

CHAPTER 1

Unconstrained Data Acquisition Frameworks and Protocols

João C. Neves^{*}, Juan C. Moreno^{*}, Silvio Barra[†], Fabio Narducci[‡],
Hugo Proença^{*}

^{*}IT – Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal

[†]Department of Mathematic and Computer Science, University of Cagliari, Cagliari, Italy

[‡]BIPLab – University of Salerno, Fisciano, Italy

Contents

1.1	Introduction	1
1.2	Unconstrained Biometric Data Acquisition Modalities	3
1.3	Typical Challenges	4
1.3.1	Optical Constraints	5
1.3.2	Non-comprehensive View of the Scene	5
1.3.3	Out-of-Focus	6
1.3.4	Calibration of Multi-camera Systems	7
1.4	Unconstrained Biometric Data Acquisition Systems	8
1.4.1	Low Resolutions Systems	9
1.4.1.1	Super-resolution Approach	9
1.4.1.2	Gait Recognition	10
1.4.2	PTZ-Based Systems	11
1.4.3	Face	15
1.4.3.1	Typical Design of Master–Slave Systems	17
1.4.3.2	Systems Based on Logical Alignment of the Cameras	19
1.4.3.3	QUIS–CAMPI System	21
1.4.3.4	Other Biometrics: Iris, Periocular, and Ear	23
1.5	Conclusions	25
	References	26

1.1 INTRODUCTION

The recognition of individuals either from physical or behavioral traits, usually denoted as biometric traits, is common practice in environments where the subjects cooperate with the acquisition system.

In the last years, the focus has been placed on extending the robustness of recognition methods to address less constrained scenarios with non-cooperative subjects. Researchers have introduced different strategies to address high variability in the age [1], pose [2], illumination [3], expres-

sion [4] (A-PIE), and other confounding factors such as occlusion [5] and blur [6]. Also, these improvements have been evidenced by the performance advances reported in unconstrained biometric datasets, such as the Labeled Faces in the Wild (LFW) [7].

Despite these achievements, the recognition of humans in totally wild conditions observed in visual surveillance scenarios has not been achieved, yet. In this kind of setup, images are captured from large distances, and the acquired data have limited discriminative capabilities, even when using high-resolution cameras [8]. Considering that in unconstrained environments data resolution may have a greater impact on the performance than A-PIE factors, several authors have been particularly devoted to extending the workability of biometric data acquisition frameworks to unconstrained scenarios in which human collaboration is not assumed.

In this chapter, we review most of the relevant frameworks and protocols for acquiring biometric data in unconstrained scenarios. In Section 1.2, we provide a comparative analysis between the different acquisition modalities. The advantages of using magnifications devices, such as PTZ cameras, are evidenced by the difference in the resolution between the eyes (interpupillary distance). The resolution of biometric data acquired by PTZ cameras at the maximum zoom is five times higher than typical surveillance cameras. Also, the minimum resolution required to acquire high-resolution face images (interpupillary distance greater than 60 pixels) with a stand-off distance larger than 5 m can only be attained using PTZ devices.

Section 1.3 discusses the advantages and drawbacks of the different acquisition modalities with special attention given to the use of magnification devices in unconstrained environments. The use of a highly narrow field of view introduces a multitude of challenges that restrict the workability of PTZ-based systems in outdoor scenarios (inter-camera calibration) and degrade image quality (e.g., out-of-focus, incorrect exposure).

In Section 1.4, we present a comprehensive collection of the state-of-the-art systems for unconstrained scenarios. The systems are organized according to the modalities of unconstrained biometric data acquisition: (i) low-resolution systems; and (ii) PTZ-based systems. In the former, most approaches rely on soft biometrics (e.g., gait) for recognizing individuals in unconstrained scenarios, since the reduced discrimination of data inhibits the use of hard biometrics. The use of these traits is only feasible when relying on super-resolution. In the latter, the systems are grouped with respect to the biometric trait that they were designed to acquire. Despite the advantages of iris regarding recognition performance, its reduced size

curtails the maximum stand-off distance of these systems. Consequently, most approaches have introduced multiple strategies to acquire facial imagery at large stand-off distances. The workability of these approaches in real unconstrained environments is discussed by analyzing their feasibility in surveillance scenarios, which are among of the most representative examples of these environments. Among these systems, particular attention is given to the works of Park et al. [9] and Neves et al. [10], which are two representative examples of PTZ-based systems capable of acquiring high-resolution face images in surveillance scenarios. Finally, Section 1.5 concludes the chapter.

1.2 UNCONSTRAINED BIOMETRIC DATA ACQUISITION MODALITIES

The acquisition of biometric data in unconstrained environments is generally performed in two distinct manners: (i) using wide-view cameras and (ii) using magnification devices, such as PTZ cameras.

The former strategy is usually associated with CCTV cameras, which have increased exponentially in number during the last years [11]. The large number of such devices deployed in outdoor scenarios and their reduced cost are the major reasons for relying on CCTV surveillance systems. However, in wide open scenarios, the obtained resolution is not adequate to faithfully represent biometric data [12], restraining the recognition of humans at-a-distance. With the rise of high-resolution cameras, they have been considered as substitutes of old CCTV cameras and suggested as the solution for remote human recognition. Even though high-resolution cameras can be a practical solution for mid-term distances, they still cannot outperform PTZ-based systems. Fig. 1.1 illustrates the relation between the interpupillary distance (assumed to be 61 cm) and the stand-off distance (the distance between the front of the lens and the subject) when using different optical devices. In this comparison, the angle of view (AOV) of wide-view cameras was assumed to be 70° , while the AOV of PTZ cameras was assumed as 2.1° when set at the maximum zoom. This comparison shows that, apart from the notorious differences between the resolution of the data, the minimum number of pixels required to acquire high-resolution face images (interpupillary distance greater than 60 pixels) with a stand-off distance larger than 5 m can only be attained when using PTZ devices.

The use of PTZ cameras has been pointed as the most efficient and practical way to acquire biometric data in unconstrained scenarios, since

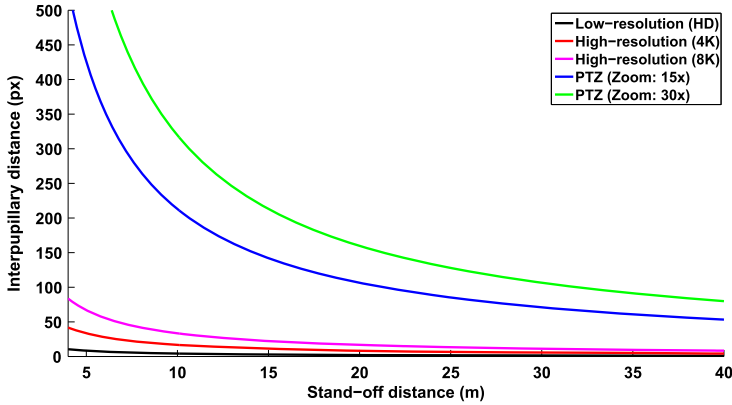


Figure 1.1 Relation between the inter-pupillary resolution and the stand-off distance when using different acquisition devices. The number of pixels between the eyes is determined with respect to the stand-off distance and four acquisition devices: (i) a typical surveillance camera (720p, 70°); (ii) a high-resolution camera (4K, 70°); (iii) a high-resolution camera (4K, 70°); (iv) a PTZ camera with 15× zoom (1080p, 4.2°); and (v) PTZ camera with 30× zoom (1080p, 2.1°). Note the evident advantages of using PTZ cameras, the resolution of face traits is more than 5× the resolution of 8K cameras.

the mechanical properties of these devices permit the acquisition of high resolution imagery at arbitrary scene locations. Although these devices can be used independently, they are usually exploited as a foveal sensor assisted by a wide-view camera. This strategy, known as a master–slave architecture, is regarded as the most efficient for acquiring biometrics at-a-distance in unconstrained scenarios, but at the same time it also introduces a multitude of challenges (refer to Section 1.3) that have been progressively addressed by different approaches (refer to Section 1.4).

