	[20]
Spatiotemporal information	Super-voxel, super-pixel	Handle partial occlusion and pose variations		[19][22]

Depth	3D	Handle drift and occlusion		[21]
Multi-feature	Color with texture	Robust	Computation expensive	[22]

4.1.1.2 Discriminative appearance model

This model will select different features to construct a classifier. The model will classify the features of the object and the background to detect the object. [22] used SVM as the appearance model to classify the objects.

4.1.2 Motion model

The motion model is used to describe the dynamic motion of the objects in the scene. It is used to determine the future position of each objects in order to reduce searching space of the tracking algorithm. Motion model can be categorized into 2 groups: linear motion model and non-linear motion model.

1) Linear model is based on the assumption that in the very short distance all motion can be viewed as linear motion. This model is much more popular that the non-linear model [29]. There are three notions that can be used to construct the model
    a. Velocity smoothness, this notion considers that the velocity of the object in the successive frames changing smoothly. In [30], the author used this notion to implement cost term as following:

    $$C_{dyn} = \sum_{t=1}^{N-2} \sum_{i=1}^{M} \| v_i^t - v_i^{t+1} \|^2, \qquad (2)$$

    b. Position smoothness, this model enforces the discrepancy between the observed position and estimated position.
    c. Acceleration smoothness, in reference [31], the probability distribution of motion of a state $\hat{s}_k$ at time k given the observation $o_k$is model as:

    $$P(\hat{s}_k | o_k) = \prod_k N(x_k - \hat{x}_k; 0, \Sigma_p) \prod_k N(v_k; 0, \Sigma_v) \prod_k N(a_k; 0, \Sigma_a), \qquad (3)$$

    Where $v_k$is velocity, $a_k$is the acceleration and N is a zero mead Gaussian distribution.
2) Non-linear model, this model can archive higher accuracy in motion affinity between targets. [32] used motion dynamic as observation feature to classify objects with similar appearance. [33] used relative motion between multiple objects to construct a Relative motion network (RMN) to reduce the effect of unexpected camera movement.

4.1.3 Exclusion model

Exclusion model is a constraint that use to prevent the collisions of the objects and trajectories. The model is based on the notion that two objects cannot occupy the same physical space. There are two types of exclusion which are detection-level exclusion and trajectory-level exclusion. The first one state that the two detections in the same frame cannot assign to the same target. The later states that two trajectories cannot overlap each other.

Detection-level exclusion can be model by soft model and hard model. For soft model, it can be done by minimizing the cost function to penalize if the exclusion is violent [35]. The hard model is implemented by define an explicit constraint to the model to prevent the collision.

Trajectory-level exclusion is defined by modeling the penalty to prevent the detector from assign the target with two different trajectory labels. For example, from [36], the penalty is defined as the inverse proportion of the distance between two detection responses with different trajectories labels. If two trajectories are overlap, the penalty will be infinite.

4.1.4 Integrated model

In multiple object tracking, there are some cases when we detect the multiple object with the same type. In this case, using only one appearance model is not enough, because the shape and textures of the object are the same. The integrated model will use multiple appearance features combined to from the detection model. In reference [37], the author proposed an appearance model of object combined a color model, motion model, sparse appearance model and spatial information.

4.2 Tracking method

In most cases, multiple object tracking is done in two main steps: Detect the objects and link the data to from trajectories. In the first step, the detector is trained to find all the target objects in every frames of the input video and represent the found objects by using observation model, after that tracking work is only link all the observations to the target to from the trajectories. In this section, we present two categories of tracking method depend on how they link observation to target: Bayesian theory-based and data association based.

4.2.1 Bayesian theory-based

In this section, we will present three tracking methods based on Bayesian theory including particle filter, Kalman filter and Bayesian framework

Kalman filter is efficient for tracking small number of objects. Because of the recursive nature of the method, it become difficult to track large number of objects using Kalman filter. In [38], the author used this method to construct motion model of the objects by using its current observations to predict object next position. By doing this, the search range of the algorithm is reduced.

Particle filter can handle non-linear system and multi-modal distribution very well. However, according to [39], it is not suitable for high dimension state space.

