Face anti-spoofing systems optimal threshold selection criteria

Ostap Stets[1,*,†], Ihor Konovalenko[1,†], Tomasz Gancarczyk[2,†] and Artur Mykytyshyn[1,†]

[1] *Ternopil Ivan Puluj National Technical University, Ruska str., 56, Ternopil, 46001, Ukraine*
[2] *University of Bielsko-Biala, Willowa St. 2, Bielsko-Biala, 43-300, Poland*

Abstract

This article is devoted to the problem of criteria definition for optimal threshold selection in face anti-spoofing systems based on common conventional metrics in the area. Analysis of previous studies has shown that live applications of presentation attack detection methods most often rely on common methods of threshold selection while tending to ignore domain and problem-specific requirements. Therefore, the main purpose of this research is to determine the criteria for optimal threshold selection in production-applied biometric authentication systems.

To address these limitations, the paper proposes an approach for automated threshold selection that incorporates an "Environmental Adjustment" factor. This factor takes into account the specific context of the PAD system's deployment, including security needs and user experience considerations.

Keywords

Face anti-spoofing, presentation attack, person identification

1. Introduction

Biometric identification systems have become ubiquitous in today's technological landscape, introducing both convenience and problems to solve. With the rise of automated or semi-automated biometric authentication processes, such as face recognition, the possibility of attacks on this particular aspect increases too. One of the significant attack types is the presentation attack (PA), which has become more prevalent due to the ease of execution. Face anti-spoofing, a critical component of biometric security systems, plays a pivotal role in safeguarding against such fraudulent activities. Its applications span across various domains where accurate facial recognition is paramount for authentication and access control. Such solutions are widely used across different

CITI'2024: 2nd International Workshop on Computer Information Technologies in Industry 4.0, June 12–14, 2024, Ternopil, Ukraine
* Corresponding author.
† These authors contributed equally.
ostap.stets@gmail.com (O.Stets); aicxxan@gmail.com (I.Konovalenko); tgan@ath.bielsko.pl (T.Gancarczyk); mykytyshyn21@gmail.com (A. Mykytyshyn)
0009-0007-9147-4728 (O.Stets); 0000-0002-2529-9980 (I.Konovalenko); 0000-0002-9709-0860 (T.Gancarczyk); 0009-0001-5999-5490 (A. Mykytyshyn)

domains. Among those some security-critical areas relying on biometric authentication are facing heightened risks, due to a wide set of possible attacks, e.g. videos, printed pictures, masks, and especially from sophisticated attacks like morphing, where fake identities can be generated by blending images of genuine and fraudulent subjects. These attacks pose significant threats requiring robust countermeasures to ensure secure and reliable identification processes.

In less security-critical environments, such as social media platforms or online shopping websites, the problem of presentation attack detection (PAD) still holds importance but is approached differently. Here, the focus is more on general user experience and convenience rather than stringent security measures. This imposes a problem of balancing security measures with user friction at least until the moment when used PAD solution performs perfectly on relevant datasets.

2. Problem formulation

While no face anti-spoofing system is error-prone, researchers and developers are forced to select the optimal error threshold. Setting thresholds too high for biometric authentication may result in increased user frustration and abandonment of the authentication process where it is not critical to him. On the other hand, setting thresholds too low may compromise security by allowing unauthorized access which is unacceptable in certain cases.

Another consideration is the trade-off between security and computational resources. Higher thresholds often require more computational power for accurate authentication, potentially increasing processing times and costs. Balancing the need for security with resource efficiency is crucial in these environments to provide a seamless user experience without compromising on security standards.

Whenever a person passes a face recognition-enabled biometrical identification system, their biometrical data is passed to the PAD subsystem. This data is most commonly analyzed in a pre-trained convolutional neural network (CNN) to distinguish real identity from attackers. The result of the PAD subsystem is a confidence score on whether this data was genuine or fraudulent. The final decision is made by comparing this score with a classification threshold. This value depends on special metrics described by ISO/IEC 30107-3:2023 on training datasets [1]. Two central parameters to ensure that test results are accurate are:

- Attack presentation classification error rate (APCER). It measures the error rate in classifying attack presentations as genuine.
- Bona fide presentation classification error rate (BPCER). It measures the error rate in classifying genuine presentations as spoofed.