1.3 TYPICAL CHALLENGES

As discussed in Section 1.2, PTZ-based approaches are currently the best strategy for reliable acquisition of biometric data in outdoor environments. The zooming capabilities of such systems make it possible to image facial biometric traits that usually ask for dedicated hardware and user’s collaboration to be properly framed and processed at long distances. State-of-the-art PTZ cameras can achieve optical zoom magnifications up to 30× with an angle of view of about 2°. Despite these advantages, the use of a highly narrow field of view also entails additional challenges.

1.3.1 Optical Constraints

The use of high zoom levels has a tremendous impact on the quality of the acquired images since optical magnification is achieved by increasing the focal distance of the camera (f) and reducing its angle of view (AOV). As a consequence, the amount of light reaching the sensor is considerably less as the AOV decreases, which is particularly critical in outdoor scenarios where illumination is non-standard.

To compensate for this effect, most cameras increase the aperture of the diaphragm (D) in the same proportion of f . The ratio between f and the aperture of the camera is denoted as the F-number (see Eq. (1.1)), and is commonly used in photography to maintain image brightness along different zoom magnifications.

$$\text{F-number} = \frac{f}{D} \quad (1.1)$$

However, its side effect is the reduction of the depth of field, which increases the chances of obtaining blurred images. As an alternative, it is possible to increase the exposure time (E) thus balancing the impact that extreme f values may have on the amount of light that reaches the sensor. However, higher values of E also increase the motion-blur level in the images.

A more robust solution is to simultaneously adjust both D and E , which is, in general, the strategy adopted by PTZ devices. However, as illustrated in Fig. 1.2, the number of ideal configurations for (D, E) is greatly dependent on zoom magnification.

1.3.2 Non-comprehensive View of the Scene

While zooming enables close inspection of narrow regions in the scene, it also inhibits scene monitoring. As a consequence, the detection and tracking of individuals can be hardly attained when using extreme zoom levels.

To mitigate this shortcoming, some systems alternate between different zoom levels, i.e., subject detection and tracking is performed at minimal zoom levels; close-up imaging of interesting subjects using maximum zoom levels is indeed used to process the details. However, zoom transition is the most time-consuming task of PTZ devices, which significantly restricts the efficiency of using a single PTZ camera for biometric recognition purposes.

As an alternative, several authors have exploited a master-slave architecture where the PTZ camera is assisted by a wide-view camera. The

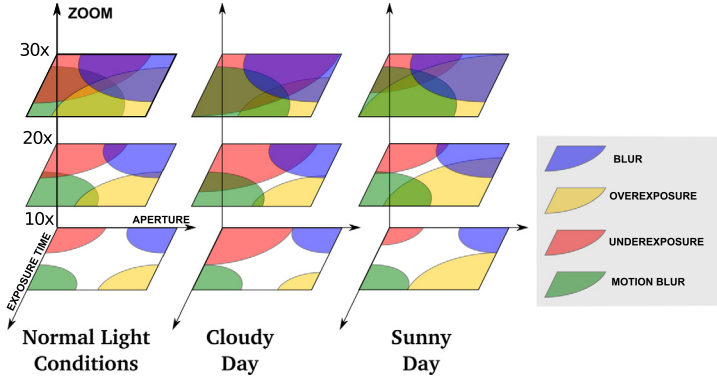


Figure 1.2 The spectrum of optical constraints with respect to the exposure time and the aperture and zoom level. The set of (E, D) combinations that produce non-degraded images decreases significantly as the focal distance increases. Besides, it is worth noting that the ideal set of (E, D) values (in white) varies with respect to the illumination conditions.

wide-view camera stream, acting as the *master*, is used to globally monitor the scene and to detect and track the subjects. On the other hand, the PTZ camera is then used as the *slave*. It operates as a foveal sensor acquiring close-up shots from a set of interesting regions provided by the master camera. In this manner, pointing and zooming over a specific region allows acquiring a detailed view of detected subjects. It is important to ensure that the two cameras are strongly related to each other so that it is possible to find a proper correspondence of any point in 2D coordinates of PTZ camera to the wide-camera, and vice versa. Despite being the most efficient and practical architecture to acquire high-resolution biometric data in unconstrained environments, a master-slave system also entails additional challenges.

1.3.3 Out-of-Focus

As previously discussed in Section 1.3.1, the use of extreme zoom levels significantly reduces the depth of field. To correctly adjust the focus distance to the subject position in the scene, two different strategies can be exploited: (i) auto-focus and (ii) manual focus.

In the former, the focus adjustment is guided by an image contrast maximization search. Despite being highly effective in wide-view cameras, this approach fails to provide focused images of moving subjects when

using extreme zoom magnifications. Firstly, the reduced field of view of the camera significantly reduces the amount of time the subject is imaged (approximately 1 s), and the auto-focus mechanism is not fast enough (approximately 2 s) to seamlessly frame the subject over time. Secondly, motion also introduces blur in the image which may mislead the contrast adjustment scheme.

As an alternative, the focus lens can be manually adjusted with respect to the distance of the subject to the camera. Given the 3D position of the subject, it is possible to infer its distance to the camera. Then, focus is dynamically adjusted using a function relating the subject's distance and the focus lens position. In this strategy, the estimation of a 3D subject position is regarded as the major bottleneck since it depends on the use of stereo reconstruction techniques. However, this issue has been progressively addressed by the state-of-the-art methods since 3D information is critical for accurate PTZ-based systems.

1.3.4 Calibration of Multi-camera Systems

Camera calibration typically refers to establishing the relationship between the world and camera coordinate systems. Several tools have been developed to address this problem with effective results. However, when using more than one camera, the issues increase, thus turning the calibration into a harder problem.

In a multi-camera system, the cameras are supposed, in general, to cooperate and share the acquisitions of the scene. Therefore, apart from calibrating each camera separately, in such systems a mapping function between the camera streams must be defined that can turn a point in the coordinate system of a camera into that of another.

It represents a non-trivial goal to achieve because of an important constraint of multi-camera systems that is related to epipolar geometry. The *epipolar geometry* [13] is used to represent the geometric relations of two points onto 2D images that come from two cameras when pointing at the same location in the world coordinates (in a 3D space). Fig. 1.3 shows a typical example of epipolar geometry where a shared 3D point X is observed by both O_1 and O_2 . We can see that, by changing the position of X (see dots along the view-axis of O_1), its projection X_1 remains the same but it changes in X_2 . Only if the relative position of the two cameras is known, it is possible to estimate the match between the two image planes and therefore obtain the exact measure for both cameras. Assuming that

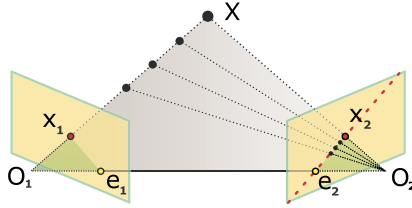


Figure 1.3 Epipolar geometry of a 3D point X over two image planes. Two cameras with their respective centers of projection points O_1 and O_2 observe a point X . The projection of X onto each of the image planes is denoted x_1 and x_2 . Points e_1 and e_2 are the epipoles.

O_1 is the wide-view camera of a master–slave system and O_2 is the PTZ camera, it would not be possible to determine the pan–tilt angle necessary to observe X by only using the information of its projection X_1 .

Multi-camera video surveillance systems are particularly suited to be exploited in person re-identification scenarios. Re-identification regards the task of assigning the same identifier to all the instances of the same object or, more specifically of the same person [14], by means of the visual aspect obtained from an image or a video. One of the most critical challenges of person re-identification is to recognize the same person viewed by disjoint, possibly non-overlapping cameras, at different time instants and locations [15]. Issues like tracking and indexing, camera–subject distance, and recognition-by-parts highly degrade the performance of re-identification. Relying on well calibrated cameras is therefore critical in order to have an efficient video surveillance system. The challenges and approaches of re-identification in surveillance is out of the scope of this study. Interested readers might find a useful source in [16].

