Online Inter-Camera Trajectory Association Exploiting Person Re-Identification and Camera Topology

Na Jiang, SiChen Bai, Yue Xu, Chang Xing, Zhong Zhou, Wei Wu
State Key Lab of Virtual Reality Technology and Systems, Beihang University, Beijing, China
{jiangna,bhbaisichen,xuyuevr,xingchang,zz,wuwei}@buaa.edu.cn

ABSTRACT

Online inter-camera trajectory association is a promising topic in intelligent video surveillance, which concentrates on associating trajectories belong to the same individual across different cameras according to time. It remains challenging due to the inconsistent appearance of a person in different cameras and the lack of spatio-temporal constraints between cameras. Besides, the orientation variations and the partial occlusions significantly increase the difficulty of inter-camera trajectory association. Targeting to solve these problems, this work proposes an orientation-driven person re-identification (ODPR) and an effective camera topology estimation based on appearance features for online inter-camera trajectory association. ODPR explicitly leverages the orientation cues and stable torso features to learn discriminative feature representations for identifying trajectories across cameras, which alleviates the pedestrian orientation variations by the designed orientation-driven loss function and orientation aware weights. The effective camera topology estimation introduces appearance features to generate the correct spatio-temporal constraints for narrowing the retrieval range, which improves the time efficiency and provides the possibility for intelligent inter-camera trajectory association in large-scale surveillance environments. Extensive experimental results demonstrate that our proposed approach significantly outperforms most state-of-the-art methods on the popular person re-identification datasets and the public multi-target, multi-camera tracking benchmark.

KEYWORDS

Inter-Camera Trajectory Association, Person Re-identification, Camera Topology Estimation

ACM Reference Format:
Na Jiang, SiChen Bai, Yue Xu, Chang Xing, Zhong Zhou, Wei Wu. 2018. Online Inter-Camera Trajectory Association Exploiting Person Re-Identification and Camera Topology. In *2018 ACM Multimedia Conference (MM '18), October 22–26, 2018, Seoul, Republic of Korea*, Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3240508.3240663

MM '18, October 22–26, 2018, Seoul, Republic of Korea

ACM ISBN 978-1-4503-5665-7/18/10…$15.00
https://doi.org/10.1145/3240508.3240663

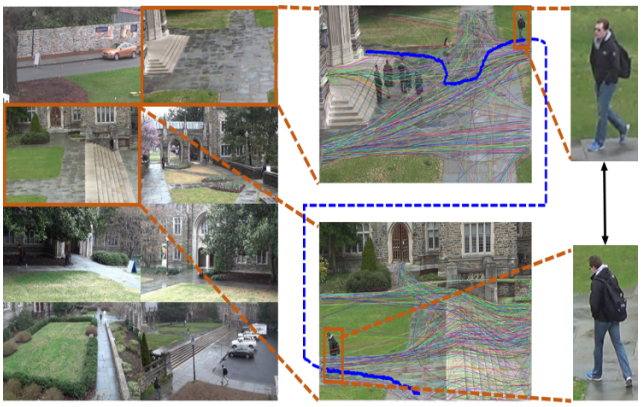


Figure 1: An illustration of inter-camera trajectory association on DukeMTMC dataset.

1 INTRODUCTION

Online inter-camera trajectory association is an important fundamental problem in computer vision, especially for intelligent video surveillance system. It focuses on generating complete trajectories of multiple identities across cameras according to the travel time. As the number of surveillance videos increases, the traditional inter-camera trajectory association based on manual analysis requires revolution due to its high labor cost, low efficiency and accuracy. We are thus strongly motivated to look for ways to improve efficiency, robustness, and accuracy of inter-camera trajectory association, which is the focus of this paper.

As a promising topic, multi-object multi-camera tracking for intelligent monitoring and analysis has recently attracted increasing attentions in academia and industry [18, 23, 24, 28, 31]. Abundant works have greatly fostered the development of multi-object tracking in single camera, but the online inter-camera trajectory association for object handover across cameras remains to explore. It is challenging due to inconsistent appearance of a person in different cameras and the missing of spatio-temporal constraints between cameras. As illustrated in Fig.1, the solid blue curves depict the trajectories of the same pedestrian in two different cameras while other curves denote trajectories of others. The unpredictable orientation variations of the pedestrians and amounts of blind-gaps in real surveillance videos significantly increase the complexity of the online inter-camera trajectory association. Targeting to solve these problems, the discriminative feature representations and accurate spatio-temporal constraints between cameras become the critical cues.

The appearance features of the pedestrians extracted from deep learning architecture have achieved significant improvements in

person re-identification. Inspired by it, the deep learning-based person re-identification is exploited to learn appearance features for similarity metric of trajectories across cameras. Considering that orientation variations and occlusions are the main reasons for causing inconsistent appearance, we propose an orientation-driven person re-identification (ODPR) to extract discriminative appearance features. Different from most existing methods [6, 30, 35] that focus on designing the feature extraction modules, the orientation-based similarity constraint is considered in ODPR. We introduce the pedestrian orientations to design an orientation-driven loss function for training the feature extraction architecture and estimate the orientation aware weights for measuring the trajectory similarity, which impose the orientation-based similarity constraint to alleviate the orientation variations. Meanwhile, the appearance features are generated with the combination of global features and stable torso features, which are conducive to improving the discriminative feature representations in the face of partial occlusions and inaccurate detections.

For the missing of spatio-temporal constraints, this paper chooses the camera topology to provide the accurate spatio-temporal constraints for the online inter-camera trajectory association. It consists of the connections and the transfer time between cameras, which can eliminate the huge redundancy for each probe trajectory. Most previous methods [4, 20] of camera topology estimation accumulate amounts of incorrect associations caused by inconsistent appearance, which make it impossible to obtain accurate transfer time from initial statistical distribution. The camera topology without accurate transfer time cannot provide the effective temporal constraints for inter-camera trajectory association. To solve the problem, the appearance features from ODPR is introduced to reduce incorrect associations and a nearest neighbor accumulation strategy is employed to optimize the initial statistical distribution. The cooperation makes the connectivity and the transfer time between cameras more clear, which is beneficial to narrow the retrieval range of trajectory association and improve the time efficiency.

In summary, an inter-camera trajectory association framework exploiting person re-identification and camera topology estimation is presented in this paper. Assuming that the pedestrians who disappear from a camera will enter a connected camera within the corresponding transfer time range, according to the priori known that any target in the three-dimensional world will only appear in one spatial position at a point in time. The spatio-temporal constraints provided by camera topology are used to generate the corresponding gallery trajectory set for each probe trajectory that leaving from any camera. The appearance features extracted from the ODPR are exploited to identify the identities of trajectories between probe set and gallery set. The incorporation improves the efficiency and accuracy of online inter-camera trajectory association, which provides the important foundation for inter-camera trajectory association extended to large-scale surveillance environments.

