Online Multiple View Tracking: Targets Association Across Cameras

Quoc Cuong Le, Donatello Conte, Moncef Hidane

Online Multiple View Tracking: Targets Association Across Cameras

Quoc Cuong LE[1]
quoccuong.le@etu.univ-tours.fr
Donatello CONTE[1]
donatello.conte@univ-tours.fr
Moncef HIDANE[2]
moncef.hidane@insa-cvl.fr

[1] LIFAT
University of Tours,
Tours, France

[2] Computer Science Department
INSA Centre Val de Loire,
Blois, France

Abstract

Most multiple object tracking algorithms relying on a single view have failed to follow the trajectories of targets when they have been completely hidden by obstacles. In this paper, we introduce a novel method of collaborative tracking in a synchronized overlapping cameras network. We propose an efficient target association method between cameras based on the tracking results of each target on each view. Our framework naturally handles obstacle occlusions and mutual target occlusions.

We implemented our multiple object tracking algorithm by Decision Making algorithm [30] on each view. The tracking outcomes on each camera are collected and associated into targets. The feedback from the central association helps the individual cameras in tracking hidden targets, even in the case of complete occlusion. We use the standard MOT metric to validate our method. The experimental results on each view show that the multiple view tracking system outperforms the single view ones. The source code will be available publicly.

1 Introduction

Recent years have seen important improvements in single object tracking (SOT) algorithms, effectively conforming to real time and accuracy requirements. The huge challenge in SOT is to deal with an object whose shape and appearance might change in time. To adapt to target and environment changes, as well as to background clutter and mutual occlusion, almost all tracking methods developed update strategies such as model learning through the negative/positive samples around/inside target [1, 11, 34] or discriminative learning in the frequency domain such as correlation filters [12, 25].

Multiple Objects Tracking (MOT) is a related fundamental computer vision problem with a variety of real world applications such as visual surveillance, traffic monitoring, human identification, and autonomous driving. The main purpose of MOT is to determine the trajectories of a *number* of targets in a video. In almost all MOT applications, the objects of interest are either pedestrians or vehicles whose movement happens in a fixed background or a changing environment.

Implementing a SOT algorithm to track multiple targets simultaneously must face serious problems. First, MOT algorithms must handle the interaction between the targets or track the targets in crowded scenes, and usually, the targets slightly have the same appearance. These cause the single trackers to confuse their targets and their update mechanisms mostly fail since targets overlap with each other in a single view. Second, in MOT tracking videos, the targets are frequently occluded, partially or totally, by obstacles. Most of single tracking algorithm cannot handle this issue since the tracking result is basically based on the confidence score and a low score does not indicate whether it is an appearance change or an occlusion. Third, when the target has been hidden for a long time, it is considered as having disappeared of the scene and then reappeared. A MOT algorithm must then reconnect the target with the previous tracking results instead of initializing a new tracker. This issue can be considered as re-identification problem. Finally, a MOT algorithm has also to proceed to a multiple objects management to determine when and where the tracker should be terminated.

Modern MOT algorithms, especially those targeted at pedestrians tracking, followed a *tracking-by-detection* strategy. This is due to the efficiency of recent detectors which have proved their capability in detecting people with high precision rate in any kind of complex scenarios with several issues such as illumination changes and cluttered background. The next step then involves associating detections from different frames in order to obtain full objects trajectories. These association methods usually solve the problem by collecting the *entire* detections in the videos, leading to offline methods formulated as global graph optimization problems (multicut graph or k-bipartite graph problems).

However, offline methods do not fit the online and real-time requirements of many applications. Online matching methods generate the trajectories by using *causal* measures such as previous detections or last velocities. Recently, Markov Decision Process (MDP) [30] has been used to control the start/end or temporal appearance/disappearance of the targets. These events are treated as states in the MDP and make tracking multiple target more efficient. Notwithstanding, it still cannot handle cases in which the targets are completely hidden behind obstacles. Obviously, these information can not be figured out in a single view setting and it is then natural to resort to multiple overlapping cameras in order to increase robustness.

