Gait-based Person Re-identification: A Survey

ATHIRA NAMBIAR, ALEXANDRE BERNARDINO, and JACINTO C. NASCIMENTO,
Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal

The way people walk is a strong correlate of their identity. Several studies have shown that both humans and machines can recognize individuals just by their gait, given that proper measurements of the observed motion patterns are available. For surveillance applications, gait is also attractive, because it does not require active collaboration from users and is hard to fake. However, the acquisition of good-quality measures of a person's motion patterns in unconstrained environments, (e.g., in person re-identification applications) has proved very challenging in practice. Existing technology (video cameras) suffer from changes in viewpoint, daylight, clothing, accessories, and other variations in the person's appearance. Novel three-dimensional sensors are bringing new promises to the field, but still many research issues are open. This article presents a survey of the work done in gait analysis for re-identification in the past decade, looking at the main approaches, datasets, and evaluation methodologies. We identify several relevant dimensions of the problem and provide a taxonomic analysis of the current state of the art. Finally, we discuss the levels of performance achievable with the current technology and give a perspective of the most challenging and promising directions of research for the future.

CCS Concepts: • **Security and privacy → Biometrics**; • **Computing methodologies → Computer vision**; *Computer vision problems*; *Machine learning*; • **Applied computing → Surveillance mechanisms**;

Additional Key Words and Phrases: Video surveillance, computer vision, gait analysis, biometrics, person re-identification, machine learning

ACM Reference format:
Athira Nambiar, Alexandre Bernardino, and Jacinto C. Nascimento. 2019. Gait-based Person Re-identification: A Survey. *ACM Comput. Surv.* 52, 2, Article 33 (April 2019), 34 pages.
https://doi.org/10.1145/3243043

1 INTRODUCTION

With the increase in security and forensics concerns, as well as improved access to multimedia technology, surveillance camera networks are proliferating in both public and private areas, including airports, railway stations, university campuses, shopping complexes, housing apartments, supermarkets, and workplaces. Only in the United Kingdom, there are between 4 million and

This work was partially supported by the FCT projects [UID/EEA/50009/2013], AHA CMUP-ERI/HCI/0046/2013, SPARSIS [PTDC/EEIPRO/0426/2014] and FCT doctoral grant [SFRH/BD/97258/2013].

Authors' addresses: A. Nambiar, A. Bernardino, and J. C. Nascimento, Institute for Systems and Robotics, Instituto Superior Técnico, Av. Rovisco Pais, 1, 1049-001, Lisbon, Portugal; emails: anambiar@isr.tecnico.ulisboa.pt, alexandre.bernardino@tecnico.ulisboa.pt, jan@isr.tecnico.ulisboa.pt.
Author's current address: A. Nambiar, Post-doctoral scholar, Vision lab, Dept. of Computer Science & Engineering, Indian Institute of Technology, Madras, Chennai, India; email: anambiar@cse.iitm.ac.in.

0360-0300/2019/04-ART33 $15.00
https://doi.org/10.1145/3243043

5.9 million CCTV cameras, according to the British Security Industry Association (BSIA); one for every 11 people (Barrett 2013). Each Londoner is caught on camera on average 300 times each day,[1] which reveals the real influence of surveillance systems on our daily lives (Wiegler 2008). In addition to providing video footage, surveillance cameras also act as a visible deterrent to criminals. Usually, they cover vast areas with non-overlapping fields of view.

The automatic analysis of data collected in surveillance camera networks enables us to preempt suspicious events and to provide real-time alarms and situational awareness to the security personnel. The security paradigm can shift from reaction/investigation of incidents to a more pro-active prevention of potentially catastrophic events (Hampapur et al. 2003). These kinds of proactive steps not only serve toward public safety but also act as primary evidence in identifying the criminals, as in the 7/7 London Bombings (2005) or in the Boston Marathon bombing terrorist attack (2013). The advances in computer vision, as well as machine-learning techniques in the recent years, have ameliorated this expedition toward smart surveillance at a fast pace and as a result, a plethora of algorithms for the automatic analysis of the video sequences have been proposed.

Among them, person re-identification (Re-ID) is one of the very interesting yet challenging problem. One of the earliest definitions of person re-identification owes to metaphysics (Plantinga 1961), where Alvin Plantinga provided the definition to Re-ID in 1961 while discussing the relationship between mental states and behavior, as "*To re-identify a particular, then, is to identify it as (numerically) the same particular as one encountered on a previous occasion.*" Afterward, many works have been encountered in various fields such as psychology, logic, computer vision, and so on Zheng et al. (2016). From vision and surveillance point of view, person Re-ID is a hot topic with a high research and application significance, where the system has to *re-identify* persons in camera networks, under unconstrained conditions.

From the multitude of personal traits that characterize an individual, one of the most interesting for re-identification is human gait. It includes both the body posture and dynamics while walking (Lee 2002). Human gait has been mentioned in many famous early works, i.e., Aristotle (384–322 BC) in his book "*De Motu Animalium*" on the movement of animals, and Leonardo Da Vinci (1452–1519) in his anatomic paintings. In cognitive science, gait is considered as one of the cues that humans exploit to recognize people (Stevenage et al. 1999).

Human gait is very promising research for surveillance applications, because it does not require active collaboration from users and is hard to fake. In addition, gait is unobtrusive as well as perceivable from a distance. A rich literature on gait analysis has been produced in the past decade, and a new trend is to exploit gait to *re-identify* people. This is quite young, novel, and promising field with a wide spectrum yet to explore. Hence, in this survey, we review the main challenges and approaches taken in the past few years on the *Gait-based person re-identification* problem. In particular, we present a systematic review of the human gait applications reported in the person Re-ID scenario, categorizing the main approaches in a taxonomy of relevant dimensions and presenting the datasets and evaluation methodologies in the state of the art.

We organize the article as follows. In Section 1, the terminologies and the basic definitions of the gait-based Re-ID problem are analyzed. In Section 2, a multidimensional overview of the use of human gait for person Re-ID is presented. The research conducted toward addressing each of the dimension is presented in length, and the pros and cons of various methods are critically analysed. In Section 3, available datasets toward gait-based Re-ID application are detailed. The performance evaluation strategy and the metrics are discussed in Section 4. Finally, in the concluding Section 5, the open issues and promising directions for future research are discussed.

[1] http://www.ibtimes.co.uk/britain-cctv-camera-surveillance-watch-london-big-312382.

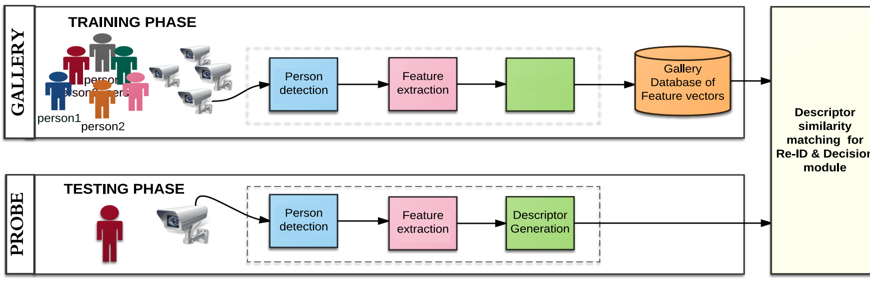


Fig. 1. A classical person re-identification (Re-ID) diagram.

1.1 Re-identification

Person Re-ID is a crucial tool for intelligent surveillance and security systems, being the process of establishing correspondences between images of a particular person taken both at different locations and time instances (Gala and Shah 2014b; Roy et al. 2012). This process of establishing connections and thus extending the tracking beyond *"blind gaps"*[2] is known as Re-ID. In Re-ID, the detection of a subject is combined with a unique label, so that the same person at different instances can be re-identified.

The European Commission has provided certain definitions to surveillance terms *viz.*, *"detect, classify, identify, recognize, and verify"* in EUROSUR-2011 (Frontex 2011). In line with those definitions, Re-ID is found to be lying in between identification and recognition (Vezzani et al. 2013).[3] In the identification context, the goal of Re-ID is to group the observations incoming from the surveillance network according to subject, like in a unsupervised learning algorithm. There is no prior knowledge on the identity of the users and the system automatically creates unique labels for the groups of observations that likely belong to the same subject. This has important applications in tracking long-term trajectories in wide area surveillance networks, particularly when trajectory discontinuities exist (Vezzani et al. 2013). In the recognition case, there is prior knowledge about the identity of the users (see recent available survey Connor and Ross (2018)). Typically, this is acquired in an enrollment phase where the characteristics of the subjects are acquired and stored in a gallery set. For recognition, a query is formulated about the target person and all the possible instances matching the target are retrieved from the gallery. The result of such a query is a set of ranked items, with the hypothesis that one and only one element of the gallery will correspond to the query (Vezzani et al. 2013). In contrast to the identification scenario, recognition demands the probe to be within the gallery (closed-set identification[4]).

The most significant fact that distinguishes Re-ID from classical identification and recognition methods is the much more relaxed set of conditions where it must operate. In typical surveillance scenarios, the subjects are not aware of the surveillance task and behave in an arbitrary fashion. Therefore, the range of conditions that must be taken into account is much less constrained. The set of techniques employed for Re-ID must be more robust to pose, scale, and lighting changes, as well as observation direction, terrain, clutter, and clothing accessories.

Figure 1 shows a diagram of a typical person Re-ID system that includes person detection, feature extraction, and descriptor matching. In the training phase, the video sequences for all

---

[2]Blind gaps correspond to the time lapses within which the subject does not appear in the fields of view of any adjacent cameras (Doretto et al. 2011).
[3]EUROSUR-2011 defines *"identification as the process to establish the unique identity of the object (name, number), as a rule without prior knowledge, whereas recognition is defined as the process to establish that a detected object is a specific pre-defined unique object"* (Frontex 2011).
[4]Closed-set identification is where every input image has a corresponding match in the database.

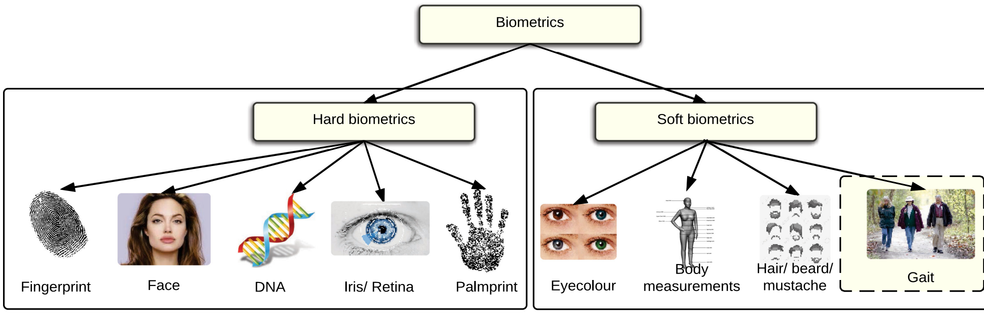


Fig. 2. Overview of biometrics classified according to their physiological characteristics (hard biometrics) and physical, behavioural, or adhered characteristics (soft biometrics). Human gait is highlighted as an instance of soft biometric.

individuals appearing in the surveillance scenario are acquired via one or more cameras in the network using, for instance, pedestrian detection algorithms (Dollár et al. 2009). Afterward, feature extraction is carried out and robust feature[5] descriptors are generated and stored in a gallery database, to be used in the Re-ID stage. Whenever a test person (probe) enters into the system, his/her feature vector is generated in the same way as for the gallery feature vectors. Then, the probe feature descriptor will be compared against the gallery database of feature descriptors by some similarity matching or classification technique. At this stage, the Re-ID decision is made and the re-identified person ID is retrieved.

Most of the traditional Re-ID approaches are based on the overall human appearance in the multimedia content, *viz. Appearance-based Re-ID*. They leverage visual features concerning not only the appearance (colour and texture) but also the objects that the subject may carry. The visual descriptors include either color/texture features or local features such as key points and edges. Rich and vast literature surveys have been conducted on these approaches in Doretto et al. (2011), Riccio et al. (2014), and Bialkowski et al. (2012). A common problem with such techniques the assumption of *colour constancy*,[6] which is not easy to achieve in practice (Maloney and Wandell 1986). Another limitation of appearance-based techniques is their short-term time span, during which, the appearance described by the clothing and other attributes are considered to be constant known as "*Appearance constancy hypothesis*." However, if Re-ID is to be performed for many days/ weeks, then the techniques above will be quite ineffective, since the holistic appearance will undergo drastic variations. For such long-term scenarios, methods based on biometric traits are found to be more suitable to be applied.

1.2 Human Gait as a Soft Biometric

Biometrics is defined as "the science of establishing the identity of an individual, based on his/her inherent physical and behavioural traits" (Ross and Jain 2007). The term *biometrics* is coined from two Greek words: *bios*, meaning "life," and *metrics*, meaning "to measure." A biometric-based surveillance system identifies or validates the person by extracting the characteristic features of the people and comparing them with the registered gallery samples (see Figure 1). Figure 2 shows various biometrics commonly used in applications. The most acclaimed and popular biometrics are fingerprint, iris, face, palm print, and voice, used in access control systems. These biometrics are

[5]Features are the values derived from the original data, intended to be informative and non-redundant with respect to the person's identity, facilitating the subsequent learning and generalization steps.
[6]The ability to assign the same colour to the same object under different lighting conditions.

invariant in time, thus termed *hard-biometrics*, but demand the necessity for well-controlled environments and detailed computational processing, which is difficult to attain in real-world surveillance conditions. Instead, in typical video surveillance scenarios, people move freely in ways that may prevent the acquisition of hard biometrics. Another genre of biometrics *viz.*, soft biometrics sounds more promising in these scenarios.

Soft biometrics are defined by Dantcheva et al. (2010) as "*the physical, behavioral or adhered human characteristics, classifiable in predefined human compliant categories that are established by humans with the aim of differentiating individuals.*" They encode characteristic human traits such as anthropometric measurements, height, body size, and gait, which are coherent and reliable for long-term Re-ID (Nixon et al. 2015). Comparing to hard biometrics, in the soft biometric context the individuals are not so distinguishable, lacking strong indicators of their identity. Nevertheless, there are some advantages to Re-ID, namely the ability to gather cues from a distance, without disturbing the user or requiring his/her cooperation.

Among the soft biometric cues, gait is very relevant to Re-ID in surveillance networks. Gait is the most prevalent human movement in typical surveillance spaces. It is unique for each human and hard to fake. Several studies in neuroscience and psychology also highlight the importance of gait in human perception of the identity of others. For instance, in medical situations like Prosopagnosia (face blindness), the victims use secondary cues such as gait and body appearances for person identification[7] (Kress and Daum 2003). Besides, observation of gait is believed to be an important aspect of diagnosis for several musculo-skeletal and neurological conditions, such as cerebral palsy, multiple sclerosis, parkinsonism, and stroke (Whittle 1996).

