

**RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ**

zaprasza na

OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr. inż. Jalil Nourmohammadi Khiarak’a

która odbędzie się w dniu **28 kwietnia 2026 roku**, o godzinie **14:00** w trybie hybrydowym

Temat rozprawy:

„Countermeasure algorithms against subterfuge in mobile biometric systems”

Promotor: prof. dr hab. inż. Włodzimierz Kasprzak – Politechnika Warszawska

Recenzenci: dr hab. inż. Adam Czajka– Uniwersytet Notre Dame - USA

prof. dr hab. inż. Khalid Saeed– Politechnika Białostocka

prof. dr hab. inż. Krzysztof Ślot – Politechnika Łódzka

Obrona odbędzie się zdalnie na platformie MS Teams oraz stacjonarnie w Audytorium Centralnym Wydziału Technik Informatycznych Politechniki Warszawskiej w Warszawie. Osoby zainteresowane uczestnictwem w obronie proszone są o zgłoszenie chęci uczestnictwa w formie elektronicznej na adres sekretarza komisji: dr hab. inż. Grzegorz Pastuszek, email: grzegorz.pastuszek@pw.edu.pl, do dnia 27.04.2026r. godz. 15:00.

Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Warszawa, Plac Politechniki 1.

Streszczenie rozprawy doktorskiej i recenzje są zamieszczone na stronie internetowej: <https://www.bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego/Doktoraty/Wszczete-po-30-kwietnia-2019-r/Rada-Naukowa-Dyscypliny-Informatyka-Techiczna-i-Telekomunikacja/Jalil-N.-Khiarak>

Przewodniczący Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej
prof. dr hab. inż. Jarosław Arabas

Streszczenie

Biometryczna autentykacja, wspierana przez zaawansowane telefony komórkowe oraz technologie takie jak rozpoznawanie ucha, wywarła znaczący wpływ na współczesne społeczeństwa. Rozpoznawanie ucha, umożliwione dzięki sztucznej inteligencji, Internetowi Rzeczy (IoT) oraz wizji komputerowej, stało się powszechną techniką biometrycznej autentykacji w urządzeniach mobilnych. Wraz z szerokim wykorzystaniem aplikacji biometrycznych w smartfonach, laptopach, usługach lotniskowych i bankowych, zapewnienie ich bezpieczeństwa staje się kluczowe. Obfitość wrażliwych informacji przechowywanych na tych urządzeniach podkreśla potrzebę ochrony danych i tożsamości. W związku z tym konieczne jest badanie metod wykrywania ataków prezentacyjnych oraz technik weryfikacji żywotności jako środków przeciwdziałania atakom fałszywym. Nowe techniki muszą uwzględniać zmieniające się scenariusze akwizycji oraz rosnący poziom zakłóceń w zebranych danych biometrycznych.

Niniejsza praca doktorska koncentruje się na badaniu metod przeciwdziałania atakom prezentacyjnym w rozpoznawaniu ucha, w szczególności w kontekście ataków bazujących na dotyku ucha oraz zdjęciach ucha, poprzez opracowanie systemu PAD i jego analizę. Poruszono dwa kluczowe wyzwania biometryczne: PAD dla dotyku ucha oraz PAD dla zdjęć ucha.

Biometria ucha, jako poddziedzina wizji komputerowej, zyskała dużą uwagę ze względu na swoje niezawodne i skuteczne możliwości rozpoznawania osób. Jednak systemy rozpoznawania ucha (zarówno zdjęć, jak i dotyku) są podatne na ataki prezentacyjne lub fałszywe, w których atakujący próbują ukryć swoją prawdziwą tożsamość i oszukać system biometryczny. Aby chronić te systemy przed atakami, rozwijane są nowe metody PAD, czyli przeciwdziałania oszustwom w rozpoznawaniu ucha. W tej pracy zbadano i porównano różne nowatorskie podejścia do wykrywania ataków prezentacyjnych na ucho, wykorzystując zarówno tradycyjne cechy, jak i techniki głębokiego uczenia. Dane czasowe, odzwierciedlające dynamiczne właściwości ataków prezentacyjnych, zostały przeanalizowane pod kątem różnych charakterystyk. Zbadano zastosowanie głębokich architektur neuronowych oraz cech ręcznie definiowanych, a także potencjalne rozszerzenia w zakresie PAD. Ponadto zbadano metody głębokiego uczenia wykorzystujące różne wstępnie wytrenowane architektury sieci neuronowych w celu poprawy

Prof. dr hab. inż. Khalid Saeed
Wydział Informatyki
Politechnika Białostocka
ul. Wiejska 45A, 15-351 Białystok
Tel. (+48-85) 746 9196
k.saeed@pb.edu.pl

Białystok, 28.01.2026 r.

RECENZJA rozprawy doktorskiej
mgr. inż. Jalila Nourmohammadiego Khiaraka
z Wydziału Elektroniki i Technik Informacyjnych
Politechniki Warszawskiej
zatytułowanej *Developing Countermeasure algorithms against subterfuge in
mobile biometric systems*

Promotor:

Prof. dr hab. inż. Włodzimierz Kasprzak
Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska

Niniejszą recenzję przygotowałem na zlecenie zawarte w piśmie z dnia 01.12.2025, które otrzymałem od profesora Jarosława Arabasa przewodniczącego Rady Naukowej Dyscypliny informatyka techniczna i telekomunikacja Politechniki Warszawskiej na podstawie uchwały Rady nr 65/2025 z dnia 24.06.2025.

I. Omówienie, opis i ocena zawartości rozprawy

Rozprawa doktorska mgr. inż. Jalila Nourmohammadiego Khiaraka koncentruje się na czterech problemach: wprowadzeniu nowej cechy biometrycznej rozpoznawania ucha dotykowego, zaproponowaniu metody wykrywania ataków prezentacji (Presentation Attack Detection - PAD) opartej na zdjęciu i dotyku ucha oraz zbieraniu danych na urządzeniach mobilnych dla biometrii ucha. Doktorant przedstawia kilka nowych użytecznych algorytmów. Najważniejsze z nich dotyczą badania metod wykrywania ataków prezentacyjnych (PAD lub tak zwany *spoofing*) oraz technik weryfikacji żywotności jako środków przeciwdziałania fałszywym atakom. Dzięki opracowaniu nowej bazy danych oraz badaniom nad nowymi algorytmami rozpoznawania użytkownika i detekcji ataku, badania autora wnoszą znaczący wkład w dziedzinę biometrii oferując rozwiązania problemów uwierzytelniania w urządzeniach mobilnych. Integracja trybu dotyku ze zdjęciem ucha okazuje się obiecującym podejściem, poprawiając dokładność metod zachowania bezpieczeństwa systemów biometrycznych. Rozprawa napisana jest w języku

angielskim i ma charakter teoretyczno-doświadczalny. Zawiera 145 stron formatu C5 tekstu, rysunków i tabel. Składa się z siedmiu rozdziałów, bibliografii oraz czterech załączników. Zawartość załącznika Appendix B jest kontynuacją części merytorycznej rozprawy, gdyż tam znajdują się opisy algorytmiczne (pseudokody) implementacji pięciu programów komputerowych.

Rozdział pierwszy przedstawia wprowadzenie do systemów PAD na urządzeniach mobilnych, wskazując najnowsze trendy i odnosząc się do pojawiających się wyzwań w tej dziedzinie. Dodatkowo, w rozdziale tym, autor opisuje założenia i cel pracy oraz zakres i strukturę rozprawy doktorskiej.

W *rozdziale drugim* doktorant podaje szczegóły wykorzystanej bazy danych WUT-Ear V1.0 (Warsaw University of Technology Ear Version 1.0). Rozdział zaczyna od przeglądu, który przedstawia zakres bazy danych oraz uzasadnienie jej stworzenia, a następnie szczegółowo opisuje cztery kluczowe zestawy danych w bazie. Każdy zestaw koncentruje się na różnych aspektach biometrii ucha i PAD, tj. rzeczywistych zdjęciach ucha, fałszywych zdjęciach ucha, rzeczywistych dotyków ucha oraz zestawie fałszywych dotyków ucha. Rozdział kończy się podsumowaniem, które podkreśla kluczowe omówione kwestie oraz znaczenie bazy danych WUT-Ear V1.0 dla rozwoju badań w dziedzinie biometrii ucha i PAD.

