Multimedia Tools and Applications
https://doi.org/10.1007/s11042-018-6638-5

People tracking in multi-camera systems: a review

Rabah Iguernaissi[1] **· Djamal Merad**[1] **· Kheireddine Aziz**[2] **· Pierre Drap**[1]

Received: 18 December 2017 / Revised: 24 August 2018 / Accepted: 31 August 2018 /
Published online: 14 September 2018

Abstract

Ubiquitousness of multiple cameras in surveillance systems is very beneficial for studying peoples behavior. The multiple views of the observed scene permit the management of dynamic occlusions and failures that may affect any sensor. The multi-camera tracking of objects is considered as the basic step in the design of intelligent surveillance applications. This thematic had been addressed in several researches. Various methods had been proposed to achieve an accurate tracking in the most challenging conditions as occlusions and lighting variations. These methods are addressed in two main research lines: the centralized and the distributed tracking approaches. In this paper, we propose an overview of the multi-camera tracking of objects which summarizes and classifies the most used existing methods.

Keywords Multi-camera tracking · Person re-identification · People tracking

1 Introduction

Nowadays, utility of video surveillance becomes a common fact in many fields. Among these surveillance systems, multiples cameras are used to improve coverage and accuracy in the surveyed area. Due to the huge amount of data generated by these video surveillance systems; the need for automatic analysis techniques that may be used to exploit these data becomes more indispensable to study the behavior of people without needing many human operators to analyze the captured videos.

People's tracking in video stream is the basic step for intelligent surveillance applications in different fields such as the automatic behavioral analysis of people, and the security surveillance systems. Thus, many researches had been conducted, and a variety of approaches had been used to solve this problem. The problem of people tracking is addressed in two main thematics. The mono-camera configuration in which only data captured from a single camera are processed during tracking process, and the multi-camera

✉ Rabah Iguernaissi
rabah.iguernaissi@lsis.org

[1] LSIS - UMR CNRS 7296, Aix-Marseille University, 163 Avenue of Luminy, Marseille, 13288 Cedex 9, France

[2] SeaTech School, Toulon University, Avenue de l ' Universite - BP 20132, La Garde, 83957, France

configuration in which the correspondences between tracked individuals among the whole system are considered.

Both mono and multi-camera tracking of people are defined as the ability of automatically recording the trajectories of tracked individuals in the surveyed scene. The main difference between these configurations reside in the fact that in multi-camera systems the correspondences between the tracked individuals in different cameras are considered whereas in the mono-camera systems these correspondences are just ignored. This tracking may be achieved either in real time which is called the online tracking or from a recorded video which is known as the offline tracking. [31]

The issue of people tracking in a multi-camera system is an extension of the problem of people tracking in mono-camera system. Thus, most of the algorithms which are used in multi-camera tracking are based on the well-known algorithms of mono-camera tracking such as the motion detection, object modeling and feature extraction methods.

The use of multi-camera has many advantages over the use of a single camera. Among these advantages, we may mention the reduction of errors that are caused by occlusions in crowded environment and failures that may affect any sensor and the ability of monitoring large areas. But, in the other hand, people tracking in multi-camera system is very challenging, because the tracking process insure to fuse information coming from the different sensors. In this context, many researches had been conducted in the last few years to solve this problem. [31, 86]

In this paper, we present an overview of different approaches, and present some of the proposed methods which are used in solving the problem of multi-camera tracking of people. The paper is organized in two main sections. The first section is about the centralized approaches, whereas the second section is dealing with the distributed approaches for the multi-camera tracking of people.

2 Different methods for people tracking in multi-camera system

Due to its importance, many methods had been proposed for people tracking in multi-camera system in the last decade. These tracking methods had been classified in different manners according to different criteria. They can be classified according to the methods used for motion detection into two categories the temporal difference methods and the background subtraction methods. They may also be classified according to the position of cameras into overlapping and non-overlapping camera views. The third classification that we adopted in this paper, is based on the data fusion between different sensors. In this approach, methods are classified into two categories which are the early fusion of data or the centralized approach in which detection and tracking is performed after combining data from different sensors and the second category which is the late fusion of data or the distributed approach in which detection and tracking is done for each camera and the results are combined to get the final trajectories of people. [31, 66, 86]

Figure 1 represents the classification of the different multi-camera tracking methods that we adopted in this paper.

3 Centralized approaches

There exist many methods for people tracking in multi-camera systems that rely on the use of the centralized approach which consists mainly in the combination of data from different

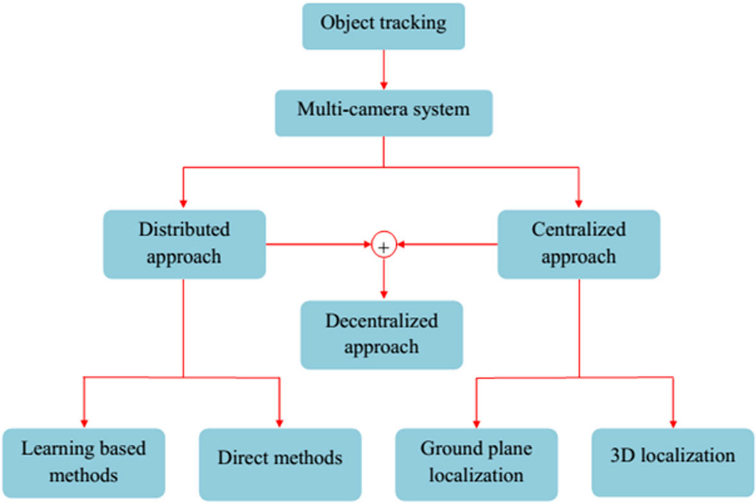

Fig. 1 Adopted classification for multi-camera tracking methods

cameras before tracking (early fusion of data). The centralized approach is generally used with overlapping camera views to reduce the effect of occlusions and noisy observations in crowded environment or in the monitoring of small areas. [56, 66]

3.1 Basic approach

Most tracking methods that are using the centralized approach are based on the same basic steps which consist in performing the people detection for each camera view, fuse the data obtained from the different cameras, and then perform the tracking to get the trajectories of the tracked individuals. This basic approach is illustrated in Fig. 2.

In the first step, the well-known methods for foreground object detection such as the background subtraction methods are used to detect the tracked people in each camera view for specific frame of the videos (this frame should correspond to the same instant in all

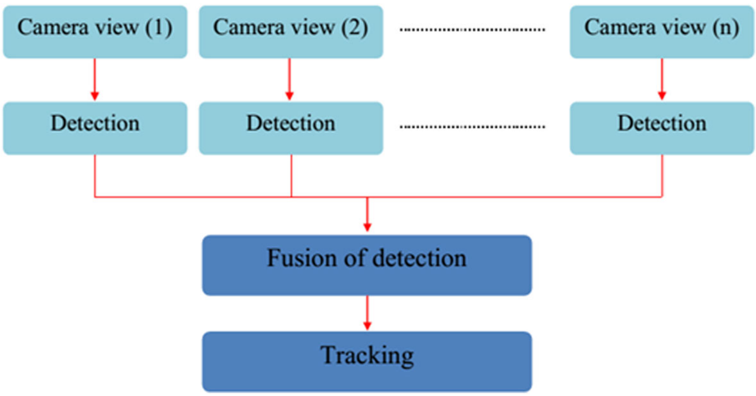

Fig. 2 Centralized approach

camera views, thus the cameras should be synchronized). Then, these detections will be used in the second step either to found the 3D localization of the tracked individuals in the actual scene, or to establish the ground plane occupancy map in which the tracked individuals are located in the ground plane of the scene (for this step also the geometric relationship between different cameras should be known). The third and last step consists in repeating the previous steps over a period of time (for successive frames of the videos) and recording the locations of the tracked individuals at each time to reconstruct their trajectories. A synthesis of methods is presented in Table 1, and more details are gives below.