We demonstrate the effectiveness of the proposed online inter-camera trajectory association, referred as TAREIDMTMC, using challenging DukeMTMC dataset [23] on MOT benchmark [1]. Extensive experimental results on person re-identification datasets and the MOT benchmark demonstrate that our proposed approach is superior to most state-of-the-art methods.

2 RELATED WORK

Inter-camera trajectory association can be classified into two categories: with overlapping field of views [12, 17] and without overlapping field of views [18, 23, 24, 28, 31]. Inter-camera trajectory association between cameras with overlapping field of views is relatively simple due to the existence of the handover zones between different cameras. Inter-camera trajectory association among non-overlapped cameras is very complicated because of the open blind areas. For the real-world monitoring environments, non-overlapped cameras are more ubiquitous than cameras with overlapping views. Therefore, we mainly review the existing trajectory association approaches with non-overlap field of views in the section.

A large amount of works have been proposed to improve inter-camera trajectory association [5, 13, 29]. Ristani et al. publish the DukeMTMC dataset [23] and design appearance features [24] to accelerate progress in multi-object multi-camera tracking. Maksai et al. [18] propose a Non-Markovian approach to imposing global consistency and behavioral patterns to guide the multi-object tracking. Zhang et al. [34] introduce re-ranking and hierarchical clustering to realize multi-target multi-object tracking on DukeMTMC project. Yoon et al. [31] form multiple track-hypothesis trees to solve the multi-target multi-camera tracking. Zhang et al. [33] regard multi-camera tracking as network flow problems and utilize spatio-temporal group events to analyze the trajectory association. These efforts which focus on the appearance cues or the spatio-temporal cues have fostered the development of inter-camera trajectory association. If integrating the appearance cues with the spatio-temporal cues, the performance of algorithms will be further optimized. Therefore, this paper focuses on extracting the discriminative feature representations and estimating camera topology, which explicitly leverages the appearance cues and the spatio-temporal cues to improve the inter-camera trajectory association.

To exploit the appearance cues, Porikli et al. [22] derive a color distortion function for pair-wise camera correlation analysis. Javed et al. [11] design brightness transfer functions to optimize color descriptors. Extracting these color features and hand-crafted features is simple and convenient, whereas the discrimination of these features will be weakened when pedestrian orientations or monitoring environments change among different cameras. To enhance the discrimination of appearance features, the person re-identification [15, 16, 36, 38, 39] is introduced to identify the trajectories from different cameras [23, 29]. The selected methods and strategies are mostly based on the deep learning architecture. For example, Li et al [15] propose the Deepreid that learns to encode photometric transforms between the filter pairs. Zhong et al. [39] design the k-reciprocal encoding using the Jaccard distance to re-rank the re-identification results. Zheng et al. [38] exploit generative adversarial networks (GANs) to generate unlabeled data for data augmentation. They contribute to re-identification from different perspectives, but they neglect the crucial orientation factor that has significant influence on the consistency of appearance. This work sheds a new light on the exploiting of the orientation, and propose an orientation-driven person re-identification to learn appearance features and similarity metric.

Except the appearance cues, some spatio-temporal cues have also been also applied to the inter-camera trajectories association across

non-overlapping views [3, 40]. Dick et al. [7] propose a stochastic transition matrix to describe pedestriansḿotion patterns as the spatio-temporal constraints. Chen et al. [3] analyze path probabilities of objects as the spatio-temporal cues to implement the offline trajectory association. Niu et al. [21] combine the normalized color and overall size to count observed persons for the camera topology estimation. Zehavit et al. [19] divide the moving objects into blocks and refine the relationship between blocks belong different cameras. These spatio-temporal constraints improve the offline/online inter-camera trajectory association, nonetheless, they remains to be further optimized due to the inaccurate detections and the fragmented trajectories. They are also required to retrain when the monitoring layout changes. To avoid these troubles, the camera topology with the accurate transfer time is estimated to provide the spatio-temporal constraints. The transfer time is helpful to improve the time efficiency of inter-camera trajectory association. And the estimated camera topology only need update the changed cameras in the face of the monitoring layout variation.

3 OUR APPROACH

In this paper, a novel inter-camera trajectory association exploiting person re-identification and camera topology is presented. As shown in Fig.2, it consists of single camera tracking, person re-identification, camera topology estimation and inter-camera trajectory association. Given any video, an improved online multi-object tracking is conducted on each frame to generate the trajectories in single camera. Then the appearance features of each images in the achieved trajectories are extracted from the proposed ODPR. According to the estimated camera topology, the redundant trajectories that do not satisfy the spatio-temporal constraints are eliminated. For those trajectories or bounding boxes that are in the connected cameras and meet the transfer time constraints, the trajectory similarity is computed to implement the inter-camera trajectory association.

3.1 Formulation of Inter-Camera Trajectory Association

Given a probe trajectory T_p, the inter-camera trajectory association to identify and return trajectories containing the identical identity in T_p from a set of gallery trajectories $T_G = \{T_{g_1}, T_{g_2}, ..., T_{g_i}, ..., T_{g_N}\}$, where T_{g_i} denotes the i-th gallery trajectory. The inter-camera trajectory association can be tackled by the global trajectory retrieval. As the number of the gallery trajectories increases, the time efficiency of the global retrieval mode will drop significantly, and the accuracy will also decrease due to the noise from the large-scale gallery set. To overcome these drawbacks, we make the following improvements.

Firstly, the candidate set $T_C = \{T_{c_1}, T_{c_2}, ..., T_{c_M}\}$ is generated from the T_G according to the spatio-temporal constraints provided by the camera topology. The camera where any candidate trajectory T_{c_i} is located has the connection with the camera of T_p. The travel time interval between any T_{c_i} and T_p is close to transfer time between the corresponding entry/exit zones. The spatio-temporal constraints are defined as follows.

$$Cont(T_p, T_{g_i}) = \begin{cases} 1, & connected \\ 0, & otherwise \end{cases} \tag{1}$$

$$S_{time}(T_p, T_{g_i}) = \begin{cases} 1, & Time(T_p, T_{g_i}) \geq \delta \\ 0, & Time(T_p, T_{g_i}) < \delta \end{cases} \tag{2}$$

$$Time(T_p, T_{g_i}) = exp(-(Interval(T_p, T_{g_i}) - \mu)^2 / 2\sigma^2) \tag{3}$$

$$Interval(T_p, T_{g_i}) = max(T_p^s, T_{g_i}^s) - min(T_p^e, T_{g_i}^e) \tag{4}$$

where $Cont(T_p, T_{g_i})$ represents the connection between the cameras where T_p and T_{g_i} are located, $S_{time}(T_p, T_{g_i})$ demonstrates whether the time interval between T_p and T_{g_i} meets the transfer time constraint. $Time(T_p, T_{g_i})$ indicates the score of the transfer time between two trajectories. δ is an elastic threshold of limiting the number of trajectories that meet transfer time constraint, which is empirically set to 0.7. $Interval(T_p, T_{g_i})$ denotes the time interval between T_p and T_{g_i}. The superscript s represents the start time of trajectory, and e denotes the end time. If $Interval(T_p, T_{g_i})$ is negative, the two trajectories are overlapping.

