IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 March 2024, accepted 5 April 2024, date of publication 10 April 2024, date of current version 18 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3386955

RESEARCH ARTICLE

Information Fusion and Hand Alignment to Improve Hand Recognition in Forensic Scenarios

LAZARO JANIER GONZALEZ-SOLER [1], DURITA KVILT JÓNSDÓTTIR [2], CHRISTIAN RATHGEB [1], AND DANIEL FISCHER [1]

[1]da/sec-Biometrics and Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany
[2]Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Lyngby, Denmark

Corresponding author: Lazaro Janier Gonzalez-Soler (lazaro-janier.gonzalez-soler@h-da.de)

This work was supported in part by the Hessian Ministry of the Interior and Sport in the course of the Bio4ensics Project, and in part by German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

ABSTRACT In many forensic scenarios, criminals often attempt to conceal their identity by covering their face and other distinctive body parts. In such situations, physical evidence may, however, reveal other unique characteristics, e.g. hands, which can be used to identify offenders. In this context, several state-of-the-art biometric recognition systems have been proposed recently. These recognition systems offer high identification performance in restricted environments. However, in forensic scenarios, the environment is often unconstrained, making biometric identification considerably more difficult, with a consequent decrease in accuracy. In this article, we explore methods (e.g. hand alignment and information fusion) to improve the identification of subjects within forensic investigations. Experimental results show that explored techniques play an important role in the improvement of the identification performance of existing schemes: the combination of hand alignment and information fusion results in the highest Rank-1 identification performance improvement of up to 13.10% (i.e., 26.30% vs. 13.20%) and 16.30% (i.e., 77.00% vs. 60.70%) with respect to the baseline for the unconstrained databases NTU-PI_v1 and HaGRID, respectively (https://github.com/ljsoler/IF-HA-HandRecognition).

INDEX TERMS Hand recognition, hand alignment, information fusion, forensic investigations, uncontrolled scenarios.

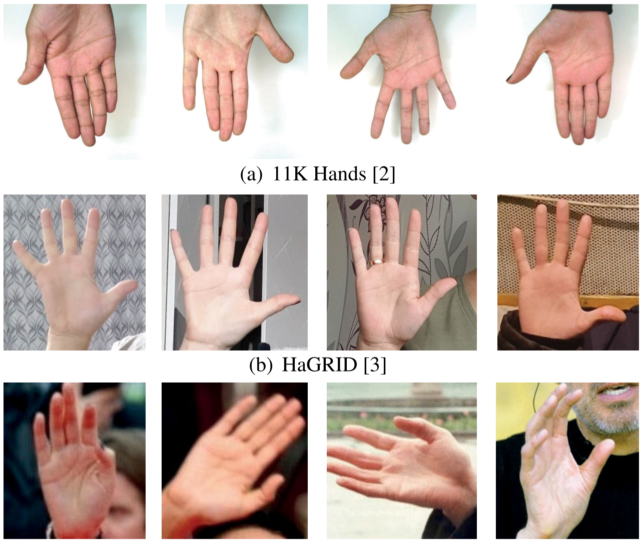
I. INTRODUCTION

The demand for biometric systems has increased considerably in recent years due to their high performance in verifying or identifying subjects, their efficiency and, in many cases, their convenience for the users. They are therefore used in sectors as diverse as healthcare technology, access control and forensics. A biometric system enables the automated recognition of individuals based on their biological and behavioural characteristics [1]. Some of the most commonly used biometric characteristics for such purposes are fingerprints, face, iris and hand.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Jin.

To assist law enforcement, biometric systems have been applied since the mid-1960s, when the Automated Fingerprint Identification System (AFIS) was formally introduced. Since then, significant progress has been made not only in the use of fingerprints but also in the use of other types of biometric characteristics such as face or gait [5]. Despite these advances, offenders often attempt to conceal their identity by covering their face and other distinguishing features in forensic scenarios. In such cases, physical evidence can, however, reveal other unique biometric characteristics, e.g. the perpetrator's hand, which can be used for biometric identification.

The anatomy of the hand represents an important source of information utilised by hand-based biometric systems either to verify the identity a subject. In general, a hand



(a) 11K Hands [2]

(b) HaGRID [3]

(c) NTU-PI_v1 [4]

FIGURE 1. Examples of controlled hand images in 11K Hands [2] and uncontrolled images in HaGRID [3] and NTU-PI_v1 [4], respectively.

consists of a broad palm with five fingers, each attached to the joint called the wrist. The back of the hand is formally called the dorsum or dorsal of the hand [6]. Earlier hand-based approaches mainly used the geometric characteristics of the hand for recognition purposes [7], [8]. However, with the success of deep neural networks (DNNs) in various computer vision and pattern recognition tasks, current systems analyse a variety of features ranging from texture to appearance properties. In 2019, Afifi [2] proposed a public database of hand images together with a DNN-based system for subject identification. To improve the baseline identification performance reported in [2], the most recent DNN-based algorithms have exploited attention mechanisms [9], [10] or vision-transformers [11]. These schemes were mainly evaluated in closed-set identification scenarios where the searched identity is known to be included in the database of enrolled references. In addition, the databases used in such studies contain controlled hand images with little variation in background context, image quality, hand pose and finger gestures, properties that often cause large variations in images processed during forensic investigations. Examples of controlled and uncontrolled images as well as the identification rates for the existing hand-based systems on these images are presented in Fig. 1 and Tab. 1, respectively. A high-performance deterioration can be observed when uncontrolled hand images are evaluated: identification rates (IR) at Rank-1 for uncontrolled hand images (i.e., NTU-PI_v1) decrease by about 80 percentage points compared to the ones computed on controlled images (i.e., 11K Hands).

In this paper, we propose a study on several pre-processing steps and information fusion techniques to improve the identification rate of existing hand-based recognition systems in uncontrolled scenarios, specifically within the context of

TABLE 1. Identification rates (%) for controlled and uncontrolled images in 11K Hands [2] and uncontrolled NTU-PI_v1 [4], respectively.

