Michał Włodarczyk
Damian Kacperski
Wojciech Sankowski
Kamil Grabowski

COMPACT: BIOMETRIC DATASET OF FACE IMAGES ACQUIRED IN UNCONTROLLED INDOOR ENVIRONMENT

Abstract

Biometric databases are important components that help improve the performance of state-of-the-art recognition applications. The availability of more and more challenging data is attracting the attention of researchers, who are systematically developing novel recognition algorithms and increasing the accuracy of identification. Surprisingly, most of the popular face datasets (like LFW or IJBA) are not fully unconstrained. The majority of the available images were not acquired on-the-move, which reduces the amount of blurring that is caused by motion or incorrect focusing. Therefore, the COMPACT database for studying less-cooperative face recognition is introduced in this paper. The dataset consists of high-resolution images of 108 subjects acquired in a fully automated manner as people go through the recognition gate. This ensures that the collected data contains real-world degradation factors: different distances, expressions, occlusions, pose variations, and motion blur. Additionally, the authors conducted a series of experiments that verified the face-recognition performance on the collected data.

Keywords   biometrics, face recognition, face database, less-cooperative identification

1. Introduction

Over the past few years, biometric authentication systems have become highly popular and widespread. They are widely used in such areas as automated teller machines [28], the automotive industry [18], or at check-in points at airports [2]. However, when moving to less-cooperative scenarios where the acquisition conditions are significantly lower and subject cooperation is not acceptable, a huge drop in recognition performance can be observed.

In such environments, even fundamental problems like feature extraction and matching become more difficult and challenging. For these reasons, the currently available solutions cannot be used in certain situations (such as people entering stadiums or crowded surveillance environments) [16]. In fact, biometric recognition under uncontrolled conditions has still yet to be realized, making it a major challenge for a large number of researchers; the face area is the most promising trait that can be successfully used in such scenarios.

Biometric databases are important instruments that allow us to improve the performance of state-of-the-art applications. The availability of more and more challenging data is attracting the attention of researchers, who are systematically developing novel recognition algorithms and increasing the accuracy of identification. As an example, the case of the Labeled Faces in the Wild (LFW) [9] dataset can be presented. The release of this database turned out to be a key step towards advancing unconstrained face recognition in which the accuracy increased from 87% in 2009 [7] to 98% in 2017 [14]. Since the inception of LFW, many similar datasets have been released, like PubFig [17] and IJB-A [15].

Surprisingly, LFW and other popular face datasets are not fully unconstrained. They contain images downloaded from the Internet instead of being acquired on-the--move and in an automatic way. This reduces the amount of blur caused by motion or incorrect focusing. Therefore, a novel database for studying less cooperative face recognition is introduced in this paper. The dataset's name is COMPACT, and it was collected during the 2015–2016 academic year. All of the participants were informed about the purpose of the dataset and have granted their written permission so as to allow the use their images for research activities. This set aims at fostering the development of face-recognition systems that work indoors under fully unconstrained conditions. When compared to the existing biometric datasets, the COMPACT database has the following contributions:

- Images are automatically acquired as people go through a recognition gate. The obtained data is available for various distances (1–4 meters) as well as on-the--move.
- Images are taken in the near infrared light spectrum using a high-resolution camera (3520 × 2200 pixels).
- For each recording, the corresponding tracking images are also available. Such data can be used for preliminary classification or research on tracking and prediction algorithms.

- The collected data is divided into four subsets based on the degrees of difficulty. The easiest contain clear frontal facial images, while the hardest introduce pose variations, facial expressions, and occlusions. A detailed illustration of the collected degradation factors is presented in Figure 1.
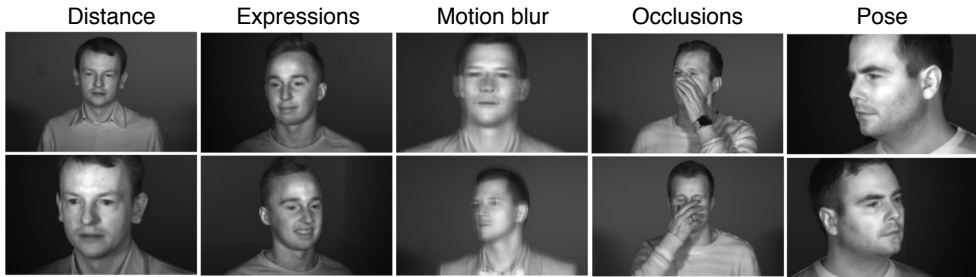

Figure 1. Degradation factors in COMPACT dataset

The remainder of this paper is organized as follows: Section 2 provides an overview of the biometric datasets used for assessing the recognition performance in uncontrolled environments. Section 3 describes the developed system for data acquisition. A detailed description of the introduced dataset is given in Section 4. Section 5 presents the baseline experimental results. Finally, Section 6 concludes the paper.

2. Biometric datasets

There are a large number of databases available to researchers for studying recognition performance in less-cooperative scenarios. These datasets vary in size, scope, and purpose. A comparative analysis is summarized in Table 1. The detailed description of each set is presented below.

Table 1

Comparison between databases designed for studying uncontrolled biometric recognition. Datasets are compared based on number of subjects and degradation factors: expression (E), occlusion (O), illumination (I), pose (P), motion blur (M), on-the-move (OM), and automated acquisition (AA)

Database	Number of subjects	Covariate Factors	Highlights
XM2VTS	295	E, I, P, OM	Multi-modal dataset containing face images, video sequences, and speech recordings
BANCA	26	E, I, P, OM	Face videos containing 12 recordings per subject captured under controlled, degraded, and adverse conditions
FRGC	688	E, I	Still face images taken under controlled (laboratory) and uncontrolled (e.g., outdoors) conditions with one extra 3D image

Table 1 (cont.)