Bayesian framework can be viewed as a probability graph model. It can construct the trajectory by maximizing the posterior distribution of state of target given the past and current information. The works that used this method can be found in [34]

4.2.2 Data association-based

The tracking part in these methods usually perform by linking each detected object together based on their observation model. In general, in order to find the associations, we have to decide what is the cost function, and then we minimize the cost to find the association. Data association-based methods can be divided into two groups: Local optimization-based data association and global optimization-base data association. The differences are that the local one will associate data in just two or a few more frames opposed to the global which consider input as batches of the frames.

4.2.2.1 Local optimization-based data association

Bipartite matching and its extensions are the best example of this group of method. In this group of algorithms, the Hungarian algorithm is the most popular one to use to find affinity to link the object to trajectory. In reference [40], the author used Hungarian algorithm to link object hypotheses and detection responses based on similarity measured from position, size and color.

4.2.2.2 Global optimization-based data association

Grape-based approach, hierarchical data association-based, quadratic boolean problem-based and binary integer programing-based method are the most common global optimization methods.

5. Multiple object tracking evaluation

Because multiple object tracking is difficult to categorize due to complex nature of algorithm, it is also hard to define the metric to evaluate the performance of the algorithm. In [10], the author proposed two groups of metrics according to the most popular tracking style today which is tracking by detection. The first group of metrices aims to measure the detection performance of the algorithm, and the second one has a goal to measure the performance of tracking. The summarize of the metrics can be found on Table 5.

Based on the presented metrics, the quantitative results comparison of recent multiple object tracking algorithms on public dataset, PETS2009-S2L1, can be found in reference [10].

6. CONCLUSION

In this paper, we give an idea about how the multiple object tracking problem is formulated, how to categorize its algorithms, What are the building blocks to consider in design the algorithm, what are the metrics to evaluate the performance of the algorithms and along the way we presented the previous works that have been done in the research community. Finally, we point out to the quantitative comparison results of the multiple object tracking algorithms on the public dataset.

7. Future direction

One area that can be improve in the future is the adaptive detection in video. Since one of the main tasks of multiple object tracking is to detect the objects using the pretrained detector, the adaptive detector that can detect a wide range of objects is still needed.

Another interesting topic is the 3D multiple object tracking, since the works on multiple object tracking nowadays focus only on 2D, the 3D one is the way to go in the future. It can provide more accuracy position and size estimation of the object. Furthermore, it can handle occlusion much more efficient than 2D Tracking.

Deep learning is another hot topic that multiple object tracking can be benefit from. It is a very powerful framework that efficiently solve the problem of object classification, object detection and single object tracking. If we can use deep learning framework as the observation model of objects detection, it will increase the performance of tracking a lot.

TABLE 5: An overview of evaluation metrics for multiple object tracking adapted from [10].

Metric	Type	Measured property	Description	Direction of better performance
Recall	Detection	Accuracy	Ratio of correctly matched detections to ground-truth detections	Greater
Precision	Detection	Accuracy	Ratio of correctly matched detections to total result detections	Greater
FAF/FPPI	Detection	Accuracy	Number of false alarms per frame averaged over a sequence	Smaller
MODA	Detection	Accuracy	Combines missed detections and FAF	Greater
MODP	Detection	Precision	Average overlap between true positives and ground truth	Greater
MOTA	Tracking	Accuracy	Combines false negatives, false positives and mismatch rate	Greater
IDs	Tracking	Accuracy	Number of times that a tracked trajectory changes its matched ground-truth identity (or vice versa)	Smaller

MOTP	Tracking	Precision	Overlap between the estimated positions and the ground truth averaged over the matches	Greater
TDE	Tracking	Precision	Distance between the ground-truth annotation and the tracking result	Smaller
OSPA	Tracking	Precision	Cardinality error and spatial distance between ground truth and the tracking results	Smaller
MT	Tracking	Completeness	Percentage of ground-truth trajectories which are covered by the tracker output for more than 80% of their length	Greater
ML	Tracking	Completeness	Percentage of ground-truth trajectories which are covered by the tracker output for less than 20% of their length	Smaller
PT	Tracking	Completeness	1.0 - MT - ML	-
FM	Tracking	Completeness	Number of times that a ground-truth trajectory is interrupted in the tracking result	Smaller
RS	Tracking	Robustness	Ratio of tracks which are correctly recovered from short occlusion	Greater
RL	Tracking	Robustness	Ratio of tracks which are correctly recovered from long occlusion	Greater

REFERENCES

[1] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher, “Detecting, tracking and counting fish in low quality unconstrained underwater videos,” Proc. Int. Conf. Comput. Vis. Theory Appl., pp. 514–519, 2008..