W *rozdziale trzecim* doktorant analizuje szczegółowe badanie na temat oceniania skuteczności uwierzytelniania ucha na urządzeniach mobilnych. W szczególności bada wydajność algorytmów wykrywania ataków prezentacji oraz metod weryfikacji użytkownika na podstawie nowo opracowanej bazy danych. Bezpieczeństwo systemów uwierzytelniania na podstawie ucha spotyka poważne wyzwania, szczególnie w przypadku ataków prezentacji, gdy fałszerze próbują oszukać system, prezentując fałszywe lub przerobione obrazy ucha. Autor analizując problem, zajął się tym aspektem oraz opracował odpowiednie algorytmy z obiecującymi wynikami.

Rozdział czwarty bada różne sposoby i możliwości wykorzystania dotyku ucha jako metody uwierzytelniania użytkownika telefonu komórkowego oraz kontroli ich dostępu. Przedstawiono propozycję metody rejestrowania danych dotyku ucha. Autor przedstawia rozwiązanie bardzo ważnego problemu w aspekcie dotyku ucha przez telefon, a mianowicie problemu brakujących punktów dotyku (*missing data points*), które są zasadnicze do obliczania elementów wektora cech. Doktorant twierdzi, iż jego zaproponowana metoda rozwiązuje ten problem skutecznie, z wysokim poziomem wydajności. Twierdzi, że wspomniany brak nie ma dużego wpływu na jego metodę: „*By focusing on the consistent elements of the ear's structure, the proposed system can extract meaningful biometric data even when some points are missing*” – cytat ze str. 125 rozprawy doktorskiej. Moim zdaniem, problem ten nie może być rozwiązany na tej podstawie i dalej jest poważnym wyzwaniem, gdyż brak niektórych punktów dotyku może pojawić się w różnych miejscach i na różne sposoby u użytkowników, w zależności od wielu czynników. Chciałbym więc, aby doktorant ustosunkował się do tej uwagi podczas obrony, z naciskiem na wpływ braku tych punktów na wektor cech.

W rozdziale piątym autor przedstawia nowatorskie podejście łączące zdjęcia ucha z jego dotykiem, aby dostać biometryczny multimodalny system do uwierzytelniania użytkownika telefonu komórkowego. Otrzymany model jest ciekawym rozwiązaniem – fuzją dwóch cech. Zdjęcie ucha rejestruje wizualne szczegóły unikalnej budowy ucha, natomiast dane dotykowe ucha rejestrują konkretne wzory powstające, gdy ucho styka się z ekranem dotykowym telefonu. Jednak, moim zdaniem, obawy wymienione powyżej odnośnie „missing points” będą również nadal przeszkodą przy wdrożeniu algorytmów takiej metodologii.

Rozdział szósty skupia się na opracowaniu nowatorskiej techniki wykrywania ataków prezentacji z użyciem dotyku ucha i jego fotografii. Wszystkie szczegóły są podane z naciskiem na wydajność algorytmu biometrii ucha odnośnie *spoofingu*. Osiągnięcia doktoranta oraz wyniki podane w tym rozdziale są pozytywne i na dobrym poziomie. Autor słusznie zauważył, iż systemy biometryczne stają się coraz powszechniej stosowane, napotykają coraz bardziej wyrafinowane próby oszustw, gdzie atakujący wykorzystują wysokiej jakości obrazy, filmy, a nawet modele 3D, aby wprowadzić w błąd systemy rozpoznawania. Tradycyjne metody detekcji *spoofingu* poczyniły znaczące postępy w przeciwdziałaniu tym zagrożeniom, ale zmieniająca się natura ataków wymaga opracowania bardziej zaawansowanych, odpornych i kompleksowych strategii detekcji PAD. Stąd uważam, że pomysł doktoranta w kierunku opracowania nowego podejścia jest słuszny. Pomimo to, rozdział nie jest pozbawiony usterek. Moim zdaniem, pokazane wyniki w tabeli 6-3 nie są do końca zgodne z informacją prezentowaną w tabeli 2-5 w rozdziale drugim. W tabeli 6-3 podane, że dla Dell-GA7 oraz PADNet-1 osiągnięto miarę 12,2 APCER, dla Dell-NL1020 - 0.91, natomiast według tabeli 2-4 liczebność ataków dla pierwszego scenariusza to 2134, a dla drugiego 101. Czy oznacza to, że w pierwszym przypadku jest 260.348 błędnych rozpoznań, a w drugim 0.91? Dodatkowo dla S3D-GA7 osiągnięto 0.37, co przy liczbie 16 uczestników daje 0,0592 uczestnika niesłusznie akceptowanego. Rachunki liczbowe wyglądają wątpliwie i być może wystąpił tu błąd numeryczny. Prosiłbym o wyjaśnienie i interpretację tabeli 6-3.

W rozdziale siódmym doktorant konkluduje swoją pracę podając jej streszczenie, które zawiera opis osiągnięć autora oraz dodatkowe informacje o ograniczeniach swoich metod takich, jak praktyczne kwestie - koszty i dostępność użytkownika.

Całość merytorycznej pracy kończy *Bibliografia*, która zawiera 74 pozycje wybranych referatów i artykułów z literatury światowej pokazującej stan wiedzy i odzwierciedlającej dobrą wiedzę doktoranta. Wszystkie pozycje są cytowane w pracy. Można zauważyć brak pewnych prac na temat rozpoznawania ucha oraz baz danych ucha. Przykładem jest brak pracy Dariusza Frejlichowskiego i Natalii Tyszkiewicz zawierającej bazę uszu do możliwego wykorzystania. Poza tym cytowano dwie prace [54] i [55], które należą do tego samego źródła, które z resztą nie jest opublikowane. Praca pod numerem [46] cytowana w rozprawie jako „recent study” sugeruje, że jest nową, chociaż pochodzi z 2006 roku.

II. Opinia o rozprawie doktorskiej

Wyniki rozprawy doktorskiej mgr inż. Jalila Nourmohammadiego Khiaraka chciałbym oceniać w dwóch płaszczyznach: technicznej i merytorycznej oraz klarowności i czytelności rozprawy.

A. Techniczne brzmienie i merytoryczna kompletność rozprawy

Autor wykazał w swojej pracy umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników. Cytowana literatura jest interesująca, chociaż, nie jest kompletna.

Merytoryczne osiągnięcia doktoranta są na dobrym akceptowanym poziomie. Najważniejsze z nich to:

- (1) Rozwinięcie metod mobilnej biometrii z wykorzystaniem algorytmów maszynowego uczenia.
- (2) Badanie biometryczne oparte na uchu (zdjęcie ucha i dotyk ucha) w celu wzmacniania systemów bezpiecznego uwierzytelniania mobilnego.
- (3) Opracowanie przyzwoitego systemu PAD do wykrywania ataków prezentacyjnych w rzeczywistych scenariuszach.
- (4) Wprowadzenie zbioru danych WUT-Ear V1.0 do trenowania i testowania modeli rozpoznawania.
- (5) Ważnym osiągnięciem przy dopasowaniu wzornika dla scenariusza bez brakujących punktów (*matching scenario without missing points*) jest wykorzystanie wyników prac profesora Andrzeja Pacuta (str. 72 rozprawy) przy rozwiązywaniu problemu optymalizacji równania:

$$(R^*, l^*, P^*) = \underset{(R, l, P) \in \mathcal{R} \times \mathcal{L} \times \mathcal{P}}{\operatorname{argmin}} \sum_{i=1}^N \|t_i - (Rx_{\pi i} + l)\|^2$$

Andrzej Pacut opracował swoją metodę optymalizacji na bazie algorytmu Kabsch-Umeyama, ale nie uwzględniał permutacji. Doktorant rozwinął tę metodę i opracował odpowiedni algorytm, żeby pasował do wymagań swojego podejścia. Pseudokod B1 pokazuje szczegóły algorytmu doktoranta.

