D4FLY multimodal biometric database: multimodal fusion evaluation envisaging on-the-move biometric-based border control

D4FLY Multimodal Biometric Database: multimodal fusion evaluation envisaging on-the-move biometric-based border control

Lulu Chen[1], Jonathan Boyle[1], Antonios Danelakis[2], James Ferryman[1], Simone Ferstl[3], Damjan Gicic[3], Artur Grudzień[4], André Howe[3], Kowalski Marcin[4], Krzysztof Mierzejewski[4], and Theoharis Theoharis[2]

[1] University of Reading, Whiteknights, Reading, UK
{l.chen, j.n.boyle, j.m.ferryman}@reading.ac.uk

[2]NTNU, Norwegian University of Science and Technology, Trondheim, Norway
{antonios.danelakis, theotheo}@ntnu.no

[3] Veridos GmbH, Munich, Germany
{simone.ferstl, damjan.gicic, andre.howe}@veridos.com

[4] WAT, Military University of Technology, Warsaw, Poland
{artur.grudzien, marcin.kowalski, Krzysztof.mierzejewski}@ntnu.no

Abstract

This work presents a novel multimodal biometric dataset with emerging biometric traits including 3D face, thermal face, iris on-the-move, iris mobile, somatotype and smartphone sensors. This dataset was created to resemble on-the-move characteristics in applications such as border control. The five types of biometric traits were selected as they can be captured while on-the-move, are contactless, and show potential for use in a multimodal fusion verification system in a border control scenario. Innovative sensor hardware was used in the data capture. The data featuring these biometric traits will be a valuable contribution to advancing biometric fusion research in general. Baseline evaluation was performed on each unimodal dataset. Multimodal fusion was evaluated based on various scenarios for comparison. Real-time performance is presented based on an Automated Border Control (ABC) scenario.

1. Introduction

A biometric system uses physiological and/or behavioural traits of individuals to recognise their identities. Unimodal biometric systems rely on the evidence from a single biometric trait. Multimodal biometric systems use more than one biometric trait and are expected to be more reliable due to combining multiple fairly independent biometric characteristics [11] by applying specific data fusion schemes. Advantages of using multimodal biometric systems over unimodal biometric systems have been discussed in the literature [20], including overcoming the limitations from noisy data, non-universality, intra-class variations, inter-class similarities and presentation attacks.

The EU H2020 project D4FLY focuses on enhancing the quality and efficiency of identity verification across a range of border crossing types. One of the main objectives of the D4FLY project is to improve the quality of biometric verification and reduce time spent by travellers and border staff in engaging with the verification process. Enabling a real on-the-move border crossing experience for travellers is essential to achieve the goal. Most current commercial biometric verification systems are deployed based on static data capture using the common biometrics, *i.e*. 2D face, fingerprint or iris. The D4FLY project explores and assesses the potential of using emerging modalities that are less intrusive and contactless allowing travellers to walk through the biometric verification system without stopping.

The design of the D4FLY biometric verification corridor concept is illustrated in Figure 1. The traveller walks through the corridor area and multiple biometric traits are captured without the traveller needing to stop. The green arrows show the traffic flow from the entrance to the exit of the corridor. The yellow box represents the sensor suite where all biometric sensors (except the smartphone sensors) are integrated together and is positioned facing the traveller entering the corridor. The zigzag-shaped design is to ensure optimal biometric acquisition, *i.e*. the biometric sensors can

978-1-6654-3396-9/21/$31.00 ©2021 IEEE

capture good frontal images of the person. The smartphone will be carried by the traveller and the continuous smartphone sensor-based verification process is completed prior to the traveller entering the corridor. The NFC reader is used for communication between the corridor system and the smartphone. Before the traveller exits the corridor, a final decision is calculated as to whether the traveller should be subject to further manual checks or else is eligible for entry.

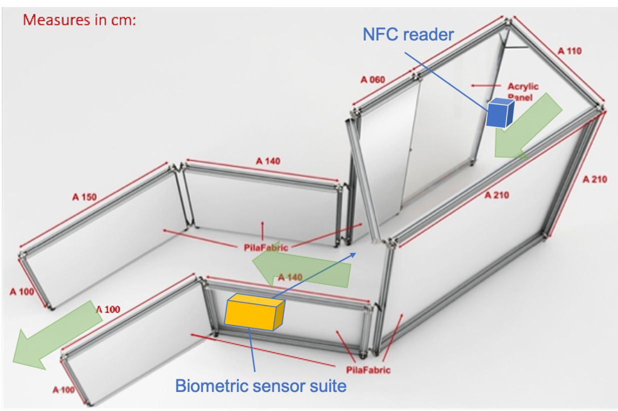

Figure 1. D4FLY design of the on-the-move biometric verification corridor for border control

