QUIS-CAMPI: An Annotated Multi-biometrics Data Feed From Surveillance Scenarios

João Neves[1,*], Juan Moreno[2], Hugo Proença[3]

[1]IT - Instituto de Telecomunicações, University of Beira Interior
[2]Department of Computer Science, University of Beira Interior
[3]IT - Instituto de Telecomunicações, University of Beira Interior
[*]jcneves@ubi.pt

Abstract: The accuracy of biometric recognition in unconstrained scenarios has been a major concern for a large number of researchers. Despite such efforts, no system can recognize in a fully automated manner human beings in totally wild conditions, such as in surveillance environments. In this context, several sets of degraded data have been made available to the research community, where the reported performance by state-of-the-art algorithms is already saturated, suggesting that these sets do not reflect faithfully the conditions in such hard settings. To this end, we introduce the QUIS-CAMPI data feed, comprising samples automatically acquired by an outdoor visual surveillance system, with subjects on-the-move and at-a-distance (up to 50 m). Also, we supply a high-quality set of enrollment data. When compared to similar data sources, the major novelties of QUIS-CAMPI are: 1) biometric samples are acquired in a fully automatic way; 2) it is an *open* dataset, i.e., the number of probe images and enrolled subjects grow on a daily basis; and 3) it contains multi-biometric traits. The ensemble properties of QUIS-CAMPI ensure that the data span a representative set of covariate factors of real-world scenarios, making it a valuable tool for developing and benchmarking biometric recognition algorithms capable of working in unconstrained scenarios.

1. Introduction

Over the last years, biometric research has been evolving towards the development of systems that work in adverse conditions. This trend has been mainly supported by the increasing number of commercial systems relying on biometric recognition, and, at the same time, in the interest in extending these systems to unconstrained scenarios. These efforts have proven fruitful, as in the case of the Verilook Surveillance system [1], where persons are recognized on-the-move using face features. However, this system is confined to indoor scenarios and demands a large amount of enrollment data. In fact, the recognition of humans in totally wild conditions, such as in surveillance environments, is still to be accomplished [2], making biometric recognition in the wild a highly popular topic, and, at the same time, one of the most ambitious goals for the research community.

Biometric datasets are an important asset to push forward the state-of-the-art recognition performance. As an example, we highlight the evolution of face recognition datasets, which have moved towards more challenging conditions, as novel algorithms surpass the challenges

Fig. 1. Illustration of the six covariate factors in the QUIS-CAMPI dataset. *The use of an automated master-slave system for acquiring facial imagery in a fully non-cooperative and covert manner assures an effective representativeness of the covariates of biometric recognition in the wild.*

of the hardest sets. This fact is particularly evident in the case of Labeled Faces in the Wild (LFW) [3], which was originally proposed to meet the satisfactory performance of state-of-the-art algorithms in addressing the typical covariates of facial recognition. The high variability of the data has caught researchers attention, who have progressively advanced the robustness of face recognition algorithms (87% accuracy in 2009 [4] to 95% accuracy in 2015 [5]). LFW has paved the way for biometric recognition in the wild, and fostered the development of even more challenging datasets. Nonetheless, as noted by Klare *et* al. [6], one explanation for unconstrained face recognition being still far from solved is that the LFW and similar datasets are not fully unconstrained. To close this gap, Klare *et* al. [6] introduced the IJB-A dataset, which follows the spirit of LFW but includes high variability in pose. However, even this challenging dataset does not encompass the complete set of covariate factors present in real surveillance scenarios, as the majority of the images were not acquired on-the-move and in an automated manner, reducing the levels of blur caused either by motion or incorrect focusing.

In this paper, we provide a tool to bridge the gap between surveillance and biometric recognition, by announcing the QUIS-CAMPI data feed, whose acronym derives from Latin and summarizes its goals: 'Quis' stands for 'Who is' and 'Campi' refers to a delimited space. Hence, this set aims at fostering the development of biometric recognition systems that work outdoors, in fully unconstrained and covert conditions. To this end, we designed an automated master-slave surveillance system to capture both full body video sequences and high-resolution head samples of subjects in a parking lot. The particularities of the surveillance system permit the continuous acquisition of novel biometric samples that are supplied to the dataset after being manually screened and associated to the corresponding gallery subjects. This singularity is the rationale for considering that QUIS-CAMPI is the first *open* biometric dataset, since the number of biometric samples grow on a daily basis. For the same reason, we argue that the proposed set can be considered a data feed, where researchers can obtain novel biometric data captured in a realistic surveillance scenario. In spite of the multiple advantages of the QUIS-CAMPI dataset, it also raises important questions on evaluation and privacy. As such, the impact of the continuous supply of new probes has been carefully planned by introducing dataset versioning (refer to section 4.3).

Regarding privacy, while the data collection has been authorized by the Portuguese data protection authority, we believe that the surveillance system could be used in a public space without compromising privacy by adopting a watchlist-based recognition, i.e., the goal is not to identify the person, but determine if the probe data corresponds to a subject in the watchlist.

It should be also mentioned that a subset of the QUIS-CAMPI dataset has already been published before [7] to promote the ICB-RW (International Challenge on Biometric Recognition in the Wild) competition. When compared to ICB-RW, QUIS-CAMPI has the following advantages: 1) the ICB-RW challenge provided only 10 face images from each of the 90 subjects, whereas QUIS-CAMPI contains more than 3,000 images from 320 subjects; 2) ICB-RW was based on a single evaluation metric, whereas QUIS-CAMPI provides a comprehensive evaluation protocol both for the verification and identification modes; 3) QUIS-CAMPI provides a comprehensive evaluation along the proposed evaluation protocols; and 4) QUIS-CAMPI provides a continuous feed of biometric samples, as well as a version control strategy for obtaining new data.

Contributions: When compared to the existing biometric datasets, the QUIS-CAMPI data feed has four major novelties: 1) biometric traits are automatically acquired by a master-slave surveillance system in a fully non-cooperative and covert manner. This allows the data to be acquired at-a-distance (up to 50 m) and on-the-move, and assures an effective representativeness of the covariates of biometric recognition in the wild; 2) it is an *open* dataset, i.e., new samples are continuously and automatically being added to the dataset and supplied to the research community. This singularity inhibits biased performance estimation - usually caused by parameter adjustment in the test set - without the burden of sequestered test data; 3) it is surveillance representative, i.e., the probe images truly encompass all the singularities of surveillance environments, and thus, advances in the recognition accuracy of these data have a direct impact on the deployment of a fully automated biometric recognition surveillance system.