3.2 Used methods

The tracking methods which are based in the use of centralized approach rely in the fusion of data from different cameras before tracking process. This step of data fusion is the most important step by which the proposed methods differ.

Among the proposed methods for people tracking that rely on the use of centralized approach; two main categories may be distinguished based on data fusion. The first category is based on the fusion of data from different sensors to establish the ground plane occupancy map of the scene as proposed in [11, 22, 35, 49, 70], and the second category is based on the fusion of data to get information about the 3D position of the tracked individuals in the scene as proposed in [11, 15, 18, 40, 56, 83, 87].

Table 1 Centralized approaches

Reference	Feature	Association method	Calibration
Kim et al. Du et al.	Binary map	Intersection of the central vertical axis	Yes
Fleuret et al. Huang et al.	Binary map	Occupancy map	Yes
Santos et al. Khan et al.	Binary map	Homgraphy constraints	No
Liem et al. Guan et al. Arsic et al.	Color	3D reconstruction	Yes
Chen et al.	Tracklets	Global graph optimization	No
Kroeger et al.	Color	SFM	No
Yao et al.	Color appearance	3D model	Yes
Hofmann et al. Leal-Taixe et al.	Color	Deformable part model, 3D projection	Yes
Bredereck et al.	HSV color and positions	3D positions	Yes
Jiaung et al. Balteiri et al.	Color appearance	Deformable part model, 3D distances	Yes
Wen et al.	Appearance, motion continuity	3D positions	Yes
Du et al.	Color appearance	Probabilistic algorithm	No
Zhang and Cheng	Haar-like features, HOG, and LBP	Combining views	Yes

In [49], Kim and Davis proposed a method that can be classified in the first category. They used standard background subtraction method to detect the foreground map in each camera view. The detected foreground pixels are classified into classes that represent individuals who are present in the scene using a Bayesian pixel classification. These segmented classes (blobs) across the different camera views are integrated to get the ground plane locations of people by constructing the top view of the scene. In order to find these ground locations they used the intersection of the central vertical axis of the detected blobs. In a similar way, Du and Piater [29] used the intersection between principles axis of targets to integrate information from different cameras and used particle filters to track people in each camera view and in the ground plane.

Fleuret et al, in their paper [35], proposed a method that combines the probabilistic occupancy map which provides estimation on the occupancy of the ground plane at each single frame of the video with a global optimization of trajectories for detected individuals over 100 frame batches. The estimation of the probabilistic occupancy map is done by first using the background subtraction method to detect the foreground objects in the scene and then a generative model which is based on the combination of the color and motion model is used to estimate the occupancy map. In the same context, Huang and Wang [43] defined a target detection probability to estimate the probability of having a moving object in ground locations based on the foreground images.

In the same category, Santos and Morinoto [70] proposed a method for people tracking in multi-camera system. The algorithm starts by using a standard background subtraction method to detect the foreground objects (individuals) in all camera views. Then, segmented foreground objects (individuals) are used to compute the evidence of people presence in each pixel of the reference image which represents the ground plane of the scene. The ground pixels of people are localized by computing the support (foreground mass) for each foreground pixel in all camera views, then by combining the computed support of each pixel in all cameras the ground pixels of people are located in the local maximums, and the last step consists in matching the detected candidates with previously tracked people to get the trajectories. In the same way, Khan and Shah [48] used novel planar homographic constraints to determine the locations of feets on the ground plane for the tracked individuals. Whereas, Eshel and Moses [30] choose to track only the heads of people, where heads are presented by a single feature point that is projected to the ground plane and tracked to recover trajectories.

As said previously, the second category of methods which are using the centralized approach are methods which try to recover 3D information from different cameras to estimate the 3D locations of tracked individuals.

In this approach, several methods had been proposed based on the 3D reconstruction of either the scene or the tracked individuals [2, 37, 50, 56, 85], Liem and Gavrila [56] proposed a method for multi-person tracking in overlapping cameras. They proposed an algorithm in which people detection and tracking is done using a volumetric 3D reconstruction of the scene that is computed from the foreground maps that are detected in overlapping cameras. In this method, the foreground maps are detected using the background subtraction method and the projection of the reconstructed objects to the ground plane allows the estimation of the number of people who are present in each detected blob according to the surface that it occupies in the ground plane. Then the tracking is done by associating these detections with previously tracked individuals. In the same context, Guan et al [37] used a method for dynamic scene reconstruction based on estimating the shape of objects from their silhouettes obtained from 2d detections, Arsic et al [2] use a homographic transformation applied for multiple layers to obtain the 3D reconstruction of the scene, and Kroeger et al [50] merged methods which are used in multi-view tracking by detection and localization methods developed for SFM to establish a tracking methods which can be used for dynamic cameras without any restrictions about planar motion of tracked individuals, whereas Yao et Odobez [85] adopted a 3D approach in which the tracked individuals are represented by body models in the common 3D world.

For the same category, Hofmann et al proposed in [40] an algorithm in which the 3D reconstruction and tracking are combined and solved in a joint framework. They start by performing people detection from each camera view using the discriminatively trained deformable part model described in [33]. Then, the 2D detections are projected to a common 3D coordinate system to get a 3D reconstruction of the tracked objects. The trajectories of tracked objects are defined as the ordered list of the 3D reconstruction. In the same way, Leal-Taixe et al [53] combined the 3D reconstruction and tracking in a single global optimization problem where 2D detections are coupled for 3D localization and tracking purposes.

Jiang et al [45] used the 3D information in another way by proposing a method which is based on two stages graph based tracking. The first graph uses hypotheses in the ground plane obtained from 2D appearances and 3D distances of tracked individuals to extract a set of tracklets. The second graph is used to link the tracklets in order to recover the whole trajectories of the tracked individuals. In the same way, Wen et al [84] used hyper-graph to realize the 3D tracking in multi-camera system based on 2D tracklets.The last method that we will present here about the use of 3D information in multi-camera tracking is the method proposed by Del-Blanco et al [18]. They proposed an algorithm in which moving objects are detected in each camera view based on the use of a standard background subtraction method, this 2D view are also used to extract color information about these moving objects. Then, these information are fused in the 3D world to estimate the volumetric occupancy probability density for each tracked individual over time using a 3D particle filter.