1.4 UNCONSTRAINED BIOMETRIC DATA ACQUISITION SYSTEMS

This section presents a comprehensive collection of the state-of-the-art systems for biometric data acquisition in unconstrained video surveillance scenarios. These frameworks can be broadly divided into two groups: (i) CCTV systems and (ii) PTZ-based systems.

In the former, cameras are arranged using a maximum coverage strategy to monitor multiple subjects in a surveillance area. These systems are popular for their flexibility and reduced cost; however, the limited resolu-

tion of biometric data is regarded as their major drawback. The reduced discriminability of the data inhibits the use of hard biometrics for recognition purposes. Consequently, two feasible approaches are commonly used to recognize individuals in low resolution data: (i) the use of soft biometric traits (e.g., gait) and (ii) the use of super-resolution approaches to infer a higher level of details from poor acquisitions. A short overview of such systems is provided in Section 1.4.1.

The second group comprises systems using PTZ cameras for acquiring high-resolution imagery of regions of interest in the scene. Challenges are numerous but it is commonly accepted that such systems (described in detail in Section 1.4.2) represent the most efficient and mature solution for acquiring biometric data at-a-distance (e.g., face, iris, periocular).

1.4.1 Low Resolutions Systems

In surveillance scenarios, cameras are typically arranged in a way that maximizes the coverage area, thus making the biometric data acquired hard to discriminate. Despite the vast number of factors affecting recognition performance, the low-resolution of data is one of the major causes for the hardness of human identification in surveillance environments. To overcome this limitation, methodologies like the super-resolution or gait-based recognition have been explored. In the following sections, some examples of recent works of these two groups are discussed. We provide a short overview of such systems because we believe that current disadvantages of both make them still infeasible for video surveillance scenarios.

1.4.1.1 Super-resolution Approach

Super-resolution approaches infer a high-resolution image from low-resolution data using a pre-learned model that relates both representations [17]. Even though the majority of the works focus on improving data quality, some approaches have also tried to boost biometric recognition performance [18–20]. Ben-Ezra et al. [21] presented a *Jitter Camera* that exploits a micro actuator for enhancing the resolution provided by a low-resolution camera. Fig. 1.4 depicts the results attained for a low-resolution frame acquired in a surveillance scenario. Despite these improvements, it is commonly accepted that these approaches have to be extended to more realistic scenarios as described in [22]. This becomes particularly evident when trying to use such approaches to biometric recognition at-a-distance.



Figure 1.4 Example of the effect of a super-resolution approach on a low resolution image. The resolution enhancement achieved by Ben-Ezra et al. [21] starting from a low-resolution acquisition in a surveillance scenario.

1.4.1.2 Gait Recognition

The way humans walk can also be used for identification purposes and is usually known as gait recognition [23,24]. The advantages of gait can be summarized in the following: (i) it can be easily measured at-a-distance, (ii) it is difficult to disguise or occlude, and (iii) it is robust to low-resolution images. Moreover, a recent study about the covariate factors affecting recognition performance has found that gait is time-invariant in the short and medium term, thus gaining special attention among reliable biometric traits. On the other hand, gait strongly depends on the control over clothing and footwear [25], which negatively impacts its feasibility in surveillance scenarios.

Notwithstanding, many methods have been introduced in the literature to optimize gait recognition systems. Ran et al. [26] used human walking to segment and label body parts that can be helpful to perform real-time recognition. Gait patterns were captured and stacked in a 3D data cube containing all possible deformations. The symmetries between the patterns were analyzed in order to measure all possible changes and correctly label different parts of the human body.

Venkat and De Wilde [27] faced the problem of low-resolution in video data focusing on potential information from sub-gaits that can contribute to the recognition system. Moustakas et al. [28] exploited the height and stride length as soft biometric traits. These features were combined in a probabilistic framework to accurately perform gait recognition system.

Conversely, Jung et al. [29] exploited gait to estimate the head pose in surveillance scenarios. In this approach, a 3D face model was also inferred to improve recognition performance.

Choudhury and Tjahjadi [30] exploited human silhouettes extracted from a gait system to perform recognition. By analyzing the shape of the contours, they were able to overcome some typical side effects introduced by the presence of noise in the gait recognition system. Considering that this strategy is highly dependent on clothing, the authors introduced an extended version of the previous work [31] handling occlusion factors related to variations of view, subject's clothing and the presence of a carried item. Nevertheless, the system requires the availability of a matching template for any possible view of interest.

Kusakunniran [32] proposed a recognition model that directly extracts gait features by from raw video sequences on a spatio-temporal feature domain. They introduced the *space-time interest points* (STIPs) that represent a point of interest of a dominant walking pattern, which is used to represent the characteristics of each individual gait. The advantage of this method is that it does not require any pre-processing of the video stream (e.g., background subtraction, edge detection, human silhouettes, and so on). This makes the proposed method robust to partial occlusion caused by, among others, carrying items or hair/clothes/footwear variations over time.

1.4.2 PTZ-Based Systems

In this section, a detailed description of PTZ-based system for unconstrained biometric data acquisition is provided. Existing PTZ-based systems can be broadly divided into two groups: master–slave configuration and single-camera configuration.

Single PTZ systems feature the advantage of trivial calibration issues. Due to the zooming capability of these acquisition systems, once the object of interest in the scene is detected, the pan-tilt motor can be easily managed to keep track of it (thus ensuring that it is seamlessly centered in the video frame). However, the engineering limitations of the pan-tilt engines should be considered in the design. PTZ-motor introduces a significant delay that negatively impacts the tracking performance. When using the maximum zoom of the camera, a strong or too fine change in pan-tilt angles may easily imply a tracking failure (details are explained in Section 1.3).

Taking into account the limitations of single PTZ systems, the majority of works have focused on master–slave approaches. The typical design of this architecture is described in Section 1.4.3.1. In spite of the multiple advantages of the master–slave architecture, its feasibility is greatly dependent on the accurate inter-camera calibration (see Section 1.3.4). The lack

of depth information poses the mapping between both devices as an ill-posed problem. To that end, several approximations have been proposed to minimize the inter-camera mapping error.

Table 1.1 provides a comparison between PTZ-based systems for surveillance purposes. It must be noted that these systems were not designed specifically for acquiring biometric data. Notwithstanding, the performance and the control of the camera(s) makes them suitable to face the challenges of biometric detection/recognition at-a-distance.

As already mentioned, single PTZ systems have no particular constraints. They can be freely disposed in the working environments and do not pose any calibration issue. The work of Kumar et al. [45] and Varcheie and Bilodeau [46,48] are two examples of approaches using a single PTZ device in surveillance scenarios. Pan-tilt values are adjusted to keep the tracked subject in the central region of the camera view. In both proposals, the zoom feature is not implemented. Therefore, they could not be used for biometric recognition of traits like face or iris, but are indeed usable for gait recognition.

Tracking methods based on traditional cameras or with a fixed zoom level have the drawback of providing a variable amount of details while an object moves far/close from/to the camera. When using PTZ cameras, the consequence is that the details of the target become unrecognizable at certain distances, and a larger zoom is then required. A reduced zoom is required in the opposite condition, that is, when the high zoom level implies strong panning and tilting that might not ensure a continuous tracking. Yao et al. [47] proposed a vision-based tracking system that exploits a PTZ camera for real-time size preserving tracking. Therefore, the authors proposed to adjust, frame by frame, the zoom level of a PTZ so that the ratio of object's pixels and background's pixels is constant over time, thus preserving the resolution at which the object is tracked. Challenges are numerous: (i) varying focal length implies a loop of parametrizations; (ii) practical implementation of the relation between the system's focal length and the camera's zoom control; (iii) feature extractions is affected by the differentiation between the target's motion and the background motion caused by camera zooming. The authors exploited 3D affine shape methods for fast target feature separation/grouping and a target scale estimation algorithm based on a linear method of Structure From Motion (SFM) [49] with a detailed perspective projection model.