In a famous study of biological motion (Johansson 1973), using Moving Light Displays (MLDs), they instrumented the main joints of a human with bright light spots. Then, just from the observation of the motion patterns of 10–12 points, subjects reported a vivid impression of human locomotion. That work postulated that observers were able to recognize human activity (walking, running, etc.) using MLDs in less than one-10th of a second and were able to make judgments on the gender and identity checking whether the gait pattern is familiar. Later, follow-up studies were conducted in the paradigm by altering data acquisition conditions such as blurring the dots and relocating the position of dots (Blake and Shiffrar 2007), which further confirmed that, even under indistinct conditions, motion perception is remarkably robust. In one of the famous studies (Sumi 1984), a hallmark attribute associated with human motion perception was proposed that it is vulnerable to inversion. In that study, it was observed that, with space-time reverted MLD patterns (i.e. inverted patterns played backwards), subjects did not perceive natural biological motions. For instance, the human arms were interpreted as legs and vice versa. All these studies strongly suggest that motion signals constitute valuable information from which the human brain can reliably perform detection and identification of persons, supporting the discriminative and unique nature of human gait. This has led to a large body of work being developed in the past few years toward recognition (DeCann et al. 2014, 2013; DeCann and Ross 2010) and identification of humans using gait (Gafurov 2007; Nixon et al. 2010; Makihara et al. 2015). This also accentuates the significance of gait pattern as a potential biometric tool in the surveillance application realms. The key advantages and challenges of the use of gait in surveillance applications are presented in Table 1.

Despite the past work on gait analysis, the application of gait to re-identification only spawned about ~8 years ago. There are fundamental differences between the gait-based recognition and Re-ID problems, which lie in the structure of the domains of application. In recognition, usually the operator has the control over most of the acquisition conditions, such as camera viewpoint (often single camera), background, subject pose, illumination, the number of persons in the

[7] https://en.wikipedia.org/wiki/Prosopagnosia.

Table 1. Pros and Cons of Gait as a Soft Biometric

Advantages	Disadvantages
• unobtrusive • cooperation of the user not necessary • measured at far distance • unique for each individual • cannot be easily concealed • hard to fake	• varying with illness, aging and emotional states • varying with walking surface, shoe, cloth types, carrying objects and clutter in the scenario

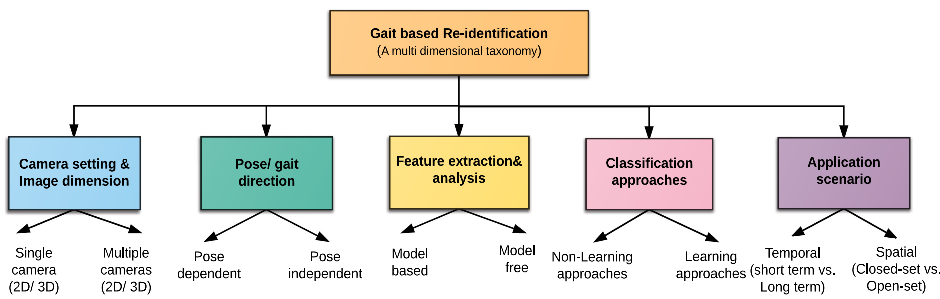


Fig. 3. A multidimensional overview of the gait-based Re-identification algorithms.

acquisition, chance of occlusion, to mention a few. On the contrary, in Re-ID, most of the conditions are uncontrolled, e.g., changes in background and illumination over a large number of different cameras, no control on the number of people and possible occlusions, also subjects' direction vary a lot. Hence, due to the more realistic and unconstrained application scenarios, gait-based person Re-ID has been receiving enormous attention from the computer vision and biometric communities (Lee et al. 2014) and several works endorsed quite promising results.

2 GAIT FOR RE-ID: A MULTIDIMENSIONAL OVERVIEW AND WORKS TILL DATE

To better understand the state-of-the-art techniques, as well as the challenges in Gait-based Re-ID, we categorize the paradigm into several dimensions (see Table 2) as shown in Figure 3. In this section, we address each of these dimensions in detail by conducting an extensive survey of various state-of-the-art approaches reported in the literature and discuss their strength and weakness.

2.1 Camera Setting and Image Dimension

The nature and characteristics of the data acquisition setup, i.e., number and type of cameras and dimensionality of the acquired imagery, clearly influence the algorithm to be employed toward Gait-based Re-ID. Re-ID systems exploit either two-dimensional (2D) or 3D information, depending on which image acquisition methods are employed. For example, depth sensors (e.g., Kinect) and motion capture systems (MOCAP) can record directly 3D data of the environment (Josiński et al. 2014; Nambiar et al. 2017b), albeit the 2D-based image sequence are the de-facto standard in real scenarios (Gala and Shah 2014a; Bouchrika et al. 2016).

A typical Re-ID scenario consists of many cameras (overlapping or non-overlapping) distributed across the surveillance network. Among them, some cameras are used for training, i.e., to create a gallery database, and some others are used for testing. Depending upon the number of the cameras used to acquire data and their coverage of the space (overlapping or not), the Re-ID systems differ. In the majority of cases, Re-ID systems consider networks composed of multiple cameras with

Table 2. Examples of Gait Analysis Techniques Applied for Person Re-identification, Classified under Multidimensional Taxonomy

Reference	Camera & dimension♣	View/gait direction	Feature model	Feature analysis; classification△	Application scenario	Dataset
(Liu et al. 2015)	Single, 2D	Independent (11 different views)	Model free (2D silhouettes)	Gait+appearance features (with PCA); Metric Learning to Rank	Short-term-based (appearance features incorporated)	CASIA
(John et al. 2013)	Single, Kinect 3D	Dependent (top-down/ lateral)	Model-free (point cloud)	Frequency response of the height dynamics+ KL-Divergence (Feature selection); ML classifier	Long-term able (but not tested)	TUM-GAID + local studio datasets
(Chattopadhyay et al. 2015)	Multiple, Kinect 3D	Dependent (front and back)	Model based + model free	Soft biometric cues for Re-ID and point cloud voxel-based width image for recognition; LMNN classifier	Long-term able (but not tested)	Local dataset
(Gala and Shah 2014a)	Single, 2D	Independent (random directions in 8 cameras)	Model-free (2D silhouettes)	Gait feature (GEI/FDEI)+ colour (HSV histogram); combined similarity measure	Short-term (colour dependent)	local MCID Database + SAIVT SoftBio dataset
(Wei et al. 2015)	Single, 2D	Dependent (constrained poses were experi- mented)	Model-free (2D silhouettes)	Swiss-system-based cascade ranking; NN/SVM	Long-term able (better results verified w.r.to others)	Indoor CASIA, outdoor SOTON, local PKU datasets
(Kawai et al. 2012)	Single, 2D	Independent (near side view as the query)	Model-free (silhouette)	Fusion of gait feature (STHOG) and colour information; Score-level fusion	Short term (colour dependent)	local dataset
(Nambiar et al. 2016b)	Single, 2D	Dependent (frontal)	Model free (optic flow)	Histogram Of Flow Energy Image (HOFEI); NN classifier	Long-term able (but not test ed)	CASIA and HDA datasets
(Nambiar et al. 2017a)	Single, Kinect 3D	Independent (Five different views)	Model-based (3D Joint info)	Context-based ensemble fusion, SFS feature selection; NN classifier	Long-term able	Vislab KS20 dataset
(Wang et al. 2014, 2016)	Single, 2D	Independent (arbitrary viewpoints)	Model-free (2D silhouette)	Appearance and space-time feature (ColHOG3D); DVR model for cross-view Re-ID	Short-term (colour is integrated)	PRID2011, iLIDS-VID and HDA+
(Bouchrika et al. 2016)	Single, 2D	Independent (arbitrary viewpoints)	Model-based (motion model)	Haar-like template for localization+ magnitude and phase of the Fourier components for gait signature; KNN classifier	Long term (but not tested)	i-LIDS
(Roy et al. 2012)	Single, 2D	Dependent (lateral)	Model-free (2D silhouettes)	Pose Energy Image (PEI)+phase of motion; graph-based path searching	Long-term able (but not tested)	local studio dataset
(Iwashita et al. 2010)	Single, 2D	Independent (arbitrary viewpoints)	Model-free (2D silhouette)	Virtual 3D sequential model generation + affine moment invariance from virtual images; kNN classifier	Long-term able (but not tested)	Local dataset

(Continued)

Table 2. Continued

Reference	Camera & dimension♣	View/gait direction	Feature model	Feature analysis; classification△	Application scenario	Dataset
(Josiński et al. 2014)	MOCAP, 3D	Pose independent (3D data)	Model-based(3D joint info)	MPCA for dimensionality reduction and exIWO meta heuristic for feature selection; 1NN classifier	Long-term able	Local data via 10 NIR Cameras
(Balazia and Sojka 2017)	MOCAP, 3D	Pose independent (3D data)	Model-based(3D joint info)	Maximum Margin Criterion (MMC) method, PCA+LDA; Mahalanobis distance function	Long-term able	MOCAP database, CMU lab

(♣) Camera settings at runtime. Refer to the text for training settings.
(△) ";" separates feature and classifier, "+" refers to combination, and "/" refers to alternatives.

non-overlapping fields of view. Thus, the basic unit of data for analysis is a short video snapshot, typically containing a few gait cycles, taken from a single view. For training the Re-ID system, many of the cameras can be used to collect a gallery of video snapshots that represent the identity of the subject from different views. Then, during runtime, a single camera is used to collect the probe image from which the subject will be re-identified. There exist many publicly available datasets composed of multiple non-overlapping cameras, e.g., CASIA, HDA+, SAIVT (see Table 3), so many works fall in this context of *single-view multiple camera non-overlapping setup*. For example, Gala and Shah (2014b) applied a gait assisted Re-ID algorithm to the SAIVT dataset. This dataset contains views from eight non-overlapping cameras, but to simulate a real-world Re-ID scenario, they form gallery and probe sets from different cameras. During runtime a single camera is used for the probe set and the others are used for the gallery set. A similar approach was carried out in Wei et al. (2015) on the CASIA, SOTON, and PKU datasets, all containing multiple non-overlapping cameras. In Liu et al. (2015), they used 11 different views of the CASIA dataset. During the runtime, they matched individuals from any random viewpoint against the stored RGB images in the gallery database collected from any other viewpoint. Bouchrika et al. (2016) applied gait-based Re-ID methods in two different cameras of the i-LIDS datasets: one camera for gallery and the other for probe. Instead, Wang et al. (2016) randomly splits each sequence pair of the datasets[8] into two subsets of equal size, one for training and one for testing. That work was tested on the i-LIDS-VID, PRID2011, and HDA+ datasets. Roy et al. (2012) and Kawai et al. (2012) employed single views for gait-based Re-ID locally created datasets of non-overlapping cameras. Roy et al. (2012) shows two camera and three camera topologies. In Kawai et al. (2012), seven non-overlapping views, from front to rear-oblique, have been collected. The lateral view is tested against the gallery set of all other views.

Overlapping camera views have also been reported in the literature. thus creating *multi-view multiple camera overlapping* data acquisition setup. Such scenarios are comparatively rare due to the practical constraints on installing and calibrating a large number of cameras. Nevertheless, they can be of interest in scenarios where the imagery is collected at the entrance of buildings or at security checking points using many cameras, so as to provide a multi-view of the subject. An example is the work of Iwashita et al. (2010), where 16 overlapped cameras were used to reconstruct the 3D shape of the person and generate 3D gait models, which were obtained via the volumetric intersection of the extracted silhouettes from walking images. Then, synthetic images

[8] A multi instance ranking technique (Discriminative Video Ranking) has been exploited in this work, so that pairs of image sequences from different camera views were used for learning.

Table 3. Characteristics of the Main Public Datasets Applicable to Gait-based Re-ID

Nameand Ref	#Camera	#Image resolution	#People	Scenario	Main application
CASIA-datasetA	1 (3 views)	352 × 240	20	Outdoor	Gait recognition
CASIA-datasetB	11	320 × 240	124	Indoor	Gait recognition
SOTON (large)	6	20	114	Indoor	Gait Recognition
USF	2	720 × 480	122	Outdoor	Gait recognition
MoBo (Gross and Shi 2001)	6	640 × 480	25	Indoor	Gait analysis
HID-UMD (dataset2)	2	Data unavailable	55	Outdoor	Human identification
TUM-GAID	1 (Kinect)	640 × 480	305	Indoor	Gait recognition
AVAMVG	6	640 × 480	20	Indoor	Gait recognition
OU-MVLP	7	1,280 × 980	10,307	Indoor	Cross-view gait recognition
KinectREID	1 (Kinect)	vary	71	Indoor	Re-ID
Vislab KS20	1 (Kinect)	3D data only	20	Indoor	Re-ID
SAIVT	8	704 × 576	150	Indoor	Person recognition and Re-ID
HDA dataset	13	2,560 × 1,600 (max)	85	Indoor	Person detection and Re-ID
i-LIDS (MCT)	5	576 × 704	119	Indoor	Person tracking
PRID2011	2	64 × 128	245	Outdoor	Person Re-ID
PETS2009	8	768 × 576	(NA)	Outdoor	Person detecion
3DPeS	8	704 × 576	200	Outdoor	People Tracking and Re-ID
MARS	6	1,080 × 1,920 (max)	1,261	Outdoor	Video-based Person re-identification

from arbitrary viewpoints were created and stored as a gallery. During runtime, 2D images of the walking subject in any random direction are acquired via one single camera. Other works consider the use of non-classical data input devices, such as Kinect and MOCAP systems, that acquire directly 3D information about the environment. Recently, many works have been reported with a Kinect sensor. In most cases, the same Kinect is used for both the training and the testing phases (John et al. 2013; Nambiar et al. 2017a, 2017b). In John et al. (2013), they used the TUM-GAID dataset, which includes sequences of people walking in lateral views. In addition, they used a top-down mounted colour-depth camera to acquire a studio dataset for the experiments. The works of Nambiar et al. (2017b) and Nambiar et al. (2017a) consider people walking in various directions with respect to the 3D sensor. Overlapped Kinect views have been reported in Chattopadhyay et al. (2015), where information from a frontal and rear views are fused to bypass the restriction of limited range sensing of individual RGBD cameras.

