Artif Intell Rev
DOI 10.1007/s10462-016-9474-x

Biometric recognition in surveillance scenarios: a survey

João Neves[1] · **Fabio Narducci**[2] · **Silvio Barra**[2] · **Hugo Proença**[1]

Abstract Interest in the security of individuals has increased in recent years. This increase has in turn led to much wider deployment of surveillance cameras worldwide, and consequently, automated surveillance systems research has received more attention from the scientific community than before. Concurrently, biometrics research has become more popular as well, and it is supported by the increasing number of approaches devised to address specific degradation factors of unconstrained environments. Despite these recent efforts, no automated surveillance system that performs reliable biometric recognition in such an environment has become available. Nevertheless, recent developments in human motion analysis and biometric recognition suggest that both can be combined to develop a fully automated system. As such, this paper reviews recent advances in both areas, with a special focus on surveillance scenarios. When compared to previous studies, we highlight two distinct features, i.e., (1) our emphasis is on approaches that are devised to work in unconstrained environments and surveillance scenarios; and (2) biometric recognition is the final goal of the surveillance system, as opposed to behavior analysis, anomaly detection or action recognition.

Keywords Human motion analysis · Surveillance · Biometric recognition · Scene understanding · Detection · Tracking · Recognition · Unconstrained scenarios

✉ João Neves
joao_crn@hotmail.com; jcneves@ubi.pt

Fabio Narducci
fabio.narducci@na.icar.cnr.it

Silvio Barra
silvio.barra@na.icar.cnr.it

Hugo Proença
hugomcp@di.ubi.pt

[1] IT - Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal

[2] Institute of High Performance Computing and Networking, National Research Council of Italy (ICAR-CNR), Naples, Italy

Published online: 22 March 2016

1 Introduction

The deployment of video surveillance cameras has grown astonishingly in recent years, with more than 4 million CCTV cameras reported in the UK alone (McCahill and Norris 2002). Surveillance data have become easily available as the number of real-time streams increased in recent years (EarthCam 2014; Terena 2014), which raises interest in automated surveillance systems that are capable of analyzing and understanding human behavior and even performing human identification.

At the same time, biometrics has experienced a popularity growth, while novel algorithms have minimized user cooperation and relaxed recognition systems constraints, e.g., Iris-On-The-Move (Matey et al. 2006). Despite the achievements that were attained in recent years, no automated system has yet been able to perform reliable biometric recognition in surveillance scenarios. These scenarios are typically harsh for recognition purposes and are usually denoted as "wild" scenarios, for a number of reasons: (1) environments are non-standard and are subject to irregular lighting changes according to daylight, weather conditions and reflections; (2) the background regions are complex, and the human resolution could differ significantly in distinct scene locations; (3) subjects move freely through the scene, which could induce occlusions; and (4) the system should work covertly and not require subjects to be cooperative, which hinders the capture of facial biometric data. For these reasons, biometric identification "in the wild" is still considered to be the "grand challenge" (Jain et al. 2004).

However, the recent advances in human motion analysis and biometric recognition suggest that both fields can be combined in a joint approach to develop a fully automated system for biometric recognition purposes.

Human motion analysis refers to a broad area that is mainly devoted to describing and understanding human actions (Moeslund and Granum 2001; Gavrila 1999). Despite the multitude of applications in this field (Moeslund et al. 2006), such as the analysis of human conditions (e.g., athletic performance, medical diagnosis) and human computer interaction, more and more studies have focused on surveillance applications, including people counting (Hou and Pang 2011), crowd analysis (Feris et al. 2013), recognition of actions and behaviors and detection of abnormal activities. Surveillance systems that rely on human motion analysis usually share three main stages: pre-detection, detection and tracking. With regard to pre-detection, an increasing number of background subtraction algorithms have been especially interested in providing additional robustness to surveillance scenarios (Maddalena and Petrosino 2008; Barnich and Droogenbroeck 2011). Additionally, this trend is confirmed by the development of benchmarks that are specifically focused in assessing the performance of background subtraction in these scenarios (Brutzer et al. 2011). Similarly, in the detection phase, robustness to surveillance scenarios is confirmed by the increasing interest in extending human detection to highly challenging conditions, where a large number of subjects move freely in outdoor scenarios. In the tracking field, in spite of the majority of the approaches being not specifically focused on surveillance scenarios, a large effort has been made to benchmark state-of-the-art algorithms with the VOT challenges (Vot 2015), which has consequently contributed to propelling forward the performance of tracking algorithms in complex scenes.

On the other hand, biometrics research was also capable of improving the performance of recognition algorithms in non-ideal conditions. These advances are especially evident in face recognition approaches, whose performance has moved forward remarkably (e.g., the progress of the verification accuracy reported by the LFW dataset). Such developments are

Table 1 Previous surveys on human motion analysis or on surveillance systems

References	Focus
Cédras and Shah (1995)	Motion analysis/action recognition
Aggarwal et al. (1998)	Motion analysis/action recognition
Gavrila (1999)	Human motion analysis/action recognition
Aggarwal and Cai (1999)	Human motion analysis/action recognition
Moeslund and Granum (2001)	Human motion analysis/action recognition
Wang et al. (2003)	Human motion analysis/action recognition
Hu et al. (2004)	Visual-surveillance/activity analysis
Davies and Velastin (2005)	Surveillance systems
Pantic et al. (2006)	Human computer interaction/action recognition
Poppe (2007)	Human motion analysis/action recognition
Moeslund et al. (2006)	Human motion analysis/action recognition
Krger et al. (2007)	Action recognition
Zhou and Hu (2008)	Human motion analysis
Haering et al. (2008)	Visual-surveillance
Turaga et al. (201)	Action recognition
Ji and Liu (2010)	Human motion analysis/action recognition
Poppe (2010)	Action recognition
Kim et al. (2010)	Visual-surveillance
Weinland et al. (2011)	Action recognition
Raty (2010)	Surveillance systems
Turaga et al. (201)	Action recognition
Ko (2008)	Visual-surveillance/activity analysis
Aggarwal and Ryoo (2011)	Action recognition
Popoola and Wang (2012)	Abnormal behaviour
Sodemann et al. (2012)	Visual-surveillance

due to the large number of novel datasets that were specifically devised to study the problem of unconstrained face recognition, which again demonstrates the interest in identifying humans in surveillance scenarios.

The increasing number of surveys and reviews, as shown in Table 1, which specifically cover the advances in human motion analysis in surveillance scenarios, also confirms the increasing importance of surveillance applications. Moreover, the large number of surveillance-oriented human motion analysis studies have proven to be fruitful and have resulted in automated surveillance systems, such as the W4 (Haritaoglu et al. 2000), which is intended to recognize human actions.

In contrast, this survey aims to contribute to the development of a fully automated surveillance system for human identification purposes by reviewing the most recent advances that have been attained both in human motion analysis and biometric recognition, with special emphasis placed on surveillance scenarios. When compared to previous surveys, as described on Table 1, two distinctive features can be highlighted: (1) the emphasis is placed on approaches that are devised to work in unconstrained environments / surveillance scenarios; and (2) biometric recognition is regarded as the final goal of a surveillance system rather

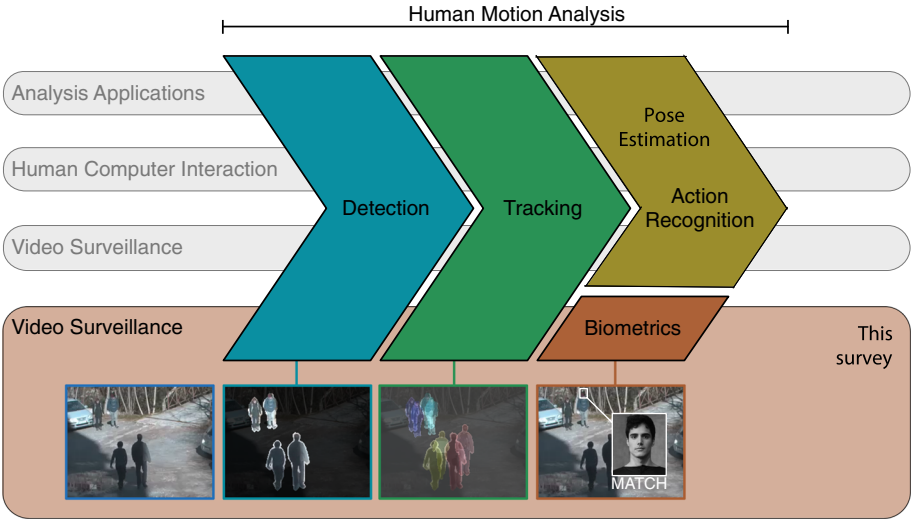


Fig. 1 Previous surveys have specifically covered the developments of human motion analysis for action recognition purposes throughout several different domains (Moeslund et al. 2006; Aggarwal and Ryoo 2011; Weinland et al. 2011; Poppe 2010). On the contrary, this survey is particularly focused on covering the recent advances of human motion analysis in surveillance scenarios for biometric recognition purposes

than as behavior analysis (Pantic et al. 2006; Ko 2008), anomaly detection (Popoola and Wang 2012; Sodemann et al. 2012) or action recognition (Aggarwal and Ryoo 2011; Weinland et al. 2011; Poppe 2010). The novelty of our survey is further justified in Fig. 1, where this paper distinguishes itself from the others with regard to the application (surveillance scenarios) and purpose (biometric recognition).

The remainder of this paper is organized according to the typical phases of a human motion analysis system. Human detection and tracking are reviewed in Sects. 2 and 3, respectively. Section 4 reviews the progress that has been made toward recognizing subjects under non-ideal conditions with respect to the different biometric traits. Section 5 summarizes the major conclusions with regard to the achievements attained in each phase.

2 Detection

Most visual surveillance approaches rely, initially, on locating objects of interest, allowing the removal of unnecessary information and also reducing the processing time of subsequent phases. In visual surveillance scenarios, because movement is a feature that is broadly shared by the objects of interest, temporal information is widely exploited by detection approaches. Indeed, the motion information is commonly used to prune the scene in a pre-detection phase, providing regions of interest to the detection phase. Typically, the pre-detection step relies on background subtraction to highlight the regions of interest, while some alternatives are also possible, such as optical flow. The detection phase attempts to locate humans by searching the scene for a specific model or cue. The taxonomy proposed for this phase is illustrated in Fig. 2.