In this paper, we focus on dealing with complete occlusion problems and mutually occluded targets on a single view by setting up a multiple views tracking system. Our main contributions are: first, we extend the MDP algorithm to a multiple views framework and second, we introduce a novel targets association method across cameras. We want also to highlight that the proposed framework could also leverage existing online MOT methods.

The plan of the paper is as follows. We review in Sec. 2 related works pertaining to multiple object tracking in single and multiple view settings. Sec. 3 is dedicated to the presentation of our multiple view tracking approach which extends the MDP framework to a multi-view setting. The details concerning the proposed target association method allowing to link targets across different views are also presented therein. We assess the performance of our method by presenting experimental results on the PETS and EPFL datasets in Sec. 4 and we emphasize the significant advantage of multiple view tracking compared to the ones based on a single view. We finally conclude the paper in Sec. 5.

2 Related Works

2.1 Single View MOT

Following the tracking-by-detection paradigm [2, 31], graph optimization problems are formulated to link the detections of targets in order to form complete trajectories. These methods have become popular because they simplified the classic issues mentioned above such as object management, interaction, initialization and update strategy. The data association problem is formulated as a graph whose nodes represent the detections/features and the edges are weighted by the similarity between detections. The association methods usually collect all detections/features over the videos, the current position of a target (a node of the graph) being thus determined by adjacent nodes that represent the previous and next detections over the time. The goal of data association methods is to optimize the cost made by the edges of the graph. There are various methods such as global optimization using maximum clique problem [31], minimum cost subgraph multicut problem [14, 27], flow network optimization problems [32]. The problem is formulated in [2] as a k-shortest paths problem, as a subgraph decomposition in [26], and using a conditional random field (CRF) model in [4, 18]. In order to deal with the occlusion issue, most association methods proposed to introduce virtual nodes or hypothetical tracklets that represent the positions of a target on a single view during occlusion. However, these methods cannot fit to online application such as surveillance system.

Online methods determine the target's position by using the information of previous frames up to the current frame. Online re-identification based on sparse coding is used to match the current detections into the current tracklets [8]. Meanwhile, in [4], a CRF model is used to match a target to the best hypothesis among all combination of hypotheses created by new and existing targets. Despite these advances, hard occlusions in tracking still remains a serious problem. This motivates the development of our multiple view tracking algorithms.

It is also important to highlight that there are some MOT algorithms that are designed from a SOT perspective. We mention the CNN-based spatio-temporal attention mechanisms for online MOT introduced in [5], and the work of [30] that considers each state of the tracker as a state transition of a Markov Decision Process (MDP) thus allowing to deal with the birth/death and appearance/disappearance of targets. Our algorithm is heavily inspired from this latter framework.

2.2 Multiple Views MOT

MOT approaches based on a single view have been recently extended to multiple views. These approaches have been proposed in an attempt to fully cover the observation of the objects. Multi-camera tracking can solve the problem of occlusion where the interesting targets are frequently occluded by the environment or by the other objects. First attempts in using multiple cameras dealt with the re-identification problem, in order to track the objects between cameras [28]. Then, many other researchers studied the problem of collaborative tracking between the cameras. Almost all authors use the hypothesis that the exact positions of the cameras network is known and camera calibration has been done before applying tracking. In the tracking phase, the trackers of different views usually pool their tracking results on a 3-D coordinate system via the projection from image plane to ground plane in real world [17, 19, 23], so as to associate its results with the others, reconnect the missing trajectories or make their final result more stable.

According to the computational model used, there are two branches of multiple camera tracking methods: centralized algorithms and distributed ones. In centralized computing, data association is interpreted in the central node, with the purpose of connecting the incomplete trajectories from multiple views [16, 32, 33]. These approaches are generally suitable for offline tracking. On the other hand, several algorithms have used a distributed computational model in an attempt to create a probability map (see *e.g.* [10, 20]). Both distributed methods [7] and centralized ones [21] used a graphical model which relates the trackers on the different views and assumes some independence conditions between appearances on different views to induce the global observation likelihood of target given its occurrence on corresponding views. Similarly, [24] proposed the use of particle filters both in local and global configuration. The likelihood is then computed from all views based on a global appearance model. In the next section, we propose a novel framework for multiple views multiple objects tracking.

