Explication d'Attaques par Présentation d'Empreintes Digitales

Augustin DIERS Christophe ROSENBERGER

University Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR6072, F-14000 Caen, France

Résumé – Les systèmes biométriques sont utilisés dans notre vie quotidienne mais ils font l'objet d'attaques visant à les contourner en tant que solution de sécurité. Les attaques par présentation d'empreintes digitales se produisent lorsqu'un imposteur tente d'utiliser un faux échantillon lors de l'étape d'acquisition pour se faire passer pour une autre personne ou pour ne pas être identifié. Fournir une explication à l'opérateur (qui n'est pas un expert en biométrie) pourrait être d'un grand intérêt pour de nombreuses applications (contrôle des frontières, contrôle d'accès physique). Dans cet article, nous proposons une méthode de détection des attaques par présentation d'empreintes digitales avec un retour d'information explicatif pouvant être compris par n'importe quel utilisateur. Les expériences ont été réalisées sur le jeu de données issues de Fingerprint Liveness Detection Competition (LivDet) de 2015 et contient plus de 58 000 images d'empreintes digitales de bonne foi et d'attaques. La méthode proposée atteint un taux de reconnaissance de 95.7% sur LivDet2015 avec un retour utilisateur qui peut être compris par n'importe quel utilisateur.

Abstract – Biometric systems are used in our daily life but are subject to attacks to bypass them as a security solution. Presentation attacks in digital fingerprints occur when an imposter tries to use a fake sample at the acquisition step to impersonate another individual or not to be identified. Providing an explanation for the operator (who is not an expert in biometrics) could be of great interest for many applications (border control, physical access control). In this paper, we propose a fingerprint presentation attack detection method with explainability feedback that can be understood by any user. The experiments has been realized on the Fingerprint Liveness Detection Competition (LivDet) dataset in 2015 and contains more than 58,000 bona fide and attack fingerprint images. The proposed method reaches an accuracy rate of 95.7% on LivDet2015 with feedback that can be understood by any user.

1 Introduction

La biométrie a pour objectif d'identifier automatiquement un utilisateur ou de vérifier son identité en utilisant des caractéristiques morphologiques ou comportementales. De nos jours, l'empreinte digitale est utilisée comme l'une des modalités biométriques les plus sécurisées et fiables pour l'authentification des utilisateurs. Cette utilisation massive comme solution de sécurité a donc conduit à l'apparition et à la multiplication des attaques contre ces systèmes. Il est donc important d'ajouter une fonction à tout système biométrique, un Détécteur d'Attaque par Présentation (PAD) ou Anti-Spoofing [2]. Les autres fonctions étant l'enrôlement, l'authentification et l'identification, comme décrit dans [3]. Les attaques par présentation se produisent au niveau du capteur biométrique, où l'imposteur tente de falsifier la donnée biométrique d'une autre personne ou de créer une nouvelle empreinte digitale afin d'accéder à des informations confidentielles auxquelles il/elle n'a pas droit.

Les méthodes coopératives et non coopératives sont les deux approches utilisées par les imposteurs pour fabriquer de fausses empreintes digitales appelées instruments d'attaque par présentation (PAI). Dans les méthodes coopératives, l'imposteur collabore avec l'individu dont il souhaite usurper l'identité afin d'obtenir un moule parfait de son empreinte digitale (par exemple, pour le pointage horaire). Dans les méthodes non coopératives, l'imposteur tente de falsifier l'empreinte digitale d'un individu sans son consentement à partir de différentes sources (empreintes latentes, cadavres, copie synthétique). Dans les deux cas, les imposteurs utilisent des matériaux malléables (latex, silicone, pâte à modeler...) pour reproduire la forme de l'empreinte digitale à usurper.

FIGURE 1 : Exemple d'une véritable empreinte (à gauche) et une contrefaçon (à droite) issues de la base LiveDet2015.