As shown in Eq.(1), when the cameras where T_p and T_{g_i} are located have the connections, $Cont(T_p, T_{g_i})$ is set to 1. In contrast, $Cont(T_p, T_{g_i})$ is equal to 0. For the connected entry/exit zone pairs, μ denotes the transfer time and σ means the standard deviation of statistical distribution. We assume that the statistics of associated trajectories with different time intervals is subject to normal distribution $N(\mu, \sigma^2)$. Hence, the score of the transfer time between two trajectories can be calculated through the probability density function $f(x) = exp(-(x-\mu)^2/2\sigma^2)/\sqrt{2\pi}\sigma$. To normalized the score, the results from the probability density function need to be divided by $f(\mu)$. The derivative formula of the transfer time score is defined in Eq.(3). If $Time(T_p, T_{g_i})$ of the trajectory T_{g_i} is greater than or equal to the threshold δ, the trajectory T_{g_i} in the connected camera of T_p is considered to meet the transfer time constraints. According to Eq.(1) ~ Eq.(4), a large number of trajectories that have not spatio-temporal relationship with the T_p are eliminated.

Secondly, we calculate the trajectory similarity $Sim(T_p, T_{c_i})$ between T_p and T_{c_i}, which is the major evidence of the trajectory association. It depends on the feature distance between images and is easily influenced by the orientation variations of a pedestrian across cameras. In this case, feature distances between images with the same orientation are more important for the trajectory similarity than those between images with the different orientations. Therefore, the orientation aware weights described in Eq.(6) are introduced to make the trajectory similarity calculated in Eq.(5) more reasonable and reliable. The trajectory similarity with orientation aware weights $Sim(T_p, T_{c_i})$ is formulated as following:

$$Sim(T_p, T_{c_i}) = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} w_{I_p^x I_{c_i}^y} d(f(I_p^x), f(I_{c_i}^y))}{\sum_{x=1}^{M} \sum_{y=1}^{N} w_{I_p^x I_{c_i}^y}} \tag{5}$$

$$W_{I_p^x I_{c_i}^y} = \begin{cases} w_{fs}, & (I_p^x, I_{c_i}^y) \in \{fs, sf\} \\ w_{fb}, & (I_p^x, I_{c_i}^y) \in \{fb, bf\} \\ w_{bs}, & (I_p^x, I_{c_i}^y) \in \{bs, sb\} \\ w_{same}, & (I_p^x, I_{c_i}^y) \in \{bb, ff, ss\} \end{cases} \tag{6}$$

where M is the number of images in T_p, and N is the number of images in T_{c_i}. I_p^x represents the x-th image of T_p, and $I_{c_i}^y$ represents the y-th image in T_{c_i}. $f(I)$ indicates the apparent eigenvector of the image I. $d(x, y)$ represents the L2-norm distance between x and y. W denotes the orientation-aware weights, which are achieved

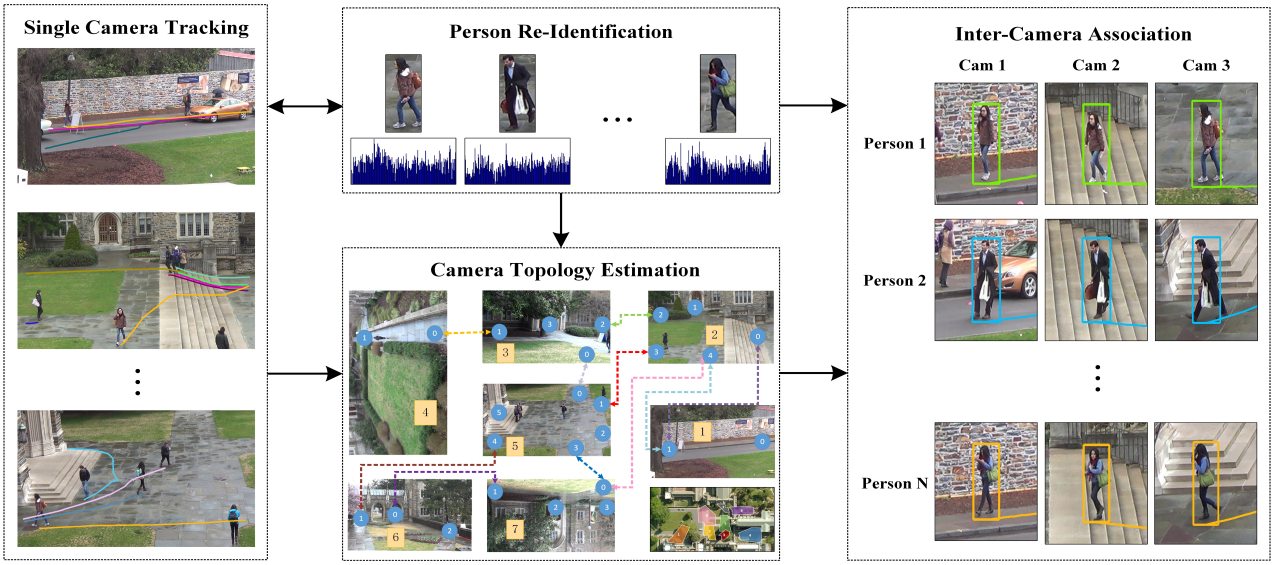


Figure 2: Overview of the proposed approach.

by a two-layer fully connected shallow network. The f, s, and b represent the front, side, and back, respectively.

Finally, we achieve the list of trajectories according to these $Sim(T_p, T_{c_i})$, which is denoted as $T_R = \{T_{r_1}, T_{r_2}, \ldots, T_{r_L}\}$. Considering the uniqueness of pedestrian in the three-dimensional world, the T_{r_i} with the highest similarity in each connected camera is identified as the inter-camera trajectory association of T_p. The objective function of our online inter-camera trajectory association is summarized as Eq.(7).

$$T^* = argmax_{i \in 1,2,\ldots,N}(Cont(T_p, T_{g_i}) \cdot S_{time}(T_p, T_{g_i}) \cdot exp(-Sim(T_p, T_{g_i})) > \tau) \quad (7)$$

where T^* denotes the associated trajectory of T_p along the time line. τ is a decision threshold of trajectory association. In our learned feature space through densely connected blocks, it is set to 0.65.