In addition to the previously mentioned categories, there exist some methods which are hybrid between the two categories such as the method proposed by Balteiri et al in [11], in which 3D information from all cameras are used to detect people and their ground plane location by approximating human bodies by cylinders. Then, a short term tracking is performed by exploiting 3D geometrical information and constraint between cameras. These short term trajectories are merged to get the long term trajectories using a re-identification method which is based on the 3D body model which is presented in [9] which takes into account both the color and appearance features. Another method that we will mention here is the method proposed by Du and Piater [28] which rely in a limited share of information between different sensors. Their algorithm is based on the use of a dedicated particle filters based local trackers for each camera view, and the use of a global belief propagation algorithm to share information between the different trackers. . We can also mention Wang et al [80] who proposed a system to fuse information from multi-camera systems for augmented reality applications based on prioritization.

4 Distributed approaches

The distributed approach is generally used in applications which are dealing with non-overlapping camera views that are designed most of the time to monitor large areas. In this kind of applications either the camera views are non-overlapping or the geometric relationship between cameras is difficult to recover. In this approach, no restrictions on the synchronization or the positions of cameras are required which is a very important characteristic [66, 74].

4.1 Basic approach

The methods which are based on the distributed approach are methods that rely in a limited share of information between different sensors (cameras). In general, these methods are made up of three basic steps that allow establishing the final trajectories of tracked individuals. They start by detecting and tracking people in each camera view independently, then getting the correspondences between the tracked individuals from different cameras, before merging these trajectories obtained in the first step using the correspondences obtained in the second step to construct the whole trajectories of the tracked individuals. This approach is simplified in Fig. 3.

In this approach, the first step is a standard mono-camera people tracking problem in which the well-known methods of tracking such as Kalman and Particle filters may be used to establish the trajectories of the tracked individuals in each camera view independently from other cameras. Then, the second step consists in finding correspondences between the tracked individuals over all views which is generally done by using some re-identification techniques which are based on shape or appearance descriptors of the tracked individuals. The third and last step consists in fusing the trajectories obtained in the first step using the correspondences obtained in the second step to establish the final global trajectories. An overview of these methods is presented in Table 2, and more detailed descriptions are given below.

4.2 Re-identification methods

In distributed approach, people's tracking is done independently for each camera view without taking into account any information from other cameras. Thus, the problem is viewed as a standard mono-camera tracking system with additional step that consists in re-identifying tracked individuals which is the most important part in a distributed multi-camera tracking system.

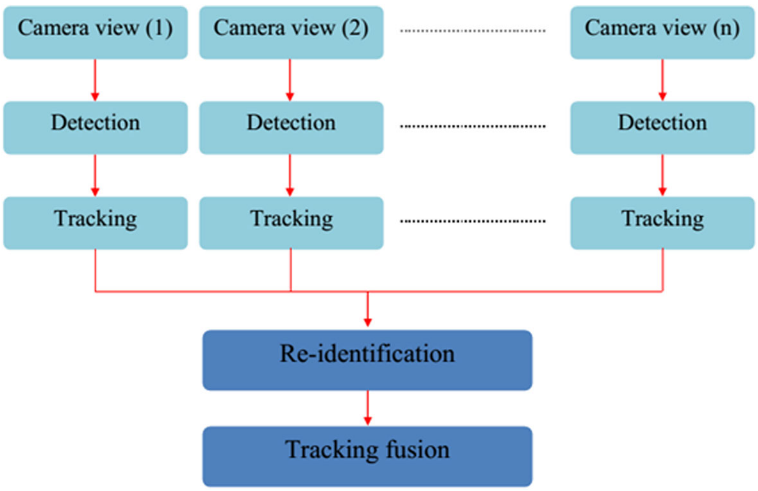

Fig. 3 Distributed approach

Table 2 Distributed approaches

Reference	Feature	Association method	Calibration
Back et al.	Haar-like	Adaboost	–
Kuo et al.	color histogram, covariance matrix, HOG		
Avraham et al.	HOG	SVM	-
Kenk et al.	Color histogram		
Nakajama et al.	HOG	Multi-class SVM	–
Prosser et al.	Color, texture	RankSVM	–
Loy et al.	Color	Laplacian graph	–
Bazzani et al.	Histogram plus epitome	K-NN	–
Bak et al.	Covariance descriptor	SVM	–
Javed et al.	Color	Epipolar geometry	Yes
Aziz et al.	SIFT, SURF, SpinImage	2D model, k-NN	Yes
Cheng et al.	Color, texture	–	–
Wang et al.			
Cong et al.	Color	2D model, k-NN	–
Chen et al.	Color spectrum	Tracklet association	–
Hamdoun et al.	SURF	Voting algorithm, K-NN	–
Farenzena et al.	Several appearance features	Asymmetric and symmetric axis of human body model	–
Mazzon and Cavallaro	Position and velocity	Social force model	–
Bouma et al.	Color and spatial information	Histogram intersection	–
Nie et al.	HSV, HOG, spatial information	Graph matching	–
Brendel et al.	Position, size, color histograms, intensity gradient, motion	MWIS algorithm	–
Li et al.	Color	–	–
Colombo et al.	spatio-temporal properties, Color	Distance measure	Yes

Re-identification may be defined as the ability of deciding whether a person who is currently detected in a camera view is already detected and can be matched to some individual who had been already detected either in the same camera view or in other camera views. This problem arises generally in the case of non-overlapping camera views. But it can also arise with overlapping cameras in case the geometric relationship between different cameras is difficult to recover.

The recent re-identification methods can be classified into two main categories the learning based methods and the direct approaches or the non-learning based methods.

This work is focused in re-identification methods which are proposed in the context of multi-camera tracking of people. A more general study about re-identification methods is proposed by Vezzani et al in [76].

4.2.1 Learning based methods

Learning based methods for re-identification are based on the use of training dataset of individuals in order to analyze the extracted features from these specific samples and then generalize the concept to re-identify new unseen tracked individuals. These methods differ in the used classifiers. Thus, many methods were proposed based on different classifiers. Among these methods we can mention the boosting method proposed in [6], the support vector machine (SVM) proposed in [67], the descriptors learning methods [34], the distance learning based methods [51, 78, 82], the deep learning approach [23, 68, 73, 81], the nearest neighbor, and the least square reduction proposed in [57] and [72] respectively.

In [6] Bak et al proposed an algorithm for people re-identification based on the use of Adaboost scheme with Haar-like features. The detection and tracking of people are done using the histogram of oriented gradients (HOG). Then, a set of detections of the interest individual are accumulated over a short period of time. These detections are used to extract the Haar-like features that are used in the Adaboost scheme to extract a visual signature of tracked individual and which will be stored in database to be compared with other detections.

