PROTECT Multimodal DB: fusion evaluation on a novel multimodal biometrics dataset envisaging Border Control [1]

Ana F. Sequeira[2], Lulu Chen[2], James Ferryman[2], Chiara Galdi[3], Valeria Chiesa[3], Jean-Luc Dugelay[3], Patryk Maik[4], Piotr Gmitrowicz[4], Lukasz Szklarski[4], Bernhard Prommegger[5], Christof Kauba[5], Simon Kirchgasser[5], Andreas Uhl[5], Artur Grudzien[6], Marcin Kowalski[6]

Abstract: This work presents a novel multimodal database comprising 3D face, 2D face, thermal face, visible iris, finger and hand veins, voice and anthropometrics. This dataset will constitute a valuable resource to the field with its number and variety of biometric traits. Acquired in the context of the EU PROTECT project, the dataset allows several combinations of biometric traits and envisages applications such as border control. Based upon the results of the unimodal data, a fusion scheme was applied to ascertain the recognition potential of combining these biometric traits in a multimodal approach. Due to the variability on the discriminative power of the traits, a leave the n-best out fusion technique was applied to obtain different recognition results.

Keywords: Biometric recognition, multimodal, fusion, database, border control.

1 Introduction

Biometric recognition systems that rely on a single source of biometric information to perform recognition are called unimodal. Biometric systems that include multiple sources of information for establishing an identity are known as multimodal. The analysis of advantages and disadvantages of individual traits leads to the conclusion that there is no "gold-standard" biometric trait and that some biometric traits seem to present advantages that counterbalance other trait's disadvantages.

In the context of border control, the EU PROTECT project [H2] aims at building an advanced biometric-based person identification system that works robustly across a range of border crossing types and that has strong user-centric features. An effort is made to optimize currently deployed biometric modalities as well to apply emerging biometrics in a multimodal approach. To assess the effectiveness of any developed technique in the described context, there is a need for multimodal data comprising the specific biometric traits to be investigated. This paper contributions are: (i) the novel *PROTECT Multimodal DB* database that will represent a valuable resource for research and will innovate by providing a new combination of biometric traits which cannot be found in any of the existing databases; (ii) the evaluation of the data in a multimodal fusion approach.

[1] Supported by PROTECT EU project (grant agreement H2020-700259). All authors contributed equally.
[2] University of Reading, Reading, UK, Email: {a.f.p.sequeira, l.chen, j.m.ferryman}@reading.ac.uk
[3] EURECOM, France, Email: {galdi, chiesa, dugelay}@eurecom.fr
[4] ITTI, Poznan, Poland Email: {patryk.maik,piotr.gmitrowicz,lukasz.szklarski}@itti.com.pl
[5] PLUS, University of Salzburg, Salzburg, Austria, Email: {bprommeg,ckauba, skirch, uhl}@cs.sbg.ac.at
[6] WAT, Military University of Technology, Warsaw, Poland, Email:{artur.grudzien,marcin.kowalski}@wat.edu.pl

In the remainder of the paper, in Section 2 the *PROTECT Multimodal DB* database is presented including an overview of multimodal recognition and databases, the acquisition setup and the characterisation of the subjects. Section 3 comprises the experimental setup used for the unimodal and multimodal evaluations. In Section 4 the metrics used are described along with the unimodal and multimodal results and their discussion. Finally, in Section 5 the work is concluded with the final remarks.

2 The *PROTECT Multimodal DB* database

2.1 Multimodal Biometric Recognition and Databases

Extensive studies have been performed on several biological traits, regarding their capacity to be used for unimodal biometric recognition. When analysing a biometric trait considering the four critical factors: Universality, Uniqueness, Collectability and Permanence, a couple of general conclusions can be withdrawn: (1) there is no "gold-standard" biometric trait, i.e. the choice of the best biometric trait will always be conditioned by the means at our disposal and the specific application of the recognition process; (2) some biometric traits seem to present advantages that counterbalance other trait's disadvantages. Marked advantages might be found by exploring the synergistic effect of multiple statistically independent biometric traits. Using biometric evidence obtained from multiple sources of information will result in an improved capability of tackling some of the more relevant known problems of unimodal systems such as dealing with noisy data; intra-class variations; inter-class similarities; non-universality; and spoof attacks.