Even though single PTZ systems impose few constraints for calibration and can be freely mounted everywhere in the environment (refer to column

Table 1.1 A list of PTZ-based video surveillance systems

System	Architecture	Master camera	Pan/tilt est.	Cam. disp.	Zoom	Calib. marks
<i>Lu and Payandeh [33]</i>	Master–slave	Wide	Exact	Arbitrary	Yes	Yes
<i>Xu and Song [34]</i>	Master–slave	Wide	Exact	Arbitrary	Yes	No
<i>Bodor et al. [35]</i>	Master–slave	Wide	Approximated	Specific	No	Yes
<i>Scotti et al. [36]</i>	Master–slave	Catadioptric	Exact	Specific	Yes	Yes
<i>Tarhan and Altug [37]</i>	Master–slave	Catadioptric	Approximated	Specific	No	No
<i>Chen et al. [38]</i>	Master–slave	Omnidirectional	Approximated	Arbitrary	No	Yes
<i>Krahnstoever et al. [39]</i>	Master–slave	PTZ (multiple)	Exact	Arbitrary	No	No
<i>Zhou et al. [40]</i>	Master–slave	PTZ	Exact	Specific	Yes	No
<i>Yang et al. [41]</i>	Master–slave	PTZ	Exact	Arbitrary	No	Yes
<i>Del Bimbo et al. [42]</i>	Master–slave	PTZ	Approximated	Arbitrary	Yes	No
<i>Everts et al. [43]</i>	Master–slave	PTZ	Approximated	Arbitrary	No	No
<i>Liao and Chen [44]</i>	Master–slave	PTZ	Approximated	Specific	Yes	No
<i>Kumar et al. [45]</i>	Single PTZ	–	–	–	–	–
<i>Varcheie and Bilodeau [46]</i>	Single PTZ	–	–	–	–	–
<i>Yao et al. [47]</i>	Single PTZ	–	–	–	–	–
<i>Varcheie and Bilodeau [48]</i>	Single PTZ	–	–	–	–	–

Camera Disposal in Table 1.1), they also have several limitations. Master–slave systems indeed represent the most appropriate solution to address the challenges of biometric recognition in video surveillance scenarios. As described in Table 1.1, there are diverse configurations for master–slave systems. Most of them use two PTZ cameras where one acts like a master and the second one like a slave. The master camera is used as a wide-view camera, and therefore it is responsible for detecting and tracking objects in the scene. The slave receives the tracking information and tracks the objects of interest providing an alternative view of them (Yang et al. [41]).

A very complex and effective calibration procedure is proposed by Del Bimbo et al. [42]. They exploited a pre-built map of visual 2D landmarks of the wide area to support multi-view image matching. The landmarks were extracted from a finite number of images taken from non-calibrated PTZ cameras. At run-time, the features that were detected in the current PTZ camera view are matched with those of the base set in the map. The matches were used to localize the camera with respect to the scene and hence estimate the position of the target body parts. Self-calibration is regarded as the major advantage of this approach (see column *Calib. Marks* in Table 1.1). On the other hand, the dependency of stationary visual landmarks for calibration may be problematic in dynamic surveillance scenarios (in a crowded scene or in the presence of moving objects that significantly change the appearance of the scene).

In [40] and [44], the authors implemented dual-PTZ systems with high resolution images of subjects obtained by exploiting the zooming capability of the PTZ cameras (refer to the *Zoom* column). No specific biometric traits were detected, but they could be reasonably used for face detection and tracking. Other approaches using different cameras for the wide view of the scene have been proposed in the literature. Omnidirectional (Chen et al. [38]) or catadioptric cameras¹ (Scotti et al. [36,37]) have been exploited in surveillance scenarios. The added value of using an omni/catadioptric camera is that they make it feasible to seamlessly track a scene at about 360°.

Biometric recognition at-a-distance in surveillance (unconstrained) scenarios poses numerous challenges. Although the methods discussed are good candidates, none of them has been formally proved to be effective for human recognition. The following section explores the state-of-the-art

¹ A catadioptric optical system is one where refraction and reflection are combined in an optical system, usually via lenses (dioptrics) and curved mirrors (catoptrics).

and presents a collection of notable systems that achieved significantly high level of accuracy in recognition of strong biometric traits, e.g., face and iris, thus proving the feasibility and the potentials of such a line of research.

1.4.3 Face

By observing [Table 1.2](#), it is evident that most systems opt to use face for recognizing individuals in surveillance. The robustness and detectability at long distances makes the human face the biometric trait of choice for the surveillance scenario.

The work of Stillman et al. [\[51\]](#) represents one of the first attempts where multiple cameras were combined for biometric data acquisition in surveillance scenarios. Simple skin-color segmentation and color indexing methods were used to locate multiple people in a calibrated space. The proposed method demonstrated the feasibility of face detection in uncontrolled environment exploiting a multi-camera system. As we can see in [Table 1.2](#), the use of a wide-view as master camera is the most preferred option. Wide-view cameras ensure a wide coverage area thus representing the most efficient solution in surveillance scenarios. Background subtraction is the approach that is typically adopted for people detection and tracking. Hampapur et al. [\[50\]](#) and Marchesotti et al. [\[53\]](#) used both background subtraction techniques to extract the people silhouettes from the scene and used face colors to detect and track people's faces. Color-based techniques are in general computationally inexpensive but are also affected by several limitations related to illumination and occlusions. However, in surveillance scenarios with a wide-view camera, color-based detection techniques become almost the unique solution to adopt. Bernardin et al. [\[56\]](#) performed human detection using fuzzy rules to simulate the natural behavior of a human operator that allowed obtaining smoother camera handling. A KLT tracker [\[63\]](#) was used to track face's features over the time. In any case, the detection phase of the proposed tracker relied on face colors. Mian [\[57\]](#) also proposed a single PTZ-camera system to detect and track faces over the video stream by exploiting the Camshift algorithm. As already discussed in previous sections, using a single camera for detection and tracking avoids the problems related to excessive calibration. However, especially when facing with biometrics, multi-camera systems become necessary to deal with the problem of off-pose or occlusions. According to this perspective view of the problem, Amnuaykanjanasin et al. [\[55\]](#) used stereo-matching and triangulation between a pair of camera streams to estimate the 3D position

Table 1.2 A list of biometric video surveillance systems

System	Architecture	Master camera	Pan/tilt est.	Cam. disp.	I.Z.S.	Calib. marks
FACE						
<i>Hampapur et al. [50]</i>	Master–slave	Wide (multiple)	Exact	Arbitrary	No	Yes
<i>Stillman et al. [51]</i>	Master–slave	Wide (multiple)	Approximated	Specific	No	No
<i>Neves et al. [10]</i>	Master–slave	Wide	Exact	Arbitrary	Yes	No
<i>Wheeler et al. [52]</i>	Master–slave	Wide	Approximated	Arbitrary	No	Yes
<i>Marchesotti et al. [53]</i>	Master–slave	Wide	Approximated	Arbitrary	Yes	Yes
<i>Park et al. [9], [54]</i>	Master–slave	Wide	Exact	Specific	Yes	No
<i>Amnuaykanjanasin et al. [55]</i>	Master–slave	Wide	Exact	Specific	No	No
<i>Bernardin et al. [56]</i>	Single PTZ	–	–	–	–	–
<i>Mian [57]</i>	Single PTZ	–	–	–	–	–
IRIS						
<i>Wheeler et al. [58]</i>	Master–slave	Wide (multiple)	Exact	Specific	Yes	No
<i>Yoon et al. [59]</i>	Master–slave	Wide + light stripe	Approximated	Specific	Yes	Yes
<i>Bashir et al. [60]</i>	Master–slave	Wide	Exact	Specific	No	No
<i>Venugopalan and Savvides [61]</i>	Single PTZ	–	–	–	–	–
PERIOCLAR						
<i>Juefei-Xu and Savvides [62]</i>	Single PTZ	–	–	–	–	–

of a person. The proposed method still relies on color information of the skin to detect the faces. On the other side, the depth information from stereo-matching ensures good estimation of the PTZ parameters to point the camera.