This metric's purpose is to assess the PAD subsystem's ability to identify bona fide presentation attacks, its instruments, attack frequency, and error rate [1]. They cover such factors as presentation attack instruments and species (PAIS), artifacts and present non-

conformant characteristics, and description of output information provided by the PAD subsystem.

The APCER for a given PAIS is calculated using the formula:

$$APCER_{PAIS} = 1 - \left(\frac{1}{N_{PAIS}}\right) \sum_{i=1}^{N_{PAIS}} Res_i \quad (1)$$

where N_{PAIS} is the number of attack presentations for given PAI species; Res_i takes value 1 if the corresponding presentation is classified as an attack presentation and value 0 if classified as a bona fide presentation [1, 2].

As mentioned in the papers [1, 2], performance metrics for the set of bona fide presentations captured with the evaluation target shall be calculated and reported as BPCER using the formula:

$$BPCER = \frac{\sum_{i=1}^{N_{BF}} Res_i}{N_{BF}} \quad (2)$$

where N_{BF} is the number of bona fide presentations; Res_i takes value 1 if the corresponding presentation is classified as an attack presentation and value 0 if classified as a bona fide presentation.

The overall accuracy of the PAD subsystem is measured by using the Average Classification Error Rate (ACER) defined as [1]:

$$ACER = \frac{APCER + BPCER}{2} \quad (3)$$

As with all biometric identification systems, both error rates, APCER and BPCER, can't be minimized at the same time, as a decrease of one means an increase of another because it is impossible to completely separate responses of bona fide presentations as presentation attacks.