Methods that are using SVM classifiers are, generally, based on the same principle which consists in learning an SVM classifier using a set of train images and then use the classifier to re-identify people that are detected in other camera views. In this context, Avraham et al proposed in [3] an algorithm that is trained to find correspondences between people captured by two non-overlapping cameras. It is based on the use of a classifier that is trained to distinguish between positive and negative pairs of detections based on common simple descriptors such as color histograms knowing that the positive pairs represent the case where the detected individuals in both cameras is the same person and negative pairs represents the case where the detected individuals in the two cameras are different. Then, the binary SVM classifier classifies each new pair of detections in one of these two classes. Martinel et al [60] used the same basic idea by proposing a method based on the training of multiple re-identification experts on an extracted robust feature representation from pairs of images that represents either the same individual or different individuals. Then, these experts are used to evaluate representations from new pairs of images and the re-identification is decided depending on the combination of answers from different experts. In the same category, Nakajama et al proposed in [64] a multi-class SVM classifier that is based on the combination of binary classifiers. This classifier is used for person full-body recognition by training the SVM classifier in color and shape-based features from images of detected individuals. In the same way, Kenk et al [47] used color histograms in the creation of multiclass classifier that is based on the use of binary classifiers to solve the problem of re-identification in distributed smart camera environment. In the same context, Prosser et al proposed new reformulation of the problem of person re-identification in [67] as a learning to rank problem. They used RankSVM, which is trained on multi-dimensional feature vectors representing the appearances of detected individuals (color, texture), to order the possible matches of a detected person instead of using the conventional template matching using a distance measure. Loy et al [58] adopted the same principle by proposing a manifold ranking framework for re-identification, for this purpose, they studied two manifold ranking models the normalized graph Laplacian, and the unnormalized iterated graph Laplacian.

Another category of approaches is the descriptors learning based methods. This category of methods relies in the use of a weighted combination of several features that describe the appearance and shape of the tracked individual. The learning phase is used to weight the different features and extract a single signature that describes the tracked individual. Among these methods we can mention the method proposed by Figueira et al [34], in which, the algorithm extracts a set of appearance based descriptors from the body parts of each individual. These set of features is used to estimate the parameters of the model by using the multi-feature learning framework of [69]. In the same way, Kuo et al [52] proposed a method that relies in online discriminative appearance learning. They used the Adaboost algorithm in a set of appearance feature (color histograms, covariance matrix, HOG features) calculated in different locations of the detected individual to combine the most discriminative feature in an appearance model that is used to decide whether two tracklets belong to the same person or not. In the same context, Bazzani et al [12] proposed a novel descriptor that condenses information from different images in single signature, called histogram plus epitome, which embeds global and local appearance features of the detected individual. Another category of methods relying on the use of weighted features are those where the weighting procedure of used feature is based on the use of a deep learning approach. In this context, Wang et al [81] proposed a method based on the use of a deep learning approach to associate individuals in non-overlapping cameras. They proposed an algorithm based on the use of 4 sub-CNNs trained on different parts of the human body with different scales. They also trained a list-wise loss function to give larger margin for the hard negative samples in order to improve the re-identification. Su et al [73] proposed a 3 stages dCNN learning approach for re-identifying individuals, whereas Chen et al [23] proposed a pyramid multi-scale deep learning approach used with the fusion of multi-scale features to re-identify people.

The methods based on the distance learning approach aim to learn the similarity measures between pairs of images or features that represent detected individuals. So, the algorithms try, in general, to learn the distances (differences) between images in both positive pairs that represent the same person and negative pairs that represent different persons instead of learning the visual features of individuals. In this category, we can mention the method proposed by Zheng et al [89] in which the problem of people re-identification is reformulated as a relative distance comparison learning problem in which the pairs of true matches had a smaller distance than the pairs of wrong matches. The relative distance comparison model is applied to an appearance representation of people captured by a set of basic features (texture and color histograms). This type approach is affected by the zero-shot problem that is addressed by Wang et al [78] using the concept of cross-view consistency to improve the performance of a learnt uniform metric. They addressed the problem by using a cross-view support adaptive factor and a cross-view projection adaptive factor obtained from the associations between the training samples and the new pairs of detections. In the same context, Bak et al [8] used the covariance descriptor of [75] which encodes the information about the feature variances inside image regions to characterize the appearances of detected individuals. But instead of using just one feature as done in the latter, they used the most discriminative feature for each region of the image by a machine learning technique. Similarly, Wang et al [82] proposed to transform the feature vectors into discrepancy matrix since the use of discrepancy matrices is more performant than simple feature vectors. They used these matrices with a matrix metric that is trained to pull two discrepancy matrices representing the same individual close and push two discrepancy matrices representing different individuals far away from each other. Most of these methods may boost their performance by using the tool proposed by Bai et al [5] which is called the supervised smoothed manifold (SSM).

For the same category, Javed et al [44] proposed a method that is based on learning both the space-time and the appearance relationships between each two cameras to be used in calculating the probability that observations in the two cameras belong to the same individual. The learnt appearance relationship consists in the brightness transfer function between the two cameras whereas the space- time relationship consists in entrances, exists, directions, and the average time needed to reach one camera view after leaving the other.

4.2.2 Direct methods

Direct methods for people re-identification are based on extracting a discriminative signature for each tracked individual that is directly used to match the detected persons in different cameras without using a training dataset. There exist many methods which are based on this approach. Among these techniques, we can distinguish the single-shot methods which are based on using a single image for each tracked individual and the multiple-shot methods which are using a sequence of images for each tracked individual. The effectiveness of these techniques is directly dependent on the signature which is used to describe the persons. Thus, different global and local features are used in extracting these signatures. The used features are, generally, based on appearance properties such as color, shape, and texture. In this category of re-identification, we can mention the methods proposed in [7] for the texture, [38, 77] for the shape, and [32] for a combination of a set of local features.

Bak et al [7] proposed an appearance model based on spatial covariance regions which are extracted from human body parts. In their method, they used the Histogram of Oriented Gradients (HOG) to detect the human body parts and covariance descriptors to extract the color information for each of these parts. Then, the invariant signatures for comparing different parts are generated by combining the color and structural information. Whereas, Chen et al [21] used the major color spectrum histogram with two similarity measures a positive similarity (PS) to indicate linked tracklets and a negative similarity (NS) to indicate unlinked tracklets for improving the inter camera object tracking. In the same way, Aziz et al [4] proposed a method for person re-identification in which appearances of the detected individuals are classified into frontal or back appearance, before segmenting their silhouettes into 3 parts (head, torso, and legs). Then, for each of these parts a discriminative signature is extracted based on the SIFT, SURF, and spin images. These signatures are used to match the different parts from the detected individuals for re-identification. The use of pose estimation before re-identification improves drastically the performance of algorithms. In this context, more sophisticated pose estimation strategies may be used such as those proposed by Hong et al [41, 42]. In their works, Hong et al proposed the use of non-linear mapping between the 2D and the 3D environments to recover the human pose based on a deep learning approach. Cheng et al [27], for the same principle, used a custom pictorial structure in order to localize the body parts and extract their descriptors; whereas Alahi et al [1] proposed a method based on the creation of an object descriptor which is constructed from the cascade of grids of regions descriptors.

In the same category, Wang et al [77] introduced the concept of shape and appearance context which is used to compute the occurrence matrix that represents the descriptor of the object (person) of interest. The descriptor is created by capturing the spatial relationship among the appearance labels. They also proposed an integral computations framework in which the occurrence matrix is computed in real-time; whereas Cong et al [25] proposed new appearance signature called color-position histogram which is extracted from the silhouette of the detected individual by dividing it to several horizontal regions and taking the normalized RGB values from each region to get a discriminative signature that is used in matching the detected individuals from different views. In the same context, Bouma et al [14] used a histogram intersection method to associate descriptors based on both color and spatial information, while, Nie et al [65] used a graph matching approach.