Approach	Database	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
ABD-Net [13]	11K Hands	92.80	97.80	95.80
	NTU-PI_v1	11.00	21.20	27.70
RGA-Net [14]	11K Hands	84.90	91.20	93.50
	NTU-PI_v1	4.60	8.40	11.80
MBA-Net [10]	11K Hands	96.80	99.80	100
	NTU-PI_v1	13.20	22.70	28.40

forensics. Current recognition systems treat hand images as they are and do not apply any pre-processing procedures, which might hinder network optimisation. In essence, the main contributions of this work are:

- An in-depth analysis of different hand-alignment techniques, including palmprint and affine alignments in order to improve the robustness of hand recognition (note the differences in terms of alignment between images of 11K Hands and NTU-PI_v1 in Fig. 1).
- The exploration of different information fusion methods, which is motivated by the fact that complementary information can improve the overall recognition accuracy. We perform the fusion between different DNN-based systems and different hand-alignment algorithms to improve the final identification performance.
- A thorough evaluation of the proposed systems compliant with the international standard ISO/IEC 19795-1 [12] for biometric testing and reporting is conducted on challenging hand images varying their pose, gesture and image quality.

The remainder of this work is organised as follows: a review of existing hand-based biometric systems is provided in Sect. II. Sect. III and Sect. IV present fundamentals about hand alignment and information fusion techniques used in this scientific article. The experimental setup is explained in Sect. V. The experimental results and derived findings are discussed in Sect VI. Finally, conclusions and future work directions are presented in Sect. VII.

II. RELATED WORK

Hand biometric recognition focuses on verifying or identifying individuals based on the unique properties of their hands. In the last century, hand-based biometric systems analysed the geometric properties of the hand as the main feature for recognition, and the first device using such features was patented in 1971 [15]. Subsequently, the first commercial hand geometry scanner based on Miller's patent, named Identimat, was developed in 1976 [16].

While the former hand recognition systems were mainly based on hand geometry, ongoing research tries to exploit further properties of the hand [17]. Recently, a growing trend in the use of deep learning is observable that, instead of manually extracting specific features, automatically extracts discriminative information (e.g. appearance and texture features) from the hand. The advantage of these approaches

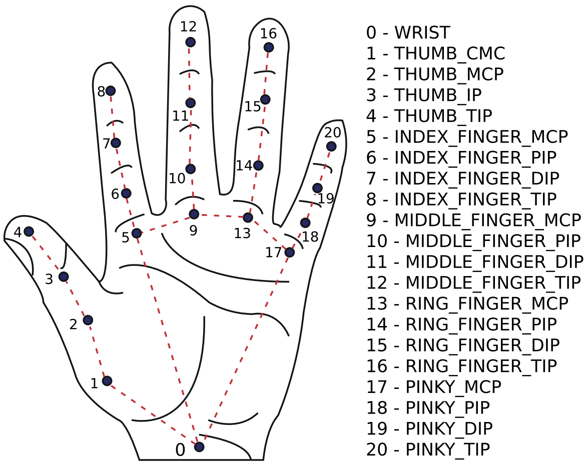


FIGURE 2. Hand landmarks computed by Google MediaPipe.

is that they can leverage all available features of the hand, rather than being limited to a single feature type. In order to improve the balance between applicability, user convenience, and recognition performance of previous approaches, the latest techniques map whole hand images acquired in the visible spectrum into a latent representation using DNNs.

Afifi [2] introduced an annotation-rich hand database (referred to in the scientific literature as 11K Hands) consisting of 11,076 high-quality hand images captured in the visible spectrum. In addition, the same author proposed a dual-stream convolutional neural network (CNN)-based algorithm whose recognition performance values (i.e., correct identification rates (CIRs) ranging from 94% to 97% in Rank-1 for the palmar and dorsal area, respectively) provided a starting benchmark for future investigations. Following the above idea, Baisa et al. [9] recently proposed a dual-stream CNN-based approach which learns both global and local features of the hand image. The experimental results reported a CIR in terms of Rank-1 of around 95% on 11K Hands [2]. Baisa et al. [10] extended this architecture by including an extra stream and incorporating both channel and spatial attention modules in branches. An improvement in recognition of around 3% (i.e., CIR of 98.05%) was achieved on the right palm images compared to the 95.83% obtained in [9]. In the same study, other CNNs were evaluated, e.g., ABD-Net [13] and RGA-Net [14], resulting in similar recognition performance compared to the one in [9]. Finally, Ebrahimian et al. [11] evaluated the feasibility of using vision transformers for hand recognition, resulting in a CIR of 99.4% on a small subset of 11K hands consisting of 30% of the images.

Despite the results achieved for the above approaches, a proper evaluation is still needed including more realistic and challenging hand images (e.g., images in NTU-PI_v1-v1 [4] or HaGRID [3]) and evaluation scenarios (e.g., open-set scenario). Note that the above methods were only evaluated in a closed-set scenario over controlled images stemming from the following datasets: 11K Hands [2], Hong Kong Polytechnic University Hand Dorsal (HD) [18] and IIT Delhi Touchless Palmprint Database [19].

In summary, research in the field of hand-based identification has primarily focused on controlled environments, while there is limited research available for uncontrolled scenarios [4]. Existing studies in uncontrolled environments indicate poor performance due to images having varying backgrounds, hand positions, or resolution [4]. Similar challenges have been identified in other related fields such as face recognition, which have received more research attention [20].

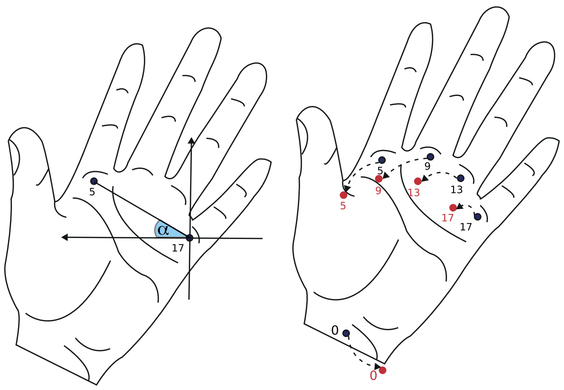
III. HAND ALIGNMENT

Recently proposed hand recognition systems treat images as they are and do not apply any pre-processing steps, such as hand pose alignment. This is partly because hand recognition is an emerging research field that can be used to assist law enforcement in recognising perpetrators of crimes. Apart from hand recognition, pose variations have been studied in more detail for the face as a biometric characteristic in the last decade. Grimmer et al. [21] analysed the impact of variables representing face pose (i.e. yaw, pitch and roll) on existing face recognition performance and demonstrated that a variation of these parameters from those associated with a frontal pose resulted in significant performance degradation of the systems. That is, similar effects could also be expected for hand pose variations.