3 Proposed Framework

3.1 Targets Association Across Cameras

In this paper, the target association problem across cameras is formalized as an optimization problem involving an undirected weighted graph, as proposed in the paper [31]. We use the notation $G = (V, E, w)$, where V, E and w respectively denote the set of nodes, set of edges and weights of the edges. In our formulation, the "alive" targets (i.e. targets that are visible and in a "tracked" state in MDP) are considered as the set V of nodes. Let C_k be a cluster of tracking targets in the view k and v_m^k denote the m^{th} target within the view k. Therefore, $C_k = \{v_1^k, ..., v_M^k\}$. The edges of the graph are defined as $E = \{(v_m^k, v_n^l) | k \neq l\}$ with the condition $k \neq l$ indicating that two nodes in same camera cannot be connected. This also means that in each view, all targets are tracked separately. A node v_m^k is associated with its trajectory $\mathbf{x}_m^k = \{x_m^{k,1} ..., x_m^{k,F}\}$ which is the last appearance of a target in the previous frame, where the location records are the 2-dimensional coordinates (x, y) on the ground plane $z = 0$. The vector Φ_m^k denotes the patch surrounding the target m on the view k. The weight of an edge between two nodes is defined by the following equation:

$$w(v_m^k, v_n^l) = \alpha\, e_{medBF}(\Phi_m^k, \Phi_n^l) + f_{dist}(\mathbf{x}_m^k, \mathbf{x}_n^l) \tag{1}$$

where e_{medBF} is the Forward-Backward error defined in [13] to evaluate the similarity between a target in different views. In detail, by implementing similar ideas as in [30] through views, first we use the Lucas-Kanade (LK) tracker to find the correspondence points both between the target template Φ_m^k in the view k and Φ_n^l in the view l and between target template Φ_n^l in the view l and Φ_m^k in the view k. Then we calculate the Forward-Backward error from these two correspondences. In addition, the distance function f_{dist} is defined as the average distortion of the trajectories between two targets at the previous frame:

$$f_{dist}(\mathbf{x}_m^k, \mathbf{x}_n^l) = \frac{1}{L} \sum_{i=0}^{L-1} \left\| x_m^{k,F-i} - x_n^{l,F-i} \right\|, \tag{2}$$

where $L = \min\left(\min\left(|\mathbf{x}^k|, |\mathbf{x}_n^l|\right), 30\right)$. Moreover, the coefficient α allows us to regularize the contributions between Forward-Backward error and distance function.

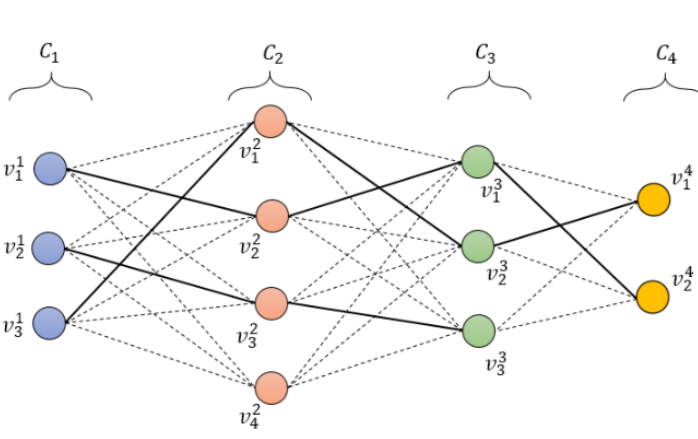


Figure 1: Finding the corresponding targets. For visualisation purpose, we draw only a subset of the edges within the subgraph indicating target across views, the others edges are ignored. In this example, there are three subgraphs indicating three people in the tracking process.

As mentioned in [31], the process of matching a target in different views requires identifying correspondences of the target in all different views. Hence, the solution of the problem can be described as a subgraph of G in which each node (target) is selected from only one cluster (view). Therefore, the subgraph for a particular tracked person can be denoted by $G_s = (V_s, E_s, w_s)$. The set of nodes V_s has a general form $V_s = \{v_m^k | k \in \{1,...,K\}\}$, $E_s = \{E(p,q) | p,q \in V_s\}$ and $w_s = \{w(p,q) | p,q \in V_s\}$. Fig. 1 shows an illustration of subgraph problem for associating the targets through views. To construct this subgraph, we search for nodes that are connected in different views by edges whose weight is not higher than a fixed threshold.