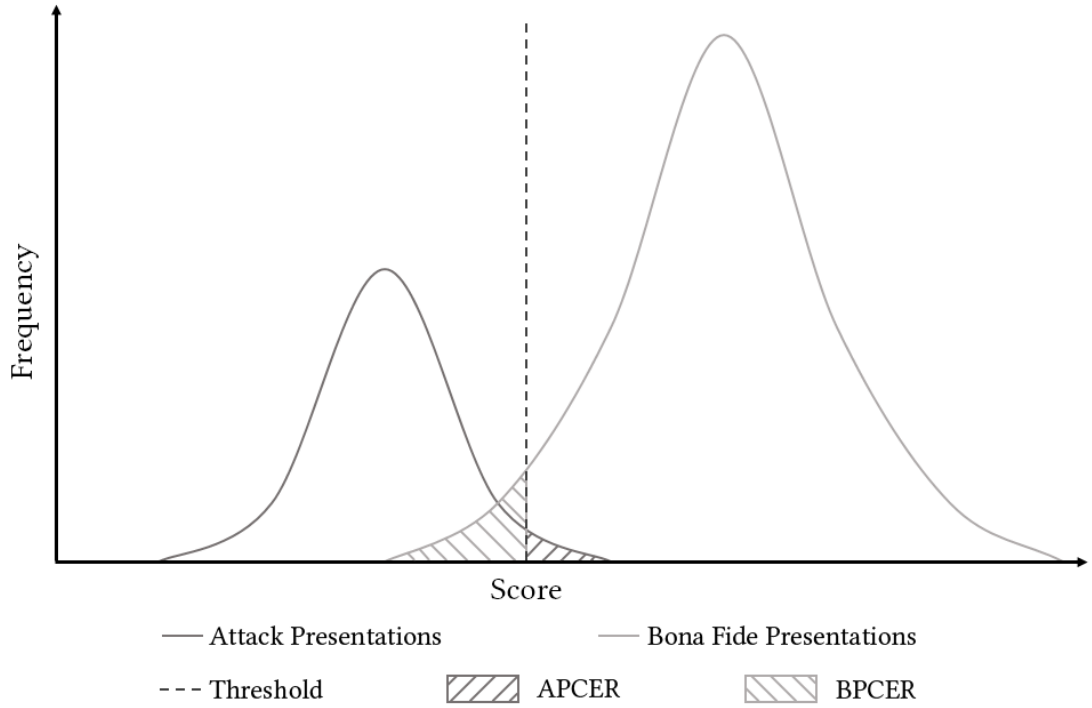

Figure 1: Example histogram of classification for bona fide and attack presentations. Shaded areas correspond to classification errors

As the outcome of the PAD system purpose, attack presentations tend to receive lower scores while bona fide presentations receive higher scores. However, because these scores overlap in most cases, a specific threshold should be selected. This paper is devoted to the research of the selection method of this threshold value in different environments where security importance can vary. The purpose of the analysis is to determine criteria, which would allow easier balancing between PAD safety and general user experience and convenience.

3. Comparative analysis of known solutions and suggested improvements

3.1. ABC4EU

In paper [2] followed by research [3] authors describe the pilot of a new Automatic Border Control (ABC) system which was developed in the ABC4EU European project and conforms to the laws established in the Schengen zone. These new ABCs have specific characteristics, such as a structural configuration divided into two devices: self-enrolment kiosk and biometric gate, one for enrolment and the other for verification, which entails two capture stages and two weaknesses where it is possible to attack the system [1, 2]. Researchers describe three different presentation attack types in their experimental setup:

- "Enrolment PA, when a presentation attack occurs at the self-enrolment stage. For example, an attacker provides the system with documentation that belongs to someone else and therefore tries to impersonate the true holder of the documents" [1].
- "Verification PA, when a presentation attack occurs at the verification stage. An attacker tries to impersonate a traveler who has previously enrolled in the system. For example, a correctly registered traveler loses or steals his/her documents between the self-stage and the verification stage. Then an attacker uses those documents to try to pass the verification" [1].
- "Enrolment and Verification PA. In this case, an impersonation has occurred at the enrolment and the attacker continues impersonating the true traveler at the verification stage (double attack). For example, an attacker presents travel documentation that belongs to someone else and gets successfully enrolled. After that, in the verification stage, the attacker continues to impersonate the true holder of the documents to cross the e-gate" [1].

As mentioned by the authors, security is a top priority in ABC systems while convenience and user experience are secondary, so they decided to set a threshold value that returns a low APCER value even if it increases the BPCER [1]. In this case, it is not critical as ABC systems are controlled by an agent, who can verify and correct bona fide presentations which were considered as an attack. They came up with an experimental setting of threshold values which led to a threshold of 80 at self-enrolment and a threshold of 95 at the biometric gate. Results of these experiments are displayed in Table 1 and Table 2.

Table 1
APCER, BPCER, and ACER values for different thresholds at self-enrolment [1]

Threshold	40	70	80	90	95
APCER	0.7609	0.3261	0.1739	0.1087	0.0217
BPCER	0.0	0.0	0.0667	0.7333	1.0
ACER	0.3804	0.1630	0.1203	0.421	0.5217

Even considering ABC4EU PAD as a security-critical subsystem, a BPCER value of 0.7333 is too high for the automatic system to be effective. This could indicate a discrepancy between training datasets and real-life bona fide presentations.