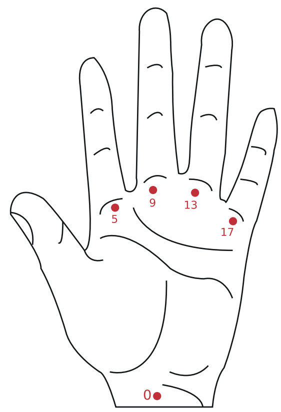
Based on the fact that some face recognition approaches have reported considerable performance improvement in uncontrolled environments when face pose alignment is adjusted, we explored whether such improvements are also observable in hand-based recognition. In this context, four different alignment methods are implemented and analysed: Simple Alignment with Cropping (SAC), Affine Alignment with and without cropping (AAC and AA) and Palmprint Alignment (PA) using the ROI-LAnet approach in [4]. Note that all hand pose alignment methods, except PA, are based on the landmarks detected by Google MediaPipe [22] as shown in Fig. 2. Notice also that if MediaPipe does not detect any landmarks, the original image is not processed, i.e. no alignment method is applied to it. MediaPipe is a state-of-the-art hand detection algorithm that was trained on uncontrolled databases taking into account the complexities that can arise in forensic scenarios, such as the influence of hand pose, lighting conditions and other factors. In our experiments, we achieved landmark detection rates of 95% and 97.50% on the NTU-PI_v1 and HaGRID databases, respectively. This indicates that most of the hand images in these databases can be correctly aligned.

A. SIMPLE ALIGNMENT WITH CROPPING (SAC)

Inspired by the work of Yoruk et al. [23], we propose an alignment method which performs both rotation and cropping of an image. The proposed approach uses landmark points 5 and 17 provided by Google MediaPipe to define a pivot line to compute the angle of symmetry with the x-axis. This



(a) Simple Alignment with Cropping (SAC) (b) Affine Alignment (AA)



(c) Aligned Hand Image

FIGURE 3. Illustrative examples of SAC 3(a) and AA 3(b) which result in an aligned hand image 3(c).

angle α is then used as a parameter for computing the rotation matrix of the image. After estimating the rotation matrix, the image is rotated and cropped by identifying the smallest bounding box enclosing the 21 landmarks. Fig.3(a) illustrates the procedure applied by SAC.

B. AFFINE ALIGNMENT WITH (AAC) AND WITHOUT CROPPING (AA)

The second proposed alignment method uses image registration to compute an affine transformation that allows for rotation, scaling, translation and shear. In contrast to SAC, this approach employs five landmarks, namely 0, 5, 9, 13, and 17 from Google MediaPipe, to estimate the affine transformation. The image registration then aims at finding an optimal mapping which minimises the least squares between the set of detected and reference landmarks. Both detected and reference landmarks were determined by manually examining various hand positions and defining the optimal positions. Similar to SAC, in this algorithm is also possible to apply cropping (AAC) to the affine alignment (AA). An example of this method is shown in Fig.3(b).

C. PALMPRINT ALIGNMENT (PA)

We adapt the ROI-LAnet approach in [4] as the third alignment method, which extracts the palmprint of the hand and applies consequently a Spatial Transformer Network (STN) [24] for hand alignment. The extraction process aims to align each palmprint in the same coordinate system to improve recognition performance. In essence, PA comprises a STN that takes an image of a hand as input, returns the palmprint and estimates normalised coordinates $\hat{\theta}$ of the hand landmarks. The coordinates are then forwarded to a grid generator which transforms a regular square grid $\mathbf{G}$ to a deformed grid $T_{\hat{\theta}}(\mathbf{G})$ based on $\hat{\theta}$. The palmprint is then sampled on a regular aligned grid using a bilinear sampler. In our experiments, we use the ROI-LAnet pre-trained weights computed on the NTU-PI_v1 database. Fig. 4 shows examples of hand samples and the respective aligned image after applying hand alignment approaches.

IV. INFORMATION FUSION

Several studies have documented the benefit of applying information fusion, which typically results in an increase in the recognition performance of biometric systems [25], [26]. In this work, we focus only on multi-algorithmic systems, where algorithms applied to the same biometric sample are fused to produce a collective decision. Since both score-level and rank-level fusion approaches have demonstrated high performance in an identification scenario [27], these are exlored in this work.

A. EXPONENTIAL WEIGHTED RANK LEVEL FUSION (EWRF)

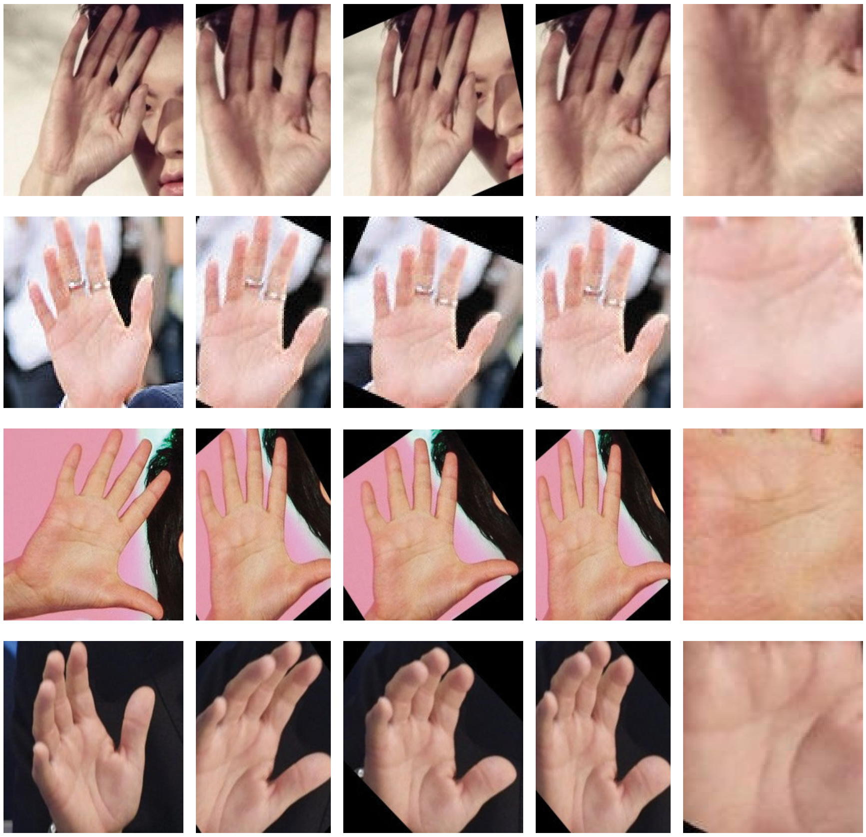
Initially, multiple non-linear rank-weighted fusion approaches were evaluated in our experiments, including Borda Count and Weighted Borda Count, among others [28]. In this case, we only report results for the best-performing technique: Exponential Weighted Rank Level Fusion (EWRF) [28]. Given N hand recognition systems, EWRF is mathematically defined as:

$$C_p = \sum_{i=1}^{N} w_i \cdot \exp(r_i(p)), \tag{1}$$

where $r_i(p)$ is the rank assigned to the p-th candidate by the i-th hand recognition system and w_i, represents the weight assigned to the i-th algorithm. To compute w_i, a logistic regression method is applied to the concatenation of different rank matrices, i.e. one rank matrix per recognition system. The use of the system-specific variable parameter w_i helps to control the shape and slope of the non-linear function. According to [28], this exponential function grows faster than polynomial functions and therefore can be more effective in exploiting the maximum variations in the output rank.