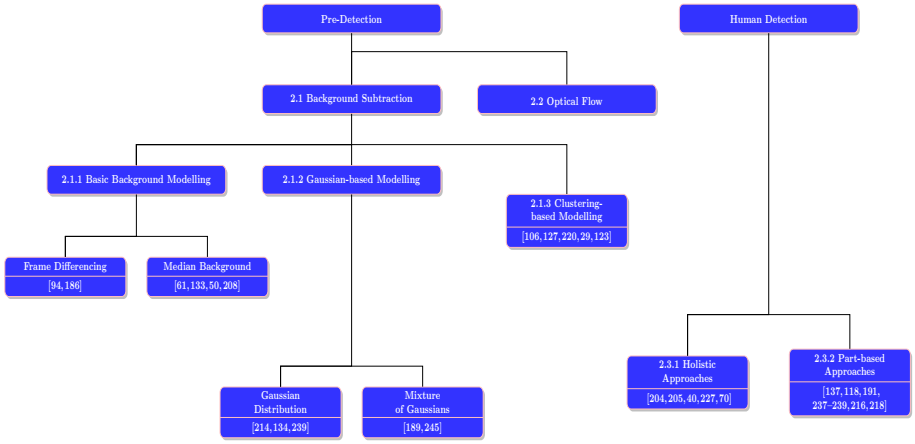


Fig. 2 Proposed taxonomy for the detection phase

2.1 Background estimation

Background Subtraction (BS) methods aim to divide the scene in foreground and background regions using the typical appearance values of static regions of the scene. Although the detection of specific objects is not attained, the scene is pruned and the computational burden of subsequent phases is reduced. For this reason, BS has been used as a pre-detection phase in different approaches such as human detection (Zhao et al. 2008), tracking (KaewTrakulPong and Bowden 2003), pose estimation (Sharma et al. 2011) and activity recognition (Bobick and Davis 2001; Weinland et al. 2006).

Despite BS popularity, this strategy suffers from performance degradation in complex environments, particularly in surveillance scenarios, and, currently, the focus is placed on providing further robustness to the several degradation factors of unconstrained scenarios (Maddalena and Petrosino 2008).

In this survey background modelling methods have been divided according to the strategy used: basic background modelling, Gaussian-based modelling and clustering-based modelling.

2.1.1 Basic background modelling

a) Temporal Differencing This strategy uses temporal differentiation to detect moving regions and is extremely dependent on the assumption of static background. Despite of low-complexity nature, it fails to detect the full object and can not cope with noisy environments.

b) Median Filter This strategy derives a coarse representation of the background from the initial frames (Gloyer et al. 1995) or from the last N frames using simple statistical measures (McFarlane and Schofield 1995). Eng et al. (2003) used the median filter to infer the background in swimming scenarios. In order to perform gait recognition, Wang et al. (2003) extracted the persons silhouette by representing the background with the least median of squares method. Despite being a good compromise between processing speed and perfor-

mance in controlled scenarios, they are not adequate in dynamic environments where they are to prone to produce a large number of false positives.

2.1.2 Gaussian-based modelling

a) Single Gaussian Assuming that intensity values of a pixel are normally distributed, Wren et al. (1997) proposed adjusting a Gaussian distribution to the observed values. Rather than use a threshold, a confidence interval is defined to perform foreground detection, allowing the correct classification of both high and low variance background pixels. Although this strategy provides further robustness to dynamic conditions, such as outdoor environments (McKenna et al. 2000; Zhao et al. 2008), it cannot model multiple sources of background.

b) Mixture of Gaussians To overcome the limitations of single Gaussian modelling, Stauffer and Grimson (1999) proposed describing each pixel as a Mixture of Gaussians (MoG) so that each background component (e.g., buildings and waving trees) could be correctly modelled by a Gaussian distribution. However, the trade-of between the robustness to quick changes and the detection of slow moving objects constitutes its main drawback. The improved MoG (Zivkovic 2004) attempts to address this problem by adaptively adjust the number of Gaussians per pixel.

c) Non-parametric model Originally proposed by Elgammal et al. (2000), this approach uses a model that can handle situations where the background of the scene is cluttered and not completely static but contains small motions. The model estimates the probability of observing pixel intensity values based on a sample of intensity values for each pixel. By sampling, this technique avoids parametric modelling and adapts quickly to changes in the scene enabling very sensitive detection of moving targets.

2.1.3 Clustering-based modelling

Clustering-based approaches estimate the background by grouping pixels in K different clusters, corresponding to multiple sources of background. The Codebook model (Kim et al. 2005) used a set of codewords to represent each cluster, while color and brightness information were used to define the distance function. Different features were used to describe clusters, such as luminance (Butler et al. 2003; Wu et al. 2011b) and chrominance (Butler et al. 2005).

Recently, unsupervised neural networks models have been explored to provide BS methods with further robustness in surveillance scenarios. Self Organizing Maps (SOM) were successfully used by Maddalena and Petrosino (2014). Each pixel was modelled by a SOM and the different background sources were represented by each neuron. Neuron's weights stored the typical RGB values, acting as clusters centroid. Competitive neural networks (Luque et al. 2008) used a similar idea by adjusting the weights of output layer neurons, however, contrary to SOM, learning reinforcement was only applied to the winner neuron.

2.2 Optical flow

Contrary to BS approaches, which compare the scene with a background model to detect moving regions, optical flow approaches rely on displacements between consecutive frames. By assuming small movement and brightness constancy, the displacement of each pixel can be computed. Horn and Schunck (1981) introduced the first technique to address this problem

being followed by many others (Lucas and Kanade 1981). In general, techniques differ in the trade-off between accuracy and speed.

The robustness to moving camera scenarios is the primary reason why optical flow techniques (Talukder and Matthies 2004) are preferred over BS, while their high complexity and inability to cope with changing illumination and fast movements restrain their use in dynamic scenarios.

2.3 Human detection

When compared to the pre-detection phase, detection algorithms are more specific because they aim to provide the location of a specific object in the scene. In general, detection algorithms do not require a pre-detection phase, yet the majority of them rely on this phase to alleviate the computational burden and ease the detection phase. Moreover, in some cases, human detection algorithms do not use pre-detection only as an attentive filter. Instead, they rely on the shape information that is yielded from BSG methods because it has been found that it greatly improves performance when combined with appearance cues (Yao and Odobez 2011).

To achieve human detection, two different strategies are commonly employed: (1) holistic detection, where a whole-body search is conducted; and (2) part-based detection, where the search is oriented to locate a single body part or a combination of parts. Currently, the second approach is attracting more attention, especially in surveillance scenarios, where the head and shoulder regions are commonly used as discriminative features.

2.3.1 Holistic approaches

Most holistic approaches train a discriminative classifier to exhaustively search for a specific object. Viola and Jones (2001) adapted their general object detector to locate humans in surveillance scenarios using motion patterns (Viola et al. 2003). In a similar fashion, Dalal and Triggs (2005) introduced the HOG features to perform human detection by training a discriminative classifier, such as a SVM. HOG features have been explored in several approaches for the purpose of increasing robustness in surveillance scenarios (Moctezuma et al. 2011; Schwartz et al. 2009). LBP features (Ojala et al. 1996) have also been widely used for human detection purposes, especially in surveillance scenarios (Zhang et al. 2007; Wang et al. 2009).

Yao and Odobez (2011) improved the performance of a cascade of detectors by including shape information that was acquired in the pre-detection phase. In the work of Gurwicz et al. (2011), moving objects were obtained with a background estimation method. Several features were extracted, such as image moments and horizontal and vertical projections, but only the features that were capable of the most discrimination were retained, based on the entropy gain. The selected features were provided to a Support Vector Machine (SVM) to distinguish between human regions and clutter in surveillance scenarios.

2.3.2 Part-based approaches

Mikolajczyk et al. (2004) used a probabilistic assembly of parts to attain human detection. A coarse-to-fine cascade approach was used for parts detection, and a parts assembly strategy pruned incorrect detections by imposing geometric constraints.

Lin et al. (2001) focused on head detection to estimate the number of people in a large crowd. Subburaman et al. (2012) also used head features for crowd counting, attaining state-of-the-art results in the PETS2012 dataset.

Zhao and Nevatia (2004), and Zhao et al. (2008) addressed human detection by analyzing the silhouette boundaries that were obtained from the foreground mask. Head detection was attained by checking local vertical peaks on the foreground contour. Detections were filtered by cross-checking silhouette information with human anthropometric data.

Wu and Nevatia (2007b) used four different body parts (full-body, head-shoulder, torso, and legs) to detect humans in non-cooperative scenarios. Parts detectors were learned by boosting a number of weak classifiers based on edgelet features (short segments of edge pixels). The detectors responses were combined to provide robustness to occlusions. Later, this work was extended not only to improve detection performance but also to achieve human segmentation using hierarchical body part detectors (Wu and Nevatia 2009).

2.4 Benchmark data

Several datasets have been proposed to evaluate the performance of human detection methods, such as the Caltech Dataset (Dollar et al. 2012), CAVIAR dataset (Fisher 2005), USC Pedestrian Dataset (Wu and Nevatia 2007a), ETZH dataset (Ess et al. 2007), INRIA Person Dataset (Dalal and Triggs 2005) and PETS databases (PETS 2015).

The PETS databases comprise surveillance data acquired by multiple cameras disposed across a university campus. Several challenges are held as new data deployments occur. In the PETS 2010 challenge, the Probabilistic Occupancy Map algorithm (Fleuret et al. 2008) outperformed the remaining methods. The CAVIAR and ETHZ datasets also contain video sequences acquired in surveillance scenarios. The former contains low resolution videos of a shopping mall, while the latter comprises outdoor sequences captured with a mobile platform. On the other side, INRIA person dataset provides a set of human/non-human cropped images in diverse scenarios, and the USC dataset contains a set of images sampled from the CAVIAR dataset.

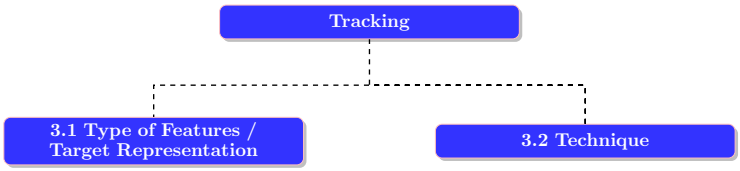
Despite the advantages of multiple datasets (e.g., data from multiple scenarios) they also hamper an objective evaluation of human detection in surveillance scenarios. Contrary to what has been done in pedestrian detection, where the Caltech Dataset was introduced as a unifying framework for evaluation purposes, a reference benchmark is still missing to evaluate human detection in surveillance scenarios.

3 Tracking

Given an initial estimation of object location, visual tracking approaches are expected to determine the correspondences between the same object in consecutive frames. In general, tracking approaches can be distinguished regarding the technique adopted and the type of information used to model target objects, usually denoted as target representation. The proposed taxonomy for tracking is depicted in Fig. 3, where both the most important tracking strategies and target representation have been included.

3.1 Type of features/target representation

Tracking algorithms should be provided with an object description that is usually obtained from distinctive features such as motion, shape or appearance. The model comprising all the information associated with interest object is denoted as the target representation.



Features \ Technique	Bayesian Tracking	Kernel Filter	Model / Shape Tracking	Tracking-by-Detection
Motion	[136,233,173,223, 26,242,238,230,112, 225,97,240,139]			[11,101,178]
Appearance	[136,233,173,223, 230,112,225,149,97, 240]	[39,103,236,183, 52,115,38,234, 119]	[143]	[64,65,11,101,178,157, 7,136,233,97,240,73]
Shape	[88,239]	[234,119]	[60,57,59,143,85, 87]	[7,157,62]

Fig. 3 Classification of tracking techniques according to the technique used and the type of features exploited. Rather than distinct families, *dashed lines* represent the two major attributes of tracking algorithms

3.1.1 Motion

Motion-based tracking exploits object dynamics. In the particular case of human tracking, different cues, such as typical human velocity, articulation constraints and periodic motion are combined to model the target.