Osiągnięcia te doprowadziły do uzyskania bezpiecznych modeli biometrycznych wykorzystujących obraz ucha oraz jego dotyk do telefonu komórkowego. Najważniejsze skutki i efekty tych prac to: uzyskanie bezpiecznego modelu fuzji dotyku i zdjęcia ucha jako systemu multimodalnego; wnioskowanie, iż podejścia głębokiego uczenia (VGG16, MobileNetV2, Vision Transformers) poprawiły dokładność systemu; osiągnięto EER na poziomie 0.04, co dowodzi przydatność biometrii ucha do uwierzytelniania.

Pomimo, że doktorant starał się, żeby praca była klarowna i czytelna, nie ustrzegł się pewnych niedoskonałości formalnych i językowych.

Uwagi

- Opis bazy WUT-Ear jest nieprecyzyjny. Doktorant podaje różne liczby w różnych miejscach: Na str 27 pisze, że rzeczywistych danych było 8000 (sztucznych 8662), ale na stronie 29, że w bazie jest 137 osób z 60 zdjęciami, czyli

8220. Liczebność jest w formie mało precyzyjnej. Natomiast na stronie 30 - pojawia się histogram bez zestawienia tabelarycznego.

- Nie podano pełnej informacji na stronie 31, gdzie doktorant pisze: "*Almost 35 images from 5 different positions are taken in one session from each side therefore totally there are almost 70*". Dalej pisze "*It should be noted that images however some subjects have more than 70. It should be noted that more than 10 subjects have earrings*." Brak informacji, dlaczego liczba zdjęć dla niektórych uczestników wzrosła do 70 oraz czy obecność kolczyków miała znaczenie.

- Sztuczne dotknięcia uszu (str. 26 - "*The dataset not only includes authentic ear images, but also features a comprehensive collection of presentation attack instruments (PAIs), such as printed images and 3D models*" oraz str 43 - "*Figure 2-13*"): Czy faktycznie wykorzystano odciski uszu? Rysunek 2-13 pokazuje tylko dotknięcia palców dłoni.

- Na tej samej stronie istnieje napis "*.... the raw images have a resolution of 4608×3456 pixels, while the preprocessed images have 1992×3120 pixels*". Tutaj wypadłoby napisać, na czym konkretnie polegało wstępne przetwarzanie (preprocessing), że doprowadziło do takich zmian.

- Na rysunku Figure 2-7 widzimy „*Sample visualization by t-SNE*”. Czy zwykłe wrzucenie wizualizacji danych do t-SNE jest poprawne? A może była wykorzystana np. biblioteka:

(https://opentsne.readthedocs.io/en/stable/examples/01_simple_usage/01_simple_usage.html) - stworzyć embedding dla zbiorów treningowych, optymalizować, a następnie użyć transformacji dla nowych danych testowych. Zwykle sklearn.manifold.TSNE, prawdopodobnie, nie ma takich możliwości. Proszę o wyjaśnienie tego aspektu, czy doktorant w ten sposób wykonał wizualizacji danych.

- Figure 3-5: na *flowchart* jest strzałka pokazująca pomijanie ekstrakcji cech. To jest błąd, gdyż można pominąć augmentację do ewaluacji, ale nie "*feature extraction*".

- Strona 67 zawiera ważną informację oraz fakty o biometrii ucha, ale bez żadnego cytatu. Autor powinien podać źródło wymienionych faktów i informacji.

- Brak lub błędne informacje podano w źródłach [52] i [53]. Powinny brzmieć tak: [52] M.D. Atkinson: An Optimal Algorithm for Geometrical Congruence, Journal of Algorithms 8 (1987), pp. 159-172.

[53] Alt, H., Mehlhorn, K., Wagener, H., & Welzl, E.: Congruence, Similarity, and Symmetries of Geometric Objects. Discrete and Computational Geometry 3 (1988), 237-256. doi:10.1007/BF02187910.

Z drugiej strony brak informacji, jak zostały te źródła wykorzystane w rozprawie w odniesieniu do biometrii ucha (strona 69).

B. Klarowność i czytelność rozprawy

Rozprawa jest napisana zrozumiałym językiem - algorytmy, twierdzenia, rysunki i tabele są czytelne i starannie opracowane. Istotne dla tematyki pracy

zagadnienia omówiono czytelnie i przejrzysto. Autor włożył dużo pracy w klarowne przygotowanie manuskryptu. Pomimo wysiłków autora, by praca prezentowała jego osiągnięcia w sposób klarowny, niemożliwym do uniknięcia są drobne usterki językowe lub błędy typograficzne, których przykłady zostały wymienione poniżej.

Drobne usterki

- Str. 18: „a biometric” oraz na str. 19 „biometrics”. Biometrics to nie liczba mnoga dla biometric. Biometrics to rzeczownik i może być używany jako przymiotnik (a biometrics feature) – ang. compound noun.
- Tabele 3-1 i 3-2 wskazują na modele neuronowe z funkcją wyjściową sigmoid czyli zakresy typowo 0-1, natomiast tabele 3-3 i 3-4 pokazują wskaźniki bez jasno określonego progu decyzyjnego.
- Figure 4-15 pokazuje wyniki dla przypadków trzech dotyków i dla jednego dotyku a w podrozdziale 4.4.1, jest informacja, że dla 17 osób nie ma *missing points*, ale brak informacji o liczbie użytkowników z jednym dotykiem.

Znalezione błędy edytorskie

- W wielu miejscach stosowano nieprawidłowe „possessive ‘s’ ” jak: PAD’s, subject’s, image’s, dataset’s, model’s, study’s, etc.
- Podobna sytuacja jest z „while”, gdzie używane w większości przypadków w sensie „whilst”, przykłady są na str. 1, 2, 4, 45, 67, ... itd.
- Str. 27 i w 7. innych miejscach autor pisze „Let’s” zamiast „Let us” – (The general rule is to avoid using contractions for academic and other formal writing.)
- Str. 69: In [51], The authors → ... authors,
to matching → to match
- Str. 70: featured → feature lub features
- Str. 103: Section 5-2, 5-3, 5-4 and 5-5 → 6-2, 6-3, 6-4 and 6-5
- Str. 110: Where → where

III. Merytoryczne osiągnięcia doktoranta

Pan mgr inż. Jalila Nourmohammadiego Khiaraka ma na koncie bardzo dobry dorobek naukowo-publikacyjny. Doktorant jest współautorem 9. publikacji naukowych w indeksowanych czasopismach oraz referatów naukowych na konferencjach krajowych i międzynarodowych. Poza tym doktorant ma 4 artykuły przygotowane do złożenia w czasopismach. Dorobek publikacyjny świadczy o znaczeniu osiągniętych wyników pracy naukowej doktoranta w dziedzinie multimodalnej biometrii ucha. Oznacza to, iż problematyka rozprawy wpisuje się bezpośrednio w bieżący nurt zagadnień z tej właśnie dziedziny. Autor uczestniczył w *European Union Horizon 2020 Grant: AMBER Project* w roli *Early Stage Researcher*.

IV. Wnioski końcowe

Wystawiam pozytywną ocenę rozprawie doktorskiej Pana mgr. inż. Jalila Nourmohammadiego Khiaraka pt. "*Developing Countermeasure algorithms against subterfuge in mobile biometric systems*". Rozprawa stanowi oryginalne rozwiązanie postawionych zadań oraz osobisty wkład Doktoranta w rozwój metod przeciwdziałania fałszerstwa w mobilnych systemach biometrycznych. Stwierdzam, że praca spełnia wymagania i warunki nakładane przez ustawę o stopniach naukowych (art. 190 ust. 2 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce - Dz. U. z 2024 r. poz. 1571, z późn. zm.).

