Katarzyna Roszczewska*
Ewa Niewiadomska-Szynkiewicz**

Ataki prezentacyjne na systemy rozpoznawania tożsamości wykorzystujące biometrię twarzy oraz metody wykrywania

Streszczenie

Biometria jako technika pomiarów istot żywych skupia się na automatycznym rozpoznawaniu jednostek na podstawie ich cech fizycznych. Jedną z najczęściej stosowanych metod biometrycznego uwierzytelniania jest biometria twarzy. Metoda ta jest dość powszechnie stosowana w smartfonach, paszportach oraz innych urządzeniach i systemach służących do weryfikacji tożsamości lub wymagających uwierzytelnienia. Wraz z rosnącą popularnością biometrii twarzy pojawiają się również obawy dotyczące bezpieczeństwa, szczególnie związane z atakami prezentacyjnymi. Celem tych ataków jest oszukanie systemów biometrycznych. Wykorzystuje się do tego celu różne materiały i dostępne dane, m.in.: wydrukowane zdjęcia, nagrania wideo i maski. Wykrywanie tego typu ataków oraz ochrona wymagają stosowania różnego typu środków bezpieczeństwa oraz zaawansowanych algorytmów detekcji. W artykule są omawiane różne metody wykrywania ataków prezentacyjnych, w tym podejścia wykorzystujące interakcję użytkownika z systemem, analizę właściwości obrazu oraz metody sztucznej inteligencji. Szczególna uwaga jest zwrócona na jakość zbiorów danych uczących wykorzystywanych do trenowania algorytmów, w tym ich zrównoważenie i zróżnicowanie zawartych w nich danych, a także na

* Mgr inż. Katarzyna Roszczewska, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska, e-mail: katarzyna.roszczewska.dokt@pw.edu.pl, ORCID: 0000-0003-2685-455X.
** Prof. dr hab. inż. Ewa Niewiadomska-Szynkiewicz, Instytut Automatyki i Informatyki Stosowanej, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska, e-mail: ewa.szynkiewicz@pw.edu.pl, ORCID: 0000-0003-4782-3816.

konieczność stałego rozwijania mechanizmów bezpieczeństwa w celu ochrony systemów uwierzytelniania biometrycznego przed ewoluującymi zagrożeniami.

Słowa kluczowe: biometria twarzy, detekcja ataków prezentacyjnych, bezpieczeństwo systemów rozpoznawania tożsamości

Wstęp

Zgodnie z definicją biometria to technika dokonywania pomiarów istot żywych. W najnowszych zastosowaniach metody biometryczne są wykorzystywane do automatycznego rozpoznawania ludzi na podstawie cech fizycznych takich, jak: odcisk palca, tęczówka oka, twarz, podpis odręczny oraz głos. Systemy rozpoznawania tożsamości wykorzystujące biometrię stały się nieodłącznym elementem naszego codziennego życia. Obecnie jest to standardowy sposób na potwierdzenie tożsamości podczas odblokowania telefonu, a dzięki specjalnym bramkom na lotnisku odczytującym dane biometryczne zapisane w paszportach możliwa jest szybka kontrola pasażerów. Według danych Biometrics Research Group Inc[1] w ostatnich latach obserwuje się znaczny wzrost wartości rynku biometrycznego. W 2023 roku jego wartość wynosiła 48,8 mld USD, prognoza na 2028 rok to 86,1 mld USD. Dane zamieszczone w raporcie CISCO[2] pokazują, że 81% telefonów dostępnych na rynku w 2022 roku było wyposażonych w narzędzia do biometrycznego uwierzytelniania tożsamości. Wraz ze wzrostem mocy obliczeniowych oraz dynamicznym rozwojem metod sztucznej inteligencji coraz większą popularność zdobywa biometria twarzy. Weryfikacja tożsamości za pomocą twarzy stanowi oprócz odcisku palca jedną z najpopularniejszych modalności biometrycznych wykorzystywanych w telefonach. Główną zaletą biometrii twarzy jest jej wygoda i intuicyjność. Z drugiej strony, twarz jest modalnością najczęściej upubliczniana w mediach społecznościowych i dobrze widoczną przez postronnego obserwatora. Obecny poziom techniki fotograficznej pozwala na robienie dobrej jakości zdjęć twarzy w przestrzeni publicznej. Zdjęcia są więc powszechnie dostępne. Adresując to zagrożenie, podczas projektowania systemów weryfikacji tożsamości wykorzystujących biometrię twarzy należy pamiętać o konieczności wprowadzenia niezbędnych zabezpieczeń przed zagrożeniem związanym z wykorzystanie wizerunku osoby do nieuprawnionego uwierzytelnienia.