Innovative sensors were also developed in the project to ensure the quality of the on-the-move biometric data capture. In the D4FLY dataset, most of the modalities collected are acquired in both static and walking scenarios which will be useful for research in both areas and in comparative study. While there is work remaining that focuses on optimising the current deployed biometric modalities, assessing the potential of using emerging modalities and innovative sensor hardware is a crucial part of the project and hence motivated the creation of the D4FLY multimodal dataset. To ensure maximum security and biometric verification accuracy, multimodal fusion is the central component that combines all the novel biometric systems. Multimodal biometric fusion can reduce biometric verification error rates (False Match rate and/or False non-Match rate) and improves resilience to presentation attacks. It can also be applied to reduce failure-to-acquire/enrol rates whilst increasing the verification throughput which is the other important aspect for a border control system.

The main contributions in this work include: 1) A novel on-the-move multimodal biometric dataset; 2) Baseline evaluation on the unimodal biometrics; 3) Multimodal fusion evaluation based on different scenarios. The main objective is to provide baseline results for the created dataset, as well as to assess the potential of how these emerging biometric modalities can perform in real-life scenarios and their contribution to a multimodal fusion system. The rest of the paper is organised as follows: Section 2 compares the D4FLY dataset with existing multimodal datasets. Section 3 introduces the details of D4FLY multimodal biometric dataset. Section 4 describes the evaluation methods for both unimodal and multimodal fusion. The evaluation results for both unimodal and multimodal fusion based on different scenarios are presented in Section 5.

2. Related work

In the current state of border control, face, fingerprint, and iris are the three biometrics that are considered in International Civil Aviation Organization (ICAO) standards for Machine Readable Travel Documents (MRTD). Face is the essential biometric trait, while additional data can be provided to the verification processes by including multiple biometrics in their travel documents, *i.e.* a combination of face and/or fingerprint and/or iris.

Dataset	**Biometric traits**	**Number subjects**
D4FLY (2021)	3D face, thermal face, iris on-the-move, iris mobile, somatotype, smartphone sensors	31 (1 session)
PROTECT (2018)	2D face, 3D face, thermal face, iris, periocular, hand vein, finger vein, voice, anthropometrics	47 (Session 1) 38 (Session 2)
Biosecure-BMDB (2010)	Face, fingerprint, hand, voice, iris, face talking/still, signature, handwriting, keystroking	400 users (4 sessions)
MBioID (2007)	Face (2D, 3D), fingerprint, iris, signature, voice	120 users (2 sessions)
JMBDC (2007)	Iris, face, voice, fingerprint, hand geometry and palm print	270 users
BIOMET (2003)	Face (2D, 3D), voice, fingerprint, hand shape and signature	91 users (3 sessions)

Table 1. Comparison of existing multimodal biometric dataset

There are a variety of existing multimodal biometric datasets such as Biosecure-BMDB, MBioID, JMBDC and BIOMET (Table 1) which all contain a large number of biometric traits (six or more). However, most of these datasets contain more traditional biometrics and were collected in a conventional capture environment (*e.g.* an office setup). The more recent PROTECT Multimodal database [5], which is also created under a border control application setting, contains on-the-move biometric capture (*i.e.* 2D face, periocular and anthropometrics), however, the rest are still static (*i.e.* finger/hand vein, thermal/3D face and iris) as per the requirements of the PROTECT border scenarios.

The U.S. Department of Homeland Security (DHS) Science and Technology Directorate (S&T) has hosted a series

