Image and Vision Computing 32 (2014) 270–286

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Editor's Choice Article

A survey of approaches and trends in person re-identification ☆

Apurva Bedagkar-Gala, Shishir K. Shah *

Quantitative Imaging Laboratory, University of Houston, Department of Computer Science, Houston, TX 77204-3010, USA

ARTICLE INFO

Article history:
Received 29 July 2013
Received in revised form 12 January 2014
Accepted 13 February 2014
Available online 21 February 2014

Keywords:
Person re-identification
Multi-camera tracking
Video surveillance
Closed set Re-ID
Open set Re-ID
Short and long period Re-ID

ABSTRACT

Person re-identification is a fundamental task in automated video surveillance and has been an area of intense research in the past few years. Given an image/video of a person taken from one camera, re-identification is the process of identifying the person from images/videos taken from a different camera. Re-identification is indispensable in establishing consistent labeling across multiple cameras or even within the same camera to re-establish disconnected or lost tracks. Apart from surveillance it has applications in robotics, multimedia and forensics. Person re-identification is a difficult problem because of the visual ambiguity and spatiotemporal uncertainty in a person's appearance across different cameras. These difficulties are often compounded by low resolution images or poor quality video feeds with large amounts of unrelated information in them that does not aid re-identification. The spatial or temporal conditions to constrain the problem are hard to capture. However, the problem has received significant attention from the computer vision research community due to its wide applicability and utility. In this paper, we explore the problem of person re-identification and discuss the current solutions. Open issues and challenges of the problem are highlighted with a discussion on potential directions for further research.

1. Introduction

Large networks of cameras are increasingly deployed in public places like airports, railway stations, college campuses and office buildings. These cameras typically span large geospatial areas and have non-overlapping fields-of-views (FOVs) to provide enhanced coverage. Such networks provide huge amounts of video data, which is either manually monitored by law enforcement officers or utilized after the fact for forensic purposes. Human monitoring of these videos is erroneous, time consuming and expensive, thereby severely reducing the effectiveness of surveillance. Automated analysis of large amounts of video data can not only process the data faster but significantly improve the quality of surveillance [1]. Video analysis can enable long term activity and behavior characterization of people in a scene. Such analysis is required for high-level surveillance tasks like suspicious activity detection or undesirable event prediction for timely alerts to security personnel making surveillance more pro-active [2].

Understanding of a surveillance scene through computer vision requires the ability to track people across multiple cameras, perform crowd movement analysis and activity detection. Tracking people across multiple cameras is essential for wide area scene analytics and person re-identification is a fundamental aspect of multi-camera tracking. Re-identification (Re-ID) is defined as a process of establishing correspondence between images of a person taken from different cameras. It is used to determine whether instances captured by different cameras belong to the same person, in other words, assign a stable ID to different instances of the person. Fig. 1 shows an example of a surveillance area monitored by multiple cameras with non-overlapping FOVs. The figure shows the top view of a building floor plan and the relative placement of the cameras with respect to the building. Colored dots depict different people and numbers besides the dots are the IDs assigned to the people. The dotted lines with arrows represent the directions in which certain people move through the camera network.

As a person moves from one camera's FOV into another camera's FOV, Re-ID is used to establish correspondence between disconnected tracks to accomplish tracking across the multiple cameras. Thus, single camera tracking along with Re-ID across cameras allows for the reconstruction of the trajectory of a person across the larger scene. Person Re-ID is a non-trivial task, but is critical in improving the semantic coherence of analysis. Re-ID is relevant for surveillance applications with a single camera as well. For example, to determine if a person visits a particular location multiple times or if the same or different person picks up an unattended package/bag. Beyond surveillance it has applications in robotics, multimedia, and more popular utilities like automated photo tagging or photo browsing [3].

Person Re-ID as a task is quite simple to understand. As humans, we do it all the time without much effort. Our eyes and brains are trained to detect, localize, identify and later re-identify objects and people in the real world. Re-ID implies that a person that has been previously seen is identified in their next appearance using a unique descriptor of the person.

☆ Editor's Choice Articles are invited and handled by a select rotating 12 member Editorial Board committee. This paper has been recommended for acceptance by Xiaogang Wang.

* Corresponding author at: University of Houston, Dept. of Computer Science, 4800 Calhoun, 501 PGH, Houston, TX 77204-3010, USA.
E-mail addresses: avbedagk@central.uh.edu (A. Bedagkar-Gala), sshah@central.uh.edu (S.K. Shah).

http://dx.doi.org/10.1016/j.imavis.2014.02.001
0262-8856/

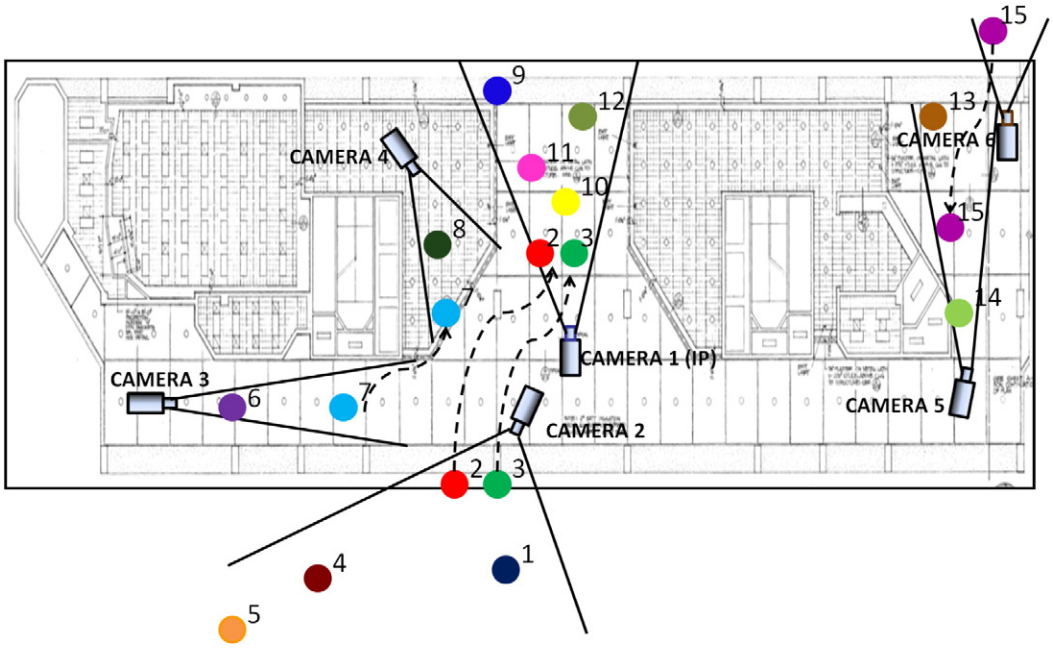

Fig. 1. Multi-camera surveillance network illustration of Re-ID.

Humans are able to extract such a descriptor based on the person's face, height and built, clothing, hair color, hair style, walking pattern, etc. A person's face is the most unique and reliable feature that humans use to identify people. Automation of person Re-ID on the other hand is quite difficult to accomplish without human intervention.

2. Person Re-ID: task and its challenges

In general, person Re-ID is difficult to automate for a number of reasons, which we will discuss later in this section, but the main challenge to Re-ID comes from the variation in a person's appearance across different cameras. Fig. 2 shows images of a person taken by different cameras on the same and different days, highlighting the variations in appearance. The top row illustrates the changes in appearance of a person across different cameras. It is also interesting to note that the appearance changes significantly within the same camera view as well.

A typical Re-ID system has two basic components: capturing a unique person descriptor or model and then comparing two models to infer either a match or a non-match. In order to learn a unique person descriptor, the ability to automatically detect and track people in images or videos is required. Fig. 3 shows a schematic representation of a Re-ID system, its two components and the sub-components within each component. To automate each component, a series of tasks need to be accomplished, which present their own challenges and contribute to the complexity of Re-ID.

2.1. System-level challenges

A typical Re-ID system may have an image (single-shot) or a video (multi-shot) as input for feature extraction and descriptor generation. For an image input the person must be reliably detected and localized for accurate feature extraction. If multiple images are available, in order ensure that the features extracted belong to the person of interest,

Fig. 2. Images of the same person taken from different cameras to illustrate the appearance changes. The top row images were captured on the same day, bottom row images were captured on different days.

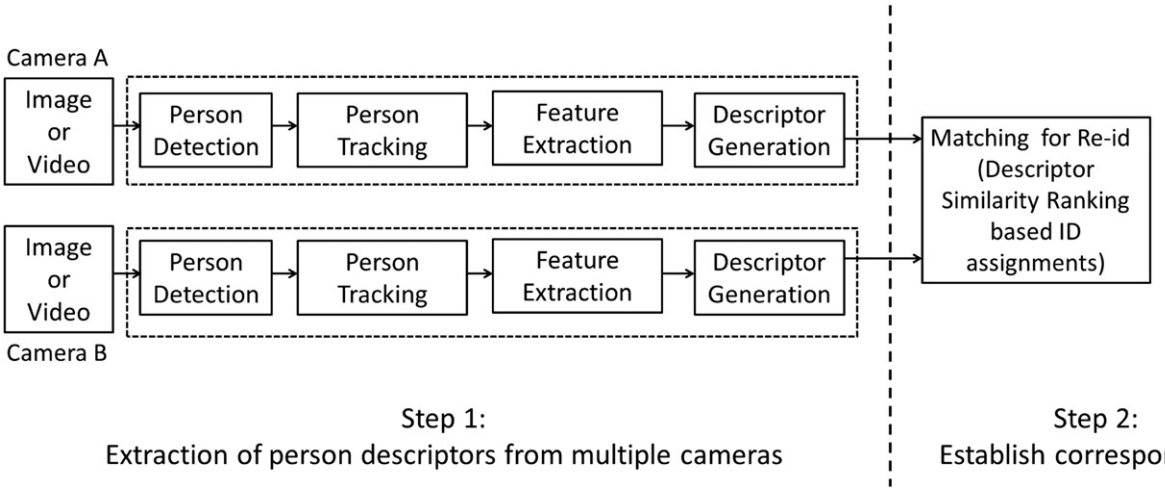

Fig. 3. Re-ID system diagram.

we need the ability to establish correspondence between detected subjects across frames. This process is also called tracking and it provides a consistent label to each subject in multiple frames. Thus, multiple instances of a person can be used for feature extraction and subsequent descriptor generation to be used for Re-ID.

Person detection and multiple person tracking are difficult problems with their own hurdles. Significant amount of work has gone into the problem of person detection over the years [4–7]. Multiple Object Tracking (MOT) within a single camera's FOV has also been widely researched and many algorithms have been proposed [8–12] over the past two decades, but sustained tracking under varying observation environments remains an open problem.

2.2. Component-level challenges: descriptor issues

Assuming that accurate person detection and single camera tracking are possible, the first step in Re-ID is to learn a person's visual descriptor. Robust and discriminative visual descriptors need to be extracted from data that is captured in unconstrained environments where the people may not be co-operative and the environments are uncontrolled. Besides, people can be partially or completely occluded due to crowds or clutter. It is difficult to ensure high quality of visual data as factors like resolution, frame rate, imaging conditions and imaging angles vary widely and cannot always be controlled. Thus, extracting a reliable descriptor is dependent upon availability of good quality observations. Incorrect detections and faulty trajectory estimation introduce errors in the descriptor extraction and generation process that directly affect the quality of Re-ID.