As evidenced by Fig. 3, motion models are usually related to Bayesian tracking approaches, where temporal dynamics are used to update the target state over time (Breitenstein et al. 2011; Zhou and Aggarwal 2006; Zhao and Nevatia 2004). However, these models can also be independently used to leverage appearance or shape information (Wu and Huang 2001; Zhou et al. 2003).

Motion information is also widely used to reduce the search space, by assuming small movements between frames. Tracking based on optical flow estimation, namely the KLT tracker (Shi and Tomasi 1994), combines this assumption with brightness constancy, in order to follow a set of keypoints. Tracking-by-detection approaches have also used this strategy. In Babenko et al. (2011) the next location is constrained to a predefined radius. In Santner et al. (2010) the optical flow is exploited to provide further robustness to discriminative classifiers. More complex methods have exploited the motion relations between different regions of the scene to attain additional robustness to occlusions (Grabner et al. 2010).

3.1.2 Appearance

Albeit different tracking techniques can use any kind of appearance descriptor, the literature evidences a relation between the technique and the type of descriptor.

Kernel tracking methods use a histogram of color intensities to represent the target (Comaniciu et al. 2003). Different color spaces, such as HSV and XYZ, were also

used (Maggio and Cavallaro 2005; Stern and Efros 2005; McKenna et al. 1999). McKenna et al. (1999) exploited Gaussian mixture models to parametrize the objects' color distributions in hue-saturation space. An adaptive learning algorithm was used to update these color models and ensure robustness under varying illumination. Since in different scenarios the performance is maximized by different color spaces, (Stern and Efros 2005) developed a method to automatically switch the color space with respect to the environment conditions.

Tracking-by-detection approaches encode appearance information to train discriminative classifiers, using multiple descriptors such as Haar wavelets (Babenko et al. 2011; Santner et al. 2010; Hare et al. 2011), Local Binary Patterns (LBP) (Kalal et al. 2012; Dinh et al. 2011) or Histogram of Oriented Gradients (HOG) (Supancic and Ramanan 2013).

Regarding Bayesian tracking, several approaches have exploited a large number of appearance descriptors (Zhang et al. 2013; Breitenstein et al. 2011; Okuma et al. 2004), but, recently, sparse representation has been widely used by the great majority (Zhong et al. 2012; Zhang et al. 2012; Mei and Ling 2011; Jia et al. 2012). Also, state-of-the-art results have been obtained by combining Bayesian inference with the Extreme Learning Machine (ELM) (Liu et al. 2014) algorithm, whose learning speed can be thousands of times faster than neural networks learning algorithms (Huang et al. 2006).

3.1.3 Shape

Compared to appearance-based tracking, shape modelling is invariant to illumination and appearance changes per se, but in turn, this cue is highly sensitive to occlusion and pose.

Although some tracking methods consider shape as a key feature (Huttenlocher et al. 1993), it is often regarded as a pruning feature or as a way to leverage other cues. This holds particularly in surveillance scenarios, where the limited number of pixels representing the object restrains the use of complex shape models. Notwithstanding, the fusion of simple shape models with other features, such as appearance and motion, proved successful in surveillance scenarios. KaewTrakulPong and Bowden (2003) combined shape cues with position, appearance and motion information to determine the temporal associations between a set of blobs, corresponding to human targets in an outdoor surveillance scenario. Wu and Yu (2006) used a Markov field to learn a prior shape model for human edges. Pedestrian tracking was considered as a posterior density estimation according to the shape model learned, where target state is propagated using a simple motion model.

Albeit edges are the most frequent shape feature used, other alternatives have been currently exploited to track objects in dynamic scenarios [e.g., the shape context descriptor (Belongie et al. 2002; Liu et al. 2012)].

3.2 Technique

Classical approaches attempted to track an object by searching for a specific pattern in the neighborhood of the previous location (Kernel/Model Tracking) or by evolving the state of the target according to a motion and appearance model (Bayesian Tracking). Recently, a new strategy-denoted as tracking-by-detection—has gained popularity as the demand for arbitrary object tracking in unconstrained scenarios increased. The recent developments of each technique are reviewed with particular attention given to the robustness in unconstrained environments.

3.2.1 Bayesian tracking

In a Bayesian framework, tracking is regarded as the estimation of the target state x_k given all the measurements $z_{1:k}$, which is equivalent to maximize the probability $p(x_k|z_{1:k})$. Bayesian filters solve this recursively using two steps: (1) *prediction* step infers the next state distribution, $p(x_k|z_{1:k-1})$, with respect to a motion model describing the target state over time; (2) *update* step uses the current observation z_k to update $p(x_k|z_{1:k-1})$, yielding $p(x_k|z_k)$. This process allows the estimation of the latent or unobservable variable x_k through noisy measurements z_k. Regarding the type of noise, different Bayesian filters can be used.

When the system is affected by Gaussian noise and the motion model is linear, the Kalman filter (Kalman 1960) can be employed. Despite being based on restrictive assumptions, some approaches used it in surveillance scenarios (Szeto and Gazis 1972; Zhao and Nevatia 2004; Zhou and Aggarwal 2006). Zhao and Nevatia (2004) used the Kalman filter with a constant velocity model to estimate the state of humans. In Zhou and Aggarwal (2006) a multi-camera approach was exploited, where the combined observations of each camera were provided to the Kalman filter to obtain a more accurate target state.

The Extended Kalman Filter (EKF) (Julier and Uhlmann 2004) was introduced to handle non-linear systems. Mittal and Davis (2003) used this technique in a multi-view approach so that severe occlusion could be handled. Oliver et al. (2000) combined the EKF predictions with appearance information to track persons in outdoor scenes for action-recognition purposes.

In general, particle filters or sequential Monte Carlo methods are preferred in Bayesian tracking (Ross et al. 2008; Wu et al. 2011a; Zhang et al. 2012; Kwon and Lee 2010; Xiao et al. 2013), since they can handle any kind of noise and do not require the motion model to be linear. Okuma et al. (2004) used appearance cues by combining the particle filter with AdaBoost. Hu et al. (2009) combined appearance, shape and motion information to track occluded people also using the particle filter. Sparse representation was also exploited by some state-of-the-art tracking methods (Mei and Ling 2011; Zhang et al. 2012; Jia et al. 2012; Zhong et al. 2012). Each candidate location was represented as a combination of the training templates so that the smallest projection error candidate was chosen. Mei and Ling (2011) used this strategy in the L1 tracker. The target motion in consecutive frames was modelled as an affine transformation and was estimated in a particle filter framework. The importance of each transformation (i.e., the particle weights) was a function of the sparse reconstruction error. The MTT tracker Zhang et al. (2012) was later introduced as a generalization of L1 since it accounted for the dependences between transformations.

3.2.2 Kernel filter

Kernel-based tracking gathers appearance information over an image patch by constructing a weighted feature histogram. The first representative kernel-based method was proposed by Comaniciu et al. (2003), where the Mean Shift (Cheng 1995) technique was adapted to track objects based on their appearance. Target location was achieved by maximizing a similarity measure and the mean shift procedure guided the search for conditional probability maximum, avoiding a brute force search.

Although this strategy provides invariance to some pose changes, the loss of spatial information is the primary drawback of kernel-based approaches. To address this issue, Kang et al. (2003) divided the object according to its polar representation and modelled the typical RGB color of each part with a Gaussian distribution. Zhao and Tao (2009) included spatial information in the appearance model using the Correlogram technique (Huang et al. 1998),

allowing to infer not only the objects trajectory but also their orientation. Recently, distribution fields (Sevilla-Lara 2012; Felsberg 2013) have also been introduced to preserve the spatial information by constructing a histogram at each pixel.

Robustness to dynamic environments has also been recently proposed. Chu et al. (2013) used multiple kernels to improve tracking under occlusion. Zhang et al. (2013) devised a head tracker using a kernel-Bayesian framework, where appearance and shape information were combined. Mixture of Gaussians were used to model the appearance and the Chamfer distance Barrow et al. (1977) was used for shape comparison. Liu et al. (2011) approached human tracking using eigenshape. The arbitrarily shaped kernel allowed the tracker to adapt to the object shape avoiding background noise.

3.2.3 Model/shape tracking

Maximizing the similarity between the shape model and the contour-map of the image is the rationale of shape tracking. In general, contour information is provided by an edge-map representation and shape similarity is evaluated either with the Chamfer matching (Barrow et al. 1977) or with the Haussdorf distance (Huttenlocher et al. 1993). Both shape matching techniques are computationally expensive and not suitable to work in real time systems.

To efficiently compute the Chamfer matching or the Haussdorf distance, (Gavrila and Philomin 1999; Gavrila 1998) proposed a solution based on the distance transform. In a later work (Gavrila 2007), hierarchical matching was proposed to further increase the efficiency of shape matching. A set of shapes from an object, previously captured from the training data, were clustered so that a tree of shape models could be constructed with the representative model of each cluster in the first layer. Besides, a Markov transition matrix was used to encode the probabilities between shape transitions, so that, during the tracking, the most likely poses are prioritized. These approaches were combined in Munder et al. (2008) to develop a complete pedestrian detection and tracking system, where motion and appearance cues are also exploited. The tracking module used pose clusters and a tree of pose models to efficiently search for the model that best fitted the data.

In dynamic environments, shape tracking is particularly sensitive to occlusion. For this reason, Saber et al. (2005) devised a matching strategy robust to partial occlusion, the partial shape matching. Husain et al. (2006) used this technique to track objects in surveillance scenarios.

However, even these improvements fail to produce a robust solution in surveillance scenarios, mainly due to the reduced size of interest objects.

3.2.4 Tracking-by-detection

The use of detectors in tracking has gained wide notoriety, mainly driven by the possibility of tracking arbitrary objects. Tracking-by-detection algorithms estimate the target position by searching the location that maximizes a function $F(\mathbf{x}) \in [-1, 1]$, where F is usually determined by a classifier and $\mathbf{x}$ is the feature vector of the target state. Contrary to other tracking methods, no *a priori* target representation is required, postponing the learning of this representation to the online training of the classifier. Online training allows the classifier to adapt to any kind of object and also to appearance variations. Currently, the main research line in tracking-by-detection is focused both in improving the classifier learning scheme and in exploiting multiple cues.

Regarding the learning scheme, the use of online boosting classifiers was a common strategy in initial approaches (Grabner et al. 2006, 2008). At each frame, the target location was

sampled for positive examples while its neighborhood was sampled for negative examples. However, this strategy is highly sensitive to appearance changes, since small displacements from the ground truth location may introduce incorrect positive examples in the learning process.

Babenko et al. (2011) exploited Multiple Instance Learning (MIL) to overcome this problem, where examples were presented as bags containing a set of instances. Bags containing at least one positive instance, corresponding to the instances sampled at the target location, were labelled as positive, otherwise they were labelled as negative. Although this strategy required the classifier to distinguish between positive and negative instances in some bags, previous results had shown that, in fact, it was more flexible and outperformed the traditional learning strategies (Viola et al. 2005). In a similar fashion, the Struck tracker (Hare et al. 2011) used a structured output SVM (Tsochantaridis et al. 2005) to perform learning.