Comparison between databases designed for studying uncontrolled biometric recognition. Datasets are compared based on number of subjects and degradation factors: expression (E), occlusion (O), illumination (I), pose (P), motion blur (M), on-the-move (OM), and automated acquisition (AA)

Database	Number of subjects	Covariate Factors	Highlights
GBU	437	E, I	Subset of FRVT dataset divided into three groups: face pairs that had above-average, close-to-average, and below-average performance, respectively
MBGC	570 (images) 140 (videos)	E, I	Part of Multiple Biometrics Grand Challenge. challenge problem relies on frontal and off-angle face images uncontrolled lighting. Video challenge contains videos in hallways, atria, and outdoor environments
SC-FACE	130	E, I, P	Static face images taken in uncontrolled indoor environmen five video surveillance cameras of various qualities
LFW	5749	E, O, I, P	First database of face images designed directly for studying less cooperative face recognition. All images downloaded from Internet
PubFig	200	E, O, I, P	Similar to LFW dataset but containing more images per subject
FaceScrub	530	E, O, I, P	Large dataset with more than 100,000 images. Data con large variations in pose, expression, and lighting
IJB-A	500	E, O, I, P, M	Face images manually labelled with face-bounding boxes fiducial landmarks for center of two eyes and base of Wide geographical distribution of included people
YouTube Faces	1595	E, O, I, P, M, OM	Face videos downloaded from YouTube. Average of 2,15 videos available for each subject
Choke Point	25	E, O, I, P, M, OM, AA	First dataset with automatically acquired face images by three cameras placed above several portals (natural choke points in terms of pedestrian traffic). Similar in spirit to COMPACT database
COMPACT	108	E, O, P, M, OMAA	Face images automatically acquired for people walking through recognition gate. Images were taken in near infrared light and contain large variations in pose, motion blur, expressions, and occlusions

2.1. XM2VTS [19]

The XM2VTS dataset is a multi-modal database consisting of face images, video sequences, and speech recordings taken from 295 subjects at one-month intervals. The data is available at the cost of distribution from the University of Surrey; it is intended for cooperative scenarios. The data was acquired over a long period of time, which means that variations in the subjects' appearances are noticeable; e.g., changes in hair style or amount of facial hair. The XM2VTS dataset contains four sessions, with each session consisting of eight images per subject.

2.2. BANCA [20]

The BANCA dataset was also collected at the University of Surrey; it contains 52 subjects (26 males and 26 females). It was captured under three different operating scenarios. Each subject was asked to participate in 12 recording sessions. Sessions 1–4 contain data under *Controlled* conditions, while Sessions 5–8 and 9–12 contain *Degraded* and *Adverse* scenarios, respectively.To create independent experiments, the images from various sessions have been divided into independent groups. Based on this, the authors have specified seven distinct experimental configurations; e.g., *Matched Controlled* or *Matched Adverse*.

2.3. FRGC [26]

The FRGC dataset contains 688 subjects; it was collected at the University of Notre Dame. For each subject, four controlled still images, two uncontrolled still images, and one 3D image were captured. The controlled images were taken in a laboratory setting, while the uncontrolled data was acquired under varying lighting conditions (e.g., outdoors). Both sets contain two expressions: neutral and smiling. The 3D images were acquired under controlled illumination. The gallery data was taken during the 2002–2003 academic year, while the verification set was collected during the 2003–2004 academic year. The data was divided into two groups: training and validation. The training set consists of 12,776 images from 222 subjects, with 6388 controlled still images and 6388 uncontrolled still images. The validation set contains images from 466 subjects collected over 4007 sessions.

2.4. GBU [24]

The GBU dataset consists of 437 subjects and was constructed from the Notre Dame multi-biometric data used in the FRVT 2006 database [27]. All of the utilized images are frontal still-face images collected either outside or with ambient lighting in hallways. The *Good* set consists of face pairs that featured above-average performance, the *Bad* set is comprised of face pairs that featured average performance, and the *Ugly* set includes face pairs that featured below-average performance. After constructing the divided groups, the performance of the fusion algorithms was computed. The authors confirmed that the *Ugly* partition is significantly more difficult to match than the *Good* or *Bad* partitions.

2.5. MBGC [25]

The MGBC dataset is part of the Multiple Biometrics Grand Challenge that was designed to investigate, test, and improve the performance of face- and iris-recognition technology on both still and video imagery. The dataset was also constructed at the University of Notre Dame. The still-face challenge problem consists of 570 subjects; it relies on compromising frontal and off-angle face images taken in uncontrolled indoor and outdoor lighting. The video challenge contains videos of 147 subjects captured in hallways, atria, and outdoors under unconstrained illumination, poses, and camera angles. It was the first-ever challenge presented to address face recognition from unconstrained videos.

2.6. SC-FACE [6]

The SC-Face dataset was collected at the University of Zagreb; it consists of static images of human faces taken in an uncontrolled indoor environment using five video surveillance cameras of various qualities. The data was collected in both the visible and infrared spectrum. The constructed dataset contains 130 subjects and 4160 images; the images were taken from various distances and under uncontrolled illumination conditions. The individuals were asked not to look at a fixed point. Ultimately, the acquired images contain nine different poses, which makes the data suitable for head-pose modeling and estimation.

2.7. LFW [9]

The LFW dataset was constructed at the University of Massachusetts in 2007. It consists of 13,233 images of 5749 people. For 1680 subjects, there are two or more images in the database. All of the pictures were downloaded from the Internet. The provided data is organized into two views: a development set of 3200 pairs for building models and choosing features, and a 10-fold cross-validation set of 6000 pairs for evaluation. There are four versions of the LFW dataset: the original as well as three different types of aligned versions. LFW is the first database of face images designed directly for studying less-cooperative face recognition.

2.8. PubFig [17]

The PubFig database was constructed at Columbia University and consists of 58,797 images of 200 subjects. The images were downloaded from the Internet using a person's name at the search query on a variety of image search engines (such as Google Images or Flickr). The PubFig dataset is similar in spirit to LFW, but it contains more images per subject. The collected data contains large variations in pose, lighting, expression, scene, camera, imaging conditions, and parameters.

2.9. FaceScrub [21]

The FaceScrub dataset constructed at the University of Illinois is comprised of 106,863 images of 530 celebrities (with about 200 images per person). The images were re-

trieved from the Internet and were taken during real-world situations. The gender is noted for each subject, so the obtained data can be used for both recognition and gender classification. The collected dataset is similar to LFW and PubFig; however, as a contribution, the authors proposed a new method that automatically removes outliers from the set of faces where the majority are assumed to belong to a particular individual.

2.10. IJB-A [15]

The IJB-A dataset was collected with the cooperation of the Institute of Standards and Technology. The database contains a mix of images (5712) and videos from 500 subjects taken in unconstrained environments. As in PubFig, FaceScrub, or LFW, the images were retrieved from image search engines. The authors claim that the development of recognition algorithms is limited by the face detector's performance; therefore, for each subject, the images were manually labelled with face-bounding boxes and fiducial landmarks for the center of the two eyes and base of the nose. More importantly, this database contains a wider geographical distribution of people than previous sets (e.g., LFW or PubFig), as the subjects were also specified manually.