B. SCORE LEVEL FUSION USING WEIBULL NORMALIZATION (SFWN)

A näive solution for score-level fusion is the sum of the scores produced by different recognition systems. Since the scores of multiple algorithms may vary in their respective distributions, a normalisation step must be performed before



(a) Hand Images (b) SAC (c) AA (d) AAC (e) PA

FIGURE 4. Examples of original and hand images aligned by applying the proposed hand-alignment methods.

the summation to address these variations. In our work, a score-level fusion with min-max normalisation and score-level fusion with Weibull normalisation are analysed [27]. In our experiments, score level fusion with Weibull normalisation (SFWN) obtained the best identification results and was therefore adapted in all evaluations. According to [27], Weibull normalisation is well suited for identification scenarios, as it is an outlier analysis. The Weibull distribution fits the upper n scores, where the best case is when the mated score appears in the tail. In those cases, the best score in the tail is an outlier with respect to the Weibull distribution model. To generate the normalised scores, the cumulative distribution function (CDF) of the Weibull distribution is used:

$$CDF(s) = 1 - e^{(\frac{kt}{\lambda})} \tag{2}$$

where k and λ are the shape and scale parameters estimated from the Weibull distribution, respectively.

V. EXPERIMENTAL SETUP

The experimental evaluation goals are manifold: *i*) study the impact of the proposed hand-alignment approaches on the identification performance of the recognition systems *ii*) assess the feasibility of fusing the complementary information provided by the systems, and *iii*) evaluate the overall improvement of the schemes' performance for challenging scenarios when combining fusion and alignment techniques. For the latter, we follow a cross-database protocol where one database is used to train the systems and the other one for evaluating identification performance. To evaluate the performance stability of the methods, in all experiments, the test set is randomly divided into five subsets of biometric enrolment and identification transactions following a closed-set scenario, i.e., the enrolled image and the biometric transaction of each subject are randomly sampled 5 times from the subject's entire image set. The average of the identification rates together with the standard deviation computed from the above five individual experiments are then presented as a Cumulative Matching Characteristic (CMC) curve for each model. The standard deviation can be considered a measure of stability, with low values indicating that approaches are consistent across different enrolment and transaction configurations. By definition, the CMC is a graphical presentation of results of mated searches in a closed-set identification test, plotting the true positive identification rate (hereinafter referred to as the identification rate or IR) as a function of a rank value [12]. It is worth noting that we only focus on the closed-set scenario, as any

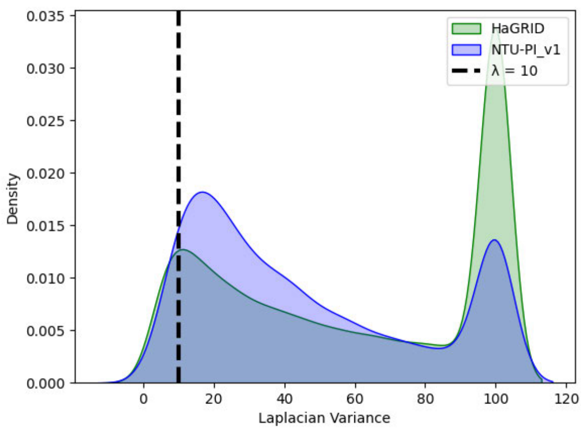


FIGURE 5. Image quality distributions for HaGRID and NTU-PI_v1 and the respective filtered threshold (black dashed line).

TABLE 2. Characteristics of databases used in the experimental evaluation.

Database	Condition	Partition	#Subjects	#Images
HaGRID [3]	Palm gestures	Train	1,032	6,205
		Transactions	555	2,604
		Enrolment	555	555
NTU-PI_v1 [4]	Natural gestures	Train	366	1,969
		Transactions	365	505
		Enrolment	365	365

performance improvement in this configuration has a direct impact on an open-set scenario.

A. DATABASES

To reach the above goals, two challenging databases are used: NTU-PI_v1 [4] and HaGRID [3]. The former contains 7,781 hand images from 2,035 different palms and 1,093 different subjects. These images were acquired under challenging settings, such as uncontrolled backgrounds, finger positions, and image quality. Since HaGRID [3] was initially proposed for gesture recognition purposes, it contains many different gestures. We only selected those similar to NTU-PI_v1 from both its train and testing set, resulting in 28,326 images. In the experiments, hand images of extremely low quality were filtered out by using a Laplacian variance approach [29]. A low Laplacian variance indicates an edge absence and therefore a blurred image. Fig. 5 shows the image quality distributions computed on HaGRID and NTU-PI_v1 as well as the selected threshold ($\lambda = 10$) to remove the low-quality images. λ was selected by a visual inspection of the images. Tab. 2 provides details of the databases used in the performance benchmark after the removal of the low-quality images and single-sample subjects. In the experimental evaluation, only the right-hand images are considered.

1) IMPLEMENTATION DETAILS

All algorithms used in the performance benchmark were implemented in PyTorch [30] and trained utilising a Nvidia A100 Tensor Core GPU with 40GB of GPU Memory. For the training and testing of the systems, we took the parameters as indicated in their corresponding articles. The image size was set to 256 × 256 pixels for ABD-Net [13] and RGA-Net [14], and 356 × 356 for MBA-Net [10]. In addition, the networks were initialised with their pre-trained weights in ImageNet [31] and trained for 100 epochs using the Adam optimiser. As indicated in [14], the RGA-Net architecture was trained for 600 epochs in all cases.