[2] E. Fontaine, A. H. Barr, and J. W. Burdick, “Model-based tracking of multiple worms and fish,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops, 2007, pp. 1–13.

[3] W. Luo, T.-K. Kim, B. Stenger, X. Zhao, and R. Cipolla, “Bi-label propagation for generic multiple object tracking,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2014, pp. 1290–1297.

[4] X. Yin, X. Liu, J. Chen, and D. M. Kramer, “Joint Multi-Leaf Segmentation, Alignment, and Tracking for Fluorescence Plant Videos,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, pp. 1411–1423, Jan. 2018.

[5] G. Viaud, O. Loudet, and P.-H. Cournède, “Leaf Segmentation and Tracking in Arabidopsis thaliana Combined to an Organ-Scale Plant Model for Genotypic Differentiation,” Frontiers in Plant Science, vol. 7, Nov. 2017.

[6] M. Betke, E. Haritaoglu, and L. S. Davis, “Real-time multiple vehicle detection and tracking from a moving vehicle,” Mach. Vis. Appl., vol. 12, no. 2, pp. 69–83, Feb. 2000.

[7] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, “You’ll never walk alone: Modeling social behavior for multi-target tracking,” in Proc. IEEE Int. Conf. Comput. Vis., 2009, pp. 261–268.

[8] X. Wang, “Intelligent multi-camera video surveillance: A review,” Pattern Recognit. Lett., vol. 34, no. 1, pp. 3–19, Jan. 2013.

[9] H. Uchiyama and E. Marchand, “Object Detection and Pose Tracking for Augmented Reality: Recent Approaches,” in Proc. Korea-Japan Joint Workshop Frontiers Comput. Vis., 2012, pp. 721– 730.

[10] W Luo, X Zhao, and TK Kim. Multiple object tracking: A review. arXiv preprint arXiv:1409.7618, 2014

[11] B. Yang and R. Nevatia, “Online learned discriminative partbased appearance models for multi-human tracking,” in Proc. Eur. Conf. Comput. Vis., 2012, pp. 484–498

[12] B. Bose, X. Wang, and E. Grimson, “Multi-class object tracking algorithm that handles fragmentation and grouping,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2007, pp. 1–8.

[13] B. Yang and R. Nevatia, “Online learned discriminative partbased appearance models for multi-human tracking,” in Proc. Eur. Conf. Comput. Vis., 2012, pp. 484–498.

[14] L. Zhang and L. van der Maaten, “Structure preserving object tracking,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2013, pp. 1838–1845.

[15] J. Zhang, L. L. Presti, and S. Sclaroff, “Online multi-person tracking by tracker hierarchy,” in Proc. IEEE Int. Conf. Advanced Video Signal-Based Surveillance, 2012, pp. 379–385.

[16] J. F. Henriques, R. Caseiro, and J. Batista, “Globally optimal solution to multi-object tracking with merged measurements,” in Proc. IEEE Int. Conf. Comput. Vis., 2011, pp. 2470–2477.

[17] B.Yang, R.Nevatia,“Multi-target tracking by online learning a CRF model of appearance and motion patterns,” International Journal of Computer Vision,vol. 107, no.2,pp. 203-217,2014.

[18] L. Fan, Z. Wang, B. Cail, C. Tao, Z. Zhang, Y. Wang, S. Li, F. Huang, S. Fu, and F. Zhang, “A survey on multiple object tracking algorithm,” 2016 IEEE International Conference on Information and Automation (ICIA), 2016.

[19] C. H. Kuo, C.Huang, and R.Nevatia, “Multi-target tracking by on-line learned discriminative appearance models,” IEEE Conference on Computer Vision & Pattern Recognition, vol. 238, pp. 685-692, 2010.