3.2 Orientation-Driven Person Re-Identification

The orientation variations and the frequent occlusions are the main reasons for leading to the inconsistent appearance of the same person, hence we propose an orientation-driven person re-identification (ODPR) architecture with local features to overcome them. The architecture introduces the pose estimation to implement orientation estimation, designs orientation-driven loss function, and achieve orientation aware weights. These orientation cues can alleviate the orientation variations. Meanwhile, the stable torso features are extracted and fused to enhance the generalization of appearance features. The ODPR architecture is summarized in Fig.3.

As shown in Fig.3, the orientations of pedestrians are divided into three categories: side, front, and back. On the basis, the negative sample, the positive sample with the same orientation and the positive sample with different orientation of every anchor image can be selected for training. Besides, two weighted loss functions are exploited to drive ODPR architecture learning discriminative

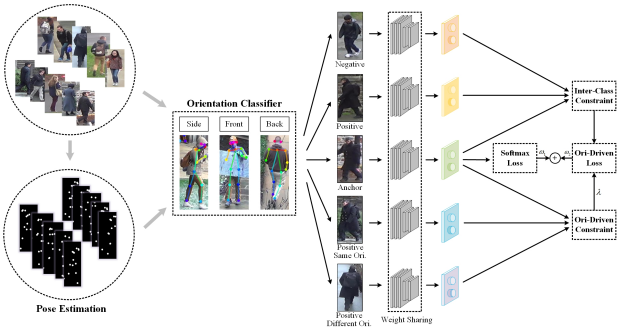


Figure 3: Architecture of orientation-driven person re-identification (ODPR). ω_1 and ω_2 denote the loss weights between softmax loss and orientation-driven loss. They are set to 1 and 0.5, respectively.

feature representations. The weighted loss functions include the softmax loss function and the proposed orientation-driven loss function. The softmax loss is only connected to central anchor branch, and all branches share weights. The expatiations of orientation estimation, feature extraction and similarity metric with orientation are described below.

3.2.1 Orientation Estimation and Feature Extraction. Pose estimation that detects the body joints plays an important role in the person re-identification. It is necessary for our proposed ODPR in term of estimating the pedestrian orientations and extracting local features.

1) *Orientation Estimation.* The designed orientation-driven loss function requires datasets to be divided into three subsets of the front, back and side, which prepare for constituting training inputs. To meet this demand, we firstly utilize part affinity fields (PAFS) [2] to generate the color skeletons based on body joints, then exploit

raw images with the color skeleton to learn the orientation classifier. As shown in Fig.3, the skeleton connections between different joints are labeled by different color segments, which is beneficial to learn the position relationships between symmetrical joints. In the training phase of the orientation estimation, DenseNet121 [10] is directly selected as the feature extraction network and the RAP dataset [14] that provides the orientation annotations is chosen as the training set. On the test set of RAP dataset, the accuracy of orientation estimation achieves 87.33%. Although the accuracy of orientation estimation does not achieve 100%, it is sufficient to help the ODPR learning the influence of orientation variations on apparent consistency through the orientation-driven loss function.

2) *Feature Extraction.* In ODPR, we extract torso features to assist the global features since the torso region is relatively stable and the limbs have continual movements. The locations of the torso local regions of interest (ROIs) are inferred by the body joints detected by the pose estimation, and then are transformed into ROI Pooling layer for local feature extraction. The local branch and the global backbone share the parameters on the stem CNNs. And the subsequent feature extraction modules can use any convolutional structure, especially residual blocks [9], inception blocks [27], or densely connected blocks [10].

3.2.2 Similarity Metric with Orientation. For the training of deep learning, loss functions are the primary tools of classification and similarity metric learning. The triplet loss function is widely used in person re-identification. It is known as the inter-class similarity constraint and teaches the network that the positive samples should have smaller feature distances than negative samples. However, the influence of the orientation variations on the appearance features dose not be considered in the triplet loss function. To make up for this deficiency, an orientation-driven loss function $L_{od}(I, w)$ is designed on the basis of triplet loss function. It consists of inter-class similarity constraint based on triplet loss function and the introduced orientation-driven constraint.

$$L_{od}(I,w) = \frac{1}{N}(\sum_{i=1}^{N}[d(f(I_i^a), f(I_i^p)) - d(f(I_i^a), f(I_i^n)) + \alpha]_+ + \lambda \sum_{i=1}^{N}[d(f(I_i^a), f(I_i^{ps})) - d(f(I_i^a), f(I_i^{pd})) + \beta]_+) \tag{8}$$

where I_i^a denotes the anchor image, I_i^p demonstrates the positive sample of anchor image, I_i^n indicates the negative sample of the anchor image, I_i^{ps} refers to positive sample of the anchor image with the same orientation, and I_i^{pd} denotes the positive sample of the anchor image with different orientations. α and β are the limit margin. $[x]_+ = max(x, 0)$. N is depending on the settings of batchsize and max iteration. λ is a weight of balancing the two constraints, and w represents the current network parameters. The fusion loss function defined in Eq.(8) can pull the instances of the identical person with same orientation closer, and meanwhile push the instances belonging to different persons with same orientation farther from each other. According to the comparative analysis of multiple experiments, the overall performance of ODPR is superior when α, β, and λ are set to 1, 0.01, and 0.03, respectively.

Although the pose effects have been considered in feature learning and training, the orientation aware weights W defined in Eq.(6) still play an important role in test phase. To learn the weights, we firstly collect training samples from DukeMTMC dataset. We regard the multiple trajectories of identical person from different cameras as a complete trajectory and then randomly select the bounding boxes with complete posture skeleton and various orientation to compose a training sample. Every training sample contains 12 bounding boxes. For any two training samples, In_{fs-sf}, In_{fb-bf}, In_{bs-sb}, and In_{same} will be obtained as four inputs. Every input In is saved as $\{v, l\}$. Taking In_{fs-sf} as an example, the v denotes the mean value of feature distances between all image pairs belong to $\{front\text{-}side, side\text{-}front\}$. If there are one/some image pair(s) belong to $\{front\text{-}side, side\text{-}front\}$ between two samples, the l is set to 1. Otherwise l is marked as 0. Push all inputs into a two-layer shallow network for training. Then the pairwise feature distance $Sim_{train}(T_i, T_j)$ between two training samples is formulized as:

$$Sim_{train}(T_i, T_j) = \frac{\sum_{x=1}^{4} l_x w_x v_x}{\sum_{x=1}^{4} l_x w_x} \tag{9}$$

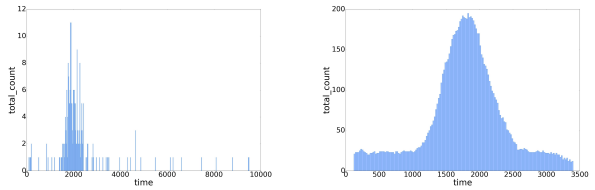
where w_1, w_2, w_3, and w_4 denote w_{fs}, w_{fb}, w_{bs}, and w_{same}, respectively. We hope that the orientation aware weights cam make $exp(-Sim_{train}(T_i, T_j))$ of the positive samples be mapped to 1 through the activation layer and meanwhile guide the ones of negative samples to 0. On DukeMTMC training set, $[w_{fs}, w_{fb}, w_{bs}, w_{same}]$ = [0.72, 1.21, 0.69, 2.82]. The w_{fs} is close to w_{bs}, and is slightly lower than w_{fb}. Meanwhile, w_{same} is assigned the highest value. Their values reflect the importance of different pose pairs in similarity metric, which alleviates the negative impact of orientation variations on trajectory similarity.