Motion-capture systems have been used for gait Re-ID in Josiński et al. (2014) and Balazia and Sojka (2017). In the former work, gait sequences of the subjects were recorded with a Vicon Motion Kinematics Acquisition and Analysis System, enabled with 10 near-infrared (NIR) cameras with acquisition speeds of 100fps to 2000fps. The subjects had to wear a special suit with attached markers for the data recording. In Balazia and Sojka (2017), they used the MOCAP database of the CMU Graphics lab,[9] which was collected using 12 Vicon infrared MX-40 cameras. Subjects wear

[9]CMU Graphics Lab. Carnegie-Mellon Motion Capture (Mo-Cap) Database, 2003. http://mocap.cs.cmu.edu.

a black jumpsuit containing 41 markers, which are visible in infra-red. Then the images collected via Vicon cameras are triangulated and produce the 3D data input.

2.1.1 Critical Analysis: Pros and Cons. Most of the classical Re-ID techniques used single-view camera networks without overlap. However, we witnessed a paradigm shift in recent years toward utilising multiple cameras with overlapping fields-of-view or sophisticated devices like Kinect/MOCAP. These systems can acquire 3D information of the subjects's motion, either directly with the 3D sensors or indirectly with the multiple simultaneous 2D views, which is quite promising in terms of producing high-quality pose-invariant data. However, research on incorporating such 3D algorithms with the popular real-world 2D surveillance networks is still underway.

2.2 Pose/Gait Direction

In a typical surveillance scenario, the pose of a person with respect to a camera can vary due to the camera viewpoint and the person's walking direction. As a result, the information extracted from the images—both static and dynamic body cues—may change drastically both across cameras and along time. Hence, the pose/gait direction of the subject is a very significant criteria to be considered in the gait-based Re-ID scheme. Here we explain various approaches reported in gait-based Re-ID, depending on pose of the input data acquired. In this regard, we classify the state-of-the-art works into two major groups: (a) **Pose-dependent** approaches, where the input data are constrained to a specific pose/viewpoint, and (b) **Pose-independent** approaches, where the input data can be in any arbitrary direction.

2.2.1 Pose-dependent Approaches. Pose-dependent approaches consider that the gait direction is constant at each camera; thus the system only deals with a limited set of poses. Although the classical Re-ID paradigm encounters arbitrary poses quite often, there are certain unconstrained natural scenarios where pose-dependent approaches are still useful, for instance, in many indoor scenarios like shopping mall corridors, subway entrance, where walking directions are quite regular.

In many works, a single camera is used for the video acquisition (John et al. 2013; Nambiar et al. 2016b). Multiple camera systems using pose-dependent approaches include Kawai et al. (2012), Roy et al. (2012), and Chattopadhyay et al. (2015). Most works found in the literature perform human gait-based Re-ID in the side view (Roy et al. 2012; Kawai et al. 2012). This is probably because human gait analysis is better observed on the side (lateral) view due to the constant perspective and the similar degree of self-occlusion during the person's trajectory. The number of studies with other views is high except for Nambiar et al. (2016b), which uses a frontal view, and John et al. (2013), which uses side and top views. Similarly, frontal and rear views have been used in Chattopadhyay et al. (2015) for gait-based Re-ID and recognition. Wei et al. (2015) used some specific viewpoints depending on the dataset: side views in the CASIA and SOTON datasets and frontal-back views in the PKU dataset.

2.2.2 Pose-independent Approaches. In this class of approaches, the human walking direction can be arbitrary, which is the most frequent case in a real-world uncontrolled scenario. It often demands more computationally expensive techniques and better-quality data than pose-dependent approaches. Many works in this category acquire random walking directions of subjects collected in networks of cameras without overlap. Thus, whenever a test image of an arbitrary pose is available, it is compared against the gallery stored samples, possibly containing multiple viewpoints, and the correct subject is re-identified. In Liu et al. (2015), the gallery contained subjects walking in 11 different gait directions (CASIA dataset), and during the testing, the view of each probe image was randomly chosen. The cases in Wang et al. (2016), Wang et al. (2014), Gala and Shah (2014a), and Bouchrika et al. (2016) were similar, where the gallery contained unconstrained walking directions of people collected via multiple cameras. Some works exploited projection techniques to achieve pose invariance. For instance, in Bouchrika et al. (2016), a view transformation model (VTM) was employed to transform multiple data samples onto the same common view angle. Likewise, Gala and Shah (2014a) uses a sparsified representation-based cross-view method, and Wang et al. (2016) uses a discriminative video ranking model for cross-view gait-based person Re-ID.

While using 3D data, the alignment of views can be achieved with more simple geometrical transformations. Iwashita et al. (2010) generates pose-invariant 3D models from 2D images. Multiple cameras with overlap allow the use of multi-view algorithms to generate 3D models of the subjects. Virtual images from arbitrary viewpoints can be synthesized from these models to create a gallery with multiple viewpoints. In runtime, the image collected using a single-view camera is compared against those virtual images in the gallery. Similarly, MOCAP technology was also used toward pose-independent 3D data acquisition in Josiński et al. (2014) and Balazia and Sojka (2017). Three-dimensional depth sensors like Kinect provide directly a volumetric information of the body, including skeleton coordinates. Nambiar et al. (2017b) use anthropometric and gait features from Kinect skeleton data to make an actual demonstration of the impact of viewpoint on gait-enabled Re-ID. Their work shows that, despite the fact that 3D skeleton data are naturally viewpoint invariant, the data provided by Kinect are not—the noise level changes with the viewpoint due to self-occlusions. Based on these observations, they propose to use viewpoint as context in a context-aware gait-based person Re-ID study (Nambiar et al. 2017a) and its extension to cross-contextual analysis for gait-based Re-ID (Nambiar et al. 2018).

2.2.3 Critical Analysis: Pros and Cons. Pose-dependent gait-based Re-ID techniques are easy to realise, computationally less expensive, and faster. However, they are not geared toward most real-world unconstrained surveillance applications. On the contrary, albeit pose-independent methods are time-consuming and computationally expensive (mainly due to the algorithms to provide pose invariance), they are the more realistic application-oriented frameworks for Re-ID. Context-aware Re-ID, view-mapping techniques, 3D data from multiple overlapping 2D cameras, or Kinect/MOCAP devices are the most-current approaches used to study pose-invariant gait-based Re-ID.

2.3 Gait Features: Extraction and Analysis

Gait feature (also denoted as gait signature) is the essential characteristic extracted from the sample images corresponding to a person. Usually, gait features are extracted over a gait cycle. A gait cycle is defined as the sequence of events/movements between two consecutive contacts of the same foot with the ground. It is considered the fundamental unit of gait. Hence, gait features computed over a gait cycle define the representative sample pattern of the posture and walking of an individual.

We categorize this dimension into two main types of algorithms namely *model based* and *model free*. Model-based techniques (refer to Section 2.3.1) make use of an explicit kinematics model of the human gait that is fit to the data, whereas model-free approaches (refer to Section 2.3.2) extract information directly from gait image sequences (e.g., silhouette shape, optical flow) by establishing a correspondence between successive frames. In Table 2, we summarized different feature extraction techniques used in gait-based Re-ID, highlighting the model dependency (model based or model free) as well as the type of extracted features.

The feature descriptors extracted from the image sequences may contain high dimensionality, which not only requires large training sets but also leads to computational constraints. Thus, in many of the works, a feature dimensionality reduction stage (e.g., PCA or LDA) accompanies the feature analysis phase. Some popular dimensionality reduction techniques as well as multi-modal fusion strategies employed in the domain are also discussed here.

2.3.1 Model-based Approaches. Model-based approaches make use of *structural models* (2D or 3D) to define the human kinematic features and the *motion models* that determine the temporal evolution of each body part. In Bouchrika et al. (2016), a motion model is encoded in the angular motion of the knee and hip at different gait phases. These features are extracted directly in the 2D images using templates. Then, a viewpoint rectification stage projects these features onto a common normal plane to extract the gait parameters (magnitude and phase of the Fourier components over a gait cycle). This is one of the few works in the literature using a model-based approach in a single 2D image.

With the arrival of MOCAP systems and depth sensors such as Kinect, human body analytics from 3D skeleton data became a reality. The initial full body gait analysis with Kinect was introduced by Microsoft Research (Gabel et al. 2012) by measuring standard stride information and arm kinematics, using the 3D virtual skeleton. Chattopadhyay et al. (2015) was the first to leverage these data to automate gait-based Re-ID followed by gait recognition. In that work, they fuse information from two Kinect sensors: one acquiring frontal and the other acquiring rear views of the subject. To ensure the unique identity of the subject as he moves across the fields of view of different cameras, they used soft biometric cues, derived from the Kinect skeleton streams, that are robust to incomplete gait cycles. Afterward, recognition was carried out by extracting some model-free features named "width image"[10] at the granularity of small fractions of a gait cycle. In Nambiar et al. (2017b), a single Kinect camera was used to implement a novel viewpoint-invariant Re-ID paradigm, leveraging anthropometric and gait features. This is achieved by analysing the subject data collected in different walking directions using the KS20 dataset.[11] Most of the state-of-the-art works assume that skeleton coordinates provided by Kinect data are viewpoint invariant.[12] However, this study showed that the Kinect skeleton computation is viewpoint dependent, since the skeleton reconstruction process depends on viewpoints and self-occlusions. In other words, although the signal is the same, its computation is not, due to change in noise level depending on the viewpoints and self-occlusions. A subsequent work proposed a context-aware ensemble fusion Re-ID framework (Nambiar et al. 2017a) that considers viewpoints as contexts and develops classifiers customized to the viewpoints. That study confirmed that gait features are better for lateral views and anthropometric features (size of body parts) are better for frontal views.

Even though the direct application of the motion capture (MOCAP) system is inconvenient in a real-world Re-ID surveillance system, MOCAP has been used in the developmental stages of gait-based Re-ID algorithms, mainly due to their high precision in recordings and model fitting (Josiński et al. 2014). Gait sequences were recorded, and the corresponding skeleton model was generated. To reduce the high-dimensional motion data and to retain the most relevant features, a data dimensionality reduction (multilinear principal component analysis) followed by feature selection (exIWO) was executed. The performance evaluation in terms of the accuracy of person Re-ID based on the selected feature subset resulted in a correct classification rate of 99.84%. Similarly, there was another recent work, i.e., Balazia and Sojka (2017), that leveraged a MOCAP system used toward gait-based identification. In that particular work, they addressed the person Re-ID problem via an unsupervised approach i.e., clustering them into potential walker identities.

[10] A 2D frame corresponding to the distance transformed voxel volume by obtaining the width information of the subject from his fronto-parallel view.

[11] http://vislab.isr.ist.utl.pt/vislab_multiview_ks20/.

[12] In principle, the signal from the Kinect sensor can be normalised to a canonical viewpoint by a roto-translation transformation.

Additionally, location traces of people within the surveillance area also played a significant role in the identification process, based on the rationale that *You are how you walk*.

2.3.2 Model-free Approaches. In contrast to model-based approaches, model-free approaches do not require intermediate 2D or 3D geometric or kinematic models. The early approaches in gait analysis used 2D model-free techniques based on the analysis of the human silhouette during a gait cycle. The Gait Energy Image (GEI) (Han and Bhanu 2006) is a representation of the spatio-temporal description of a human gait into a single image template by averaging the binary silhouette over a gait cycle. Afterward, a large number of variants of GEI's were introduced, which formed the basis of many recent model-free gait Re-ID works. One of them, Gala and Shah (2014a), employed both GEI and Frame Difference Energy Image (FDEI) as the gait features. Other works based on GEI used Pose Energy Image (PEI) (Roy et al. 2012) and Histogram Of Flow Energy Image (HOFEI) (Nambiar et al. 2016b). In another work by Liu et al. (2015), fusion of multiple features, i.e., appearance feature (HSV histogram and Gabor feature) and gait feature (GEI), have been used for shape and temporal information. Recent work by Wei et al. (2015) proposed to improve the robustness of gait-based person Re-ID on multi-covariate conditions by formulating a cascade ranking model for multi-feature ensemble learning.

Since human silhouettes are difficult to extract from images in general cases, dense optic flow-based methods were employed toward gait analysis. Nambiar et al. (2016b) and Castro et al. (2014) derived new optic flow-based descriptors; respectively, the HOFEI and the Pyramidal Fisher Motion via Divergence-Curl-Shear. Kawai et al. (2012) proposed a spatio-temporal histogram of oriented gradient (STHOG) as a feature that represents both shape and motion information. In Wang et al. (2016), simultaneous selection and matching of reliable features, i.e., multi-fragment-based appearance and space-time feature representation of image sequences, comprising HOG3D, colour, and optic flow energy profiles of image sequences, was carried out. Iwashita et al. (2010) extracted features from synthetic images. Multiple overlapping 2D cameras were used to reconstruct the 3D volume of a walking subject, and virtual images at arbitrary viewpoints are adaptively synthesized. Then, from these synthetic walking images, affine moment invariants were calculated as gait features. During runtime, the walking sequences of a target in any arbitrary direction are captured with single camera, and gait features were computed in the same manner. In the test phase, with a single camera, the features (affine moment invariants) are extracted and matched against those in the gallery for person identification. In John et al. (2013), person Re-ID using height-based gait in colour depth cameras was introduced, where the feature descriptors correspond to the frequency response of the person's height temporal information computed from 3D colour-depth blobs. This was one of the few works exploiting 3D sensors in a model-free fashion.

2.3.3 Dimensionality Reduction. Features extracted from images and video are often high dimensional, containing both redundant and irrelevant information for the task. To address this issue, some popular technique such as dimensionality reduction[13] or feature selection is usually carried out.

One of the most widely used methods toward feature dimensionality reduction is principal component analysis (PCA) (Liu et al. 2015; Roy et al. 2012). Variants of PCA, i.e., multilinear principal component analysis (MPCA) (Josiński et al. 2014) as well as fusion with other feature reduction technique, i.e., principal component analysis and Linear Discriminant Analysis (PCA+LDA) (Balazia and Sojka 2017), have also been used. Feature selection was conducted in some works as a means of selecting the most relevant feature subset from the whole set of

[13]Dimensionality reduction is defined in Burges (2010) as "the mapping of data to a lower dimensional space such that uninformative variance in the data are discarded, or such that a subspace in which the data lives is detected."

features, as in Nambiar et al. (2017a), via the Sequential Forward Selection Algorithm, as well as in John et al. (2013), via the KL-divergence algorithm.

2.3.4 Multi-modal Fusion. Fusion of multiple features is a common strategy to improve Re-ID results. Some works have successfully complemented gait-based features with other types of features in the Re-ID realm. In Kawai et al. (2012), they fuse gait signature features (spatio-temporal HOG) with colour features (HSV histogram). The fused result outperformed either gait or colour features used alone. Similarly, John et al. (2013) combined colour, person-height, and gait, and Wang et al. (2016) combined colour with space-time features. From the various strategies of feature fusion, the most widely used are score-level fusion and feature-level fusion (Ross et al. 2006).[14] In Liu et al. (2015), both score-level fusion and feature-level fusion were applied to various features extracted from the gait sequence for improving the aggregate performance. In Nambiar et al. (2017b) and Nambiar et al. (2017a), they fused gait features with anthropometric features with feature-level and score-level fusion either in a holistic manner or adaptive to viewpoint.