Les solutions matérielles et logicielles sont les deux approches proposées dans la littérature pour contrer les attaques par présentation [11, 10]. Dans [1], les auteurs ont proposé différentes solutions matérielles pour la détection d'attaque par présentation (PAD). Ces solutions nécessitent l'intégration de composants spécifiques sur le capteur afin de mesurer la distorsion des crêtes, l'élasticité, la température et la conductivité corporelle. Les solutions logicielles sont subdivisées en deux groupes : les méthodes dynamiques et statiques [4]. Les méthodes PAD dynamiques utilisent les caractéristiques de l'empreinte digitale qui sont censées varier sur un flux vidéo du capteur. Pour implémenter cette solution, deux ou plusieurs images de l'empreinte digitale sont capturées à des instants très rapprochés (de 0 à 5 secondes) comme illustré dans [5]. De nos jours, la plupart des méthodes PAD statiques reposent sur l'apprentissage profond, notamment les réseaux de neurones convolutifs (CNN) ou les transformers

[9]. Une limite importante de ces solutions en boîte noire est la difficulté de faire confiance au résultat de la décision sans retour d'information compréhensible (même lorsque les performances sont généralement très élevées). Nous avons l'intention de contribuer à cette tendance dans cet article pour donner des éléments d'explications de la sortie du PAD.

Cet article est organisé comme suit. La section 2 est consacrée à la méthode proposée. La section 3 concerne le protocole de l'étude et les résultats expérimentaux. Nous concluons et donnons quelques perspectives dans la section 4.

2 Méthode proposée

L'objectif de la méthode proposée est de construire une chaîne de traitement d'apprentissage machine pour la détection d'attaques par présentation des empreintes digitales, en ajoutant des explications intelligibles pour un opérateur. Pour cette tâche, nous analysons préalablement la fiabilité des caractéristiques utilisées. Dans ce qui suit, nous détaillons d'abord les caractéristiques utilisées, puis la chaîne traitement proposée.

2.1 Extraction de caractéristiques

Dans ce travail, nous utilisons des caractéristiques de texture issues du descripteur *Local Binary Pattern* (LBP), choisi pour sa simplicité, sa robustesse aux variations et sa capacité à capturer des motifs locaux caractéristiques des empreintes digitales. L'objectif est d'extraire, à partir d'images en niveaux de gris, des signatures numériques décrivant la répartition locale des textures. Les images, indépendamment de leur format initial sont subdivisées en une grille régulière de 4×4 sous-blocs. Pour chaque sous-bloc, nous appliquons le descripteur LBP en configuration circulaire avec un rayon de 1 pixel et 8 points voisins (*uniform pattern*), puis nous calculons l'histogramme de distribution des motifs obtenus. Chaque histogramme est normalisé pour refléter la distribution relative des motifs dans le sous-bloc. Les 16 histogrammes ainsi obtenus sont concaténés pour former un vecteur de caractéristiques global. Afin d'éviter les doublons dans les bases de données, pour chaque image, seul un fichier est conservé, avec priorité au format `.bmp` lorsqu'il est disponible. L'étiquette de classe (vivante ou falsifiée), l'année, le capteur utilisé, le type d'attaque (*medium*) ainsi que le sous-ensemble (*Training* ou *Testing*) sont automatiquement déduits de l'arborescence des dossiers. L'ensemble des caractéristiques extraites est ensuite stocké dans un fichier CSV, chaque ligne représentant une image et contenant à la fois ses métadonnées et ses descripteurs LBP. Cette procédure permet de générer un ensemble cohérent et non redondant de données exploitables pour la détection d'attaques par présentation.
Le choix du LBP repose sur son aptitude à capturer les irrégularités locales introduites par les fausses empreintes, lesquelles présentent généralement une homogénéité plus marquée et des motifs répétitifs, contrairement aux empreintes vivantes qui exhibent une variabilité naturelle plus grande.

2.2 Conception du PAD

Le PAD est une chaîne de traitement consistant à réaliser un apprentissage par un modèle statistique classique (Adaboost, SVM (RBF), Knn (K nearest neighbours), Random Forest, naïve Bayes et réseau de neurones) sur les données extraites précédentes. Un méta-modèle est également utilisé à partir des précédents pour combiner les scores de décision. Pour l'entraînement et le test, nous adoptons un processus de validation croisée (66% en apprentissage et 44% en test).

2.3 Explicabilité

Nous adoptons trois approches d'explication de la décision du PAD. La première est basée sur une analyse statistique des données réalisée pour identifier les variables les plus pertinentes pour l'identification d'attaques. Elle permet de projeter les données d'un échantillon inconnu dans l'espace des variables pertinentes pour voir son positionnement par rapport à la frontière de décision. On peut également estimer la probabilité conditionnelle d'une attaque considérant la valeur d'une variable pertinente. La seconde approche considère la confiance des différents modèles statistiques pour estimer une confiance dans la décision. Nous utilisons la même chaîne de traitement pour identifier le PAI (méthode d'attaque) par apprentissage. Enfin, la troisième approche consiste à générer une carte de saillance pour identifier les zones de l'empreinte digitale contribuant à la décision finale.