2.11. YouTube Faces [34]

The YouTube dataset was constructed at Tel Aviv University; it contains 3425 videos of 1595 different subjects. An average of 2.15 videos are available for each subject. The shortest clip duration is 48 frames, the longest is 6070 frames, and the average length of a video clip is 1813 frames. The authors began by using 5749 names of subjects included in the LFW database to search YouTube for videos of those individuals. For each query, the top six results were downloaded. Automatic screening was performed to remove duplicates and low-quality data. This was then manually verified by the authors to ensure that the data was correctly labelled.

2.12. Choke Point [36]

The Choke Point dataset was collected at the University of Queensland. The authors placed an array of three cameras above several portals (natural choke points in terms of pedestrian traffic) to capture people walking through each place in a natural way. The acquired images have variations in terms of illumination, pose, sharpness, and misalignments due to the automatic face detection. The dataset consists of 25 subjects (19 males and 6 females) in Portal 1 and 29 subjects (23 males and 6 females) in Portal 2. In total, the dataset consists of 54 video sequences and 64,204 labelled face images.

3. Acquisition system

3.1. Recognition gate

The COMPACT database was collected with the dedicated acquisition system constructed by the authors. It is based on two wide-field-of-view (WFOV) cameras and

one narrow-field-of-view (NFOV) camera. The WFOV cameras operate within the visible light spectrum, observe the entire scene, and locate the potential subjects to be identified. They form a stereo pair, so they need to be precisely synchronized with each other. Once the system detects that the subject's face is visible and close enough to perform the recognition, the NFOV camera is directed to capture high-resolution biometric images. In order to provide stable lighting conditions, the NFOV images are acquired in near-infrared light. The described concept is presented in Figure 2.
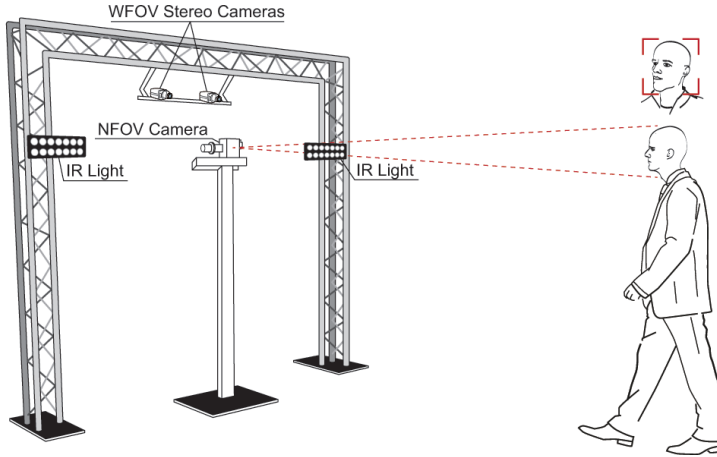

Figure 2. Design concept of COMPACT recognition gate

In the authors' realization, the WFOV vision system consists of two monochromatic Imaging Source DMK33GR0134 cameras. The NFOV vision system is realized with the use of a custom-designed device proposed by the authors. This device is based on the Teledyne DALSA Genie TS-C3500 camera with a specialized lens featuring an advanced pupil and galvanometric motors equipped with mirrors. This allows the system to change the camera-view position in a similar fashion to Pan-Tilt-Zoom (PTZ) cameras; however, the motion is much faster and free of undesirable vibrations in this case. This is somewhat similar to the solution introduced by Okumura et al. [23]. The NFOV camera is also equipped with a specialized focusing mechanism that allows it to capture sharp images within a distance range of 0.8 to 3.0 meters. The realized construction of the recognition gate is presented in Figure 3.

The image-acquisition procedure is divided into five steps; these are presented in Figure 4. The processing starts with the application of the adaptive background subtraction algorithm [37] on the WFOV images. Then, the image regions are determined where the potential subjects may be found. In each of these regions, the face HOG detector [13] is executed. As the WFOV cameras form a stereo pair, the 3D position (x, y, z) of each person's face is obtained. Having this, a 3D tracking record is constructed for each subject. Then, the Kalman filter is used to predict the person's future position at the time of acquiring the NFOV image. This allows us to

compensate for the subject's motion and the time shift introduced by the delay of the lens-focusing mechanism. The final step is to transform the subject's predicted 3D position into a coordinate system for the galvanometers. The WFOV stereo vision system returns the position as a Cartesian coordinate system, while the NFOV camera uses a polar coordinate system. Therefore, the authors developed a dedicated calibration procedure [11] that allows us to determine the required transformation parameters.

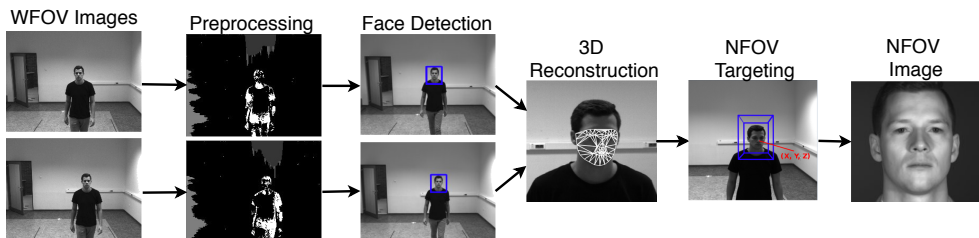**Figure 3.** Practical realization of COMPACT recognition gate

Figure 4. Processing steps of COMPACT system

The WFOV and NFOV cameras are synchronized with each other, so the NFOV camera can capture the images in between the frames of the WFOV cameras. The near-infrared illumination is activated only when the NFOV camera acquires a set of frames. To ensure stable lighting conditions, the illumination intensity varies as the distance of the subject changes. As a person is approaching the NFOV camera, the lighting intensity is reduced.

3.2. Rotating platform

In a biometric system, the verification images are matched against high-quality registration data. For this, the authors constructed a dedicated turntable to collect multi-pose face images [10]. The developed device takes a photo every 5°. The Dalsa TS-C2500 camera was used for image acquisition. The camera was set at a distance of three meters from the turntable. As the NFOV images are acquired in near-infrared light, the gallery data needs to be taken in the same spectrum. Therefore, the authors mounted two illumination modules directed towards the center of the turntable.