2.3.5 Critical Analysis: Pros and Cons. Albeit the model-free approaches circumvent the difficulties in fitting models to data and are computationally simpler compared to the model-based approaches, they are sensitive to view angle, pose, and scale. On the contrary, model-based techniques are more computationally expensive and require higher-quality data than model-free methods but show better robustness to a variety of factors (changes in the appearance of gait due to clothing, carrying goods, background). More recently, they have been tested with success in many works, yielding better gait-based Re-ID when compared to model-free approaches in inter-class conditions (see Table 4). Especially the RGBD cameras (like KINECT) and precision MOCAP systems allowed model-based approaches to achieve significantly high Re-ID accuracy (Josiński et al. 2014). However, not many surveillance networks can afford to upgrade legacy video cameras to new MOCAP systems or depth cameras, or to increase the number of cameras to achieve overlapping view-fields required for model-based techniques. Instead, we envisage that recent methods on pose estimation from single 2D images such as OpenPose[15] (Cao et al. 2017) will facilitate the application of model-based techniques using standard video technology.

2.4 Classification Approaches

Another important dimension is the classification methods used in the matching process, which is the task of assigning a test probe in a given feature space to corresponding patterns contained in the gallery set. A multitude of classification algorithms has been employed in Re-ID studies. The choice of classification methodologies and distance metrics are highly dependent on the type and amount of data and may have a significant impact on the classification accuracy. In this section, we discuss two broad categories of classification approaches used in gait-based Re-ID: (i) Non-learning-based and (ii) Learning-based classifiers. We highlight the most important distance metrics as well.

2.4.1 Non-learning-based Classification Schemes. k-nearest neighbor (kNN) is one of the most used classifiers in gait-based person Re-ID. It is a simple instance-based classifier that predicts the class of the test feature according to the labels of the k closest training examples. During the training phase, all the known subjects are stored in the gallery along with their labels. In the testing phase, the distances for a particular probe to the known subjects are computed using a distance measure, e.g., Euclidean distance. The calculated distances are ordered in the ascending

[14] In Feature-level fusion, different biometric features of an individual are concatenated after an initial normalization scheme, whereas in Score-level fusion, individual match scores of each biometric features are are evaluated separately and are fused at the end to provide an aggregate score result.

[15] https://github.com/CMU-Perceptual-Computing-Lab/openpose.

Table 4. Performance Analysis of State-of-the-Art Approaches on Gait-based Person Re-ID

Reference	Approach	No. of people used for gait-based person Re-ID	Dataset used	Rank-1 CMC rate/ mean CCR (unless otherwise stated)
(Gala and Shah 2014a)	GEI	≤19	MCID Database	47%
	FDEI	≤19	MCID Database	62%
	GEI(Sparse)	23	SAIVT Database	17.40%
	FDEI(Sparse)	23	SAIVT Database	30.43%
	GEI(NN)	23	SAIVT Database	13.04%
	FDEI(NN)	23	SAIVT Database	17.4%
(Wei et al. 2015)	Swiss-System-based Cascade Ranking	124	CASIA (Indoor)	52%
	"	116	SOTON (outdoor)	42% (Set E) & 26% (Set F)
	"	18	PKU (outdoor)	40% (Cam 1) & 21% (Cam 2)
(Wang et al. 2016)	Discriminative Selection in Video Ranking	200	PRID2011	40%
	"	300	iLIDS-VID	39.5%
	"	83	HDA+	54.3% (HDA 5fps) and 52% (HDA 2fps)
(Kawai et al. 2012)	STHOG (gait feature)	27	Local dataset	39%
(Nambiar et al. 2017b)	multi-modal feature fusion of 3D soft biometric cues	20	Vislab KS20 dataset	39.6% (on average for fully view-invariant case)
(Nambiar et al. 2017a, 2018)	Context-aware ensemble fusion	20	Vislab KS20 dataset	74.67% (no context), 88% (1 context) and 88.67% (2 contexts) and 82.33% (cross-context)
(Liu et al. 2015)	Metric Learning to Rank	124	CASIA dataset	15.27%
(John et al. 2013)	frequency response of the height dynamics+KL-Divergence	75	TUM-GAID dataset	over 80% (KNN classif. accuracy)
(Roy et al. 2012)	Pose Energy Image and phase of motion	≤30	Local dataset	94%
(Bouchrika et al. 2016)	markerless feature extraction	20	iLIDS Cam 2 & 3	92.5%
(Chattopadhyay et al. 2015)	Soft biometric cues (Re-ID) + width image (recognition)	29	Local dataset	70%–90%
(Iwashita et al. 2010)	virtual 3D sequential model + affine moment invariance	5	Local database	50%
(Josiński et al. 2014)	Multilinear PCA (MPCA)	25	MOCAP Local database	97%–99%
(Balazia and Sojka 2017)	Fisher Linear Discriminant + Maximum Margin Criterion (MML)	464	CMU Graphics MOCAP	−75%–85% (ROC)

In this comparative chart, the results exploiting appearance-based features, e.g., colour, are not considered, instead we report only the Re-ID results of gait (or gait+other biometric) reported toward long-term Re-ID paradigm. The numbers are read from the results in the cited papers.

order and the first k subjects in the gallery will be selected. The most frequent subject in the k nearest neighbours is returned as the matching class, via majority voting. If $k = 1$, then the subject is matched to the class of its immediate neighbor. In many gait-based Re-ID works, i.e., Iwashita et al. (2010), Nambiar et al. (2016b), John et al. (2013), Wei et al. (2015), Bouchrika et al. (2011), and Nambiar et al. (2017a), kNN was used as the classifier. Other variations of nearest neighbor were also found in gait-based Re-ID approaches, e.g., the large margin nearest-neighbour classifier (LMNN) (Chattopadhyay et al. 2015).

A different classification scheme using a *sparsified representation* has been presented in Gala and Shah (2014a) for gait-assisted person re-identification. A dictionary matrix is created from the labeled gallery features of many subjects. It is assumed that a test feature of a specific subject can be expressed as the linear combination of the dictionary features from the dictionary. By computing the best linear combination for fitting the test subject to the dictionary while minimizing the l_1 norm of the linear combination coefficients, a sparse representation is obtained and is used for the re-identification.

Distance Metrics In the past, a fair number of distance metrics has been reported for classification, typically as proxies for NN and kNN classifiers. The Euclidean distance metric is the one used by default, in the absence of prior information about the structure of the data. In Roy et al. (2012), PEI feature vectors were compared using Euclidean distance measure. Similarly, Nambiar et al. (2016b) and Bouchrika et al. (2016) also employed Euclidean distance metrics as the similarity measure. In Gala and Shah (2014a), another widely used distance metric, Bhattacharyya distance, was applied in measuring the colour similarity for Re-ID. $L_{0.5}$norm also has been used as the distance metric in gait-based Re-ID works (Kawai et al. 2012).

2.4.2 Learning-based Classification Schemes. Learning-based classification schemes were also employed in gait-based Re-ID. A traditional classifier commonly used is the Support Vector Machine (SVM), which performs classification tasks based on the concept of decision planes (Cortes and Vapnik 1995). Some of the gait-based Re-ID works using SVM were found in Wei et al. (2015) and Wang et al. (2016). A view-independent method was used to compute similarities in Liu et al. (2015), where Metric Learning to Rank (MLR) was used to train a distance function. Similarly, learning a video ranking function for person ReID, viz. Discriminative video ranking (DVR) was presented in Wang et al. (2016). In John et al. (2013), a maximum likelihood classification scheme for a particular class or person was used to identify the test person. Also, in Kawai et al. (2012), a score-level fusion function was trained from a joint distribution of the colour and gait distances and positive (the same person) and negative (different person) labels. For score-level fusion, linear logistic regression (LLR) of the likelihood ratio between positive and negative samples was chosen. Thus, the person in one camera image was re-identified to the person with the minimum fused distance in another camera image.

In gait-based Re-ID, learning-based classifiers such as HMM and deep learning techniques have not yet found applications, mainly due to the constraint of the big amount of data depicting video sequences of multiple subjects walking in unconstrained scenario. Some recent works propose the generation of large amounts of synthetic data through human avatar simulations, as a proxy for dealing with real images of people. SOMAset (Barbosa et al. 2017) presents about 100K samples of subject-clothing-pose combinations in a realistic scene. Albeit it contains huge amount of data applicable to one-shot Re-ID purposes, cannot be deployed toward gait-based Re-ID, due to the lack of gait image frames. A boost in the performance of gait-based Re-ID is envisaged in the near future, as soon as realistic gait simulators are available to generate large amounts of distinct walking patterns suitable to train learning-based models.

2.4.3 Critical Analysis: Pros Cons. Both classifications classes described above have advantages and disadvantages. In the non-learning class of approaches, the advantage is that they exhibit faster running time figures at Re-ID classification, not requiring intensive training procedures. However, a metric should be used. This can be an issue, since not all metrics are appropriate for a given problem. Concerning the learning-based approaches, they have the advantage of appropriately learning the metric that should be used for the classification task. The main limitation of these methods is that they require large training datasets, since the robustness will increase as the training sets becomes larger, thus, requiring high computational cost for a robust classifier learning.

2.5 Application Scenario

Yet another significant dimension is the application scenario, in terms of the temporal and spatial aspects of the patterns to analyse. Temporal aspects deal with the time gap between the gallery and probe samples under analysis, which may vary from a few hours (short-term) to days/weeks/months (long-term). Depending on this dimension the features required for Re-ID may vary significantly. The spatial aspect considers closed-set vs. open-set scenarios, depending on the presence or not of all test subjects in the gallery. In a open-set scenario, the probe subject may be unknown and the algorithm should be able to deal with it (e.g., detect it is unknown and add it to the gallery).

2.5.1 Temporal: Short Term vs. Long Term. In a short-term scenario, the data collected are only valid for some short period of time (from a few hours to a single day), in which the "*appearance constancy hypothesis*" holds well, i.e., there are no changes in the clothing or other appearance traits of the person (e.g., Re-ID in a busy airport, supermarket, shopping mall). Hence, it is plausible to exploit both the appearance cues (color/texture) as well as the gait features toward re-identification purposes. In Kawai et al. (2012), Liu et al. (2015), Gala and Shah (2014a), and Wang et al. (2016), gait and appearance features were fused in Re-ID applications. Most of the existing datasets fall under this category (refer Section 3); SAIVT, HDA, ETZH, i-LIDS, PETS2009, and AVAMVG are some of the short-term datasets.

In contrast to the short-term surveillance, long-term surveillance extends for a longer period, in which the appearance cues are highly volatile, i.e., Re-ID on different days/weeks or regardless of the clothes, accessories, and hairstyle. As mentioned in Gong et al. (2014), "the longer the time and space separation between views is, the greater the chance will be that people may appear with some changes of clothes or carried objects in different camera views." As a result, for Re-ID, we cannot make use of the appearance-based techniques. Instead, soft biometric information is the most suited for these scenarios. As shown in Table 2, most of the works belong to long-term approaches (John et al. 2013; Chattopadhyay et al. 2015; Wei et al. 2015; Nambiar et al. 2017b; Bouchrika et al. 2016; Roy et al. 2012; Iwashita et al. 2010; Josiński et al. 2014; Balazia and Sojka 2017). Some Re-ID datasets were collected explicitly for long-term applications. For instance, USF and 3DPeS were collected over several days and provided variants such as appearance and viewpoint change, change in light conditions, and so on.

2.5.2 Spatial: Closed-set vs. Open-set. Most of the traditional Re-ID approaches, including gait based, rely on the closed-set paradigm. Only a few works on the person Re-ID literature addressed the open space scenario. For instance, Bäuml and Stiefelhagen (2011) first verify whether a detected person matches the gallery or not. If classified as known, then they further determine the identity among the trained persons. In another work (Bedagkar-Gala and Shah 2011), an appearance-(colour) based person Re-ID work was proposed in an open-set scenario upon a dynamically evolving dataset. Re-ID was established by solving a linear assignment problem, where each gallery ID was compared to each probe ID, and the minimum assignment cost entitled the best Re-ID match. Another recent work (Liao et al. 2014) considered the open-set Re-ID problem as two sub-tasks:

detection and identification. In the former, the system decides whether the probe subject is present in the gallery or not. In the latter, the system determines the identity of the accepted probe. All the aforementioned works addressed appearance-based Re-ID approaches within the open-set scenarios. However, the issue of open-set Re-ID has not yet been addressed on a gait-based Re-ID paradigm but is expected to be a relevant research direction in the coming years.

In Table 2, we deem all the works that exploit colour/texture as "short term" and others as "long term." Among the latter, we distinguish whether they have been tested and verified for long-term sequences or not. Regarding the spatial aspect, all the works mentioned fall under the closed-set scenario and hence are not explicitly highlighted in the table.

2.5.3 Critical Analysis: Pros and Cons. In the temporal case, short-term methodologies have the advantage of leveraging both the appearance and gait-based features. In long-term applications, however, appearance-based approaches are less discriminative, and biometric features seem to be the best choice. A caveat of long-term methods is the lack of datasets having large time spans. Also, the training of the long-term cues are still open issues. On the spatial aspect, most of the existing datasets consider closed-set scenario, in which probe-gallery matching is not a complex task assuming that probe is a subset of the gallery. On the contrary, in open-set scenarios, databases are dynamically evolving in nature, and the probe may not exist in the gallery thus making the Re-ID process very challenging by demanding complex training procedures and evaluation criteria. In summary, both long-term and open-set scenarios poses new challenges that are not entirely addressed in the literature.

3 AVAILABLE DATASETS FOR GAIT-BASED RE-IDENTIFICATION

Among the plethora of publicly available datasets for video surveillance applications, only a few can be adopted for gait-based Re-ID. Since it demands spatio-temporal evaluation of the walking pattern, gait-based Re-ID datasets require multiple-shot image sequences of subjects walking in different directions. In this section, we categorize the datasets into two major classes based on the scenario constraints: (i) datasets collected in controlled scenarios and (ii) datasets collected in uncontrolled scenarios.

3.1 Datasets Collected in Controlled Scenarios

These datasets are collected in a controlled scenario where people walk in predefined paths. Despite the majority of works in gait-based re-ID demand unconstrained scenarios, the following datasets are still useful for algorithm benchmarking, since they contain important covariates to consider in gait-based Re-ID (different clothing, different poses/viewpoints, different backgrounds, etc.). Some of them have actually been used in gait bsed Re-ID works, and others have the potential to do so (see Figure 4 for some datasets collected in controlled scenarios).