The remainder of this paper is organized as follows: Section 2 overviews the datasets for assessing the recognition performance in these environments. Section 3 describes the master-slave surveillance system developed for data acquisition. A detailed description of the announced dataset is given in Section 4. Section 5 describes the evaluation protocol and compares the results attained by state-of-the-art face recognition algorithms in the QUIS-CAMPI and LFW datasets. Finally, Section 7 concludes the paper.

2. Biometric Datasets

About 25 years ago, biometric recognition emerged as an interesting topic, leading to the development of many novel algorithms, usually validated in small, non-representative and proprietary databases, according to distinct evaluation protocols. To meet the growing demands for objective evaluation tools, sets of biometric samples comprising different covariate factors were introduced as a solution. The ORL database of faces [16], the AR face database [17] and the Yale face database were pioneer sets on face recognition, while FERET [18] was the first benchmark on this topic. Despite their valuable contribution in providing objective and trustworthy tools for assessing recognition performance, these sets soon became outdated as

Table 1. Comparative analysis between the datasets particularly devised for studying unconstrained biometric recognition. *Datasets are compared with respect to the number of subjects available and the key covariate factors of recognition in the wild: expression (E), occlusion (O), illumination (I), pose (P), motion-blur (M) and out-of-focus (F). Also, the key aspects to ensure that the data realistic result from real-world scenarios are also included. The abbreviations of these aspects refer to non-cooperative (NC), on-the-move (OM), at-a-distance (AD), outdoor (OU) and automated image acquisition (AA).*

	Number of subjects	Covariate Factors	NC	OM	AD	OU	AA	Observations
XM2VTS [8]	295	E, I, P		✓				A multi-modal database comprising face images, video sequences and speech recordings acquired at one month intervals.
BANCA [9]	26	E, I, P		✓				A database of face videos comprising twelve recordings per subject were acquired under controlled, uncontrolled and adverse conditions.
FRGC [10]	688	E, I				✓		Four controlled still images, two uncontrolled still images and a 3D face model.
FRVT 2006 [11]	$>$ 35 000	E, I				✓		The first independent performance benchmark for 3D face recognition technology. Also, it comprises still frontal face images acquired under controlled and uncontrolled illumination.
SC-FACE [12]	130	E, I, P			✓	✓		Facial imagery acquired in an indoor surveillance scenario using five video surveillance cameras of various qualities.
LDHF-DB [13]	100	I, F			✓	✓		This set comprises both visible and near-infrared face images at distances of 60m, 100m, and 150m acquired outdoors.
LFW [3]	5749	E, O, I, P	✓			✓		The first database of face photographs designed for the study of unconstrained face recognition.
IJB-A [6]	500	E, O, I, P,M	✓			✓		Similar in spirit to the LFW dataset, but containing high-variability in pose.
YouTube Faces [14]	1595	E, O, I, P, M	✓	✓				A database of face videos designed for the study of unconstrained face recognition in videos.
Choke Point [15]	25	E, O, I, P, M	✓	✓			✓	A database of face videos acquired indoors in a non-cooperative manner.
QUIS-CAMPI	268 (v1) 320 (v2)	E, O, I, P, M, F	✓	✓	✓	✓	✓	The first data feed of biometric samples automatically acquired by an outdoor surveillance system, with subjects on-the-move and at-a-distance.

novel algorithms reported almost ideal accuracy on these data.

These improvements fostered the development of more challenging datasets, such as the CMU PIE [19], the Multi-PIE [20], the XM2VTS [8] and the BANCA [9] databases, comprising biometric samples with significant variations in illumination, pose and expression. Also, different challenges were introduced for assessing the accuracy of state-of-the-art face recognition methods in less constrained scenarios (e.g., the Face Recognition Grand Challenge (FRGC) [10] and the Face Recognition Vendor Test (FRVT 2006)).

Aiming at providing more realistic data, the research has advanced towards the acquisition of unconstrained samples along the diverse biometric traits, such as iris [21], periocular [22, 23] and face [24]. Regarding face recognition, LFW was the first database particularly devised for studying face verification in the wild and was, therefore, responsible for promoting the development of more robust algorithms (an increment of 10% in the recognition accuracy in last years), as well as for fostering the emergence of more challenging collections of data (e.g., PubFig [25], FaceScrub [26], IJB-A [6] and Disguise and Makeup Faces Database [27]).

Simultaneously, still to video modality has also gained increasing attention leading to the development of novel datasets and biometric challenges on this topic. The video challenge portion of the Multiple Biometrics Grand Challenge (MBGC) contained subjects walking towards the camera and non-frontal footage of subjects performing an activity. Later, the Point and Shoot Face Recognition Challenge (PaSC) was introduced, comprising unconstrained video sequences of subjects performing multiple activities outdoors. The YouTube Faces database [14] contains unconstrained recordings obtained from the internet, and it was particularly designed for studying the problem of unconstrained face recognition in videos. On contrast, the SC-FACE [12] and the ChokePoint [15] datasets were originally intended to provide data acquired in realistic indoor surveillance scenarios. Regarding surveillance scenarios, the PETS [28], i-LIDS [29], CAVIAR [30] datasets and the VISOR [31] repository comprise video sequences of pedestrians in realistic surveillance scenarios. Even though the low resolution of data inhibits its use for face recognition purposes, it has been showed that the fusion with gait information can significantly increase the performance [32, 33].

The evolution of gait databases over time has varied over three major factors: the number of subjects and sequences, data covariates (e.g., clothing, carried items, speed), and the acquisition scenario (e.g., indoor, outdoor). The SOTON large database [34] was the first set to comprise over 100 subjects and has contributed to study the impact of inter-subject variation on gait recognition. The USF dataset [35] was introduced to provide a representative dataset for benchmarking gait recognition in challenging conditions, and investigating which covariates significantly degrade the gait recognition performance. CASIA B [36] is a commonly used gait database containing large view variations, as well as variations in clothing and carrying status. For this reason, this set is usually exploited for the evaluation of cross-view gait recognition and for evaluating the impact of clothing and carrying status on the performance of gait recognition methods. While the remaining databases did not exceed 200 subjects, OU-ISIR LP [37] has overcome the remaining sets by providing gait sequences from more than 4000 subjects with a wide age range. In spite of not containing any covariates, it is useful for estimating the performance of the gait recognition with high statistical reliability.

A comparative analysis between state-of-the-art databases concerning unconstrained face recognition is given in Table 1. It is interesting to note that despite these sets comprise highly challenging biometric data, most of them were manually captured by human operators and they still lack several crucial covariate factors of surveillance environments, such as motion-blur.