In the TLD (Kalal et al. 2012) and the PROST (Santner et al. 2010) methods a different solution combined an optic flow tracker with an online learned random forest. Negative examples were only sampled from unlikely locations of object presence based on motion constraints. Besides, new examples required an appearance confirmation to be provided to the classifier. ConTra (Dinh et al. 2011) improved this strategy by taking in account distracters, i.e., objects sharing the same appearance as the target.

3.3 Multi-target tracking

Despite multiple instances of each algorithm could be used to address multiple target tracking, these methods would require an additional data association module. The Joint Probabilistic Data Association Filter (Fortmann et al. 1983) and Multiple Hypothesis Tracking (Reid 1979) are two classical approaches for this purpose, however the exponential growth of computational complexity restrains their use when the number of targets is high. Greedy strategies have been used as an alternative, where correspondences are regarded as an assignment problem based on spatial distance (Wu and Nevatia 2007b; Cai et al. 2006) or appearance similarity (Breitenstein et al. 2009).

Offline or batch techniques methods are an alternative solution for multiple target tracking, which, in contrast to online methods, use the complete set of detections before perform trajectory estimation. This problem is usually regarded as an optimization problem, where a function describes the cost of each solution (Leibe et al. 2007; Zhang et al. 2008; Andriyenko and Schindler 2010). Linear programming was employed by several works (Jiang et al. 2007; Berclaz et al. 2009, 2011; Andriyenko and Schindler 2010) to solve this problem, where the possible target locations were discretized and modelled as graph. A continuous formulation of the problem was later introduced by Andriyenko and Schindler (2011), Milan et al. (2014), Andriyenko et al. (2012). The main drawback of these approaches is the high latency required to analyse a video, which is incompatible with real-time surveillance requirements. To address this issue, Benfold and Reid suggested the use of a small subset of frames. In Benfold and Reid (2011) the most recent six seconds of video were analysed to track multiple pedestrians by combining information from a HOG-based detector and a KLT tracker.

3.4 Benchmark data

A multitude of tracking datasets has been proposed to cover specific scenarios. While general tracking approaches are usually tested against a collection of videos with a wide variety of environments (Babenko et al. 2011; Kalal et al. 2012), surveillance oriented approaches are

typically evaluated in specific datasets such as the CAVIAR dataset (Fisher 2005), the I-lids datasets (Maggio et al. 2007) and the PETS databases.

The VOT challenges (Kristan et al. 2013; Vot 2015) represented a joint effort to establish a benchmark dataset for tracking evaluation purposes. The performance of state-of-the-art methods was compared and the results were presented in Kristan et al. (2013). Although none of the trackers has stood out globally, the results provide insight about the best strategies with respect to the environment specificities.

Simultaneously, Wu et al. (2013) also introduced a useful tool for tracking benchmarking comprising an evaluation kit of several state-of-the-art tracking methods. Moreover, a dataset was introduced along with the algorithms performance in these data.

4 Recognition

In a typical human motion analysis system, recognition is regarded as the ultimate goal to which every preceding phase should contribute by providing pre-processed information. In general, recognition aims at finding a correspondence between the observed data and a gallery of exemplars, which can be actions, activities or biometric traits. As previously discussed in Sect. 1, this survey is especially focused on biometric recognition, and thus, the recognition of human activities is not covered in this section. The reader is referred to Aggarwal and Ryoo (2011) for a detailed review on action recognition.

4.1 Biometric recognition

Biometric recognition refers to the use of human traits, either physical or behavioral, to perform identification of individual people. Several distinct traits have been exploited in the literature, such as fingerprint (Bolle and Pankanti 1998), face (Turk and Pentland 1991), iris (Daugman 1993), hand geometry (Sanchez-Reillo et al. 2000) and voice (Squires and Sammut 1995). To be considered to be a valid biometric trait, four main requirements must be fulfilled: (1) universality—should be shared by every human; (2) distinctiveness—no similar instances should exist; (3) permanence—should be invariant to time; and (4) collectability—should be easy to collect. Although some traits ensure that all of these requirements are met and attain high accuracy levels (e.g., fingerprints), in this survey, focus is placed on traits that can be recognized at a distance.

4.1.1 Iris

Compared to other biometric traits, such as face and gait, iris is one of the most discriminative traits for identification purposes (Proença 2007). Daugman (1993) introduced a pioneering approach for iris recognition in which Gabor filters were used to encode iris patterns. Daugman showed that the distinctiveness of a 256-byte iris code could afford 1 error in approximately 10^{31}. Another classical iris segmentation method was presented by Wildes (1997), where the Hough transform was applied to the image edge map instead.

Nevertheless, the performance of these approaches is highly dependent on the data quality and consequently on the subjects cooperation during the acquisition process. To achieve robust iris segmentation in unconstrained scenarios, Proença and Alexandre (2007) proposed an iris recognition system that was capable of addressing noisy data. This work used multiple signatures by dividing the iris into six independent regions, in such a way that the corruption of the whole signature by localized noisy regions could be avoided.

Even though this approach attained good results in noisy images (e.g., UBIRISv1), surveillance systems cannot rely on this trait in wide open scenarios. Daugman (2004) stated that a minimum of 70 pixels in the iris radius is required to capture the rich details of the iris patterns. A recent work of Tan and Kumar (2012) attempted iris recognition at a distance; however, high-resolution facial images were used. Moreover, the authors stated that despite the superior performance that was attained, further improvements are required to address surveillance scenarios. Boddeti et al. (2011) addressed more challenging images with approximately 50 pixels in the iris diameter that were captured, in the context of Iris-On-The-Move system (Matey et al. 2006; Phillips 2014b). However, under such conditions, the performance was greatly reduced (the error rate was approximately 30 %).

4.1.2 Periocular

Considering the drawbacks of iris recognition at a distance and in surveillance scenarios, Park et al. (2009) suggested that the facial region in the vicinity of the human eye—the periocular region could be used as a discriminant biometric trait between individuals. A set of local descriptors [LBP, HOG and Scale-Invariant Feature Transform (SIFT)] were used to extract features from the periocular region. Later, the authors evaluated the role of periocular components in recognition performance (Park et al. 2011), such as the eyebrows, eyes and iris.

Lyle et al. (2010, 2012) used the periocular region to perform gender and ethnicity classification and reported similar performance to the performance attained using the facial region. Moreover, a comparative study between the iris and ocular region concluded that the latter attains higher recognition performance in unconstrained scenarios (Boddeti et al. 2011).

These results fostered the use of the periocular region in unconstrained scenarios (Santos and Proença 2013) and drove the development of algorithms that were robust to noisy data. Padole and Proença (2012) analyzed the role of different degradation factors in the periocular recognition. Tan and Kumar (2013) attempted to perform biometric identification at a distance in unconstrained images of periocular and facial regions. The authors exploited a joint iris and periocular strategy to improve recognition accuracy.

4.1.3 Face

The search for algorithms that are capable of recognizing humans using the facial region has occurred over more than 50 years. The first attempt dates back to 1964, when Bledsoe (1964) developed a facial recognition system that was based on a set of 20 distances measured from facial keypoints. During his experiments, Bledsoe stressed that the "great variability in head rotation and tilt, lighting intensity and angle, facial expression and aging" make face recognition an extremely difficult challenge. To date, these variability factors remain the primary focus of face recognition research studies.

Turk and Pentland (1991) introduced the notion of eigenfaces to represent facial features in a low-dimensional space. Recognition was attained by projecting the new image, which is considered to be a point in N-dimensional space, in the face space and determining the nearest neighbor. Although the eigenfaces method is regarded as one of the first facial recognition technologies, robustness to degradation factors, such as lighting and pose, is barely attained. Later, Belhumeur et al. (1997) improved this idea by using LDA instead of PCA to represent the facial features.

To address the pose variation, Blanz and Vetter (2003) introduced morphable models. Still images, captured at different poses, were used to build a 3D face model that contained shape

and texture information. The model was used to infer synthetic images under varying poses, with a view to enlarging the training set with representative images of all possible variations.

The use of LBP (Ahonen et al. 2006) to encode facial features has made a significant contribution toward increasing facial recognition performance in non-ideal scenarios. This strategy attained state-of-the-art results not only in frontal faces but also in faces that were subjected to varying illumination and expression. Again, several studies used this idea to provide further robustness to unconstrained face recognition. Li et al. (2007) developed an illumination-invariant face recognition system by combining near-infrared imaging with an LBP-based face description. Tan and Triggs (2010) extended the LBP to LTP to address difficult lighting conditions. Recent methods (Chan et al. 2013; Heikkila et al. 2014) have found the LPQ descriptor (Ojansivu et al. 2008) to be more robust than LBP to specific degradation factors, such as blur.

Occlusions are another typical degradation factor of face recognition systems, and this factor has been addressed in several studies (Martinez 2002). Nevertheless, robustness to occlusion was attained only when sparse representation techniques were introduced in facial recognition (Wright et al. 2009). These results were subsequently improved and the processing time decreased by combining sparse coding with the ELM algorithm (He et al. 2014).

The advances in face recognition performance in less constrained conditions have paved the way for face recognition in real-world scenarios, whose popularity has exponentially rose with the introduction of LFW database (Huang et al. 2007). The particularities of this set, such as the large variability in expression, pose, illumination and the objective evaluation protocol, established it as the reference benchmark for unconstrained face recognition and fostered the development of approaches robust to non-cooperative scenarios (Li et al. 2013; Schroff et al. 2015; Zhu et al. 2015).

Nonetheless, one explanation for unconstrained face recognition being still far from solved is that the LFW and similar datasets are not fully unconstrained. In fact, most sets comprise manually captured data, and thus they do not provide a faithful representation of biometric traits acquired by fully automated surveillance systems.

Facial recognition in surveillance scenarios is mainly plagued by the reduced resolution of the data. To overcome this problem, the use of PTZ cameras has been increasing (Cai et al. 2013; Xu and Song 2010; Yao et al. 2008; Senior et al. 2005; Wheeler et al. 2010). The mechanical properties of these devices allow the acquisition of high-resolution images of arbitrary scene locations. Park et al. (2013) presented a PTZ-based system that is capable of acquiring high-resolution face images at a distance of 15 m. In spite of the authors had reported encouraging results (91 % rank-1 identification) for the recognition accuracy of this system, the prototype can be barely used in outdoor scenarios due to the restrictive configurations between the cameras. To address this problem, Neves et al. (2015) have recently introduced an innovative PTZ-based surveillance system that is flexible enough to be deployed in any surveillance scenario while maintaining an accurate mapping between cameras.

4.1.4 Gait recognition

In spite of the recent developments, facial biometrics performance decreases significantly when using low-resolution images. This fact motivated the search for non-invasive biometric traits that can be identified at a distance. As such, special attention has been given to the walking pattern of humans, the gait, which has been found to be very discriminative (Murray 1967). Although gait distinctiveness cannot compare with hard biometrics (Jain et al. 2004), it has proven to be a good compromise in surveillance scenarios (Jean et al. 2009).