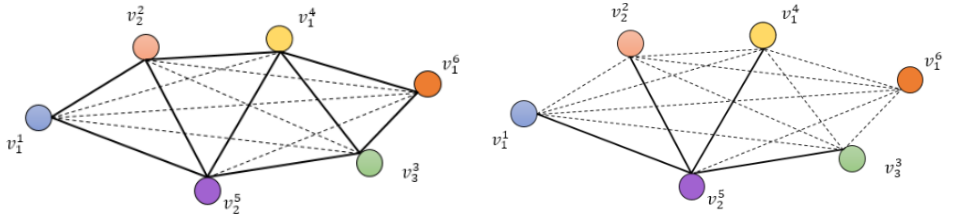
Let's chose a fixed threshold $M \in \mathbb{R}$, the optimal solution can be formulated as follows:

$$G_s = \left\{ V_s, E_s, w_s | \forall (v_p, v_q) \in E_s, w(v_p, v_q) \leq M \in \mathbb{R} \right\}. \tag{3}$$

The authors of [31] use the Generalized Minimum Clique Graph (GMCP) for matching association in time. Notwithstanding, GMCP does not fit to our problem because targets do not necessarily appear in all views, as opposed to the assumption done in [31]. In this paper, we do not focus on finding the optimal solution. Instead, we propose a novel fast algorithm to find an approximate solution. The details of this new algorithm are given in the next Section.

3.2 Proposed Algorithm

We propose a fast algorithm for finding the subgraphs satisfying the condition mentioned above. Instead of comparing all nodes, we fix a node v_0 in graph G and we include in the subgraph G_s all other nodes that are adjacent with v_0 and whose weight is under a fixed threshold. Obviously, our solution is sub-optimal, but the evaluation in terms of tracking performance done in Sec. 4 reveals that it is sufficient. The comparison between our approximate solution and the optimal one has been illustrated in the Fig. 2. The detail of the



(a) (b)

Figure 2: Comparison between the optimal solution (a) and our approximate solution (b). Note that node v_1^6 is included in the optimal solution, but not in the approximate one, because v_1^6 is not adjacent to the fixed node v_2^5.

proposed algorithm is described in Algorithm 1. In order to implement our targets association algorithm into a MOT algorithm, we introduce a MOT multiple views framework based on the MDP method [30]. The target association through views helps the system to recover targets on each view even in case of total or partial occlusion.

Partial occlusion cases. In the original method, the association step in the case of tracking failure will try to associate the target to a nearby detection. But it usually fails when target started being occluded, even if the detector still works well. We propose an extra step by adding the detection to the trajectory of target. The positions of all targets (nodes) in different views (subgraph) will be projected into the current image view and then the detection nearest to these positions will be assigned to the target in the current view, if the projected positions are not farther from the detected box than a fixed threshold.

Recovery after hard occlusion. The association strategy in the original MDP method cannot recover the tracking state of target when it has been hidden for a long time. We propose a recovery step to reconnect the newborn target to the lost target. The new detection will be compared with the targets' positions in different views. If its position on ground plane is near enough to the targets in all views, then the newborn detection will be added into the closest target. The detail of the method can be seen in Algorithm 2.

4 Experimental Results

This section presents the experimental result verifying the efficiency of our multiple views tracking method. To evaluate MOT performance, the benchmark MotChallenge [15] has been released with 2 datasets (MOT15 and MOT16), which contain a number of single view video sequences recorded by static or dynamic cameras, and the evaluation metrics of CLEAR MOT [3] and ID measures for MOT [22]. Additionally, the MotChallenge also provides a Multi-Target Multi-Camera Tracking benchmark [22] with mostly non-overlapping cameras. Unfortunately, these datasets do not fit to our problem, so for our benchmark we still used the same metric on MotChallenge, but replaced the dataset by the dataset PETS.

Implementation. We keep the same parameters for MDPs as the original paper [30], the others parameters in our approach are detailed in the above algorithms. For target association algorithm, we chose the view 1 as the dominant view k_0. Our proposed method is implemented in MATLAB on two cores CPU i7-6700HQ with 16GB RAM. Our algorithm processes approximately 0.5 fps for each view on average.