3.3 Appearance-Based Camera Topology Estimation

To achieve the effective spatio-temporal constraints, we propose an appearance-based camera topology estimation. The Gaussian Mixture Model (GMM) is used to cluster the entry/exit zones. Then the trajectories through each entry/exit zone are identified by appearance features to estimate the connectivity and the transfer time between every two entry/exit zones. Due to the occlusions and the inaccurate detections, the time intervals of different persons across the same entry/exit zones have obvious differences. Therefore, the initial statistical distribution according to the time intervals is very discrete, and the accurate transfer time between entry/exit zones cannot be obtained. To make the statistical results suitable for the camera topology estimation with accurate transfer time, the nearest neighbor accumulation strategy is exploited to filter the noise and optimize the distribution. The accumulated equation between entrance E_i and exit E_j is defined as follows:

$$Count_p(E_i, E_j) = \sum_{\eta_0=\eta_n-\xi}^{\eta_n+\xi} N^{ij}(\eta_0), \eta_n > \xi \tag{10}$$

where $Count_p(E_i, E_j)$ represents the optimized statistical results between E_i and E_j. $N^{ij}(\eta_0)$ denote the number of associated trajectories with time interval η_0. $[\eta_n - \xi, \eta_n + \xi]$ indicates the cumulative



(a) Original Statistics (b) Accumulated Statistics

Figure 4: An illustration of nearest neighbor accumulation on DukeMTMC between camera1E1 and camera2E0. The abscissa indicates the time interval, and the ordinate represents the statistical number of associated trajectories with different time intervals. Fig.4(a) shows the original statistical results between camera1E1 and camera2E0. Fig.4(b) shows the optimized $Count_p(E_i, E_j)$ by the nearest neighbor accumulation.

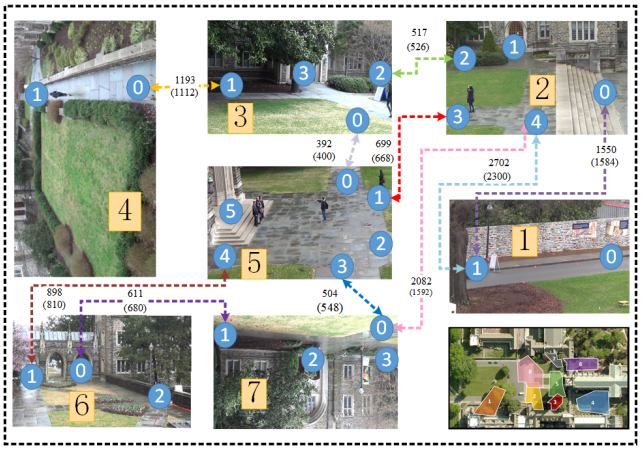


Figure 5: Our estimated camera topology of DukeMTMC dataset.

range, which is determined by video resolution and pedestrian density. After the cumulative calculation, the count value equal to 1 is eliminated as noise. An example of our nearest neighbor accumulation is shown in Fig.4, which ξ is empirically set to 5.

As shown in Fig.4, the original data with discrete distribution cannot be used directly to infer the transfer time and connectivity between cameras. But the accumulated data with obvious peak value can estimate the connection and the transfer time between entry/exit zones. The decision threshold of peak value for data distribution is calculated as Eq.(11).

$$threshold = avg(Count_p(E_i, E_j)) + \omega \cdot std(Count_p(E_i, E_j)) \quad (11)$$

where avg calculates the mean value of optimized $Count_p(E_i, E_j)$, std computes the standard deviation, and ω is set to 2 according to the experience.

We assume that there are no more than 2 transfer time between two entry/exist zones. When there are more than one peak in the transition statistics, the relationship between the entry/exist zones is considered disconnected. When there is one peak, entrance E_i and exit E_j are connected and the peak value is approximately equal to the transfer time. Considering that the accumulated data is close to normal distribution, Gaussian Fitting is utilized to optimize it. For the fitted data distribution, the mean value μ and standard deviation σ can be estimated to calculate the score of the transfer time between two trajectories as Eq.(3). μ is set to the transfer time between the connected entry/exit zones. Our estimated camera topology of DukeMTMC dataset is shown in Fig.5.

As illustrate in Fig.5, our estimated camera topology is basically consistent with the actual camera relationship which displayed in bottom right corner. The camera 8 does not be drew in the estimated topology due to the fact that it has overlapping view with camera 2, camera 3, and camera 5. The values on connecting lines denote the inferred transfer time between entry/exit zones, and the values in brackets indicate the manual statistics transfer time. The max error of transfer time does not exceed 500 frames, which is approximately equal to 8.1s (frame rate = 60fps). According to the estimated camera topology, we merely need to identify the trajectories that meet the transfer time constrains in connected cameras. It effectively narrows the retrieval range and improves the time efficiency of the online trajectory association.

3.4 Trajectory Association across Multiple Cameras

After achieving ODPR model and camera topology, we revisit POI [32] to obtain the trajectories in single camera. It is a detection-based multi-object tracking algorithm with appearance feature association. The outstanding MaskCNN [8] on ImageNet is chosen as the detector. PAFs [2] is introduced to split the bounding boxes including multiple pedestrians. The appearance features are extracted by our proposed ODPR with densely connected blocks. The location and frame number of each bounding box are regarded as the spatio-temporal constraints.

For all tracking trajectories in single camera, the one leaving from any entry/exit zone will be regarded as a probe trajectory. According to the estimated camera topology, its candidate set T_c that satisfy the transfer time constraint defined in Eq.(2) are selected from the connected cameras. In the achieved candidate set T_c, the associated trajectory T^* of the probe trajectory can be generated by Eq.(7). When a probe trajectory is associated in other cameras, it is removed from the probe trajectory set. The pseudocode of our proposed online trajectory association across multiple cameras is described in Algorithm 1, where t_{best} represents temporary optimal candidate trajectory.