3. QUIS-CAMPI Data Acquisition System

For the acquisition of QUIS-CAMPI, we rely on a master-slave surveillance system capable of acquiring face imagery of subjects at-a-distance and on-the-move (Fig. 2). Even though this type of configuration is not novel and it has also been used in surveillance scenarios [38], our approach has several singularities, making the system particularly suitable for working in real-world conditions.

When compared with the existing master-slave systems, which were particularly devised for the acquisition of biometric data at-a-distance [39, 40, 41], our approach has two major advantages: 1) a novel calibration algorithm avoids the use of extra optical devices [39] or stringent configurations for the cameras [40, 41]; 2) camera scheduling is performed using a general graph model that determines in real-time the best tour to acquire the targets in the scene and can be easily customized to incorporate several prioritization rules, avoiding the use of manually defined rules [40].

The proposed surveillance system is divided in five major modules, broadly grouped in three main phases: 1) human motion analysis; 2) inter-camera calibration; 3) camera scheduling. The workflow chart of the surveillance system used for acquiring the QUIS-CAMPI dataset is given in Fig. 2 and described in detail in the next sections.

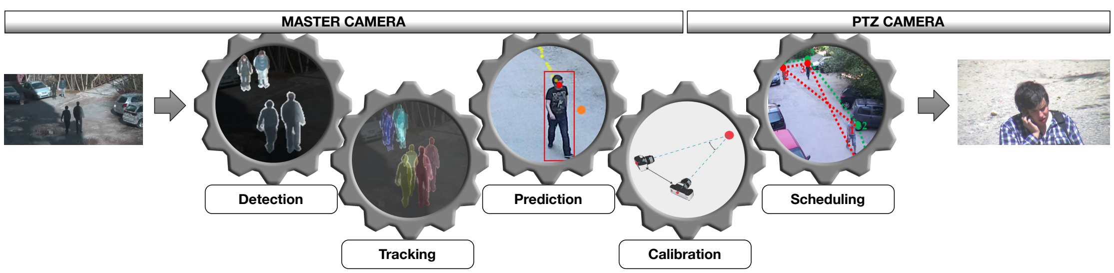

Fig. 2. Processing chain of the QUIS-CAMPI surveillance system. *A master-slave architecture is adopted for the proposed surveillance system, where the master camera is responsible for monitoring a surveillance area and providing a set of interest regions (in this case the location of subjects face) to the PTZ camera.*

As illustrated in Fig. 2, the rationale behind the surveillance system is to use the PTZ camera as a foveal sensor, i.e., the video stream obtained from the wide camera is analyzed to obtain the location of subjects' head, so that the PTZ camera can image the facial region at a high-magnification state. In the former phase, the master camera is responsible for detecting and tracking multiple subjects in the surveillance area. In every frame, a background subtraction algorithm prunes the search area inspected by a pre-traihed human shape detector, whose output instantiates a tracking algorithm. Multi-person tracking is achieved by using multiple instances of the algorithm running simultaneously. Subsequently, the tracking record of each subject is analyzed for inferring their position some seconds ahead. This step is particularly important to counterbalance the time offset introduced by

Fig. 3. Illustrative example of the biometric data available in QUIS-CAMPI. *For each subject of the database distinct biometric traits are acquired during enrollment, comprising soft biometrics, full body imagery (a) and a 3D model (b). Subsequently, high-resolution face images (c) are automatically collected each time a subject enters the surveillance area in a non-cooperative way. Note that these data are acquired under varying lighting and weather conditions, at different times of the day, while subjects are on-the-move and at-a-distance.*

the mechanical delay of PTZ devices. In the calibration module, the image coordinates in the master camera referential need to be converted to the correspondent pan-tilt angle. To this end, we relied on a novel calibration algorithm [42] that exploits geometric cues (the vanishing points available in the scene) to automatically estimate subjects' height and thus determine their 3D position (refer to [42] for additional details). Finally, the calibration module allows the PTZ camera to determine the sequence of observations which 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 not a known solution that runs in polynomial time, we relied on a method capable of inferring an approximate solution in real-time [43].

Table 2. ***List of the soft biometric traits collected during enrollment.***

Trait	Labels
Age	ℕ
Height	ℕ
Weight	ℕ
Sex	Male, Female
Ethnicity	Caucasian, African, Hispanic, Asian, Indian
Skin Color	White, Tanned, Oriental, Black
Hair Color	None, Black, Brown, Red, Blond, Grey, Dyed
Hair Length	None, Shaven, Short, Medium, Long
Facial Hair Color	None, Black, Brown, Red, Blond, Grey
Facial Hair Length	None, Stubble, Mustache, Goatee, Full beard
Hair Style	None, Straight, Curly, Wavy, Frizzy

4. Description of the QUIS-CAMPI Dataset

When planning the QUIS-CAMPI dataset, we had two main concerns: 1) to acquire biometric data of subjects in a real surveillance scenario, covertly, on-the-move and at-a-distance; and 2) to provide multiple biometric enrollment data to perceive the advantages of using media collection [44] for identifying humans in the wild. For that purpose, we collected multiple biometric samples that where organized into two distinct groups: 1) enrollment data; 2) probe data. In the former, we have enrolled volunteers who have provided written authorization for image acquisition and distribution. The enrollment was conducted in a indoor controlled scenario and the following samples were collected: soft biometrics, full-body imagery and a 3D face model. In the latter, the data were acquired by the system described in Section 3, while subjects walked throughout a surveillance area. Probe samples comprise high-resolution face images automatically captured by the PTZ camera. It is important to note that the large majority of subjects use this area in their normal routine, which ensures a faithful representation of surveillance covariates. Figure 3 illustrates the biometric data available for each subject.

4.1. Enrollment Data

Enrollment data provide good quality samples acquired indoor: soft biometrics, full-body imagery and a 3D face model.

Soft biometrics: Eleven types of soft biometric labels were registered for each subject. The full list is presented in Table 2 and the rationale behind the choice of these features was their discernibility at a distance and the discrimination power reported in the study of Tome *et* al. [45]. The distribution of each trait with respect to the labels adopted is depicted in Fig. 4.