- **CASIA:** This is one of the largest and more popular databases in gait analysis and related research.[16] It contains four different datasets: Dataset A (Wang et al. 2003b) is an outdoor gait dataset consisting of 20 people walking in three directions (lateral (90°), frontal (0°), and 45°). Dataset B is an indoor gait dataset composed of 13,640 samples acquired from 124 subjects at 11 different views (Yu et al. 2006). Dataset C is collected using an infrared camera from 153 subjects at four different conditions (normal, slow, fast, normal with a bag). Dataset D was collected simultaneously with a camera and a Rscan Footscan[17] on 88 subjects. In most of the gait analysis works, Dataset A and Dataset B are widely used due to their realistic mode of data acquisition (Wang et al. 2003a; Sivapalan et al. 2012).

[16] http://www.cbsr.ia.ac.cn/english/Gait\%20Databases.asp.
[17] http://www.rsscan.com/footscan/.

Fig. 4. Shot examples from (a) a CASIA indoor dataset (Yu et al. 2006), (b) a CASIA outdoor dataset (Wang et al. 2003b), and (c) SOTON large dataset (Nixon and Carter 2006). Both the CASIA and SOTON databases contain the indoor and outdoor scenarios. Nevertheless, they are collected in highly constrained conditions, unlike in a realistic surveillance environment. Hence, these databases are more suited for gait recognition than Re-ID.

Specifically for **gait-based Re-ID**, Dataset B is well suited since it addresses the issue of pose by acquiring the scene with a network of 11 cameras, each with a view angle separation of 18°, and was used in Wei et al. (2015), Nambiar et al. (2016b), and Liu et al. (2015).

— **SOTON:** The SOTON[18] database (Nixon and Carter 2006) was developed at the University of Southampton, with the principal aim of developing new technologies for recognising people at a distance. It is composed of a large and a small database. The former consists of nearly 114 subjects and over 5,000 samples but contains little variability (each subject was filmed from only two different views over three separate scenarios). The small database contains only 12 persons but is more complete regarding the covariates (change in clothing, accessories, and different speeds). For the large database, three scenarios are analyzed, namely outside, inside track, and inside treadmill, whereas the small database contains subjects walking with different appearances and at various speeds, all collected in the indoor scenario. The SOTON dataset has been used for **gait-based Re-ID** in Wei et al. (2015).

— **USF:** The USF[19] dataset (Sarkar et al. 2005) contains 1,870 sequences acquired from 122 subjects. It comprises elliptical movements of people walking in front of cameras. For each person, up to five covariates were manipulated such as shoe type, bag carried, type of surface, viewpoint, and time instants. The data are composed of sequences of 33 subjects collected over 4 days (two acquisitions in May and the rest in November 2001). Although no work was reported in gait-based Re-ID using the USF database, it has the potential to be used due to the various covariates of data and multiple subject walking directions.

— **MoBo:** The MoBo[20] dataset (Gross and Shi 2001) contais video data from multiple synchronized cameras. The CMU Motion of Body (MoBo) database was collected in a studio setup, CMU 3D Room). The MoBo dataset contains 25 subjects performing four different walking activities (slow walk, fast walk, incline walk and walking with a ball) on a treadmill. Six cameras were evenly distributed around the treadmill capturing more than 8,000 images per subject. Different viewpoints and walking activities provide ample potential for its possible application in gait-based Re-ID.

— **HID-UMD database:** The HID-UMD [21] database (Kale et al. 2003) is exploited in gait recognition and face recognition systems for human identification at a distance. It comprises gait sequences of around 25–50 people in four poses, i.e., walking toward, walking away, toward left, and toward right. Specifically, the first dataset has 25 people collected using one camera, and the second dataset has 55 people collected using two cameras (simultaneously from orthogonal views). To test the view synthesis, another dataset was acquired from 12 people

[18] http://www.gait.ecs.soton.ac.uk/.
[19] http://figment.csee.usf.edu/GaitBaseline/.
[20] http://www.ri.cmu.edu/publication_view.html?pub_id=3904.
[21] http://www.umiacs.umd.edu/labs/pirl/hid/data.html.

walking at angles of 0°, 15°, 30°, 45°, and 60° with respect to the camera. Despite its potential to study pose invariance aspects of gait-based re-ID, we could not find works using dataset for this purpose.

— **TUM-GAID:** A new freely available database or multimodal gait recognition was proposed in Hofmann et al. (2014). It is denoted GAID[22] (Gait from Audio, Image and Depth) and contains RGB video, depth, and audio concurrently. It is composed of recordings from 305 people in three variations, making it one of the largest to date. A second subset of 32 people was recorded to further investigate challenges of temporal variability. **Gait-based Re-ID** research work was done with the TUM-GAID dataset in John et al. (2013).
— **AVAMVG:** The AVA Multi-View Dataset for Gait Recognition AVAMVG[23] (López-Fernández et al. 2014) is another recent dataset directed toward robust recognition. It collects data of 20 people walking along 10 trajectories each, using six calibrated cameras with different views angles. Images have a resolution of 640×480 pixels and are acquired at 25Hz. The database has been specifically designed to test gait recognition algorithms based on 3D data. The binary silhouettes of each video sequence are also provided. Some gait recognition works were reported in the past using the AVAMVG dataset (Castro et al. 2014; López-Fernández et al. 2016). Multiple views and various gait trajectories enable the dataset to be leveraged toward gait-based Re-ID in future.
— **OU-MVLP:** The OU-ISIR Gait Database, Multi-View Large Population Dataset (OU-MVLP)[24] was created by the Institute of Scientific and Industrial Research (ISIR), Osaka University (OU), toward the research in developing, testing, and evaluating algorithms for cross-view gait recognition (Takemura et al. 2018). The dataset comprises 10,307 subjects (5,114 males and 5,193 females with various ages between 2 and 87 years), observed from from 14 view angles, i.e., ranging from 0° to 90° and from 180° to 270°. Seven networked cameras are placed at intervals of 15-deg azimuth angles along a quarter of a circle, and are used to collect gait images of $1{,}280 \times 980$ pixels at 25fps. The subject repeats forward and backward walking twice, thus making 28 gait image sequences per subject, i.e., (7 (cameras) $\times$ 2 (forward and backward) $\times$ 2 (twice)). The OU-MVLP database is one of the very latest database available so far, and applications in gait-based Re-ID have not been reported so far.
— **KinectREID:** One of the few person Re-ID datasets collected using the Kinect sensor in an unconstrained environment is KinectREID[25] (Pala et al. 2015). The purpose of the dataset is to provide data to test and evaluate algorithms of **person re-identification** using features extracted from the Kinect sensor: anthropometry, **gait**, and appearance of the clothes using both the skeleton features and RGB-D data. It is composed of many video sequences of 71 people, acquired indoor at various illumination conditions and various angles: three front, three behind, and a side. Also, appearance variations, i.e., carrying backpacks, bags, or other accessories were incorporated in the dataset. All these variables make the dataset a possible candidate to evaluate gait-based Re-ID algorithms.
— **Vislab KS20:** The KS20 Vislab Multi-view Kinect Skeleton dataset[26] (Nambiar et al. 2017b) is a new dataset collected by the authors in the context of long-term person re-identification. It comprises Kinect skeleton data sequences (3D coordinates of joints)

[22] https://www.mmk.ei.tum.de/verschiedenes/tum-gaid-database/.
[23] http://www.uco.es/investiga/grupos/ava/node/41.
[24] http://www.am.sanken.osaka-u.ac.jp/BiometricDB/GaitMVLP.html.
[25] http://pralab.diee.unica.it/it/PersonReIdentification.
[26] http://vislab.isr.ist.utl.pt/vislab_multiview_ks20/.

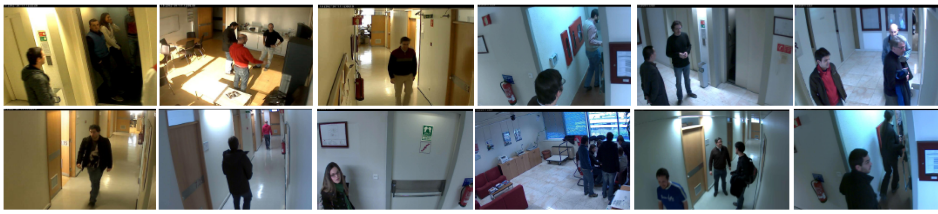


Fig. 5. Example of video frames from HDA dataset (Nambiar et al. 2014). It is collected indoor, under a multi-camera network. It provides data of realistic uncontrolled conditions, with significant variation in the pose, illumination, and camera view angle.

collected from 20 subjects walking at different directions, using Kinect v2. The major motivation behind the creation of this dataset was the lack of similar Kinect datasets consisting of people walking in different viewpoints (other than just the view angles of ~0° and ~90°) and to actually demonstrate the real impact of viewpoints and self-occlusions on the Re-ID paradigm. In this regard, multiple walking sequences along five different directions, i.e., Left lateral (LL at ~0°), Left diagonal (LD at ~30°), Frontal (F at ~90°), Right diagonal (RD at ~130°), and Right lateral (RL at ~180°) were collected. Altogether it has 300 skeleton image sequences collected from 20 subjects (3 video sequences per person in a particular viewpoint) in the aforementioned directions. It was deployed in **gait-based Re-ID** works in Nambiar et al. (2017b), Nambiar et al. (2017a), and Nambiar et al. (2018).

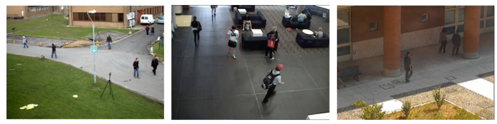
3.2 Datasets Collected in Uncontrolled Scenarios

This category contains datasets acquired in more realistic scenarios, i.e., people walking in natural everyday environments without being instructed to behave in any particular way. Usually, such datasets are collected in real-world scenarios like office building, airports, and so on, and are quite fundamental to analyse the performance of the algorithms in practical applications.

— **HDA person dataset:** The HDA[27] dataset is a multi-camera video dataset mainly dedicated to benchmarking video surveillance algorithms such as person detection and Re-ID (Nambiar et al. 2014). It is a fully labeled image sequence dataset, collected using 13 indoor cameras for a duration of 30 minutes (Figure 5). More than 64,000 annotations were performed on a total of more than 75,000 frames. The dataset is quite diverse in terms of types of cameras (standard, high, and very high resolution), environment types (corridors, doors, open spaces), and frame rates (5fps, 2fps, 1fps). Several of the acquired image sequences are in the HR range (1,280 × 800 pixel and 2,560 × 1,600 pixel), which makes the HDA dataset the first one to include labeled video sequences of such resolution. Extended versions of the dataset have been published *viz.*, HDA+ dataset (Figueira et al. 2014) along with a novel framework toward fully automated person Re-ID (Taiana et al. 2014). Some **gait-based Re-ID** works employing HDA datasets are Nambiar et al. (2016b) and Wang et al. (2016).

— **i-LIDS:** The Imagery Library for Intelligent Detection Systems (i-LIDS) is the U.K. government's benchmarking dataset for video analytics systems. It comprises a library of CCTV video footage collected from various scenarios mainly categorized as event detection and object tracking scenarios. Among them, the i-LIDS multiple camera tracking (MCT) scenario was collected inside a busy hall using five cameras at 25fps. One hundred nineteen people were captured, but the average image count per person is four, which is very few

[27] http://vislab.isr.ist.utl.pt/hda-dataset/.



(a) (b) (c)

Fig. 6. Image samples from (a) PETS2009, (b) SAIVT, and (c) 3DPeS datasets.

for gait-based applications. The presence of occlusions and quite large illumination changes make this dataset very challenging for the Re-ID task. An extended versions of the i-LIDS dataset, iLIDS-VID,[28] is presented in Wang et al. (2014). Bouchrika et al. (2016) presented identity tracking across multiple cameras using i-LIDS, and Wang et al. (2016) presented **gait-based Re-ID** using the i-LIDS-VID dataset.

— **PRID2011:** This dataset was created for the purpose of testing person Re-ID approaches[29] (Hirzer et al. 2011). It consists of image frames extracted from two static camera recordings depicting people walking in different directions. Images from both cameras contain variations in viewpoint, illumination, background, and camera characteristics. Four hundred seventy-five– and 856-person trajectories were recorded via individual cameras, with 245 persons appearing in both views. The dataset has two versions: a single-shot scenario and a multi-shot scenario. PRID2011 has been employed in **gait-based Re-ID** applications in Wang et al. (2014) and Wang et al. (2016).
— **PETS2009:** A widely known dataset is PETS[30] (Ferryman and Shahrokni 2009), presented at the 2009 edition of the international workshop on performance evaluation of tracking and surveillance. It was recorded in a public space outdoor scene at University of Reading, UK. It is a multi-camera system consisting of eight cameras, and it contains three sequences with different crowd activities in a real-world environment (Figure 6(a)). The partial dataset that addresses person tracking consists of three subclasses based on their subjective difficulty level, associated with the density of the crowd. Refer to Baltieri et al. (2011a) for some benchmarking results. Since the dataset provides multi-shot sequences with multiple viewpoints, this is useful for **gait-based Re-ID**. In Bouchrika et al. (2016), gait-based Re-ID and tracking across multiple non-intersecting cameras have been applied to the PETS2009 dataset.
— **SAIVT:** Recently, a multi-camera surveillance database, SAIVT[31] (Bialkowski et al. 2012), was created for the evaluation of person recognition and Re-ID models in realistic surveillance scenarios. The database consists of unconstrained walking video sequences of 150 people, collected inside a building (Figure 6(b)). Eight surveillance cameras acquired images of resolution 704 × 576 pixels at a frame rate of 25 frames per second. The dataset provides a highly unconstrained environment for testing person Re-ID models in conditions that are closer to real scenarios. **Gait-based Re-ID** employing SAIVT dataset has been published in Gala and Shah (2014a).
— **3DPeS:** The 3D People Surveillance Dataset 3DPes[32] (Baltieri et al. 2011b) is a dataset designed mainly for person Re-ID and tracking (Figure 6(c)). The dataset was captured by a multi-camera network of eight different cameras within a real surveillance scenario. Data were collected on different days. Since it is an outdoor dataset, it presents high variations of

[28] http://www.eecs.qmul.ac.uk/~xz303/downloads_qmul_iLIDS-VID_ReID_dataset.html.
[29] https://lrs.icg.tugraz.at/datasets/prid/.
[30] http://www.cvg.reading.ac.uk/PETS2009/a.html.
[31] https://wiki.qut.edu.au/display/saivt/SAIVT-SoftBio+Database.
[32] http://www.openvisor.org/3dpes.asp.

light conditions. Background models of the cameras are available and the 1,012 snapshots of 200 persons are provided with silhouette masks and bounding box information. Some more details and benchmarking results on person Re-ID can be found in Vezzani et al. (2013) and Baltieri et al. (2015). Some works have already proposed to use 3DPeS dataset in future work for **gait-based Re-ID** (Kawai et al. 2012).