The simplest and most obvious descriptor of a person that can be easily obtained from video data is *appearance*, characterized by features like color and texture. Shape is another descriptor that is extractable. However, these descriptors are hardly unique and prone to variations. Color/texture descriptors are not sufficiently descriptive and vary drastically due to cross view illumination variations, pose variations or view angle or scale changes inherent in a multi-camera setting. Articulated nature of human body leads to deformable shapes of silhouettes and different camera geometries make shape descriptors less discriminative.

Since the person descriptors come from different cameras, the nature of separation between the cameras dictates the difficulty in Re-ID. For example, if the two images are taken only a few minutes or hours apart then appearance based descriptors could prove reasonable to use in Re-ID. The assumption being that people will most probably be in the same clothes, as clothing is a major contributor to appearance. This does not mean that clothing is the best descriptor in this scenario but is a reasonable one. We will refer to this type of Re-ID scenario as *short-period Re-ID*. Whereas, if the images/video are taken days or months apart, the Re-ID is called *long-period Re-ID*. In Fig. 2, the images shown in the bottom row are of the same person captured from different camera on different days. This figure perfectly illustrates the fragile nature of appearance based descriptors. The temporal separation between the images is a factor in the complexity of Re-ID. Thus, person Re-ID requires robust yet unique descriptors, which are extremely difficult to extract automatically.

2.3. Component-level challenges: correspondence issues

Comparing person descriptors is challenging due to the uncertainty attributed to the possible lack of prior known spatio-temporal relationships between cameras. Appearance of the same person can change dramatically due to other objects like bags, unzipped jackets across front and back views, etc. At the same time appearance of different people might be rather similar. This implies that within class variations can be large where as inter-class variations may be relatively smaller. Moreover, even if the person's descriptors can be captured effectively, matching them across cameras in the presence of large number of people observed is non-trivial. Comparing person descriptors across large number of potential candidates is a hard task as the descriptors are captured in different locations, time instants, and over different durations. Complexity of the matching process further increases, as increase in the number of candidates leads to loss of descriptor specificity, increasing the possibility of matching errors. It is also a compute and memory intensive process.

Person Re-ID is a broad and difficult problem with numerous open issues. In this section we discussed the general problem of person Re-ID and its broader challenges. However, person Re-ID can be constrained by the context in which it is applied. In the next section, we will explore the context specific nature of the problem. Section 3 provides an overview of the current research work in the field. We adopt a methodology-based taxonomy to classify the methods and better understand the current trends. In Section 4, we discuss the evaluation techniques and present the datasets currently used to conduct Re-ID experiments. Section 5 highlights the deficiencies of current Re-ID models, and more importantly, points out the unaddressed issues in person Re-ID. Section 6 concludes the paper.

3. Person Re-ID scenarios

In the previous section, we presented the general definition of person Re-ID and discussed the implementation pipeline and associated challenges. However, the Re-ID problem can be split into two scenarios: *open set Re-ID* and *closed set Re-ID*. A Re-ID system is similar to a recognition system, which comprises of a gallery set (set of known people) and

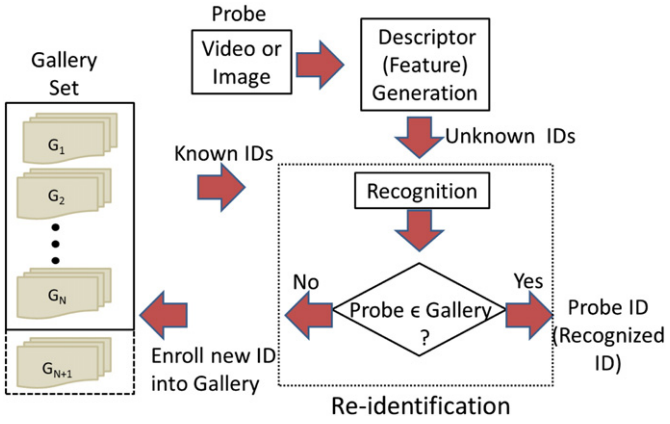

Fig. 4. Re-ID as a recognition system.

the probe (unknown person) on which the recognition has to be performed. Fig. 4 depicts the Re-ID system setup as a recognition system.

Let the gallery set be represented as $G = (g_1, g_2, \ldots, g_N)$. Thus, the set of known IDs is given by $id(G) = (id(g_1), id(g_2), \ldots, id(g_N))$, where the function $id(.)$ specifies the ID assigned to its argument. Let $P = (p_1, p_2, \ldots, p_M)$ represent the probe set, which means that the set of unknown IDs is given by $id(P) = (id(p_1), id(p_2), \ldots, id(p_M))$. Typically in a recognition framework, when the probe is presented to the system, it is compared to each gallery and similarity measures are computed. The gallery is ranked using the similarity in order to determine the probe ID. The same setup applies to the problem of Re-ID. Closed set Re-ID is the scenario where the probe is a subset of the gallery, i.e. the probe ID exists in the gallery and the objective is to determine the true ID of the probe. Thus, given that $id(P) \subseteq id(G)$, the true probe ID for a given probe p_j is $id(p_j) = id(g_{i*})$, such that,

$$i* = \underset{i \in 1,\ldots,N}{\arg\max} p\left(g_i|p_j\right) \tag{1}$$

where, $p(g_i|p_j)$ is the likelihood that $id(p_j) = id(g_i)$ and is most often represented by a similarity measure. This implies that the top ranked gallery ID is assigned to the probe. In open set Re-ID on the other hand, the probe may or may not be a subset of the gallery. This implies the open set Re-ID objective is to first establish if the probe ID is a part of the gallery, and if so, determine the true probe ID. Thus, in order to find the true ID, in addition to ranking the gallery elements and determining $i*$ using Eq. 1, the following condition also needs to be satisfied,

$$p\left(g_{i*}|p_j\right) > \tau. \tag{2}$$

In Eq. 2, τ is the acceptable level of certainty above which we can be reasonably assured that $id(p_j) \subseteq id(G)$. If this condition is not satisfied, then it is determined that the probe is not a part of the gallery. If so, the probe ID is then to be enrolled into the gallery. The process of determining a previously unknown ID is called *novelty detection*. Similar to identification tasks, the closed set Re-ID is a constrained form of open set Re-ID. The Re-ID application dictates the matching scenario. For instance, to achieve consistent tracking over multiple cameras for global trajectory of a person over a camera network requires open set Re-ID. On the other hand, identity based retrieval (for forensic applications), i.e. the ability to identify multiple observations of a particular person is a closed set Re-ID problem.

3.1. Open set Re-ID

Person Re-ID in the context of tracking across multiple cameras is an open set matching problem where the gallery evolves over time, the probe set dynamically changes for each camera FOV, and all the probes within a set are not necessarily a subset of the gallery. Additionally, there might be several subjects that co-exist in time and need to be re-identified simultaneously. Thus, it is not a single person but a multiple person Re-ID problem. We will explore the open set Re-ID problem by illustrating Re-ID in multiple camera tracking.

Fig. 5 shows a schematic of a camera network (with 4 cameras) and the evolution of the gallery, where the Re-ID is done across each camera pair. For ease of illustration, let us assume that all subjects in the figure are moving in the direction depicted by the red arrow and the tracking across the network starts at $t = 0$ from camera A. In other words, there is no prior gallery set and tracking (Re-ID) progresses from camera A through D. Additionally, the subjects appear in the FOV of camera C before they appear in the FOV of camera D. The first time a person is seen in camera A, his/her appearance model is learned, and the subject is enrolled in the gallery set. Thus, all people observed in the camera B form the probe set. After Re-ID, all the people observed in the camera B who were previously unseen are enrolled into the gallery. As the Re-ID moves to the next camera pair (camera B and C), the gallery set is extended. The open set Re-ID can be summarized as a many-to-many matching problem. In tracking scenario, Re-ID provides a means of connecting subjects' tracks that were disconnected due to the subject entering an area not in the FOV of the camera network.

3.2. Closed set Re-ID

Person Re-ID in the context of identity retrieval is closer to the classic closed set matching problem, where a single probe is presented and the gallery size is fixed. In a typical multi-camera identity retrieval scenario, the person whose multiple observations throughout the network are to be detected is the probe subject and his/her appearance model is assumed to be available. The gallery set is a set of people IDs seen in selected or all the cameras over a specified period of time. The time interval specified can be different for different cameras. In other words, the gallery is comprised of subjects seen in many different cameras constrained by time and space. Additionally, the probe can simultaneously match to gallery subjects coming from different cameras, i.e. multiple instances of the probe can be detected within the gallery. After Re-ID, multiple observations of the probe subject across the gallery cameras are detected. As the probe subject to be re-identified changes, the cameras and time intervals to be searched change and so does the gallery. However, for Re-ID of a particular probe the gallery remains fixed. Thus, the closed set Re-ID is a one-to-many matching problem.

4. Current work in person Re-ID

Re-ID has been a topic of intense research in the past five years [13–17]. In almost all of the research, the problem of Re-ID has been widely treated as a retrieval or recognition problem. Given an image or multiple images of an unknown person (probe) and a gallery set that consists of a number of known people, the objective is to produce a ranked list of all the people in gallery based on their visual similarity with the unknown person. The expectation is that the highest ranked match in the gallery will provide an ID for the unknown person, thereby identifying the probe. Here the assumption is that the probe ID is a subset of the gallery of known individuals, i.e. closed-set Re-ID. Current state-of-the-art methods attempt to solve the closed set Re-ID problem.

Most of the current approaches rely on appearance based similarity between images to establish correspondences. The typical features used to quantify appearance are low level color and texture extracted from clothing. A review of appearance descriptors for Re-ID is presented in [18]. However, such appearance features are only stable over short time intervals as people dress differently on different days. Thus, appearance based models are only suited for short-period Re-ID. All of the state-of-the-art approaches attempt solutions to short period

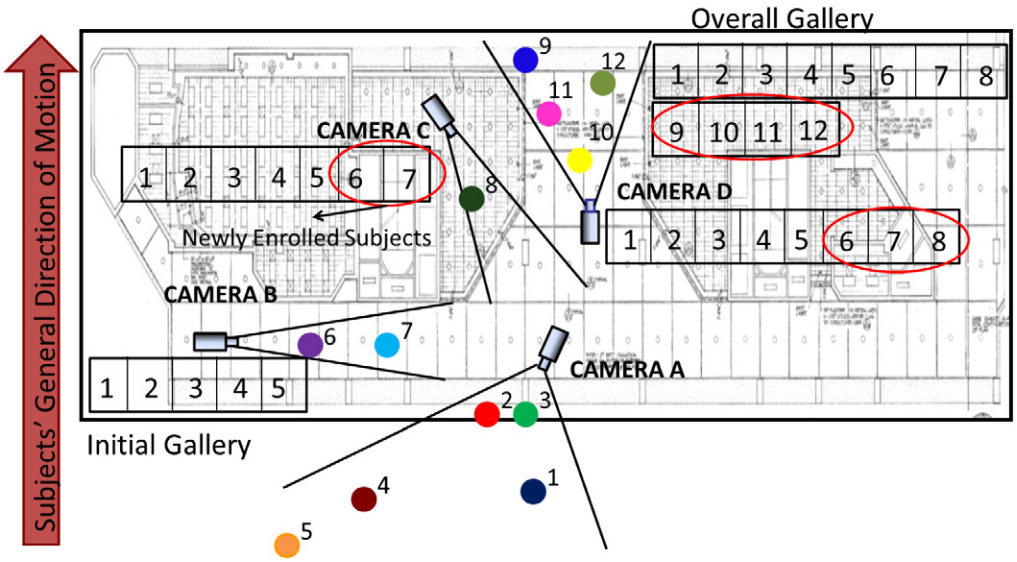

Fig. 5. Multi-camera tracking scenario based on open set Re-ID.