Full-body shots: A high-resolution image of the person body was acquired at three different angles (frontal, left-side and right-side). Also, the intrinsic and extrinsic parameters of the camera were registered, along with five keypoints of the body in the frontal view. These data can be used to infer real-world measurements of body components (e.g. height and face

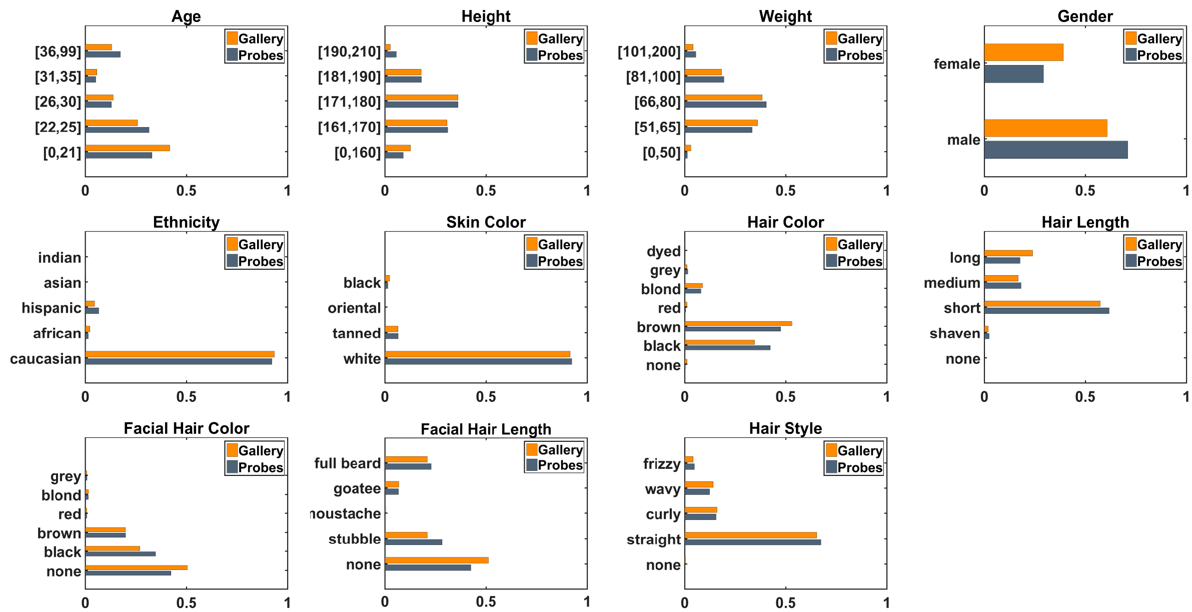

Fig. 4. ***QUIS-CAMPI statistics.*** *A set of statistics was collected for distinct types of biometric samples: 1) distribution of the soft traits along the enrollment data (denoted as gallery) and probe data; 2) distribution of the interpupillary distance in the probe images; 3) distribution of the tracking sequences width collected outdoors; and 4) distribution of the number of days elapsed between the acquisition of probe data and the enrollment process.*

metrology).

3D face model: A set of images acquired at different viewing angles was used to construct a textured 3D model of face using Visual-SFM [46].

4.2. Probe Data

Fully unconstrained biometric samples are the key novelty of the QUIS-CAMPI dataset and comprehend face images automatically acquired by the PTZ camera.

High-resolution facial shots: The master-slave surveillance system described in Section 3 is used to automatically acquire high-resolution face images of the enrolled subjects, while they walk throughout the surveillance area. Considering that not all the acquired data contain the facial region (e.g., incorrect human detection or tracking) and that the surveillance rig is not able to distinguish between enrolled and non-enrolled subjects, the data are manually screened before being supplied to the database. Also, the face location of the interest subject in the image is provided as metadata. These annotations are determined by a state-of-the-art face detection algorithm [47] and cross-verified manually. On average, the interpupillary distance of a face image is 116 px with a standard deviation of 35 px, and about 99% of the images have an interpupillary distance higher than 60 px (the minimum resolution required for commercial face recognition engines).

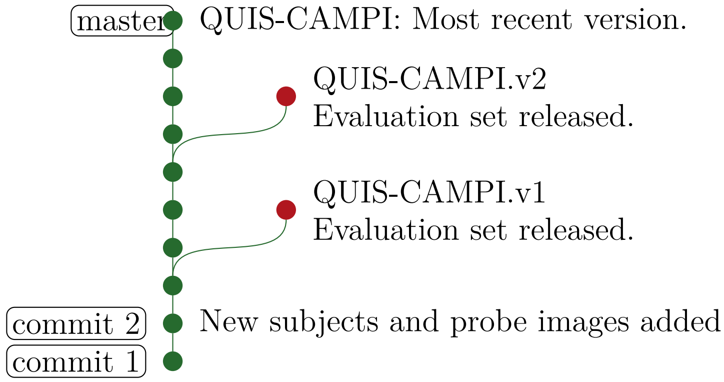

Fig. 5. History graph of the QUIS-CAMPI data feed using a version control software. *A git repository was used to deploy new samples acquired by the data feed (represented by the master branch), while maintaining static evaluation sets released at much lower rate (represented by branches). This strategy permits researchers to access any state of the QUIS-CAMPI data feed for development purposes, while it also ensures that algorithms can be compared by reporting performance on the different evaluation set versions.*

4.3. Database Versioning

The automated acquisition of biometric samples and their regular deployment to the dataset is the reason for denoting QUIS-CAMPI as a data-feed and at the same it is one of the key novelties of this tool. Moreover, this singularity is the rationale to argue that QUIS-CAMPI is the first *open* dataset, which is particularly advantageous to avoid inappropriately fitting classifiers to the final test data. Despite the advantages of this choice, it also introduces significant challenges that have to be carefully addressed to ensure that the performance reported in this dataset can be compared in a practical and fair manner.

To this end, we relied on *git* - one of the most commonly used version control systems - to organize the QUIS-CAMPI data feed in two distinct types of branches: 1) the master branch comprises the most updated version of the entire biometric data; 2) the evaluation branches encompass a former snapshot of the master branch plus the evaluation files defined according to the evaluation protocol of QUIS-CAMPI (refer to Section 5.1). This structure is depicted in Figure 5, where the advantages of this strategy can be easily perceived. First, the version control capabilities allow users to navigate through any state of the QUIS-CAMPI data feed using the master branch, which is useful to obtain new biometric samples without the burden of redownloading the entire set. Second, the evaluation branches are static and independent of any updates on master branch, allowing researches to compare their approaches by referring to a specific evaluation branch.