In the same context, Hamdoun et al [38] proposed an identification scheme based on the matching of interest points which are collected in several images during short video sequence. They used the SURF algorithm to extract the feature points and their descriptors to build a signature for each tracked individual. In the same way, Jungling and Arens [46] build a re-identification system that uses the SIFT features and the implicit shape model (ISM) proposed in [54]. For the same context of finding the most discriminative signature and instead of looking for feature points [59, 79, 88] exploited the image saliency. In [59], they proposed a strategy based on the identification of salient regions that may be used for people re-identification. They used a kernelized graph-based technique to compute the saliency of each pixel depending on its surrounding. This saliency is used to weight the extracted features that will be used for re-identification. In the same context, [88] proposed a re-identification method based on saliency matching. The first step of this approach is the estimation of the salience distribution over the whole body of the individual. Then, the person re-identification problem is formulated as a salience matching where the matching between individuals is done based on both the spatial distribution of salience and the visual similarity between patches that have the same saliency.

Among methods combining several local features, we can mention the method proposed by Farenzena et al [32] which is based on accumulating several local features to get a discriminative signature that can differentiate different individuals. The algorithm is designed in three steps and can be used as either a single or multiple shot. The first step consists in detecting the body parts using the asymmetric and symmetric axis of human body. Then, several appearance based features are extracted from each part. In the third step, the matching between the different candidates is done by accumulating the distances between different features and assuming that the accumulation of all features forms a single signature that describes the detected individual. In the same way, Mazzon et al [62] combined appearance features extracted from the upper part of the tracked person body and the position of candidates for the re-identification; whereas Brendel et al [16] proposed a method for person re-identification based on extracting a descriptor that record the position, the size, the color histograms, the intensity gradient, and the optical flow for each detected individual, then, the tracking is achieved based on data association algorithms.

For the same category of re-identification methods, Li et al [55] combined the strengths of matching based and classification based algorithm by extracting new features which are used in the matching process from the learned metric space instead of extracting these features from the original data space; whereas Colombo et al [24] proposed a technique for estimating the transformation between cameras color spaces to achieve good color constancy between the colors values associated to each camera. The use of appropriate color constancy improves the performance of the different color descriptors that may be used for re-identification across several camera views. To recover the trajectories of tracked individuals Mazzon and Cavallaro [61] proposed a method that is based on exploiting the environment. Their strategy is based on the use of a social force model that encodes the points of interest in the studied scene (exists, seats, points of meeting). This model facilitates the prediction of people trajectories based on goals and barriers that may influence the trajectory of the tracked person. In the same way, the semantic representation proposed by chang et al [19, 20] for event detection may be used to improve the tracking. In their works, chang et al proposed to use a semantic representation and a semantic pooling approach for detecting event in videos based on the use of SVMs trained on semantic saliency. This kind of approaches may add some restrictions that will improve most exiting re-identification strategies.

5 Decentralized approaches (clustered tracking)

In general over a large network of cameras, the use of the centralized approach leads to high data transfer; whereas the use of distributed approach does not exploit the information from the multi-camera. Thus, in order to overcome these limitations, some researches adopted a hybrid system in which the cameras are grouped into clusters with local fusion centers. In these methods, the multi-camera tracking in each cluster (group of cameras) is done using the centralized approach and then re-identification algorithms are used to match tracklets from different clusters to get the trajectory of the tracked individuals over the whole network [74].

Among methods that are using the decentralized approach we can mention the method proposed by Cosar and Cetin [26], in which, the image features are extracted from each image. Then, the tracking is performed in the fusion node by combining the extracted features. They used the tracking methods proposed in [35] in which tracking is performed using ground plane occupancy map and ground plane color map in a centralized approach. But, instead of using centralized approach, the ground plane occupancy map is evaluated in the fusion node whereas the ground plane color map is evaluated in a distributed way by computing the color likelihood function for the tracked individual in each camera and then combining these functions in the fusion node to get the multiview color likelihood function. The tracking is achieved by combining information from the ground plane occupancy map and the ground plane color map; whereas, Medeiros et al [63] used a clustered kalman filter to achieve a decentralized tracking of objects in wireless network of cameras. In this method, the targets positions are estimated in the cluster head by receiving information about the tracked object from the cameras which belong to the same cluster (group of cameras) and using a kalman filtering. Then, the estimated information about the position of the tracked object in the cluster head is transmitted to the base station which gathers information from the different clusters to get the trajectory of the tracked object. In the same context, Bhuvana et al [13] proposed a strategy to limit the information exchange between different cameras in the network in order to improve the performance of the tracking system. The method is based on the selection of a limited number of cameras which are the most informative with respect to the tracked object (surprisal cameras) to reduce the amount of information that should be transferred to the fusion center.

6 Usage and comparison

In this section, the Table 3 summarizes a comparison of the suitable usage cases for both categories, some of their benefits and drawbacks, and some applications domains in which the multi-camera tracking of objects can be used.

Table 3 Summary of usage and comparison between the centralized and the distributed approaches

	Centralized approach	Distributed approach
Usage	- Monitoring of small areas. - Cameras with overlapping views. - Crowded areas.	- Monitoring of large areas. - Non-overlapping cameras views.
Benefits	- Exploit the multi-views of the scene. - Well handling of occlusions. - Reduced effect of noisy observations and failure in a single sensor.	- No need for synchronization. - Accurate Geometric relationship between cameras is not necessary for tracking. - Just few cameras are needed. - Suitable for online applications.
Drawbacks	- Synchronization of cameras is needed. - Accurate Geometric relationship between cameras is necessary for tracking. - Large number of cameras is needed for a small area. - Not very suitable for online applications.	- Use just a single view even in the availability of multi-views. - Problem in case of severe occlusions. - Noisy observations or failure in any sensor affects the whole tracking system.
Some applications	- Intelligent surveillance systems. - Urban traffic control. - Monitoring of people in public places. - Pedestrian counting. - Behavioral analysis of peoples. - Sport analysis. - Statistical studies for commercial purposes.	

As described in Table 3, there are no discriminative criteria that allow deciding whether the centralized approach is better than the distributed one or vice-versa. Both the centralized and the distributed approaches have some benefits and drawbacks. Thus, the choice of adopted approach depends, generally, in the specifications of the designed application as shown in the Table 3.

The effectiveness of different approaches depends on the specifications of the application under design. The centralized approaches are more convenient in the case of a network with overlapping cameras while the distributed approaches are more convenient in networks with non-overlapping cameras. In general manner, the centralized approaches are more effective for occlusions' management and for precise localization of tracked individuals in the 3D world that corresponds to the actual positions of individuals in the real world. But this precision induces more calculation charge and time consumption, in addition to some restrictions in cameras' positioning (calibrated, geometric relationship between cameras). In the other hand, the distributed approaches can be used directly with existing networks of cameras since they do not require special specifications on the cameras. They are also more convenient for online applications since less time consuming. But they are less robust in the occlusions' management and less precise in the interpretation of positions to the real

Fig. 4 Some illustrative examples from the VIPeR dataset

world. Then, as a conclusion, we can say that the centralized approaches are more effective for dense crowds in overlapping cameras, whereas the distributed approaches are more effective in sparse crowds or in non-overlapping cameras.