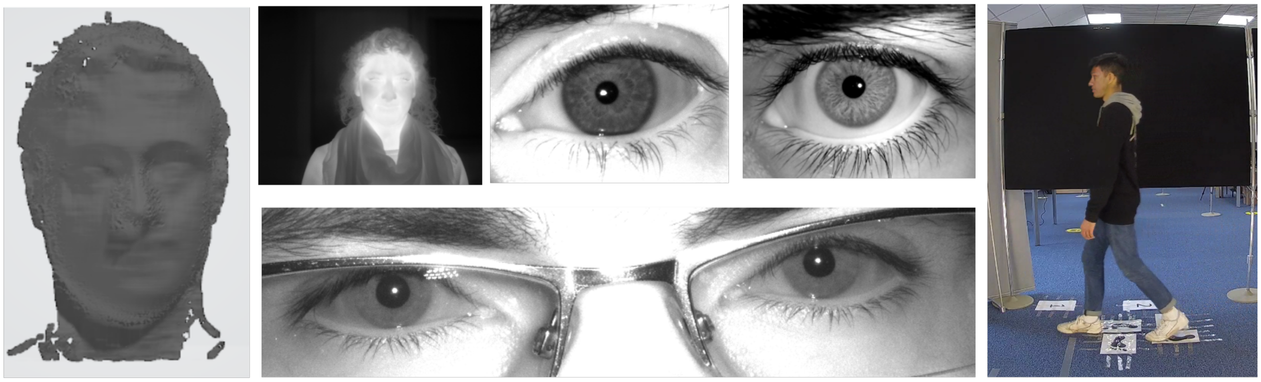

Figure 2. Sample images from the D4FLY multimodal dataset: rendered 3D face, thermal face, iris on-the-move, iris mobile, somatotype

Biometric	Description	Scenario	Number of samples per subject	Data format
3D face	3D facial meshes	Static/ Walking	2 samples	.ply files
Thermal face	Image sequences	Static/ Walking	1000-2000 images (640×512)	.bmp
Iris on-the-move	NIR images	Static/ Walking	5 images (640×480) per eye On-the-move: 20 dual-eye images	.png and .bmp
Iris mobile	NIR images	Static	5 images per eye (640×480)	.raw and .jpg
Somatotype	RGB images	Static/ Walking	3 images (1920×1080 pixels)	.jpg
Smartphone sensors	Sensor readings	Walking	5 sessions, 3 repeats per session	.txt

Table 2. List of biometric traits included in the dataset

of Biometric Technology Rallies since 2018 that aimed to challenge current biometric technologies in operational settings. The most relevant to this work are the 2018 and 2019 challenges [9, 6] that focussed on use cases such as traveller identification in a high-throughput security environment using an unmanned system for face and iris recognition (fingerprint was additionally included in 2019). The Biometric Technology Rally environment setting is similar to the D4FLY project concept, however, only the three traditional biometrics were tested in the rally and the data is only available to each participant using their own system.

This D4FLY dataset contains a set of novel biometrics that have not been broadly applied in the context of border control. Moreover, this dataset focuses on on-the-move data capture, despite it containing a relatively small number of subjects compared to the previous works. To the authors' best knowledge, this is the first multimodal biometric dataset where all types of biometric traits were captured on-the-move. The dataset has been made publicly available and can be downloaded via D4FLY website [1].

3. Multimodal biometric dataset

Six modalities including 3D face, thermal face, iris (2 types), somatotype, and smartphone sensors were captured for the dataset. Iris was captured in two different settings: iris on-the-move capture and iris mobile which used a compact portable iris scanner. As this work focusses on on-the-move multimodal biometric fusion analysis, iris mobile is not included in the evaluation, however, the description of the data is provided. Table 2 lists all the modalities in the dataset and their corresponding data format.

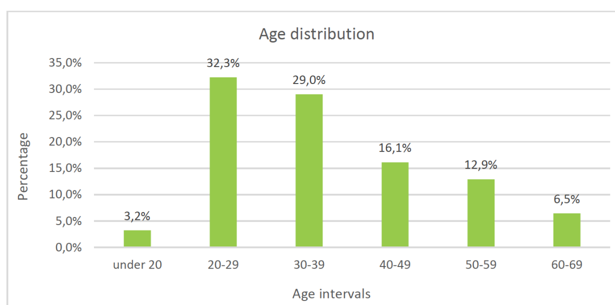

Figure 3. Subject age distribution in 10-year intervals

3.1. Dataset acquisition setup

Each modality was recorded with dedicated equipment setup as described below. All biometrics except iris mobile have been captured with an on-the-move scenario to align with the D4FLY project concept. Each subject attending the data capture had all six types of biometrics captured.

3.2. Dataset characteristics

The dataset consists of a total number of 31 subjects. The set of subjects included diverse age, gender, and ethnicity background. The age interval was from 19 to 64. The age distribution is depicted in Figure 3. The gender distribution is: male (58.1%), female (38.7%), and non-binary (3.2%). Figure 2 shows some sample images of the collected data.

4. Experiment setup: unimodal and multimodal evaluation

This section describes the baseline evaluation for each unimodal biometric dataset (a brief description and methods used) and the evaluation protocol for the multimodal fusion.