4.4. Database Availability

Regarding the dataset structure, the file names correspond to the acquisition date in the following format: Y<a>M<b>D<c>h<d>m<e>s<f>, where a, b, c, d, e, and f denote the acquisition year, month, day, hour, minute and second, respectively. The correspondences between the files and enrolled subjects, as well as the soft biometric traits, are provided in a relational database, which is deployed as a backup SQL file. For convenience, we include a view in the database that eases the access to biometric data using simple SQL queries. For

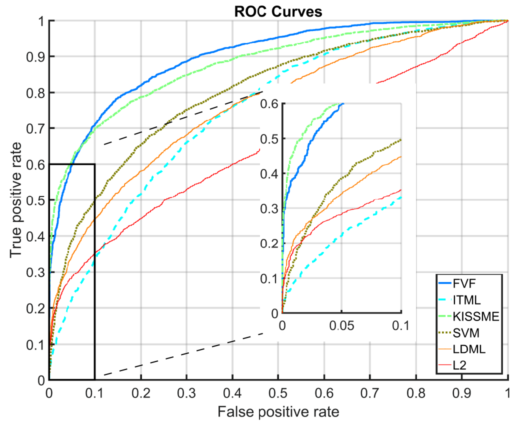

Fig. 6. Verification performance attained in the QUIS-CAMPI dataset under the unrestricted setting with no outside data.

additional information on how to get and use the dataset, please refer to the QUIS-CAMPI web site [1].

5. Experimental Evaluation

In this section, we introduce the evaluation protocols that should be adopted for reporting the algorithms performance in the QUIS-CAMPI set. We believe that the proposed guidelines for the different recognition modalities are adequate for the majority of biometric recognition algorithms. However, as in the recent case of the updated guidelines of LFW [24], additional protocols may be included in the future to meet novel requirements.

5.1. Evaluation Protocol

Having in mind the main purpose of QUIS-CAMPI, i.e., to provide an objective tool for assessing the performance of biometric recognition algorithms in surveillance scenarios, we introduce two evaluation protocols for the two recognition modalities: 1) verification and 2) identification.

5.1.1. Verification: Regarding the verification paradigm, we adopt the protocol defined in LFW [3, 24], which is an objective, simple and well established way of assessing face verification algorithms. Accordingly, the PTZ face images are used to form pairs of matched images (positive pairs) and mismatched images (negative pairs) organized in two groups: 1) model selection and algorithm development; and 2) performance reporting. In the first group, random pairs are used as training (2200 pairs) and test (1000 pairs) sets. This group is particularly intended for tuning algorithm parameters, and thus, preventing the bias introduced by adjusting the method to the final evaluation set. In the second group, 10 splits - containing 300 positive and negative pairs of PTZ face images - are built for evaluating algorithms performance using leave-one out validation. In the training phase, two distinct paradigms are available: image-restricted, where only the training split pairs can be used,

[1] http://quiscampi.di.ubi.pt

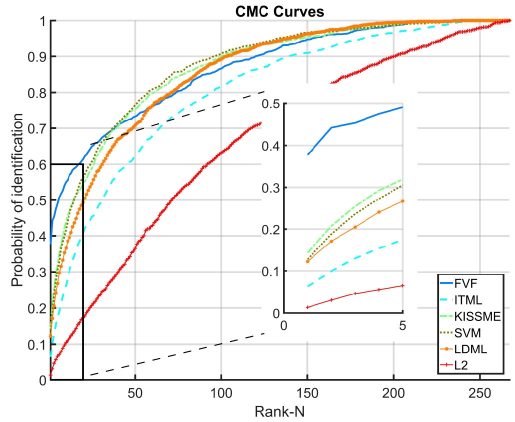

Fig. 7. ***Identification performance attained in the QUIS-CAMPI dataset under the closed-set setting with no outside data.***

Table 3. ***Description of the QUIS-CAMPI evaluation protocols under the verification paradigm.*** *As in LFW [24], the protocols are defined according to the use of the image-restricted or unrestricted setting, and to the use of additional training data.*

	Unrestricted, No Outside Data	Image-restricted, Label Free Outside Data	Unrestricted, Label Free Outside Data	Unrestricted, With QUIS-CAMPI Biometric Data	Unrestricted, With Labelled Outside Data
Identity information in training images	✓		✓	✓	✓
Annotations in training images		✓	✓	✓	✓
External images		✓	✓	✓	✓
Binary label for external image pairs				✓	✓
Identity label for external images				✓	✓

and unrestricted, where identity information is provided, which allows forming additional training pairs. In addition, one can increase the algorithms robustness by exploiting metadata from outside of QUIS-CAMPI or external training images. The ensemble of training paradigms for face verification in the QUIS-CAMPI dataset is listed in Table 3. During the test phase, algorithms must be evaluated using mean classification accuracy (ACC) over the 10 splits. If possible, the ROC curve and its corresponding Area Under Curve (AUC) should be also reported. The metrics that shall be adopted under the different recognition paradigms are described in Table 4.

5.1.2. Identification: Under the identification paradigm, a probe image must be matched against all gallery subjects. This task can be approached in two distinct manners: 1) assuming that all probe images correspond to one subject in the gallery (closed-set recognition); and 2) assuming that not all the probe identities are represented in the gallery (open-set recognition).

Probe data comprise the PTZ face images, while the frontal mugshots acquired during enrollment are used as gallery. Probes are divided in two mutually exclusive sets: 1) training

Table 4. Description of the performance metrics adopted for the different evaluation paradigms.

Paradigm	Setting	Performance Metric
Verification	Image-restricted	ROC plot AUC ACC
	Unrestricted	ROC plot AUC ACC
Identification	Closed-set	CMC plot AUC Rank-1 accuracy
	Open-set	ROC plot AUC ACC

(containing 70% of the subjects); and 2) test (containing 30% of the subjects). Besides, two distinct sets of gallery subjects are provided. One comprising all the available subjects of the dataset (closed-set) and the other containing just 80% of the subjects. At last, in order to avoid bias in the subject separation, this process is repeated 10 times defining 10 independent evaluation sets. At the training phase three distinct paradigms can be exploited: 1) no outside data; 2) QUIS-CAMPI biometric data; and 3) biometric data from external sources. At the test phase, algorithms must be evaluated using leave-one out validation and the average performance should be reported according to the metrics listed in Table 4.