Gait recognition can be coarsely divided into two distinct strategies: (1) model-based approaches (Lee and Grimson 2002; Gu et al. 2010), which recover the human structure to provide information about the walking dynamics; and (2) model-free approaches (Han and Bhanu 2006; Chen et al. 2009; Iwama et al. 2012), which directly analyze motion features from image sequences. Despite being more accurate, model-based methods are computationally expensive and sensitive to appearance and occlusion issues, and thus, they are not adequate to handle surveillance scenarios. The gait energy image (GEI) (Han and Bhanu 2006) is a model-free strategy that is commonly used in several studies (Iwama et al. 2012; Okumura et al. 2010; Bashir et al. 2010).

Gait recognition in surveillance scenarios has been progressively addressed by the development of speed-invariant (Priydarshi et al. 2013), cloth-invariant (Hossain et al. 2010), view-invariant (Jean et al. 2009; Goffredo et al. 2010) methods in low-resolution (Zhang et al. 2010) data.

4.1.5 Soft biometrics

Soft biometric traits differ from typical biometric traits, which are usually denoted by hard biometrics, in distinctiveness and permanence, i.e., they cannot be used to uniquely identify a person, but they can provide informative cues to describing an individual. In contrast to hard biometrics, these traits are not as dependent on subject cooperation and can be acquired from low-resolution and poor quality data, which makes them especially suitable to surveillance scenarios. Gender, ethnicity, hair, height and weight are some examples of soft biometric traits.

Considering the lack of distinctiveness of soft biometrics, they were originally proposed as complementary traits in biometric recognition systems (Jain et al. 2004). In Jain et al. (2004), the authors presented a methodology for incorporating soft biometric information in a fingerprint recognition system at the decision level. This idea was further exploited in Ailisto et al. (2006). In Denman et al. (2009), the feasibility of recognition solely based on soft traits was evaluated in surveillance scenarios using the PETS 2006 database. Although reliable authentication could not be afforded, the authors described the results as encouraging with respect to eventually providing coarse authentication at a distance. The use of soft biometrics as a single biometric trait, rather than as an ancillary trait, was introduced by Dantcheva et al. (2011), using a bag of facial soft biometrics.

Tracking, re-identification (Vezzani et al. 2013) and semantic classification of surveillance videos are examples of other typical applications of soft biometrics. With regard to the last topic, recent approaches have focused on learning relations between gait and soft biometrics to automatically perform annotation or content-based retrieval of surveillance videos (Samangooei and Nixon 2008, 2010). This work was extended in Reid and Nixon (2010), Reid et al. (2014) by exploiting imputation techniques, which comprise statistical methods for inferring missing data. Unavailable soft biometric traits, due to occlusion or other factors, were extrapolated from the available traits based on correlations between them. The same authors also presented a strategy to avoid traits subjectivity by labeling each subject according to an annotated database.

4.2 Benchmark data

Considering the multitude of visual biometric traits, a wide number of datasets exists to cover the demand for evaluation data in distinct scenarios. Throughout the years the focus has been put on providing data acquired in more unconstrained and challenging scenarios in order

Table 2 Summary of the main biometric datasets and the average resolution of the different traits

Database	Content	Iris	Eye	Periocular	Face
CASIAv3 CASIA (2014)	Near-infrared iris images	189 × 221	253 × 515	NA	NA
CASIAv4 CASIA (2014)	Iris images captured at-a-distance	151 × 160	198 × 335	312 × 1015	2352 × 1728
FRGC Phillips et al. (2005)	Low resolution face images	20 × 22	33 × 67	77 × 220	329 × 243
FOCS Phillips (2014b)	Iris images captured from the Iris-On-The-Move system	150 × 159	175 × 345	NA	NA
UBIRISv2 Proença et al. (2010)	Ocular region images with iris subjected to several noise factors	125 × 132	180 × 315	NA	NA
PUT Kasinski et al. (2008)	Pose-varying faces containing facial contours and facial landmarks	57 × 62	115 × 195	230 × 700	960 × 880
SCface Grgic et al. (2011)	Face images captured in indoor surveillance scenarios	I	I	I	67 × 86
FERET Phillips (2014a)	Large database of facial images with some variability in age, illumination and expression	19 × 21	28 × 60	82 × 213	375 × 290
MULTI-PIE Gross et al. (2010)	Facial images across 13 different poses and 4 different expressions with different illumination conditions	11 × 12	21 × 38	48 × 140	260 × 200
LFW Huang et al. (2007)	Faces acquired in unconstrained scenarios	I	I	37 × 90	135 × 102
IJB-A Klare et al. (2015)	Faces acquired in unconstrained scenarios with high-variability in pose	I	I	90 × 105	438 × 295
QUIS-CAMPI Neves (2015)	The first dataset of biometric samples automatically acquired by an outdoor surveillance system, with subjects on-the-move and at-a-distance	I	I	70 × 178	276 × 203

I denotes insignificant resolution and *NA* denotes not available

to promote the development of biometric recognition methods robust to non-cooperative scenarios and also to surveillance environments.

Table 2 summarizes the main datasets of facial biometric data. Regarding soft biometric traits, the TunnelDBSoftBio (Tome et al. 2014) comprising 23 physical traits from 58 users is considered a reference.

5 Conclusions

The interest in the automated visual surveillance of human beings has significantly increased and is strongly driven by security concerns. Although no fully automated surveillance system for biometric recognition purposes exists, recent developments in human motion analysis and biometric recognition can contribute to the development of such a system. The typical phases of human motion analysis—detection, tracking and recognition—were covered in this survey, with a special focus placed on the effort to address surveillance scenarios.

In the pre-detection phase, it is important to highlight the advances of background subtraction algorithms performance in surveillance scenarios, which are the direct result of the development of datasets particularly focused on surveillance scenarios (Brutzer et al. 2011).

With regard to human detection, a large number of studies have focused on surveillance scenarios that provide both fast and accurate solutions for real-time systems. The work of Yao and Odobez (2011) can process 20 frames/s in a 384x288 video, and their system attains accurate results on the CAVIAR and PETS datasets. Robustness to occlusions was also successfully achieved by part-based approaches (Wu and Nevatia 2009), which showed promising results in surveillance videos. On the opposite side, it is important to highlight the lack of a reference benchmark for human detection in surveillance scenarios.

In the tracking phase, the rise of tracking-by-detection approaches allowed the development of methods that are capable of tracking arbitrary objects under dynamic conditions. Furthermore, the use of offline multi-target tracking techniques is also especially interesting for addressing surveillance scenarios, even though some delay is always associated (Benfold and Reid 2011). Regarding benchmark evaluations, PETS stands out as the reference dataset for accessing performance in surveillance scenarios, particularly for multi-tracking approaches. However, no effort has been done yet for benchmarking the performance of multi-tracking algorithms in these scenarios.

With regard to the recognition phase, biometric identification was the main focus of this survey, in contrast to the large number of surveys of human motion analysis (Moeslund et al. 2006; Aggarwal and Ryoo 2011; Weinland et al. 2011; Poppe 2010). In the recent years, different approaches were introduced to address the typical challenges of unconstrained biometric recognition. The almost ideal performance reported on unconstrained datasets by these approaches contrasts with the fact that biometric recognition in surveillance scenarios is far from being solved (Klontz and Jain 2013). This suggests that state-of-the-art datasets are not fully unconstrained, since most of them comprise manually captured data and do not provide a faithful representation of biometric traits acquired in surveillance scenarios. This fact constitutes a chief limitation in the development of fully automated human recognition systems.

The use of PTZ cameras (Chen et al. 2013; Neves et al. 2015) might be the missing piece of the surveillance/biometrics jigsaw puzzle because they can enable the acquisition of high-resolution biometric data at a distance. Besides, it important to highlight that the development of biometric datasets automatically acquired by PTZ-based systems is already

in progress (Neves 2015), and these are definitely the best tools to correctly assess how far research has come in biometric recognition in the wild.

On the other hand, the use of soft biometrics and gait in surveillance scenarios has shown encouraging results, which suggests that they could be used in a multi-modal recognition system with other hard biometric traits.

References

Aggarwal J, Cai Q, Liao W, Sabata B (1998) Nonrigid motion analysis: articulated and elastic motion. Comput Vis Image Underst 70(2):142–156

Aggarwal J, Cai Q (1999) Human motion analysis: a review. Comput Vis Image Underst 73(3):428–440

Aggarwal J, Ryoo M (2011) Human activity analysis: a review. ACM Comput Surv 43(3):16:1–16:43

Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. IEEE Trans Pattern Anal Mach Intell 28(12):2037–2041

Ailisto H, Vildjiounaite E, Lindholm M, Mkel SM, Peltola J (2006) Soft biometrics combining body weight and fat measurements with fingerprint biometrics. Pattern Recogn Lett 27(5):325–334

Andriyenko A, Schindler K (2010) Globally optimal multi-target tracking on a hexagonal lattice. In: Proceedings of the 11th European conference on computer vision: part I. pp 466–479

Andriyenko A, Schindler K (2011) Multi-target tracking by continuous energy minimization. In: IEEE conference on computer vision and pattern recognition. pp 1265–1272

Andriyenko A, Schindler K, Roth S (2012) Discrete-continuous optimization for multi-target tracking. In: IEEE conference on computer vision and pattern recognition. pp 1926–1933

Babenko B, Yang MH, Belongie S (2011) Robust object tracking with online multiple instance learning. IEEE Trans Pattern Anal Mach Intell 33(8):1619–1632

Barnich O, Van Droogenbroeck M (2011) Vibe: a universal background subtraction algorithm for video sequences. IEEE Trans Image Process 20(6):1709–1724

Barrow HG, Tenenbaum JM, Bolles RC, Wolf HC (1977) Parametric correspondence and chamfer matching: two new techniques for image matching. In: Proceedings of the 5th international joint conference on artificial intelligencem, IJCAI'77, vol. 2. Morgan Kaufmann Publishers Inc., San Francisco, pp 659–663

Bashir K, Xiang T, Gong S (2010) Gait recognition without subject cooperation. Pattern Recogn Lett 31(13):2052–2060

Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720

Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(4):509–522

Benfold B, Reid I (2011) Stable multi-target tracking in real-time surveillance video. In: Proceedings of the 2011 IEEE conference on computer vision and pattern recognition, CVPR '11. IEEE Computer Society, Washington, DC. pp 3457–3464

Berclaz J, Fleuret F, Turetken E, Fua P (2011) Multiple object tracking using k-shortest paths optimization. IEEE Trans Pattern Anal Mach Intell 33(9):1806–1819

Berclaz J, Fleuret F, Fua P (2009) Multiple object tracking using flow linear programming. In: Twelfth IEEE international workshop on performance evaluation of tracking and surveillance (PETS-Winter). pp 1–8

Blanz V, Vetter T (2003) Face recognition based on fitting a 3d morphable model. IEEE Trans Pattern Anal Mach Intell 25(9):1063–1074

Bledsoe WW (1964) The model method in facial recognition. Tech. Rep. PRI 15. Panoramic Research, Inc., Palo Alto

Bobick A, Davis J (2001) The recognition of human movement using temporal templates. IEEE Trans Pattern Anal Mach Intell 23(3):257–267

Boddeti V, Smereka J, Kumar B (2011) A comparative evaluation of iris and ocular recognition methods on challenging ocular images. In: International joint conference on biometrics. pp 1–8