7 Used datasets

Many methods for peoples tracking in multi-camera systems and people re-identification had been proposed in the last decade. Thus, the need for testing datasets that may be used in the evaluation of these methods become more indispensable. For this purpose, many

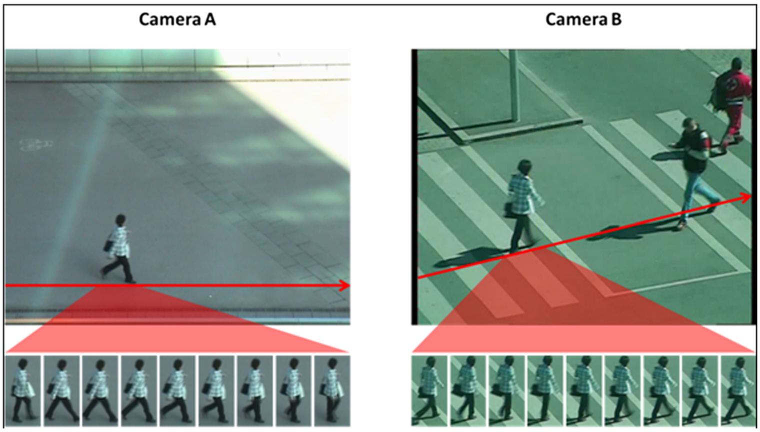

Fig. 5 Illustrative example from the PRID dataset

Fig. 6 Illustrative example from the 3DPeS dataset

datasets had been constructed [71]. Among these datasets we can mention some of the most used such as the VIPeR dataset [36] for person single-shot re-identification techniques, and CAVIAR dataset [17] for people tracking, the PRID dataset [39] for multiple-shot re-identification techniques, and the 3DPeS dataset [10] for peoples re-identification in non-overlapping multi-camera system.

One of the most used datasets for person re-identification is the Viewpoint Invariant Pedestrian Recognition (VIPeR) which contains 632 pairs of pedestrian images taken from 2 camera views with variation in pose and lighting conditions. These variations make it a very challenging dataset for re-identification. But in the other hand, just one image per person per camera is given. Thus, the use of this dataset is restricted to the single-shot re-identification techniques.Some illustrative examples of this dataset are shown in Fig. 4.

The CAVIAR dataset consist of videos recorded for people tracking. These videos are recorded for different scenarios such as persons walking alone, persons meeting with others, persons entering/leaving shops, and persons leaving packages in public places. A variant for this dataset was created for person re-identification which is the CAVIAR4REID [27].

We can also mention the PRID dataset which is created to test the multiple-shot re-identification techniques. It consists of images extracted from trajectories of peoples which were recorded using two different cameras. The 475 trajectories had been recorded from the first camera view, and 856 had been recorded from the second one. Among these recording 245 persons had been recorded in both cameras. The dataset provides at least 5 images of the same person in order to test the multiple-shot re-identification techniques. Illustrative example from this dataset is shown in Fig. 5.

A more recent dataset for person re-identification in multi-camera system is the 3DPeS dataset. It provides videos from 8 surveillance cameras mounted in outdoor environment for a hundred of individual, in addition to the camera setting and the 3D reconstruction of the environment. Illustrative example from this dataset is shown in Fig. 6.

8 Conclusion

In this paper, we presented an overview of the existing methods for people's tracking in multi-camera systems. We had classified these methods into two main categories the centralized and the distributed approaches depending in data fusion between different cameras. The centralized approaches are combining data from different sensors in order to recover the ground plane occupancy map or the 3D coordinates of the tracked individuals whereas the distributed approaches are combining just the trajectories obtained by independent tracking in each camera to obtain the global trajectory of the tracked individual. In each of these categories a variety of methods were proposed. Thus, we summarized the most known method for each of these categories to get an overview that mention most of the used approaches without mentioning all the proposed methods.

References

1. Alahi A, Vandergheynst P, Bierlaire M, Kunt M (2010) Cascade of descriptors to detect and track objects across any network of cameras. Comput Vis Image Underst 114(6):624–640
2. Arsic D, Hristov E, Lehment NH, Hörnler B, Schuller B, Rigoll G (2008) Applying multi layer homography for multi camera person tracking. In: ICDSC. IEEE, pp 1–9
3. Avraham T, Gurvich I, Lindenbaum M, Markovitch S (2012) Learning implicit transfer for person re-identification. In: Fusiello A, Murino V, Cucchiara R (eds) ECCV Workshops (1), springer, lecture notes in computer science, vol 7583, pp 381–390
4. Aziz KE, Merad D, Fertil B (2011) People re-identification across multiple non-overlapping cameras system by appearance classification and silhouette part segmentation. In: AVSS. IEEE Computer Society, pp 303–308
5. Bai S, Bai X, Tian Q (2017) Scalable person re-identification on supervised smoothed manifold. In: CVPR, vol 6, pp 7
6. Bak S, Corvée E, Brémond F, Thonnat M (2010a) Person re-identification using haar-based and dcd-based signature. In: AVSS. IEEE Computer Society, pp 1–8
7. Bak S, Corvée E, Brémond F, Thonnat M (2010b) Person re-identification using spatial covariance regions of human body parts. In: AVSS. IEEE Computer Society, pp 435–440
8. Bak S, Charpiat G, Corvee E, Bremond F, Thonnat M (2012) Learning to Match Appearances by Correlations in a Covariance Metric Space. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C (eds) 12th European Conference on Computer Vision, vol 7574. Springer, Florence, pp 806–820. https://doi.org/10.1007/978-3-642-33712-3_58, https://hal.inria.fr/hal-00731792
9. Baltieri D, Vezzani R, Cucchiara R (2010) 3d body model construction and matching for real time people re-identification. In: Puppo E, brogni A, Floriani LD (eds) Eurographics Italian Chapter Conference, Eurographics, pp 65–71
10. Baltieri D, Vezzani R, Cucchiara R (2011a) 3dpes: 3d people dataset for surveillance and forensics. In: Proceedings of the 1st International ACM Workshop on Multimedia access to 3D Human Objects, Scottsdale, pp 59–64
11. Baltieri D, Vezzani R, Cucchiara R, Utasi A, Benedek C, Szirányi T (2011b) Multi-view people surveillance using 3d information. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE, pp 1817–1824
12. Bazzani L, Cristani M, Perina A, Murino V (2012) Multiple-shot person re-identification by chromatic and epitomic analyses. Pattern Recogn Lett 33(7):898–903
13. Bhuvana VP, Schranz M, Regazzoni CS, Rinner B, Tonello AM, Huemer M (2016) Multi-camera object tracking using surprisal observations in visual sensor networks. EURASIP J Adv Signal Process 2016(1):1–14