In our proposed algorithm, the online single camera multi-object tracking and inter-camera trajectory association are run simultaneously. Compared with the inter-camera trajectory association based on the global retrieval mode, the camera topology effectively narrows the retrieval range and improves the time efficiency of inter-camera trajectory association. Meanwhile, the orientation aware weights and discriminative appearance features are explicitly leveraged to increase the accuracy of the inter-camera trajectory association.

Algorithm 1 Online Trajectory Association Pseudocode

Input: Topology information, Synchronous camera frames
1: **for** $f \in frames$ in all cameras **do**
2: single camera multi-object tracking
3: generate *probeset* according to tracking trajectories
4: **if** $\sim isEmpty(probeset)$ **then**
5: **for** $T_{p_i} \in probeset = \{T_{p_1}, T_{p_2}, \cdots, T_{p_N}\}$ **do**
6: generate candidate trajectory set T_c according to topology information and Eq.1-4
7: $t_{best} = T_c$ where $max(Sim(T_c, T_{p_i}))$ refer to Eq.5-7
8: **if** $Sim(t_{best}, T_{p_i}) > \tau$ **then**
9: unified ID number of t_{best} and T_{p_i}
10: delete T_{p_i} from *probeset*
output: multi-camera multi-object trajectories

4 EXPERIMENTAL RESULTS

In this section, we conduct two sets of experiments to validate our proposed contributions, which consist of the analysis of appearance features and the evaluation of inter-camera trajectory association. All experimental results are from the server with Titan X and dual E5, where red and blue values in bold highlight the first and second results, respectively. '↑' means that higher is better and '↓' represents that lower is better.

4.1 Impact of Appearance Features

We conduct a comparative analysis with some state-of-the-art algorithms on popular CUHK03 [15], Market1501 [36], and Duke-reID [38] datasets to verify whether the designed person re-identification architecture can extract discriminative feature representations. Each selected dataset is collected from multiple cameras, and includes the occlusions, orientation variations and many other issues. The comparative results are shown in Table 1.

Table 1: Comparative Results of Person Re-Identification.

Method	CUHK03		Market1501		Duke-reID	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
Re-rank [39]	38.10	40.3	77.11	63.63	–	–
Gan [38]	73.10	77.40	78.06	56.23	67.68	47.13
OIM [30]	77.50	–	82.10	–	68.10	–
SVDNet [26]	81.80	84.80	82.30	62.10	76.70	56.80
ACRN [25]	62.63	70.20	83.61	62.60	72.58	51.96
D-loss [37]	73.10	68.20	79.51	59.87	68.9	49.3
Deepcc [24]	–	–	**89.46**	**75.67**	**79.80**	63.40
Ours(R)	**90.64**	**87.93**	85.57	67.97	72.89	63.04
Ours(I)	88.95	78.95	86.10	68.40	79.62	**63.48**
Ours(D)	**89.29**	**79.03**	**87.23**	**76.35**	**81.64**	**72.34**

Our architecture can incorporate with any feature extraction module. In Table 1, 'R' denotes that the backbone network is the ResNet50 [9], 'I' represents that we use Inception V3 [27] as the backbone network, and 'D' indicates that DenseNet121 [10] are introduced into the person re-identification. The structure of the local feature extraction network is the same as the global feature extraction network except for the stem CNNs. As illustrated in Table 1, the proposed network outperforms most of the state-of-the-art algorithms on CUHK03 datasets. The improvement owes to the exploiting of the orientation cues and local features.

4.2 Evaluation of Inter-Camera Trajectory Association

An online trajectory association analysis experiment is conducted on DukeMTMC benchmark [23], which provides 8 surveillance videos captured from different cameras and contains more than 7,000 single camera trajectories belong to over 2,000 unique identities. Every video lasts for 85 mins, where 50 minutes data is for training and 35 minutes long is for testing. The test set is split into test-easy and the test-hard. The test-easy is 25 minutes long and has statistics similar to the train dataset. The test-hard is 10 minutes long and contains a group of over 60 people traversing 4 cameras. The comparison results of inter-camera trajectory association on the test-easy and the test-hard datasets are displayed on Table 2.

In Table 2, ID F1 score (IDF1) represents the ratio of correctly identified detections over the average number of ground-truth and computed detections. ID Precision (IDP) is the fraction of computed detections that are correctly identified. ID Recall (IDR) is the fraction of ground truth detections that are correctly identified. BIPCC [23], lx_b [1], MYTRACKER [31], and our method are online trajectory association, which are immediately available with each incoming frame and do not any modification at the later time. PT_BIPCC [18] and MTMC_CDSC [28] are offline approaches, which recover missing detections and reduce the fragmented trajectories by offline optimization. '$W = Ones$' denotes the evaluation results when all orientation weights described in Eq.(6) are set to 1. 'Trained W' refers to the results using the trained orientation weights. As shown in Table 2, the introduction of the orientation weights obviously improves the performance of our inter-camera trajectory association. And our proposed approach outperforms most state-of-the-art methods on the test set. Especially on the test-hard, our approach achieves superior performance on all evaluation protocols. This effectively proves the robustness of our method in the face of high-density populations.

So far, there are two offline methods on the DukeMTMC benchmark [1] that outperform us. The first method summarized in a technical report [34] utilizes the re-ranking strategy. The second method [24] introduces person re-id and global motion correlation into the inter-camera trajectory association. They have achieved excellent performance in accuracy, but they do not exploit the spatio-temporal cues, which makes it difficult to be applied to large-scale monitoring environments.

In addition to the above discussed accuracy, the time efficiency is also critical to the online inter-camera trajectory association, especially in large-scale surveillance environments. To verify the effect of spatio-temporal cues in our proposed framework, we analyze the matching times of trajectory association before and after using camera topology. The specific time efficiency analysis is shown in Table 3. To eliminate the contribution of the orientation aware weights, we set W to $\mathbb{1}$ in the following experiments.

As shown in Table 3, the topology information reduces the matching times to 92.83% of the number without the camera topology and improves the Rank-1 accuracy by 2.17% on Duke-reID dataset

Table 2: Comparative Results of Inter-Camera Trajectory Association on DukeMTMC Benckmark.