Na tej podstawie wnioskuję o dopuszczenie Autora wymienionej rozprawy doktorskiej do jej obrony.



Khalid Saeed

Łódź, 12.02.2026

Prof. dr hab. inż. **Krzysztof Ślot**
Instytut Informatyki Stosowanej
Politechnika Łódzka

Recenzja rozprawy doktorskiej

Jalil Nourmohammadi Khiarak

Praca pt.

Countermeasure algorithms against subterfuge in mobile biometric systems

1. Tematyka, struktura i kontekst rozprawy

Przedstawiona do recenzji rozprawa doktorska dotyczy zagadnienia biometrycznej weryfikacji użytkownika dokonywanej na podstawie analizy kształtu małżowiny usznej oraz, stanowiącej naturalne uzupełnienie tego wątku, problematyki detekcji prób tzw. ataków prezentacji (ang. Presentation Attack Detection – PAD) polegających na przedstawianiu sensorowi spreparowanych danych. Obydwa przedstawione kierunki prac są osadzone w kontekście procedury uwierzytelniania użytkowników urządzeń mobilnych, mogąc stanowić alternatywę lub uzupełnienie innych powszechnie stosowanych metod biometrycznych, bazujących na analizie obrazu twarzy lub linii papilarnych. Istotnym wyróżnikiem uzasadniającym celowość podjęcia prac w rozważanym obszarze jest uniwersalność implementacji wynikowych metod w środowisku dowolnego współczesnego urządzenia mobilnego (smartfona), z uwagi na nieograniczany dostęp do sensorów rozważanych przez Doktoranta modalności (kamera i ekran dotykowy). Podjęte przez Doktoranta badania mają bardzo aktualny i użyteczny wymiar, a postawione problemy: zarówno poprawna weryfikacja rejestrowanej struktury ucha jak i skuteczna ochrona przed atakami prezentacji, należą do zadań nietrywialnych. W konsekwencji, należy jednoznacznie stwierdzić, że podjęte zagadnienie badawcze odpowiada swoją złożonością i zasadnością wymaganiom stawianym przed rozprawami doktorskimi, a tematyka prac znajduje się w obszarze dyscypliny naukowej informatyka techniczna i telekomunikacja.

Struktura rozprawy jest przejrzysta – w kolejnych rozdziałach Doktorant opisuje prace wykonane w ramach różnych wątków podjętej przez siebie tematyki. Rozprawa rozpoczyna się od prezentacji opracowanej przez Doktoranta bazy danych (część 2). Następnie Doktorant przedstawia zastosowane metody biometrycznej weryfikacji użytkownika na podstawie obrazu ucha i powiązane z tym metody detekcji ataków prezentacji (część 3). Oryginalna propozycja wykorzystania jako biometryki zbioru punktów styku ucha z powierzchnią ekranu smartfona, stanowiącego sensor dotyku jest treścią

czwartej część rozprawy. Dwa ostatnie rozdziały są poświęcone zagadnieniom związanym z weryfikacją tożsamości wykorzystującą dwie modalności powiązane z uchem: obrazów rejestrowanych kamerą smartfona i rozkładów punktów kontaktu ucha z powierzchnią ekranu smartfona. W części 5 pracy została przedstawiona autorska propozycja dwumodalnej metody weryfikacji tożsamości, zaś w rozdziale 6 – metoda dwumodalnej ochrony przed atakami prezentacji.

Przedstawione przez Doktoranta prace są właściwie osadzone w kontekście bieżących prac prowadzonych w obszarze biometrycznej weryfikacji tożsamości na podstawie obrazów ucha i metod detekcji ataków prezentacji na proces akwizycji rozważanej biometryki. Doktorant wykazuje dobrą orientację w aktualnym stanie wiedzy, trafnie identyfikując i korzystając w swoich rozwiązaniach z najnowszych metod i narzędzi algorytmicznych, w tym, zaawansowanych, głębokich architektur neuronowych, jak również poprawnie diagnozując ograniczenia istniejących zasobów danych, prowadzące do opracowanego przez Niego nowego repozytorium danych biometrycznych.

2. Ocena formalnej strony rozprawy

Pomimo nie budzącego zastrzeżeń, klarownego układu rozprawy i czytelnego sposobu prezentacji treści, techniczna strona rozprawy pozostawia sporo do życzenia. Pierwszym, drobnym ale utrudniającym jej śledzenie mankamentem jest użycie szablonu, który w sposób mało przejrzysty uwypukla podział tekstu na poszczególne moduły. Drugorzędny charakter, chociaż również utrudniający jej studiowanie, mają również niedociągnięcia językowe tekstu rozprawy. Język rozprawy nie jest językiem ojczystym Doktoranta, co stanowi oczywistą przeszkodę w klarownym formułowaniu myśli, ale na etapie przygotowania tekstu Doktorant mógł skorzystać z powszechnie dostępnych bezpłatnych narzędzi pozwalających na jego korektę.

Istotniejszym mankamentem formalnej strony pracy są liczne powtórzenia tych samych treści, wynikające ze sposobu przygotowania monografii – składa się ona z fragmentów tekstów publikacji współautorstwa Doktoranta, które jako źródła niezależne zawierały często identyczne lub bardzo zbliżone fragmenty. Całe serie akapitów powielają wcześniej przedstawione informacje dotyczące bazy danych, np. w części 3.3.2, w praktycznie całym wprowadzeniu do części 4, w części 5.4 lub na początku rozdziału 6.4. Również opisy stanu wiedzy w zakresie związanym z proponowanymi metodami się powtarzają (np. 3.2 i 6.1), tak jak i podawane kilkakrotnie wyjaśnienia używanych metryk. Testowaniem cierpliwości Czytelnika jest powtórzenie w rozdziale 5.5.1 i 5.5.2, treści przedstawionych wcześniej w rozdziałach przedstawiających wyniki analiz uzyskiwane osobno dla każdej z dwóch rozważanych modalności – obrazu ucha i punktów kontaktu ucha.

3. Cele i tezy rozprawy

Koncentrując swoje prace badawcze na problematyce biometrycznego uwierzytelniania użytkowników urządzeń mobilnych z wykorzystaniem specyficznej cechy osobniczej, jaką jest morfologia ucha, Doktorant formułuje w rozprawie cztery tezy ('Statements') identyfikujące osiągnięcia pracy i jednocześnie określające szczegółowe kierunki podjętych badań. Pierwsza z nich:

S1: We introduced and investigated on an ear-photo presentation attacks detection and recognition system on mobile device,

informuje o przeprowadzeniu prac nad rozpoznawaniem osób na podstawie obrazów ucha pozyskiwanych kamerą smartfona oraz nad detekcją ataków prezentacji dokonywanych z użyciem podstawionych w postaci obrazów danych. Druga teza:

S2. Ear-touch characteristic is an effective biometric, which is based on mobile devices with multiple touchscreens

dotyczy osiągnięć Doktoranta w zakresie biometrycznej analizy powierzchni kontaktu ucha i ekranu dotykowego smartfona. Trzecia teza:

S3. Ear photo and ear-touch multimodal biometrics improve ear recognition system accuracy in comparison with its single model,

informuje o możliwości zwiększenia poprawności klasyfikacji rozważanej biometryki z użyciem podejścia dwumodalnego, uwzględniającego analizę obrazu ucha i analizę rozkładu punktów kontaktu ucha z powierzchnią sensora. Ostatnia teza:

S4. Ear photo and ear-touch multimodal biometrics improve ear PAD accuracy concerning for to a single model.

sygnalizuje możliwość poprawy skuteczności detekcji ataków prezentacji z wykorzystaniem analizy informacji pochodzących jednocześnie z dwóch rozważanych modalności – obrazu i rozkładu punktów styku. Przedstawione wątki tematyczne w kompletny sposób rysują ramy zrealizowanych przez Doktoranta prac badawczych, uszczegółowionych w dalszej części rozprawy w metodycznie poprawny sposób i poprzedzanych kompetentną prezentacją aktualnego stanu prac dla każdego z rozważanych obszarów.

