ICB-RW 2016: International Challenge on Biometric Recognition in the Wild

João Neves, Hugo Proença
IT - Instituto de Telecomunicações
University of Beira Interior, Covilhã, Portugal
jcneves@ubi.pt,hugomcp@di.ubi.pt

Abstract

Biometric recognition in totally wild conditions, such as the observed in visual surveillance scenarios has not been achieved yet. The ICB-RW competition was promoted to support this endeavor, being the first biometric challenge carried out in data that realistically result from surveillance scenarios. The competition relied on an innovative master-slave surveillance system for the acquisition of face imagery at-a-distance and on-the-move. This paper describes the competition details and reports the performance achieved by the participants algorithms.

1. Introduction

While biometric recognition in constrained setups has been proving to be a successful case of pattern recognition, a long road is still ahead before of the deployment of automata capable of recognizing human beings in totally unconstrained conditions and without requiring cooperation from the subjects [5]. Under such type of less constrained conditions, the quality of the resulting data substantially decreases, and no recognition algorithm is able to handle the data covariates of such large magnitude (e.g., varying pose, severe occlusions, poor resolution and blur). In recent years, different approaches have been introduced to address the typical challenges of biometric recognition in such *wild* conditions, and - simultaneously - an increasing number of so-called *degraded* datasets has been released. Among such sets, the Labelled Faces in the Wild (LFW) [6] is well known for fostering the advances on the accuracy of face recognition algorithms and for promoting the development of even more challenging databases. Nonetheless, one reason for unconstrained biometric recognition being still far from solved is that the LFW and similar datasets are not fully unconstrained. In fact, most sets comprise manually captured data and thus they fail at in providing a faithful representation of biometric traits acquired by fully automated surveillance systems.

Perhaps the most obvious application of unconstrained biometrics is the deployment of such type of systems in the conditions that are nowadays associated to outdoor visual surveillance scenarios. In this kind of setups, images are captured from large distances and the acquired data have limited discriminability, even when using high-resolution cameras [3]. Several authors have suggested the use of PTZ (Pan Tilt Zoom) based systems, which acquire high-resolution imagery on arbitrary scene locations using master-slave architectures. However, inter-camera calibration is considered a major bottleneck of this data acquisition paradigm, and most master-slave systems rely either on 2D-based approximations or stringent configurations between the cameras to alleviate this issue, which, in turn, prevents their deployment in outdoor surveillance scenarios. Aiming at resolving this shortcoming, we have recently introduced the QUIS-CAMPI system [10], a master-slave surveillance system capable of imaging subjects up to 50m in outdoor scenarios.

The singularities of the QUIS-CAMPI surveillance system permit its deployment in real surveillance scenarios, and enable the automated acquisition of face imagery of subjects at-a-distance and on-the-move. These data are the basis of the ICB-RW challenge [1], whose primary goal is fostering the development of biometric recognition algorithms capable of working in real surveillance scenarios. The main contributions of ICB-RW are the following: 1) biometric traits are automatically acquired by a master-slave surveillance system in a fully non-cooperative and covert manner; 2) 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 development of a fully automated biometric recognition surveillance system.

The remainder of this paper is organized as follows. Section 2 summarizes the most relevant approaches that have attempted the recognition of humans in unconstrained scenarios. The competition impact, the dataset used and protocol adopted in the challenge are described in section 3, 4 and 5, respectively. The recognition performance attained

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

978-1-5090-1869-7/16/$31.00 ©2016 IEEE

by the algorithms submitted by ICB-RW participants is presented and discussed in section 6. Finally, section 7 outlines the major conclusions of the competition.

2. Biometric Recognition in the Wild

Reliable biometric identification in controlled conditions is considered to be solved, and the focus is now placed on unconstrained scenarios, where research efforts proved to be fruitful [9]. Different authors have introduced recently robust strategies to address the A-PIE factors, and almost ideal performance was reported in unconstrained biometric datasets [19], such as the LFW [6]. However, these achievements contrast with the results reported in real surveillance scenarios [5].