Tracker	Test-easy						Test-hard					
	Multi-Camera			Single-Camera			Multi-Camera			Single-Camera		
	IDF1↑	IDP↑	IDR↑	IDF1↑	IDP↑	IDR↑	IDF1↑	IDP↑	IDR↑	IDF1↑	IDP↑	IDR↑
PT_BIPCC[18]	34.9	41.6	30.1	71.2	84.8	61.4	32.9	41.3	27.3	65.0	**81.8**	54.0
MTMC_CDSC[28]	60.0	68.3	53.5	77.0	87.6	68.6	50.9	**63.2**	42.6	65.5	81.4	54.7
BIPCC[23]	56.2	67.0	48.4	70.1	83.6	60.4	47.3	59.6	39.2	64.5	81.2	53.5
lx_b[1]	58.0	**72.6**	48.2	70.3	**88.1**	58.5	48.3	60.6	40.2	64.2	80.4	53.4
MYTRACKER[31]	**65.4**	71.1	60.6	**80.3**	87.3	74.4	50.1	58.3	43.9	63.5	73.9	55.6
Ours($W = Ones$)	62.4	64.1	**60.8**	76.5	78.6	**74.5**	**55.6**	60.6	**51.3**	**69.4**	75.7	**64.0**
Ours(Trained W)	**68.8**	**71.8**	**66.0**	**83.8**	**87.6**	**80.4**	**61.2**	**68.0**	**55.5**	**77.9**	**86.6**	**70.7**

Table 3: Time Efficiency Analysis Before and After the Addition of Camera Topology.

Duke-reID	Rank-1(%)↑	Matching Times↓
Without topology	81.64	39,348,708
With topology	**83.81**	**36,527,932**
Test-easy	IDF1(%)↑	Matching Times↓
Without topology	58.8	40,018,276
With topology	**62.4**	**113,868**
Test-hard	IDF1(%)↑	Matching Times↓
Without topology	51.8	16,900,321
With topology	**55.6**	**161,329**

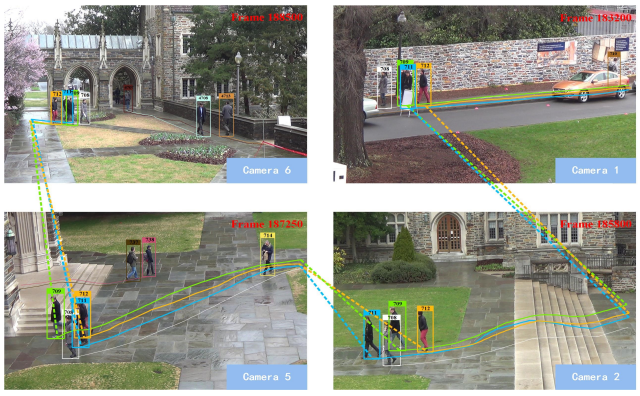


Figure 6: Visual results of online trajectory prediction on DukeMTMC. Bounding boxes and trajectories with the same color represent the identical pedestrian (Best viewed in color).

[38]. The improvement of time efficiency is not obvious due to that the transfer time constraint is unavailable on image-based re-identification dataset. Whereas, the time efficiency of the online trajectory association is significantly improved by the camera topology. For the test-easy set, the accurate connection and transfer time between cameras reduce the matching times from 40,018,276 to 113,868. For the test-hard set, the estimated camera topology decreases the matching times to 161,329, which is approximately one percent of the matching times without topology. The improvement in time efficiency is of great significance to promoting online or real-time intelligent monitoring and analysis. Some visual experiment results of the online trajectory association on DukeMTMC are shown in Fig.6.

As illustrated in Fig.6, the four non-overlapping cameras have obvious differences in illumination, pedestrian orientation and size. Nonetheless, our proposed approach still successfully achieves the correct inter-camera trajectory association of the pedestrians with back-and-forth occlusion. This is mainly because that we extract discriminative appearance features and introduce effective spatio-temporal constraints for identifying trajectories across cameras.

5 CONCLUSION

In this paper, we propose an inter-camera trajectory association framework exploiting person re-identification and camera topology. The contributions of the framework focus on learning the discriminative appearance features and achieving the spatio-temporal constraint between cameras. An orientation-driven person re-id architecture is proposed to reduce mismatching of trajectories from different cameras and alleviate the pedestrian orientation variations. Meanwhile, the appearance features are introduced into camera topology estimation to analyze the connections and the transfer time between cameras. On this basis, we explicitly leverage spatio-temporal constraints provided by the camera topology to narrow the retrieval range and utilize appearance features extracted from ODPR to associate the trajectories belong to the same identities across different cameras. These steps together achieve the goal of improving the efficiency and accuracy of trajectory association. As a result, our proposed person re-identification outperforms the state-of-the-art methods significantly. The estimated camera topology is basically consistent with the actual base map. More importantly, the online inter-camera trajectory association framework shows superior performance in term of accuracy and efficiency.

This paper sheds light on intelligent monitoring analysis challenges and advances. In future, we will continue to work on the combination of deep learning and surveillance video analysis in large scale of camera network.

ACKNOWLEDGMENTS

This work is supported by the Natural Science Foundation of China under Grant No. 61572061, 61472020, 61502020, and the China Postdoctoral Science Foundation under Grant No. 2013M540039.

REFERENCES

[1] Milan Anton, Ristani Ergys, and Laura. 2018. MOT Challenge Benchmark. Retrieved April 8, 2018 from https://motchallenge.net/results/DukeMTMCT/

[2] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. 1302–1310. https://doi.org/10.1109/CVPR.2017.143

[3] Weihua Chen, Lijun Cao, Xiaotang Chen, and Kaiqi Huang. 2015. An equalized global graph model-based approach for multi-camera object tracking. *IEEE Transactions on Circuits and Systems for Video Technology* PP, 99 (2015), 1–1.

[4] Xiaotang Chen, Kaiqi Huang, and Tieniu Tan. 2014. Object tracking across non-overlapping views by learning inter-camera transfer models. *Pattern Recognition* 47, 3 (2014), 1126–1137. https://doi.org/10.1016/j.patcog.2013.06.011

[5] De Cheng, Yihong Gong, Jinjun Wang, Qiqi Hou, and Nanning Zheng. 2017. Part-aware trajectories association across non-overlapping uncalibrated cameras. *Neurocomputing* 230 (2017), 30–39. https://doi.org/10.1016/j.neucom.2016.11.038

[6] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng. 2016. Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. 1335–1344. https://doi.org/10.1109/CVPR.2016.149

[7] Anthony R. Dick and Michael J. Brooks. 2004. A Stochastic Approach to Tracking Objects Across Multiple Cameras. In *AI 2004: Advances in Artificial Intelligence, 17th Australian Joint Conference on Artificial Intelligence, Cairns, Australia, December 4-6, 2004, Proceedings*. 160–170. https://doi.org/10.1007/978-3-540-30549-1_15

[8] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. 2980–2988. https://doi.org/10.1109/ICCV.2017.322

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. 770–778. https://doi.org/10.1109/CVPR.2016.90

[10] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely Connected Convolutional Networks. In *CVPR*, Vol. 1. 3.