4. Merytoryczna ocena pracy

Przedstawione w rozprawie efekty prac badawczych Doktoranta mają w mojej ocenie bardzo zróżnicowany poziom naukowy – trzy z przedstawionych rozwiązań ograniczają się do wykorzystania szablonowej metodyki postępowania i nie stanowią żadnego wkładu o charakterze naukowym do dyscypliny, ale dwie to oryginalne, interesujące i wartościowe propozycje o dużym potencjale aplikacyjnym i badawczym. Zdecydowanie najważniejszym osiągnięciem Doktoranta jest nowatorska propozycja biometrycznej weryfikacji tożsamości wykorzystującej jako cechę biometryczną rozkład punktów styku ucha z powierzchnią ekranu dotykowego smartfona. Jest to pomysł zupełnie nowego, naturalnego, wygodnego i nie wymagającego świadomej kooperacji, sposobu uwierzytelnienia użytkownika, który dotychczas nie pojawił się w powszechnym obiegu. Z uwagi na pionierski charakter koncepcji, Doktorant stanął w obliczu konieczności zgromadzenia nieistniejącego w dostępnych zasobach zbioru danych niezbędnych do opracowania odpowiedniego algorytmu. W efekcie przeprowadzonych przez Doktoranta prac powstała unikatowa baza danych (WUT-Ear v1) stanowiąca kompleksowe repozytorium materiału eksperymentalnego wspierającego prace nad biometrią wykorzystującą kształt ucha, obejmująca oprócz rejestracji rozkładów punktów kontaktu ucha z sensorem, również obszerny zestaw obrazów ucha oraz równie obszerny zbiór przykładów ataków na procedurę weryfikacji, zarówno obrazów jak i sztucznie wygenerowanych rozkładów punktów kontaktu ucha z sensorem.

Ponieważ opracowanie bazy danych wymagało zastosowania metodyki badawczej: identyfikacji luk w istniejących repozytoriach, określenia scenariuszy akwizycji i scenariuszy ataków, statystycznej ewaluacji zgromadzonego materiału w celu oceny spełnienia stawianych przed nim wymagań oraz zapewnienia odpowiedniego zróżnicowania materiału eksperymentalnego, ten efekt prac Doktoranta, nie wymieniony jawnie w zbiorze kluczowych elementów dorobku rozprawy, uważam za istotne osiągnięcie mające znamiona naukowe. Jak wspomniałem, najcenniejszym, trzecim komponentem

opracowanej bazy danych, są unikatowe rejestracje położenia punktów styku powierzchni ucha z powierzchnią dotykowego sensora smartfona. Również zaproponowana metodyka ataków prezentacji i zgromadzony w tym celu materiał w czwartym komponencie bazy, są pomysłowe i wartościowe. Dwa pierwsze komponenty, zawierające materiał odnoszący się do rejestracji obrazów ucha są mniej oryginalne, chociaż przedstawione wizualizacje dystrybucji obrazów rzeczywistych i spreparowanych w celu przeprowadzenia ataku prezentacji, w przekonujący sposób uzasadniają wysoką wartość materiału eksperymentalnego przygotowanego przez Doktoranta.

W kolejnych podrozdziałach recenzji ustosunkowuję się do czterech wymienionych we wstępie pracy obszarów dokonań Doktoranta. W odniesieniu do materiału zawartego w kolejnych częściach rozprawy, chcę zwrócić uwagę na mankament metodyki ewaluacji przeprowadzanych eksperymentów, którym jest niekompletne statystyczne podsumowanie, pomijające informację o odchyleniach standardowych (lub przedziałach ufności) uzyskiwanych rezultatów. Przedstawiane na wykresach i w tabelach wartości średnie wyników eksperymentów nie dają pełnego poglądu o stabilności (wrażliwości) stosowanych metod w funkcji zmienności danych używanych do ich testowania, co ma znaczenie z punktu widzenia oceny ich przydatności.

4.1. Weryfikacja tożsamości z użyciem obrazów ucha

Pierwszym wątkiem prac poruszonych w rozprawie jest procedura weryfikacji tożsamości na podstawie obrazów ucha. Prace Doktoranta przedstawione w tej części rozprawy uważam za szablonowe i nie wnoszące żadnych nowych elementów do istniejącego stanu wiedzy. Doktorant używa znanych metodyk ekstrakcji cech: uczenie kontrastowe, stosowane dla budowy głębokich modeli przy skąpych zasobach przykładów treningowych oraz metoda LBP. W pierwszym przypadku, procedura analizy to użycie pretrenowanej sieci syjamskiej jako ekstraktora reprezentacji obrazu testowanego i wzorca reprezentującego weryfikowanego użytkownika. Dla obydwu metod ekstrakcji, wynik analizy jest wyznaczany poprzez ocenę podobieństwa wektorów cech, dokonywaną z użyciem odległości kosinusowej i konfrontowaną z wyznaczoną wcześniej eksperymentalnie wartością progową. Standardowa jest też metodyka ewaluacji działania algorytmu, a z jej wyników nie wynikają żadne zaskakujące wnioski. Na marginesie chcę zauważyć, że przy okazji prezentacji opracowanej przez siebie bazy danych, Doktorant kilkakrotnie podkreśla obecność etykiet płci jako istotny walor bazy danych, ale w przeprowadzonych eksperymentach nie podejmuje ciekawego wątku przetestowania hipotezy o potencjalnych różnicach w strukturze ucha kobiecego i męskiego.

Podobnie jak to miało miejsce w przypadku metody weryfikacji użytkownika, przedstawionej w części 3.3, również metoda detekcji faktu podstawienia w procesie akwizycji danych spreparowanych (PAD) wykorzystuje typowy schemat uczenia transferowego, bez żadnych ciekawych lub oryginalnych pomysłów autorskich. Podejście do budowy algorytmu weryfikacji jest schematyczne i pozbawione wnikliwości – Doktorant nie próbuje optymalizować architektury dodanych przez siebie warstw MLP, stosując arbitralnie przyjęte liczby neuronów w warstwach. Jest to mankamentem, bo MobileNetv2 w swojej bazowej wersji zawiera ok. 3.5 mln. parametrów, zaś przyjęta przez Doktoranta architektura dodanego, gęstego klasyfikatora reprezentacji generowanej przez ekstraktor to sporo ponad 2.5 miliona parametrów (ok. $1280 \times 1024 + 1024 \times 1024 + 1024 \times 512$). W konfrontacji ze stosunkowo skromnym zbiorem uczącym, skala niezbędnych do estymacji parametrów budzi podejrzenie nieuchronnego dopasowania sieci do zbioru danych, pochodzącego mimo starań Doktoranta, z bardzo specyficznego rozkładu prawdopodobieństwa. Brak refleksji nad

relacją liczby parametrów względem ilości dostarczanej informacji uczącej, uważam za uchybienie metodologiczne omawianego wątku pracy. Problem dysproporcji staje się jeszcze bardziej ewidentny w drugim przyjętym przez Doktoranta wariancie, arbitralnego zwiększenia głębokości douczania sieci MobileNetV2. Doktorant sugeruje tu zamrożenie tylko pierwszych 16 warstw, co oznacza dodanie do puli parametrów podlegających uczeniu kolejnych ponad 2.5 miliona elementów. W efekcie problem uczenia przed jakim stawia algorytm może być sprowadzony do zadania: określ ponad 5 mln parametrów na podstawie 8 tysięcy przykładów. Ponieważ w moim przekonaniu architektura sieci jest zbyt złożona w stosunku do posiadanej liczby przykładów, przedstawione próby wyjaśnienia wyników eksperymentów są niewłaściwie ukierunkowane i nie prowadzą do zasadnych wniosków.