This incongruence can be explained by the fact that state-of-the-art datasets do not faithfully represent the covariates of surveillance scenarios. Insufficient illumination, blur and low resolution at a long distance are regarded as the most important challenges for face recognition in these scenarios. Currently, the use of magnification devices is broadly accepted as the most convenient and efficient strategy to acquire high-resolution biometric samples at a distance. Wheeler *et* al. [17] developed a master-slave system capable of acquiring high-resolution face images up to 20m. However, a specific configuration between the cameras had to be assumed to ease inter-calibration, restraining its workability in surveillance scenarios.

As an alternative, Park *et* al. [14] proposed the use of a beam splitter to determine the 3D position of the subject. Also, a fully automated acquisition system was introduced and the authors reported a 95% rank-1 identification rate for close-up shots automatically acquired from moving subjects at a distance of 15 meters. Nevertheless, the restrictive configuration between the optical devices restrains its use in outdoor environments.

2.1. Datasets

Aiming at providing more realistic biometric datasets, the research has advanced towards the acquisition of unconstrained samples along the diverse biometric traits, such as iris [16], periocular [13] and face [6]. Regarding face recognition, LFW [6] 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 five years), as well as for fostering the emergence of more challenging collections of data. It is interesting to note that despite these sets comprise highly challenging biometric data, they still lack several crucial covariates of surveillance environments, such as high variability in pose and motion-blur.

First, the use of an automatic face detector is a chief limitation in state-of-the-art benchmark datasets, since pose and other confounding factors are not represented. The IJB-A dataset [4] was introduced to address this problem, containing manually localized face images with large variability in expression, occlusion, illumination and pose. However, most images were acquired by human operators, reducing motion-blur and under-exposure levels.

Second, images are not acquired at-a-distance and on-the-move. The LDHF-DB [8] attempted to solve this problem by introducing a set with cross-distance (up to 150m) and cross-spectral data (visible and near infrared). The UTK-LRHM [18] is similar in spirit to LDHF-DB, but the data were acquired on-the-move with dynamic magnification devices. Although both sets were acquired in an outdoor environment, the acquisition system depends on subjection cooperation, restraining their applicability for studying face recognition in surveillance scenarios.

On contrast, the data acquired by the QUIS-CAMPI surveillance system is different from the remaining sets by comprising high-resolution face imagery automatically acquired in outdoor environments, while subjects are at-a-distance and on-the-move. These particularities ensure the collected biometric data faithfully represent all the covariates of surveillance scenarios.

3. ICB-RW Competition

The ICB-RW competition took place from September to December, 2015. The website had more than 700 visitors from 24 countries. There was a total of 19 registrations in the competition. Most of the registered users were members of academic institutions, whereas a smaller number was from and private companies. At the end, nine participants have submitted their executable to be evaluated in the ICB-RW sequestered data.

4. ICB-RW Dataset

4.1. Acquisition System

Considering the disadvantages of acquiring biometric data at a long distance with typical surveillance cameras, we have relied on an innovative PTZ-based architecture capable of acquiring high-resolution face imagery up to 50m, the QUIS-CAMPI system. In this system, a surveillance camera monitors a wide surveillance area using automated human detection and tracking modules, while a PTZ camera acquires close-up shots of the head. The use of online human height estimation for determining subjects 3D position is regarded as the major novelty and advantage of the system, as it does not depend on extra optical devices (e.g., mirrors [14]) or stringent configurations between cameras (e.g., side-by-side cameras [17]). Additional details on the QUIS-CAMPI system can be found at [10, 11]. Considering the flexibility of this strategy, the QUIS-CAMPI system was deployed in a typical surveillance scenario (about