4.1. Unimodal evaluation methods

3D face The 3D face dataset contains two 3D facial model captures for each identity: one static, and one on-the-move. The data were captured using a single innovative Raytrix R26 monochromatic light-field camera, equipped with a ZEISS Interlock 2/135 lens. Initially, for both the static and the on-the-move scenarios, multiple captures were recorded but only the one with the best quality was retained. The best quality capture was manually selected using the RxLive 5.0 Raytrix Software. During each capture, the subjects were looking directly towards the camera. In general, as expected, the static captures were of better quality than the on-the-move captures. For baseline evaluation, the static captures were used as the gallery samples, while the on-the-move captures were used as the probe samples. An optimised variation of the UR3D-C method [16] was implemented for evaluating the verification performance.

Thermal face Visible 2D face and thermal face were captured simultaneously for the purpose of thermal to visible face recognition. Due to privacy protection requirement, only the thermal faces are included in the dataset. Face in frontal position and with head turning slightly left and right were captured while the person was walking towards the camera. Thermal face images were acquired using a FLIR A65 thermal camera. Visible face images were acquired using a Basler acA2040-90uc camera with a $16\,\mathrm{mm}$ focal length. Distance from subject to cameras is 1.5 metres. 20 thermal images from the recorded image sequence were selected for each subject (20 visible face images were also selected) for the evaluation, where 10 images present the face in the frontal position. For baseline evaluation, face detectors [12] were firstly applied for pre-processing the thermal and visible images, respectively, and are based on a Faster R-CNN architecture [18] using a ResNet-50 neural network model [7]. The feature extraction is performed using a MobileNetV2 network model [22], which has previously been trained, inter alia, on several datasets. The experiments are conducted in accordance with the leave-one-out cross validation for splitting ratio of dataset with 80% for the training data and 20% for the test data, i.e. 4 thermal images were matched against all 4 visible face images. The training and test data each consists of two sets. The first set can be considered as the genuine class, and combines two vectors from the same person. The second set is the imposter set, and the two vectors are extracted from different people. Each combined feature vector is a concatenation of two feature vectors extracted from separate images. The decision function used for experiments is based on the Support Vector Machines (SVM) algorithm [17]. The obtained results for the five-fold cross-validation are averaged.

Iris on-the-move Both enrolment and verification images were captured for the iris-on-the-move scenario. The enrolment images were captured using a commercial IrisID device with the subjects standing still at a close distance from the sensor. From each subject 5 high quality enrolment images were recorded for each eye. For the iris on-the-move verification, an iris camera system developed by Raytrix was employed. The camera system is comprised of three identical 20MP sensors combined with lens optics featuring three different focal lengths ($75\,\mathrm{mm}$, $50\,\mathrm{mm}$ and $35\,\mathrm{mm}$) and Near Infra-Red filters. For illumination an array of 24 high power flashing LEDs emitting NIR light were positioned on top of the camera system. To simulate the on-the-move scenario the subjects were asked to walk at a normal pace along a straight trajectory towards the camera system starting at a distance of $1.8\,\mathrm{m}$. whereby dual-eye images were recorded at 10 frames per second. For the presented multimodal dataset, 5 of the resulting verification images were selected for each user. Iris templates were extracted from both enrolment and verification images. All enrolment templates were matched against all verification templates. For template extraction, as well as for template matching, software from Neurotechnology VeriEye SDK was employed. A separate eye detection algorithm from OpenCV [2] was applied.

Somatotype Somatotype defines a body type classification methodology and focusses on the measurement of the structural aspects of the human body. A person's somatotype can be determined from full-body images that can be captured at a distance on-the-move [3]. The somatotype dataset contains two static, and one on-the-move somatotype capture for each identity, giving three somatotype captures in total for each identity. A high-resolution camera, which was set up along the length of the recording area, was utilised. One of the static captures is recorded from the frontal view of the individual, while the second static capture is recorded from the side view. The on-the-move capture is recorded from the side of the individual. For the baseline evaluation, only the side captures were used. More specifically, static side captures were used as the gallery samples, while the on-the-move side captures were used as the probe samples. A Siamese deep learning network, optimized for extracting somatotype-related features [3], was implemented for evaluating the verification performance.

Smartphone sensors Person verification based on smartphone sensor for border control is relatively unexplored in the literature. To investigate the potential using smartphone sensors for verification, a smartphone sensor dataset has been collected, including accelerometer, gyroscope, magnetometer, barometer, and luminosity. The sensors were continuously read through each recording session using a developed Android app. Users were asked to walk continuously at their normal pace with the phone in a defined position (phone held in hand, making a call, phone in trousers' pocket, reading the screen, and switching between the other activities freely) to mimic the scenario of a traveller walking to the border control point after disembarking the plane. Each session is repeated three times and lasts 30s or 60s depending on the session parameters. For the baseline evaluation, the smartphone sensor data are firstly pre-processed with re-sampling and noise reduction. Considering the nature of the smartphone sensor data, which are multi-axis time-series data, a Long Short-Term Memory (LSTM) [8] based network has been applied for learning and extracting motion features on combined accelerometer and gyroscope data. The network was trained using the HMOG dataset [24] and evaluated on the unseen D4FLY dataset.