Input : Set of tracked/lost (MDP states) targets $\mathcal{T}^k = \{t_i^k\}_{i=1}^M$ in all views, a dominant view k_0
Output: Central $\mathcal{C}$ = set of subgraphs G_s indicating the tracked targets across views

Associated trackers $\mathcal{U}$ contain K empty sets corresponding to K views;
$\mathcal{C}$ contains M subgraphs G_s corresponding to M targets in view k_0;
foreach *Tracked/lost target* $t_i^{k_0}$ *in view* k_0 **do**
  foreach *each view* $j \neq k_0$ **do**
    Register $\leftarrow \emptyset$;
    foreach *Tracked/lost target* $t_{i'}^j$ *in view* j *and* $t_{i'}^j \notin \mathcal{U}^j$ **do**
      Weight of the edge connecting two targets $t_i^{k_0}$ and $t_{i'}^j$ (by eq. 1);
      Save the edge $(t_i^{k_0}, t_{i'}^j)$ and its weight to register;
    end
    Optimal cost $\leftarrow min(register)$;
    if *Optimal cost* < 6 & $f_{dist}(t_i^{k_0}, t_{i'}^j) < 1.5$ **then**
      Subgraph $\mathcal{C}^i \leftarrow \{\mathcal{C}^i, t_{i'}^j\}$;
      Associated trackers in view j : $\mathcal{U}^j \leftarrow \{\mathcal{U}^j, t_{i'}^j\}$;
    end
  end
end

Algorithm 1: Across views target association algorithm

Dataset. We used the well-known dataset PETS2009 [9] and the EPFL dataset (Multi-camera Pedestrian Videos) [2]. Among all sequences of PETS2009, the most relevant and suitable for our multiple views tracking system is "PETS09-S2L1" with 7 views from 7 synchronized and calibrated cameras. For our experiment, we only used 1 main view (from the camera 1) and 4 close-up views (from the cameras 5, 6, 7 and 8). Meanwhile, the EPFL dataset provided a multiple of video sequences recording the pedestrians indoor and outdoor by 4 cameras. Because of the similarity of sequence scenarios, we just selected the sequence "Terrace1" for our dataset. Fig. 3 shows that the observable zone on each view contains about 15-20% the common overlapping zone covered by all cameras. The ground truth and detection data on all views will be published on our project page (github.com/quoccuongLE/MDP_MTMC_tracking).

Detection. In all tracking-by-detection approaches, the detector plays an important role for tracking performance. As the same public detector used in MotChallenge [15], we applied the Aggregate Channel Features (ACF) pedestrian detector [6] on all views of the sequences "PETS09-S2L1" and "Terrace1" using the pre-trained Caltech model [29].

Evaluation metric. To validate the efficiency of our multiple views multiple object tracking method, we adopt the popular metrics used in MotChallenge including CLEAR MOT metric [3] and ID measures [22]. The metrics are the scores MOTA (multiple object tracking accuracy), MOTP (multiple object tracking precision), IDs (identity switches), IDF1, IDP (ID precision), IDR (ID Recall), False Positive (FP) and False Negative (FN). (For further details on the metric, we recommend the website https://motchallenge.net)

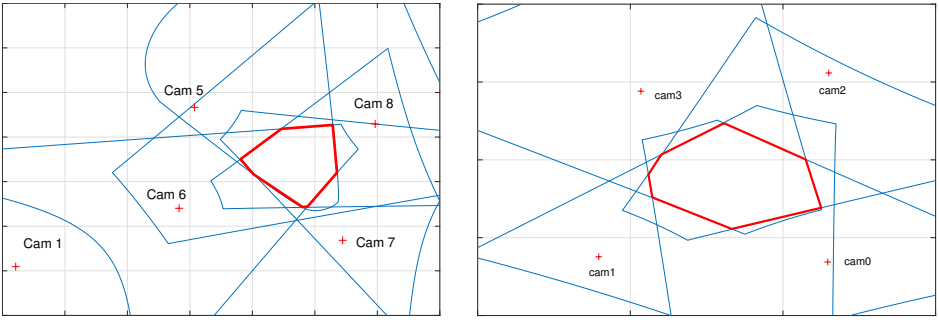
Performance analysis. We compare the MDP original method with our proposed MDP multi-view one. The overall tracking results on PETS sequence can be seen in the table