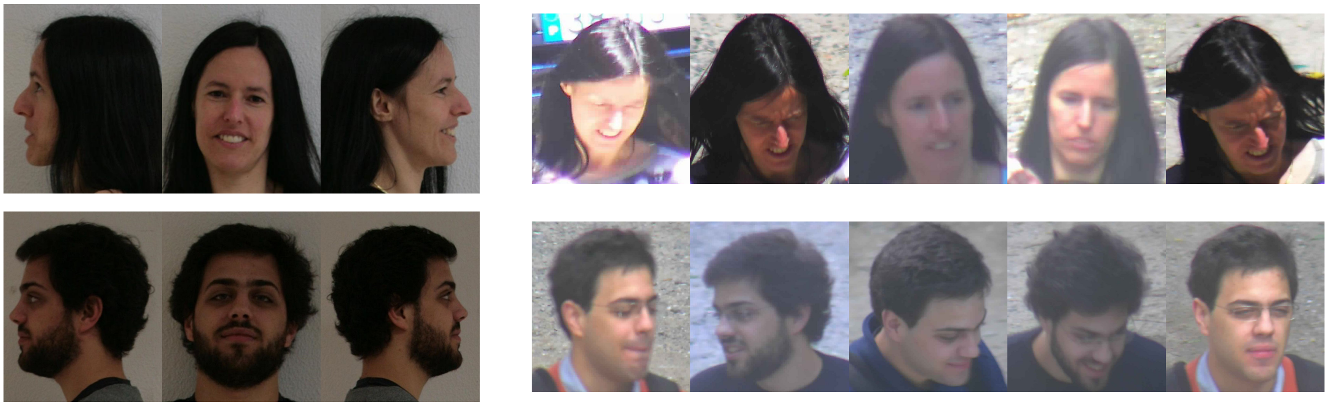


Figure 1. **Example of the gallery and probe data of two subjects in the ICB-RW dataset.** Gallery data comprise one frontal view and two side view images of the subject (left), while five probe images were provided to the participants for algorithm training (right).

650 m^2) and was used for acquiring biometric samples of non-cooperative subjects while they are at-a-distance and on-the-move. These data were acquired at a daily basis and a subset of these images was used as ICB-RW probe data. Also, multiple biometric samples of the probe subjects were also acquired in a controlled manner to constitute the gallery data.

4.2. Data Summary

The ICB-RW evaluation set comprises biometric data from 90 volunteers who have provided written authorization for image acquisition and distribution. The data were organized into two sets: 1) gallery data; and 2) probe data. Gallery data comprise 3 images of the subject's head acquired in a controlled scenario: 1) frontal image; 2) left-side image; and 3) right-side image. Probe data were automatically acquired by the QUIS-CAMPI surveillance in a covert and non-cooperative manner while subjects walk throughout the surveillance area. For each subject, ten probe images were selected aiming at preserving the following degradation factors throughout the dataset: 1) variations in pose; 2) variations in expression; 3) varying illumination; 4) occlusions; and 5) blur.

Moreover, the face location of the interest subject in both probe and gallery images was provided as a bounding box. These data were automatically inferred from a state-of-the-art head-landmark localization method [20] and corrected manually. The annotations were provided in the dataset as metadata. Fig. 2 illustrates the gallery and probe images of two subjects in the dataset.

5. ICB-RW Protocol

For evaluation purposes, the probe images were randomly divided into two subsets comprising five images each. Participants were provided with gallery data and one probe data subset, while the other subset was kept as sequestered data. During the competition period, participants were expected to build and train their algorithms using the publicly available data, while the final evaluation would be performed by the competition organizers upon algorithm submission.

5.1. Evaluation Metrics

The algorithm performance was determined by the area under curve (AUC) of the cumulative match characteristic (CMC) curve. For each probe image P_i , a rank-K list was constructed by selecting the K most similar gallery subjects according to the algorithm scores. The CMC curve relates the percentage of correct identification for all probe images with the size K of the rank-K list.

It is worth noting that the evaluation was conducted using the sequestered subset of probe images, which are disjoint from the ones used by the participants and ensure a non-biased evaluation.

6. Results and Discussion

In this section, we report the results attained in the ICB-RW challenge. The performance of the nine algorithms submitted to ICB-RW is presented in Fig. 2 and summarized in Table 1 with respect to the rank-1 and rank-5 identification rate and the AUC of the CMC curve. Also, a brief description provided by the authors is included.