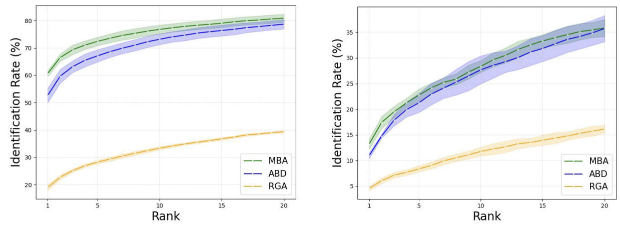


FIGURE 6. CMC curves of the reference systems on the 5 random enrolment-transaction sets extracted from the test set. The mean IR per model is represented by lines, while the coloured area indicates the standard deviation.

TABLE 3. Mean identification rates (in %) at different rank values for different systems (best IRs are highlighted in bold).

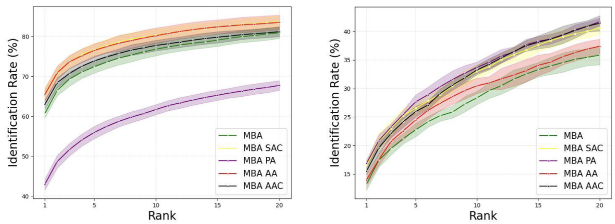
Systems	Databases	Rank-1	Rank-5	Rank-10
ABD-Net	HaGRID	52.70	67.10	73.20
	NTU-PI_v1	11.00	21.20	27.70
RGA-Net	HaGRID	19.00	28.20	33.30
	NTU-PI_v1	4.60	8.40	11.80
MBA-Net	HaGRID	**60.70**	**72.40**	**76.80**
	NTU-PI_v1	**13.20**	**22.70**	**28.40**

VI. RESULTS AND DISCUSSION

The goal of this scientific study is to investigate how to improve the performance of hand-based recognition systems in forensic scenarios, specifically in uncontrolled environments (Sect. VI-A). To achieve the objectives, the experimental results after applying the proposed hand-alignment (Sect. VI-B) and fusion (Sect. VI-C) techniques on uncontrolled hand images are summarised in this section. In addition, an ablation study on the use of the best-performing methods is presented (Sect. VI-D). In the above experiments, an intra-database protocol is followed, i.e. the training and testing subjects stem from the same database. Further, we present a cross-database evaluation that aims to analyse the utility of different databases to improve the generalisability of systems to unseen images of hands (Sect. VI-E).

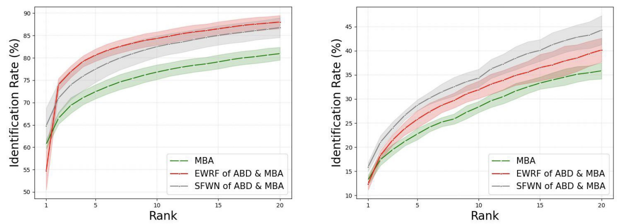
A. BASELINE DETECTION PERFORMANCE

The initial evaluation scenario aims to establish a baseline of the hand-based identification models MBA-Net, ABD-Net and RGA-Net on the NTU-PI_v1 and HaGRID databases.



(a) HaGRID (b) NTU-PI_v1

FIGURE 7. **CMC curves of MBA-Net after applying the proposed hand-alignment approaches.**



(a) HaGRID (b) NTU-PI_v1

FIGURE 8. **CMC curves after applying SFWN and EWRF to baseline models on HaGRID and NTU-PI_v1.**

TABLE 4. **Mean identification rates (in %) at different rank values for different systems. The best IRs are highlighted in bold.**

Systems	Databases	Rank-1	Rank-5	Rank-10
MBA-Net	HaGRID NTU-PI_v1	60.70 13.20	72.40 22.70	76.80 28.40
MBA-Net + SAC	HaGRID NTU-PI_v1	**65.50** ↑ 16.60 ↑	**76.90** ↑ 26.50 ↑	**80.40** ↑ 33.00 ↑
MBA-Net + AA	HaGRID NTU-PI_v1	65.40 ↑ 14.00 ↑	76.50 ↑ 24.50 ↑	80.20 ↑ 30.50 ↑
MBA-Net + AAC	HaGRID NTU-PI_v1	63.00 ↑ 15.50 ↑	73.90 ↑ 26.00 ↑	77.80 ↑ 33.20 ↑
MBA-Net + PA	HaGRID NTU-PI_v1	42.90 ↓ **16.90** ↑	56.00 ↓ **27.60** ↑	61.70 ↓ **33.70** ↑

↑ shows an improvement in performance over the first row baseline.
↓ represents a deterioration in performance over the first row baseline.

This baseline serves as a benchmark to evaluate the improvement of hand alignment and fusion. By comparing the performance of the models before and after applying these methods, the observed improvements can be identified. Fig. 6 shows the CMC curves computed by the systems for HaGRID and NTU-PI_v1. The respective mean IR values are summarised in Tab. 3. Note that the MBA-Net reports the best IRs for different rank values ranging from 60.70% to 76.80% for HaGRID and 13.20% to 28.40% for NTU-PI_v1. Also a substantial difference in the performance of NTU-PI_v1 compared to HaGRID can be observed: the IRs are notably higher in the latter. This is because HaGRID is a simpler uncontrolled database, characterised by fewer variations in hand position and background (see Fig. 1). HaGRID also includes one gesture, the palm of the hand, while NTU-PI_v1 includes natural gestures with significantly more variation.

B. HAND ALIGNMENT

The second set of experiments aims to evaluate the performance of the systems when the proposed hand-alignment approaches are independently applied. To do that, we select the best-performing scheme MBA-Net and compute the CMC curves per hand-alignment method. Fig. 7 illustrates the identification performance obatined by applying the hand-alignment methods on the best performance system, i.e. MBA-Net, while Tab. 4 summarises the mean IRs for three rank values.