The rotating platform idea is presented in Figure 5, while Figure 6 shows the constructed station.
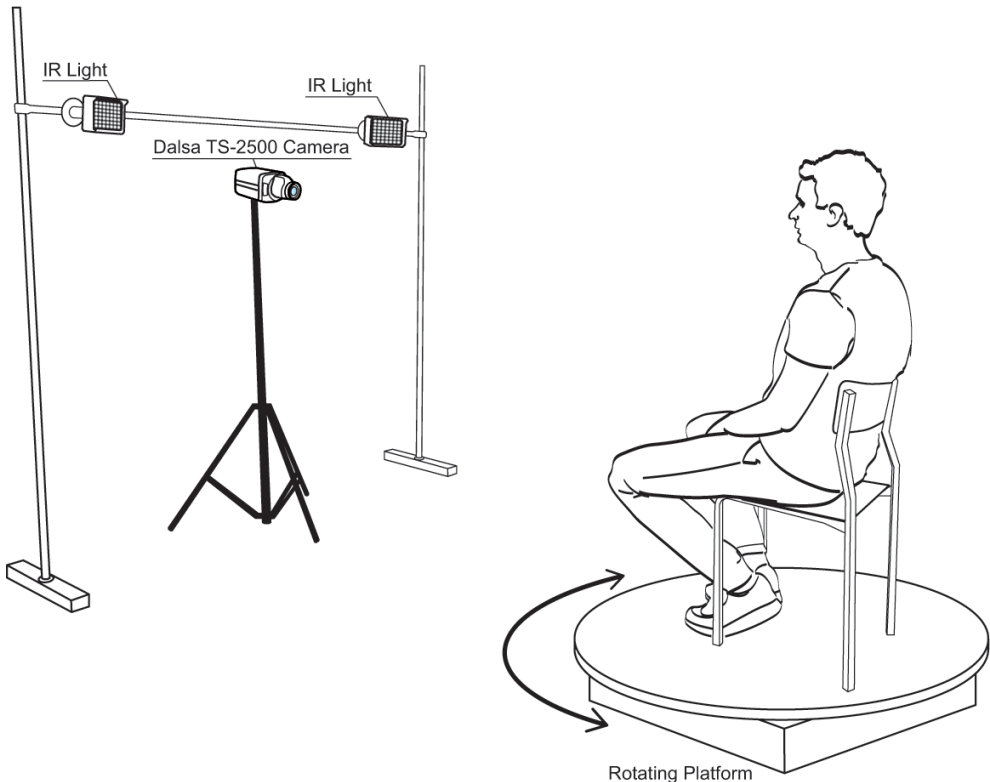

Figure 5. Design concept of rotating platform

Figure 6. Practical realization of rotation platform used to capture multi-pose registration data

4. COMPACT dataset

The COMPACT dataset was collected in an indoor laboratory environment. The acquisition consists of three parts: a soft biometric survey, multi-pose registration images, and probe data. The registration and probe images were taken on different days to ensure reliable comparisons. The most important features of the COMPACT database are summarized in Table 2.

Table 2
Features of COMPACT dataset

Spectrum	Near Infrared
Number of subjects	108
Number of registration images	12,312
Number of probe images	31,078
Avg. no. of probe images per subject	287
Image size	2560×2048 px (registration) 3520×2200 px (probe)

The authors acquired the data to meet the following requirements:

- Automatically acquire real-world images of people walking through recognition gate. Obtained data should include numerous variations in poses, expressions, and occlusions.
- Acquire images that allow design of multi-biometric recognition strategies (e.g., face and periocular fusion).
- Categorize obtained data into independent groups representing different degrees of difficulty.

4.1. Soft biometrics

Each person was asked to fill out a soft biometric survey. The prepared questionnaire contained nine multiple-choice questions. The obtained data can be used as a pre-processing step; e.g., to restrict the number of possible matches in the identification process. The list of questions included in the survey is presented in Table 3. The responses' distribution with respect to the possible answers is presented in Figure 7.

Table 3
List of questions included in soft biometrics survey with possible answers

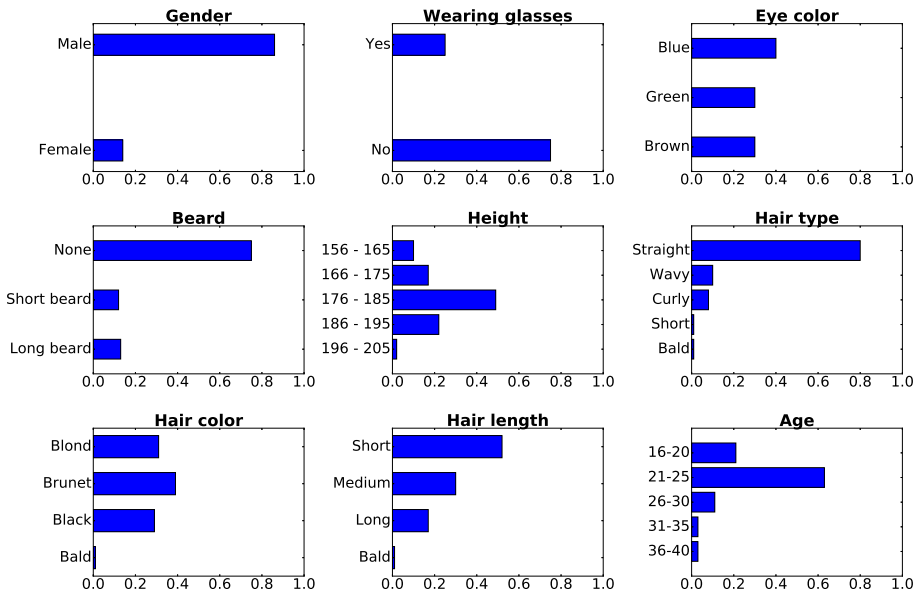| Trait	Labels
Age	$\mathbb{N}$
Height	$\mathbb{N}$
Gender	Male, Female
Hair color	Bald, Blond, Black, Brown
Hair type	Bald, Short, Curly, Wavy, Straight
Hair length	Bald, Short, Medium, Long
Beard	None, Short beard, Long beard
Eye color	Blue, Green, Brown
Wearing glasses	Yes, No

Figure 7. Distribution of soft biometric survey answers within collected data

4.2. Registration data

The registration data consists of facial images acquired with the use of the rotating platform described in Section 3.2. The images were collected within a range of $-45°$

to 45°, every 5°, and within three vertical poses: looking ahead, looking up, and looking down. This means that a single set consists of $19 \cdot 3 = 57$ images. For each person, two consecutive sessions were acquired. The reason for duplicating the data was that the likelihood of acquiring a degraded image of a selected pose (e.g., the subject blinks or twitches during frame acquisition) is significantly smaller. The sample images included in the registration data are presented in Figure 8.