Face recognition at-a-distance, although more explored than other hard biometrics, can be still considered as an unfulfilled and promising field of research in which improvements are expected in a recent future. In the following sections, the design of a typical master-slave biometric system for surveillance scenario is presented. Section 1.4.3.2 presents an innovative solution to face recognition for video surveillance while the last subsection (Section 1.4.3.3) discusses in more details a recent master-slave system, called *QUIS-CAMPI*, that exploits a novel calibration technique and automatic detection and tracking of people in-the-wild (outdoors) video surveillance scenario.

1.4.3.1 Typical Design of Master-Slave Systems

In Fig. 1.5, an overview of a typical video surveillance system aimed at biometric recognition is depicted. Such a system is a generalization of the face recognition system proposed by Wheeler et al. [52], in which two cameras cooperate in a master-slave architecture for the tracking of an individual and for the cropping of the face to achieve biometric recognition. In master-slave architectures, the hardware usually consists of:

- A Wide Field of View camera (WFOV) that acts as a *master*. By providing a wide view of the scene, such cameras allow actions like the tracking of objects/persons and detection of events of interest.
- A Narrow Field of View camera (NFOV) that acts as a *slave*. This camera provides a narrowed view of the scene and allows focusing on a single element of the scene. If such cameras provide good resolution images, the acquisition of several biometric traits will be possible (face, ear, periocular area, in descending size order).

The WFOV is responsible of providing the view of the whole scene in which the system will operate. Since it is a stationary device, a background/foreground segmentation approach is applicable and thus detects moving objects in the scene. Intrinsic and extrinsic parameters of the camera have to be determined by means of a calibration procedure, so that a mapping with the real world coordinates is provided. As well as for the wide camera, a calibration procedure is also required for the NFOV camera. Firstly, it needs to be calibrated with the WFOV camera:

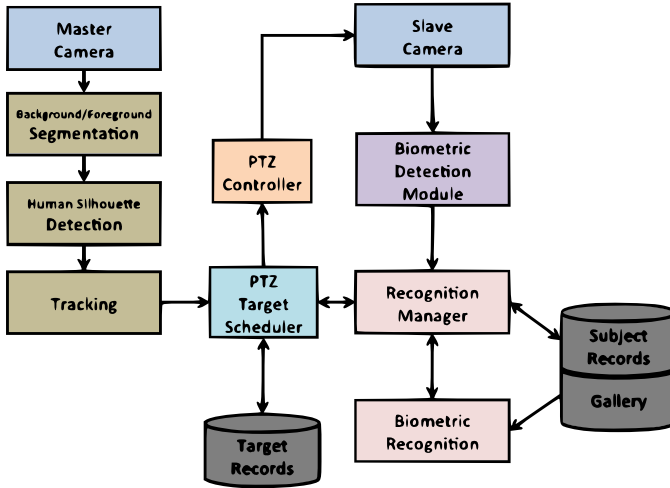


Figure 1.5 Overview of a typical video surveillance system aimed at biometric recognition. The system architecture shows a Wide-View camera and a PTZ camera operating in a master–slave configuration to detect, track and recognize biometric data in a surveillance scenario.

- The pan, tilt, and zoom values of the NFOV camera are set such that it is in its home position;
- A homography matrix is then estimated, by creating correspondences among the points in the wide scene and those in the narrow scene.

A further calibration is usually applied to the NFOV camera, in order to determine how the pan, tilt, and zoom values affect the field of view of the camera. The *zoom point* is calibrated in order to reduce the offset between several levels of zoom. The concept of a *zoom point* is introduced in [52] and indicates the pixel location that points at the same real world coordinates, even if the zoom factor is changing. Once the full calibration is accomplished, it is simple to determine the pan, tilt, and zoom values of the NFOV from the region of interest in the WFOV.

Since multiple subjects/objects may be detected and tracked in the WFOV video, a *Target Scheduler* module is needed in order to keep trace of the position of the targets (Target Records) in the video and their current state. The scheduler passes the information regarding the position of the target to a *PTZ controller* that calculates the PTZ values and zooms-in on the detected target (the zoom value may vary depending on the resolution of the video and on the size of the biometric trait). Once the biometric trait is cropped, the *Recognition Module* handles the recognition activities

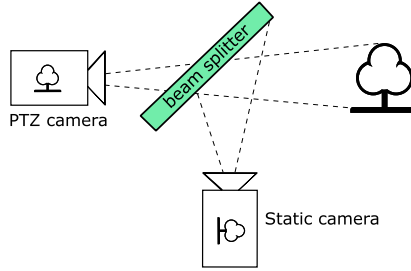


Figure 1.6 Schematic view of a multi-camera system using a beam splitter. The beamer splits the light into two so that the PTZ Camera and the Static Camera can share the same view of the scene.

(segmentation of the trait, feature extraction, matching). If a trait matches one present in the gallery, an ID number is associated, and the *Target Records* dataset is updated.

1.4.3.2 Systems Based on Logical Alignment of the Cameras

In Section 1.4.2, we described multiple surveillance systems designed to face the issues of the calibration between pairs of cameras. Ideally, if two identical cameras were mounted at the same point so that they could collect the same view of the scene, a calibration between them would not be necessary. Even though this configuration is not possible, the use of beam splitter² can mimic this process and ease the calibration between the static and PTZ cameras. Solutions that use a beam splitter [64] are perfect examples of how to approach the problem of low-calibration constraints in multi-camera systems. To better understand how the beam splitter works and how a multi-camera system can be configured, see the schematic view in Fig. 1.6.

A particularly interesting approach that relaxes the constraints of calibration was presented by Park et al. [9]. They proposed different multi-camera systems that indirectly solve the problem of sharing the same view between two cameras. In this approach, designated as *coaxial-concentric configuration*, the cameras are mounted in a way that they are all logically aligned along a shared view-axis. Therefore, it overcomes the problems related to the *epipolar geometry* (for details, refer to Section 1.3.4). A picture of the system proposed by Park et al. [9] is shown in Fig. 1.7.

² A beam splitter is an optical device that splits a beam of light into two.

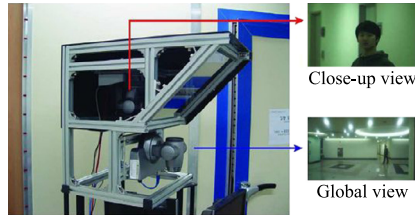


Figure 1.7 A coaxial-concentric multi-camera system. It represents the solution proposed by Park et al. [9] that uses a beam splitter in combination with a wide-view camera and a PTZ camera to achieve a coaxial configuration.

In the system of Park et al. [9] (Fig. 1.7), the multi-camera consisted of a hexahedral dark box with one of its sides tilted by 45 degrees and attached to a beam splitter. PTZ camera was configured inside the dark box and the static camera was placed outside the box. The incident beam was split at the beam splitter and captured by PTZ and static cameras to provide almost the same image to both of the cameras. All the camera axes were effectively parallel in this configuration (it enables the use of a single static camera to estimate the pan and tilt parameters of the PTZ camera). It is worth noting that such a system ensures a high level of matching between the two camera streams. However, the field of view does not completely overlap due to different camera lens and optics. As such, the authors introduced a calibration method for estimating with minimum error the pan-tilt parameters of the PTZ camera after a user-assisted one-time parametrization.

A similar solution was presented by Yoo et al. [65], where the wide-view and narrow-view cameras were combined with a beam splitter to simultaneously acquire facial and iris images. The authors combined two sensors (an image sensor for face and infra-red sensor for irises) with a beam splitter. The integrated dual-sensor system was therefore able to map rays to same position in both camera sensors, thus avoiding excessive calibration and the need of depth information.

Compared to other camera systems proposed in the literature, the approaches based on a logical alignment of the cameras feature interesting advantages:

1. World coordinates and their matching between pairs of camera streams are not involved in the calibration process;
2. Just a simple calibration, which mainly consists in a visual alignment between camera streams, is required;
3. The calibrated system can be easily deployed at a different location with no need of re-calibration.