14. Bouma H, Borsboom S, den Hollander RJ, Landsmeer SH, Worring M (2012) Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination. In: SPIE Defense, Security, and Sensing. International Society for Optics and Photonics, pp 83,590Q–83,590Q
15. Bredereck M, Jiang X, Körner M, Denzler J (2012) Data association for multi-object tracking-by-detection in multi-camera networks. In: 2012 Sixth International Conference on Distributed smart cameras (ICDSC). IEEE, pp 1–6
16. Brendel W, Amer M, Todorovic S (2011) Multiobject tracking as maximum weight independent set. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1273–1280. https://doi.org/10.1109/CVPR.2011.5995395
17. Caviar:context aware vision using image-based active recognition (2004) http://homepages.inf.ed.ac.uk/rbf/CAVIAR/
18. del Blanco CR, Mohedano R, García NN, Salgado L, Jaureguizar F (2008) Color-based 3d particle filtering for robust tracking in heterogeneous environments. In: ICDSC. IEEE, pp 1–10
19. Chang X, Ma Z, Yang Y, Zeng Z, Hauptmann AG (2017a) Bi-level semantic representation analysis for multimedia event detection. IEEE Trans Cybern 47(5):1180–1197
20. Chang X, Yu YL, Yang Y, Xing EP (2017b) Semantic pooling for complex event analysis in untrimmed videos. IEEE Trans Pattern Anal Mach Intell 39(8):1617–1632
21. Chen W, Cao L, Chen X, Huang K (2014) A novel solution for multi-camera object tracking. In: 2014 IEEE International Conference on Image Processing (ICIP). IEEE, pp 2329–2333
22. Chen W, Cao L, Chen X, Huang K (2016)
23. Chen Y, Zhu X, Gong S et al (2018) Person re-identification by deep learning multi-scale representations
24. Colombo A, Orwell J, Velastin S et al (2008) Colour constancy techniques for re-recognition of pedestrians from multiple surveillance cameras. In: Workshop on multi-camera and multi-modal sensor fusion algorithms and applications-m2SFA2
25. Cong DNT, Khoudour L, Achard C, Meurie C, Lezoray O (2010) People re-identification by spectral classification of silhouettes. Signal Process 90(8):2362–2374
26. Cosar S, Çetin M (2013) A sparsity-driven approach to multi-camera tracking in visual sensor networks. In: AVSS. IEEE, pp 407–413
27. Dong Seon C, Marco C, Michele S, Loris B, Vittorio M (2011) Custom pictorial structures for re-identification. In: Proceedings of the British Machine Vision Conference. BMVA Press, pp 68.1–68.11. https://doi.org/10.5244/C.25.68
28. Du W, Piater J (2006) Data fusion by belief propagation for multi-camera tracking. In: 2006 9th International Conference on Information Fusion, pp 1–8. https://doi.org/10.1109/ICIF.2006.301712
29. Du W, Piater JH (2007) Multi-camera people tracking by collaborative particle filters and principal axis-based integration. In: Yagi Y, Kang SB, Kweon IS, Zha H (eds) ACCV (1), Springer, lecture notes in computer science, vol 4843, pp 365–374
30. Eshel R, Moses Y (2010) Tracking in a dense crowd using multiple cameras. Int J Comput Vis 88(1):129–143
31. Evans M, Osborne CJ, Ferryman JM (2013) Multicamera object detection and tracking with object size estimation. In: AVSS. IEEE, pp 177–182
32. Farenzena M, Bazzani L, Perina A, Murino V, Cristani M (2010) Person re-identification by symmetry-driven accumulation of local features. In: CVPR. IEEE, pp 2360–2367
33. Felzenszwalb PF, Girshick RB, McAllester DA, Ramanan D (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
34. Figueira D, Bazzani L, Minh HQ, Cristani M, Bernardino A, Murino V (2013) Semi-supervised multi-feature learning for person re-identification. In: AVSS. IEEE, pp 111–116
35. Fleuret F, Berclaz J, Lengagne R, Fua P (2008) Multicamera people tracking with a probabilistic occupancy map. IEEE Trans Pattern Ana Mach Intell 30(2):267–282
36. Gray D, Brennan S, Tao H (2007) Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International workshop on performance evaluation of tracking and surveillance, Citeseer
37. Guan L, Franco JS, Pollefeys M (2010) Multi-view occlusion reasoning for probabilistic silhouette-based dynamic scene reconstruction. Int J Comput Vis 90(3):283–303
38. Hamdoun O, Moutarde F, Stanciulescu B, Steux B (2008) Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In: ICDSC. IEEE, pp 1–6
39. Hirzer M, Beleznai C, Roth PM, Bischof H (2011) Person re-identification by descriptive and discriminative classification. In: Heyden A, Kahl F (eds) SCIA, vol 6688. Springer Lecture Notes in Computer Science, pp 91–102

40. Hofmann M, Wolf D, Rigoll G (2013) Hypergraphs for joint multi-view reconstruction and multi-object tracking. 2013 IEEE Conf Comput Vis Pattern Recogn 0:3650–3657. https://doi.org/10.1109/CVPR.2013.468
41. Hong C, Yu J, Tao D, Wang M (2015a) Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval. IEEE Trans Ind Electron 62(6):3742–3751
42. Hong C, Yu J, Wan J, Tao D, Wang M (2015b) Multimodal deep autoencoder for human pose recovery. IEEE Trans Image Process 24(12):5659–5670
43. Huang C, Wang SJ (2012) A bayesian hierarchical framework for multitarget labeling and correspondence with ghost suppression over multicamera surveillance system. IEEE T Autom Sci Eng 9(1):16–30
44. Javed O, Shafique K, Rasheed Z, Shah M (2008) Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views. Comput Vis Image Underst 109(2):146–162
45. Jiang X, Körner M, Haase D, Denzler J (2014) A graph-based map solution for multi-person tracking using multi-camera systems. In: Proceedings of the 9th International Conference on Computer Vision Theory and Applications (VISAPP), pp 343–350
46. Jüngling K, Arens M (2011) View-invariant person re-identification with an implicit shape model. In: AVSS, IEEE Computer Society, pp 197–202
47. Kenk VS, Kovačič S, Kristan M, Hajdinjak M, Perš J et al (2015) Visual re-identification across large, distributed camera networks. Image Vis Comput 34:11–26
48. Khan SM, Shah M (2006) A multiview approach to tracking people in crowded scenes using a planar homography constraint. In: Leonardis A, Bischof H, Pinz A (eds) ECCV (4), Springer, Lecture notes in computer science, vol 3954, pp 133–146
49. Kim K, Davis LS (2006) Multi-camera tracking and segmentation of occluded people on ground plane using search- guided particle filtering. In: In ECCV, pp 98–109
50. Kroeger T, Dragon R, Van Gool L (2014) Multi-View Tracking of Multiple Targets with Dynamic Cameras. Springer, pp 653–665. https://doi.org/10.1007/978-3-319-11752-2_54
51. Köstinger M, Hirzer M, Wohlhart P, Roth PM, Bischof H (2012) Large scale metric learning from equivalence constraints. In: CVPR. IEEE, pp 2288–2295
52. Kuo CH, Huang C, Nevatia R (2010) Multi-target tracking by on-line learned discriminative appearance models. In: CVPR. IEEE, pp 685–692
53. Leal-Taixé L, Pons-Moll G, Rosenhahn B (2012) Branch-and-price global optimization for multi-view multi-target tracking. In: CVPR. IEEE, pp 1987–1994
54. Leibe B, Leonardis A, Schiele B (2008) Robust object detection with interleaved categorization and segmentation. Int J Comput Vis 77(1-3):259–289
55. Li W, Wu Y, Mukunoki M, Minoh M (2012) Common-near-neighbor analysis for person re-identification. In: 2012 19th IEEE International Conference on Image Processing (ICIP), pp 1621–1624. https://doi.org/10.1109/ICIP.2012.6467186
56. Liem MC, Gavrila DM (2014) Joint multi-person detection and tracking from overlapping cameras. Comput Vis Image Underst 128:36–50
57. Lin Z, Davis LS (2008) Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance. In: Bebis G, Boyle RD, Parvin B, Koracin D, Remagnino P, Porikli FM, Peters J, Klosowski JT, Arns LL, Chun YK, Rhyne TM, Monroe L (eds) ISVC (1), Springer, lecture notes in computer science, vol 5358, pp 23–34
58. Loy CC, Liu C, Gong S (2013) Person re-identification by manifold ranking. In: ICIP, Citeseer, vol 1, pp 5
59. Martinel N, Micheloni C, Foresti GL (2015) Kernelized saliency-based person re-identification through multiple metric learning. IEEE Trans Image Process 24(12):5645–5658
60. Martinel N, Micheloni C, Foresti GL (2016) A pool of multiple person re-identification experts. Pattern Recogn Lett 71:23–30
61. Mazzon R, Cavallaro A (2013) Multi-camera tracking using a multi-goal social force model. Neurocomputing 100:41–50
62. Mazzon R, Tahir SF, Cavallaro A (2012) Person re-identification in crowd. Pattern Recogn Lett 33(14):1828–1837
63. Medeiros H, Park J, Kak A (2008) Distributed object tracking using a cluster-based kalman filter in wireless camera networks. IEEE J Sel Top Signal Process 2(4):448–463. https://doi.org/10.1109/JSTSP.2008.2001310
64. Nakajima C, Pontil M, Heisele B, Poggio T (2003) Full-body person recognition system. Pattern Recogn 36(9):1997–2006