Figure 8. Sample images of registration data acquired using constructed rotating platform

4.3. Probe data

The probe data is divided into two groups: high-resolution facial images captured by the NFOV camera, and the tracking sequences recorded by the WFOV vision system. A single recording contains images of one person walking through the recognition gate. All of the data was acquired in a fully automated way using the constructed acquisition system described in Section 3.1.

In order to have the proper gradation of difficulty in the acquired data, the authors proposed four scenarios for the recordings. The first two scenarios are controlled, and the other two are fully uncontrolled. People wearing glasses were asked to wear them only during the last scenario. Each subject performed each scenario twice. In total, eight recordings were collected for each person. The degradation factors for each scenario are illustrated in Figure 9.
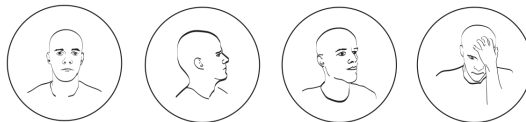

Figure 9. Degradation factors for each designed scenario: looking straight ahead, looking left or right, looking around, and looking around with occlusions

The following scenarios were proposed:

- **Scenario 1:** A person walks through the recognition gate looking straight ahead. This allows us to capture clear frontal facial images. Such data can be used for the development process or to verify algorithm implementation, for example.
- **Scenario 2:** A person walks through the recognition gate turning his/her head to the left (first recording) or right (second recording) without changing the pitch or roll angles. The head is always turned away from the camera to make the images more difficult. This data can be used to study the impact of pose variations on the recognition performance or head-pose modeling.
- **Scenario 3:** A person walks through the recognition gate looking around the scene. This involves turning the head to the left and right as well as up and down. These images represent the real-world uncontrolled data, including a mix of frontal and pose images.
- **Scenario 4:** A person walks through the recognition gate looking around the scene (as in the third scenario). However, during these recordings, people rubbed their eyes, adjusted their hair, or scratched their cheeks. This allowed for the acquisition of real-world unconstrained images with occlusions.

As for the WFOV sequences, the authors saved all of the available images together with the stereo calibration matrices. This data can be used for various purposes; for example, it is possible to reconstruct the tracking sequences and develop position prediction algorithms. Figure 10 presents sample NFOV and WFOV image sets.

Figure 10. Sample images from probe data (WFOV LEFT, WFOV RIGHT, NFOV)

4.4. Data availability

The main directory of the COMPACT database contains the following directories: *Registration* and *Verification*. Each of these stores images for the registration and

probe data, respectively. Within them, there are directories for all of the subjects stored in the database. In each subject's individual directory, all of the sessions with acquired images are stored. A detailed description of the directory structures for each subject is presented in Figure 11.
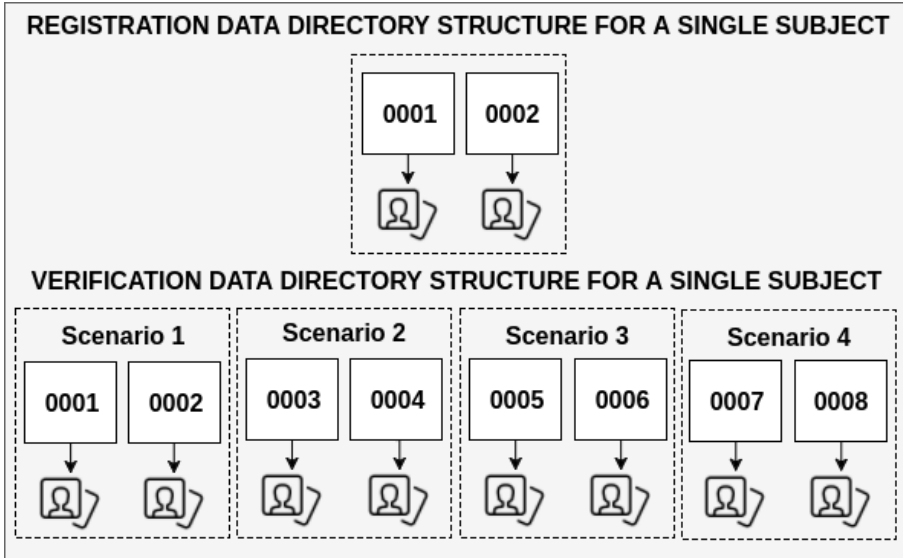

Figure 11. COMPACT dataset directory structure for each subject

The names of the registration images utilize four-digit positive integers, while the probe image-naming convention uses the following format: D_C, where D represents the timestamp of the image in the Linux time format and C stands for the utilized device (*NFOV*, *WFOV_LEFT*, or *WFOV_RIGHT*). The timestamps allow us to easily match the WFOV images with the corresponding NFOV images. For additional information on how to obtain and use the database, please refer to the COMPACT web page [33].

5. Experimental evaluation

5.1. Verification procedure

The authors conducted a series of experiments to verify the face-recognition performance on the COMPACT database. The probe data was filtered using the following procedure: for each NFOV image, the HOG face detector was executed [13]; any images that detected no faces were ignored in the further processing. The number of rejected images for each scenario is reported in Table 4.

In most cases, a face was unavailable due to detection and tracking errors in the WFOV stereo vision system. Consequently, the NFOV camera was not correctly targeted (see Section 3.1).