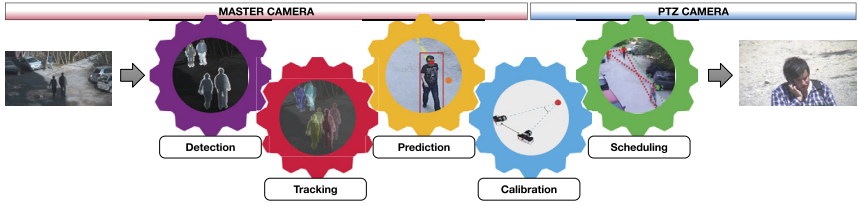


Figure 1.8 Processing chain of the QUIS-CAMPI surveillance system. A master-slave architecture is adopted for the proposed surveillance system, where the master camera is responsible for monitoring a surveillance area and providing a set of regions of interest (in this case the location of subjects face) to the PTZ camera.

Moreover, these approaches were already demonstrated to be feasible for human recognition at-a-distance (rank-1 face recognition accuracy of 91.5% in case of single person tracking with a probe set of 50 subjects against a notably larger gallery set of 10.050 subjects). However, the strict configuration required having the camera focal points aligned which might represent a limitation of the proposed approach in some video surveillance scenarios since the dimensions of the system inhibit its deployment in outdoor scenarios.

1.4.3.3 QUIS-CAMPI System

Recently, Neves et al. [10,66] have introduced an alternative solution to extend PTZ-assisted facial recognition to surveillance scenarios. The authors proposed a novel calibration algorithm [10,67] capable of accurately estimating pan-tilt parameters without resorting to intermediate zoom states, multiple optical devices or highly stringent configurations. This approach exploits geometric cues, i.e., the vanishing points available in the scene, to automatically estimate subjects height (h) and thus determine their 3D position. Furthermore, the authors have built on the work of Lv et al. [68] to ensure robustness against human shape variability during walking.

The proposed surveillance system is divided into five major modules, broadly grouped into three main phases: (i) human motion analysis, (ii) inter-camera calibration, and (iii) camera scheduling. The workflow chart of the surveillance system used for acquiring the QUIS-CAMPI dataset is given in Fig. 1.8 and described in detail afterwards.

The master camera is responsible for covering the whole surveillance area (about 650 m²) and for detecting and tracking subjects in the scene, so that it can provide to the PTZ camera a set of facial regions. In the calibra-

tion phase, the coordinates $(x_i(t), y_i(t))$ of the i th subject in the scene need to be converted to the correspondent pan-tilt angle. However, 3D positioning is required, which involves solving the following underdetermined equation:

$$\lambda \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \underbrace{\mathbf{K}_m [\mathbf{R}_m | \mathbf{T}_m]}_{:= \mathbf{P}_m} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \quad (1.2)$$

where $\mathbf{K}_m$ and $[\mathbf{R}_m | \mathbf{T}_m]$ denote the intrinsic and extrinsic matrices of the master camera, whereas $\mathbf{P}_m$ represents the camera matrix.

To address this ambiguity, the existing systems either relied on highly stringent camera disposals [52,54] or on multiple optical devices [9]. In contrast, the authors introduced a novel calibration algorithm [10] that exploited geometric cues, i.e., the vanishing points available in the scene, to automatically estimate subjects' height and thus determine their 3D position. By considering the ground as the XY plane, the Z coordinate equals the subject height, and therefore, Eq. (1.2) can be rearranged as

$$\lambda \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad h\mathbf{p}_3 + \mathbf{p}_4] \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}, \quad (1.3)$$

where $\mathbf{p}_i$ is the set of column vectors of the projection matrix $\mathbf{P}_m$.

The corresponding 3D position in the PTZ referential can then be calculated using the extrinsic parameters of the camera as

$$\begin{pmatrix} X_p \\ Y_p \\ Z_p \end{pmatrix} = [\mathbf{R}_p | \mathbf{T}_p] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (1.4)$$

The corresponding pan and tilt angles are given by

$$\theta_p = \arctan \left(\frac{X'_p}{Z'_p} \right) \quad (1.5)$$

and

$$\theta_t = \arcsin \left(\frac{Y'_p}{\sqrt{(X'_p)^2 + (Y'_p)^2 + (Z'_p)^2}} \right). \quad (1.6)$$

When multiple targets are available in the scene, a camera scheduling approach determines the sequence of observations that minimizes the cumulative transition time, in order to start the acquisition process as soon as possible and maximize the number of samples taken from the subjects in the scene. Considering that this problem has no known solution that runs in polynomial time, the authors have introduced a method capable of inferring an approximate solution in real-time [69].

1.4.3.4 Other Biometrics: Iris, Periocular, and Ear

Commercial iris recognition systems can identify subjects with extremely low error rates. However, they rely on highly restrictive capture volumes, reducing their workability in less constrained scenarios. In the last years, different works have attempted to relax the constraints of iris recognition systems by exploiting innovative strategies to increase both the capture volume and the stand-off distance, i.e., the distance between the front of the lens and the subject. Successful identification of humans using iris is greatly dependent on the quality of the iris image. To be considered of acceptable quality, the standards recommend a resolution of 200 pixels across the iris (ISO/IEC 2004), and an in-focus image. Also, sufficient near infra-red (IR) illumination should be ensured (more than 2 mW/cm^2) without harming human health (less than 10 mW/cm^2 according to the international safety standard IEC-60852-1). The volume of space in front of the acquisition system where all these constraints are satisfied is denoted as the capture volume of the system. Considering all these constraints, the design of an acquisition framework capable of acquiring good quality iris images in unconstrained scenarios is extremely hard, particularly at large stand-off distances. This section reviews the most relevant works and acquisition protocols for iris and periocular recognition at-a-distance.

In general, two strategies can be used to image iris in less constrained scenarios: (i) the use of typical cameras and (ii) the use of magnification devices. In the former, the Iris-on-the-Move system is notorious for having significantly decreased the cooperation in image acquisition. Iris images are acquired on-the-move while subjects walk through a portal equipped with NIR illuminators. Another example of a widely used commercial device is the LG IrisAccess4000. Image acquisition is performed at-a-distance; however, the user has to be directed to an optimal position so that the system can acquire an in-focus iris image. The need for fine adjustment of the user position arises from the limited capture volume of the system.

Considering the reduced size of periocular region and iris, several approaches have exploited magnification devices, such as PTZ cameras, which permit extending the system stand-off distance while maintaining the necessary resolution for reliable iris recognition. Wheeler et al. [58] introduced a system to acquire iris at a resolution of 200 pixels from cooperative subjects at 1.5 m using a PTZ camera assisted by two wide view cameras. Dong et al. [70] also proposed a PTZ-based system, and due to a higher resolution of the camera they were capable of imaging iris at a distance of 3 m with more than 150 pixels. As an alternative, Yoon et al. [59] relied on a light stripe to determine 3D position, avoiding the use of an extra-wide camera. The eagle eye system [60] uses one wide-view camera and three close-view cameras for capturing the facial region and the two irises. This system uses multiple cameras with hierarchically-ordered field of view, a highly precise pan-tilt unit, and a long focal length zoom lens. It is one of a few example systems that can perform iris recognition at a large stand-off distance (3–6 m). Experimental tests show good acquisition quality for single stationary subjects of both face and irises. On the other hand, the average acquisition time is 6.1 s which does not match with the requirements of real-time processing in non-cooperative scenarios.