— **MARS:** The Motion Analysis and Re-identification Set (MARS)[33] dataset was published for video-based person re-identification. MARS contains multiple frames of video sequences, which enables it to be used for gait-based Re-ID. Six near-synchronized cameras (five 1,080 × 1,920 HD cameras and one 640 × 480 SD camera) were installed in the campus of Tsinghua University for dataset acquisition. MARS consists of 1,261 different pedestrians who are captured by at least two cameras, and around 20,000 video sequences, making it the largest video Re-ID dataset to date. Albeit many works on video based Re-ID have been proposed[34] on the MARS dataset, gait-based methods are yet to be reported.

There are many other datasets individually available for Re-ID (e.g., Viper (Gray et al. 2007), CAVIAR4REID (Cheng et al. 2011), CUHK03 (Li et al. 2014), and Market-1501 (Zheng et al. 2015)), as well as for gait analysis (e.g., TUM-IITKGP Gait Database (Hofmann et al. 2011) and Multi Biometric Tunnel (Seely et al. 2008)). However, since we focus on the gait-based Re-ID, we consider only those datasets containing video sequences of multiple people walking in various directions and in various appearances, either in a controlled or uncontrolled manner. Only those datasets made available for public access are described in this section; local datasets are not considered. The details of the datasets above are summarized in Table 3.

4 PERFORMANCE EVALUATION METRICS

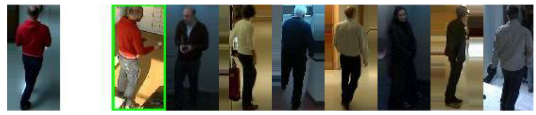
Depending on the scenario and context of the application, the evaluation metrics employed in the Re-ID task may also vary accordingly (Vezzani et al. 2013). One noteworthy point is that for either appearance-based or biometric-based Re-ID, the evaluation metrics used are the same and, therefore, in this section we analyse the performance evaluation metrics used for person Re-ID in general. Here we present the different alternatives available for particular implementations of Re-ID as either recognition or identification.

4.1 Re-ID as Recognition

CMC curve and nAUC: To evaluate the performance of Re-ID algorithms in closed-set identification scenarios, the cumulative matching characteristic (CMC) curve is the most acclaimed and popular method of choice. CMC measures how well the system ranks the identities in the enrolled database given the unknown probe image. As described in Nambiar et al. (2014), "the CMC curve shows how often, on average, the correct person ID is included in the best K matches against the training set for each test image." Hence, the Re-ID task is considered as a recognition problem, with the assumption that exactly one sample class in the gallery corresponds to the query. As a result, the Re-ID output is given as a ranked list of gallery classes, based on some matching similarity to the query probe.

A comprehensive characterization of the CMC curve for evaluation or recognition problems was explained in Moon and Phillips (2001), where it was originally proposed for the evaluation of face-recognition algorithms (on FERET image sets). Most works in the area of gait-based Re-ID employed CMC (e.g., Gala and Shah (2014a), Nambiar et al. (2017b), Wei et al. (2015), Wang

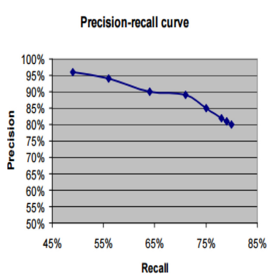
[33] http://www.liangzheng.com.cn/Project/project_mars.html.
[34] http://www.liangzheng.com.cn/Project/state_of_the_art_mars.html.



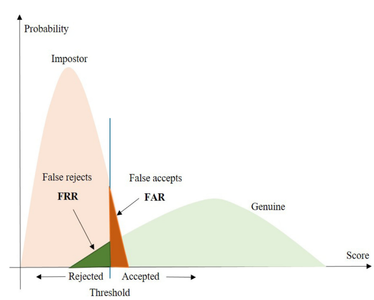
(a)

(b)

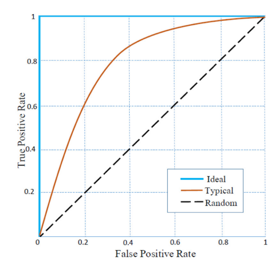
Fig. 7. Performance visualization: Examples of queries made in HDA person dataset reidentified by the methodology described in Figueira et al. (2013). The probe query is shown at the left side and the top eight results in the ranked Re-ID order is shown. The correct match is highlighted in green. (a) Person is correctly identified in rank 1 (b) Person correctly identified in rank 2.



(a)



(b)



(c)

Fig. 8. Traditional curves used to evaluate the performance of Re-ID as a biometric identification/verification system; (a) Precision-Recall curve in the Re-ID experiment (Hamdoun et al. 2008); (b) FAR and FRR measures; (c) receiver operating characteristic (ROC) curve.

et al. (2016), Nambiar et al. (2016a), Chattopadhyay et al. (2015), Liu et al. (2015), and Roy et al. (2012)). Also, in some works only the Rank-1 accuracy, also called the Correct Classification Rate (CCR), was used to represent the system performance (e.g., Josiński et al. (2014), Kawai et al. (2012), Iwashita et al. (2010), John et al. (2013), and Bouchrika et al. (2016)).

Derived from the CMC curve, another metric called *normalised area under the CMC curve* (nAUC) is also used to characterize the overall algorithm performance with a single score (e.g., Gala and Shah (2014a)). The nAUC illustrates the performance of a method independent of the dataset size, with perfect and chance nAUC values as 1.0 and 0.5, respectively.

Confusion Matrix: Another way of depicting the results of the Re-ID is with the help of the confusion matrix. A confusion matrix has as many columns and rows as the number of classes. Each entry i, j of the matrix contains the fraction of cases of class i classified as class j. Diagonal terms express the accuracy of recognizing each class, and off-diagonal elements represent false classifications. The more "diagonal" the matrix, the more accurate the method. The confusion matrix inherits its name by the ability to inspect the cases of confusion, i.e., which classes are more prone to erroneous classifications. Application of confusion matrices in gait-based Re-ID analysis were found in Nambiar et al. (2015), Middleton et al. (2005), and Bialkowski et al. (2013).

Performance visualization: A common qualitative method for representing the ranked list of gallery classes against the test query is via visual representation. This is usually accomplished by plotting the ranked set of persons' bounding boxes (see Figure 7). This type or representation, although qualitative, is very useful for a on-the-fly subjective interpretation of the achieved results. Some cases of Re-ID performance visualization are shown in Vezzani et al. (2013) and Layne et al. (2012).

4.2 Re-ID as Identification

Precision-Recall (P/R) statistics: In Re-ID scenarios related to identification, there may be instances to classify that are outside of the knowledge base. In this case, evaluation typically uses precision and recall (P/R) statistics. Another scenario where these metrics are suitable is, for instance, in a shopping mall, where we want to track people along a camera network but we are not particularly interested in their real identity. This scenario is similar to data clustering, where the samples of multiple subjects have to be grouped without prior knowledge. Ideally, each cluster relates to a single individual.

The performance evaluation of such a system works similarly to a biometric verification system that checks if two instances belong to the same person. This analysis includes checking occurrences of false positives[35] and missed detections.[36] To conduct a fair evaluation of the influence of false positives and missed detections, P/R statistics are more suitable than the CMC.

$$\text{Precision} = \frac{\text{CorrectIdentifications}}{\text{TruePositiveDetections} + \text{FalsePositiveDetections}} = \frac{\text{CorrectIdentifications}}{\text{NumberofDetections}} \quad (1)$$

$$\text{Recall} = \frac{\text{CorrectIdentifications}}{\text{TruePositiveDetections} + \text{MissedDetections}} = \frac{\text{CorrectIdentifications}}{\text{NumberofPersonAppearances}} \quad (2)$$

An excerpt from Hamdoun et al. (2008) depicting a P/R curve is presented in Figure 8(a). Also, some other Re-ID works employing PR curves are Figueira et al. (2014) and Riccardo Satta and Roli (2014). One work in gait-based Re-ID-exploited PR is Balazia and Sojka (2017). Another useful metric derived from precision-recall is the F1-score, which indicates a test's accuracy. As detailed in Tsai and Kwee (2011), "the F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst at 0, which means the higher the F-score, the more accurate the test" (see Figueira et al. (2014) and Cancela et al. (2014)).

FAR and FRR: Some other standard biometric evaluation measures used in particular identification/verification problems (e.g., biometric access control) are the false acceptance rate (FAR), the false rejection rate (FRR), the Receiver Operating Characteristic (ROC) curve, and the Decision Error Tradeoff (DET) curve. FAR is the percentage of accepted non-genuine (impostor) individuals with the total acceptance made by the system. It measures the probability that a system wrongly verifies an authentication trial by an unauthorized user. Similarly, FRR is the percentage of rejected genuine individuals compared to total rejects made by the system. It measures the probability that a system wrongly rejects an authentication attempt by an authorized user. Figure 8 (b) shows a pictorial representation of FAR and FRR.

An ideal human identification system requires the recognition performance with both FAR and FRR at zero level. Since this is hard to achieve in real-world applications, a threshold is determined by the type of application. For example, if Re-ID is to provide access control/ authentication purposes, then the system prefers to keep FAR as low as possible (lower the access chance for impostors). In some other situations like forensic scenarios, the preference would be to reduce FRR, since

[35] False positive is an error in data reporting in which a test result wrongly indicates that a condition has been attained, when in reality it is not.
[36] Missed detection (false negative) is an error in which a test result wrongly indicates that a condition has not been attained, when in reality it is.

we do not want to reject genuine individuals connected to the crime activity. For the real-world Re-ID scenarios such as Open-set Re-ID, these identification metrics are of great use, wherein the system has to check whether gallery contains the probe ID or not. FAR-FRR application to person Re-ID was reported in Jungling and Arens (2010), as well as in gait recognition (Wang et al. 2003a). There are still no reported uses in gait-based Re-ID.

Receiver Operating Characteristics: The Receiver operating characteristic (ROC) curve is a well-accepted measure to express the performance of 1:1 matches. It shows the measurements of true-positive rate (TPR, genuine users accepted) against the false-positive rate (FPR, impostor users accepted) at various threshold settings (see Figure 8(c)). The ROC curve is also known as the sensitivity vs. (1 - specificity) plot, since TPR and FPR are equivalent to sensitivity and 1 - specificity, respectively. The ideal result yields a point in the upper left coordinate (0,1) of the ROC space, as shown in Figure 8(c), corresponding to the no false negative and no false positive scenario (i.e., 100% sensitivity and 100% specificity). Another alternate representation of ROC is the Detection Error Tradeoff (DET) graph. A DET curve plots the false-negative rate against the false-positive rate on non-linearly transformed axes, using either standard normal deviates or logarithm. Some instances of the applications of ROC and DET measures in gait analysis and Re-ID scenarios can be found in Sivapalan et al. (2012), Liao et al. (2014), DeCann and Ross (2013), DeCann and Ross (2015), and in particular (Balazia and Sojka 2017) for gait-based Re-ID.

4.3 Re-ID in Forensics

Beyond the application of FAR and FRR evaluation metrics in forensic scenarios, there are other standard measures commonly used.

Likelihood Ratio: The likelihood ratio (LR) is a standard measure of information that summarizes, in a single number, the data support for a hypothesis (Perlin 2010). It is a good legal and scientific standing that underlies the credibility of forensic science in court by quantifying the belief in a hypothesis. Basically, LR is defined as "the ratio of two probabilities of the same event under different hypotheses. For two events, say A and B, the probability of A given B is true, divided by the probability of event A given B is false, is termed as a likelihood ratio" (Vezzani et al. 2013).

$$\mathbf{Likelihoodratio(LR)} = \frac{Pr(A|B)}{Pr(A|\neg B)} \tag{3}$$

In crime scenarios, $Pr(E|S)$ is the probability of the evidence if the suspect is the source of evidence, and $Pr(E|U)$ is the probability of the evidence if an unknown (unrelated) is the source of evidence, then likelihood is calculated as follows:

$$\mathbf{Likelihoodratio(LR)} = \frac{Pr(E|S)}{Pr(E|U)}. \tag{4}$$

The likelihood ratio metric has found its application in gait recognition (Muramatsu et al. 2014) as well as in classical image-based Re-ID framework (Koestinger et al. 2012). However, its application has not yet been reported in the gait-based Re-ID literature.

4.4 Performance Evaluation of State-of-the-Art Works and Comparative Analysis

As a part of our survey, we carried out a performance evaluation of various state-of-the-art techniques in gait-based Re-ID works, using the information provided in the surveyed papers. Table 4 provides a detailed summary of the SOA works in the literature, reporting the approach, dataset used, number of people involved in the evaluation, and the performance reported as Rank-1 CMC rate (CCR accuracy), whenever available.

From the table, it is observed that most of the SOA approaches carried out in realistic datasets (like SAIVT, iLIDS, HDA+, PRID2011, etc.) are still struggling to overcome 50% CCR accuracy. However, some recent approaches exploiting local datasets (acquired in their laboratory) have reported higher performances. For instance, Roy et al. (2012) achieved up to 94% via hierarchical combination of gait with spatiotemporal and phase of motion, Nambiar et al. (2017a) reached up to 88% Rank-1 CMC rate via Context-aware ensemble fusion and Chattopadhyay et al. (2015) performed up to 70–90% using a model-based+model-free framework. Nevertheless, due to the low variance and the small size of dataset, such works may be overshadowed by the issues like overfitting. Among the whole set of works, Josiński et al. (2014) produced the best result of a correct classification rate of 99.84% using high-quality recording and model fitting using a MOCAP system under a controlled environment. Similar improved results were observed in most of the 3D data- (Kinect/MOCAP) based Re-ID techniques as well (Chattopadhyay et al. 2015; John et al. 2013; Nambiar et al. 2017a; Balazia and Sojka 2017). With the increase in the quality of sensors, availability of data, and better algorithms, further improvements in performance in realistic scenarios are envisaged in the near future.

5 CONCLUSION AND DISCUSSION

Gait-based Re-ID is a recent field in pattern recognition that aims at recognising and identifying people by their gait in unconstrained scenarios typical of video surveillance systems. In contrast to the classical appearance-based person Re-ID approaches, only applicable in short-term scenarios, gait-based Re-ID can be utilised toward long-term applications. Hence, gait-based Re-ID entails a large potential for applications in video surveillance, human–robot interaction, ambient assisted living, among others. However, due to the unconstrained nature of the scenario and variability of factors such as subject pose, appearance, camera viewpoint, background diversity, illumination and occlusions, gait-based Re-ID is a more challenging problem than the traditional gait recognition in controlled setups.