Re-ID. Earlier research on Re-ID focused on combining inter-camera relationships with the matching process, but more recent efforts have focused on developing discriminative features, learning distance models, or both, for robust matching. In general, recent approaches have focused on two aspects of the solution: 1) design of discriminative, descriptive and robust visual descriptors to characterize a person's appearance; and 2) learning suitable distance metrics that maximize the chance of a correct correspondence. Overall the methods for Re-ID can be broadly classified into Contextual and Non-contextual methods. Fig. 6 provides a methodology based taxonomy that summarizes the state-of-the-art research in person Re-ID.

4.1. Contextual methods

These methods rely on external contextual information either for pruning correspondences or extracting features for Re-ID. They can be further classified as those that utilize camera geometry information or those that incorporate camera calibration as the context.

4.1.1. Camera geometry as context

The early work in person Re-ID focused on leveraging spatial and temporal relationships between cameras to reduce the Re-ID errors by limiting the size of the gallery set. Space-time cues are exploited in [19] to learn inter-camera relationships that are in turn used to constrain correspondences across cameras. These relationships are modeled as a probability density function of space-time parameters like entry and exit locations, velocities, and transition times between cameras. Entry–exit points of each camera and transition times between cameras are learned in [20], in order to calibrate all the cameras in the network. The calibrated cameras are used to learn the topology of the camera network as a bipartite graph. The topology is further augmented with temporal information to achieve a tempo-topographical model of the camera network. A similar approach is used to calibrate the camera network and estimate trajectory of targets in the network using MAP estimation in [21]. Propagation of people trajectories is used in [22] to identify areas of interest in the unobserved regions within cameras. These areas are further used to choose potential paths people might take, limiting the reappearance areas in the subsequent camera's FOV to constrain Re-ID.

The topology of cameras is determined by correlating activities across cameras with disjoint FOVs in [23,24] and hence do not rely on tracking information. The FOVs of the cameras are segmented into regions within which activity patterns are similar. Spatial and temporally causal relationships across these regions in different cameras are modeled using canonical correlation analysis [25]. Affinity matrices are used to infer camera spatio-temporal camera topologies to aid Re-ID. A similar idea is built on in [26]. Here the relationships between activities are learned using MAP estimate that is continually updated at each time instant. A comprehensive review of camera topology estimation methods is presented in [27] and a study of scalability of topology estimation is performed in [28].

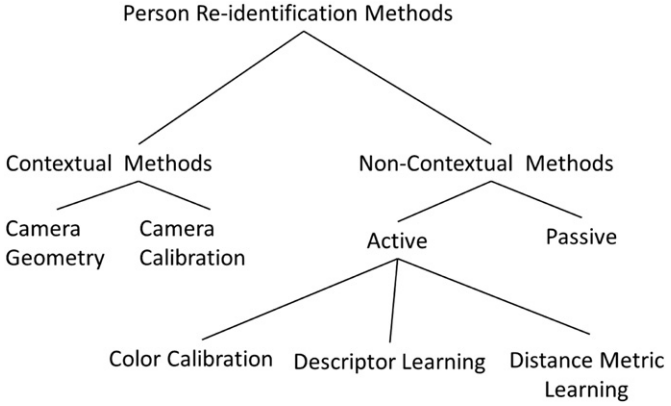

Fig. 6. Methodology based taxonomy of Re-ID approaches.

4.1.2. Camera calibration as context

In these methods, camera calibration or homography is exploited to extract unique and discriminative features to augment the visual descriptors used for Re-ID. The height of a person is determined using homography based 3D position estimation in [29]. The human silhouette is divided into three parts from top to bottom at specified proportions to the blob height. Each region is then represented with dominant color and edge energy texture descriptors. Integrated region matching is used for human silhouette similarity computations to achieve Re-ID. Similar height extraction method is used in [30]. The height along with clothing color and body build is used as a feature to establish a match.

A panoramic appearance map (PAM) proposed in [31] extracts and combines information from all the cameras that view the object to generate a single object signature. Multiple camera triangulation is used to determine the position of the object and a cylindrical surface model is

placed at its location. A parametric surface grid is projected onto all cameras where the object is visible and corresponding image patches are extracted. The features or pixel colors from these extracted patches are integrated to form the PAM, which is used for Re-ID. Maps from different tracks are compared using weighted sum of a squared distance metric. However, to generate the appearance signature, the object needs to be visible in at least 3 cameras with overlapping views simultaneously. Camera calibration and accurate 3D positioning are needed to accomplish Re-ID.

The principal axis of each person, i.e. the axis of symmetry of the human body is detected in [32] to match people across camera views. Landmarks on the ground plane shared by two cameras are used to estimate homography. The intersection of the principal axis of a person in one view and transformed principal axis of a person in another view using homography is used to compute a degree of match between people from different cameras. The degree of match is used to compute correspondence likelihood to establish Re-ID. However, the accuracy of the detected principal axis depends on the accurate segmentation of the human silhouette from the foreground and hence is prone to errors in crowded scenes and cluttered backgrounds.

A 3D point process model is used for detection and representation for person matching in [33]. The placement and orientation of the 3D model is determined using camera calibration and tracking information. Each model vertex is represented by a number of appearance features, namely, HSV histogram, mean color, direction of normal to vertex, optical reliability of vertex and vertex uniqueness. Re-ID score is a product of distances between HSV histograms weighted by vertex reliabilities and vertex saliency distances. As is evident, the main drawback of these methods is their reliance on camera calibration. With large camera networks, calibration of all cameras is not feasible.

4.2. Non-contextual methods

Several approaches have been proposed that rely entirely on the analysis of visual descriptors and no external contextual information is incorporated to assist the correspondence process. These methods can be further classified as active and passive methods. Most of the recent research is focused on non-contextual methods. A popular classification within this class is based on whether single image (*single-shot*) or multiple images (*multi-shot*) are used to generate and compare the appearance descriptors. There are many different non-contextual techniques for Re-ID and in order to provide an overview of some of the more prominent approaches, a tabular summary is presented in Table 1. The approaches are distinguished based on the type of features used, single/multi-shot incorporation and the incorporation of false match rejection (novelty detection) in the matching framework.

4.2.1. Passive methods

These methods deal with design of descriptive visual descriptors to characterize the person's appearance and compare these by computing similarity measures to achieve Re-ID. These methods are termed as passive as they do not rely on learning techniques, supervised or unsupervised, for descriptor extraction and matching.

A color and shape features based appearance model from the detected blob is proposed in [50]. The blob is segmented into multiple polar bins and the color Gaussian model and edge pixels counts from each bin form the descriptor. A match is established using three similarity measures and the optimal match is the one that maximizes all the similarity measures. A spatiotemporal segmentation algorithm based on watershed segmentation and graph partitioning is used in [16] to detect stable spatiotemporal edges called edgels. The appearance of a person is a combination of color (hue and saturation) and edgel histograms and the intersection histogram is used to establish a match between observations. A non-surveillance application of person Re-ID was explored in [3], where the objective was to find all occurrences of person in a sequence of photographs taken over a short period of time. A two step approach is adopted, where the first step, identifies different people that exists in the photographs by clustering frontal face detections. The clustering is based on 16-bin RGB histograms extracted from clothing. In the second step, color features based pictorial structures are used to find each person identified in the previous step, even in photographs where their frontal faces cannot be seen. Each part identified by the pictorial structure is represented by 5 component Gaussian mixture model. This approach assumes that each person is facing the camera in at least one photograph in the sequence and that people are distinguishable by their clothing color.

The human silhouette is represented by two complementary appearance features in [14]. The first feature is an HSV histogram that encodes the global appearance while the local appearance is encoded using a set of recurrent local patches using epitomic analysis. The appearance matching is based on a weighted sum of feature similarities. The features are extracted over multiple images of a person and are termed as Histogram Plus Epitome (HPE). The human silhouette is divided into head, torso and leg regions by detecting 2 horizontal axes of asymmetry and one vertical axis of symmetry in [15]. Each part is then described using 3 features, weighted HSV histogram, maximally stable color regions (MSCR) [51] and recurrent highly textured local patches. Again, the appearance matching is based on a weighted sum of feature similarities. The features extracted are combined over

Table 1
Non-contextual person Re-ID approaches.

Approach type	Approaches	Structural information	Images used for descriptor	Features	False match rejection
Passive	Spatiotemporal model [16]	✓	Multiple	Color, edges	×
	SDALF [15]	×	Single/multiple	Color, texture	×
	SCR [34]	✓	Single	Position, color, gradients	×
	Multi-feature Model [35]	✓	Multiple	Color, face	✓
	BiCov [36]	✓	Multiple	Color, texture	×
	CPS [37]	✓	Single/multiple	Color	×
Descriptor learning	ELF [38]	✓	Single	Position, color, gradients	×
	PLS [39]	×	Single	Color, texture, HOG	×
	Shape context [17]	✓	Single	Shape, color, texture, HOG	×
	Group context [40]	×	Single	Color, texture, HOG	×
	Boosted Re-ID [41]	✓	Multiple	Position, color, gradient	×
	Re-ID with attributes [42]	✓	Single	Color, texture	×
	Correlation space Re-ID [43]	✓	Multiple	Position, color, gradient	×
	Re-ID by saliency [44]	✓	Single	Position, color, texture	×
Metric learning	LMNN-R [45]	×	Single	Color	✓
	PRDC [46]	×	Single	Color, texture	×
	RankSVM [47]	×	Single	Color, texture	×
	Impostor learning [48]	✓	Single	Color, texture	✓
	Fisher vector [49]	✓	Multiple	Position, color, texture	×

multiple images of a person to form an appearance model called SDALF. The Re-ID performance of the features proposed in [14] and augmented by applying it as a human part descriptor adopting the asymmetry driven part detection proposed in [15], thus defining a structure feature named Asymmetry-based HPE [52]. Pictorial structures model [53] was employed in [37] for part based human detection and each part is used to extract HSV histograms and MSCRs. The part based representation is then used for Re-ID. They proposed a slight modification of pictorial structures to better localize body parts using multiple images of a person to guide the MAP estimates of the body configuration. This approach is known as Custom Pictorial Structures (CPS). This aids the extraction of more reliable visual features to improve Re-ID.

Spatial covariance regions (SCR) are extracted from human body parts in [34] and spatial pyramid matching is used to design a dissimilarity measure. HOG based body part detector is used to detect 4 parts: torso, left arm, right arm and legs. Each detected body part is characterized by a covariance descriptors based on region colors, gradient magnitudes and orientations. These descriptors encode variances in region features, their covariances, and spatial layout. Covariance matrix distance is used to compute dissimilarity between descriptors. Covariance matrices are also adopted in [36] but the underlying features are Gabor filter response magnitude images extracted from different spatial scales (BiCov). Neighboring scale responses are grouped to form a single band and magnitude images are computed using the MAX operator within each band. The appearance model is not the covariance matrices but differences between matrices between consecutive bands.

A multiple component matching approach inspired by multiple component learning concept in object recognition is proposed in [54]. Multiple frames of a person are treated as multiple instances of the person. Thus, the descriptor is basically a collection of features extracted from multiple frames. Each image feature set is treated as an instance and gallery and probe are considered a match if at least a few pairs of instances match. The body is represented by randomly selected rectangular patches whose appearance is captured by HSV histograms. This framework is extended in [55] where person descriptors are formed by a vector of dissimilarity values to a set of predefined visual prototypes.