Regarding periocular recognition at-a-distance, few works have been developed. In general, the periocular region is significantly less dependent on face distortions (i.e., neutral expression, smiling expression, closed eyes, and facial occlusions) than the whole face for recognition across all kinds of unconstrained scenarios. The work by Juefei-Xu and Savvides [62] is considered the only notable proposal to perform periocular recognition in highly unconstrained environments. The authors utilized the 3D generic elastic models (GEMs) [71] to correct the off-angle pose to recognize non-cooperative subjects. To deal with illumination changes, they exploited a parallelized implementation of the anisotropic diffusion image preprocessing algorithm running on GPUs to achieve real-time processing time. In their experimental analysis, they reported a verification rate of 60.7% (in the presence of facial expression and occlusions) but, more notably, they attained a 16.9% performance boost over the full face approach. Notwithstanding the encouraging results achieved, the periocular region at-a-distance still represents an unexplored field of research. The same holds for ear recognition. Ear is another interesting small biometric that has been proved relatively stable and has drawn researchers' attention recently [72]. However, like other similar biometrics (e.g., iris and

periocular), it is particularly hard to be managed in uncontrolled and non-cooperative environments. Currently, the recognition of human ears, with particular regard to challenges of at-a-distance scenarios, has not been faced yet, thus representing a promising and uncharted field of research which could reserve interesting opportunities and achievements in the recent future.

1.5 CONCLUSIONS

Biometric recognition in-the-wild is a challenging topic with numerous open issues. However, it also represents a promising research field that is still unexplored nowadays. In this chapter, we reviewed the state-of-the-art in biometric recognition systems in unconstrained scenarios discussing the main challenges as well as the existing solutions.

Despite the advances on biometric research, fully automated biometric recognition systems are still at very early stages, particularly due to the limitations of the current acquisition hardware. As such, we discussed the typical problems related to optics distortions, out-of-focus, and calibration issues of multi-camera systems. Also, particular attention was given to the system stand-off distance, which is a sensible aspect of unconstrained scenarios.

The relation between the interpupillary resolution and the stand-off distance can vary significantly among different acquisition devices. Wide-field of view cameras do not represent feasible solutions for unconstrained biometric environments. Indeed, PTZ acquisition devices have been recently proven effective to improve the performance of surveillance systems supported by biometrics. We provided a comprehensive review of the state-of-the-art master-slave surveillance systems for acquiring biometric data at-a-distance in non-cooperative environments. In particular, we provided a comparison of the most representative works in the literature highlighting their strengths and weaknesses as well as their suitability to biometric recognition in unconstrained scenarios.

We observed that face is the most mature and reliable biometric trait to be recognized at-a-distance. The detectability of this trait in challenging conditions as well as its robustness and identifiability justify the vast number of PTZ-based systems designed for acquiring face imagery in unconstrained scenarios.

Simultaneously, the recognition of iris at-a-distance represents a new field of research that has gained significant attention. State-of-the-art ac-

quisition frameworks are capable of collecting high-quality iris images up to 5 m.

Despite all these achievements, biometric recognition in uncontrolled environments is still to be achieved. We hope that this review can contribute to advance this area, particularly the development of novel acquisition frameworks.

REFERENCES

1. D. Gong, Z. Li, D. Lin, J. Liu, X. Tang, Hidden factor analysis for age invariant face recognition, in: IEEE International Conference on Computer Vision, 2013, pp. 2872–2879.
2. H. Li, G. Hua, Z. Lin, J. Brandt, J. Yang, Probabilistic elastic matching for pose variant face verification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3499–3506.
3. F. Juefei-Xu, M. Savvides, Pokerface: partial order keeping and energy repressing method for extreme face illumination normalization, in: IEEE Conference on Biometrics Theory, Applications and Systems, 2015, pp. 1–8.
4. X. Zhu, Z. Lei, J. Yan, D. Yi, S. Li, High-fidelity pose and expression normalization for face recognition in the wild, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 787–796.
5. J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
6. P. Vageeswaran, K. Mitra, R. Chellappa, Blur and illumination robust face recognition via set-theoretic characterization, *IEEE Trans. Image Process.* 22 (4) (2013) 1362–1372.
7. G.B.H.E. Learned-Miller, Labeled Faces in the Wild: Updates and New Reporting Procedures, Tech. Rep. UM-CS-2014-003, University of Massachusetts, Amherst, May 2014.
8. A.K. Jain, S.Z. Li, *Handbook of Face Recognition*, vol. 1, Springer, 2005.
9. U. Park, H.-C. Choi, A. Jain, S.-W. Lee, Face tracking and recognition at a distance: a coaxial and concentric PTZ camera system, *IEEE Trans. Inf. Forensics Secur.* 8 (10) (2013) 1665–1677.
10. J. Neves, J.C. Moreno, S. Barra, H. Proença, Acquiring high-resolution face images in outdoor environments: a master–slave calibration algorithm, in: IEEE 7th International Conference on Biometrics Theory, Applications and Systems, 2015, pp. 1–8.
11. M. McCaill, C. Norris, CCTV in Britain, Center for Criminology and Criminal Justice–University of Hull–United Kingdom, 2002, pp. 1–70.
12. J. Neves, F. Narducci, S. Barra, H. Proença, Biometric recognition in surveillance scenarios: a survey, *Artif. Intell. Rev.* (2016) 1–27.
13. Z. Zhang, Determining the epipolar geometry and its uncertainty: a review, *Int. J. Comput. Vis.* 27 (2) (1998) 161–195.
14. R. Vezzani, D. Baitieri, R. Cucchiara, People reidentification in surveillance and forensics: a survey, *ACM Comput. Surv.* 46 (2) (2013) 29:1–29:37.
15. M. De Marsico, R. Distasi, S. Ricciardi, D. Riccio, A comparison of approaches for person re-identification, in: ICPRAM, 2014, pp. 189–198.

16. S. Gong, M. Cristani, S. Yan, C.C. Loy, *Person Re-identification*, vol. 1, Springer, 2014.
17. C. Liu, H.-Y. Shum, W. Freeman, Face hallucination: theory and practice, *Int. J. Comput. Vis.* 75 (1) (2007) 115–134.
18. P. Hennings-Yeomans, S. Baker, B. Kumar, Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
19. K. Jia, S. Gong, Multi-modal tensor face for simultaneous super-resolution and recognition, in: *Tenth IEEE International Conference on Computer Vision*, vol. 2, 2005, pp. 1683–1690.
20. O. Arandjelovic, R. Cipolla, A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution, in: *ICCV 2007: Proceedings of the International Conference on Computer Vision 2007*, IEEE, 2007, pp. 1–8.
21. M. Ben-Ezra, A. Zomet, S.K. Nayar, Jitter camera: high resolution video from a low resolution detector, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2004, pp. 135–142.
22. S. Biswas, G. Aggarwal, P. Flynn, Pose-robust recognition of low-resolution face images, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 601–608.
23. W. Gong, M. Sapienza, F. Cuzzolin, Fisher tensor decomposition for unconstrained gait recognition, in: *Proceedings of Tensor Methods for Machine Learning, Workshop of the European Conference on Machine Learning*, 2013.
24. A. Dantcheva, P. Elia, A. Ross, What else does your biometric data reveal? A survey on soft biometrics, *IEEE Trans. Inf. Forensics Secur.* 11 (3) (2015) 441–467.
25. D. Matovski, M. Nixon, S. Mahmoodi, J. Carter, The effect of time on gait recognition performance, *IEEE Trans. Inf. Forensics Secur.* 7 (2) (2012) 543–552.
26. Y. Ran, Q. Zhen, R. Chellappa, T.M. Strat, Applications of a simple characterization of human gait in surveillance, *IEEE Trans. Syst. Man Cybern., Part B, Cybern.* 40 (4) (2010) 1009–1020.
27. I. Venkat, P.D. Wilde, Robust gait recognition by learning and exploiting sub-gait characteristics, *Int. J. Comput. Vis.* 91 (1) (2010) 7–23.
28. K. Moustakas, D. Tzovaras, G. Stavropoulos, Gait recognition using geometric features and soft biometrics, *IEEE Signal Process. Lett.* 17 (4) (2010) 367–370.
29. S.-U. Jung, M. Nixon, On using gait to enhance frontal face extraction, *IEEE Trans. Inf. Forensics Secur.* 7 (6) (2012) 1802–1811.
30. S.D. Choudhury, T. Tjahjadi, Gait recognition based on shape and motion analysis of silhouette contours, *Comput. Vis. Image Underst.* 117 (12) (2013) 1770–1785.
31. S.D. Choudhury, T. Tjahjadi, Robust view-invariant multiscale gait recognition, *Pattern Recognit.* 48 (3) (2015) 798–811.
32. W. Kusakunniran, Recognize gaits on spatio-temporal feature domain, *IEEE Trans. Inf. Forensics Secur.* 9 (9) (2014) 1416–1423.
33. Y. Lu, S. Payandeh, Cooperative hybrid multi-camera tracking for people surveillance, *Can. J. Electr. Comput. Eng.* 33 (3/4) (2008) 145–152.
34. Y. Xu, D. Song, Systems and algorithms for autonomous and scalable crowd surveillance using robotic PTZ cameras assisted by a wide-angle camera, *Auton. Robots* 29 (1) (2010) 53–66.