In the context of border control, the three biometric traits considered by the standards developed within the International Civil Aviation Organization (ICAO) are face, fingerprints and iris. Face is the primary biometric trait chosen for most states nevertheless any state can provide additional data input to the identity verification processes by including multiple biometrics in their travel documents, i.e., a combination of face and/or fingerprint and/or iris. For example, the Homeland security program in the USA collects face images and fingerprints from all visitors.

Like in other fields of research, the existence of suitable databases is crucial for the development and evaluation of methodologies. Multimodal biometric databases were not always available for several reasons. To overcome this difficulty, some works use chimeric or virtual-subject databases with several drawbacks like the fact that these databases contain users that do not exist in the real world which the system will never encounter. This practice also disregards one of the goals of fusion systems which is to describe the relationship among different modalities.

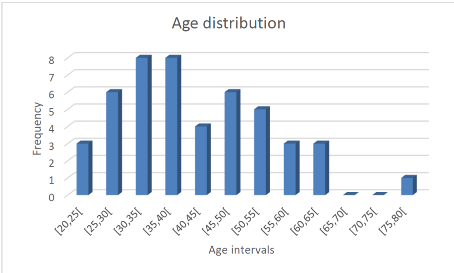
When considering the construction of a multimodal database it is necessary to take in account the complementarity of the chosen traits. The *PROTECT Multimodal DB* database presents a novel combination of 9 biometric traits: 2D face; 3D face (RGB and Depth Field); thermal face; iris; voice; finger-veins; hand-veins; and anthropometrics, some samples are depicted in Figure 1. To the best of our knowledge, this is the largest number of traits, when compared to previous databases with 6 (Biosecure-BMDB; MBioID; JMBDC; and BIOMET); 7 (MyIdea); and 8 (BiosecurID) different traits.

Fig. 1: Samples from the *PROTECT Multimodal DB* database: 2D face, anthropometrics, 3D face, thermal face, iris, hand veins, finger veins.

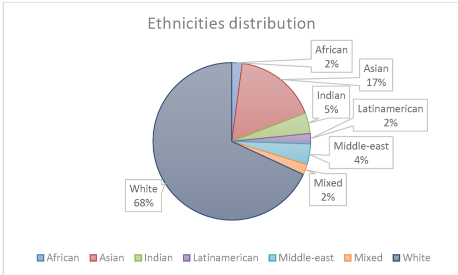
Acquired in a context focused on border control, the *PROTECT Multimodal DB* database aims at being representative of the universe of travellers that cross the borders thus including a wide range of variety in age, gender, ethnicity and skin/eye colour types. The data was collected in one site and in a single session. A subset of the *PROTECT Multimodal DB* was released freely to the academia and industry upon request and the complete dataset will be released in a near future (for details please see projectprotect.eu).

Acquisition setup: One of the main goals of the construction of the *PROTECT Multimodal DB* dataset was to proportionate data acquired in conditions aligned with the PROTECT EU project concept which is a less constrained and intrusive biometric capture for the passenger on the move through any kind of border. Specifically, envisaging a biometric corridor use case, a 10 metre long area was created to mimic a walk-through border crossing. This biometric capture area was used to capture simultaneously 2D face and gait/anthropometrics. The subjects were asked to walk naturally while looking forward. In other locations of the same room the subjects provided their other types of biometric data. The sequence of collection of the biometric data was not fixed due to time operational constraints.

Characterization of the subjects: Biometric data was recorded from a total number of 47 subjects. The distribution male/female is 57%/43%. In Figure 2(a) the age distribution is depicted in 5-year intervals. The variety of ethnicities (therefore of eye/skin types) can be visualised in Figure 2(b).