Interest point based descriptors collected over a number of images of a person are utilized in [56] to characterize the person's appearance. Hessian based interest points are detected using efficient integral image implementation. A histogram of Haar wavelet responses in a 4 $\times$ 4 region centered around the interest points are used as the descriptors. The descriptors are matched using sum of absolute differences metric and Re-ID is established using a best bin first (BBF) search on a KD-tree containing all gallery models. The Re-ID model is generated from tracking data in [57]. The model is generated by encoding SIFT features extracted during tracking by Implicit Shape Model [58] codebook learned offline. The spatial distance between the SIFT points also contributes towards the model. Matching high dimensional models is computationally very expensive and the major drawback of this approach. A comparative study of local features for the task of Re-ID was reported in [59] and concluded that GLOH [60] and SIFT features outperform other local features.

A part-based spatio-temporal model based on HS color histograms and representative colors was proposed in [61]. The person's body is divided into stable body parts [6] using HOG based body part detectors. The color histograms are extracted for each body part and are combined into an active color model inspired by the active appearance model [62]. Representative colors are also extracted from each body part and combined over multiple images by clustering to generate representative meta colors. The active color model and representative meta colors are used as the appearance descriptor and similarity is computed as a weighted sum over the two features. This is the only paper that tackles multiple person Re-ID problem and presents an open set matching framework for Re-ID. The same model was extended in [35] to include facial features from low resolution face images in order to assist Re-ID.

FisherFaces [63] based facial features and dense sampling of colors in luminance–chrominance space (LCC) along with horizontal and vertical edges from clothing is combined for Re-ID in [64]. Person recognition is based on a nearest neighbor classifier. Color position histogram is constructed in [65] by splitting the silhouette into fixed number of horizontal bands and characterizing each band with its mean color. Sparsified representation [66] is utilized for Re-ID. A person is represented as a graph in [67] with color features representing the nodes and region proximity dictating the graph edges. Graph edit distance based graph kernel is used for classification and hence Re-ID. Covariance descriptors based on color, Gabor and LBP features are used to characterize appearance in [68]. Fuzzy color quantization in the Lab color space is used to extract probabilistic histograms in [69]. Re-ID is based on a k-nearest neighbor classifier. Color position histogram is used to characterize silhouettes in [70] and is subsequently subjected to nonlinear dimensionality reduction to form the descriptor vector.

4.3. Active methods

These methods are termed as active as they employ supervised or unsupervised learning techniques for descriptor extraction or matching. Such learning based methods can be further classified into color calibration methods, descriptor learning and distance metric learning methods. The last two sub-categories depend on learning is employed whether to learn optimal appearance features or to learn optimal distance metrics for Re-ID.

4.3.1. Color calibration

These methods attempt to model the color relationships between a given camera pair using color calibration techniques and they need a learning stage to develop the calibration model that needs to be updated frequently to capture all desired relationships. In order to model the changes in appearance of objects between two cameras a brightness transfer function (BTF) between each pair of cameras is learned from training data in [71]. The BTF is used as a cue in establishing appearance correspondence. Learning the brightness transfer function between a pair of cameras was first proposed in [72]. Once such a mapping between cameras is learned, the Re-ID problem is reduced to one of matching transformed appearance models. However, such a mapping is not unique and it changes from frame to frame depending on varying factors like illumination, scene geometry, focal length, exposure time and aperture size of each camera. Thus, a single BTF cannot be used for matching models consistently. Javed et al. [19] show that all the BTFs between a given camera pair lie in a low dimensional subspace even in the presence of large number of unknown parameters. They propose a method to learn the low dimensional subspace from training data and use this information to determine the probability that observations taken from two different cameras belong to the same person. Prosser et al. [73] propose a cumulative BTF computation method that requires only a spare color training set and the BTF is computed by relying on the mean operation taken over multiple learned BTFs. A novel bi-directional matching criterion is also proposed for comparing individuals in order to reduce false matches.

4.3.2. Descriptor learning

This class of methods either attempt to learn the most discriminative features or a discriminative weighting scheme for multiple features to achieve Re-ID or employ a learning stage to generate descriptive dictionaries of features that better represent a person's appearance using a bag-of-features approach.

Shape and appearance context models were used in [17] where co-occurrences between a priori learned shape and appearance words form the person descriptor. The human silhouette is split into parts using a modified shape context algorithm that builds on shape dictionary learned a priori. Bag-of-features based approach is used to learn code words to characterize appearance based on HOG [4] features

computed in the log-RGB space. The human silhouette is first represented as a collection of shape labels and then the appearance descriptor is constructed using the spatial occurrence of the appearance words with respect to each shape label. Since the model is based on appearance words, learned on a training dataset, the applicability of the model is limited. Similar appearance words and visual context using local and global spatial relationships (group context) are used to describe an individual's appearance in [40]. The appearance words are based on SIFT [74] and average RGB color as features. Groups of people are represented by two descriptors. The first one aims to describe the ratio information of appearance words within and across rectangular ring regions centered on the group. The second descriptor captures more local spatial information between the labels. The obtained group descriptors are utilized as a contextual cue for person Re-ID by combining person descriptor matching cost and group matching cost. Group information is used to reduce ambiguity in person Re-ID if a person would appear in the same group. Cai and Pietikinen [13] utilize spatial distributions of self similarities with respect to learned appearance words and combine them to form the appearance descriptor. The appearance words are based on a weighted hue histogram and then for each label, its occurrence frequency in each bin of a log-polar grid centered on the image center is computed. It is used as a global color context representation to model self similarities of image patterns. Re-ID is established between person images using a nearest neighbor classifier.

Adaboost learning is employed in [38] to simultaneously learn discriminant features and ensemble of weak classifiers (ELF) for pedestrian recognition. Weak classifiers are learnt on a training set to determine the features that impart maximum discriminative ability. The color features used are histograms of RGB, HSV and YCbCr channels and texture features are histograms of Schmid and Gabor filter responses. The most discriminative characteristics of these features, such as location of features, most discriminative bin, and likelihood ratios determined by boosting, are used. The Adaboost classifier assigns a positive label to pair of images from the same person and a negative label to pair of images from different people. This study concluded that the Hue, Saturation, R, G and B channels are most discriminative in that order. Partial least squares (PLS) technique is employed to not only find discriminative weighting for color, texture and edge features but also as a means to reduce descriptor dimensionality in [39]. The observation that appearance variations across multiple cameras are multi-modal in nature is utilized in [75] by learning multiple classifiers in the joint appearance feature space across multiple cameras. Re-ID is achieved by ranking combined scores generated by all the learned classifiers.

A two step process is applied to learn discriminative features in [76]. In the first step, covariance descriptors are used to rank the gallery images as per similarity to a probe image. The first 50 images are shown to a human operator who decides whether the true match exits in this set for the probe. If not, as a second step, boosting is performed over a set of covariance descriptors based on RGB color and Haar features to select a fixed number of discriminative features that are then used to establish a match.

Haar like features and dominant color descriptors are used as features for Re-ID and to guide detection of upper and lower body of a person in [77]. Adaboost classifier is used to find the most appropriate appearance model to use for matching images of people. An extension of the covariance descriptors used in [34] is proposed in [41], where instead of using predefined body parts to extract low level features, a spatio-temporal grid over multiple images is used to extract the features. Each region of an image is used to extract the covariance matrices and these are combined over multiple frames using Riemannian mean of the covariances (MRC). The spatio-temporal MRCs that contribute to the final descriptor are selected by a boosting algorithm and a matching scheme based on Riemannian manifolds is used for Re-ID.

A binary SVM classifier is trained in [78] to learn camera-pair specific variations in the feature space. The features used to train the SVM are formed by concatenating the appearance descriptors of people across a given camera pair. If both the descriptors belong to the same person, then these are considered positive samples otherwise they are treated as negative samples. In other words, it solves the camera-specific Re-ID problem. HSV histograms extracted from 5 horizontal regions of the silhouette are used as the appearance feature. This same idea is extended in [79], given classifiers trained on camera pairs A–B and B–C, they attempt to infer the classifier for camera pair A–C. The inference is based on the notion of statistical marginalization which is approximated by summation over descriptors coming from camera B.

A view that different regions of the subject should be matched using different matching strategies and features is explored in [43]. The location of different regions of the body are represented by their distance from the body center along with color and texture features. Covariance matrices are used as a feature and human body specific matching criteria are learned using correlation based feature selection. The distance model for matching exists in the covariance feature space.

4.3.2.1. Attribute based person re-identification. The idea that certain features can be more important than others is explored in [80]. In the context of Re-ID, they attempt to determine features that are unique that distinguish a given subject from another even if their overall appearances are very similar. To determine such features they link low level features to attributes. Attributes are defined a *midlevel* features or visual concepts that have semantics attached to them, namely: stripped, furry, tall, short and so on [81]. An unsupervised approach for learning adaptive weights of different features based on their unique and inherent appearance attributes is proposed. First, a clustering stage is applied to a set of training images to discover representative prototypes of attributes. The assumption being that each prototype represents certain attributes specific to that subject. The feature's weights are then computed based on its ability to discriminate between different prototypes. An incoming probe image is then assigned to one of the prototypes to generate an attribute driven representation. A combination of all appearance features and attribute weighted features is used to rank gallery images and establish Re-ID. The underlying features are the color and texture features proposed in [38]. A similar idea is explored in [44,82]. Here distinctive or salient regions are defined as regions that discriminate an individual's appearance and are general enough to identify the person across different views. For instance, these regions could be a distinctive textured backpack or bright jacket. Person images are represented by appearance characterized patches, and patch matching under spatial adjacency constraints is used to generate an appearance descriptor in an unsupervised fashion. Re-ID is achieved by combining the salient patches with global appearance (SDALF features). We term this approach as Re-ID by saliency.

Attributes have been successfully applied over the past few years to various problems like face recognition [83] and object recognition [84–86]. In order to augment the insufficient discriminative ability of low level features, the concept of an attribute for Re-ID is refined in [87,42]. For instance, a human observer can distinguish between two people wearing very similar clothing based on distinct shoes or hair styles. Thus, attributes represent features that can be interpreted distinctly based on their perceptual semantics. 15 binary attributes: type of clothing (skirts, jeans, etc.), shoes, hair, gender and accessories are defined based on human operators and their detection using SVM based on color and texture features [38] has also been proposed. A combination of these attributes along with global appearance descriptors (SDALF features) defined in [15] is used to compute weighted similarity between gallery and probe images to establish Re-ID. We term this approach as Re-ID with attributes.

4.3.3. Distance metric learning

These methods shift the focus from feature selection based efforts to improve Re-ID to learning appropriate distance metrics that can maximize the matching accuracy regardless of the choice of appearance representation. Distance Metric learning methods [88] are extensively explored in the recognition and image retrieval problems, and they

attempt to learn a metric in the space defined by image features that keep features coming from same class closer, while, the features from different classes are farther apart. In the context of Re-ID, the image features are appearance descriptors across camera views and the aim is to learn a distance metric in the appearance space that maximizes the distance between descriptors of different people and minimizes the distance for descriptors of the same person. Metric learning is done in a supervised fashion under pairwise constraints. The training features are paired appearance descriptors and the training labels are either positive or negative depending on whether the appearance descriptors belong to the same person or different, respectively.