4.2. Uwierzytlenianie na podstawie rozkładu punktów kontaktu ucha z sensorem

Pierwszym elementem prezentacji opracowanych przez Doktoranta metod biometrycznych, który stanowi oryginalne podejście do rozwiązania złożonego problemu, jest część poświęcona biometrycznej analizie zbiorów punktów styku ucha z powierzchnią sensora – ekranu smartfona. Istotą rozwiązywanego problemu jest znalezienie transformacji najlepiej dopasowującej zarejestrowany zbiór punktów kontaktu do utworzonych w fazie treningu algorytmu wzorców, a następnie, podjęcie decyzji o pozytywnej lub negatywnej weryfikacji próbki. Istnieje kilka metod realizacji tak sformułowanego zadania, o których Doktorant nie wspomina w przeglądzie metod, skupiającym się na kompleksowych algorytmach biometrycznych, a nie na metodach rozwiązania komponentów składowych. Dla porządku więc, należy wskazać, że adekwatnymi metodami mogą tu być tzw. Geometric Hashing, metoda RANSAC czy metoda ICP (Iterative Closest Point). Przedstawione w pracy podejście jest bliskie ostatniej z wymienionych metod, przy czym Doktorant nie przedstawia szczegółów użytego algorytmu wyznaczania parametrów transformacji położeń zbioru dopasowywanych punktów, a jedynie wskazuje na źródło – algorytm opracowany przez Jego Promotora, do którego niestety, nie byłem w stanie dotrzeć (szkoda, że nie pojawił się w załączniku). Autorska modyfikacja tego algorytmu to ekstensywne powtarzanie procedury estymacji macierzy obrotu i wektora translacji dla wszystkich możliwych permutacji zbioru punktów styku. Oprócz metodyki analizy nieznanego rozkładu punktów styku, Doktorant proponuje również metodykę budowy wzorców klas, które są wyznaczane jako regresory zbioru przykładów treningowych klasy. Metoda zaprezentowana w pierwszej części rozdziału 4 jest następnie uogólniona przez Doktoranta na bardziej realistyczny i zarazem trudniejszy przypadek, w którym liczba punktów styku dla różnych rejestracji w zbiorze treningowym jest zmienna, a ich ‘pochodzenie’ podobnie jak poprzednio - nieznanne. W tym miejscu należy zauważyć, że Doktorant nie ułatwia śledzenia swoich wywodów, bo definicję wielkości E , kluczowej dla zrozumienia treści części 4.3.3, zamieszcza dopiero w części 4.3.4. Mimo, że przyjęta przez Doktoranta metoda ma charakter ekstensywny, liczba rozważanych punktów styku jest na tyle mała, że nie stanowi to istotnej bariery obliczeniowej.

Podsumowując ocenę Rozdział 4.3, uważam go zdecydowanie za najbardziej wartościową część pracy, w której Doktorant przedstawia metodę rozwiązania złożonego problemu dopasowania do siebie niekompletnych podzbiorów punktów, pochodzących z tego samego lub z różnych rozkładów i poddawanych rotacjom oraz translacjom z dodatkowym komponentem szumu. Tak sformułowany problem realistycznie odzwierciedla uwarunkowania biometrycznej analizy rozkładów punktów styku ucha z powierzchnią sensora. Metodyka rozwiązania problemu zaproponowana przez

Doktoranta jest poprawna (choć oczywiście, nie jedyna) i oryginalna, a z uwagi na niewielkie liczności rozważanych zbiorów punktów, całkowicie akceptowalna z punktu widzenia złożoności obliczeniowej. Dodatkowo, przydatne z praktycznego punktu widzenia i wykraczające poza standardową ewaluację metody, jest przeprowadzenie analizy poprawności weryfikacji w zależności od liczby posiadanych punktów styku używanych w budowie wzorców.

4.3. Dwumodalna weryfikacja użytkownika

Przedstawione w części 5 pracy dwumodalne podejście do rozpoznawania ucha ma wymiar nowości jedynie z punktu widzenia rodzaju wykorzystywanych źródeł informacji – uwzględnienie obok obrazu, wiedzy o punktach styczności ucha z ekranem dotykowym, wprowadza oryginalny, nowy komponent do arsenału rozważanych dotychczas metod. Na tym jednak oryginalność podejścia się kończy – przyjęta metodyka weryfikacji bazującej na fuzji informacji dokonywanej na etapie podejmowania decyzji jest standardowa, a sposoby dokonywania indywidualnych analiz danych pochodzących z obydwu źródeł w niczym nie wykraczają poza przedstawione wcześniej w literaturze podejścia. Z pewnością na uwagę zasługują raportowane przez Doktoranta efekty zastosowania proponowanego podejścia, stanowiąc potwierdzenie przydatności pomysłu wykorzystania obszarów kontaktu ucha z sensorem dla analizy biometrycznej. Niedosyt pozostawia zaproponowana przez Doktoranta arbitralna metodyka fuzji wyników częściowych i brak próby uwzględnienia narzucających się dla stosowanej strategii mechanizmów dywersyfikacji znaczenia informacji pochodzącej z obydwu modalności. Z punktu widzenia badawczego, trudne do akceptacji jest przyjęte z góry założenie, że wynik weryfikacji na podstawie obrazu jest ważniejszy od wyniku uzyskanego dla analizy punktów styku ucha. Oczekiwanym podejściem, które niestety nie zostało podjęte, jest łatwa do realizacji próba oceny jakości danych obrazowych (np. przez prostą ocenę kontrastu lub rozmycia) jako przesłanki do określenia znaczenia tych danych w dokonywanej na etapie podejmowania decyzji syntezie.

4.4. Ochrona przed atakami prezentacji z wykorzystaniem dwóch źródeł informacji

Ostatnia część pracy jest poświęcona problemowi dwumodalnej detekcji ataków prezentacji wykorzystującej osobne analizy danych pochodzących z rozważanych źródeł i fuzję uzyskanych wyników analizy. W przeciwieństwie do metodyki przyjętej dla oceny efektów dwumodalnego rozpoznawania, tym razem Doktorant wzbogaca swoje prace używając dla PAD w odniesieniu do obrazów ucha nowego algorytmu klasyfikacji: transformera (a nie jak, w przypadku PAD dla samych obrazów – sieci konwolucyjnej). Metodą stosowaną w odniesieniu do detekcji ataków prezentacji na podstawie rozmieszczenia punktów styku ucha staje się metoda XGBoost. Ponieważ obydwie metody klasyfikacji są od siebie całkowicie niezależne aż do momentu agregacji wyników końcowych, rodzi się pytanie o zasadność umieszczenia rozważań dotyczących metod w części 6, a nie w częściach 3 i 4, poświęconych zagadnieniom PAD dla każdej z modalności z osobna. Przedstawione w rozdziale wyjaśnienia pozostawiają niedosyt: opis koncepcji transformera DeIT, przedstawiony w p. 6.3.1 jest chaotyczny i zawiera irytujące powtórzenia tych samych treści. W części dotyczącej ewaluacji, Doktorant nie zdradza szczegółów procesu dostrajania transformera (jakie moduły i ile parametrów zostały użyte w douczaniu?).

Rozważenie przez Doktoranta scenariusza z podwójną ścieżką analizy oryginalności danych jest uzasadnione, a uzyskiwana poprawa skuteczności obrony przed takimi atakami stanowi

potwierdzenie przydatności hipotezy sformułowanej przez Doktoranta. Na uznanie zasługuje ostrożność Doktoranta w formułowaniu wniosków wypływających z jego eksperymentów – zdaje sobie sprawę ze szczupłości posiadanego przez siebie zbioru danych, więc przyjmuje prawdopodobne pogorszenie efektów rozpoznawania w konfrontacji z próbkami z innych źródeł. Jednakże z punktu widzenia naukowego, poza samym pomysłem dodania jako źródła informacji o potencjalnym ataku danych z sensora dotyku, pozostała część metodyki nie wykracza poza utarte standardy dziedziny.

5. Uwagi szczegółowe

W pracy znajdują się niejasności wymagające doprecyzowania i usterki o różnym charakterze, które przedstawiam w formie poniższej, być może niekompletnej listy:

22. Nie rozumiem zdania: „*As previously said, we have classified PAD approaches into models inside faces, which include ear-touch, and ear photo biometrics*”

27: Zdanie „*The dataset is specifically designed for research in ear verification and PAD, offering a balanced representation of genders ...*” nie ma pokrycia w przytoczonych danych, z uwagi na proporcje 4:1 między przedstawicielami płci.