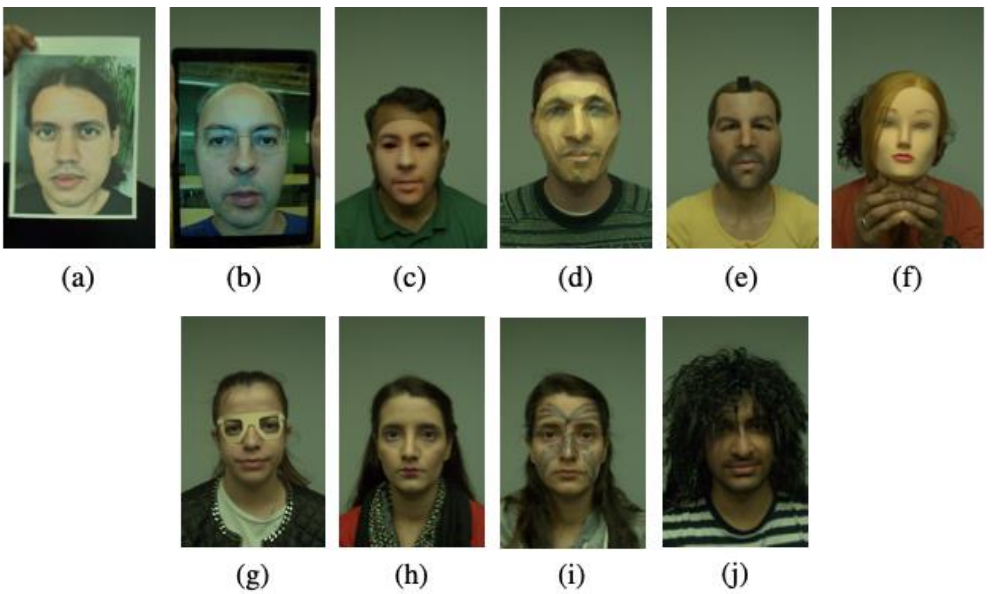
1 *Biometrics Market Reports*, https://www.biometricupdate.com/biometric-news/biometric-research [dostęp: 12.04.2024].

2 *The 2022 Duo Trusted Access Report. Logins in a Dangerous Time*, https://duo.com/resources/ebooks/the-2022-duo-trusted-access-report [dostęp: 29.02.2024].

Ataki prezentacyjne

Ataki prezentacyjne to grupa zagrożeń cybernetycznych dotyczących systemów biometrycznych, w których atakujący podszywa się pod inną osobę. Nie wymagają one od atakującego żadnej ingerencji w system uwierzytelniania. Stanowią typowy przypadek użycia tego systemu. Schemat działania przy próbie złamania zabezpieczeń nie różni się od standardowego korzystania z systemu uwierzytelniania. Atakujący musi jedynie dysponować danymi, które pozwolą na zidentyfikowanie go jako inną osobę (ang. *Presentation Attack Instrument* – PAI).