The competition results are useful for two major reasons: 1) perceive how far research has come in fully automated human recognition; 2) provide insight into what are the most adequate strategies for extracting discriminant information from extremely degraded images.

In the former, it is interesting to note that the best performing approaches reach relatively high levels of accuracy, particularly if more than one match is considered (85.3% accuracy by retrieving 5% of the database). Despite these results suggest that biometrics is quite close of recognizing humans in completely unconstrained scenarios in an immediate future, there exist some issues that should be taken into account. First, it would be important to study the impact of gallery size on the algorithms performance, as the number of gallery subjects of ICB-RW was rather limited when compared with real-world scenarios. Second, it is worth noting that in this challenge, a closed-set identification was adopted. Again, it is important to study the impact on the performance when considering open-set identification.

In the latter, we use the general description of the methods, their performance in the ICB-RW competition and the correlation between the similarity scores (Fig. 3) to conclude about the most valuable strategies to address highly degraded biometric data. Regarding the four best performing approaches, it is interesting to note that all of them rely on deep learning, in particular they use deep convolutional neural networks. The similarity between them is also corroborated by high correlation values presented in Fig. 3. On the other hand, the approaches using typical descriptor-based features such as SIFT [7] and LBP [12] have attained poor identification rates. Even though these approaches compensated for illumination variations, it is likely that these descriptors are not invariant to the extreme variations in blur, occlusion and expression of unconstrained scenarios. Regarding the high variability in pose, this factor is addressed in most approaches with the use of face landmark localization and frontalization techniques, whose performance in unconstrained scenarios has been significantly improved [20, 2].

In addition, we analyzed the most easy-to-identify and hard-to-identify probes and subjects to provide additional insight into the most discriminant and confounding factors of unconstrained biometric recognition. For this purpose, we relied on the similarity scores of the three best performing approaches to determine the percentage of gallery subjects that have to be retrieved for correctly identifying a probe image. The percentages of methods were fused using the maximum value and the four best and worst performing images are depicted in Fig. 4a) and 4b), respectively. By averaging the results per subject we also determined the most easy-to-identify and hard-to-identify subjects, which are depicted in Fig. 5. It is interesting to note that high variability in pose greatly affects the methods performance, whereas almost frontal poses with neutral expression provided the best results. On the other hand, it is worth noting that the subjects sharing the same age group and facial features, such as facial hair style, were the most likely to be incorrectly classified.

7. Conclusions

The limitations of state-of-the-art unconstrained biometric datasets and the lack of benchmark tests on biometric recognition in surveillance scenarios were the rationale

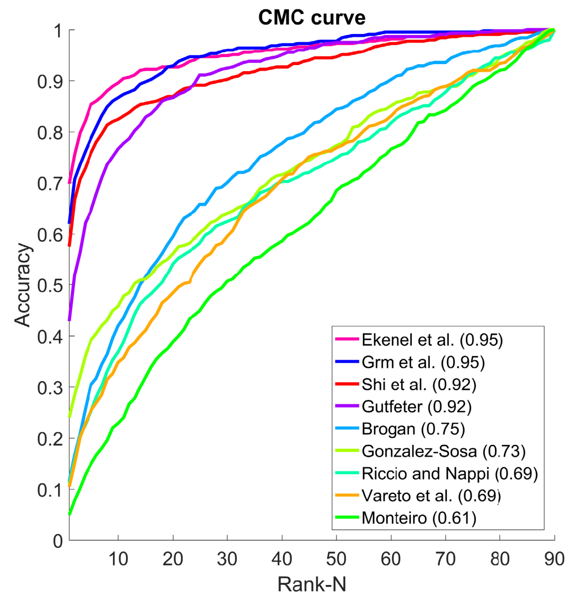


Figure 2. **Algorithms performance in the ICB-RW competition.** Algorithms were evaluated in the sequestered probe data and the CMC curves were used to assess the identification performance.