[20] W. Hu, W. Li, X. Zhang, et al,“Single and Multiple Object Tracking Using a Multi-Feature Joint Sparse Representation,” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 37, no. 4, pp. 816-833,2015.

[21] Y. Chen, Y. Shen, X. Liu, et al,“3D object tracking via image sets and depth-based occlusion detection,” Signal Processing,vol. 112, pp. 146-153, 2015.

[22] A. Milan, L. Lealtaixe, K. Schindler, et al,“Joint tracking and segmentation of multiple targets,”Computer Vision and Pattern Recognition. IEEE, 2015.

[23] Y. Li, C. Huang, and R. Nevatia, “Learning to associate: Hybridboosted multi-target tracker for crowded scene,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2009, pp. 2953–2960.

[24] W. Brendel, M. Amer, and S. Todorovic, “Multiobject tracking as maximum weight independent set,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2011, pp. 1273–1280.

[25] Y. Liu, H. Li, and Y. Q. Chen, “Automatic tracking of a largenumber of moving targets in 3d,” in Proc. Eur. Conf. Comput. Vis., 2012, pp. 730–742.

[26] J. Berclaz, F. Fleuret, and P. Fua, “Robust people tracking with global trajectory optimization,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2006, pp. 744–750.

[27] D. M. Gavrila and S. Munder, “Multi-cue pedestrian detection and tracking from a moving vehicle,” Int. J. Comput. Vis., vol. 73, no. 1, pp. 41–59, Jan. 2007.

[28] B. Wu, S. Lyu, B. G. Hu, et al,“Simultaneous Clustering and Tracklet Linking for Multi-face Tracking in Videos,” IEEE International Conference on Computer Vision. IEEE Computer Society, 2013, pp. 2856-2863.

[29] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, “Robust tracking-by-detection using a detector confidence particle filter,” in Proc. IEEE Int. Conf. Comput. Vis., 2009, pp. 1515–1522.

[30] A. Milan, S. Roth, and K. Schindler, “Continuous energy minimization for multitarget tracking,” IEEE Trans. Pattern Anal. Mach. Intel., vol. 36, no. 1, pp. 58–72, Jan. 2014.

[31] C.-H. Kuo and R. Nevatia, “How does person identity recognition help multi-person tracking?” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2011, pp. 1217–1224.

[32] C. Dicle, O. I. Camps,M. Sznaier,“The Way They Move: TrackingMultiple Targets with Similar Appearance,” ICCV, 2013, pp. 2304-2311.

[33] J. H. Yoon,M. H. Yang, J. Lim, et al, “Bayesian multi-object tracking using motion context from multiple objects,”Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. IEEE, 2015, pp. 33-40.

[34] L. Kratz , K.Nishino “Tracking Pedestrians Using Local SpatioTemporal Motion Patterns in Extremely Crowded Scenes,” Pattern Analysis & Machine Intelligence IEEE Transactions on, vol. 34, no.5, pp. 987-1002, 2011.

[35] A. Milan, K. Schindler, and S. Roth, “Detection- and trajectorylevel exclusion in multiple object tracking,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2013, pp. 3682–3689.

[36] A. Andriyenko and K. Schindler, “Multi-target tracking by continuous energy minimization,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2011, pp. 1265–1272.

[37] D. Riahi, G.A.Bilodeau, “Multiple object tracking based on sparse generative appearance modeling,” IEEE International Conference on Image ProcessingIEEE, 2015.

[38] X. Li, K. Wang, W. Wang, et al,“A multiple object tracking method using Kalman filter”IEEE International Conference on Information andAutomation. IEEE, 2010, pp. 1862-1866.

[39] A. Vatavu, R. Danescu, S.Nedevschi,“Stereovision-Based Multiple Object Tracking in Traffic Scenarios Using Free-Form Obstacle Delimiters and Particle Filters,”Intelligent Transportation Systems, IEEE Transactions on, vol. 16, no. 1, pp. 498-511,2015.

[40] B.Wu,R.Nevatia,“Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors,” International Journal of ComputerVision, 2007.