Bolle R, Pankanti S (1998) Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society. Kluwer Academic Publishers, Norwell, MA, USA

Breitenstein M, Reichlin F, Leibe B, Koller-Meier E, Van Gool L (2011) Online multiperson tracking-by-detection from a single, uncalibrated camera. IEEE Trans Pattern Anal Mach Intell 33(9):1820–1833

Breitenstein M, Reichlin F, Leibe B, Koller-Meier E, Van Gool L (2009) Robust tracking-by-detection using a detector confidence particle filter. In: International conference on computer vision. pp 1515–1522

Brutzer S, Hoferlin B, Heidemann G (2011) Evaluation of background subtraction techniques for video surveillance. In: IEEE conference on computer vision and pattern recognition (CVPR). pp 1937–1944
Butler DE, Bove VM, Sridharan S (2005) Real-time adaptive foreground/background segmentation. EURASIP J Adv Signal Process 2005(14):841,926
Butler D, Sridharan S, Bove VMJ (2003) Real-time adaptive background segmentation. In: Proceedings of 2003 international conference on Multimedia and Expo, 2003. ICME '03, vol. 3. pp III-341–III-344
Cai Y, de Freitas N, Little JJ (2006) Robust visual tracking for multiple targets. In: ECCV. pp 107–118
Cai Y, Medioni G, Dinh T (2013) Towards a practical PTZ face detection and tracking systems. In: Proceedings of the IEEE workshop on applications of computer vision. pp 31–38
CASIA: Casia iris image databases (2014). http://www.idealtest.org/findTotalDbByMode.do?mode=Iris
Cédras C, Shah M (1995) Motion-based recognition a survey. Image Vis Comput 13(2):129–155
Chan CH, Tahir M, Kittler J, Pietikainen M (2013) Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. IEEE Trans Pattern Anal Mach Intell 35(5):1164–1177
Chen C, Liang J, Zhao H, Hu H, Tian J (2009) Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recogn Lett 30(11):977–984
Chen CH, Yao Y, Chang H, Koschan A, Abidi M (2013) Integration of multispectral face recognition and multi-ptz camera automated surveillance for security applications. Cent Eur J Eng 3(2):253–266
Cheng Y (1995) Mean shift, mode seeking, and clustering. IEEE Trans Pattern Anal Mach Intell 17(8):790–799
Chu CT, Hwang JN, Pai HI, Lan KM (2013) Tracking human under occlusion based on adaptive multiple kernels with projected gradients. IEEE Trans Multimed 15(7):1602–1615
Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell 25(5):564–577
Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, vol. 1. pp. 886–893
Dantcheva A, Velardo C, DAngelo A, Dugelay JL (2011) Bag of soft biometrics for person identification. Multimed Tools Appl 51(2):739–777
Daugman J (1993) High confidence visual recognition of persons by a test of statistical independence. IEEE Trans Pattern Anal Mach Intell 15(11):1148–1161
Daugman J (2004) How iris recognition works. IEEE Trans Circuits Syst Video Technol 14(1):21–30
Davies A, Velastin S (2005) A progress review of intelligent cctv surveillance systems. In: IEEE intelligent data acquisition and advanced computing systems: technology and applications. pp 417–423
Denman S, Fookes C, Bialkowski A, Sridharan S (2009) Soft-biometrics: unconstrained authentication in a surveillance environment. In: Digital image computing: techniques and applications, 2009, DICTA '09. pp 196–203
Dinh TB, Vo N, Medioni G (2011) Context tracker: exploring supporters and distracters in unconstrained environments. In: IEEE conference on computer vision and pattern recognition. pp 1177–1184
Dollar P, Wojek C, Schiele B, Perona P (2012) Pedestrian detection: an evaluation of the state of the art. IEEE Trans Pattern Anal Mach Intell 34(4):743–761
EarthCam (2014) Times square cams. http://www.earthcam.com/usa/newyork/timessquare/?cam=tsstreet
Elgammal A, Harwood D, Davis L (2000) Non-parametric model for background subtraction. In: ECCV. pp 751–767
Eng HL, Toh KA, Kam AH, Wang J, Yau WY (2003) An automatic drowning detection surveillance system for challenging outdoor pool environments. IEEE Int Conf Comput Vis 1:532
Ess A, Leibe B, Van Gool L (2007) Depth and appearance for mobile scene analysis. In: IEEE 11th international conference on computer vision. pp 1–8
Felsberg M (2013) Enhanced distribution field tracking using channel representations. In: International conference on computer vision workshops. pp 121–128
Feris R, Datta A, Pankanti S, Sun MT (2013) Boosting object detection performance in crowded surveillance videos. In: IEEE workshop on applications of computer vision. pp 427–432
Fisher R (2005) Caviar dataset
Fleuret F, Berclaz J, Lengagne R, Fua P (2008) Multicamera people tracking with a probabilistic occupancy map. IEEE Trans Pattern Anal Mach Intell 30(2):267–282
Fortmann TE, Bar-Shalom Y, Scheffe M (1983) Sonar tracking of multiple targets using joint probabilistic data association. IEEE J Ocean Eng 8(3):173–184
Gavrila D (1998) Multi-feature hierarchical template matching using distance transforms. In: Fourteenth international conference on pattern recognition, vol. 1. pp 439–444
Gavrila D (1999) The visual analysis of human movement: a survey. Comput Vis Image Underst 73(1):82–98
Gavrila D (2007) A bayesian, exemplar-based approach to hierarchical shape matching. IEEE Trans Pattern Anal Mach Intell 29(8):1408–1421

Gavrila D, Philomin V (1999) Real-time object detection for smart vehicles. In: International conference on computer vision, vol. 1. pp 87–93
Gloyer B, Aghajan HK, Siu KY, Kailath T (1995) Video-based freeway-monitoring system using recursive vehicle tracking. pp 173–180
Goffredo M, Bouchrika I, Carter J, Nixon M (2010) Performance analysis for automated gait extraction and recognition in multi-camera surveillance. Multimed Tools Appl 50(1):75–94
Grabner H, Grabner M, Bischof H (2006) Real-time tracking via on-line boosting. In: Proceedings of BMVC. pp 6.1–6.10
Grabner H, Leistner C, Bischof H (2008) Semi-supervised on-line boosting for robust tracking. In: Proceedings of the 10th European conference on computer vision: part I. pp 234–247
Grabner H, Matas J, Van Gool L, Cattin P (2010) Tracking the invisible: learning where the object might be. In: IEEE conference on computer vision and pattern recognition. pp 1285–1292
Grgic M, Delac K, Grgic S (2011) Scface surveillance cameras face database. Multimed Tools Appl 51(3):863–879
Gross R, Matthews I, Cohn J, Kanade T, Baker S (2010) Multi-pie. Best of automatic face and gesture recognition 2008. Image Vis Comput 28(5):807–813
Gu J, Ding X, Wang S, Wu Y (2010) Action and gait recognition from recovered 3-d human joints. IEEE Trans Syst Man Cybern Part B Cybern 40(4):1021–1033
Gurwicz Y, Yehezkel R, Lachover B (2011) Multiclass object classification for real-time video surveillance systems. Patt Recogn Lett 32(6):805–815
Haering N, Venetianer P, Lipton A (2008) The evolution of video surveillance: an overview. Mach Vis Appl 19(5–6):279–290
Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(2):316–322
Hare S, Saffari A, Torr PHS (2011) Struck: structured output tracking with kernels. In: IEEE international conference on computer vision. pp 263–270
Haritaoglu I, Harwood D, Davis L (2000) W4: real-time surveillance of people and their activities. IEEE Trans Pattern Anal Mach Intell 22(8):809–830
He B, Xu D, Nian R, van Heeswijk M, Yu Q, Miche Y, Lendasse A (2014) Fast face recognition via sparse coding and extreme learning machine. Cognit Comput 6(2):264–277
Heikkila J, Rahtu E, Ojansivu V (2014) Local phase quantization for blur insensitive texture description. In: Local binary patterns: new variants and applications. pp 49–84
Horn BK, Schunck BG (1981) Determining optical flow. Artif Intell 17(13):185–203
Hossain MA, Makihara Y, Wang J, Yagi Y (2010) Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. Patt Recogn 43(6):2281–2291
Hou YL, Pang GH (2011) People counting and human detection in a challenging situation. IEEE Trans Syst Man Cybern Part A Syst Hum 41(1):24–33
Hu W, Tan T, Wang L, Maybank S (2004) A survey on visual surveillance of object motion and behaviors. IEEE Trans Syst Man Cybern Part C Appl Rev 34(3):334–352
Hu W, Zhou X, Hu M, Maybank S (2009) Occlusion reasoning for tracking multiple people. IEEE Trans Circuits Syst Video Technol 19(1):114–121
Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst
Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: Theory and applications. Neurocomputing 70(13):489–501
Huang J, Kumar S, Mitra M, Zhu WJ (1998) Spatial color indexing and applications. In: Sixth international conference on computer vision, 1998. pp 602–607
Husain M, Saber E, Misic V, Joralemon S (2006) Dynamic object tracking by partial shape matching for video surveillance applications. In: IEEE international conference on image processing. pp 2405–2408
Huttenlocher D, Klanderman G, Rucklidge W (1993) Comparing images using the hausdorff distance. IEEE Trans Pattern Anal Mach Intell 15(9):850–863
Huttenlocher D, Noh J, Rucklidge W (1993) Tracking non-rigid objects in complex scenes. In: International conference on computer vision. pp 93–101
Iwama H, Muramatsu D, Makihara Y, Yagi Y (2012) Gait-based person-verification system for forensics. In: IEEE fifth international conference on biometrics: theory, applications and systems. pp 113–120
Jain AK, Dass S, Nandakumar K (2004) Soft biometric traits for personal recognition systems. In: Biometric authentication. pp 731–738
Jain AK, Pankanti S, Prabhakar S, Hong L, Ross A (2004) Biometrics: a grand challenge. In: 17th International conference on pattern recognition, ICPR '04. IEEE Computer Society, Washington, DC, pp 935–942