29: Niejasne jest znaczenie jednostki ‘unit’ (nie została ona wyjaśniona) stosowanej w opisie statystyk bazy obrazów uszu, zamieszczonych w Tabeli 2-3. Czy przedstawione informacje dotyczą liczby zdjęć przypadających na jednego użytkownika w przetworzonym zbiorze danych?

str. 30: Nie rozumiem zdania: „*The dataset contains a relatively distribution of photos across subjects, which can be useful for analyzing trends and patterns in the data.*”

str. 31: W zdaniu: „*The dimensions of the preprocessed images remained the same as their raw counterparts.*” Autor ma chyba na myśli proporcje (aspect ratio).

str. 31: Jaki ‘model’ został użyty do ekstrakcji wektorów cech, które są przedstawione na rys. 2-5?

35: Nie bardzo widzę potrzebę istnienia tabeli 2-5 – nie ma w niej niczego istotnego ponad informacje podane w tabeli 2-4.

38 Zdanie: „*The average distances for printed photos are considered as genuine images (~15cm)*” miało pewnie informować o tym, że odległość zdjęcia była podobna do odległości autentycznie autoryzowanej osoby.

40: Nie rozumiem fragmentu opisującego zawartość komponentu zawierającego punkty styczności ucha z ekranem: „*there were 92 subjects – 72 men and 20 women. If two ears (left and right ear) were taken for each subject, we would have a total subject of 92.*”

41: Wyjątkowo niezręczne sformułowanie myśli: „*According to our data collection experience, we could record ear-touches with fewer touch points (less than four touch points); hence, we captured ear-touches with at least four touch points.*”

str.43: Wyjątkowo niezręczne sformułowanie myśli: „*To make fake data four hands (Two a man hands and two a woman hands) were used.*”

str.55: Wbrew temu co pisze Autor: „*To train a Siamese network using VGG16 for ear photo recognition, we follow a systematic approach ... used for MobileNetV2, but with certain key differences due to the specific characteristics of the VGG16 architecture.*” – w przedstawionej procedurze nie ma żadnych istotnych różnic.

60: „*Figure 3-5: The flowchart of ear photo PAD using a deep neural network structure*” - nie bardzo wiadomo, co dokładnie przedstawia schemat blokowy: argumentem oceny są jednocześnie wektory cech i augmentowane obrazy?

61: „*The layers highlighted in orange represent the added top layers.*” – nie widzę na rysunku niczego wyróżnionego kolorem pomarańczowym.

str.80: „*... then D1 can be reformulated as D2 in Eq.6 to ...*” - złe odwołanie, równanie 6 nie zawiera wspomnianych symboli.

str.80: Kolejne błędne odwołanie - „*The definition of E is presented in Eq.7...*” wyrażenie 7 nie ma związku z definicją E (definicja pojawia się w równaniu 9).

str.81: Dla pracy naukowej w obszarze informatyki, dość amatorsko wygląda szacowanie złożoności algorytmu opisywane w sposób prezentowany przez Doktoranta, np. „*... exceeds 10, the computational complexity becomes*

prohibitive due to the factorial growth in the number of permutations ... This exponential increase in permutations would make the algorithm computationally infeasible, ...” Analiza złożoności powinna być precyzyjna, czas obliczeń – podany jawnie (liczba 10 na pewno nie jest granicą, powyżej której obliczenia nie są możliwe, więc lepiej było określić próg złożoności przyjętej przez Doktoranta).

str.84: Miara FMR to to samo co FAR, po co więc definicja FAR?

str.85: „*We used single enrollment and multi enrollment ...*” - Czy single-enrollment to budowa wzorca dla pojedynczej rejestracji?

99. Co to jest „*VGG-16 SiaNet fusion model*”. Nie wiadomo, czy jest to jakaś osobna sieć, czy jest to po prostu wyznaczenie ważonej sumy wyników generowanych przez sieć syjamską dla obrazu i przez moduł oceny podobieństwa punktów styku.

103: „*The quality of ... ear spoofing attacks might be because of ...*” - zgubiony wyraz?

136 (Appendix B) Autor powielił błąd przedstawiony w swojej publikacji – zamiast ‘best match of X w.r.t T’, mamy ‘best match of X w.r.t X’

5. Wniosek końcowy

Podsumowując opinię o przedłożonej do recenzji rozprawie doktorskiej, chciałbym stwierdzić, że spełnia ona warunki określone dla rozpraw doktorskich przez odpowiednie przepisy. W konsekwencji tej konkluzji, **wnioskuję o dopuszczenie Autora rozprawy do dalszych etapów przewodu doktorskiego.**

REVIEW OF THE DOCTORAL DISSERTATION

Warsaw University of Technology, Warsaw, Poland

Academic Discipline: Technical Computer Science and Telecommunications
(PL: Informatyka Techniczna i Telekomunikacja)

Candidate: Jalil Nourmohammadi Khiarak
Thesis Title: *Countermeasure Algorithms Against Subterfuge
In Mobile Biometric Systems*
Supervisor: Prof. Andrzej Pacut, Warsaw University of Technology
Reviewer: Adam Czajka, Ph.D., Dr. Habil.
Department of Computer Science and Engineering
University of Notre Dame, IN 46556, USA

1. Introduction and Relevance

This dissertation focuses on biometric Presentation Attack Detection (PAD) methods for ear biometrics, specifically examining two Presentation Attack Instruments (PAI): physical ear imitations and ear photographs. Despite ear recognition being a relatively established biometric technology, ear PAD has received limited research attention to date. However, due to the ubiquity of mobile phones, which involve physical contact with the ear, as well as the recent popularity of wearable devices capable of measuring ear features (such as virtual reality headsets or headphones), this technique has the potential to facilitate continuous user authentication. Furthermore, it can complement more common modalities, such as face or iris recognition, in scenarios where such mobile devices or headsets are used. This creates a clear need for appropriate security measures to equip the technology with anti-spoofing countermeasures, including PAD techniques. Consequently, this dissertation is relevant to the biometric research community and addresses current industry needs in a timely manner.

2. General Characteristics and Editorial Quality

The thesis is written in clear, high-quality English, with a small number of grammatical errors and a few missing references (“Error! Reference source not found”). However, these do not significantly hinder the comprehension of the content.

The author guides the reader through the entire development process of a new biometric and PAD techniques. This journey spans from the initial problem formulation (Chapter 1) to the introduction of newly collected photo (Chapter 2) and touch-based (Chapter 4) ear datasets. The author then estimates the utility of the collected data in ear recognition and PAD (Chapter 4) and proposes a fusion strategy for photo- and touch-based recognition (Chapter 5) and PAD (Chapter 6) approaches. The thesis concludes with a summary of key findings, research contributions, and the implications of the work (Chapter 7). In my opinion, this structure and content demonstrate research maturity at the level expected of a PhD candidate.

I like that abbreviations were introduced once at the beginning and used consistently throughout the thesis. The author also added several appendices, including biometric vocabulary, error metrics, and testing protocols (Appendix A), selected pseudo codes

(Appendix B), the author’s list of publications and conference presentations (Appendix C), and a list of projects, in which the candidate participated (Appendix D).

3. Substantive Evaluation of the Dissertation

3.1 Research Objectives and Hypotheses

The thesis does not explicitly state research hypotheses. Instead, the author provides the following four “statements” around which the thesis is structured. These (if I understand their role correctly) serve as the functional research hypotheses:

- (S1) “We introduced and investigated on an ear-photo presentation attacks detection and recognition system on mobile device.”
- (S2) “Ear-touch characteristic is an effective biometric, which is based on mobile devices with multiple touchscreens.”
- (S3) “Ear photo and ear-touch multimodal biometrics improve ear recognition system accuracy in comparison with its single model.”
- (S4) “Ear photo and ear-touch multimodal biometrics improve ear PAD accuracy concerning for to a single model.”