TABLE 5. **Mean identification rates (in %) at different rank values for different systems. The best IRs are highlighted in bold.**

Systems	Databases	Rank-1	Rank-5	Rank-10
MBA-Net	HaGRID NTU-PI_v1	60.70 13.20	72.40 22.70	76.80 28.40
EWRF (MBA-Net + ABD-Net)	HaGRID NTU-PI_v1	54.70 ↓ 12.40 ↓	**80.60** ↑ 25.80 ↑	**84.30** ↑ 31.90 ↑
SFWN (MBA-Net + ABD-Net)	HaGRID NTU-PI_v1	**64.80** ↑ **15.90** ↑	77.50 ↑ **28.70** ↑	82.50 ↑ **34.30** ↑

Note that all proposed hand alignment approaches improve the MBA-Net baseline identification performance (first row), except PA over HaGRID. In particular, improvements of up to 3 and 6 percentage points over the baseline are observed for NTU-PI_v1 and HaGRID, respectively. Among the hand alignment methods, SAC reports the highest performance for HaGRID and the second highest for NTU-PI_v1. For the latter, the PA solution appears to be the best choice. PA is mainly proposed to align NTU-PI_v1 images and therefore shows a low generalisability in aligning images with different characteristics to the database used during training. That is, a performance deterioration of PA is noticeable when it is applied to images in HaGRID. Notice also that the difference in performance between SAC and PA is marginal for NTU-PI_v1: PA slightly outperforms SAC by 0.30 percentage points (i.e., 16.90% vs. 16.60%). As in the case of the face, hand alignment seems to be an important preprocessing step to improve recognition performance. It should therefore be strongly considered to improve hand recognition, especially in uncontrolled images as it is the case in forensic investigations.

C. FUSION

Similar to the Sect. III, in this section we evaluate the identification performance when different fusion approaches are applied. As mentioned in Sect. IV, two different fusion methods, namely EWRF and SFWN, are evaluated. In this context, EWRF and SFWN are applied to two different combinations of schemes: the two and the three best-performing systems. For each fusion technique, the CMC curves of the best-performing combination is presented in Fig. 8 and the summarised results are reported in Tab. 5.

Note that the different fusions between MBA-Net and ABD-Net lead to an improvement of the best baseline performance, i.e. MBA-Net in most cases. In particular, SFWN appears to be the best fusion method, achieving an IR in the ranges of 15.90%-34.30%. Notice that EWRF yields IRs higher than the ones attained by MBA-Net for ranks $\geq$ 5. In fact, IRs from both fusion approaches for rank values $\geq$ 5 outperform the ones yielded by both the best hand alignment method (i.e., SAC) and the baseline MBA-Net for the same ranks.

Note also that the combination between alignment and the fusion methods, which take place at different processing steps of the recognition pipeline, results in a considerable performance improvement over the best baseline. In particular, in the case of HaGRID, we observe that a wanted perpetrator can be identified with an accuracy of 90% if only the former 20 entries of the candidate list are taken into consideration. These results are beneficial from a forensic point of view, since in a forensics scenario it is likely that a short candidate list is checked by human examines, e.g. to detect suspects or missing persons.

D. ABLATION STUDY

In the above experiments, a performance improvement is observed when applying alignment and fusion techniques individually. Building upon those findings, we evaluate the potential performance increase when both techniques are combined. Since RGA-Net reported poor performance, this system was not considered in the ablation study.

The ablation study consists of multiple experiments involving the fusion of various systems with prior hand alignment. Initially, we run all possible single combinations of hand alignment and fusion. Since most of the hand alignment methods improve the identification performance of the baseline in Tab. 4, we also combine them with the proposed fusion methods. Tab. 6 reports the identification performance of the best combinations after enabling hand alignment and fusion. In fusions that enable prior hand alignment, MBA-Net is merged with ABD-Net plus SAC for HaGRID and ABD-Net plus PA for NTU-PI_v1.

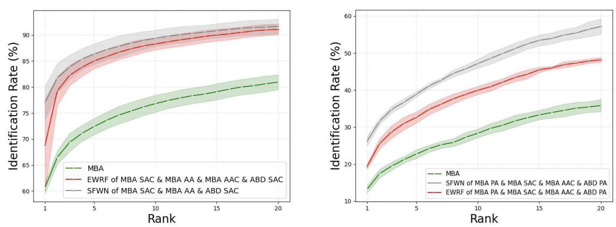
In most cases, hand alignment or fusion improves the baseline performance of MBA-Net by up to 3 and 4 percentage points for NTU-PI_v1 and HaGRID, respectively. However, when the two are combined, the increase in performance is even greater, resulting in approximately up to 13 and 17 percentage points of gain for NTU-PI_v1 and HaGRID, respectively. In particular, for the most complex database, NTU-PI_v1, the combination of MBA-Net with ABD-Net, which enables several previous hand alignment components, shows an IR of 26.30%, an increase in relative performance of more than 50% over the baseline IR obtained by MBA-Net.

To validate the identification performance improvement of the best performing combination (i.e., SFWN plus hand alignment) over the baseline, a non-parametric MannWhitney [32] test is applied with 95% confidence. As a result, we do confirm that the identification performance obtained by the combination of SFWN and hand alignment is statistically superior to that obtained with MBA-Net for both databases. Contrary to previous trends, the independent use of some hand alignment techniques (e.g., AA) does not yield statistically superior results with respect to the baseline.

TABLE 6. Ablation study of different components. The best IRs are highlighted in bold. Fusions enabling prior hand alignment are performed with ABD-Net plus SAC for HaGRID and ABD-Net plus PA for NTU-PI_v1.

Systems	Databases	Hand Alignment				Fusion		Rank-1
		SAC	AA	AAC	PA	EWRF	SFWN	
MBA-Net	HaGRID	✗	✗	✗	✗	✗	✗	60.70†
		✓						65.50 ↑
			✓					65.40 ↑
				✓				63.00 ↑
					✓			42.90 ↓
						✓		54.70 ↓
							✓	64.80 ↑
		✓	✓	✓		✓		68.80 ↑
		✓	✓				✓	**77.00** ↑
	NTU-PI_v1	✗	✗	✗	✗	✗	✗	13.20†
		✓						16.60 ↑
			✓					14.00 ↑
				✓				15.50 ↑
					✓			16.90 ↑
						✓		12.40 ↓
							✓	15.90 ↑
		✓		✓	✓	✓		19.50 ↑
		✓		✓	✓		✓	**26.30** ↑

† represents the baseline IR.



(a) HaGRID (b) NTU-PI_v1

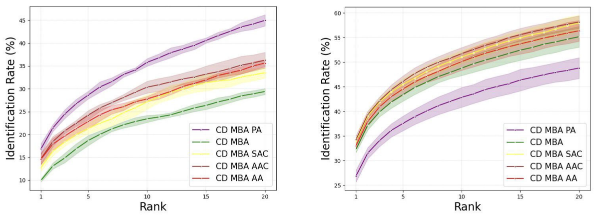
FIGURE 9. CMC curves of MBA-Net after applying the prior hand alignment and fusion approaches.