Table 4
Number of ignored images in COMPACT dataset

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Input data	4174	3987	4142	3182
Valid data	4145	3856	4037	2980
Coverage	99.31%	96.71%	97.47%	93.65%

The remaining images were aligned in the following way: first, the image was rotated with the use of face landmarks [12, 13] so that the centers of the eyes were located on the same level. Second, the face was cropped and scaled to a fixed size of 150×150 pixels. Finally, the well-known TanTriggs algorithm [31] was executed to normalize the illumination. The described procedure is summarized in Figure 12.
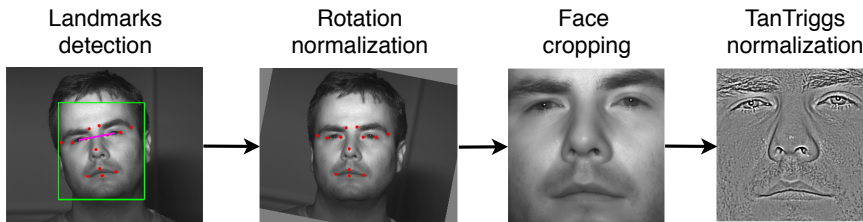

Figure 12. Face-normalization procedure

As described in Section 4.2, a single registration session consists of 57 images. The authors treat this set of images as a single sample. Each probe image is compared with all 57 images, and each comparison returns a dissimilarity score. As a final result, the minimum score is selected. Unfortunately, the proposed brute-force method turned out to be considerably time consuming. For this reason, the authors decided to reduce the number of comparisons by utilizing the head pose estimation algorithm [32]. In the improved solution, the yaw angle is estimated for both registration and probe images. Then, only the comparisons between those images where the estimated yaw angle differences are smaller than $15°$ are processed.

5.2. Evaluation algorithms

The performance of the following algorithms was evaluated: ThreePatchLBP [35], HighDimLBP [3], and ResNet [14]. The choice of these methods was not accidental: they operate on significantly different principles.

The ThreePatchLBP algorithm is based on an LBP (Local Binary Pattern) descriptor [22]. The face image is divided into a ten-by-ten grid of small regions. For each region, the ThreePatchLBP descriptor is performed individually, and the histogram that contains the shape and texture information is calculated. These histograms are then concatenated into a single-dimensional feature vector. Such vectors are then compared using the Bhattacharyya distance [1].

The HighDimLBP algorithm constructs the feature vector by extracting multi-scale patches centered at dense face landmarks. The multi-scale image patches are extracted around each landmark; then, each patch is divided into a four-by-four grid. For each region, the LBP descriptor is extracted; finally, all of the descriptors are concatenated to form a high-dimension feature vector. In total, the authors used 27 landmarks, which were the silent points on the eyebrows, eyes, nose, and mouth [5]. The obtained feature vectors are then compared using the Bhattacharyya distance [1].

The ResNet algorithm is a modified version of the ResNet-34 neural network [8]. It contains 27 convolution layers; its implementation and model comes from the *dlib* library [14]. The utilized model was trained on a dataset of about three million faces. The training set was constructed from a number of public databases: FaceScrub, VGG, and a large number of images downloaded from the Internet. For each face, the ResNet network returns a 128-dimension feature vector. Such arrays are then compared using the Euclidean distance [4].

5.3. Result discussion

The experimental results are summarized in Table 5. The ROC (Receiver Operating Characteristic) and CMC (Cumulative Match Characteristic) curves are presented in Figure 13. When analyzing the obtained results, a number of conclusions can be drawn.

First, it can be noticed that the utilized verification scenarios were prepared properly and met their requirements. The best performance was obtained for the first scenario (which includes mostly frontal images), while the worst performance was reported for the fourth scenario (which consists of pose variations, occlusions, and expressions). A remarkable drop in performance was also observed for the second scenario (where a person walks through the recognition gate, turning his/her head to the left or right). This shows that the impact of pose variations is significant and the state-of-the-art face recognition algorithms are not fully resistant to this factor. For the third scenario (which includes a mix of frontal and pose images), the obtained results were close to the results obtained for the first scenario. This means that this subset contains a relatively large number of frontal images.

Second, a significant difference in recognition performance can be noticed between the ResNet network and the ThreePatchLBP and HighDimLBP algorithms. The results show that neither the ThreePatchLBP nor HighDimLBP algorithms is fully resistant to pose variations and occlusions. For instance, when a face is partially covered with a hand, the features are extracted from significantly different areas. On

the other hand, the ResNet model (trained on around three million images) is more resistant to such distortions. However, when analyzing the results obtained for the fourth scenario, it can be noticed that the ResNet algorithm still needs to be improved for high-security use cases (FAR values lower than 10^{-3}), as the GAR result of approx. 90% may not be enough in certain applications.

Finally, as for the recognition in less-cooperative conditions, the obtained results confirm that the deep-learning algorithms significantly outperform the classic approaches. This indicates that nowadays access to large training sets is one of the main limitations of commercial systems for developing reliable recognition in uncontrolled scenarios. This is particularly evident when one analyzes the results published by companies such as Google [29] or Facebook [30], who have extensive collections of training data.