65. Nie W, Liu A, Su Y, Luan H, Yang Z, Cao L, Ji R (2014) Single/cross-camera multiple-person tracking by graph matching. Neurocomputing 139:220–232
66. Patino Luis FJ (2014) Multicamera trajectory analysis for semantic behaviour characterisation. In: 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, p 369–374
67. Prosser B, Zheng WS, Gong S, Xiang T (2010) Person re-identification by support vector ranking. In: Labrosse f, Zwiggelaar R, Liu Y, Tiddeman B (eds) BMVC, British Machine Vision Association, pp 1–11
68. Qian X, Fu Y, Jiang YG, Xiang T, Xue X (2017) Multi-scale deep learning architectures for person re-identification. arXiv:170905165
69. Quang MH, Bazzani L, Murino V (2013) A unifying framework for vector-valued manifold regularization and multi-view learning. In: Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp 100–108
70. Santos TT, Morimoto CH (2011) Multiple camera people detection and tracking using support integration. Pattern Recogn Lett 32(1):47–55. https://doi.org/10.1016/j.patrec.2010.05.016
71. Schumann A, Monari E (2014) A soft-biometrics dataset for person tracking and re-identification. In: 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp 193–198. https://doi.org/10.1109/AVSS.2014.6918667
72. Schwartz WR, Davis LS (2009) Learning discriminative appearance-based models using partial least squares. In: SIBGRAPI. IEEE Computer Society, pp 322–329
73. Su C, Zhang S, Xing J, Gao W, Tian Q (2018) Multi-type attributes driven multi-camera person re-identification. Pattern Recogn 75:77–89
74. Taj M, Cavallaro A (2011) Distributed and decentralized multi-camera tracking. IEEE Signal Process Mag 28(3):46–58
75. Tuzel O, Porikli F, Meer P (2006) Region covariance: a fast descriptor for detection and classification. In: Leonardis A, Bischof H, Pinz A (eds) ECCV (2), Springer, lecture notes in computer science, vol 3952, pp 589–600
76. Vezzani R, Baltieri D, Cucchiara R (2013) People reidentification in surveillance and forensics: A survey. ACM Comput Surv 46(2):29:1–29:37. https://doi.org/10.1145/2543581.2543596
77. Wang X, Doretto G, Sebastian T, Rittscher J, Tu PH (2007) Shape and appearance context modeling. In: ICCV. IEEE, pp 1–8
78. Wang Z, Hu R, Liang C, Yu Y, Jiang J, Ye M, Chen J, Leng Q (2016) Zero-shot person re-identification via cross-view consistency. IEEE Trans Multimed 18(2):260–272
79. Wang H, Yan Y, Hua J, Yang Y, Wang X, Li X, Deller JR, Zhang G, Bao H (2017a) Pedestrian recognition in multi-camera networks using multilevel important salient feature and multicategory incremental learning. Pattern Recogn 67:340–352
80. Wang J, Qian L, Azimi E, Kazanzides P (2017b) Prioritization and static error compensation for multi-camera collaborative tracking in augmented reality. In: 2017 IEEE Virtual Reality (VR). IEEE, pp 335–336
81. Wang J, Wang Z, Gao C, Sang N, Huang R (2017c) Deeplist: Learning deep features with adaptive listwise constraint for person reidentification. IEEE Trans Circuits Syst Video Techn 27(3):513–524
82. Wang Z, Hu R, Chen C, Yu Y, Jiang J, Liang C, Satoh S (2017d) Person reidentification via discrepancy matrix and matrix metric. IEEE transactions on cybernetics
83. Wen L, Lei Z, Chang MC, Qi H, Lyu S (2016) Multi-camera multi-target tracking with space-time-view hyper-graph. Int J Comput Vis 122(2):1–21
84. Wen L, Lei Z, Chang MC, Qi H, Lyu S (2017) Multi-camera multi-target tracking with space-time-view hyper-graph. Int J Comput Vis 122(2):313–333
85. Yao J, Odobez JM et al (2008) Multi-camera multi-person 3d space tracking with MCMC in surveillance scenarios. In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008, http://hal.inria.fr/inria-00326747/
86. Youlu W (2013) Distributed multi-object tracking with multi-camera systems composed of overlapping and non-overlapping cameras. PhD thesis, University of Nebraska-Lincoln
87. Zhang Z, Cheng J et al (2013) Multi-camera tracking helmet system. Journal of Image and Graphics 1(2):76–79
88. Zhao R, Ouyang W, Wang X (2013) Person re-identification by salience matching. In: Proceedings of the IEEE International Conference on Computer Vision, pp 2528–2535
89. Zheng WS, Gong S, Xiang T (2013) Reidentification by relative distance comparison. IEEE Trans Pattern Anal Mach Intell 35(3):653–668

Rabah Iguernaissi is a PhD student in the University of Aix-Marseille, France. He received a master degree in international biometrics from the university of Paris-Est Val-de-Marne, France, in 2014. He also received an electrical engineer degree specialized in control systems from the University of Boumerdes, Algeria, in 2011. He is working in the behavioral analysis of customers based in image and video processing. His research interests include computer vision and image processing.