Iris mobile Iris was also captured using a commercial compact portable iris scanner IriShield MK212OU connected to an Android smartphone. Five images were captured for each eye per user. Considering this is the same biometric trait with iris on-the-move and it is the only one that is not suitable for on-the-move capture, it has been excluded from the evaluation in this work. However, the data is included in the D4FLY multimodal dataset.

4.2. Multimodal biometric fusion

Multimodal biometric fusion that combines multiple sources of biometric information has been widely investigated in the literature in the past to improve confidence and accuracy in a biometric recognition system over a single biometric recognition system [4, 13]. The fusion of evidence from various biometric modalities can be performed at sensor level, feature level, score level and decision level [20]. Score-level fusion has been more widely applied in all biometric fusion applications and is generally preferred because it offers the best trade-off in terms of the information content and the ease in fusion [15].

Score level fusion combines multiple matching scores using a fusion scheme to form a single score which is used to make the final decision. The selection of the fusion method can depend on various factors, including the application area, the modality types and their score distributions, and the amount of data available for learning, etc. [20].

In the D4FLY project, multiple biometric traits of the traveller are captured through the biometric corridor while the traveller approaches the exit. Each unimodal biometric system compares their captured biometric data on-the-move with the corresponding templates obtained at the enrolment stage. All the matching score outputs are then combined to produce a final single result before the traveller exits the corridor. Therefore, a fusion scheme that is fast and robust is important in this border control scenario. A score-level fusion process was applied for the fusion evaluation using the multiple biometric traits collected in the dataset.

Score Normalization In score level fusion, selection of a normalisation scheme is an important step for obtaining reliable performance [14] due to the match scores from individual modalities potentially being heterogeneous, *e.g.* the output score from a specific modality may be either a dissimilarity measure or a similarity measure; individual match scores may not be on the same numerical scale; or individual match scores may present different statistical distributions and hence induce thresholds at different scales. Therefore, it is essential to transform the individual scores into a common domain prior to combining them [10]. Normalisation methods have been previously investigated and compared in various works [19, 23, 25]. The Hyperbolic Tangent ($tanh$) normalisation has been proven to be robust to outliers and highly efficient in general compared to other methods (min-max, z-score, median/median absolute deviation (MAD) and double sigmoid function, etc.) [25]. The tanh calculation (Eq. 1) requires two pre-calculated statistics on a known dataset: S_μ^G is the mean average of all genuine match scores (TP (True Positive) + FN (False Negative)), and S_σ^A is the standard deviation of all scores for the known dataset. This provides a final score with positive match scores being centred around 0.5 with a standard deviation of 0.01.

$$S' = \frac{1}{2}\left[tanh\left(0.01 \cdot \frac{(S - S_\mu^G)}{S_\sigma^A}\right) + 1\right] \tag{1}$$

Score-level fusion The score-level fusion can be performed using different methods, such as fusion rule based approaches (*e.g.* sum rule, user weight, min or max rule) and classification based approaches (where a classifier is trained using a feature vector combining all the individual matching scores) [13, 25]. Extensive study has suggested that fusion rule based approaches perform better than classification approaches [21].

$$F = \frac{\sum_{n=0}^{N} S'_n}{N} \tag{2}$$

The sum-rule fusion is a simple yet efficient approach for fusing multiple score values. In this work, the sum-rule fusion is applied to combine the output from individual modalities following a tanh score normalisation to determine the final fused score F in Eq. 2, where, S' is the normalised match score calculated from n^{th} unimodal recognition output, and N is number of available unimodal biometric recognition systems (*i.e*. number of available scores).

5. Results and discussion

The results for both unimodal recognition and multimodal fusion are presented and discussed in this section. The metrics used for assessing the performance are Equal Error Rate (EER), FMR1000, and ZeroFMR (defined upon False non-match rate (FNMR)) (ISO/IEC 19795-1:2021). EER is estimated at when $FMR = FNMR$. FMR1000 is the lowest FNMR for $FMR \leq 0.1\%$; and the ZeroFMR is given by the lowest FNMR for $FMR = 0\%$.