(a) 5-year intervals distribution



(b) Ethnicities distribution

Fig. 2: Age and ethnicities distributions of subjects in the corpus.

3 Experimental Setup: Unimodal and Multimodal Evaluation

In this section are presented: the baseline evaluation for each unimodal modalities[1] (a brief description and methods used) and the evaluation setup used for the multimodal fusion.

3.1 Unimodal Evaluation Methods

2D face: The 2D face dataset contains videos captured from three cameras, which were set up along the length of the recording area, roughly the same distance apart and with variations in height, yaw and pitch. Frames with better image quality were manually selected from all videos: full-frontal face, no glasses, person looking directly to the camera if possible. Due to limited number of such frames, only one image per user was selected as gallery sample. For the probe samples, an automatic detection tool was used to select the frames where the face could be detected as the majority of the frames contain faces in low quality. A commercial software by Visage Technologies [Vi18] with state-of-the-art face tracking and analysis was used for evaluating the recognition performance (which underlining algorithms cannot be disclosed).

3D face: The 3D face raw data is processed with the Lytro Power software which allows the RGB picture and the corresponding depth map to be automatically extracted. An alignment algorithm is applied to all RGB data. Depth maps are aligned with the same transformation applied to the correspondent texture image. Because of the complexity of the database due to occlusions or challenging position, only 88% of the faces are detected. To perform the face recognition, the OpenFace [ALS16] features are chosen which are based on the deep neural network described in [SKP15] (implemented on Python and Torch).

Thermal face: Thermal facial recognition is based on the analysis of individual heat patterns emitted by the human face on the form of an image presenting a map of apparent values of temperature on its surface. The thermal image contains sufficient amount of information for distinguishing individuals. Emission dominated, passive imaging does not

[1]The voice data was not processed because throughout the project this trait was discarded from research.

require additional illuminator and is independent from illumination non-uniformities. The thermal face recognition process is composed of various stages comprising alignment, face detection, feature extraction and comparison. The face detection was performed using the Viola-Jones algorithm [VJ04]. Several feature extraction methods were investigated and the best results were obtained with Local Binary Patterns [OPH96] which perform well with thermal facial images, combined with various distance metrics.

Iris: The iris images collected present a good iris pattern quality for light pigmented irises but a lower quality for dark irises despite the additional lighting used which in turn caused specular reflections. The method for segmentation [HK15] was the best ranked in the MICHE I competition. For the feature extraction and comparison, it was used a novel approach designed for iris recognition on smartphones submitted to MICHE II. The FIRE method [GD16] was chosen for its good performance with mobile low-quality images. Among the three possible comparisons: left, right or left-left and right-right eye patterns, the best results were obtained for the right eye pattern comparisons.

Finger-vein: The finger-vein images are collected from both right and left index and middle fingers. The ROI extraction is done manually and then the images are pre-processed in order to improve the visibility using High Frequency Emphasis Filtering, Circular Gabor Filter and simple CLAHE (Local Histogram Equalisation). For the performance evaluation some well-established finger-vein recognition schemes were used. The Maximum Curvature (MC) [MNM07] combined with the correlation-based comparison approach proposed by Miura et al. [MNM07] achieved the best results. For more details see [KRU14].

Hand-vein: Dorsal hand-vein images of both hands have been acquired under different illumination conditions: two reflected light illuminators (850nm and 950nm) and one trans-illumination light source (850nm). The same processing tool-chain as for finger-vein is used to conduct the hand-vein performance evaluation. In addition, a rotation correction has been adopted in the comparison step. The best results were obtained with the MC for the 950nm reflected light acquisition scheme.