Jain A, Ross A, Prabhakar S (2004) An introduction to biometric recognition. IEEE Trans Circuits Syst Video Technol 14(1):4–20
Jean F, Albu AB, Bergevin R (2009) Towards view-invariant gait modeling: computing view-normalized body part trajectories. Patt Recogn 42(11):2936–2949
Jia X, Lu H, Yang MH (2012) Visual tracking via adaptive structural local sparse appearance model. In: IEEE conference on computer vision and pattern recognition. pp 1822–1829
Jiang H, Fels S, Little J (2007) A linear programming approach for multiple object tracking. In: IEEE conference on computer vision and pattern recognition. pp 1–8
Ji X, Liu H (2010) Advances in view-invariant human motion analysis: a review. IEEE Trans Syst Man Cybern Part C Appl Rev 40(1):13–24
Julier S, Uhlmann J (2004) Unscented filtering and nonlinear estimation. Proc IEEE 92(3):401–422
KaewTrakulPong P, Bowden R (2003) A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes. Image Vis Comput 21(10):913–929
Kalal Z, Mikolajczyk K, Matas J (2012) Tracking-learning-detection. IEEE Trans Pattern Anal Mach Intell 34(7):1409–1422
Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82(Series D):35–45
Kang J, Cohen I, Medioni G (2003) Continuous tracking within and across camera streams. In: IEEE computer society conference on computer vision and pattern recognition, vol. 1. pp I-267–I-272
Kasinski A, Florek A, Schmidt A (2008) The put face database. Image Process Commun 13(3–4):59–64
Kim K, Chalidabhongse TH, Harwood D, Davis L (2005) Real-time foreground background segmentation using codebook model. Real Time Imaging 11(3):172–185
Kim I, Choi H, Yi K, Choi J, Kong S (2010) Intelligent visual surveillance a survey. Int J Control Autom Syst 8(5):926–939
Klare BF, Klein B, Taborsky E, Blanton A, Cheney J, Allen K, Grother P, Mah A, Burge M, Jain AK (2015) Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In: Conference on computer vision and pattern recognition (CVPR)
Klontz J, Jain A (2013) A case study of automated face recognition: the boston marathon bombings suspects. IEEE Comput 46(11):91–94
Ko T (2008) A survey on behavior analysis in video surveillance for homeland security applications. In: 37th IEEE applied imagery pattern recognition workshop. pp 1–8
Krger V, Kragic D, Ude A, Geib C (2007) The meaning of action: a review on action recognition and mapping. Adv Robot 21(13):1473–1501
Kristan M, Pflugfelder R, Leonardis A, Matas J, Porikli F, Cehovin L, Nebehay G, Fernandez G, Vojir T, Gatt A, Khajenezhad A, Salahledin A, Soltani-Farani A, Zarezade A, Petrosino A, Milton A, Bozorgtabar B, Li B, Chan CS, Heng C, Ward D, Kearney D, Monekosso D, Karaimer H, Rabiee H, Zhu J, Gao J, Xiao J, Zhang J, Xing J, Huang K, Lebeda K, Cao L, Maresca M, Lim MK, El Helw M, Felsberg M, Remagnino P, Bowden R, Goecke R, Stolkin R, Lim S, Maher S, Poullot S, Wong S, Satoh S, Chen W, Hu W, Zhang X, Li Y, Niu Z (2013) The visual object tracking vot2013 challenge results. In: IEEE international conference on computer vision workshops. pp 98–111
Kwon J, Lee KM (2010) Visual tracking decomposition. In: IEEE conference on computer vision and pattern recognition. pp 1269–1276
Lee L, Grimson WEL (2002) Gait analysis for recognition and classification. In: Fifth IEEE international conference on automatic face and gesture recognition. pp 148–155
Leibe B, Schindler K, Van Gool L (2007) Coupled detection and trajectory estimation for multi-object tracking. In: IEEE 11th international conference on computer vision. pp 1–8
Li S, Chu R, Liao S, Zhang L (2007) Illumination invariant face recognition using near-infrared images. IEEE Trans Pattern Anal Mach Intell 29(4):627–639
Li H, Hua G, Lin Z, Brandt J, Yang J (2013) Probabilistic elastic matching for pose variant face verification. In: IEEE conference on computer vision and pattern recognition (CVPR). pp 3499–3506
Lin SF, Chen JY, Chao HX (2001) Estimation of number of people in crowded scenes using perspective transformation. IEEE Trans Syst Man Cybern Part A Syst Hum 31(6):645–654
Liu Z, Shen H, Feng G, Hu D (2012) Tracking objects using shape context matching. Neurocomputing 83:47–55
Liu H, Sun F, Yu Y (2014) Multitask extreme learning machine for visual tracking. Cognit Comput 6(3):391–404
Liu C, Hu C, Aggarwal J (2011) Eigenshape kernel based mean shift for human tracking. In: IEEE international conference on computer vision workshops. pp 1809–1816
Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: Hayes PJ (ed) IJCAI. William Kaufmann, pp 674–679

Luque R, Domnguez E, Palomo E, Muoz J (2008) A neural network approach for video object segmentation in traffic surveillance. In: Image analysis and recognition. pp 151–158
Lyle JR, Miller PE, Pundlik SJ, Woodard DL (2012) Soft biometric classification using local appearance periocular region features. Patt Recognit 45(11):3877–3885
Lyle J, Miller P, Pundlik S, Woodard D (2010) Soft biometric classification using periocular region features. In: Fourth IEEE international conference on biometrics: theory applications and systems. pp 1–7
Maddalena L, Petrosino A (2008) A self-organizing approach to background subtraction for visual surveillance applications. IEEE Trans Image Process 17(7):1168–1177
Maddalena L, Petrosino A (2014) The 3dsobs+ algorithm for moving object detection. Comput Vis Image Underst 122:65–73
Maggio E (2005) Cavallaro a multi-part target representation for color tracking. In: IEEE international conference on image processing, vol. 1. pp I-729–I-732
Maggio E, Piccardo E, Regazzoni C, Cavallaro A (2007) Particle phd filtering for multi-target visual tracking. In: IEEE international conference on acoustics, speech and signal processing, vol. 1. pp I-1101–I-1104
Martinez AM (2002) Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Trans Pattern Anal Mach Intell 24(6):748–763
Matey J, Naroditsky O, Hanna K, Kolczynski R, LoIacono D, Mangru S, Tinker M, Zappia T, Zhao WY (2006) Iris on the move: acquisition of images for iris recognition in less constrained environments. Proc IEEE 94(11):1936–1947
McCahill M, Norris C (2002) Cctv in britain. Center for Criminology and Criminal Justice, University of Hull, London
McFarlane N, Schofield C (1995) Segmentation and tracking of piglets in images. Mach Vis Appl 8(3):187–193
McKenna SJ, Jabri S, Duric Z, Wechsler H (2000) Tracking interacting people. In: Proceedings of the fourth IEEE international conference on automatic face and gesture recognition 2000, FG '00. IEEE Computer Society, Washington, DC. p 348
McKenna SJ, Raja Y, Gong S (1999) Tracking colour objects using adaptive mixture models. Image Vis Comput 17(34):225–231
Mei X, Ling H (2011) Robust visual tracking and vehicle classification via sparse representation. IEEE Trans Pattern Anal Mach Intell 33(11):2259–2272
Mikolajczyk K, Schmid C, Zisserman (2004) A Human detection based on a probabilistic assembly of robust part detectors. In: ECCV. pp 69–82
Milan A, Roth S, Schindler K (2014) Continuous energy minimization for multitarget tracking. IEEE Trans Pattern Anal Mach Intell 36(1):58–72
Mittal A, Davis LS (2003) M2tracker: a multi-view approach to segmenting and tracking people in a cluttered scene. Int J Comput Vis 51(3):189–203
Moctezuma D, Conde C, de Diego I, Cabello E (2011) Person detection in surveillance environment with hogg: gabor filters and histogram of oriented gradient. In: IEEE international conference on computer vision workshops. pp 1793–1800
Moeslund TB, Hilton A, Krger V (2006) A survey of advances in vision-based human motion capture and analysis. Comput Vis Image Underst 104(23):90–126
Moeslund TB, Granum E (2001) A survey of computer vision-based human motion capture. Comput Vis Image Underst 81(3):231–268
Munder S, Schnorr C, Gavrila D (2008) Pedestrian detection and tracking using a mixture of view-based shape-texture models. IEEE Trans Intell Transp Syst 9(2):333–343
Murray MP (1967) Gait as a total pattern of movement. American Journal of Physical Medicine 46(1):290–333
Neves J (2015) Quis-campi dataset. http://quiscampi.di.ubi.pt
Neves JC, Moreno JC, Barra S, Proenca H (2015) Acquiring high-resolution face images in outdoor environments: a master-slave calibration algorithm. In: IEEE 7th international conference on biometrics theory, applications and systems (BTAS). pp 1–8
Ojala T, Pietikinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Patt Recognit 29(1):51–59
Ojansivu V, Rahtu E, Heikkila J (2008) Rotation invariant local phase quantization for blur insensitive texture analysis. In: 19th International conference on pattern recognition. pp 1–4
Okuma K, Taleghani A, Freitas N, Little JJ, Lowe DG (2004) A boosted particle filter: multitarget detection and tracking. In: ECCV. pp 28–39
Okumura M, Iwama H, Makihara Y, Yagi Y (2010) Performance evaluation of vision-based gait recognition using a very large-scale gait database. In: Fourth IEEE international conference on biometrics: theory applications and systems. pp 1–6
Oliver N, Rosario B, Pentland A (2000) A bayesian computer vision system for modeling human interactions. IEEE Trans Pattern Anal Mach Intell 22(8):831–843

Padole C, Proença H (2012) Periocular recognition: analysis of performance degradation factors. In: 5th IAPR international conference on biometrics. pp 439–445
Pantic M, Pentland A, Nijholt A, Huang T (2006) Human computing and machine understanding of human behavior: a survey. In: Proceedings of the 8th international conference on multimodal interfaces, ICMI-ACM. New York, pp 239–248
Park U, Jillela R, Ross A, Jain A (2011) Periocular biometrics in the visible spectrum. IEEE Trans Inf Forensics Secur 6(1):96–106
Park U, Choi HC, Jain A, Lee SW (2013) Face tracking and recognition at a distance: a coaxial and concentric PTZ camera system. IEEE Trans Inf Forensics Secur 8(10):1665–1677
Park U, Ross A, Jain A (2009) Periocular biometrics in the visible spectrum: a feasibility study. In: IEEE 3rd international conference on biometrics: theory, applications, and systems. pp 1–6
PETS (2015) Performance evaluation of tracking and surveillance. http://www.cvg.reading.ac.uk/slides/pets.html
Phillips P (2014a) Color feret database. http://www.nist.gov/itl/iad/ig/colorferet.cfm
Phillips P (2014b) Face and ocular challenge series. http://www.nist.gov/itl/iad/ig/focs.cfm
Phillips P, Flynn P, Scruggs T, Bowyer K, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the face recognition grand challenge. In: IEEE computer society conference on computer vision and pattern recognition, vol. 1. pp 947–954
Popoola O, Wang K (2012) Video-based abnormal human behavior recognition a review. IEEE Trans Syst Man Cybern Part C Appl Rev 42(6):865–878
Poppe R (2007) Vision-based human motion analysis: an overview. Comput Vis Image Underst 108(1–2):4–18
Poppe R (2010) A survey on vision-based human action recognition. Image and Vision Computing 28(6):976–990
Priydarshi AN, Chakraborty P, Nandi G (2013)Speed invariant, human gait based recognition system for video surveillance security. In: Intelligent interactive technologies and multimedia. pp 325–335
Proença H (2007) Towards non-cooperative biometric iris recognition. Ph.D. thesis, University of Beira Interior
Proença H, Filipe S, Santos R, Oliveira J, Alexandre L (2010) The ubiris.v2: a database of visible wavelength iris images captured on-the-move and at-a-distance. IEEE Trans Pattern Anal Mach Intell 32(8):1529–1535
Proença H, Alexandre L (2007) Toward noncooperative iris recognition: a classification approach using multiple signatures. IEEE Trans Pattern Anal Mach Intell 29(4):607–612
Raty T (2010) Survey on contemporary remote surveillance systems for public safety. IEEE Trans Syst Man Cybern Part C Appl Rev 40(5):493–515
Reid DA, Nixon MS (2010) Imputing human descriptions in semantic biometrics. pp 25–30
Reid D (1979) An algorithm for tracking multiple targets. IEEE Trans Autom Control 24(6):843–854
Reid D, Nixon M, Stevenage S (2014) Soft biometrics; human identification using comparative descriptions. IEEE Trans Pattern Anal Mach Intell 36(6):1216–1228
Ross D, Lim J, Lin RS, Yang MH (2008) Incremental learning for robust visual tracking. Int J Comput Vis 77(1–3):125–141
Saber E, Xu Y, Tekalp AM (2005) Partial shape recognition by sub-matrix matching for partial matching guided image labeling. Patt Recognit 38(10):1560–1573
Samangooei S, Nixon M (2008) Performing content-based retrieval of humans using gait biometrics. In: Semantic multimedia. pp 105–120
Samangooei S, Nixon M (2010) Performing content-based retrieval of humans using gait biometrics. Multimed Tools Appl 49(1):195–212
Sanchez-Reillo R, Sanchez-Avila C, Gonzalez-Marcos A (2000) Biometric identification through hand geometry measurements. IEEE Trans Pattern Anal Mach Intell 22(10):1168–1171
Santner J, Leistner C, Saffari A, Pock T, Bischof H (2010) Prost: parallel robust online simple tracking. In: IEEE conference on computer vision and pattern recognition. pp 723–730
Santos G, Proença H (2013) Periocular biometrics: an emerging technology for unconstrained scenarios. In: IEEE workshop on computational intelligence in biometrics and identity management. pp 14–21
Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: IEEE computer society conference on computer vision and pattern recognition
Schwartz W, Kembhavi A, Harwood D, Davis L (2009) Human detection using partial least squares analysis. In: IEEE 12th international conference on computer vision. pp 24–31
Senior AW, Hampapur A, Lu M (2005) Acquiring multiscale images by pan-titl-zoom control and automatic multicamera calibration. In: Proceedings of the 7*th* IEEE workshop on application of computer vision, vol. 1. Breckenridge, pp 433–438
Sevilla-Lara L (2012) Distribution fields for tracking. In: IEEE conference on computer vision and pattern recognition, CVPR '12IEEE computer society. Washington, DC, pp 1910–1917