5.1. Unimodal verification results

Evaluation methods for each unimodal biometric were briefly described in Section 4.1. Table 3 presents the performance of each individual unimodal biometric verification system. The best results are given by the thermal-to-visible recognition system followed by the iris on-the-move. The two real-time on-the-move face and iris-based recognition systems have shown both good quality data capture and verification performance. Compared with these traditional biometric types, somatotype and smartphone sensors are the two worst performing modalities, which is understandable as human movement/behaviour has more degrees of freedom hence more difficult to model. The poor results from smartphone sensor-based recognition system also suggests the large variations generated in the dataset, in particular, the decision to vary the position of the phone to better emulate a range of passenger behaviours, has introduced addition challenges for recognising the uniqueness amongst identities due to the output signals being highly sensitive to the sensor location.

Biometric Trait	EER	FMR1000	ZeroFMR
3D face	25.81	70.97	96.77
Thermal face	**13.49**	**31.94**	**31.94**
Iris on-the-move	18.42	40.00	48.86
Somatotype	36.74	91.93	91.93
Smartphone sensors	46.99	98.51	99.91

Table 3. Unimodal biometric benchmark verification results (in %)

5.2. Multimodal fusion results

As introduced in Section 4.2, tanh normalisation was applied to the match scores from each individual modality followed by a sum-rule score-level fusion. As mentioned in Section 4.1, some biometrics have fewer samples hence matching results, the multimodal fusion evaluation was performed using all the combinations from all the available unimodal results. The best set of results from each unimodal system was used for fusion, if n-fold validation was used for that modality. All users are represented in each unimodal result. The fusion was evaluated in two parts.

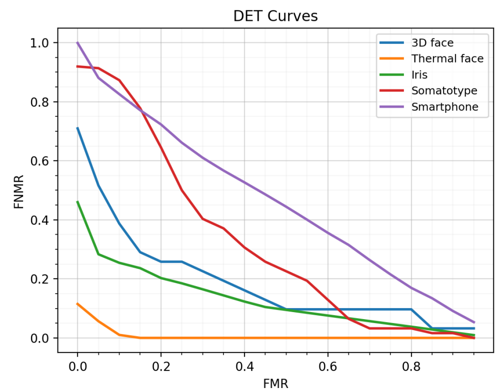

Figure 4. DET curves of each individual biometrics

Firstly, the multimodal fusion was evaluated based on different on-the-move biometric fusion scenarios including iris and thermal face, iris and 3D face, thermal face and 3D face, smartphone sensor and somatotype, and all biometrics. Figure 5 shows the Detection Error Tradeoff (DET) curves comparing the fusion results from the defined scenarios. The detailed performance is listed in Table 4.

As the dataset was captured in a similar environment setup to the PROTECT walk-through scenario [5], it is natural to compare the multimodal fusion performance, despite that different data are used for the two systems. D4FLY presents higher error rates in its fusion results than the PROTECT walk-through scenario, however PROTECT's underlying fused biometrics are primarily 2D (visible and near infra-red) face plus anthropometrics. Even the 3D face in PROTECT was implemented as multi-view 2D image processing of a light-field capture, whereas D4FLY performs recognition on the 3D face. Additionally, D4FLY system aims at obtaining a result within 4-5 seconds for all biometrics, whereas PROTECT's 3D face capture could not be achieved under the intended operational conditions, and is not suitable for deployment in a real border scenario.

In the D4FLY results presented in Table 4, the iris and thermal face combination achieved a comparable result to the best combinations in PROTECT and could be performed in a real border control setting. The performance in terms of processing time will be discussed below.

In the second part, the multimodal fusion was evaluated using the leave-best-n-out scheme. The performance of unimodal biometrics was sorted based on their EER rate and ZeroFMR/FMR1000, respectively, however, they are consistent in these evaluation results. Fusion was then calculated with excluding the n-best performing modalities, where $n = 0; ...; N-1$, with $N = 5$. Figure 6 illustrates the DET curves of the results showing the impact from fusion.