5.2. Results and Discussion

In order to perceive the robustness of state-of-the-art face recognition algorithms to the QUIS-CAMPI degradation factors, we report results for six face recognition algorithms under the unrestricted setting with no outside data (for verification) and under the closed-set setting with no outside data (for identification) using QUIS-CAMPI.v1 . The tests are conducted using three metric learning based approaches (ITML [48], LDML [4] and KISSME [49]), the Fisher vector faces (FVF) [50], a support vector machine (SVM) and the L2 distance between SIFT features. The justification for choosing these methods is twofold: 1) they are well established face recognition methods with competitive performance reported on LFW; and 2) the source code is freely available allowing, which ensures a reliable assessment of their performance. In these experiments, the QUIS-CAMPI metadata are used to provide the algorithms with 256x256 cropped images of the facial region. Considering that, apart from FVF, all the evaluated algorithms expect pre-processed face descriptors, we adopt the widely used strategy of Guillaumin *et* al. [4], which exploits the SIFT descriptors [51] of 9 automatically detected face landmarks [52].

In accordance with the QUIS-CAMPI evaluation protocol, the methods are optimized using the first group of pairs, i.e., the parameters are determined based on maximum recognition accuracy obtained. Subsequently, the performance of each method is determined using leave one-out validation in the 10 splits of the second group. Aiming at providing a comparative performance analysis between QUIS-CAMPI and state-of-the-art benchmarks, the performance of these algorithms is also assessed in LFW.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited.
Content may change prior to final publication in an issue of the journal. To cite the paper please use the doi provided on the Digital Library page.

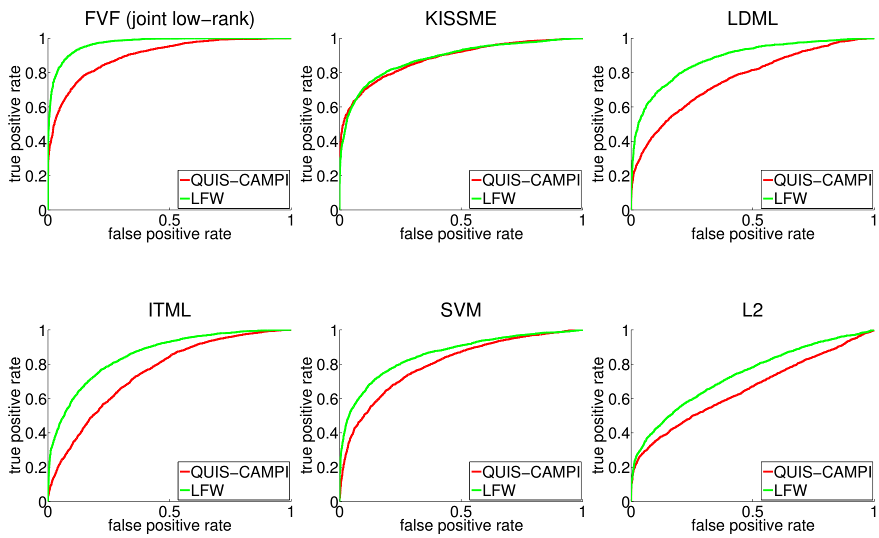

Fig. 8. Comparison between the recognition performance observed per algorithm in the QUIS-CAMPI and LFW datasets, under the unrestricted setting.

The results are reported using the ROC curves (for verification) and CMC curves (for identification), as well as, its corresponding AUC. Fig. 6 and Fig. 7 depict the ROC curves and CMC curves obtained in the QUIS-CAMPI dataset under the unrestricted setting and the closed-set setting, respectively. The comparison between the ROC curves in the QUIS-CAMPI and LFW sets is given in Fig. 8. Additionally, these results are summarized in Table 5, with respect to the AUC values.

Regarding the algorithms performance, it is important to note that FVF clearly outperforms the remaining approaches, whereas the use of L2 norm on the SIFT descriptors - without using any additional learning metric - is clearly behind the remaining methods. While the latter is not surprising, in the first the case the performance gap can be explained by the use of automatic keypoint detection. Due to the large variation in pose, it is likely that the use of such strategy in completely unconstrained environments fails, yielding thus, incorrect face descriptors. On contrast, holistic approaches, such as FVF, may be a more adequate solution for addressing these challenging conditions.

Regarding the comparison of the recognition performance per dataset, the results obtained sustain our claim that QUIS-CAMPI is much more challenging than LFW. This is particularly evident in Fig. 8, since none of the state-of-the-art algorithm was able to improve the results obtained in the LFW when addressing the QUIS-CAMPI dataset.

Besides, the verification accuracy achieved for QUIS-CAMPI justifies the need for such a dataset, since much more has to be done to close the gap between surveillance and biometric recognition.

Additionally, the discriminability of soft biometric data was assessed by determining the probability of observing the correct identity within the top K ranks. For evaluation, the age, gender, skin color and hair color of the subject were inferred automatically from the probe images. The predictions were obtained by training a classifier on features extracted from the face region. The CMC curve illustrated in Fig. 9 summarizes the performance obtained. Soft biometric traits are not capable to individually authenticate the subject, which explains its

This article has been accepted for publication in a future issue of this journal, but has not been fully edited.
Content may change prior to final publication in an issue of the journal. To cite the paper please use the doi provided on the Digital Library page.

Table 5. Recognition performance on the QUIS-CAMPI and LFW datasets under the verification and identification modalities. *Algorithms performance was determined on both datasets according to the QUIS-CAMPI evaluation protocol under the unrestricted setting (verification) and the closed-set setting (identification). The comparative analysis between the results of QUIS-CAMPI and LFW confirms the additional challenges of the proposed dataset. Performance was assessed using AUC of the ROC curve and its corresponding standard deviation.*

	QUIS-CAMPI				LFW			
	Verification		Identification		Verification		Identification	
Algorithm	AUC (%)	ACC (%)	AUC (%)	ACC (%)	AUC (%)	Rank-1 (%)	AUC (%)	Rank-1 (%)
FVF [50]	89.9 ± 2.8	81.7 ± 3.0	97.1 ± 0.8	90.2 ± 1.6	88.5	26.4	86.8	37.7
ITML [48]	74.9 ± 4.0	68.4 ± 2.9	85.7 ± 1.8	77.5 ± 2.0	93.9	4.9	80.3	6.4
KISSME [49]	88.0 ± 3.9	79.2 ± 4.0	88.2 ± 1.7	80.9 ± 2.0	95.8	10.7	86.8	14.3
LDML [4]	76.4 ± 3.1	69.1 ± 2.2	88.1 ± 1.4	79.7 ± 2.0	96.9	3.9	85.6	12.2
SVM [53]	80.2 ± 2.8	72.8 ± 2.6	85.8 ± 1.3	78.1 ± 1.0	94.9	9.4	87.3	12.7
L2	65.6 ± 4.3	62.5 ± 3.6	73.7 ± 1.5	67.3 ± 1.8	70.5	3.0	66.7	1.3