Table 5
Results obtained using COMPACT database

| Scenario 1			
Accuracy	ThreePatchLBP	HighDimLBP	ResNet
EER, %	2.54 ± 0.25	2.43 ± 0.12	0.44 ± 0.08
AUC, %	99.06 ± 0.16	98.91 ± 0.13	99.96 ± 0.01
RANK-1, %	99.20 ± 0.14	99.39 ± 0.47	99.67 ± 0.08
Scenario 2			
Accuracy	ThreePatchLBP	HighDimLBP	ResNet
EER, %	12.07 ± 0.74	10.33 ± 0.92	1.41 ± 0.65
AUC, %	92.53 ± 0.89	93.32 ± 0.85	99.25 ± 0.61
RANK-1, %	95.48 ± 0.82	95.88 ± 0.97	98.61 ± 0.67
Scenario 3			
Accuracy	ThreePatchLBP	HighDimLBP	ResNet
EER, %	9.67 ± 0.72	9.21 ± 0.72	0.62 ± 0.20
AUC, %	94.64 ± 0.59	94.56 ± 0.62	99.83 ± 0.12
RANK-1, %	95.20 ± 0.75	96.47 ± 0.86	99.56 ± 0.14
Scenario 4			
Accuracy	ThreePatchLBP	HighDimLBP	ResNet
EER, %	21.76 ± 1.84	22.67 ± 1.42	1.58 ± 0.28
AUC, %	85.44 ± 1.29	84.41 ± 1.14	99.88 ± 0.02
RANK-1, %	89.03 ± 1.42	93.69 ± 2.15	98.06 ± 0.64
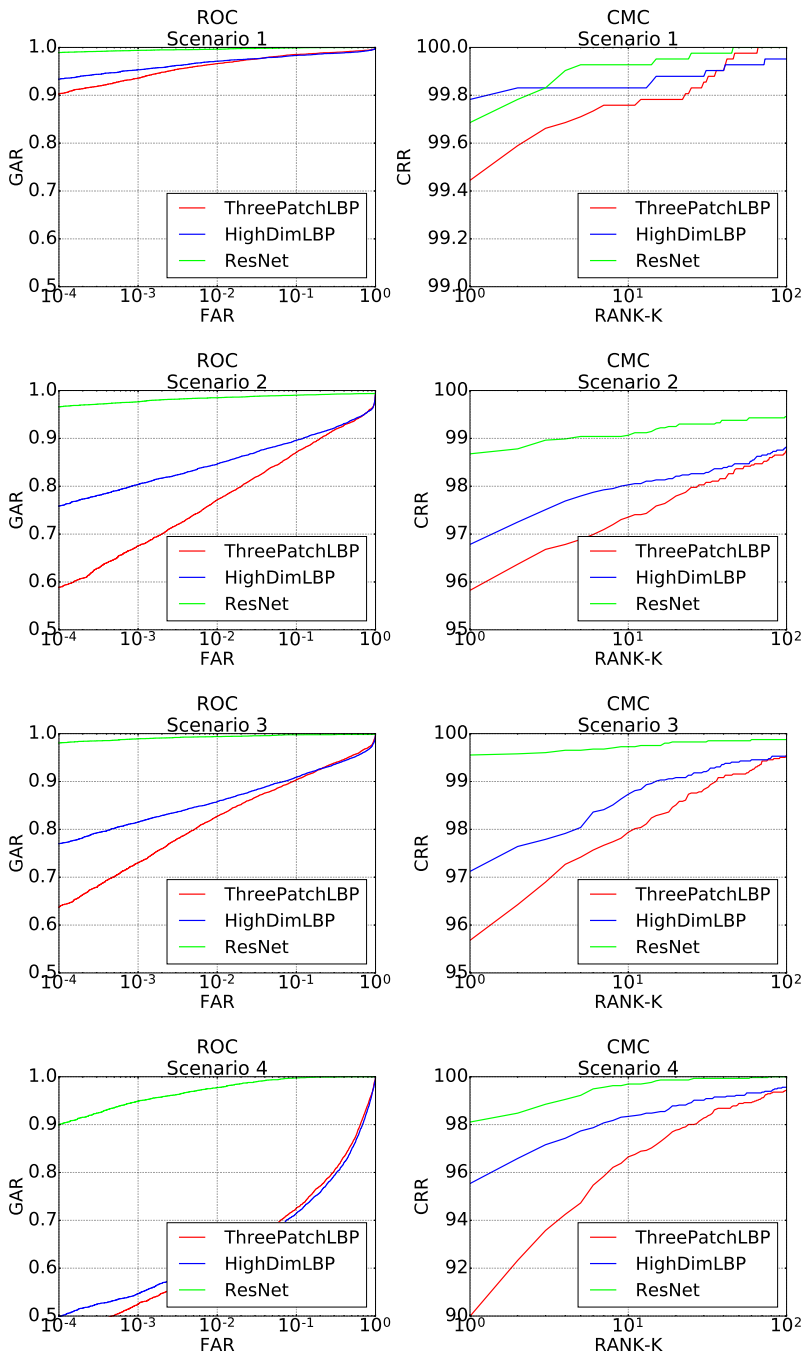
Figure 13. Obtained ROC and CMC curves on COMPACT dataset

6. Conclusions

This paper introduces the COMPACT database, whose aim is to foster the development of face-recognition systems that work indoors and under unconstrained conditions. The authors describe the constructed recognition gate and the acquisition process in detail. Also, a series of experiments were conducted to verify the performance of the state-of-the-art face recognition algorithms on the obtained data.

The COMPACT dataset consists of 108 subjects. For each individual, a soft biometric survey, multi-pose registration images, and probe data were collected. The proposed enquiry contains nine multiple-choice questions. The obtained answers can be used as a preprocessing step to restrict the number of possible matches in the identification process, for example. The registration data consists of facial images acquired in a horizontal pose within a range of $-45°$ to $45°$, every $5°$, and within three vertical poses: looking ahead, looking up, and looking down. For each individual, the total number of registration images is 114. The probe data contains images that are taken when people pass through the developed recognition gate. The average number of acquired biometric images for each person is 280. In order to have the proper gradation of difficulty in the obtained data, the authors proposed four scenarios for the recordings. The easiest contains clear frontal facial images, while the hardest introduce pose variations, facial expressions, and occlusions. Compared to the related state-of-the-art datasets, the COMPACT images were acquired in a fully automated manner. This ensures that the collected data contains real-world degradation factors such as motion blur or incorrect focusing. Additionally, the associated tracking images are also available for each recording, which allows for the development of 3D trackers and prediction algorithms.

To provide the baseline results, the authors verified the performance of face recognition on the COMPACT database. The following algorithms were evaluated: ThreePatchLBP, HighDimLBP, and the ResNet neural network. The conclusions are twofold: first, the obtained results confirm that properly trained deep-learning algorithms are more resistant to real-world degradation factors (e.g., occlusions, pose variations, and expressions) and outperform the classic approaches. Second, the COMPACT database proved to be a challenging set, as the state-of-the-art face recognition algorithms returned significant error rates for the obtained data. In future work, the authors plan to focus on periocular recognition and the fusion of facial and periocular biometrics. The authors also intend to incorporate quality assessment into the developed recognition system.

Acknowledgment

References

[1] Bhattacharyya A.K.: On a measure of divergence between two statistical populations defined by their probability distributions, *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–109, 1943.

[2] Cabello E.: Face biometrics without intrusion in airport. In: *Optical Communication Systems (OPTICS), 2011 Proceedings of the International Conference on*, 2011.

[3] Chen D., Cao X., Wen F., Sun J.: Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[4] Deza M.M., Deza E.: *Encyclopedia of Distances*, Springer, Berlin, Heidelberg, 2009.

[5] Everingham M., Sivic J., Zisserman A.: "Hello! My name is... Buffy" – Automatic Naming of Characters in TV Video. In: *Proceedings of the British Machine Vision Conference*. 2006.

[6] Grgic M., Delac K., Grgic S.: SCface – surveillance cameras face database, *Multimedia Tools and Applications*, vol. 51(3), pp. 863–879, 2011.

[7] Guillaumin M., Verbeek J., Schmid C.: Is that you? Metric learning approaches for face identification. In: *IEEE International Conference on Computer Vision*, 2009.