The benefits in performance of using both hand alignment and fusion approaches can be also observed in Fig. 9. From a forensic point of view in which not only the earlier positions in the candidate list are important, the fusion of MBA-Net with ABD-Net after applying a previous hand alignment on the images reports IRs higher than 55% and 90% for NTU-PI_v1 and HaGRID, respectively, in rank values greater than or equal to 20. This implies that forensic investigators can detect a perpetrator based on the hands with an accuracy of up to 90% by checking at most the first 20 entries of the candidate list.

Note that the complementary information provided by each of the fusion methods could also be combined in a further step to improve the identification performance reported in Tab. 6. However, the use of an additional score normalisation step makes this task more difficult, as it requires an extra database. Therefore, this was not considered in this work.

E. CROSS-DATABASE ANALYSIS

We finally evaluate the impact of the hand alignment under a cross-database scenario where one database is used



(a) HaGRID - NTU-PI_v1 (b) NTU-PI_v1 - HaGRID

FIGURE 10. **CMC curves for the cross-database evaluation.**

for training the systems and the other one for testing. Fig.10 shows the CMC curves for two train-test configurations: HaGRID - NTU-PI_v1 (Fig.10(a)) and NTU-PI_v1 - HaGRID (Fig.10(b)).

Note that the HaGRID - NTU-PI_v1 configuration shows generally a performance decrease when compared to the corresponding system trained on NTU-PI_v1. However, when comparing results with the best reference system (MBA-Net) trained on NTU-PI_v1, several schemes, including MBA-Net with PA, MBA-Net with AAC and MBA-Net with AA show improved performance. In a nutshell, MBA-Net with prior hand alignment trained on HaGRID provides higher identification rates on NTU-PI than the same pipeline trained on NTU-PI_v1. This also implies that HaGRID images are sufficient to improve the generalisability of hand recognition systems and thus to detect perpetrators based on images similar to those of NTU-PI_v1.

Contrary to the HaGRID - NTU-PI_v1, the NTU-PI_v1 - HaGRID configuration reports a significant performance deterioration compared to the corresponding models trained on HaGRID. Specifically, MBA-Net with AAC yields IRs ranging from 34.3% to 51.7% for range values between 1 and 10. Compared to the results shown for the intra-database evaluation of HaGRID, the decrease in performance is as much as 28.7 percentage points lower, indicating that the NTU-PI_v1 images do not contain the necessary variation in subjects, background and image quality to generalise well to different uncontrolled databases.

VII. CONCLUSION

This work presents a study of various approaches to improve the identification performance of existing hand recognition systems in uncontrolled environments. In such a study, different hand alignment and fusion approaches were analysed on two challenging databases, HaGRID and NTU-PI_v1. The results demonstrated that the system's performance can be improved by applying either hand alignment or fusion techniques. In particular, the hand alignment methods yielded the highest performance improvements, which achieves an IR at Rank-1 of 16.60% for NTU-PI_v1 and 65.50% for HaGRID. These IRs represent a difference or absolute increase of 3.40 and 4.80 percentage points, respectively, compared to the best baseline in the same ranks. Additionally, the SFWN-based fusion applied to MBA-Net and ABD-Net reported the highest performance increase compared to other fusion techniques and achieved IRs at Rank-1 of 15.90% and 64.80% for NTU-PI_v1 and HaGRID, respectively.

Finally, an ablation study showed that the combination of hand alignment and fusion led to the highest overall increase in identification performance: IRs of up to 26.30% and 77.00% were achieved For NTU-PI_v1 and HaGRID, respectively, at Rank-1. From a forensic point of view in which not only the earlier positions in the candidate list are important, the combination between hand alignment and fusion reported IRs higher than 55% and 90% for NTU-PI_v1 and HaGRID, respectively, in ranks ≥ 20. This implies that forensic investigators can detect a perpetrator based on the hands with an accuracy of up to 90% by checking the first 20 entries of the candidate list.

In the future, we will explore the combination of the proposed methods and workload reduction techniques to reduce the number of false positives. In forensic investigations, investigators must efficiently handle a large amount of data generated daily for the identification of offenders and victims. According to the study in [33], the increase in the number of subjects enrolled in an identification system constantly increases the response time of the system on the one hand and the false positive acceptance rates on the other hand. In addition, the quality of the image must be analysed. Extremely low quality images of hands cannot be used as evidence to incriminate an offender, even if the recognition system claims his identity. In contrast to the face, hand images do not contain aesthetic units that can be used to visually establish the identity of a wanted perpetrator. Therefore, hand image quality assessment methods focusing on forensic images should be prioritised in future work.

REFERENCES

[1] *ISO/IEC JTC1 SC37 Biometrics, Information Technology—Vocabulary—Part 37: Biometrics*, ISO/IEC Standard 2382-37:2012, Int. Org. Standardization, Geneva, Switzerland, 2382.

[2] M. Afifi, "11K Hands: Gender recognition and biometric identification using a large dataset of hand images," *Multimedia Tools Appl.*, vol. 78, pp. 20835–20854, Mar. 2019, doi: 10.1007/s11042-019-7424-8.

[3] A. Kapitanov, K. Kvanchiani, A. Nagaev, R. Kraynov, and A. Makhliarchuk, "HaGRID—HAnd gesture recognition image dataset," 2022, *arXiv:2206.08219*.

[4] W. M. Matkowski, T. Chai, and A. W. K. Kong, "Palmprint recognition in uncontrolled and uncooperative environment," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1601–1615, 2019.

[5] M. Tistarelli and C. Champod, *Handbook of Biometrics for Forensic Science*. Cham, Switzerland: Springer, 2017.

[6] A. Kumar, T. Mundra, and A. Kumar, *Anatomy Hand*. Boston, NA, USA: Springer, 2009, pp. 28–35, doi: 10.1007/978-0-387-73003-5_267.

[7] S. Sharma, S. R. Dubey, S. K. Singh, R. Saxena, and R. K. Singh, "Identity verification using shape and geometry of human hands," *Expert Syst. Appl.*, vol. 42, no. 2, pp. 821–832, Feb. 2015.

[8] M. Anitha and K. Rao, "Fusion of finger inner knuckle print and hand geometry features to enhance the performance of biometric verification system," *Intl. J. Electr. Comput. Eng.*, vol. 10, no. 10, pp. 1351–1356, 2016.

[9] N. L. Baisa, B. Williams, H. Rahmani, P. Angelov, and S. Black, "Hand-based person identification using global and part-aware deep feature representation learning," in *Proc. 11th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Apr. 2022, pp. 1–6.