[11] Omar Javed, Khurram Shafique, and Mubarak Shah. 2005. Appearance Modeling for Tracking in Multiple Non-Overlapping Cameras. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*. 26–33. https://doi.org/10.1109/CVPR.2005.71

[12] Sohaib Khan and Mubarak Shah. 2003. Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 10 (2003), 1355–1360. https://doi.org/10.1109/TPAMI.2003.1233912

[13] Cheng-Hao Kuo, Chang Huang, and Ram Nevatia. 2010. Inter-camera Association of Multi-target Tracks by On-Line Learned Appearance Affinity Models. In *Computer Vision - ECCV 2010, 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part I*. 383–396. https://doi.org/10.1007/978-3-642-15549-9_28

[14] Dangwei Li, Zhang Zhang, Xiaotang Chen, Haibin Ling, and Kaiqi Huang. 2016. A Richly Annotated Dataset for Pedestrian Attribute Recognition. *CoRR* abs/1603.07054 (2016). arXiv:1603.07054 http://arxiv.org/abs/1603.07054

[15] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. 2014. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*. 152–159. https://doi.org/10.1109/CVPR.2014.27

[16] Yiming Liang and Yue Zhou. 2017. Multi-camera Tracking Exploiting Person Re-ID Technique. In *Neural Information Processing - 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part III*. 397–404. https://doi.org/10.1007/978-3-319-70090-8_41

[17] Martijn C. Liem and Dariu M. Gavrila. 2014. Joint multi-person detection and tracking from overlapping cameras. *Computer Vision and Image Understanding* 128 (2014), 36–50. https://doi.org/10.1016/j.cviu.2014.06.003

[18] Andrii Maksai, Xinchao Wang, François Fleuret, and Pascal Fua. 2017. Non-Markovian Globally Consistent Multi-object Tracking. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. 2563–2573. https://doi.org/10.1109/ICCV.2017.278

[19] Zehavit Mandel, Ilan Shimshoni, and Daniel Keren. 2007. Multi-Camera Topology Recovery from Coherent Motion. In *2007 First ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2007, Vienna, Austria, 25-28 September, 2007*. 243–250. https://doi.org/10.1109/ICDSC.2007.4357530

[20] Niki Martinel, Gian Luca Foresti, and Christian Micheloni. 2017. Person Reidentification in a Distributed Camera Network Framework. *IEEE Trans. Cybernetics* 47, 11 (2017), 3530–3541. https://doi.org/10.1109/TCYB.2016.2568264

[21] Chaowei Niu and Eric Grimson. 2006. Recovering Non-overlapping Network Topology Using Far-field Vehicle Tracking Data. In *18th International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong, China*. 944–949. https://doi.org/10.1109/ICPR.2006.985

[22] Fatih Murat Porikli. 2003. Inter-camera color calibration by correlation model function. In *Proceedings of the 2003 International Conference on Image Processing, ICIP 2003, Barcelona, Catalonia, Spain, September 14-18, 2003*. 133–136. https://doi.org/10.1109/ICIP.2003.1246634

[23] Ergys Ristani, Francesco Solera, Roger S. Zou, Rita Cucchiara, and Carlo Tomasi. 2016. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking. In *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II*. 17–35. https://doi.org/10.1007/978-3-319-48881-3_2

[24] Ergys Ristani and Carlo Tomasi. 2018. Features for Multi-Target Multi-Camera Tracking and Re-Identification. *arXiv preprint arXiv:1803.10859* (2018).

[25] Arne Schumann and Rainer Stiefelhagen. 2017. Person Re-identification by Deep Learning Attribute-Complementary Information. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, USA, July 21-26, 2017*. 1435–1443. https://doi.org/10.1109/CVPRW.2017.186

[26] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. 2017. SVDNet for Pedestrian Retrieval. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. 3820–3828. https://doi.org/10.1109/ICCV.2017.410

[27] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2016. Rethinking the Inception Architecture for Computer Vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. 2818–2826. https://doi.org/10.1109/CVPR.2016.308

[28] Yonatan Tariku Tesfaye, Eyasu Zemene, Andrea Prati, Marcello Pelillo, and Mubarak Shah. 2017. Multi-Target Tracking in Multiple Non-Overlapping Cameras using Constrained Dominant Sets. *CoRR* abs/1706.06196 (2017). arXiv:1706.06196 http://arxiv.org/abs/1706.06196

[29] Longyin Wen, Zhen Lei, Ming-Ching Chang, Honggang Qi, and Siwei Lyu. 2017. Multi-Camera Multi-Target Tracking with Space-Time-View Hypergraph. *International Journal of Computer Vision* 122, 2 (2017), 313–333. https://doi.org/10.1007/s11263-016-0943-0

[30] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. 2017. Joint Detection and Identification Feature Learning for Person Search. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. 3376–3385. https://doi.org/10.1109/CVPR.2017.360

[31] Kwangjin Yoon, Young-min Song, and Moongu Jeon. 2018. Multiple hypothesis tracking algorithm for multi-target multi-camera tracking with disjoint views. *IET Image Processing* 12, 7 (2018), 1175–1184. https://doi.org/10.1049/iet-ipr.2017.1244

[32] Fengwei Yu, Wenbo Li, Quanquan Li, Yu Liu, Xiaohua Shi, and Junjie Yan. 2016. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature. In *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II*. 36–42. https://doi.org/10.1007/978-3-319-48881-3_3

[33] Shu Zhang, Yingying Zhu, and Amit K. Roy-Chowdhury. 2015. Tracking multiple interacting targets in a camera network. *Computer Vision and Image Understanding* 134 (2015), 64–73. https://doi.org/10.1016/j.cviu.2015.01.002

[34] Zhimeng Zhang, Jianan Wu, Xuan Zhang, and Chi Zhang. 2017. Multi-Target, Multi-Camera Tracking by Hierarchical Clustering: Recent Progress on DukeMTMC Project. *CoRR* abs/1712.09531 (2017). arXiv:1712.09531 http://arxiv.org/abs/1712.09531

[35] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. 2017. Spindle Net: Person Re-identification with Human Body Region Guided Feature Decomposition and Fusion. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. 907–915. https://doi.org/10.1109/CVPR.2017.103

[36] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable Person Re-identification: A Benchmark. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*. 1116–1124. https://doi.org/10.1109/ICCV.2015.133

[37] Zhedong Zheng, Liang Zheng, and Yi Yang. 2017. A discriminatively learned cnn embedding for person reidentification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 14, 1 (2017), 13.

[38] Zhedong Zheng, Liang Zheng, and Yi Yang. 2017. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. 3774–3782. https://doi.org/10.1109/ICCV.2017.405

[39] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. 2017. Re-ranking Person Re-identification with k-Reciprocal Encoding. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. 3652–3661. https://doi.org/10.1109/CVPR.2017.389

[40] Michael Zhu, Anthony R. Dick, and Anton van den Hengel. 2015. Camera Network Topology Estimation by Lighting Variation. In *2015 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2015, Adelaide, Australia, November 23-25, 2015*. 1–6. https://doi.org/10.1109/DICTA.2015.7371245