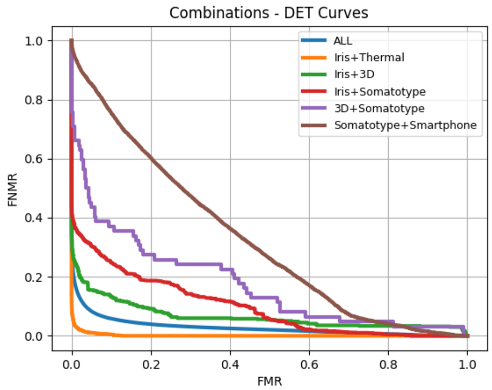

Figure 5. Multimodal fusion results based on different fusion scenarios

Scenario	EER	FMR1000	ZeroFMR
ALL	7.01	27.39	48.57
Iris+Thermal face	**2.08**	**8.54**	**16.76**
Iris+3D face	11.90	32.90	40.00
Iris+Somatotype	18.24	41.50	46.82
3D face+Somatotype	25.81	75.81	75.81
Somatotype+Smartphone	38.19	98.29	99.89

Table 4. Fusion results based on different scenarios (in %)

This work has a particular focus for border control applications, thus, processing time is also a crucial metric in the D4FLY scenario as described in Section 1 to ensure a high throughput from large volumes of travellers. Table 6 summarises the performance for each unimodal biometric recognition system and the multimodal fusion system in terms of processing time for both data capture and verification.

It takes about 5 seconds for an average person to walk through the corridor area, thus a fusion decision combining all biometric results must be produced within this short period. Four individual biometric traits are captured continuously on-the-move for 2-3 seconds as soon as the person enters the corridor. Smartphone sensor capture is initiated when the traveller disembarks from plane and completes with a verification result prior to the traveller entering the corridor. As shown in the table, individual biometric verification for the four corridor biometric traits can be completed within 4 seconds or even faster. The multimodal fusion process takes less than 50ms including networking and communication processes. This demonstrates that a system using these contactless biometric traits can achieve real-time on-the-move traveller verification based on the D4FLY biometric corridor design.

6. Conclusions

A novel multimodal biometric dataset has been introduced in this work that has been collected specifically in on-the-move scenarios for border control applications. To the authors' best knowledge, this is the first multimodal biometric dataset where all collected biometric traits have been captured in on-the-move scenarios. Innovative sensors were used in collecting the data and the emerging biometric traits are believed to bring novelty and value to research in both unimodal and multimodal biometric recognition, especially on-the-move biometric recognition.

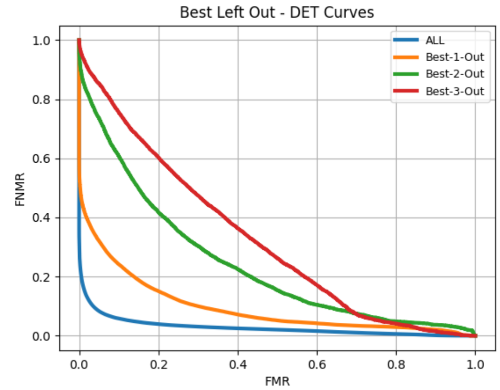

Figure 6. Multimodal fusion results based on leave-best-n-out fusion scheme

Best-n-out	EER	FMR1000	ZeroFMR
ALL	7.01	27.39	48.57
Best-1-Out	17.04	53.28	63.65
Best-2-Out	30.21	94.27	98.12
Best-3-Out	38.19	98.29	99.89

Table 5. Leave-n-best-out fusion results (in %)

Biometric	Data capture	Verification
3D face	Continuous data acquisition for 2s (10 captures)	$\sim$4s
Thermal face	Continuous data acquisition for 3s	0.7$\sim$3s
Iris on-the-move	Continuous data acquisition for 3s ($\sim$100 captures)	<1s
Somatotype	Continuous data acquisition for 2s (3 captures)	$\sim$0.9s
Smartphone sensors	Continuous data acquisition for 5-10s	5$\sim$10s
Fusion	Average person takes approx. 5s to walk through the biometric corridor	<0.05s

Table 6. Time performance for real-time D4FLY border control process

Baseline evaluations on individual modalities were performed and presented in this work. Some of the biometrics traits have shown promising results especially with an objective towards real-time operations, such as thermal-to-visible face recognition and iris on-the-move. With novel sensing technology for data capture and 3D face recognition algorithm, 3D face recognition has achieved very promising results within the desired time constraints for the D4FLY border control scenario. To the authors' knowledge, this is the first time that a 3D face verification method is assessed on data captured in a real life scenario. Existing methods were evaluated on publicly available datasets containing post-processed data, captured in a controlled environment (e.g. standardised facial expressions, movement, lighting, pose, etc.). It is expected that there will be a significant margin for further improvement in this previously unexplored scope.

Somatotype and smartphone sensors have shown poor unimodal recognition results, however, as in general human movement/behaviour related characteristics are more challenging and sensitive to intra-subject variations. Especially in the case of the smartphone sensor dataset, where variations in phone positions throughout each walk may have added unexpected challenges to an already demanding task and in the future work these considerations should be taken into account.