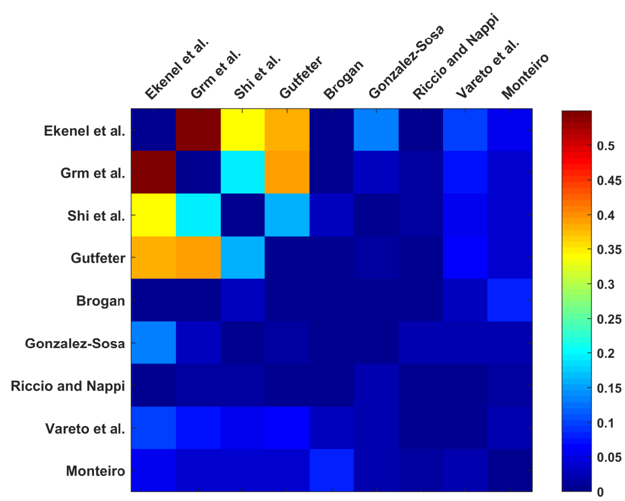


Figure 3. **Correlation between the nine algorithms submitted to ICB-RW.** The similarity scores obtained during the evaluation phase were used to determine the correlation between the different algorithms submitted to ICB-RW.

behind the ICB-RW challenge. The use of a fully automated surveillance system capable of covertly imaging subjects at-a-distance and on-the-move allowed to assess how far research has come in human recognition in totally unconstrained scenarios. The results were relatively positive, however much more has to be done to consider biometric recognition in wild as solved, particularly when consider-

Table 1. **Final results for the ICB-RW competition.** The nine valid submissions of ICB-RW are listed according to their rank in the challenge. Methods are ranked according to the AUC of the CMC curve. Also, both rank-1 and rank-5 identification rate (IR) are presented, along with a brief description provided by the authors.

Method	Description	Rank-1 IR (%)	Rank-5 IR (%)	AUC (CMC curve)
H. Ekenel, G. Özbulak, E. Ghaleb Istanbul Technical University	The probe and gallery face images are aligned with respect to eye centers. Only the frontal images are used as gallery. Face representation is extracted from a convolutional neural network (CNN) with a VGG face model [15]. In the test phase, the nearest neighbour classifier is used with the correlation distance as the similarity score.	69.8	85.3	0.954
K. Grm, S. Dobrisek, V. Struc University of Ljubljana	An augmented dataset was generated through oversampling the training images via bounding box noise and horizontal flipping. The pre-trained VGG face deep convolutional network [15] was used to extract features from the images. Then, a softmax classifier was trained on the features.	62.0	78.7	0.952
H. Shi, X. Zhu, S. Liao, Z. Lei, S. Li Institute of Automation, Chinese Academy of Sciences	Features are extracted from a deep convolutional network model trained on the CASIA-Webface database and the cosine similarity is used as score. Ten models learned from different facial parts are fused, and the gallery images of different poses are synthesized to ease the matching phase.	57.6	75.8	0.921
W. Gutfeter NASK, Warsaw University of Technology	The algorithm builds similarity scores by merging results obtained from a set of convolutional neural networks trained for recognizing faces from different angles.	42.9	64.4	0.918
J. Brogan University of Notre Dame	Gallery and probe images are frontalized using a modified version of [2]. Data features are extracted from a SLMSimple Neural Network [1] and four bins are created containing different versions of the gallery images. Probe descriptors are matched with one of the four bins according to yaw angle of the head, and the resulting pairs of feature vectors are input into an SVM trained with LFW [6] data.	11.6	30.4	0.755
E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez University Autonoma de Madrid	The LBPs [12] of nine facial regions are extracted from a frontalized image [20] followed by illumination normalization. A fused distance score is determined by only considering the five best individual facial regions at each trial.	24.0	39.1	0.725
D. Riccio, M. Nappi University of Salerno	The algorithm locates facial points through an extended Active Shape Model and remaps the face region to a 64x100 image. It applies a local normalization process to correct illumination variations, and the matching is performed with a optimized localized version of the spatial correlation index.	11.1	25.1	0.694
R. Vareto, R. Prates, W. Schwartz University Federal de Minas Gerais	A set of facial components are obtained by performing face landmark localization, and these components are described by a variant of LBP. In the learning phase, a binary classifier based on the Partial Least Squares (PLS) model is inferred for each gallery subject. During the test phase, the identity of the probe samples is determined by the classifier with the highest score.	10.4	25.3	0.688
J. Monteiro University of Porto	During enrolment, a Universal Background Model (UBM) is used to infer a model per individual describing the statistical distribution of each feature (SIFT/GIST). In the recognition phase, features are projected onto both the UBM and the individual specific models. A likelihood-ratio between both projections outputs the final recognition score.	4.9	14.9	0.613