In this article, we carried out a detailed survey of the various approaches to gait-based person Re-ID. To describe the topic in detail, we analysed its various dimensions, i.e., (i) Camera setting and Image dimension, (ii) Pose/gait direction, (iii) Gait features: extraction and analysis, (iv) Classification approaches, and (v) Application scenario. A critical analysis highlighting the strengths and weakness of various approaches was reported. In addition, existing datasets for gait-based Re-ID (datasets collected in (i) controlled and (ii) uncontrolled scenarios), evaluation metrics, and the performance analysis of the existing approaches were presented. Next, based on our survey, we discuss the challenges, limitations, and novel trends in the state-of-the-art works in the gait-based Re-ID and point out some promising directions for future research.

5.1 Main Challenges and Promising Directions

5.1.1 Pose Invariance. One of the most challenging issues in gait-based person Re-ID is the walking direction of the subject. On one hand, the walking direction will limit the number of gait cycles available for analysis (more in frontal views than lateral). However, the quality of the acquired features for analysis depends on the pose of the person, which is typically aligned with the walking direction (Nambiar et al. 2017b). Due to the open nature of the re-identification scenarios, gait-based Re-ID would benefit a lot from true pose-invariant approaches to gait analysis. Even though some pose-invariant approaches have been proposed using 2D images (Wang et al. 2016; Wei et al. 2015) or 3D models generated out of multiple 2D cameras (Iwashita et al. 2010), their levels of accuracy still lag behind what can be achieved with MOCAP data (Josiński et al. 2014), therefore leaving room for improvements. Recently, many state-of-the-art methods address the pose-invariant 3D data generated out of depth sensors like Kinect (Nambiar et al. 2017a, 2017b) and MOCAP devices (Josiński et al. 2014; Balazia and Sojka 2017). These 3D techniques have been quite revolutionary in terms of data acquisition, as well as classification accuracy; however, it is not clear how to exploit them in traditional surveillance scenarios that use 2D cameras. In this regard, novel techniques like open-pose (Cao et al. 2017) bestowing real-time pose estimation upon 2D images looks promising for future work.

5.1.2 Gait Signature. Regarding feature extraction and analysis, we discussed both the model-based and model-free techniques (Section 2.3). Albeit model-free methods are simpler and faster in leveraging the human appearance, they are highly sensitive to variations in the environment, e.g., different pose, walking directions, appearance changes, illumination, and so on. Conversely, model-based approaches provide better results and seem to be well suited for Re-ID applications where the environment varies drastically. However, the quality of the data has a significant impact on algorithm performance (noise sensitivity), and the fitting process often requires a significant computational cost. Probably the combination of model-based and model-free techniques, in a holistic approach to gait-based Re-ID, can be a future direction worth exploring.

The combination of multiple features is also of great interest. However, which features to be selected, the optimal fusion strategy, and how to adaptively select the modalities are questions still unanswered. Current approaches sometimes end up in reducing Re-ID accuracy due to redundant or noisy data. Multimodal fusion of various discriminative features have been a novel concept in the field, and many works are in the path of exploring it further. Bag of soft-biometrics to combine multiple biometric traits (Dantcheva et al. 2010) and selective context-aware Re-ID (Nambiar et al. 2017a) have been new ideas in this path for further research.

5.1.3 Open-set and Long-term Scenarios. Another significant problem is the application of gait analysis in unbounded spatio-temporal conditions, *viz.*, open-set Re-ID and long-term Re-ID. First, regarding the spatial scenario, most of the Re-ID works were treated under closed-set scenarios, i.e., assuming that the probe ID exists in the gallery. On the contrary, in open-set scenarios the system should be able to deal with novel subjects, i.e., persons not yet enrolled in the gallery (Gala and Shah 2014b). Examples of a closed-set scenarios are indoor surveillance like offices or other private spaces, where only authorized personnel can enter. Open-set scenarios are more applicable to supermarkets, shopping malls, airports, and so on, where everyone can enter. Hence, the open-set scenario is a more challenging problem than the closed-set scenario.

Second, regarding the temporal scenario, although gait is well suited for long-term person identification, only a few works have verified their performance over the longer periods. All of them suffer degradation of Re-ID performance with changes in the covariates (e.g., changes in days, carrying attributes, seasonal variations in the dressing styles) compared to their performance over short periods. Recently, some works addressed the problem of long-term Re-ID leveraging person's characteristic appearance and its variations over time (Bedagkar-Gala and Shah 2011) or some adaptation scheme for dynamic camera network (Panda et al. 2017). However, the use of gait cues in long-term Re-ID has not been attempted yet.

5.1.4 Dataset. Availability of big data is also another challenge. The acquisition of big amounts of video sequences in realistic scenarios to address the gait-based Re-ID problem is still an open problem. Although most of the traditional datasets were focused on constrained scenarios, nowadays there is a shift toward a diversification to unconstrained scenarios by considering big unconstrained data, seasonal changes, long-term scenarios, and so on. Such diversification is beneficial, as it allows for realistic recording settings while incorporating all the possible challenges in the scene. Also, such big data are quite essential toward the application of novel machine-learning algorithms such as deep learning. The largest available dataset for gait-based Re-ID are the

MARS and HDA Person datasets. In addition to the real datasets, synthetic have also been gaining attention, e.g., SOMAset (single-shot imagery, thus, not suited for gait analysis). We envision more datasets to be created in the coming years, facilitating big data analytics and deep learning toward gait-based Re-ID.

5.1.5 Improving Re-ID Accuracy. Re-ID accuracy depends critically the quality of the data and extracted features. Incomplete or noisy data of walking sequences may result in either missed or erroneous feature descriptors. Camera characteristics, illumination constraints, and occlusions influence the data quality in 2D or 3D approaches. One critical aspect of gait-based Re-ID approaches is, thus, to determine automatically which data are reliable enough for analysis. Some approaches like Taiana et al. (2014) reason about possible occlusions among subjects and improve results on person Re-ID, but such solutions are yet to be incorporated in gait-based Re-ID works. Likewise, the classification strategy also greatly influences the Re-ID performance. We can observe a new direction shift toward learning-based classification approaches from the traditional classification schemes (Wei et al. 2015; Liu et al. 2015) (see Section 2.4). We envisage the possible application of deep learning strategies to gait-based Re-ID. The main roadblock is the necessity of large training datasets, but MARS and SOMAset have just become available. Though no works were reported explicitly leveraging deep learning architectures, very soon they will find application in gait-based Re-ID.

5.1.6 Different and Varying Context. Once a gait-based Re-ID system has been trained in a certain context (activity, neighbouring persons, clutter, and crowd), the key question to answer is how to transfer the learned knowledge to different and/or varying contexts. Some works tried to exploit the idea of context to improve the performance of the algorithms by making the gallery search more efficient using contextual variables such as walking direction (Nambiar et al. 2017a) and human attributes (Zhang et al. 2014). Nevertheless, the selection of the context or contextual variable to apply in a particular scenario is a big challenge. We expect camera topology, environmental setup under surveillance, cues such as the types of activities under consideration (leisure, work, passage), and amount of interactions among persons to enrich the the quality of the methods in the next few years. Incorporating multiple contexts, adaptive selection of the contexts according to the changing scenario are also promising directions in this regard.

5.1.7 Dynamic Networks. With the proliferation of mobile cameras and the easy of deployment of new cameras in a network, the dynamic nature of a surveillance system gains importance. Such a dynamic network has to deal with the problem of transferring information from the data already gathered in the installed cameras to pair with a target camera newly introduced in the surveillance network, so that the target camera can be trained from existing data. Otherwise, retraining the whole network from the scratch is quite laborious and time-consuming. Some promising works have reported recently in the domain of person Re-ID, addressing the issue of adding new cameras in a dynamically evolving surveillance network. For instance, Panda et al. (2017) proposed to use an unsupervised adaptation scheme, viz., *domain adaptation*, to effectively find the best source camera (already installed) to adapt with a newly introduced target camera, so that new cameras can be added in a dynamic network, with minimal additional effort. This is a very practical problem to be addressed in surveillance, and we envisage that more research on the topic is underway.

To finalise this survey, we would like to stress that gait-based Re-ID is a young and challenging field with a plethora of opportunities yet to be explored. In these first years of research, several methodologies are in place and show interesting results but have been evaluated in datasets with highly disparate characteristics. The lack of a reference dataset that can challenge the algorithms with a wide variability of possible scenarios has made difficult the fair comparison among approaches until now. With the recent emergence of large datasets, better sensors, and end-to-end training systems like deep learning, we expect substantial improvements in future years.

REFERENCES

Tauseef Ali, Luuk Spreeuwers, and Raymond Veldhuis. 2012. *Forensic Face Recognition: A Survey*. Nova Publishers.

Michal Balazia and Petr Sojka. 2017. You are how you walk: Uncooperative MOCAP gait identification for video surveillance with incomplete and noisy data. In *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB'17)*. IEEE, 208–215.

Davide Baltieri, Roberto Vezzani, and Rita Cucchiara. 2011b. 3dpes: 3d people dataset for surveillance and forensics. In *Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding*. ACM, 59–64.

Davide Baltieri, Roberto Vezzani, and Rita Cucchiara. 2011a. Sarc3d: A new 3d body model for people tracking and re-identification. In *Proceedings of the International Conference on Image Analysis and Processing*. Springer, 197–206.

Davide Baltieri, Roberto Vezzani, and Rita Cucchiara. 2015. Mapping appearance descriptors on 3d body models for people re-identification. *Int. J. Comput. Vis.* 111, 3 (2015), 345–364.

Igor Barros Barbosa, Marco Cristani, Barbara Caputo, Aleksander Rognhaugen, and Theoharis Theoharis. 2018. Looking beyond appearances: Synthetic training data for deep cnns in re-identification. *Computer Vision and Image Understanding* 167 (2018), 50–62.

David Barrett. 2013. One surveillance camera for every 11 people in Britain, says CCTV survey. Retrieved from http://www.telegraph.co.uk/technology/10172298/One-surveillance-camera-for-every-11-people-in-Britain-says-CCTV-survey.html.

Martin Bäuml and Rainer Stiefelhagen. 2011. Evaluation of local features for person re-identification in image sequences. In *Proceedings of the 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS'11)*. IEEE, 291–296.

Apurva Bedagkar-Gala and Shishir K. Shah. 2011. Multiple person re-identification using part based spatio-temporal color appearance model. In *Proceedings of the Computer Vision Workshops (ICCV Workshops'11)*. IEEE, 1721–1728.

Alina Bialkowski, Simon Denman, Sridha Sridharan, Clinton Fookes, and Patrick Lucey. 2012. A database for person re-identification in multi-camera surveillance networks. In *Proceedings of the Conference on Digital Image Computing: Techniques and Applications (DICTA'12)*. IEEE, 1–8.

Alina Bialkowski, Patrick Lucey, Xinyu Wei, and Sridha Sridharan. 2013. Person re-identification using group information. In *Proceedings of the Conference on Digital Image Computing: Techniques and Applications (DICTA'13)*. IEEE, 1–6.

Randolph Blake and Maggie Shiffrar. 2007. Perception of human motion. *Annu. Rev. Psychol.* 58 (2007), 47–73.

Imed Bouchrika, John N. Carter, and Mark S. Nixon. 2016. Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras. *Multimedia Tools Appl.* 75, 2 (2016), 1201–1221.

Imed Bouchrika, Michaela Goffredo, John Carter, and Mark Nixon. 2011. On using gait in forensic biometrics. *J. Forens. Sci.* 56, 4 (2011), 882–889.

Christopher J. C. Burges. 2010. Dimension reduction: A guided tour. *Found. Trends. Mach. Learn.* 2, 4 (2010), 275–365.

Brais Cancela, Timothy M Hospedales, and Shaogang Gong. 2014. *Open-World Person Re-identification by Multi-label Assignment Inference*. British Machine Vision Association.

Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'17)*, Vol. 1. 7.

Francisco M. Castro, Manuel J. Marín-Jimenez, and Rafael Medina-Carnicer. 2014. Pyramidal Fisher Motion for multiview gait recognition. *arXiv preprint arXiv:1403.6950* (2014).

Pratik Chattopadhyay, Shamik Sural, and Jayanta Mukherjee. 2015. Information fusion from multiple cameras for gait-based re-identification and recognition. *IET Image Process.* 9, 11 (2015), 969–976.

Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino. 2011. Custom pictorial structures for re-identification. In *Proceedings of the British Machine Vision Conference (BMVC'11)*, Vol. 1. 6.

Patrick Connor and Arun Ross. 2018. Biometric recognition by gait: A survey of modalities and features. *Comput. Vis. Image Understand.* 167 (2018), 1–27.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20, 3 (1995), 273–297.

Antitza Dantcheva, Carmelo Velardo, Angela D'Angelo, and JeanLuc Dugelay. 2010. Bag of soft biometrics for person identification: New trends and challenges. *Mutimedia Tools Appl.* 51 (2010), 739–777.

Brian DeCann and Arun Ross. 2010. Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment. In *Biometric Technology for Human Identification VII*, Vol. 7667. International Society for Optics and Photonics, 76670Q.

Brian DeCann and Arun Ross. 2013. Relating roc and CMC curves via the biometric menagerie. In *Proceedings of the IEEE 6th International Conference on Biometrics: Theory, Applications and Systems (BTAS'13)*. IEEE, 1–8.

Brian DeCann and Arun Ross. 2015. Modelling errors in a biometric re-identification system. *IET Biometr.* 4, 4 (2015), 209–219.

Brian DeCann, Arun Ross, and Mark Culp. 2014. On clustering human gait patterns. In *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR'14)*. IEEE, 1794–1799.

Brian DeCann, Arun Ross, and Jeremy Dawson. 2013. Investigating gait recognition in the short-wave infrared (swir) spectrum: Dataset and challenges. In *Biometric and Surveillance Technology for Human and Activity Identification X*, Vol. 8712. International Society for Optics and Photonics, 87120J.

Piotr Dollár, Christian Wojek, Bernt Schiele, and Pietro Perona. 2009. Pedestrian detection: A benchmark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 304–311.

Gianfranco Doretto, Thomas Sebastian, Peter Tu, and Jens Rittscher. 2011. Appearance-based person reidentification in camera networks: Problem overview and current approaches. *J. Amb. Intell. Human. Comput.* 2, 2 (2011), 127–151.

J. Ferryman and A. Shahrokni. 2009. An overview of the pets 2009 challenge. In *Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*.

Dario Figueira, Loris Bazzani, Ha Quang Minh, Marco Cristani, Alexandre Bernardino, and Vittorio Murino. 2013. Semi-supervised multi-feature learning for person re-identification. In *Proceedings of the 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'13)*. IEEE, 111–116.