Input : Set of videos sequences from K views $v^1, \ldots, v^K$, object detection $\mathcal{D}_k = \{d_m^k\}_{m=1}^N$ for v_k, binary classifier (w^k, b^k) for data association
Output: Trajectories of targets $\mathcal{T}^k = \{t_i^k\}_{i=1}^M$ in the videos v^k

Initialization for all views: $\mathcal{T}^k \leftarrow \emptyset$
Initialize the central association node $\mathcal{C}$
`// main loop`
foreach *frame number l in videos* **do**
  foreach *each view j* **do**
    `// process targets in tracked states`
    foreach *tracked target t_i^k in $\mathcal{T}^k$* **do**
      Follow the policy, move the MDP of t_j^k to the next state;
    end
    `// process targets in lost states`
    foreach *tracked target t_i^k in $\mathcal{T}^k$* **do**
      Recover tracked state if found any similar detection covering the target;
      Add the nearest detection into target if reaching the agreement from other views;
    end
    Data association for the lost targets;
    foreach *lost target t_i in $\mathcal{T}^k$* **do**
      Follow the assignment, move the MDP of t_i^k to the next state;
    end
    Initialize the new targets from detection d_m^k not covered by any tracked target in $\mathcal{T}^k$;
    Connect the newborn tracklets $t_{i'}^k$ to the lost target t_i^k;
  end
  $\mathcal{C} \leftarrow$ targets association between views by Algorithm 1;
end

Algorithm 2: Multi-camera Collaborative Tracking based on MDP tracking

1. Primarily, our proposed method focuses on tracking targets in the hard occlusion case. It leads to an important reduction of identity switches and a significant improvement in ID measures. In detail, with only 15-20% overlapping zone, our approach on PETS sequence increased 17.5% in IDF1, 17.2% in IDP, 18,1% in IDR and reduced 31.8% ID switches. In terms of CLEAR MOT scores, our approach slightly improves both MOTA and MOTP scores. This can be explained by the fact that tracker's identification ability is not captured by the CLEAR MOT metric [22]. We also provide the tracking result on the main view (view 1) in the table 2 with the same consequences. On EPFL sequence, our method also remarkably improved the ID measure scores in the table 3. In contrast, the MOT scores slightly decreased, but these changes are not notable. The test on these two datasets has verified the robustness of our multi-view approach dealing with failure of re-identification and long-time occlusion. The visual effectiveness of our approach compared to the single view tracking system is shown in the Fig. 4. All the detail of the metric scores and the visualization of our entire experiment are provided in the complementary material.



(a) (b)

Figure 3: Observation zones of the cameras in PETS09-S2L1 sequence (a) and terrace1 sequence (b). The common overlapping zone has red contours. The element rectangle has a dimension of 5×5 meters in the real world

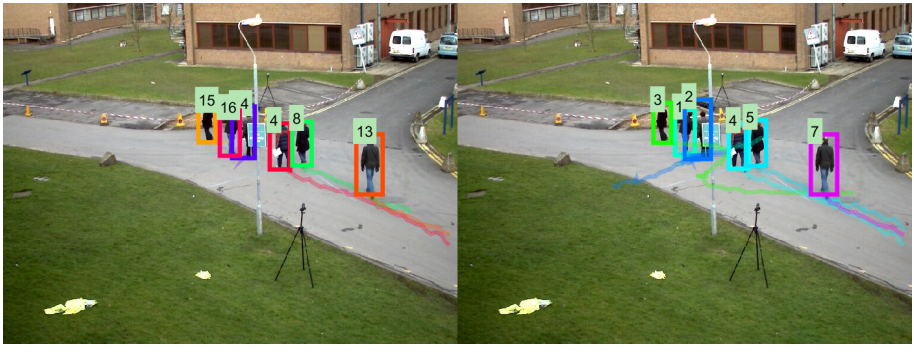


Figure 4: Tracking result at frame # 65 on view 1 of sequence "PETS09-S2L1". (Left) tracking result of MDP original method and (right) our tracking result.