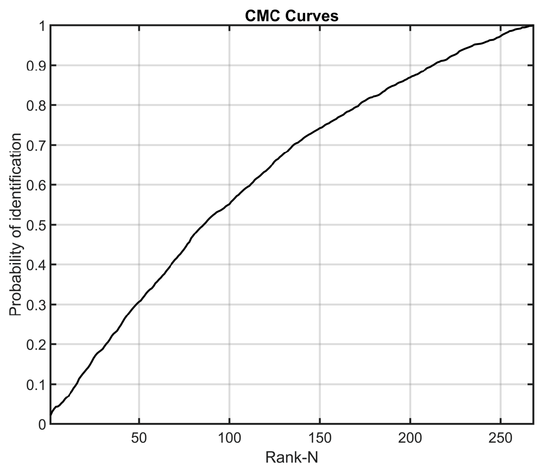

Fig. 9. Recognition performance obtained solely with soft biometric traits.

the poor identification rate. Nevertheless, it should be stressed that they can significantly increase performance when combined with hard biometric traits, and the results obtained are encouraging for these type of traits.

6. Additional biometric data

Even though face imagery and soft biometrics regard the principal data of the QUIS-CAMPI data feed, there are additional traits that can be explored by prospective users. Accordingly, this set also comprises gait recordings both on the enrollment and probe sets. Regarding

a)  b) c)

Fig. 10. Illustrative example of the additional biometric data available in QUIS-CAMPI. *For each subject of the database the gait sequence has been acquired during enrollment (a) and each time a subject entered the surveillance area, comprising the tracking sequences (b) and their corresponding foreground (c).*

the enrollment data, subjects were asked to walk naturally during 5m while being filmed at three different angles. This was performed twice to obtain six different viewing angles of the gait sequence.

Regarding the probe data, gait sequences were recorded by the master camera and correspond to each PTZ face acquisition. These tracking sequences comprise a set of videos automatically acquired by the master camera while the person is passing in the surveillance area. On average, each sequence has a resolution of 73 x 114 px with a standard deviation of 43 px for both dimensions. Additionally, the output of the background subtraction algorithm for each tracking sequence has also been stored. An illustrative example of these data is depicted in Fig. 10.

7. Conclusion

In this paper, the QUIS-CAMPI data feed has been introduced, comprising biometric samples automatically acquired outdoors, at-a-distance, on-the-move and in a fully non-cooperative manner.

A key difference between QUIS-CAMPI and related sets is that samples are automatically acquired in a real surveillance scenario, i.e., in a covert way and without any human intervention in the process. This assures that the collected data completely encompass the set of covariate factors of real-world scenarios. Additionally, to the best of our knowledge, QUIS-CAMPI is the first *open* biometric data feed, i.e., new probes and subjects are continuously being added as the system automatically acquires more data. This fact is beneficial for inhibiting bias introduced by the complete knowledge of the entire dataset, particularly in the case of open-set recognition. Moreover, QUIS-CAMPI comprises multi-biometric traits, permitting the exploitation of multi-modal recognition strategies.

To objectively justify the need for QUIS-CAMPI, six face verification algorithms were evaluated both in QUIS-CAMPI and LFW under the image-restricted and unrestricted settings. The conclusions were twofold: 1) the algorithms accuracy in the QUIS-CAMPI is much lower than in LFW, which confirms that the proposed dataset is more challenging; 2) the state-of-the-art algorithms are still far from optimal recognition rates, and substantial improvements in the recognition technology should be made before saturate the announced set.

Acknowledgments

This work is supported by '*FCT - Fundação para a Ciência e Tecnologia*' (Portugal) through the project 'UID/EEA/50008/2013', the research grant 'SFRH/BD/92520/2013', and the funding from 'FEDER - QREN - Type 4.1 - *Formação Avançada*', co-founded by the European Social Fund and by national funds through Portuguese 'MEC - *Ministério da Educação e Ciência*'.

8. References

[1] NeuroTechnology. "Verilook surveillance", `http://www.neurotechnology.com/verilook-surveillance.html`, accessed: 2015-10-03.

[2] J. Klontz, A. Jain. "A case study of automated face recognition: The boston marathon bombings suspects", *IEEE Computer*, **46(11)**, pp. 91–94, (2013).

[3] G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller. "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments", *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, (Marseille, France2008).

[4] M. Guillaumin, J. Verbeek, C. Schmid. "Is that you? metric learning approaches for face identification", *IEEE International Conference on Computer Vision*, pp. 498–505, (2009).

[5] X. Zhu, Z. Lei, J. Yan, D. Yi, S. Z. Li. "High-fidelity pose and expression normalization for face recognition in the wild", *IEEE Conference on Computer Vision and Pattern Recognition*, (2015).

[6] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, A. K. Jain. "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a", *IEEE Conference on Computer Vision and Pattern Recognition*, (2015).

[7] J. Neves, H. Proenca. "Icb-rw 2016: International challenge on biometric recognition in the wild", *International Conference on Biometrics*, pp. 1–6, (2016).

[8] J. Matas, et al. "Comparison of face verification results on the xm2vtfs database", *International Conference on Pattern Recognition*, volume 4, pp. 858–863, (2000).

[9] K. Messer, et al. "Face authentication test on the banca database", *International Conference on Pattern Recognition*, volume 4, pp. 523–532, (2004).

[10] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek. "Overview of the face recognition grand challenge", *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pp. 947–954, (2005).

[11] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, M. Sharpe. "Frvt 2006 and ice 2006 large-scale experimental results", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32(5)**, pp. 831–846, (2010).

[12] M. Grgic, K. Delac, S. Grgic. "Scface - surveillance cameras face database", *Multimedia Tools and Applications*, **51(3)**, pp. 863–879, (2011).

[13] H. Maeng, S. Liao, D. Kang, S.-W. Lee, A. K. Jain. "Nighttime face recognition at long distance: Cross-distance and cross-spectral matching", *Asian Conference on Computer Vision*, pp. 708–721, (2012).

[14] L. Wolf, T. Hassner, I. Maoz. "Face recognition in unconstrained videos with matched background similarity", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 529–534, (2011).

[15] Y. Wong, S. Chen, S. Mau, C. Sanderson, B. C. Lovell. "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition", *IEEE Computer Vision and Pattern Recognition Workshops*, pp. 81–88, (IEEE, 2011).

[16] F. S. Samaria. "Face recognition using hidden markov models", Ph.D. thesis, University of Cambridge, (1994).

[17] A. Martinez, R. Benavente. "The ar face database", Technical Report 24, (1998).