Table 2
APCER, BPCER, and ACER values for different thresholds at the biometric gate [1]

Threshold	40	70	80	90	95
APCER	0.8276	0.6552	0.5862	0.4483	0.206
BPCER	0.0	0.0	0.0	0.0	0.1429
ACER	0.4138	0.3276	0.2931	0.2241	0.1749

As displayed, BPCER value at the biometric gate is much better, however, the APCER of 0.206 at a threshold value of 95 is still too high to consider the system effective. Considering article [3] we could assume that these values were improved since the first project piloting, however latest data on PAD subsystem efficiency is not publicly available. However optimal threshold selection remains an issue in ABC4EU because manual setting during training with the dataset is imperfect for the following reasons:

1. The pre-trained model with a static dataset is limited from updates in bona fide presentation changes happening due to passenger flow shuffling (because of variable reasons, e.g. climate changes [4], economic reasons, infrastructural changes or conflicts arising [5]) imposing shifting in genuine presentation age, gender, and race. ethnicity, and other demographic PAD biases [6, 7].
2. Project scalability becomes challenging as different border control points would require different training datasets due to the same demographic reasons [5]. Additionally, specific domain considerations like different environments and devices complicate optimal threshold selection in the scenario of ABC reimplementation [8].

Considering the above, we could conclude that the optimal threshold selection process should be automated based on specific methods and parameters.

3.2. RIAPAR

The 2023 revision of ISO/IEC 30107-3:2023 [1] includes new metrics that provide better insight into the real-world performance of a complete biometric system [9]. One new metric is called "RIAPAR", used to measure how well a biometric system detects attacks without interrupting legitimate users. It is calculated using the formula:

$$RIAPAR = BPCER + FNMR + IAPAR \quad (4)$$

where *FNMR* is the proportion of the completed biometric mated comparison trials that result in a false non-match; *IAPAR* is the impostor Attack Presentation Accept Rate defined as the proportion of impostor attack presentations using the same PAIS that result in an accept [2, 4, 10, 11].

The previously common approach used by most of the PAD subsystems selected threshold minimizing equal error rate (EER), equalizing APCER and BPCER, which ignores the PAD operational environment. While the RIAPAR metric which is mandatory for PAD product certification places user experience and convenience higher in the process selection optimal threshold it is not suitable for security-critical areas and not consider the PAD system area of usage.

To mitigate this problem, optimal threshold selection automation following formula could be implemented:

$$Threshold = BT + EA \quad (5)$$

where *BT* is the Base Threshold defined during training with initial dataset APCER and BPCER; *EA* is Environmental Adjustment is the factor that accounts for variations in

environmental conditions that may impact the performance of the face anti-spoofing system.

Environmental Adjustment can be expressed as a function of environmental parameters such as general security conditions and requirements, availability of human intervention to the identification process, biometric presentation-specific conditions, and other PAD product-specific criteria. It can be further defined as:

$$EA = \sum_{i=1}^{n} W_i \times F_i \qquad (6)$$

where n is the number of environmental factors considered; W_i is the weight assigned to each environmental factor based on its importance and impact on the system's performance; F_i is the value of the environmental factor (positive or negative) at a given moment.

By incorporating environmental conditions into the threshold definition, the face anti-spoofing system can dynamically adapt its threshold to optimize performance under varying conditions, enhancing overall accuracy and robustness.

Conclusion

This article is devoted to the research of the aspect of selecting the optimal threshold in face anti-spoofing systems to bolster security while ensuring user convenience. While common methods often prioritize minimizing equal error rates (EER) [12, 13, 14], this approach fails to consider the diverse operational environments and varying security requirements

Most existent production-ready PAD systems use metrics defined in ISO/IEC 30107-3:2023, which makes them compliant with principles and methods of performance assessment of biometric presentation attack detection. Metrics like APCER, BPCER, and ACER described in the article are essential for the successful utilization of any face anti-spoofing system. When testing biometric systems for security vulnerabilities, the sheer number and variety of potential tools used to spoof the system (PAIS) can be overwhelming. It's often impractical, if not impossible, to create a model that encompasses every possible spoofing method. Consequently, leveraging between these and fine-tuning the PAD system becomes a challenge.

The analysis of existing solutions like ABC4EU demonstrates the limitations of static threshold selection based on pre-trained datasets. Therefore, this paper proposes a formula for automated threshold selection that incorporates an “Environmental Adjustment” factor (EA). This factor accounts for the specific context of the PAD system’s deployment, including security needs, and user experience considerations. By dynamically adjusting the threshold based on these environmental parameters, the face anti-spoofing system can function more effectively and securely when deployed in real-world settings.