Method	IDF1↑	IDP ↑	IDR↑	FP↓	FN↓	IDs↓	MOTA↑	MOTP↑
MDP standard	56.5	61.1	52.5	1255	3517	352	68.0	68.8
MDP multi-view	**66.4**	**71.6**	**62.0**	**1223**	**3379**	**240**	**69.8**	**68.9**

Table 1: Overall score of MOT metric on MDP standard method comparing with our MDP multi-view method on "PETS09-S2L1" sequence. Note that ↑ indicates better higher and ↓ better lower.

Method	IDF1↑	IDP ↑	IDR↑	FP↓	FN↓	IDs↓	MOTA↑	MOTP↑
MDP standard	57.8	59.6	56.1	**371**	635	95	75.4	72.2
MDP multi-view	**64.8**	**66.4**	**63.2**	372	**592**	**68**	**76.9**	**72.3**

Table 2: MOT metric scores on view 1 of MDP standard method and our MDP multi-view method on "PETS09-S2L1" sequence

5 Conclusion

In this paper, we presented a new robust online multi-view multi-object tracking method that naturally handles with the hard occlusion in tracking. Our framework is developed under the assumption that the multi-camera system is calibrated and synchronized. We ex-

Method	IDF1↑	IDP ↑	IDR↑	FP↓	FN↓	IDs↓	MOTA↑	MOTP↑
MDP standard	9.9	16.4	7.1	**727**	**14136**	**674**	**34.2**	**72.7**
MDP multi-view	**12.3**	**20.6**	**8.8**	741	14278	689	33.5	72.6

Table 3: Overall score of MOT metric on MDP standard method comparing with our MDP multi-view method on "Terrace1" sequence

tended the MDP tracking on multiple views framework and introduced a novel targets association method across views in order to track a multiple of targets collaboratively. We also showed the effectiveness of our method on the well-known sequences PETS2009 and the EPFL Multi-view Pedestrian Videos as well.

6 Acknowledgements

This work is part of the LUMINEUX project supported by a Region Centre-Val de Loire (France). We gratefully acknowledge Region Centre-Val de Loire for its support

References

[1] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie. Visual tracking with online multiple instance learning. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 983–990. IEEE, 2009.

[2] Jerome Berclaz, Francois Fleuret, Engin Turetken, and Pascal Fua. Multiple object tracking using k-shortest paths optimization. *IEEE transactions on pattern analysis and machine intelligence*, 33(9):1806–1819, 2011.

[3] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: the clear mot metrics. *Journal on Image and Video Processing*, 2008:1, 2008.

[4] Wongun Choi. Near-online multi-target tracking with aggregated local flow descriptor. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3029–3037, 2015.

[5] Qi Chu, Wanli Ouyang, Hongsheng Li, Xiaogang Wang, Bin Liu, and Nenghai Yu. Online multi-object tracking using cnn-based single object tracker with spatial-temporal attention mechanism. In *2017 IEEE International Conference on Computer Vision (ICCV).(Oct 2017)*, pages 4846–4855, 2017.

[6] Piotr Dollár, Ron Appel, Serge Belongie, and Pietro Perona. Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36 (8):1532–1545, 2014.

[7] Wei Du and Justus Piater. Multi-view object tracking using sequential belief propagation. *Computer Vision–ACCV 2006*, pages 684–693, 2006.

[8] Loïc Fagot-Bouquet, Romaric Audigier, Yoann Dhome, and Frédéric Lerasle. Improving multi-frame data association with sparse representations for robust near-online multi-object tracking. In *European Conference on Computer Vision*, pages 774–790. Springer, 2016.

[9] J Ferryman and A Shahrokni. Pets2009: Dataset and challenge. In *Performance Evaluation of Tracking and Surveillance (PETS-Winter), 2009 Twelfth IEEE International Workshop on*, pages 1–6. IEEE, 2009.

[10] Francois Fleuret, Jerome Berclaz, Richard Lengagne, and Pascal Fua. Multicamera people tracking with a probabilistic occupancy map. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):267–282, 2008.

[11] Helmut Grabner, Michael Grabner, and Horst Bischof. Real-time tracking via on-line boosting. In *Bmvc*, volume 1, page 6, 2006.