3 Protocole et résultats expérimentaux

3.1 Base de données

Les bases de données les plus utilisées pour l'évaluation de la détection des attaques par présentation sont issues de la compétition LivDet. Depuis 2009, de nombreux participants ont proposé leurs solutions de modèles PAD pour ce benchmark. Dans ce travail, nous utilisons le dataset LivDet15 [7], en référence à la compétition de 2015. Les termes *Alive*, *Spoof* désignent respectivement une empreinte vivante ou une fausse empreinte. La matière utilisée pour produire la fausse empreinte est également indiquée. Les capteurs Biometrika, Green Bit, Digital Persona et Crossmatch sont utilisés pour l'acquisition des empreintes digitales. Le jeu de données est composé de 58 583 échantillons (30 471 authentiques, 28 112 attaques).

3.2 Métriques d'évaluation

Pour la mesure de performance de la détection des attaques par présentation, nous considérons trois métriques : **APCER** (Taux d'erreur de classification des attaques par présentation) : Le pourcentage d'exemples d'attaques incorrectement identifiés comme des exemples authentiques. Une valeur élevée indique une vulnérabilité de sécurité accrue. **BPCER** (Taux d'erreur de classification des empreintes authentiques) : Le pourcentage d'empreintes authentiques incorrectement identifiés comme des attaques. Une valeur élevée indique une gêne accrue pour l'utilisateur. **CA** (Précision de classification) : Elle mesure la précision de bonne reconnaissance des exemples authentiques et des attaques.

3.3 Performance du PAD

Nous présentons les performances du PAD utilisé sur LiveDet2015. Le méta-modèle atteint une performance élevée avec

une précision de 95.7%. Nous obtenons des valeurs similaires d'APCER et de BPCER de 2.1%, montrant ainsi le bon comportement de la méthode proposée. Nous avons également appliqué le même traitement aux datasets entre 2009 et 2021. La CA pour 2009 est de 98.8%, 96.8% pour 2011, 96, 6% pour 2013, 95.7% pour 2017, 96.6% pour 2019 et 98.5% pour 2021.

3.4 Explicabilité

La méthode proposée atteint une très bonne efficacité. Même si les valeurs d'APCER et de BPCER sont faibles, des erreurs peuvent survenir. Dans des applications sensibles comme le contrôle aux frontières, il peut être intéressant de fournir une explication compréhensible à l'opérateur. À titre d'illustration, nous avons utilisé un échantillon spécifique correspondant à une attaque (voir Figure 1). Nous proposons différentes explications à l'opérateur : par exemple, avant de déployer le système PAD, d'identifier les caractéristiques fiables. Nous pouvons tracer la distribution des valeurs de chaque caractéristique pour toutes les classes (authentiques et attaques). La Figure 2 montre une illustration de ces distributions (une caractéristique LBP). Nous pouvons estimer, étant donné la valeur de la caractéristique (représentée par une ligne sombre), la valeur de probabilité conditionnelle que l'échantillon corresponde à une attaque. Dans l'exemple de la Figure 2, on peut clairement voir que plus la caractéristique est élevée, plus la confiance que l'échantillon est authentique est grande (distribution en bleu).

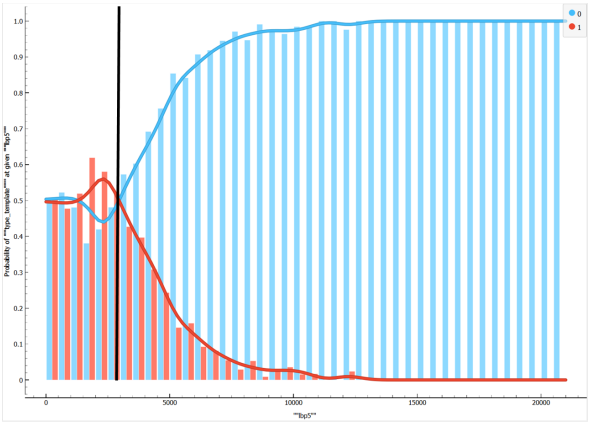

FIGURE 2 : Probabilité conditionnelle pour une caractéristique spécifique (attaque en rouge, authentique en bleu).

Pour les caractéristiques les plus informatives, nous pouvons tracer tous les échantillons dans le jeu de données d'entraînement en fonction de leurs valeurs. La Figure 3 montre un exemple pour un échantillon inconnu représenté en noir. Dans ce cas, il est assez clair que l'échantillon correspond à une attaque (puisqu'il fait partie de la zone rouge associée aux attaques). Il s'agit d'un indicateur visuel facile à comprendre.