Dario Figueira, Matteo Taiana, Athira Nambiar, Jacinto Nascimento, and Alexandre Bernardino. 2014. The HDA+ data set for research on fully automated re-identification systems. In *Proceedings of the ECCV 2014 Workshop on Visual Surveillance and Re-Identification*. 241–255.

Frontex. 2011. Application of surveillance tools to border surveillance—Concept of operations. http://ec.europa.eu/enterprise/policies/security/files/doc/conops_gmes_en.pdf.

Moshe Gabel, Ran Gilad-Bachrach, Erin Renshaw, and Assaf Schuster. 2012. Full body gait analysis with Kinect. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.

Davrondzhon Gafurov. 2007. A survey of biometric gait recognition: Approaches, security and challenges. In *Proceedings of the Annual Norwegian Computer Science Conference* Citeseer, 19–21.

Apurva Bedagkar Gala and Shishir K. Shah. 2014a. Gait-assisted person re-identification in wide area surveillance. In *Proceedings of the Asian Conference on Computer Vision Workshops (ACCV Workshops'14)*, 633–649.

Apurva Bedagkar Gala and Shishir K. Shah. 2014b. A survey of approaches and trends in person re-identification. *Image Vis. Comput. J.* 32, 4 (2014), 270–286.

Shaogang Gong, Marco Cristani, Chen Change Loy, and Timothy M. Hospedales. 2014. The re-identification challenge. In *Person Re-Identification*. Springer, 1–20.

Douglas Gray, Shane Brennan, and Hai Tao. 2007. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proceedings of the IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS'07)*, Vol. 3.

R. Gross and J. Shi. 2001. *The CMU Motion of Body (MoBo) Database*. Technical Report CMU-RI-TR-01- 18. Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu, and Bruno Steux. 2008. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *Proceedings of the 2nd ACM/IEEE International Conference on Distributed Smart Cameras*.

Arun Hampapur, Lisa Brown, Jonathan Connell, Sharat Pankanti, Andrew Senior, and Yingli Tian. 2003. Smart surveillance: Applications, technologies and implications. In *Proceedings of the IEEE Pacific-Rim Conference on Multimedia (2)*. 1133–1138.

Ju Han and Bir Bhanu. 2006. Individual recognition using gait energy image. *IEEE Trans. PAMI* 28, 2 (2006), 316–322.

Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof. 2011. Person re-identification by descriptive and discriminative classification. In *Proceedings of the Scandinavian Conference on Image Analysis*. Springer, 91–102.

Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll. 2014. The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits. *J. Vis. Commun. Image Represent.* 25 (2014), 195–206.

Martin Hofmann, Shamik Sural, and Gerhard Rigoll. 2011. Gait recognition in the presence of occlusion: A new dataset and baseline algorithms. In *Proceedings of the 19th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. 99–104.

Yumi Iwashita, Ryosuke Baba, Koichi Ogawara, and Ryo Kurazume. 2010. Person identification from spatio-temporal 3D gait. In *Proceedings of the International Conference on Emerging Security Technologies (EST'10)*. IEEE, 30–35.

Gunnar Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14 (1973), 201–211.

Vijay John, Gwenn Englebienne, and Ben Krose. 2013. Person re-identification using height-based gait in colour depth camera. In *Proceedings of the 2013 IEEE International Conference on Image Processing*. IEEE, 3345–3349.

Henryk Josiński, Agnieszka Michalczuk, Daniel Kostrzewa, Adam Witoski, and Konrad Wojciechowski. 2014. Heuristic method of feature selection for person re-identification based on gait motion capture data. In *Proceedings of the 6th Asian Conference on Intelligent Information and Database Systems (ACIIDS'14)*. 585–594.

Kai Jungling and Michael Arens. 2010. Local feature based person reidentification in infrared image sequences. In *Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'10)*. IEEE, 448–455.

Amir Kale, Naresh Cuntoor, B. Yegnanarayana, A. N. Rajagopalan, and Rama Chellappa. 2003. Gait analysis for human identification. In *Proceedings of the International Conference on Audio-and Video-Based Biometric Person Authentication*. Springer, 706–714.

Ryo Kawai, Yasushi Makihara, Chunsheng Hua, Haruyuki Iwama, and Yasushi Yagi. 2012. Person re-identification using view-dependent score-level fusion of gait and color features. In *Proceedings of the 21st International Conference on Pattern Recognition*. IEEE, 2694–2697.

Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof. 2012. Large scale metric learning from equivalence constraints. In *Computer Vision and Pattern Recognition*. IEEE, 2288–2295.

Thomas Kress and Irene Daum. 2003. Developmental prosopagnosia: A review. *Behav. Neurol.* 14, 3-4 (2003), 109–121.

Ryan Layne, Timothy M. Hospedales, Shaogang Gong, and Q. Mary. 2012. Person re-identification by attributes. In *Proceedings of the British Machine Vision Conference (BMVC'12)*, Vol. 2. 8.

Lily Lee. 2002. Gait analysis for recognition and classification. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. 155–162.

Tracey K. M. Lee, Mohammed Belkhatir, and Saeid Sanei. 2014. A comprehensive review of past and present vision-based techniques for gait recognition. *Multimedia Tools Appl.* 72, 3 (2014), 2833–2869.

Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. 2014. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 152–159.

Shengcai Liao, Zhipeng Mo, Jianqing Zhu, Yang Hu, and Stan Z. Li. 2014. Open-set person re-identification. *arXiv preprint arXiv:1408.0872* (2014).

Zheng Liu, Zhaoxiang Zhang, Qiang Wu, and Yunhong Wang. 2015. Enhancing person re-identification by integrating gait biometric. *Neurocomputing* 168 (2015), 1144–1156.

D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. 2016. A new approach for multi-view gait recognition on unconstrained paths. *J. Vis. Commun. Image Represent.* 38 (2016), 396–406.

David López-Fernández, Francisco José Madrid-Cuevas, Ángel Carmona-Poyato, Manuel Jesús Marín-Jiménez, and Rafael Muñoz-Salinas. 2014. The AVA multi-view dataset for gait recognition. In *Proceedings of the International Workshop on Activity Monitoring by Multiple Distributed Sensing*. Springer, 26–39.

Yasushi Makihara, Darko S. Matovski, Mark S. Nixon, John N. Carter, and Yasushi Yagi. 2015. Gait recognition: Databases, representations, and applications. *Wiley Encyclopedia of Electrical and Electronics Engineering* (2015).

Laurence T. Maloney and Brian A. Wandell. 1986. Color constancy: A method for recovering surface spectral reflectance. *J. Opt. Soc. Am.* 3, 1 (1986), 29–33.

Lee Middleton, Alex A. Buss, Alex Bazin, and Mark S. Nixon. 2005. A floor sensor system for gait recognition. In *Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies*. IEEE, 171–176.

Hyeonjoon Moon and P. Jonathon Phillips. 2001. Computational and performance aspects of PCA-based face-recognition algorithms. *Perception* 30, 3 (2001), 303–321.

Daigo Muramatsu, Yasushi Makihara, and Yasushi Yagi. 2014. View transformation-based cross-view gait recognition using transformation consistency measure. In *Proceedings of the International Workshop on Biometrics and Forensics*. IEEE, 1–6.

Athira Nambiar, Alexandre Bernardino, and Jacinto Nascimento. 2015. Shape context for soft biometrics in person re-identification and database retrieval. *Pattern Recogn. Lett.* 68, 2 (2015), 297–305.

Athira Nambiar, Alexandre Bernardino, and Jacinto C. Nascimento. 2016a. Person re-identification based on human query on soft biometrics using SVM regression. In *Proceedings of the 11th International Conference on Computer Vision Theory and Applications*. 484–492.

Athira Nambiar, Alexandre Bernardino, Jacinto C. Nascimento, and Ana Fred. 2017a. Context-aware person re-identification in the wild via fusion of gait and anthropometric features. In *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG'17)*. 973–980.

Athira Nambiar, Jacinto C. Nascimento, Alexandre Bernardino, and José Santos-Victor. 2016b. Person re-identification in frontal gait sequences via histogram of optic flow energy image. In *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 250–262.

Athira Nambiar, Matteo Taiana, Dario Figueira, Jacinto Nascimento, and Alexandre Bernardino. 2014. A multi-camera video data set for research on high-definition surveillance. *Int. J. Mach. Intell. Sens. Sign. Process.* 1, 3 (2014), 267–286.

Athira M. Nambiar, Alexandre Bernardino, and Jacinto C. Nascimento. 2018. Cross-context analysis for long-term view-point invariant person re-identification via soft-biometrics using depth sensor. In *Proceedings of the International Conference on Computer Vision Theory and Applications*. 105–113.

Athira M. Nambiar, Alexandre Bernardino, Jacinto C. Nascimento, and Ana L. N. Fred. 2017b. Towards view-point invariant person re-identification via fusion of anthropometric and gait features from Kinect measurements. In *Proceedings of the International Conference on Computer Vision Theory and Applications*.

Mark S. Nixon and John N. Carter. 2006. Automatic recognition by gait. *Proc. IEEE* 94, 11 (2006), 2013–2024.

Mark S. Nixon, Paulo L. Correia, Kamal Nasrollahi, Thomas B. Moeslund, Abdenour Hadid, and Massimo Tistarelli. 2015. On soft biometrics. *Pattern Recogn. Lett.* 68, 2 (2015), 218–230.

Mark S. Nixon, Tieniu Tan, and Ramalingam Chellappa. 2010. *Human Identification Based on Gait.* Vol. 4. Springer Science & Business Media.

Federico Pala, Riccardo Satta, Giorgio Fumera, and Fabio Roli. 2015. Multi-modal person re-identification using RGB-D cameras. *IEEE Trans. on Circuits and Systems for Video Technology* 26, 4 (2015), 788–799.

R. Panda, A. Bhuiyan, V. Murino, and A. K. Roy-Chowdhury. 2017. Unsupervised adaptive re-identification in open world dynamic camera networks. In *Proceedings of the IEEE Int. Conference on Computer Vision Pattern Recognition.*

Mark W. Perlin. 2010. Explaining the likelihood ratio in DNA mixture interpretation. In *Proceedings of the Promega's 21st International Symposium on Human Identification.* 11–14.

Alvin Plantinga. 1961. Things and persons. *Rev. Metaphys.* 14 (1961), 493–519.

Giorgio Fumera Riccardo Satta, Federico Pala and Fabio Roli. 2014. People search with textual queries about clothing appearance attributes. In *Person Re-Identification.* Springer, 371–389.

Daniel Riccio, Maria De Marsico, Riccardo Distasi, and Stefano Ricciardi. 2014. A comparison of approaches for person re-identification. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods.* 189–198.

Arun Ross and Anil K. Jain. 2007. Human recognition using biometrics: An overview. *Ann. Télécommun.* 62, 1 (2007), 11–35.

Arun A. Ross, Karthik Nandakumar, and Anil K. Jain. 2006. *Handbook of Multibiometrics.* Vol. 6. Springer Science & Business Media.

Aditi Roy, Shamik Sural, and Jayanta Mukherjee. 2012. Hierarchical method combining gait and phase of motion with spatiotemporal model for person re-identification. *Pattern Recogn. Lett.* 33, 14 (2012), 1891–1901.

Sudeep Sarkar, P. Jonathon Phillips, Zongyi Liu, Isidro Robledo, Patrick Grother, and Kevin Bowyer. 2005. The human id gait challenge problem: data sets, performance, and analysis. *IEEE Trans. PAMI* 27 (2005), 162–177.

Richard D. Seely, Sina Samangooei, Lee Middleton, John N. Carter, and Mark S. Nixon. 2008. The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In *Proceedings of the 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems.* 1–6.

Sabesan Sivapalan, Daniel Chen, Simon Denman, Sridha Sridharan, and Clinton Fookes. 2012. The backfilled GEI-a cross-capture modality gait feature for frontal and side-view gait recognition. In *Proceedings of the Conference on Digital Image Computing: Techniques and Applications (DICTA'12).* IEEE, 1–8.

Sarah V. Stevenage, Mark S. Nixon, and Kate Vince. 1999. Visual analysis of gait as a cue to identity. *Appl. Cogn. Psychol.* 13, 6 (1999), 513–526.

Shigemasa Sumi. 1984. Upside-down presentation of the Johansson moving light-spot pattern. *Perception* 13, 3 (1984), 283–286.

Matteo Taiana, Dario Figueira, Athira Nambiar, Jacinto Nascimento, and Alexandre Bernardino. 2014. Towards fully automated person re-identification. In *Proceedings of the International Conference on Computer Vision Theory and Applications.* 140–147.

Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. 2018. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Trans. Comput. Vis. Appl.* 10, 1 (2018), 4.

Flora S. Tsai and Agus T. Kwee. 2011. Database optimization for novelty mining of business blogs. *Expert Syst. Appl.* 38, 9 (2011), 11040–11047.

Roberto Vezzani, Davide Baltieri, and Rita Cucchiara. 2013. People re-identification in surveillance and forensics: A survey. *ACM Comput. Surv.* 46, 2 (2013), 1–36.

Liang Wang, Tieniu Tan, Weiming Hu, and Huazhong Ning. 2003a. Automatic gait recognition based on statistical shape analysis. *IEEE Trans. Image Process.* 12, 9 (2003), 1120–1131.

Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. 2003b. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. PAMI* 25, 12 (2003), 1505–1518.

Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. 2014. Person re-identification by video ranking. In *Proceedings of the European Conference on Computer Vision.* Springer, 688–703.

Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. 2016. Person re-identification by discriminative selection in video ranking. *IEEE Trans. PAMI* 38, 12 (2016).

Lan Wei, Yonghong Tian, Yaowei Wang, and Tiejun Huang. 2015. Swiss-system based cascade ranking for gait-based person re-identification. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence.* 197–202.

Michael W. Whittle. 1996. Clinical gait analysis: A review. *Hum. Move. Sci.* 15, 3 (1996), 369–387.

L. Wiegler. 2008. Big brother in the big apple [national security - video surveillance]. In *Engineering & Technology (3)*, Vol. 9. DOI: https://doi.org/10.1049/et:20080902

Shiqi Yu, Daoliang Tan, and Tieniu Tan. 2006. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, Vol. 4. IEEE, 441–444.

Liyan Zhang, Dmitri V. Kalashnikov, Sharad Mehrotra, and Ronen Vaisenberg. 2014. Context-based person identification framework for smart video surveillance. *Mach. Vis. Appl.* 25, 7 (2014), 1711–1725.

Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision.*

Liang Zheng, Yi Yang, and Alexander G. Hauptmann. 2016. Person re-identification: Past, present and future. *CoRR* abs/1610.02984 (2016). http://arxiv.org/abs/1610.02984.

Received February 2017; revised July 2018; accepted August 2018