[8] He K., Zhang X., Ren S., Sun J.: Deep Residual Learning for Image Recognition, In: *CoRR*, vol. abs/1512.03385, 2015. `http://arxiv.org/abs/1512.03385`

[9] Huang G.B., Mattar M., Berg T., Learned-Miller E.: Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. In: *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.

[10] Kacperski D., Krotewicz P., Wlodarczyk M., Grabowski K.: Pose-Oriented Face Images Acquisition Platform. In: *Mixed Design of Integrated Circuits Systems (MIXDES), 2016 23rd International Conference*, 2016.

[11] Kacperski D., Sankowski W., Włodarczyk M., Grabowski K.: Calibration of Vision Systems Operating in Separate Coordinate Systems, *International Journal of Microelectronics and Computer Science*, vol. 7(1), pp. 10–15, 2016.

[12] Kazemi V., Sullivan J.: One Millisecond Face Alignment with an Ensemble of Regression Trees. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[13] King D.E.: Max-Margin Object Detection, *CoRR*, vol. abs/1502.00046 2015. `http://arxiv.org/abs/1502.00046`.

[14] King D.E.: Dlib – open source maching learning library, 2017.

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

[16] Klontz J., Jain A.: A Case Study of Automated Face Recognition: The Boston Marathon Bombings Suspects, *Computer*, vol. 46(11), pp. 91–94, 2013.

[17] Kumar N., Berg A.C., Belhumeur P.N., Nayar S.K.: Attribute and simile classifiers for face verification. In: *2009 IEEE 12th International Conference on Computer Vision*, 2009.

[18] Lupu C., Lupu V.: Multimodal Biometrics for Access Control in an Intelligent Car. In: *2007 International Symposium on Computational Intelligence and Intelligent Informatics*, 2007.

[19] Matas J., Hamouz M., Jonsson K., Kittler J., Li Y., Kotropoulos C., Tefas A., Pitas I., Tan T., Yan H., Smeraldi F., Bigun J., Capdevielle N., Gerstner W., Ben-Yacoub S., Abeljaoued Y., Mayoraz E.: Comparison of face verification results on the XM2VTFS database. In: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 4, pp. 858–863, 2000.

[20] Messer K. et al.: Face authentication test on the BANCA database. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, vol. 4, pp. 523–532, 2004.

[21] Ng H.W., Winkler S.: A data-driven approach to cleaning large face datasets. In: *2014 IEEE International Conference on Image Processing*, 2014.

[22] Ojala T., Pietikäinen M., Harwood D.: A comparative study of texture measures with classification based on featured distributions, *Patter Recognition*, vol. 29(1), pp. 51–59, 1996.

[23] Okumura K., Oku H., Ishikawa M.: High-speed gaze controller for millisecond-order pan/tilt camera. In: *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[24] Phillips P.J., Beveridge J.R., Draper B.A., Givens G., O'Toole A.J., Bolme D., Dunlop J., Lui Y.M., Sahibzada H., Weimer S.: The Good, the Bad, and the Ugly Face Challenge Problem, *Image and Vision Computing*, vol. 30(3), pp. 177–185, 2012.

[25] Phillips P.J., Flynn P.J., Beveridge J.R., Scruggs W.T., O'Toole A.J., Bolme D., Bowyer K.W., Draper B.A., Givens G.H., Lui Y.M., Sahibzada H., Scallan III J.A., Weimer S.: Overview of the Multiple Biometrics Grand Challenge. In: Tistarelli M., Nixon M.S. (eds.), *Advances in Biometrics. ICB 2009*, Lecture Notes in Computer Science, vol 5558. Springer, Berlin, Heidelberg, pp. 705–714, 2009.

[26] Phillips P.J., Flynn P.J., Scruggs T., Bowyer K.W., Chang J., Hoffman K., Marques J., Min J., Worek W.: Overview of the face recognition grand challenge In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[27] Phillips P.J., Scruggs W.T., O'Toole A.J., Flynn P.J., Bowyer K.W., Schott C.L., Sharpe M.: FRVT 2006 and ICE 2006 Large-Scale Experimental Results. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 831–846, 2010.

[28] Sako H., Miyatake T.: Image-recognition technologies towards advanced automated teller machines. In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004.

[29] Schroff F., Kalenichenko D., Philbin J.: FaceNet: A Unified Embedding for Face Recognition and Clustering, *CoRR*, vol. abs/1503.03832, 2015. `http://arxiv.org/abs/1503.03832`.

[30] Taigman Y., Yang M., Ranzato M., Wolf L.: DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[31] Tan X., Triggs B.: Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions, *IEEE Transactions on Image Processing*, vol. 19(6), pp. 1635–1650, 2010.

[32] Wlodarczyk M., Kacperski D., Krotewicz P., Grabowski K.: Evaluation of head pose estimation methods for a non-cooperative biometric system. In: *2016 MIXDES – 23rd International Conference Mixed Design of Integrated Circuits and Systems*, 2016.

[33] Wlodarczyk M., Kacperski D., Sankowski W., Grabowski K.: COMPACT Database Description, 2018. `http://biometrics.dmcs.pl/en/databases/compact`.

[34] Wolf L., Hassner T., Maoz I.: Face recognition in unconstrained videos with matched background similarity. In: *CVPR 2011*, CVPR 2011, Colorado Springs, USA, pp. 529–534, 2011.

[35] Wolf L., Hassner T., Taigman Y.: Descriptor Based Methods in the Wild. In: *Real-Life Images workshop at the European Conference on Computer Vision (ECCV)*, 2008.

[36] Wong Y., Chen S., Mau S., Sanderson C., Lovell B.C.: Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition, *CVPR 2011 WORKSHOPS*, 2011.

[37] Zivkovic Z.: Improved Adaptive Gaussian Mixture Model for Background Subtraction. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, Cambridge, pp. 28–31, 2004.

Affiliations

Michał Włodarczyk

> Lodz University of Technology, Wolczanska 221/223, 90-924 Lodz, Poland,
> mwlodarczyk@dmcs.pl

Damian Kacperski

> Lodz University of Technology, Wolczanska 221/223, 90-924 Lodz, Poland,
> dkacperski@dmcs.pl

Wojciech Sankowski

> Lodz University of Technology, Wolczanska 221/223, 90-924 Lodz, Poland,
> wsan@dmcs.pl

Kamil Grabowski

> Lodz University of Technology, Wolczanska 221/223, 90-924 Lodz, Poland,
> kgrabowski@dmcs.pl