Tout modèle d'apprentissage peut retourner une valeur de probabilité pour chaque classe. Pour un système PAD, il est utile de calculer une mesure de confiance en considérant cette valeur ou, lorsqu'il existe plusieurs modèles, la force du consensus entre eux. La Figure 4 montre la sortie de chaque modèle pour l'exemple. Tous s'accordent à dire que

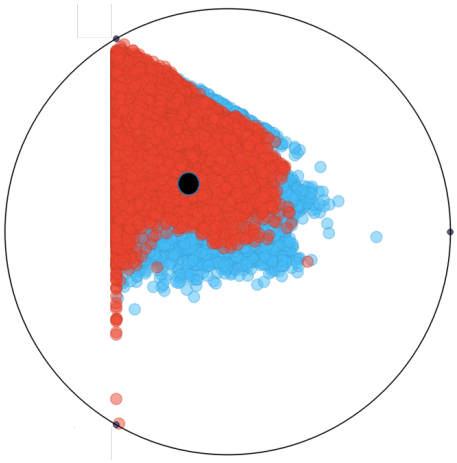

FIGURE 3 : Projection sur le sous-espace des 3 caractéristiques les plus fiables d'un échantillon inconnu représenté par un point noir (les points rouges correspondent aux échantillons d'attaque, les points bleus aux authentiques).

cet échantillon est une attaque. Nous pouvons facilement calculer une valeur de confiance en moyennant les valeurs de probabilité (avec ou sans pondération de l'efficacité de chaque modèle). Un autre retour possible est le nombre de modèles étant d'accord sur la sortie (dans ce cas, 7/7).

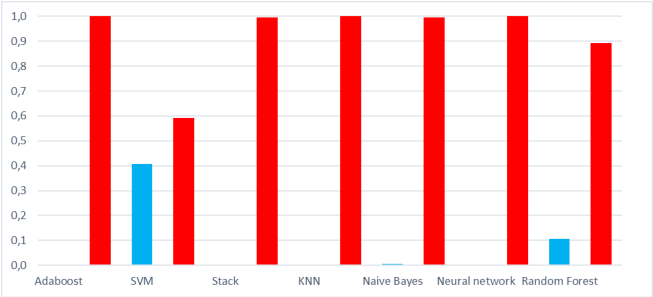

FIGURE 4 : Exemple de valeurs de confiance pour un échantillon d'attaque. Chaque modèle retourne 1 comme étant une attaque.

En cas d'attaque, l'instrument d'attaque (PAI) peut être identifié et donné comme retour à l'opérateur. La Figure 5 présente les performances de la reconnaissance. Le méta-modèle fournit le meilleur résultat avec une précision de 98.8%, tandis qu'un réseau de neurones seul atteint une précision de 98.7%. Les cartes de saillance permettent de reconnaître les zones apportant de l'information dans l'empreinte digitale. Cela permet par exemple de détecter les attaques dites "singular pixel attack".

4 Conclusion et perspectives

Nous avons montré dans ce travail qu'il est possible de proposer un système PAD très efficace (moins efficace que les systèmes PAD profonds tels que [8] avec une valeur de CA de 99.5%), mais offrant des retours utiles et faciles à comprendre à un opérateur pour expliquer la décision.

	LivDet 2009	LivDet 2011	LivDet 2013	LivDet 2015	LivDet 2017	LivDet 2019	LivDet 2021
BPCER	22.03(±9.94)	23.52(±13.85)	6.26(±9.18)	5.94(±3.21)	4.99(±0.89)	4.15(±2.65)	1.1(±0.65)
APCER	12.47(±6.12)	26.97(±18.24)	21.55(±37.21)	6.89(±3.83)	4.97(±0.62)	4.75(±5.71)	27.02(±23.31)

TABLE 1 : Moyenne et écart-type des valeurs BPCER et APCER des trois meilleurs algorithmes pour chaque concours LivDet [6]

	LivDet 2009	LivDet 2011	LivDet 2013	LivDet 2015	LivDet 2017	LivDet 2019	LivDet 2021
BPCER	1.35	3.31	3.09	7.75	4.89	3.51	1.24
APCER	1.07	2.77	3.51	4.30	3.70	3.13	1.36

TABLE 2 : Valeurs BPCER et APCER pour chaque concours LivDet avec notre méta-modèle.