Sharma A, Venkatesh KS, Mukerjee A (2011) Human pose estimation in surveillance videos using temporal continuity on static pose. In: 2011 International Conference on image information processing (ICIIP), pp 1–6
Shi J, Tomasi C (1994) Good features to track. In: IEEE computer society conference on computer vision and pattern recognition. pp 593–600
Sodemann A, Ross M, Borghetti B (2012) A review of anomaly detection in automated surveillance. IEEE Trans Syst Man Cybern Part C Appl Rev 42(6):1257–1272
Squires B, Sammut C (1995) Automatic speaker recognition: an application of machine learning. In: ICML. pp 515–521
Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real-time tracking. IEEE Comput Soc Conf Comput Vis Patt Recognit 2:246–252
Stern H, Efros B (2005) Adaptive color space switching for tracking under varying illumination. Image Vis Comput 23(3):353–364
Subburaman V, Descamps A, Carincotte C (2012) Counting people in the crowd using a generic head detector. In: IEEE ninth international conference on advanced video and signal-based surveillance. pp 470–475
Supancic J, Ramanan D (2013) Self-paced learning for long-term tracking. In: IEEE conference on computer vision and pattern recognition. pp 2379–2386
Szeto MW, Gazis DC (1972) Application of kalman filtering to the surveillance and control of traffic systems. Transp Sci 6(4):419
Talukder A, Matthies L (2004) Real-time detection of moving objects from moving vehicles using dense stereo and optical flow. IEEE RSJ Int Conf Intell Robots Syst 4:3718–3725
Tan CW, Kumar (2012) A human identification from at-a-distance images by simultaneously exploiting iris and periocular features. In: 21st international conference on pattern recognition. pp 553–556
Tan CW, Kumar A (2013) Towards online iris and periocular recognition under relaxed imaging constraints. IEEE Trans Image Process 22(10):3751–3765
Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process 19(6):1635–1650
Terena (2014) Koningsplein webcam. http://www.terena.org/webcam/
Tome P, Fierrez J, Vera-Rodriguez R, Nixon M (2014) Soft biometrics and their application in person recognition at a distance. IEEE Trans Inf Forensics Secur 9(3):464–475
Tsochantaridis I, Joachims T, Hofmann T, Altun Y (2005) Large margin methods for structured and interdependent output variables. J Mach Learn Res 6:1453–1484
Turaga P, Chellappa R, Veeraraghavan A (2010) Advances in video-based human activity analysis: challenges and approaches. Adv Comput 80:237–290
Turk M, Pentland A (1991) Face recognition using eigenfaces. In: IEEE computer society conference on computer vision and pattern recognition. pp 586–591
Vezzani R, Baltieri D, Cucchiara R (2013) People reidentification in surveillance and forensics: a survey. ACM Comput Surv 46(2):29:1–29:37
Viola P, Platt JC, Zhang C (2005) Multiple instance boosting for object detection. Adv Neural Inf Process Syst 18:1417–1426
Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, vol. 1. pp I-511–I-518
Viola P, Jones M, Snow D (2003) Detecting pedestrians using patterns of motion and appearance. In: Ninth IEEE international conference on computer vision, vol. 2. pp 734–741
Vot2015 challenge (2015). http://www.votchallenge.net/vot2015/. Accessed 21 Dec 2015
Wang L, Hu W, Tan T (2003) Recent developments in human motion analysis. Patt Recognit 36(3):585–601
Wang L, Tan T, Ning H, Hu W (2003) Silhouette analysis-based gait recognition for human identification. IEEE Trans Pattern Anal Mach Intell 25(12):1505–1518
Wang X, Han T, Yan S (2009) An hog-lbp human detector with partial occlusion handling. In: IEEE 12th international conference on computer vision. pp 32–39
Weinland D, Ronfard R, Boyer E (2006) Free viewpoint action recognition using motion history volumes. Comput Vis Image Underst 104(23):249–257
Weinland D, Ronfard R, Boyer E (2011) A survey of vision-based methods for action representation, segmentation and recognition. Comput Vis Image Underst 115(2):224–241
Wheeler F, Weiss R, Tu P (2010) Face recognition at a distance system for surveillance applications. In: Proceedings of the fourth ieee international conference on biometrics: theory applications and systems. Washington, DC, pp 1–8
Wildes R (1997) Iris recognition: an emerging biometric technology. Proc IEEE 85(9):1348–1363

Wren C, Azarbayejani A, Darrell T, Pentland A (1997) Pfinder: real-time tracking of the human body. IEEE Trans Pattern Anal Mach Intell 19(7):780–785
Wright J, Yang A, Ganesh A, Sastry S, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
Wu Y, Huang T (2001) A co-inference approach to robust visual tracking. In: Eighth IEEE international conference on computer vision, vol. 2. pp 26–33
Wu B, Nevatia R (2007a) Cluster boosted tree classifier for multi-view, multi-pose object detection. In: IEEE 11th international conference on computer vision. pp 1–8
Wu B, Nevatia R (2007b) Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. Int J Comput Vis 75(2):247–266
Wu B, Nevatia R (2009) Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses. Int J Comput Vis 82(2):185–204
Wu Y, Yu T (2006) A field model for human detection and tracking. IEEE Trans Pattern Anal Mach Intell 28(5):753–765
Wu Y, Ling H, Yu J, Li F, Mei X, Cheng E (2011a) Blurred target tracking by blur-driven tracker. In: IEEE international conference on computer vision. pp 1100–1107
Wu J, Xia J, Chen JM, Cui ZM (2011b) Adaptive detection of moving vehicle based on on-line clustering. J Comput 6(10):2045–2052
Wu Y, Lim J, Yang MH (2013) Online object tracking: a benchmark. In: IEEE conference on computer vision and pattern recognition. pp 2411–2418
Xiao J, Stolkin R, Leonardis A (2013) An enhanced adaptive coupled-layer lgtracker++. In: IEEE international conference on computer vision workshops. pp 137–144
Xu Y, Song D (2010) Systems and algorithms for autonomous and scalable crowd surveillance using robotic ptz cameras assisted by a wide-angle camera. Auton Robots 29(1):53–66
Yao Y, Abidi B, Kalka N, Schmid N, Abidi M (2008) Improving long range and high magnification face recognition: database acquisition, evaluation and enhancement. Comput Vis Image Underst 111(2):111–125
Yao J, Odobez JM (2011) Fast human detection from joint appearance and foreground feature subset covariances. Comput Vis Image Underst 115(10):1414–1426
Zhang J, Pu J, Chen C, Fleischer R (2010) Low-resolution gait recognition. IEEE Trans Syst Man Cybern Part B Cybern 40(4):986–996
Zhang X, Hu W, Bao H, Maybank S (2013) Robust head tracking based on multiple cues fusion in the kernel-bayesian framework. IEEE Trans Circuits Syst Video Technol 23(7):1197–1208
Zhang T, Ghanem B, Liu S, Ahuja N (2012) Robust visual tracking via multi-task sparse learning. In: IEEE conference on computer vision and pattern recognition. pp 2042–2049
Zhang L, Li Y, Nevatia R (2008) Global data association for multi-object tracking using network flows. In: IEEE conference on computer vision and pattern recognition. pp 1–8
Zhang L, Li S, Yuan X, Xiang S (2007) Real-time object classification in video surveillance based on appearance learning. In: IEEE conference on computer vision and pattern recognition. pp 1–8
Zhang K, Zhang L, Yang MH (2012) Real-time compressive tracking. In: ECCV. pp 864–877
Zhao T, Nevatia R, Wu B (2008) Segmentation and tracking of multiple humans in crowded environments. IEEE Trans Pattern Anal Mach Intell 30(7):1198–1211
Zhao T, Nevatia R (2004) Tracking multiple humans in complex situations. IEEE Trans Pattern Anal Mach Intell 26(9):1208–1221
Zhao Q, Tao H (2009) A motion observable representation using color correlogram and its applications to tracking. Comput Vis Image Underst 113(2):273–290
Zhong W, Lu H, Yang MH (2012) Robust object tracking via sparsity-based collaborative model. In: IEEE conference on computer vision and pattern recognition. pp 1838–1845
Zhou S, Krueger V, Chellappa R (2003) Probabilistic recognition of human faces from video. Comput Vis Image Underst 91(12):214–245
Zhou Q, Aggarwal J (2006) Object tracking in an outdoor environment using fusion of features and cameras. Image Vis Comput 24(11):1244–1255
Zhou H, Hu H (2008) Human motion tracking for rehabilitationa survey. Biomed Signal Process Control 3(1):1–18
Zhu X, Lei Z, Yan J, Yi D, Li S (2015) High-fidelity pose and expression normalization for face recognition in the wild. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR). pp 787–796
Zivkovic Z (2004) Improved adaptive gaussian mixture model for background subtraction. In: Proceedings of the 17th international conference on pattern recognition, vol. 2. pp 28–31