Anthropometrics: The collected anthropometrics data include both physiological and behavioural features of an identification subject. Behavioural features, which include parameters such as average step length, are calculated from time-based signals extracted by a network of Kinect sensors. Physiological features include parameters such as height, arm/leg length. The method used for the recognition process applies an artificial neural network to estimate the similarity between two feature vectors. The network has been based on siamese architecture [CHL05]. a representative subset of the acquired data has been used for the network training and validation purposes.

3.2 Multimodal Fusion

The multimodal evaluation has been carried out using the MATLAB BOSARIS Toolkit[2] which is a collection of functions and classes that can be used to calibrate, fuse and plot

[2]https://sites.google.com/site/bosaristoolkit/

scores for biometric recognition. For each biometric trait, two distance matrices, namely the DEV matrix and EVAL matrix have been computed and used as follow.

DEV matrix: facilitates the tuning of the weights that will then be used on the EVAL matrix. It contains the scores originating from the comparison of 47 enrolled samples (one for each subject) against 47 development samples.

EVAL matrix: is made up of scores originating from the comparison of the 47 enrolled samples against 47 evaluation samples (different from the development ones). In a "closed set" setup as all 47 subjects are both in the Gallery (enrolled samples) and in the Probe (testing samples) sets. However, the weights for fusion are computed on the DEV matrix and used on unseen test samples for final performance evaluation.

This protocol was adopted for all traits except for finger/hand-veins, where left hand was used to build the DEV matrix and the right hand for the EVAL matrix. Whenever more than one baseline was tested, the best performing configuration was chosen for fusion.

The BOSARIS toolkit capability to integrate the samples' quality scores for multimodal fusion was used. Thus, prior to fusion, all score matrices have been normalized using MinMax technique so that all scores range in $[0,1]$ interval.

4 Results and discussion

Here are presented the results obtained with the methods described in Sections 3.1 and 3.2.

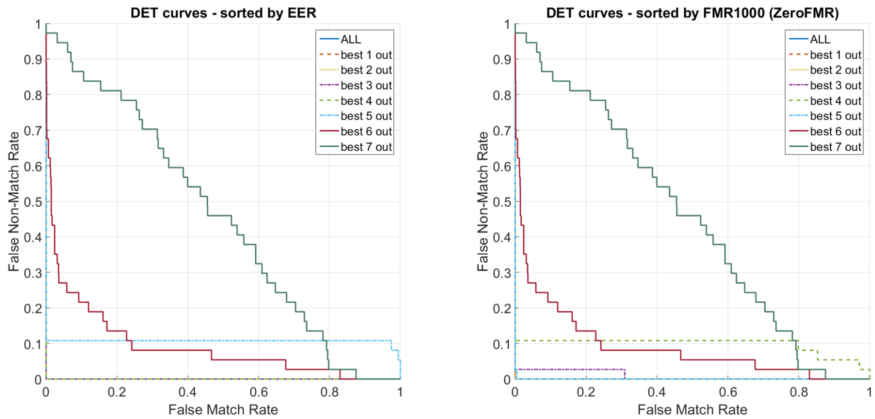
The **metrics** used are *Equal Error Rate* (EER); *FMR1000*; and *ZeroFMR*, defined upon *False non-match rate* (FNMR) and *False match rate* (FMR) as standardisation documents ISO/IEC 19795-1:2006. EER is obtained when $FMR = FNMR$; FMR1000 is the lowest FNMR for $FMR \leq 0.1\%$; and the ZeroFMR is given by the lowest FNMR for $FMR = 0\%$.

The results obtained for the **unimodal** recognition evaluation are depicted in Table 1. The first three columns show the results obtained by the benchmarking methods with all the available data. The last three columns show the results obtained by the BOSARIS method with the DEV and EVAL matrices that were the input for the fusion method. The best recognition results were obtained for 3D Face RGB and, on the opposite side, 3D Face DF, followed by thermal face and iris lead to the poorest results.