The overarching goal of this dissertation is to design, implement, and evaluate a new biometric technique based on ear characteristics. These characteristics are acquired via both touch sensors embedded in modern mobile phone screens and standard cameras, alongside appropriate PAD techniques for both acquisition methods.

3.2 Original Research Contributions

I identify the following two aspects of the thesis as the most interesting and original research contributions:

a) Collection of a new dataset (WUT-Ear V1.0): The acquisition of ear images is a relatively simple biometric collection effort, requiring no specialized equipment and avoiding many common privacy concerns. To benefit from this advantage, the candidate proactively gathered firsthand experience with biometric data collection and developed two original datasets. The collected data includes photographs of authentic ears (acquired from 137 subjects) and of three PAIs: ears displayed on monitors, 2D paper printouts, and synthetic 3D images displayed on a 3D monitor. Given my concerns regarding the limited availability of the collected dataset to the biometrics community (see my comments in Section 3.3), I would encourage the author to develop creative ways of sharing synthetic data derived from the original data, specifically in a way that still adheres to privacy regulations.

b) Demonstration of viability of touch-based ear recognition: The primary research contribution of this thesis is the design and evaluation of a touch-based ear recognition approach, which addresses typical real-world complications, such as missing or deformed ear features in the same-identity ear images. Furthermore, the author demonstrated the practical advantages of fusing this proposed method with conventional image-

and deep learning-based approaches for both ear recognition and PAD. This resulted in the construction of a fully functional, multimodal biometric system equipped with attack detection measures.

According to Appendix C, Mr. Khiarak is the co-author of nine peer-reviewed publications. The above two listed contributions have been published in two relatively high-quality IEEE conferences related to security (ICCST) and biometrics (ICB). I noted two papers published in *IEEE Transactions on Information Forensics and Security*, however, they cover topics not included in this dissertation.

3.3 Critical Remarks and Questions

In every doctoral dissertation there are inevitably research and editorial aspects that could be improved, if the PhD candidate had more time or resources. This dissertation is no exception. However, the shortcomings listed below do not significantly diminish the value of this work. They are provided as suggestions for extensions of this research, if the candidate plans future publications based on this work.

a) Acquisition methodology: I wonder how the five touch points were selected and why the entire touch map was not used to investigate optimal locations for collecting ear features. Additionally, while Fig. 2-11 illustrates the distribution of touch points (ranging from four to eight), it is unclear how more than five points were collected if the application was designed to collect data from only five locations (as shown in Fig. 2-9).

Also, if I understand Section 2.6.1 correctly, only two subjects (one male and one female) imitated the touch patterns of all other subjects in the dataset. The quality and success of these attacks are, therefore, heavily dependent on how these two specific individuals mimicked the ear structure with their hands, which is an inherently challenging task. It would be valuable to hear the author’s commentary on how the individual skill levels of these two subjects might impact the results. Moreover, it is worth discussing what conclusions can be drawn regarding the generalization of such attacks to a larger, more diverse population.

b) Selection of tools: At times, the selection and application of specific tools in the thesis are questionable. For instance, applying dimensionality reduction techniques to raw images is less effective than visualizing feature representations (either hand-crafted, such as those using Iannarelli’s ear features, or deep learning-based embeddings). While Fig. 2-5 mentions that “feature vectors ... are extracted using a feature extractor model,” the specific model is not provided. Furthermore, t-SNE is a stochastic, non-linear projection method that is highly sensitive to hyperparameters and produces different results in each run. Consequently, drawing conclusions from a single t-SNE projection is methodologically unsound. For example, the statement that “t-SNE can more effectively capture the underlying structure” is misleading. I encourage the author to review the following article regarding the potential pitfalls of stochastic visualization methods such as t-SNE: <https://distill.pub/2016/misread-tsne/>.

Another example is the selection of metrics. The Half Total Error Rate (HTER) has been deprecated by ISO/IEC JTC 1/SC 37, as it combines two metrics represented by random variables with different distributions.

c) Arbitrary choices of hyperparameters: The author does not comment on the selection of various hyperparameters. One example is the input image resolution (224×224 pixels), which was chosen to “match MobileNetV2 input requirements.” It is unclear why the author felt constrained by the architectural defaults of other researchers (a fun fact is that the 224×224 px resolution was originally picked by Alex Krizhevsky primarily due to GPU memory constraints during the design of AlexNet more than a decade ago).

Another example is the choice of five locations for measuring ear pressure on the mobile screen (*e.g.*, why not four or six?). In my opinion, a mature doctoral dissertation should avoid arbitrary choices. The candidate should provide justifications, backed by experiments or theoretical considerations, for all configurations, or at least discuss the potential impact of these selected values on the generalization capabilities of the proposed methods.

d) Evaluation methodology: The author does not provide sufficient detail regarding data splitting, beyond the percentages allocated to the training, validation, and test subsets. Specifically, in the context of PAD methodology and evaluation, these subsets must be subject-disjoint to prevent models from learning subject-specific rather than PAD-specific (and thus subject-agnostic) features.

Another concerning point is the exclusive use of point error estimators. To formally compare methods and their variants, appropriate statistical hypotheses should be defined and rigorously tested, incorporating uncertainty assessments and cross-validation techniques. There are multiple sources of uncertainty in this work. One is the stochastic nature of training deep learning models. To account for this, it would be appropriate to train and test multiple models, providing a set of point estimates that are then aggregated into summary statistics (*e.g.*, mean and standard deviation, assuming the normality of the underlying distributions). Another source of uncertainty relates to inference and the stochasticity of test data. To assess this, the test data could be resampled (with replacement) to generate a distribution of error estimates. It is surprising that the author not only omitted these calculations but also drew definitive conclusions based solely on point estimates, which may be considered a methodological flaw.

e) Reproducibility: According to the ethical statements in Section 2.7, specifically that “there were no plans to share data outside the European Union,” it appears that the collected data will not be made available to non-EU institutions. However, I could not find details regarding whether (or how) the data will be shared within the EU research ecosystem. While I understand the author must adhere to GDPR and other EU data privacy regulations, I would like the author to suggest or outline planned actions to support the claim in Section 2.9 that “the WUT-Ear V1.0 database is a valuable resource for the biometric community.” One idea could be training a modern generative model (such as StyleGAN- or a Stable Diffusion-based approaches) to synthesize synthetic ear images that mimic the original data without “leaking” identity information from the training set into the generated corpus.

I also note that no source code has been provided with the thesis. It is now standard practice in the computer science community, particularly in fields utilizing deep learning, to share code and model weights to ensure reproducibility and facilitate benchmarking of future algorithms. Perhaps the author intends to release the codes in conjunction with future publications derived from this work. While Appendix B includes pseudo codes,

most of the symbols and variables therein are not defined, making them insufficient for reproducing the actual implementations.

f) Practical applicability: The proposed biometric modality requires touch surfaces incorporated into the display glass of smartphones. It is not clear whether standard capacitive touch sensors in contemporary smartphones are suitable for the implementation of the proposed method, or if a specialized hardware design is required. I wonder if the author has any thoughts on the practical feasibility and market viability of this new biometric trait within the competitive consumer landscape.

4. Fulfillment of Statutory Requirements

The dissertation has been assessed in light of the Art. 187(1)-(2) of the Act of 20 July 2018 – The Law on Higher Education and Science.

1. Does it present an original solution to a scientific problem? **[Yes]**
2. Does it demonstrate general theoretical knowledge? **[Yes]**
3. Does it demonstrate the ability to conduct independent research? **[Yes]**

5. Final Conclusion and Motion

Based on the analysis above, I conclude that the doctoral dissertation of Jalil Nourmohammadi Khiarak meets the requirements for the PhD degree in the discipline of Technical Computer Science and Telecommunications, and **I recommend that the candidate be admitted to a public defense.**

Notre Dame, March 12, 2026
(city, date and reviewer's signature)