Multimodal biometric fusion is the key to obtaining more robust verification results whilst overcoming the limitations of unimodal recognition systems. From the evaluation, it can be observed that the fusion results can be negatively affected by poor performing biometrics, however, the main objective was to explore potential combinations of the selected on-the-move biometric modalities. Furthermore, it is expected that some biometric modalities that have been prioritised for real-time performance have a significant potential for improving their individual verification accuracy and hence the overall fusion score. This work provides insight into how these biometric traits can perform in real-life applications such as border control and to promote research using this dataset to improve both unimodal and multimodal fusion results.

7. Acknowledgments

This research was funded by EU H2020 Research and Innovation Programme under grant agreement no. 833704 (D4FLY).

References

[1] EU H2020 project D4FLY. https://d4fly.eu/. 3

[2] Open Source Computer Vision Library (OpenCV). 4

[3] A. Danelakis and T. Theoharis. Image-based somatotype as a biometric trait for non-collaborative person recognition at a distance and on-the-move. *Sensors*, 20(12), 2020. 4

[4] L. M. Dinca and G. P. Hancke. The fall of one, the rise of many: A survey on multi-biometric fusion methods. *IEEE Access*, 5:6247–6289, 2017. 5

[5] C. Galdi et al. PROTECT: Pervasive and useR fOcused biomeTrics bordEr projeCT – a case study. *IET Biometrics*, 9:297–308(11), November 2020. 2, 6

[6] J. A. Hasselgren, J. J. Howard, Y. B. Sirotin, J. L. Tipton, and A. R. Vemury. A Scenario Evaluation of High-Throughput Face Biometric Systems: Select Results from the 2019 Department of Homeland Security Biometric Technology Rally. In *The DHS ST Technical Paper*, 2020. 3

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 4

[8] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997. 5

[9] J. J. Howard, A. J. Blanchard, Y. B. Sirotin, J. A. Hasselgren, and A. R. Vemury. An investigation of high-throughput biometric systems: Results of the 2018 department of homeland security biometric technology rally. In *IEEE BTAS*, pages 1–7, 2018. 3

[10] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005. 5

[11] A. K. Jain and A. Ross. Multibiometric systems. *Commun. ACM*, 47(1):34–40, 2004. 1

[12] M. Kowalski and A. Grudzień. Detection of human faces in thermal infrared images. *Metrology and Measurement Systems*, 01 2021. 4

[13] A. Lumini and L. Nanni. Overview of the combination of biometric matchers. *Information Fusion*, 33:71–85, 2017. 5

[14] Mamta and M. Hanmandlu. Multimodal biometric system built on the new entropy function for feature extraction and the refined scores as a classifier. *Expert Systems with Applications*, 42(7):3702–3723, 2015. 5

[15] K. Nandakumar, Y. Chen, S. C. Dass, and A. Jain. Likelihood ratio-based biometric score fusion. *IEEE Transactions on PAMI*, 30(2):342–347, 2008. 5

[16] O. Ocegueda, G. Passalis, T. Theoharis, S. K. Shah, and I. A. Kakadiaris. UR3D-C: linear dimensionality reduction for efficient 3d face recognition. In *IJCB*, pages 1–6, 2011. 4

[17] P. Phillips. Support vector machines applied to face recognition. *NISTIR*, page 803–809, 11 1998. 4

[18] S. Ren, K. He, R. B. Girshick, and J. Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE TPAMI*, 39(6):1137–1149, 2017. 4

[19] S. Ribaric and I. Fratric. Experimental evaluation of matching-score normalization techniques on different multimodal biometric systems. In *IEEE Mediterranean Electrotechnical Conference*, pages 498–501, 2006. 5

[20] A. Ross. An introduction to multibiometrics. In *European Signal Processing Conference*, pages 20–24, 2007. 1, 5

[21] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, 2003. 5

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *CVPR*, pages 4510–4520, 2018. 4

[23] Y. N. Singh and P. Gupta. Quantitative evaluation of normalization techniques of matching scores in multimodal biometric systems. In *Advances in Biometrics*, pages 574–583. Springer Berlin Heidelberg, 2007. 5

[24] Z. Sitová, J. Šeděnka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani. HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users. *IEEE TIFS*, 11(5):877–892, 2016. 5

[25] K. Vishi and V. Mavroeidis. An evaluation of score level fusion approaches for fingerprint and finger-vein biometrics. *CoRR*, abs/1805.10666, 2018. 5