ing scenarios where the number of gallery subjects is much higher. We hope that the results obtained from this contest can contribute to the understanding of the challenges of biometric recognition in the wild, and further advance the research in this field.

7.1. Acknowledgements

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*'.

References

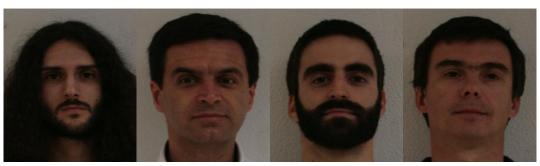
[1] D. Cox and N. Pinto. Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In *IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG)*, pages 8–15, 2011.

[2] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. In *IEEE Conference*

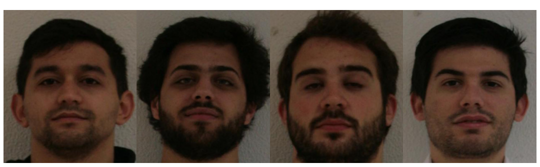
a) Easy-to-identify probe images.

b) Hard-to-identify probe images

Figure 4. **The most easily and hardly identifiable probe images according to the methods performance.** The percentage of gallery subjects necessary to correctly identify a probe image was used to rank probe images. The first and last four images are depicted in a) and b), respectively. Note that in a) the subjects are almost frontal, the expression is neutral and the face is not occluded, whereas in b) there exists extreme variation in pose, the expressions are not standard, and the facial region is degraded by shadows or self-occlusion.



a) Easy-to-identify subjects.



b) Hard-to-identify subjects

Figure 5. **The most easily and hardly identifiable subjects according to the methods performance.** The percentage of gallery subjects necessary to correctly recognize the probe images of each subject was used to rank subjects. The first and last four subjects are depicted in a) and b), respectively. Note that in a) the subjects are significantly different regarding hair, age and face structure, whereas in b) subjects share age and facial structure.

on Computer Vision and Pattern Recognition (CVPR), 2015.

[3] A. K. Jain and S. Z. Li. *Handbook of face recognition*, volume 1. Springer, 2005.

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

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

[6] 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.

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

[8] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain. Nighttime face recognition at long distance: Cross-distance and cross-spectral matching. In *11th Asian conference on Computer Vision*, pages 708–721. 2012.

[9] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W.-Y. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, 94(11):1936–1947, 2006.

[10] J. C. Neves, J. C. Moreno, S. Barra, and 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 (BTAS)*, pages 1–8, 2015.

[11] J. C. Neves and H. Proença. Dynamic camera scheduling for visual surveillance in crowded scenes using markov random fields. In *IEEE 12th International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6, 2015.

[12] T. Ojala, M. Pietikinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51 – 59, 1996.

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

[14] U. Park, H.-C. Choi, A. Jain, and 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):1665–1677, 2013.

[15] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.

[16] H. Proença, S. Filipe, R. Santos, J. Oliveira, and 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):1529–1535, 2010.

[17] F. Wheeler, R. Weiss, and 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*, pages 1–8, 2010.

[18] Y. Yao, B. R. Abidi, N. D. Kalka, N. A. Schmid, and M. A. Abidi. Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement. *Computer Vision and Image Understanding*, 111(2):111 – 125, 2008.

[19] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Li. High-fidelity pose and expression normalization for face recognition in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 787–796, 2015.

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