The results for **multimodal** fusion have been computed according to the leave-best-n-out scheme. After sorting the EER, ZeroFMR and FMR1000 values, the n-best performing biometric traits were excluded from fusion, for $n = 0,1,...,N-1$ with $N = 8$. The DET curves depicted in Figure 3 show how much the fusion results are impacted by the highest performing traits with a notorious decay in performance as the number of best performing traits excluded increases.

Tab. 1: Unimodal recognition results (results in %).

Biometric Trait	Benchmark evaluation (all data)			DEV and EVAL data		
	EER	FMR1000	ZeroFMR	EER	FMR1000	ZeroFMR
2D Face	9.12	28.10	41.09	2.69	1.28	10.81
3D Face RGB	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
3D Face Depth Field	39.37	100	100	44.27	82.50	97.30
Thermal Face	10.88	73.91	72.13	5.08	0.00	5.41
Iris VIS Mobile	15.32	45.96	65.25	16.17	4.86	70.27
Finger Veins	9.75	11.83	56.80	5.13	9.73	5.41
Hand Veins	0.12	0.25	0.25	9.76	4.77	10.81
Anthropometrics	0.88	4.44	18.66	0.47	0.00	24.32



(a) EER

(b) FMR1000/ZeroFMR

Fig. 3: Multimodal recognition results (leave-best-n-out fusion scheme).

5 Conclusions and Future Work

The current work presented a new multimodal database constructed under the scope of border control applications. Using this data, recognition results are presented for each individual trait using state-of-art methods and then a fusion method allowed to explore multimodal recognition. The results obtained showed that some biometric traits by itself, such as 3D Face, lead to very promising recognition results. On the contrary, others such as thermal face or visible iris (acquired with a mobile device) lead to poor recognition results. It was observed that when fusing good performing traits with others the recognition capability drops considerably. Nevertheless, it should be kept in mind that there are several reasons that justify the use of multimodal recognition, such as to increase the robustness to spoofing attacks or to overcome problems with lack of universality or accessibility of some particular biometric traits. The future work comprises analysing the data in subsets that envisage specific use cases, as in a real border control application it is not expected that 8 traits are fused simultaneously, there will be a selection of the more suitable traits for each situation.

References

[ALS16] Amos, Brandon; Ludwiczuk, Bartosz; Satyanarayanan, Mahadev: OpenFace: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.

[CHL05] Chopra; Hadsell; LeeCun: Learning a Similarity Metric Discriminatively, with Application to Face Verification. IEEE Comp. Vision and Pattern Recogn., 1:539 – 546, 2005.

[GD16] Galdi, C.; Dugelay, J. L.: Fusing iris colour and texture information for fast iris recognition on mobile devices. In: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 160–164, Dec 2016.

[H2] H2020, EC: , PROTECT Project. `http://projectprotect.eu/`, Last accessed on 2018-04-12.

[HK15] Haindl, Michal; Krupicka, Mikulas: Unsupervised detection of non-iris occlusions. Pattern Recognition Letters, 57:60 – 65, 2015. Mobile Iris CHallenge Evaluation Part I.

[KRU14] Kauba, Christof; Reissig, Jakob; Uhl, Andreas: Pre-processing cascades and fusion in finger vein recognition. In: Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG'14). Darmstadt, Germany, September 2014.

[MNM07] Miura, Naoto; Nagasaka, Akio; Miyatake, Takafumi: Extraction of finger-vein patterns using maximum curvature points in image profiles. IEICE transactions on information and systems, 90(8):1185–1194, 2007.

[OPH96] Ojala, T.; Pietikainen, M.; Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29(1):51 – 59, 1996.

[SKP15] Schroff, Florian; Kalenichenko, Dmitry; Philbin, James: FaceNet: A unified embedding for face recognition and clustering. pp. 815–823, 06 2015.

[Vi18] Visage Technologies: , Visage—SDK, 2018. `http://visagetechnologies.com/`, Last accessed on 2018-04-30.

[VJ04] Viola, Paul; Jones, Michael: Robust real-time face detection. International Journal of Computer Vision, 57:137–154, 2004.