35. R. Bodor, R. Morlok, N. Papanikolopoulos, Dual-camera system for multi-level activity recognition, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, 2004, pp. 643–648.
36. G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi, C. Regazzoni, Dual camera intelligent sensor for high definition 360 degrees surveillance, *IEE Proc., Vis. Image Signal Process.* 152 (2) (2005) 250–257.
37. M. Tarhan, E. Atlug, A catadioptric and pan-tilt-zoom camera pair object tracking system for UAVs, *J. Intell. Robot. Syst.* 61 (1) (2011) 119–134.
38. C.-H. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, M. Abidi, Heterogeneous fusion of omnidirectional and PTZ cameras for multiple object tracking, *IEEE Trans. Circuits Syst. Video Technol.* 18 (8) (2008) 1052–1063.
39. N. Krahnstoeber, T. Yu, S.-N. Lim, K. Patwardhan, P. Tu, Collaborative real-time control of active cameras in large scale surveillance systems, in: *Proceedings of the Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, Marseille, France, 2008.
40. J. Zhou, D. Wan, Y. Wu, The chameleon-like vision system, *IEEE Signal Process. Mag.* 27 (5) (2010) 91–101.
41. C.-S. Yang, R.-H. Chen, C.-Y. Lee, S.-J. Lin, PTZ camera based position tracking in IP-surveillance system, in: *3rd International Conference on Sensing Technology*, 2008, pp. 142–146.
42. A.D. Bimbo, F. Dini, G. Lisanti, F. Pernici, Exploiting distinctive visual landmark maps in pan-tilt-zoom camera networks, *Comput. Vis. Image Underst.* 114 (6) (2010) 611–623.
43. I. Everts, N. Sebe, G.A. Jones, Cooperative object tracking with multiple PTZ cameras, in: *International Conference on Image Analysis and Processing*, 2007, pp. 323–330.
44. H.-C. Liao, W.-Y. Chen, Eagle-eye: a dual-PTZ-camera system for target tracking in a large open area, *Inf. Technol. Control* 39 (3) (2015).
45. P. Kumar, A. Dick, T.S. Sheng, Real time target tracking with pan tilt zoom camera, in: *Digital Image Computing: Techniques and Applications*, 2009, pp. 492–497.
46. P. Varcheie, G.-A. Bilodeau, Active people tracking by a PTZ camera in IP surveillance system, in: *IEEE International Workshop on Robotic and Sensors Environments*, 2009, pp. 93–108.
47. Y. Yao, B. Abidi, M. Abidi, 3D target scale estimation and target feature separation for size preserving tracking in PTZ video, *Int. J. Comput. Vis.* 82 (3) (2009) 244–263.
48. P.D.Z. Varcheie, G.-A. Bilodeau, Adaptive fuzzy particle filter tracker for a PTZ camera in an IP surveillance system, *IEEE Trans. Instrum. Meas.* 60 (2) (2011) 1952–1961.
49. B. Tordoff, D. Murray, Reactive control of zoom while fixating using perspective and affine cameras, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (1) (2004) 98–112.
50. A. Hampapur, S. Pankanti, A. Senior, Y.-L. Tian, L. Brown, R. Bolle, Face cataloger: multi-scale imaging for relating identity to location, in: *Proceedings of the IEEE Conference on Advance Video and Signal Based Surveillance*, 2003, pp. 13–20.
51. S. Stillman, R. Tanawongsuwan, I. Essa, Tracking multiple people with multiple cameras, in: *International Conference on Audio and Video-Based Biometric Person Authentication*, 1999.
52. F. Wheeler, R. Weiss, P. Tu, Face recognition at a distance system for surveillance applications, in: *Proceedings of the Fourth IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–8.

53. L. Marchesotti, S. Piva, A. Turolla, D. Minetti, C.S. Regazzoni, Cooperative multi-sensor system for real-time face detection and tracking in uncontrolled conditions, in: *Electronic Imaging 2005*, International Society for Optics and Photonics, 2005, pp. 100–114.
54. H.-C. Choi, U. Park, A. Jain, PTZ camera assisted face acquisition, tracking & recognition, in: *Proceedings of the Fourth IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–6.
55. P. Amnuaykanjanasin, S. Aramvith, T.H. Chalidabhongse, Face tracking using two cooperative static and moving cameras, in: *IEEE International Conference on Multimedia and Expo*, 2005, pp. 1158–1161.
56. K. Bernardin, F. van de Camp, R. Stiefelhagen, Automatic person detection and tracking using fuzzy controlled active cameras, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
57. A. Mian, Realtime face detection and tracking using a single pan, tilt, zoom camera, in: *23rd International Conference in Image and Vision Computing New Zealand*, 2008, pp. 1–6.
58. F.W. Wheeler, G. Abramovich, B. Yu, P.H. Tu, et al., Stand-off iris recognition system, in: *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–7.
59. S. Yoon, H.G. Jung, J.K. Suhr, J. Kim, Non-intrusive iris image capturing system using light stripe projection and pan-tilt-zoom camera, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–7.
60. F. Bashir, P. Casaverde, D. Usher, M. Friedman, Eagle-eyes: a system for iris recognition at a distance, in: *IEEE Conference on Technologies for Homeland Security*, 2008, pp. 426–431.
61. S. Venugopalan, M. Savvides, Unconstrained iris acquisition and recognition using COTS PTZ camera, *EURASIP J. Adv. Signal Process.* 38 (2010).
62. F. Juefei-Xu, M. Savvides, Unconstrained periocular biometric acquisition and recognition using COTS PTZ camera for uncooperative and non-cooperative subjects, in: *IEEE Workshop on Applications of Computer Vision*, 2012, pp. 201–208.
63. B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, vol. 2, 1981, pp. 674–679.
64. J.Y. Han, K. Perlin, Measuring bidirectional texture reflectance with a kaleidoscope, in: *Proceedings of SIGGRAPH*, vol. 22, 2003, pp. 741–748.
65. J.-H. Yoo, B.J. Kang, A simply integrated dual-sensor based non-intrusive iris image acquisition system, in: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 113–117.
66. J. Neves, G. Santos, S. Filipe, E. Grancho, S. Barra, F. Narducci, H. Proença, Quis-campi: extending in the wild biometric recognition to surveillance environments, in: *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*, 2015, pp. 59–68.
67. J.C. Neves, J.C. Moreno, H. Proença, A master-slave calibration algorithm with fish-eye correction, in: *Mathematical Problems in Engineering*, Hindawi Publishing Corporation, 2015.
68. F. Lv, T. Zhao, R. Nevatia, Self-calibration of a camera from video of a walking human, in: *16th International Conference on Pattern Recognition*, vol. 1, 2002, pp. 562–567.

69. J.C. Neves, H. Proença, Dynamic camera scheduling for visual surveillance in crowded scenes using Markov random fields, in: 12th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2015, pp. 1–6.
70. W. Dong, Z. Sun, T. Tan, A design of iris recognition system at a distance, in: Chinese Conference on Pattern Recognition, CCPR, 2009, pp. 1–5.
71. U. Prabhu, J. Heo, M. Savvides, Unconstrained pose-invariant face recognition using 3D generic elastic models, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (10) (2011) 1952–1961.
72. H. Chen, B. Bhanu, Human ear recognition in 3D, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (4) (2007) 718–737.