Let the training appearance descriptor pairs be denoted by $x_1, x_1, \ldots, x_n$, where, n denotes the number of training samples. Let the dimensionality of each sample be denoted by m. Metric learning aims to learn a distance metric, denoted by matrix $D \in R^{mxm}$, such that, distance between two appearance pairs x_i and x_j is defined as:

$$d\left(x_i, x_j\right) = \left(x_i - x_j\right)^T D\left(x_i - x_j\right) \tag{3}$$

$d(x_i, x_j)$ is a true metric as long as matrix D is symmetric positive-semidefinite. This problem is solved using convex programming as shown below:

$$\min_D \sum_{\left(x_i, x_j\right)\epsilon Pos} \left\|\left(x_i - x_j\right)\right\|_D^2 s.t. D \geqslant 0, \text{ and } \sum_{\left(x_i, x_j\right)\epsilon Neg} \left\|\left(x_i - x_j\right)\right\|_D^2 \geqslant 1 \tag{4}$$

where, *Pos* and *Neg* denote positive and negative label training sample sets denoting appearance pairs the belong to the same person and different ones, respectively.

A large margin nearest neighbor (LMNN-R) distance metric is learnt in [45] such that it minimizes the distance between true matches and maximizes false match distances. The cost was computed using 8-bin RGB and HSV histograms after subjecting them to principal component analysis for dimensionality reduction. The metric learned has the capacity to reject matches based on a universal learnt threshold on the matching cost. It was shown in [89] that by using a slight modification in the feature vector extraction using overlapping regions of the human blob, the LMNN-R metric can provide greater robustness to Re-ID under occlusion and scale changes. A similar idea is explored in [46] that solves metric learning in a probabilistic manner termed as probabilistic relative distance learning (PRDC). They focus on maximizing the probability that a true match pair has a smaller distance than a false matched pair. Re-ID is cast as a tracklet matching problem in [90] and dynamic time warping distance is used as a metric to match tracklets. Dynamic time warping distance based large margin nearest neighbor metric learning is adopted.

The person Re-ID problem is treated as a relative ranking problem in [47], the idea being not comparing direct distance scores between correct and incorrect matches, instead to learn a relative ranking of these scores that captures the relevance of each likely match to the probe image. A set of weak SVM based rankers are learned using color and texture features [38] on small training datasets and combined to build a stronger ranker using ensemble learning. In other words, they attempt to learn a subspace where true matches are ranked highest. This method is called RankSVM [91]. A Re-ID by verification approach based on transfer learning is proposed in [92], i.e., the learning process aims at extraction of transferable discriminative information using a set of non-target people (unknown IDs). In the verification scenario, a probe's identity is verified against a small set of target people (known IDs). Re-ID is performed by learning not only the separation between target and non-target IDs, but also the separation between different targets IDs. The distance model is learned using the PRDC and the RankSVM frameworks. An iterative refinement of the ranking is proposed in [93] where gallery is modeled by a graph in the visual appearance space. The graph weights and structure are modified to accommodate the probe image and a ranking function that optimizes the graph Laplacian is computed and used for gallery ranking.

A set based discriminative ranking (SBDR) model is adopted in [94], where distance between a sequence of images of a gallery person and probe is computed using geometric distance of their approximated convex hulls. A maximum margin based ranking scheme is employed that makes distance between a true match pair smaller than the false matching pair. The metric learning process is an iterative one and hence compute intensive. The color and texture features defined in [38] are used as the underlying features. Image intensity, color, position and gradient based 7-d features are extracted and represented by Gaussian mixture models in [49]. They are combined with weighted HSV histograms and stable color regions. The sparse pairwise constraints based distance metric learning suited for high dimensional data is used for Re-ID.

Mahanolobis metric distance learning is adopted in [95] by posing correspondence detection as a two class classification problem. The learning occurs in the distance space with only two labels for each point. In other words, distances between same person's different views have the same label. By relaxing the positivity constraint on the learned matrix the above optimization problem is simplified without the need for multiple iterations. The person image is split into overlapping rectangular regions and HSV, Lab colors, and LBP [96] are extracted from each region to form the person descriptor used to learn the metric. A similar pairwise metric is learnt in [48] using the large margin nearest neighbor framework. During the learning process, the samples in the training data that are difficult to separate from the matching samples are given more priority. These samples are called impostors as they invade the perimeter of a matching pair in the distance space (Impostor learning). The features used are the same as in [95].

5. Public datasets and evaluation metrics

The visual characteristics of a person vary drastically across cameras, introducing variability in illumination, poses, view angles, scales and camera resolutions. Factors like occlusions, cluttered background and articulated bodies further add to visual variabilities. Thus, in order to develop robust Re-ID techniques it is important to acquire data that captures these factors effectively. Along with high quality data emulating real world conditions, there is also a need to compare and contrast Re-ID approaches being developed and identify improvements to techniques and the datasets. There are several available datasets that have been used to test Re-ID models. ViPER [97], i-LIDS for Re-ID [40] and ETHZ [98] are currently, most commonly used for Re-ID evaluations. Table 2 provides a summary of the widely used Re-ID datasets.

5.1. ViPER

The ViPER dataset [97] consists of images of people from two different camera views, but it has only one image of each person per camera. The dataset was collected in order to test viewpoint invariant pedestrian recognition and hence the aim was to capture as many same viewpoint pairs as possible. The view angles were roughly quantized into 45° angles and hence there are 8 same viewpoint angle pairs or 28 different viewpoint angle pairs. The dataset contains 632 pedestrian image pairs taken by two different cameras. The cameras have different viewpoints and illumination variations exist between them. The images are cropped and scaled to be 128 × 48 pixels. This is one of the most challenging datasets yet for automated person Re-ID. Fig. 7 shows some example images from this dataset.

5.2. ETHZ

ETHZ dataset consists of images of people taken by a moving camera [98] and this camera setup provides a range of variations in person

Table 2
Summary of public person Re-ID datasets.

Dataset	Multiple images	Multiple camera	Illumination variations	Pose variations	Occlusions	Scale variations
ViPER		✓	✓	✓	✓	
ETHZ	✓		✓		✓	✓
i-LIDS	✓	✓	✓	✓	✓	
CAVIAR4REID	✓	✓	✓	✓	✓	✓
i-LIDS MA and AA	✓	✓	✓	✓	✓	✓
V-47	✓	✓	✓	✓	✓	
GRID		✓	✓	✓	✓	✓

appearances. The images of pedestrians do not come from different cameras but multiple images of the person taken from a moving camera are present. The dataset has three sequences and multiple images of a person from each sequence are provided. Sequences 1, 2 and 3 have 83, 35, and 28 pedestrians respectively. The dataset consists of considerable illumination changes, scale variations and occlusions. The images are of different sizes. Fig. 8 shows some example images from this dataset.

5.3. i-LIDS for Re-ID

This dataset was extracted from the i-LIDS multi- camera tracking scenario [99] or i-LIDS MCTS dataset which is widely used for tracking evaluation purposes and was acquired in crowded public spaces. The dataset contains a total of 476 images of 119 pedestrians taken from two non-overlapping cameras. On an average there are 4 images of each pedestrian and a minimum of 2 images. The dataset has considerable illumination variations and occlusions across the two cameras. All images are normalized to 128×64 pixels. Fig. 9 shows some example images from this dataset.

5.4. CAVIAR4REID

This is extracted from another multi-camera tracking dataset [100] captured at an indoor shopping mall with two cameras with overlapping views. The dataset [37] contains multiple images of 72 pedestrians, out of which only 50 appear in both cameras, where as 22 come from the same camera. The images for each pedestrian were selected with the aim of maximizing appearance variations due to resolution changes, light conditions, occlusions, and pose changes. The minimum and maximum size of the images is 17×39 and 72×144, respectively. Fig. 10 shows some example images from this dataset.

5.5. i-LIDS MA and AA

Almost all of the above datasets contain either a single image or multiple images coming from one or in some cases two cameras. None of the datasets have a significant number of multiple images of a person coming from two separate non-overlapping cameras. In order to address this deficiency, two new datasets [41] were extracted from two non-overlapping cameras from the i-LIDS MCTS dataset. Each of the two datasets contain multiple tracked frames of a number pedestrians from two different camera views. This dataset most closely resembles a multi-camera tracking scenario and captures its characteristics better than any of the above mentioned datasets. Figs. 11 and 12 show some example images from this dataset.

- iLIDS-MA: This dataset consists of images of 40 pedestrians taken from two different cameras. 46 manually annotated images of each pedestrian are extracted from each camera. Thus, this dataset contains a total of 3680 images of slightly different sizes.
- iLIDS-AA: This dataset consists of images of 100 pedestrians taken from two different cameras. Different numbers of automatically detected and tracked images of each pedestrian are extracted from each camera. This dataset contains a total of 10,754 images of slightly different sizes. This data presents more challenges due to the possibility of errors coming from automated detection and tracking.

5.6. V-47

This dataset contains videos of 47 pedestrians captured using two cameras in an indoor setting [89]. For each camera view, two different views of each person (person walking in two different directions) are captured. The foreground masks are provided for every few frames of

Fig. 7. Example images from the VIPeR dataset.

Fig. 8. Example images from the ETHZ for Re-ID dataset, the first, second and third rows of images come from sequences 1, 2 and 3 respectively.

each video. The dataset has some illuminations variations but they are not drastic. There are few occlusions and the scene is not crowded and has very few scale variations. The dataset is important as it provides significantly large amounts of video of each pedestrian but is not sufficiently representative of typical Re-ID scenarios. Fig. 13 shows some example images from this dataset.

Fig. 9. Example images from the i-LIDS for Re-ID dataset.

Fig. 10. Example images from the CAVIAR for Re-ID dataset.

5.7. QMUL underGround Re-IDentification (GRID) dataset

This dataset was acquired by 8 cameras with non-overlapping FOVs, installed in an underground train station [23]. Thus, the images are low resolution and have significant illumination variations. The dataset has 250 pairs of images, i.e. 250 pedestrian images that appear in two different camera views and an additional 775 images of people in a single view. Even though acquisition is using 8 cameras, for a given pedestrian only 2 different views are available. Fig. 14 shows some example images from this dataset.

5.8. Additional datasets

A few other multi-camera datasets like Chokepoint [101], Terrascope [102], Person Re-ID dataset (PRID) [76], SAIVT-SoftBio [103] and CUHK02 [75] should be mentioned in this context as they can be very well applied to evaluation of person Re-ID methods. Sarc3D dataset [104] contains 200 images of 50 pedestrians taken from 4 predefined viewpoints captured with calibrated cameras to facilitate a 3D body model generation. The 3DPes dataset [105] further extends the Sarc3D dataset by including 600 videos of 200 people taken from 8 static and calibrated cameras. These datasets are geared towards evaluation of 3D human body model for tracking and identification and are applicable to Re-ID evaluation as well. Table 3 gives a summary of the Re-ID performance of some of the state-of-the-art approaches on some of the popular databases discussed above. The Re-ID performance is represented by rank 1 accuracy of Re-ID.

5.9. Limitations of datasets

The currently available Re-ID datasets are fairly reasonable in terms of encompassing multi-view variations. However, they are hardly representative of real world surveillance data. For instance, in multi-camera tracking applications, video data from large number cameras with overlapping and non-overlapping views is to be analyzed for Re-ID. The cameras and hence the data differ in resolution, frame rate and sensor characteristics. Most of the above mentioned databases are lacking in this respect. In addition, they do not provide means to analyze open set Re-ID performance or other evaluate system measures like

Fig. 11. Example images from the automatically annotated i-LIDS dataset for Re-ID.

