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)