Model	AUC	CA	F1	Prec	Recall	MCC
AdaBoost	0.952	0.952	0.952	0.953	0.952	0.904
SVM	0.584	0.529	0.517	0.548	0.529	0.084
Stack	0.999	0.988	0.988	0.988	0.988	0.976
kNN	0.965	0.959	0.959	0.959	0.959	0.917
Naive Bayes	0.698	0.629	0.627	0.644	0.629	0.276
Neural Network	0.998	0.987	0.987	0.987	0.987	0.973
Random Forest	0.996	0.975	0.974	0.975	0.975	0.949

FIGURE 5 : Résultats de reconnaissance des PAI.

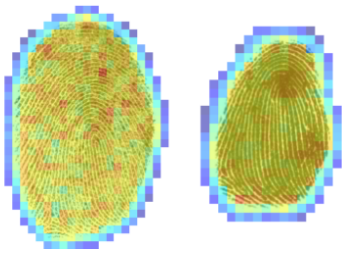

FIGURE 6 : Exemple de cartes de saillance (pour une véritable empreinte à gauche et une contrefaçon à droite)

En perspective, nous envisageons d'utiliser des CNN pour l'extraction de caractéristiques afin d'améliorer les résultats en performance et d'améliorer la qualité des cartes de saillance ainsi que de proposer des explications en langage naturel.

Références

[1] Roberto CASULA, Marco MICHELETTO, Giulia ORRÚ, Gian Luca MARCIALIS et Fabio ROLI : Towards realistic fingerprint presentation attacks : The screenspoof method. *Pattern Recognition Letters*, 171:192–200, 2023.

[2] Javier GALBALLY, Julian FIERREZ, Raffaele CAPPELLI et Gian Luca MARCIALIS : Introduction to presentation attack detection in fingerprint biometrics. *In Handbook of Biometric Anti-Spoofing : Presentation Attack Detection and Vulnerability Assessment*, pages 3–15. Springer, 2023.

[3] JOANNES FALADE, SANDRA CREMER et CHRISTOPHE ROSENBERGER : Comparative study of fingerprint database indexing methods. *In 2019 Cyberworlds Conference*, Dec. 2019.

[4] Emanuela MARASCO et Arun ROSS : *A Survey on Anti-Spoofing Schemes for Fingerprint Recgnition Systems*. ACM Computing Surveys, Vol. 47,, 2014.

[5] Emanuela MARASCO et Carlo SANSONE : *Combining perspiration and morphology based static features for fingerprint liveness detection*. Pattern Recognition Letters, 2012.

[6] Marco MICHELETTO, Giulia ORRÙ, Roberto CASULA, David YAMBAY, Gian Luca MARCIALIS et Stephanie SCHUCKERS : Review of the fingerprint liveness detection (livdet) competition series : from 2009 to 2021. *Handbook of biometric anti-spoofing : presentation attack detection and vulnerability assessment*, pages 57–76, 2023.

[7] V. MURA, L. GHIANI, G. L. MARCIALIS, F. ROLI, D. A. YAMBAY et S. A. SCHUCKERS : Livdet 2015 fingerprint liveness detection competition 2015. *In 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, Sep. 2015.

[8] Additya POPLI, Saraansh TANDON, Joshua J ENGELSMA et Anoop NAMBOODIRI : A unified model for fingerprint authentication and presentation attack detection. *In Handbook of Biometric Anti-Spoofing : Presentation Attack Detection and Vulnerability Assessment*, pages 77–99. Springer, 2023.

[9] Kiran RAJA, Raghavendra RAMACHANDRA, Sushma VENKATESH, Marta GOMEZ-BARRERO, Christian RATHGEB et Christoph BUSCH : Vision transformers for fingerprint presentation attack detection. *In Handbook of Biometric Anti-Spoofing : Presentation Attack Detection and Vulnerability Assessment*, pages 17–56. Springer, 2023.

[10] Kashif SHAHEED, Piotr SZCZUKO, Munish KUMAR, Imran QURESHI, Qaisar ABBAS et Ihsan ULLAH : Deep learning techniques for biometric security : A systematic review of presentation attack detection systems. *Engineering Applications of Artificial Intelligence*, 129:107569, 2024.

[11] Haohao SUN, Yilong ZHANG, Peng CHEN, Haixia WANG et Ronghua LIANG : Internal structure attention network for fingerprint presentation attack detection

from optical coherence tomography. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2023.