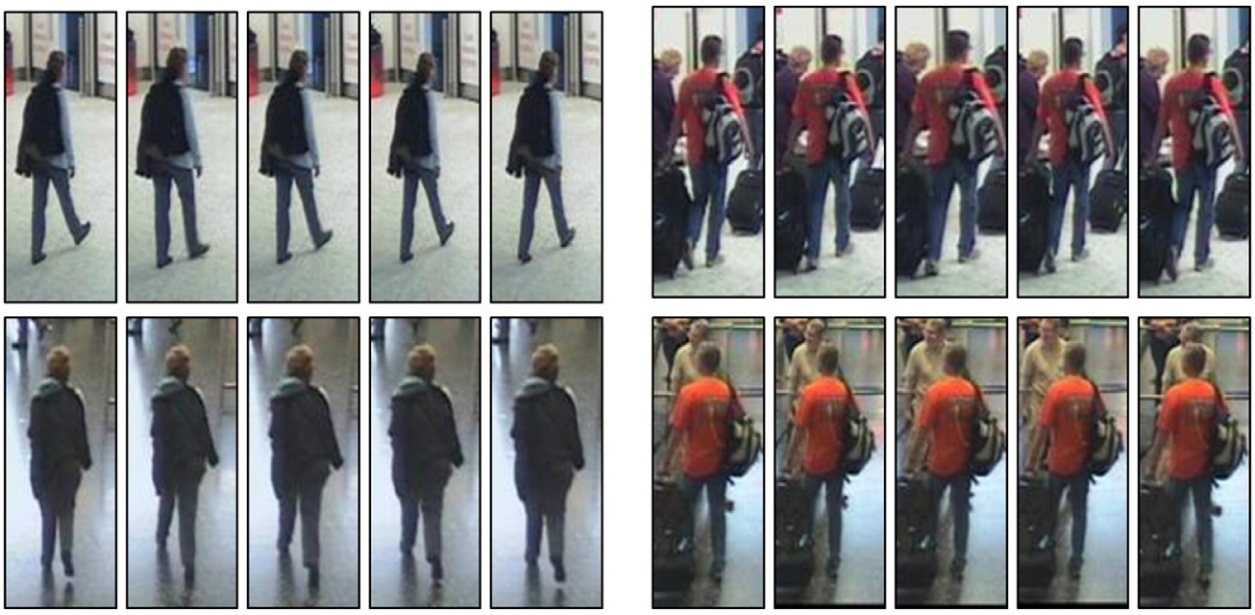

Fig. 12. Example images from the manually annotated i-LIDS dataset for Re-ID.

scalability or space time complexity. As unconstrained and long duration video data is not available, the impact of integration of temporal or sequential data into person descriptions on Re-ID cannot be judged. Availability of video data enables learning of the inter-camera relationships [23,24] that can greatly boost the Re-ID performance by pruning the incorrect (false positives) matches. With most of these datasets such experiments cannot be performed.

Evaluation of long period Re-ID requires data to be collected over several days using same or different set of cameras. None of the currently available datasets offers such instances of people collected on different days. A recent RGB-D person Re-ID dataset [106] captures depth information for each pedestrian and hence can be utilized for evaluation of depth-based features for Re-ID. Nonetheless, the data is not collected over several days and hence cannot be utilized for true long period Re-ID evaluation. Further, multi-camera tracking scenarios are by nature multiple person Re-ID problems. It implies that there exist multiple probes that have to be matched simultaneously. These datasets can be setup to test a multiple person Re-ID framework but are not truly multi probe datasets. Hence, there is a definite need for a more comprehensive and extensive Re-ID dataset.

5.10. Evaluation metrics

The most widely used evaluation methodology for person Re-ID is the performance metric known as the cumulative matching characteristic (CMC) curve. This metric is adopted since Re-ID is intuitively posed as a ranking problem, where each element in the gallery is ranked based on its comparison to the probe. The probability that the correct match in ranked equal to or less than a particular value is plotted against the size of the gallery set [97]. In order to evaluate the performance of the simultaneously matching multiple probe images of the gallery, the Synthetic Re-ID Rate (SRR) curve is derived from the CMC curve. It gives the probability that any of the given fixed number of matches is correct. The normalized area under the CMC curve (nAUC) and Rank 1 recognition rate is also important performance metrics. The nAUC is the probability that the Re-ID system will produce a true match over a

Fig. 13. Example images from the v-47 for Re-ID dataset.

Fig. 14. Example images from the GRID dataset for Re-ID.

false (incorrect) match. However, these metrics are inadequate for evaluating the open set Re-ID performance, more particularly, in evaluating the ability of the system to determine if a probe ID exists in the gallery or not (novelty detection). This point is discussed in detail in the next section.

6. Open issues in person Re-ID

As is evident, most of the work on person Re-ID leverages clothing appearance based features designed for short-period Re-ID and is evaluated in closed set Re-ID scenarios. The issue of long-period Re-ID is entirely unexplored and open set Re-ID is not completely tackled.

6.1. Long-period Re-ID

In case of large temporal separation or long-period Re-ID more stable person descriptions based on unique features like biometrics are needed. Biometrics like face and gait have been shown to have tremendous utility in person recognition/identification applications [107,108]. However, leveraging biometrics for Re-ID has its own challenges. As mentioned previously since the Re-ID data comes from uncontrolled environments with non-cooperative subjects the face of a person will not always be visible. Besides, the data is often low quality due to low sensor resolutions and low frame rates. If the face is visible, it varies greatly in pose, facial expressions, and illumination conditions. All these factors make capturing reliable facial data and subsequent face recognition very difficult. Even though the state-of-the-art face recognition techniques yield high recognition rates it is important to note these results are obtained on high resolution data captured under controlled lighting and pose settings. Automated facial recognition on low resolution images under variations in pose, age and illumination conditions is still an open problem [109,107].

Gait is a behavioral biometric that has been effective for human identification [108]. Gait is specially suited for Re-ID as it can be extracted by non-obtrusive methods and does not require co-operative subjects. Reliable gait feature extraction requires accurate silhouette extraction and sufficiently long video data. However, typical surveillance video can be low frame rate data, with people being occluded by objects or other people in the scene. Besides, the video of a given person might not be long enough to extract gait information required for identification. People are captured in different poses in different cameras and with longer video the people often tend to change their walking pose during the duration of video. Matching gait from different walking poses is an unsolved problem [110] and unless a common walking pose is present in the videos being compared, gait based Re-ID is not reliable. Another important hurdle in utilizing gait for Re-ID is that gait requires video data (multiple frames) that might not be always available. Thus, utilizing biometric information for Re-ID makes sense in theory but practical implementation is a challenging task. A combination of multiple biometrics can also be leveraged for Re-ID but fusion of multiple biometrics is an open area of research [111]. In order to boost Re-ID performance, different sensors like RGB-D [112,106] and infrared [113] that capture soft biometric cues insensitive to appearance variations are being explored.

6.2. Performance measures for open set Re-ID

The first step in open set Re-ID is to determine whether the probe ID exists in the gallery. In other words, before looking for the correct match, the system should have the ability to decide whether or not

Table 3
Summary of performance of state-of-the-art approaches on popular Re-ID datasets. The results correspond to the single-shot case, for LMNN-R, Impostor Learning, PRDC and RankSVM methods; the accuracy on i-LIDS corresponds to gallery size of 30 out of total 119 (not the complete gallery) and on VIPeR corresponds to gallery size 316 out of total 632.

Approach type	Approach	i-LIDS for Re-ID	ETHZ-1	ETHZ-2	ETHZ-3	ViPER
Passive	SDALF [15]	28%	65%	64%	76%	19.84%
	BiCov [36]	–	68%	71%	84%	20.66%
Descriptor learning	Group context [40]	23%	–	–	–	–
	ELF [38]	–	–	–	–	12%
	PLS [39]	–	79%	74%	77%	%
Metric learning	LMNN-R [45]	–	–	–	–	20%
	Impostor learning [48]	–	78%	74%	91%	22%
	PRDC [46]	42.96%	–	–	–	15.66%
	RankSVM [47]	44.05%	–	–	–	16.27%

the probe is a part of the gallery or is *unknown*. This process is known as novelty detection and it requires that the Re-ID systems have the ability of rejecting a false match. Typically in open set Re-ID, once the gallery is ranked in comparison with the probe, the probe is identified as belonging to the gallery if the similarity score is above an operating threshold. With the exception of a few methods, open set Re-ID is not entirely addressed by the current approaches.

Some approaches tackle novelty detection [59,67,61,35] by imposing an operating threshold on the similarity or distance measures between the probe and gallery IDs. Some distance metric learning methods [45,89] use the metric learning (optimization) process to determine a threshold on the distances to distinguish between true and false matches. However, to the best of our knowledge, the first attempt to formalize the open set Re-ID framework was presented in [61]. They pose the problem of Re-ID as a rectangular assignment problem, using a threshold to indicate forbidden assignments. The Re-ID is established using the Hungarian algorithm.

Closed set Re-ID adopts a ranking of the gallery with the assumption that the probe is a subset of the gallery, hence CMC curves are more suited for performance evaluation. However, the performance evaluation of open set Re-ID should be based on two measures: Re-ID accuracy (rank-1 recognition rate) and false acceptance rate (FAR). The Re-ID accuracy is expressed in terms of true positives (TPs), which represent the number of probe IDs that are correctly matched. FAR provides the performance measure when the probe is not a part of the gallery or the probe is incorrectly matched to the gallery. It is expressed in terms of mismatches (MMs) and false positives (FPs). MMs are the number of probe IDs that are incorrectly matched to gallery, when that probe ID does exist in the gallery. False positives (FPs) are the number of probes IDs that are matched to the gallery when the probe ID does not exist in the gallery. Accuracy vs FAR curves proposed in [61,35] are more suited performance evaluation of an open set Re-ID system. These metrics based on TPs, false negatives or MMs and FPs (unknown ID detection or false match rejection) are along the lines of multiple object tracking CLEAR MOT metrics [114]. Thus, in order to better understand the Re-ID models, evaluations should include accuracy vs FAR curves in addition to CMC and SRR curves.

Some other variations of Re-ID are verification and searching of people based on textual queries [115,116]. For the verification task, the system is presented by a probe that claims an ID and the system has to decide if the probe ID is the same ID that is claimed [92]. Large amounts of video data can be searched at high speeds using textual queries. Law enforcement agencies can utilize such systems for surveillance or forensic purposes. For evaluation of these Re-ID applications more specialized measures might be necessary.

6.3. Re-ID scalability

The focus of current work in Re-ID is geared towards robust descriptors and effective matching schemes but the issue of scalability is often overlooked. Scalability refers to the ability of the system to adapt itself to realistically varying factors while maintaining the performance. The following scalability issues need further research to address the specified shortcomings:

- In real world applications the gallery size is large and constantly increasing. The common similarity based ranking techniques do not scale well and hence efficient matching schemes need to be explored.
- As the gallery is ever changing, new models are added, learning based Re-ID techniques like classifiers, bag-of-words or distance metric optimization need to be recalibrated in order to incorporate the variability in the gallery set to maintain their performance.
- In order to maximize uniqueness, descriptors are often complex, high-dimensional and expensive to extract. This also makes the recognition process compute intensive and complicated. These factors affect the temporal complexity of the system making real time performance difficult to achieve.
- Large gallery sizes and high-dimensional models require large amounts of storage space and computational resources to effectively analyze the data.
- Automated video analytics can be simplified by on-camera data processing (smart cameras) and communications between cameras. However, storage and computational resource intensive Re-ID systems cannot be easily scaled to work with low power processors and narrow bandwidth transmission channels.
- All of the current approaches to Re-ID assume accurate person detection/tracking prior to feature extraction. A rigorous analysis of effects of detection/t racking errors on Re-ID performance has, to the best of our knowledge, not been performed.
- Following novelty detection, enrolling new subjects in the gallery is non-trivial. Issues like quality of model, reconciling models from several cameras and effect of Re-ID errors on gallery enrollment require further investigation.