[10] N. L. Baisa, B. Williams, H. Rahmani, P. Angelov, and S. Black, "Multi-branch with attention network for hand-based person recognition," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 727–732.

[11] Z. Ebrahimian, S. A. Mirsharji, R. Toosi, and M. A. Akhaee, "Automated person identification from hand images using hierarchical vision transformer network," in *Proc. 12th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Nov. 2022, pp. 398–403.
[12] *ISO/IEC JTC1 SC37 Biometrics, Information Technology—Biometric Performance Testing and Reporting—Part 1: Principles and Framework*, ISO/IEC Standard 19795-1:2021, Int. Org. Standardization, Geneva, Switzerland, Jun. 2021.
[13] T. Chen, S. Ding, J. Xie, Y. Yuan, W. Chen, Y. Yang, Z. Ren, and Z. Wang, "ABD-Net: Attentive but diverse person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8350–8360.
[14] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, "Relation-aware global attention for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3183–3192.
[15] R. Miller, "Finger dimension comparison identification system," U.S. Patent 3 538 576, Apr. 27, 1971.
[16] D. Sidlauskas and S. Tamer, "Hand geometry recognition," in *Handbook of Biometrics*. Springer, 2008, pp. 91–107.
[17] K. Prihodova and M. Hub, "Hand-based biometric recognition technique—Survey," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, pp. 689–698, Nov. 2020.
[18] A. Kumar and Z. Xu, "Personal identification using minor knuckle patterns from palm dorsal surface," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2338–2348, Oct. 2016.
[19] A. Kumar, "Incorporating cohort information for reliable palmprint authentication," in *Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process.*, Dec. 2008, pp. 583–590.
[20] A. K. Jain, B. Klare, and U. Park, "Face recognition: Some challenges in forensics," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Mar. 2011, pp. 726–733.
[21] M. Grimmer, C. Rathgeb, and C. Busch, "Pose impact estimation on face recognition using 3D-aware synthetic data with application to quality assessment," 2023, *arXiv:2303.00491*.
[22] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "MediaPipe hands: On-device real-time hand tracking," 2020, *arXiv:2006.10214*.
[23] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon, "Shape-based hand recognition," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1803–1815, Jul. 2006.
[24] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 28, 2015, pp. 2017–2025.
[25] L. J. González-Soler, M. Gomez-Barrero, L. Chang, A. Pérez-Suárez, and C. Busch, "Fingerprint presentation attack detection based on local features encoding for unknown attacks," *IEEE Access*, vol. 9, pp. 5806–5820, 2021.
[26] L. J. González-Soler, M. Gomez-Barrero, J. Kolberg, L. Chang, A. Pérez-Suárez, and C. Busch, "Local feature encoding for unknown presentation attack detection: An analysis of different local feature descriptors," *IET Biometrics*, vol. 10, no. 4, pp. 374–391, Jul. 2021.
[27] A. Pradhan, J. He, and N. Jiang, "Score, rank, and decision-level fusion strategies of multicode electromyogram-based verification and identification biometrics," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 1068–1079, Mar. 2022.
[28] A. Kumar and S. Shekhar, "Personal identification using multibiometrics rank-level fusion," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 5, pp. 743–752, Sep. 2011.
[29] R. Bansal, G. Raj, and T. Choudhury, "Blur image detection using Laplacian operator and open-CV," in *Proc. Int. Conf. Syst. Model. Advancement Res. Trends (SMART)*, Nov. 2016, pp. 63–67.
[30] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, Dec. 2019, pp. 8024–8035.
[31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[32] M. Asadoorian and D. Kantarelis, *Essentials of Inferential Statistics* (G—Reference, Information and Interdisciplinary Subjects Series). MD, USA: Univ. Press of America, 2005. [Online]. Available: https://books.google.de/books?id=MUyrEx2ATwkC
[33] P. Drozdowski, C. Rathgeb, and C. Busch, "Computational workload in biometric identification systems: An overview," *IET Biometrics*, vol. 8, no. 6, pp. 351–368, Nov. 2019.

LAZARO JANIER GONZALEZ-SOLER received the B.Sc. degree in mathematics and computer science from the University of Havana, in 2014, and the Ph.D. degree in applied computer science from Hochschule Darmstadt, in 2022. He is currently a Postdoctoral Researcher with the da/sec Group, National Research Center for Applied Cybersecurity ATHENE, Hochschule Darmstadt, Germany. He has been actively involved in National and European projects, such as Bio4ensics and secure and privacy-friendly mobile authentication (RESPECT). His research interest includes improving the security of biometric systems through the development of biometric presentation attack detection (PAD) techniques. Further research interests are related to the development of algorithms for biometric recognition in forensic scenarios. He has been also granted several awards, such as the Best Ph.D. Thesis Award in Computer Science of German Universities of Applied Sciences, the Best Paper Award at the International Workshop on Information Forensics and Security (WIFS) 2021, and the Best-Performing Algorithm in the LivDet 2019 and 2021. He served as a reviewer for different JRC journals and conferences.

DURITA KVILT JÓNSDÓTTIR received the master's degree in human centered artificial intelligence from the Department of Applied Mathematics and Computer Science, Danmarks Tekniske Universitet (DTU). With a primary focus on artificial intelligence, her passion lies in applying innovative solutions to real-world use cases, such as biometric applications.

CHRISTIAN RATHGEB is currently a Professor with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He is the Principal Investigator of the National Research Center for Applied Cybersecurity ATHENE. He has coauthored over 100 technical articles in the field of biometrics. His research interests include pattern recognition, iris and face recognition, security aspects of biometric systems, secure process design, and privacy enhancing technologies for biometric systems. He is a member of European Association for Biometrics (EAB). He is the Program Chair of the International Conference of the Biometrics Special Interest Group (BIOSIG) and an Associate Editor of *IET Biometrics* (IET BMT) and *Pattern Recognition* (Elsevier). He served for various program committees and conferences, such as ICB, IJCB, BIOSIG, and IWBF.

DANIEL FISCHER received the bachelor's degree in computer science from Darmstadt University of Applied Sciences in cooperation with Deutsche Telekom AG, in 2015, and the master's degree with a focus on IT-security and biometrics from Darmstadt University of Applied Sciences, in 2018. Since 2015, he has been a member of the IT-Security Group (da/sec), Darmstadt University of Applied Sciences, which is a part of the National Research Center for Applied Cybersecurity ATHENE.