References

[1] ISO/IEC JTC1 SC37 Biometrics. ISO/IEC 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting. International Organization for Standardization, 2023.

[2] D. Ortega-Delcampo, C. Conde, Á. Serrano, I. M. Diego, E. Cabello: Face Recognition-based Presentation Attack Detection in a Two-step Segregated Automated Border Control e-Gate - Results of a Pilot Experience at Adolfo Suárez Madrid-Barajas Airport. 10.5220/0006426901290138 URL: https://www.scitepress.org/papers/2017/64269/64269.pdf.

[3] D. Ortega-Delcampo, C. Conde, D. Palacios-Alonso, E. Cabello: Border Control Morphing Attack Detection with a Convolutional Neural Network De-Morphing Approach. IEEE Access, vol. 8, pp. 92301-92313, 2020, doi: 10.1109/ACCESS.2020.2994112. URL: https://ieeexplore.ieee.org/document/9091520.

[4] N. Matei, D. Garcia Leon, A. Dosio, F. Batista E Silva, R. Ribeiro Barranco, J.C. Ciscar Martinez: Regional impact of climate change on European tourism demand, EUR 31519 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-68-03925-0, doi:10.2760/899611, JRC131508. URL: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC131508/JRC131508_01.pdf.

[5] L. Gabrielli, E. Deutschmann, F. Natale: Dissecting global air traffic data to discern different types and trends of transnational human mobility. EPJ Data Sci. 8, 26 (2019). https://doi.org/10.1140/epjds/s13688-019-0204-x.

[6] ChaLearn: 2020 Looking at People Fair Face Recognition challenge ECCV. URL: https://chalearnlap.cvc.uab.cat/challenge/38/description.

[7] N. Alshareef, X. Yuan, K. Roy, M. Atay.: A Study of Gender Bias in Face Presentation Attack and Its Mitigation. Future Internet 2021, 13, 234. URL: https://doi.org/10.3390/fi13090234.

[8] K. Srivatsan, M. Naseer, K. Nandakumar. FLIP: Cross-domain Face Anti-spoofing with Language Guidance. URL: https://arxiv.org/pdf/2309.16649v1.pdf.

[9] IDR&D: What's New in the Recent Update of ISO/IEC 30107 for Biometric Presentation Attack Detection. URL: https://www.idrnd.ai/iso-30107-2023-liveness-updates.

[10] M. Ibsen, L. J. Gonzalez-Soler, C. Rathgeb, C. Busch: TetraLoss: Improving the Robustness of Face Recognition against Morphing Attacks, da/sec - Biometrics and Security Research Group, Hochschule Darmstadt, 64295, Germany. URL: https://arxiv.org/pdf/2401.11598v1.

[11] H. Lypak, A. Rzheuskyi, N. Kunanets and V. Pasichnyk, "Formation of a consolidated information resource by means of cloud technologies", 5th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology PIC S&T'2018, October 9–12, 2018.

[12] Y. Brice Wandji Piugie. Performance and Security Evaluation of Behavioral Biometric Systems. Computer Science. Université de Caen Normandie, 2023. URL: https://hal.science/tel-04397160/document.
[13] L. J. Gonzalez-Soler, M. Gomez-Barrero, C. Busch: Toward Generalizable Facial Presentation Attack Detection Based on the Analysis of Facial Regions. IEEE Access, vol. 11, pp. 68512-68524, 2023, doi: 10.1109/ACCESS.2023.3292407. URL: https://ieeexplore.ieee.org/document/10173519.
[14] Yevseiev, S., Ponomarenko, V., Laptiev, O., Milov, O., Korol, O., Milevskyi, S. Synergy of building cybersecurity systems. Kharkiv: PC TECHNOLOGY CENTER, 2021, 188. doi: http://doi.org/10.15587/978-617-7319-31-2