Consideration of scalability issues within Re-ID research can lead to better designed and more efficient systems. Most Re-ID systems produce a ranked list of gallery, but this list might need to be refined by a human to boost the accuracy of the ranking. As the gallery size increases this becomes more difficult to achieve. Thus, efficient re-ranking schemes based on human input need to be addressed [117].

7. Conclusion

In this paper we have presented the problem of person re-identification, challenging issues and an overview of current research in the computer vision community. We have considered two types of Re-ID tasks: closed set Re-ID and open set Re-ID. We have categorized the methods used and discussed their characteristics and limitations. In addition, we have provided descriptions of the available Re-ID datasets and their pros and cons. A brief discussion of popular Re-ID evaluation techniques is provided along with possibilities of extensions. We have also identified some important open issues in practical Re-ID systems: scalability and computational complexity. Unaddressed issues like open set Re-ID and long period Re-ID are also explored.

Person Re-ID is a very challenging task with wide ranging application in numerous fields. It has received a lot of attention lately and the Re-ID models and recognition techniques have come a long way but are still very narrow and specific in their application to real world problems. The most obvious next step in development of unique models is the incorporation of biometric cues. Semantic information involving human visual system based perceptual attributes can provide valuable descriptive ability to the models. Building hierarchical models that incorporate relationships between low level features and high level semantics would yield more coherent descriptors. The models should be designed keeping in mind the complexity of feature extraction and their storage footprints. Hierarchical models can be used to significantly alleviate the search space within the gallery. This will greatly ease the scalability issues as well as reduce the compute intensive nature of recognition.

In summary, person Re-ID is a broad and challenging field with vast opportunities for improvements and research. This paper attempts to provide an overview of the Re-ID problem, its challenges and issues and, at the same time, present areas of future exploration.

Acknowledgments

This work was supported in part by the US Department of Justice 2009-MU-MU-K004. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of our sponsors.

References

[1] P. Tu, G. Doretto, N. Krahnstoever, A.G.A. Perera, F. Wheeler, X. Liu, J. Rittscher, T. Sebastian, T. Yu, K. Harding, An intelligent video framework for homeland protection, Proceedings of SPIE Defence and Security Symposium—Unattended Ground, Sea, and Air Sensor Technologies and Applications IX, 2007.

[2] A. Hampapur, L. Brown, J. Connell, S. Pankanti, A. Senior, Y. Tian, Smart surveillance: applications, technologies and implications, In IEEE Pacific-Rim Conference On Multimedia, vol. 2, 2003, pp. 1133–1138.

[3] J. Sivic, C.L. Zitnick, R. Szeliski, Finding people in repeated shots of the same scene, Proceedings of the British Machine Vision Conference, 2006, pp. 909–918.

[4] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, IEEE Conference on Computer Vision and, Pattern Recognition, 2005, pp. 886–893.

[5] P.F. Felzenszwalb, D.P. Huttenlocher, Pictorial structures for object recognition, Int. J. Comput. Vis. 61 (2003) 55–79.

[6] P.F. Felzenszwalb, D.A. McAllester, D. Ramanan, A discriminatively trained, multiscale, deformable part model, IEEE Computer Society Conference on Computer Vision and, Pattern Recognition, 2008, pp. 1–8.

[7] L. Bourdev, J. Malik, Poselets: body part detectors trained using 3D human pose annotations, International Conference on Computer Vision, 2009, pp. 1365–1372.

[8] M. Andriluka, S. Roth, B. Schiele, People-tracking-by-detection and people-detection-by-tracking, IEEE Conference on Computer Vision and, Pattern Recognition, 2008, pp. 1–8.

[9] K. Okuma, A. Taleghani, N.D. Freitas, J.J. Little, D.G. Lowe, A boosted particle filter: multitarget detection and tracking, European Conference on Computer Vision, 2004, pp. 28–39.

[10] S. Pellegrini, A. Ess, K. Schindler, L. van Gool, You'll never walk alone: modeling social behavior for multi-target tracking, IEEE International Conference on Computer Vision, 2009, pp. 261–268.

[11] B. Wu, R. Nevatia, Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors, Int. J. Comput. Vis. 75 (2007) 247–266.

[12] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, ACM Comput. Surv. 38 (4) (2006) 13.

[13] Y. Cai, M. Pietikäinen, Person re-identification based on global color context, The Tenth International Workshop on Visual Surveillance (in conjunction with ACCV 2010), 2010, pp. 205–215.

[14] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, V. Murino, Multiple-shot person re-identification by hpe signature, International Conference on, Pattern Recognition, 2010, pp. 1413–1416.

[15] L. Bazzani, M. Cristani, V. Murino, Symmetry-driven accumulation of local features for human characterization and re-identification, Comput. Vis. Image Underst. 117 (2) (2013) 130–144.

[16] N. Gheissari, T. Sebastian, R. Hartley, Person reidentification using spatiotemporal appearance, IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2006, pp. 1528–1535.

[17] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, Shape and appearance context modeling, International Conference on Computer Vision, 2007, pp. 1–8.

[18] R. Satta, Appearance descriptors for person re-identification: a comprehensive review, CoRR.URL http://arxiv.org/abs/1307.5748 .

[19] O. Javed, K. Shafique, Z. Rasheed, M. Shah, Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views, Comput. Vis. Image Underst. 109 (2) (2008) 146–162.

[20] D. Makris, T. Ellis, J. Black, Bridging the gaps between cameras, IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 205–210.

[21] A. Rahimi, B. Dunagan, T. Darrell, Simultaneous calibration and tracking with a network of non-overlapping sensors, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2004, pp. 187–194.

[22] R. Mazzon, S.F. Tahir, A. Cavallaro, Person re-identification in crowd, Pattern Recogn. Lett. 33 (14) (2012) 1828–1837.

[23] C. Loy, T. Xiang, S. Gong, Multi-camera activity correlation analysis, IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1988–1995.

[24] C. Loy, T. Xiang, S. Gong, Time-delayed correlation analysis for multi-camera activity understanding, Int. J. Comput. Vis. 90 (1) (2010) 106–129.

[25] H. Hotelling, Relations between two sets of variates, in: S. Kotz, N.L. Johnson (Eds.), Breakthroughs in Statistics, Springer Series in Statistics, Springer, New York, 1992, pp. 162–190.

[26] C. Loy, T. Xiang, S. Gong, Incremental activity modeling in multiple disjoint cameras, IEEE Trans. Pattern. Anal. Mach. Intell. 34 (9) (2012) 1799–1813.

[27] X. Wang, Intelligent multi-camera video surveillance: a review, Pattern Recogn. Lett. 34 (1) (2013) 3–19.

[28] H. Detmold, A. van den Hengel, A. Dick, A. Cichowski, R. Hill, E. Kocadag, K. Falkner, D. Munro, Topology estimation for thousand-camera surveillance networks, First ACM/IEEE International Conference on Distributed Smart Cameras, 2007, pp. 195–202.

[29] M. Lantagne, M. Parizeau, R. Bergevin, Vip: vision tool for comparing images of people, vision interface, 2003. 35–42.

[30] U. Park, A. Jain, I. Kitahara, K. Kogure, N. Hagita, Vise: visual search engine using multiple networked cameras, International Conference on Pattern Recognition, vol. 3, 2006, pp. 1204–1207.

[31] T. Gandhi, M.M. Trivedi, Person tracking and reidentification: introducing panoramic appearance map (pam) for feature representation, Mach. Vis. Appl. 18 (2007) 207–220.

[32] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou, S. Maybank, Principal axis-based correspondence between multiple cameras for people tracking, IEEE Trans. Pattern. Anal. Mach. Intell. 28 (2006) 663–671.

[33] D. Baltieri, R. Vezzani, R. Cucchiara, C. Benedek Utasi, T. Szirányi, Multi-view people surveillance using 3D information, The Eleventh International Workshop on Visual Surveillance (in conjunction with ICCV 2011), 2011, pp. 1817–1824.

[34] S. Bak, E. Corvee, F. Bremond, M. Thonnat, Person re-identification using spatial covariance regions of human body parts, IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010, pp. 435–440.

[35] A. Bedagkar-Gala, S.K. Shah, Part-based spatio-temporal model for multi-person re-identification, Pattern Recogn. Lett. 33 (14) (2012) 1908–1915.

[36] B. Ma, Y. Su, F. Jurie, Bicov: a novel image representation for person re-identification and face verification, Proceedings of the British Machive Vision Conference, 2012, pp. 57.1–57.11.

[37] D. Cheng, M. Cristani, M. Stoppa, L. Bazzani, V. Murinog, Custom pictorial structures for re-identification, Proceedings of the British Machine Vision Conference, 2011, pp. 68.1–68.11.

[38] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, European Conference on Computer Vision, 2008, pp. 262–275.

[39] W.R. Schwartz, L.S. Davis, Learning discriminative appearance-based models using partial least squares, Proceedings of the XXII Brazilian Symposium on, Computer Graphics and Image Processing, 2009, pp. 322–329.

[40] W.-S. Zheng, S. Gong, T. Xiang, Associating groups of people, Proceedings of the British Machive Vision Conference, 2009, pp. 23.1–23.11.

[41] S. Bak, E. CorvéE, F. BréMond, M. Thonnat, Boosted human re-identification using Riemannian manifolds, Image Vision Comput. 30 (6–7) (2012) 443–452.

[42] R. Layne, T.M. Hospedales, S. Gong, Towards person identification and re-identification with attributes, Proceedings of the 12th European conference on Computer Vision, ECCV Workshops, 2012, pp. 402–412.

[43] S. Bak, G. Charpiat, E. Corvée, F. Brémond, M. Thonnat, Learning to match appearances by correlations in a covariance metric space, Proceedings of the European conference on Computer Vision, 2012, pp. 806–820.

[44] R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for person re-identification, IEEE Conference on Computer Vision and, Pattern Recognition, 2013, pp. 3586–3593.

[45] M. Dikmen, E. Akbas, T.S. Huang, N. Ahuja, Pedestrian recognition with a learned metric, Asian Conference in Computer Vision, 2010, pp. 501–512.

[46] W.-S. Zheng, S. Gong, T. Xiang, Reidentification by relative distance comparison, IEEE Trans. Pattern. Anal. Mach. Intell. 35 (3) (2013) 653–668.

[47] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by support vector ranking, Proceedings of the British Machine Vision Conference, 2010, pp. 21.1–21.11.

[48] M. Hirzer, P. Roth, H. Bischof, Person re-identification by efficient impostor-based metric learning, IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, 2012, pp. 203–208.

[49] B. Ma, Y. Su, F. Jurie, Local descriptors encoded by fisher vectors for person re-identification, Proceedings of the 12th European conference on Computer Vision, ECCV Workshops, 2012, pp. 413–422.

[50] J. Kang, I. Cohen, G. Medioni, Object reacquisition using invariant appearance model, Proceedings of International Conference on Pattern Recognition, vol. 4, 2004, pp. 759–762.