[12] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596, 2015.

[13] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. *IEEE transactions on pattern analysis and machine intelligence*, 34(7):1409–1422, 2012.

[14] Margret Keuper, Siyu Tang, Yu Zhongjie, Bjoern Andres, Thomas Brox, and Bernt Schiele. A multi-cut formulation for joint segmentation and tracking of multiple objects. *arXiv preprint arXiv:1607.06317*, 2016.

[15] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler. MOTChallenge 2015: Towards a benchmark for multi-target tracking. *arXiv:1504.01942 [cs]*, April 2015. URL http://arxiv.org/abs/1504.01942. arXiv: 1504.01942.

[16] Yuan Li, Chang Huang, and Ram Nevatia. Learning to associate: Hybridboosted multi-target tracker for crowded scene. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2953–2960. IEEE, 2009.

[17] Ivana Mikic, Simone Santini, and Ramesh Jain. Video processing and integration from multiple cameras. In *Proceedings of the 1998 Image Understanding Workshop, Morgan-Kaufman, San Francisco*, volume 6, 1998.

[18] Anton Milan, Laura Leal-Taixé, Konrad Schindler, and Ian Reid. Joint tracking and segmentation of multiple targets. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 5397–5406. IEEE, 2015.

[19] Roman Pflugfelder and Horst Bischof. Localization and trajectory reconstruction in surveillance cameras with nonoverlapping views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):709–721, 2010.

[20] Horst Possegger, Sabine Sternig, Thomas Mauthner, Peter M Roth, and Horst Bischof. Robust real-time tracking of multiple objects by volumetric mass densities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2395–2402, 2013.

[21] Wei Qu, Dan Schonfeld, and Magdi Mohamed. Distributed bayesian multiple-target tracking in crowded environments using multiple collaborative cameras. *EURASIP Journal on Applied Signal Processing*, 2007(1):21–21, 2007.

[22] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision*, pages 17–35. Springer, 2016.

[23] Aswin C Sankaranarayanan, Ashok Veeraraghavan, and Rama Chellappa. Object detection, tracking and recognition for multiple smart cameras. *Proceedings of the IEEE*, 96(10):1606–1624, 2008.

[24] Santhoshkumar Sunderrajan and Bangalore S Manjunath. Multiple view discriminative appearance modeling with imcmc for distributed tracking. In *Distributed Smart Cameras (ICDSC), 2013 Seventh International Conference on*, pages 1–7. IEEE, 2013.

[25] Ming Tang, Bin Yu, Fan Zhang, and Jinqiao Wang. High-speed tracking with multi-kernel correlation filters. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[26] Siyu Tang, Bjoern Andres, Miykhaylo Andriluka, and Bernt Schiele. Subgraph decomposition for multi-target tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5033–5041, 2015.

[27] Siyu Tang, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. Multi-person tracking by multicut and deep matching. In *European Conference on Computer Vision*, pages 100–111. Springer, 2016.

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

[29] Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian detection: A benchmark. In *in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. Citeseer, 2009.

[30] Yu Xiang, Alexandre Alahi, and Silvio Savarese. Learning to track: Online multi-object tracking by decision making. In *2015 IEEE international conference on computer vision (ICCV)*, number EPFL-CONF-230283, pages 4705–4713. IEEE, 2015.

[31] Amir Roshan Zamir, Afshin Dehghan, and Mubarak Shah. Gmcp-tracker: Global multi-object tracking using generalized minimum clique graphs. In *Computer Vision–ECCV 2012*, pages 343–356. Springer, 2012.

[32] Li Zhang, Yuan Li, and Ramakant Nevatia. Global data association for multi-object tracking using network flows. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[33] Shu Zhang, Yingying Zhu, and Amit Roy-Chowdhury. Tracking multiple interacting targets in a camera network. *Computer Vision and Image Understanding*, 134:64–73, 2015.

[34] Wei Zhong, Huchuan Lu, and Ming-Hsuan Yang. Robust object tracking via sparsity-based collaborative model. In *Computer vision and pattern recognition (CVPR), 2012 IEEE Conference on*, pages 1838–1845. IEEE, 2012.