[18] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss. "The feret evaluation methodology for face-recognition algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22(10)**, pp. 1090–1104, (2000).

[19] T. Sim, S. Baker, M. Bsat. "The cmu pose, illumination, and expression (pie) database", *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46–51, (2002).

[20] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker. "Multi-pie", *IEEE International Conference on Automatic Face Gesture Recognition*, pp. 1–8, (2008).

[21] H. Proença, S. Filipe, R. Santos, J. Oliveira, L. Alexandre. "The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32(8)**, pp. 1529–1535, (2010).

[22] C. Padole, H. Proença. "Compensating for pose and illumination in unconstrained periocular biometrics", *International Journal of Biometrics*, **5(3/4)**, pp. 336 – 359, (2013).

[23] F. Juefei-Xu, M. Savvides. "Unconstrained periocular biometric acquisition and recognition using cots ptz camera for uncooperative and non-cooperative subjects", *IEEE Workshop on Applications of Computer Vision*, pp. 201–208, (2012).

[24] G. B. H. E. Learned-Miller. "Labeled faces in the wild: Updates and new reporting procedures", Technical Report UM-CS-2014-003, University of Massachusetts, Amherst, (2014).

[25] N. Kumar, A. Berg, P. Belhumeur, S. Nayar. "Attribute and simile classifiers for face verification", *IEEE International Conference on Computer Vision*, pp. 365–372, (2009).

[26] H.-W. Ng, S. Winkler. "A data-driven approach to cleaning large face datasets", *IEEE International Conference on Image Processing*, pp. 343–347, (2014).

[27] T. Y. Wang, A. Kumar. "Recognizing human faces under disguise and makeup", *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, (2016).

[28] L. Li, T. Nawaz, J. Ferryman. "Pets 2015: Datasets and challenge", *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 1–6, (2015).

[29] T. Wang, S. Gong, X. Zhu, S. Wang, *European Conference Computer Vision*, chapter Person Re-identification by Video Ranking, pp. 688–703, (2014).

[30] R. Fisher. "Caviar dataset", `http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/`, (2005).

[31] R. Vezzani, R. Cucchiara. "Video surveillance online repository (visor): An integrated framework", *Multimedia Tools Appl.*, **50(2)**, pp. 359–380, (2010).

[32] M. Hofmann, S. M. Schmidt, A. Rajagopalan, G. Rigoll. "Combined face and gait recognition using alpha matte preprocessing", *International Conference on Biometrics*, pp. 1–8, (New Delhi, India2012).

[33] D. Muramatsu, H. Iwama, Y. Makihara, Y. Yagi. "Multi-view multi-modal person authentication from a single walking image sequence", *International Conference on Biometrics*, pp. 1–8, (2013).

[34] J. D. Shutler, M. G. Grant, M. S. Nixon, J. N. Carter, *On a Large Sequence-Based Human Gait Database*, pp. 339–346, (2004).

[35] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. G. ther, K. Bowyer. "The humanid gait challenge problem: Data sets, performance, and analysis", *IEEE Transactions of Pattern Analysis and Machine Intelligence*, **27(2)**, pp. 162–177, (2005).

[36] S. Yu, D. Tan, T. Tan. "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition", *Proceedings of the International Conference on Pattern Recognition*, pp. 441–444, (2006).

[37] H. Iwama, M. Okumura, Y. Makihara, Y. Yagi. "The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition", *IEEE Transactions on Information Forensics and Security*, **7(5)**, pp. 1511–1521, (Oct. 2012).

[38] J. C. Neves, G. Santos, S. Filipe, E. Grancho, S. Barra, F. Narducci, H. Proença. "Quis-campi: Extending in the wild biometric recognition to surveillance environments", *ICIAP 2015 Workshops*, pp. 59–68, (2015).

[39] 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 Transactions on Information Forensics and Security*, **8(10)**, pp. 1665–1677, (2013).

[40] F. Wheeler, R. Weiss, P. Tu. "Face recognition at a distance system for surveillance applications", *IEEE International Conference on Biometrics: Theory Applications and Systems*, pp. 1–8, (2010).

[41] H.-C. Choi, U. Park, A. Jain. "Ptz camera assisted face acquisition, tracking & recognition", *IEEE International Conference on Biometrics: Theory Applications and Systems*, pp. 1–6, (2010).

[42] J. C. Neves, J. C. Moreno, S. Barra, H. Proença. "Acquiring high-resolution face images in outdoor environments: A master-slave calibration algorithm", *IEEE International Conference on Biometrics: Theory, Applications and Systems*, (2015).

[43] J. C. Neves, H. Proença. "Dynamic camera scheduling for visual surveillance in crowded scenes using markov random fields", *IEEE Conference on Advanced Video and Signal Based Surveillance*, (2015).

[44] L. Best-Rowden, H. Han, C. Otto, B. Klare, A. Jain. "Unconstrained face recognition: Identifying a person of interest from a media collection", *IEEE Transactions on Information Forensics and Security*, **9(12)**, pp. 2144–2157, (2014).

[45] P. Tome, J. Fierrez, R. Vera-Rodriguez, M. Nixon. "Soft biometrics and their application in person recognition at a distance", *IEEE Transactions on Information Forensics and Security*, **9(3)**, pp. 464–475, (2014).

[46] C. Wu, S. Agarwal, B. Curless, S. Seitz. "Multicore bundle adjustment", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3057–3064, (2011).

[47] X. Zhu, D. Ramanan. "Face detection, pose estimation, and landmark localization in the wild", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2879–2886, (2012).

[48] J. V. Davis, B. Kulis, P. Jain, S. Sra, I. S. Dhillon. "Information-theoretic metric learning", *ICML*, pp. 209–216, (2007).

[49] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, H. Bischof. "Large scale metric learning from equivalence constraints", *IEEE International Conference on Computer Vision and Pattern Recognition*.

[50] K. Simonyan, O. M. Parkhi, A. Vedaldi, A. Zisserman. "Fisher Vector Faces in the Wild", *British Machine Vision Conference*, (2013).

[51] D. G. Lowe. "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, **60(2)**, pp. 91–110, (2004).

[52] M. Everingham, J. Sivic, A. Zisserman. "Hello! my name is... buffy – automatic naming of characters in TV video", *Proceedings of the British Machine Vision Conference*, (2006).

[53] C.-C. Chang, C.-J. Lin. "Libsvm: A library for support vector machines", *ACM Transactions on Intelligence Systems and Technology*, **2(3)**, pp. 1–27, (2011).