[51] P.-E. Forssén, Maximally stable colour regions for recognition and matching, IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[52] L. Bazzani, M. Cristani, A. Perina, V. Murino, Multiple-shot person re-identification by chromatic and epitomic analyses, Pattern Recogn. Lett. 33 (7) (2012) 898–903.

[53] M. Andriluka, S. Roth, B. Schiele, Pictorial structures revisited: people detection and articulated pose estimation, IEEE Conference on Computer Vision and, Pattern Recognition, 2009, pp. 1014–1021.

[54] R. Satta, G. Fumera, F. Roli, M. Cristani, V. Murino, A multiple component matching framework for person re-identification, Proceedings of the 16th international conference on Image analysis and processing, 2011, pp. 140–149.

[55] R. Satta, G. Fumera, F. Roli, Fast person re-identification based on dissimilarity representations, Pattern Recogn. Lett. 33 (14) (2012) 1838–1848.

[56] O. Hamdoun, F. Moutarde, B. Stanciulescu, B. Steux, Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences, 2nd ACM/IEEE International Conference on Distributed Smart Cameras, 2008, pp. 1–6.

[57] K. Jungling, C. Bodensteiner, M. Arens, Person re-identification in multi camera networks, IEEE Conference on Computer Vision and, Pattern Recognition, 2011, pp. 55–61.

[58] B. Leibe, A. Leonardis, B. Schiele, Combined object categorization and segmentation with an implicit shape model, European Conference on Computer Vision, 2004, pp. 17–32.

[59] M. Bauml, R. Stiefelhagen, Evaluation of local features for person re-identification in image sequences, International Conference on Advanced Video and Signal-Based Surveillance, 2011, pp. 291–296.

[60] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern. Anal. Mach. Intell. 27 (10) (2005) 1615–1630.

[61] A. Bedagkar-Gala, S. Shah, Multiple person re-identification using part based spatio-temporal color appearance model, The Eleventh International Workshop on Visual Surveillance (in conjunction with ICCV 2011), 2011, pp. 1721–1728.

[62] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern. Anal. Mach. Intell. 23 (6) (1998) 681–685.

[63] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern. Anal. Mach. Intell. 19 (7) (1997) 711–720 (special Issue on Face Recognition.).

[64] A. Gallagher, T. Chen, Clothing cosegmentation for recognizing people, IEEE Conference on Computer Vision and, Pattern Recognition, 2008, pp. 1–8.

[65] D.-N.T. Cong, C. Achard, L. Khoudour, People re-identification by classification of silhouettes based on sparse representation, 2nd International Conference onImage Processing Theory Tools and Applications, 2010, pp. 60–65.
[66] A. Yang, J. Wright, Y. Ma, S. Sastry, Feature Selection in Face Recognition: A Sparse Representation Perspective, Tech. Rep, University of Illinois, 2007.
[67] L. Brun, D. Conte, P. Foggia, M. Vento, People re-identification by graph kernels methods, Proceedings of the 8th International Conference on Graph-based Representations In, Pattern Recognition, 2011, pp. 285–294.
[68] Y. Zhang, S. Li, Gabor-lbp based region covariance descriptor for person re-identification, Sixth International Conference on Image and Graphics, 2011, pp. 368–371.
[69] A. D'angelo, J.-L. Dugelay, People re-identification in camera networks based on probabilistic color histograms, Electronic Imaging Conference on 3D Image Processing and Applications, vol. 7882, 2011, (78820K–78820K–12).
[70] D.N.T. Cong, L. Khoudour, C. Achard, C. Meurie, O. Lezoray, People re-identification by spectral classification of silhouettes, Signal Process. 90 (8) (2010) 2362–2374.
[71] O. Javed, K. Shafique, M. Shah, Appearance modeling for tracking in multiple non-overlapping cameras, IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2005, pp. 26–33.
[72] F. Porikli, Inter-camera color calibration by correlation model function, International Conference on Image Processing, vol. 2, 2003, (II–133–6).
[73] B. Prosser, S. Gong, T. Xiang, Multi-camera matching using bi-directional cumulative brightness transfer functions, Proceedings of the British Machive Vision Conference, 2008, pp. 64.1–64.10.
[74] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2004) 91–110.
[75] W. Li, X. Wang, Locally aligned feature transforms across views, IEEE Conference on Computer Vision and, Pattern Recognition, 2013, pp. 3594–3601.
[76] M. Hirzer, C. Beleznai, P.M. Roth, H. Bischof, Person re-identification by descriptive and discriminative classification, Proceedings of the 17th Scandinavian conference on Image, analysis, 2011, pp. 91–102.
[77] S. Bak, E. Corvee, F. Bremond, M. Thonnat, Person re-identification using haar-based and dcd-based signature, IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010, pp. 1–8.
[78] T. Avraham, I. Gurvich, M. Lindenbaum, S. Markovitch, Learning implicit transfer for person re-identification, Proceedings of the 12th European Conference on Computer Vision, ECCV Workshops, 2012, pp. 381–390.
[79] Y. Brand, T. Avraham, M. Lindenbaum, Transitive re-identification, British Machine Vision Conference, 2013.
[80] C. Liu, S. Gong, C.C. Loy, X. Lin, Person re-identification: what features are important? Proceedings of the 12th European conference on Computer Vision, ECCV Workshops, 2012, pp. 391–401.
[81] D. Parikh, K. Grauman, Relative attributes, IEEE International Conference on Computer Vision, 2011, pp. 503–510.
[82] R. Zhao, W. Ouyang, X. Wang, Person re-identification by salience matching, IEEE International Conference on Computer Vision, 2013.
[83] N. Kumar, A.C. Berg, P.N. Belhumeur, S.K. Nayar, Attribute and simile classifiers for face verification, IEEE International Conference on Computer Vision, 2009, pp. 365–372.
[84] A. Farhadi, I. Endres, D. Hoiem, D. Forsyth, Describing objects by their attributes, IEEE Conference on Computer Vision and, Pattern Recognition, 2009, pp. 1778–1785.
[85] J. Wang, K. Markert, M. Everingham, Learning models for object recognition from natural language descriptions, Proceedings of British Machine Vision Conference, 2009, pp. 1–11.
[86] Y. Wang, G. Mori, A discriminative latent model of object classes and attributes, Proceedings of the 11th European Conference on Computer Vision, 2010, pp. 155–168.
[87] R. Layne, T.M. Hospedales, S. Gong, Re-identification by attributes, Proceedings of the British Machine Vision Conference, 2012, pp. 24.1–24.11.
[88] L. Yang, R. Jin, Distance Metric Learning: A Comprehensive Survey, Tech. Rep, Michigan State University, 2006.
[89] S. Wang, M. Lewandowski, J. Annesley, J. Orwell, Re-identification of pedestrians with variable occlusion and scale, The Eleventh International Workshop on Visual Surveillance (In Conjunction with ICCV 2011), 2011, pp. 1876–1882.
[90] D. Simonnet, M. Lewandowski, S.A. Velastin, J. Orwell, E. Turkbeyler, Re-identification of pedestrians in crowds using dynamic time warping, Proceedings of the 12th European conference on Computer Vision, ECCV Workshops, 2012, pp. 423–432.
[91] T. Joachims, Optimizing search engines using clickthrough data, Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002, pp. 133–142.
[92] W.-S. Zheng, S. Gong, T. Xiang, Transfer re-identification: from person to set-based verification, IEEE Conference on Computer Vision and, Pattern Recognition, 2012, pp. 2650–2657.
[93] C.C. Loy, C. Liu, S. Gong, Person re-identification by manifold ranking, IEEE International Conference on Image Processing, 2013.
[94] Y. Wu, M. Minoh, M. Mukunoki, S. Lao, Set based discriminative ranking for recognition, Proceedings of the 12th European conference on Computer Vision, ECCV Workshops, ECCV Workshops, 2012, pp. 497–510.
[95] M. Hirzer, P.M. Roth, M. Köstinger, H. Bischof, Relaxed pairwise learned metric for person re-identification, Proceedings of the 12th European conference on Computer Vision, ECCV Workshops, 2012, pp. 780–793.
[96] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern. Anal. Mach. Intell. 24 (7) (2002) 971–987.
[97] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2007.
[98] A. Ess, B. Leibe, L.V. Gool, Depth and appearance for mobile scene analysis, International Conference on Computer Vision, 2007, pp. 1–8.
[99] U.H. Office, i-Lids Multiple Camera Tracking Scenario Definition, 2007.
[100] ([link].URL) http://homepages.inf.ed.ac.uk/rbf/CAVIAR/ 2004.
[101] Y. Wong, S. Chen, S. Mau, C. Sanderson, B.C. Lovell, Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition, Computer Vision and Pattern Recognition Workshops, 2011.
[102] C. Jaynes, A. Kale, N. Sanders, E. Grossmann, The terrascope dataset: scripted multi-camera indoor video surveillance with ground-truth, Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005, pp. 309–316.
[103] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, P. Lucey, A database for person re-identification in multi-camera surveillance networks, International Conference on Digital Image Computing Techniques and Applications, 2012, pp. 1–8.
[104] D. Baltieri, R. Vezzani, R. Cucchiara, Sarc3d: a new 3D body model for people tracking and re-identification, International Conference on Image Analysis and Processing, 2011, pp. 197–206.
[105] D. Baltieri, R. Vezzani, R. Cucchiara, 3dpes: 3D people dataset for surveillance and forensics, Proceedings of the joint ACM workshop on Human gesture and behavior understanding, 2011, pp. 59–64.
[106] I. Barbosa, M. Cristani, A.D. Bue, L. Bazzani, V. Murino, Re-identification with rgb-d sensors, Proceedings of the 12th European Conference on Computer Vision, ECCV Workshops, 2012, pp. 433–442.
[107] W. Zhao, R. Chellappa, J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.
[108] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern. Anal. Mach. Intell. 28 (2006) 316–322.
[109] M. Ao, D. Yi, Z. Lei, S. Li, Face recognition at a distance: system issues, in: M. Tistarelli, S. Li, R. Chellappa (Eds.), Handbook of Remote Biometrics, Springer, London, 2009, pp. 155–167, (Ch. 6,).
[110] K. Bashir, T. Xiang, S. Gong, Cross view gait recognition using correlation strength, Procedings of the British Machine Vision Conference, 2010, pp. 109.1–109.11.
[111] K. Nandakumar, A.K. Jain, A. Ross, Fusion in multibiometric identification systems: what about the missing data? Proceedings of the Third International Conference on Advances in, Biometrics, 2009, pp. 743–752.
[112] A. Albiol, A. Albiol, J. Oliver, J. Mossi, Who is who at different cameras: people re-identification using depth cameras, IET Comput. Vis. 6 (5) (2012) 378–387.
[113] K. Ju¨ngling, M. Arens, Local feature based person reidentification in infrared image sequences, Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010, pp. 448–455.
[114] K. Bernardin, R. Stiefelhagen, Evaluating multiple object tracking performance: the clear MOT metrics, J. IViP (2008) 1:1–1:10.
[115] R. Satta, G. Fumera, F. Roli, A general method for appearance-based people search based on textual queries, Proceedings of the 12th European Conference on Computer Vision, ECCV Workshops, 2012, pp. 453–461.
[116] D. Grangier, S. Bengio, A discriminative kernel-based approach to rank images from text queries, IEEE Trans. Pattern. Anal. Mach. Intell. 30 (8) (2008) 1371–1384.
[117] C. Liu, C.C. Loy, S. Gong, G. Wang, Pop: person re-identification post-rank optimisation, IEEE International Conference on Computer Vision, 2013.