Wyróżnia się dwa typy ataków prezentacyjnych biometrii twarzy (rys. 1), tj. podszycie się pod inną osobę i uzyskanie nieuprawnionego dostępu do systemu (ang. *impersonification*) oraz ukrycie swojej tożsamości przed systemem weryfikacji (ang. *obfuscation*).



a – zdjęcie, b – nagranie wideo, c – maska, d – maska papierowa, e – maska silikonowa, f – manekin, g – okulary zasłaniające część twarzy, h – makijaż, i – tatuaż, j – peruka

Źródło: Z. Mostaani, A. George, G. Heusch, D. Geissbühler, S. Marcel, *The High-Quality Wide Multi-Channel Attack (HQ-WMCA) database*, https://arxiv.org/pdf/2009.09703.pdf [dostęp: 15.03.2024].

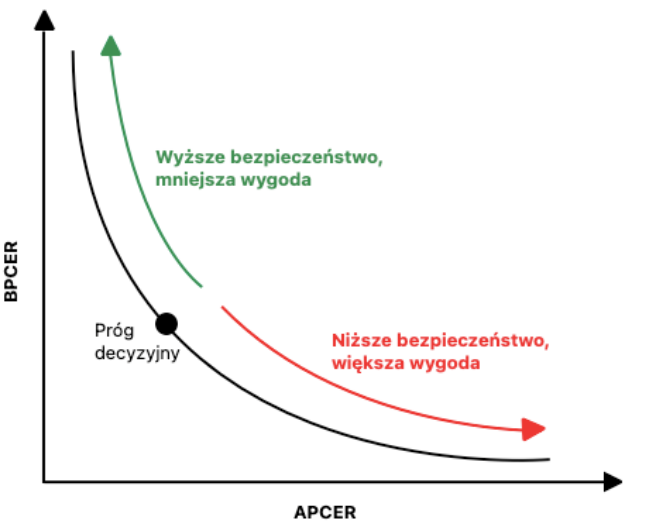
Rys. 1. Przykłady ataków prezentacyjnych

W pierwszym przypadku ataki mogą być przeprowadzone z wykorzystaniem takich materiałów, jak: wydruki, nagrania, maski papierowe, maski silikonowe oraz manekiny. W drugim atakujący wykorzystują materiały takie, jak: tatuaże, peruki czy mocny makijaż, który pozwoli oszukać system biometryczny. Do oceny odporności systemu biometrycznego na ataki prezentacyjne wykorzystywane są następujące metryki, które pokazują skalę błędnych decyzji:

– APCER (ang. *Attack presentation classification error rate*) – procent ataków uznanych za próby poprawnego uwierzytelniania,

– BPCER (ang. *Bona fide presentation classification error rate*) – procent prób poprawnego uwierzytelnienia uznanych błędnie za ataki.

W przypadku obu metryk zakłada się ustalenie pewnych wartości progowych (decyzyjnych), na podstawie których uwierzytelniający się użytkownik jest akceptowany lub odrzucany. Oczywiste jest, że wartości te wpływają na wielkości błędów. Ustawienie wysokiego progu akceptacji może skutkować częstym, niesłusznym blokowaniem użytkownika, zbyt mały próg grozi licznymi atakami. Obie miary są więc ze sobą związane. Pokazuje to rysunek 2. W związku z tym projektowane algorytmy wykrywania ataków prezentacyjnych powinny być dostosowane do planowanego zastosowania. Należy pamiętać, że zawsze będzie konieczne ustalenie pewnego kompromisu między wygodą użytkowania a bezpieczeństwem systemu.



Źródło: opracowanie własne.

Rys. 2. Krzywa Detection Error Tradeoff dla algorytmów wykrywania ataków prezentacyjnych

Detekcja ataków prezentacyjnych

Projektowanie systemu i urządzenia

Detekcja ataków prezentacyjnych jest wykonywana różnymi metodami i może wymagać zastosowania specjalnych sensorów (np. sensorów RGB), algorytmów sztucznej inteligencji lub odpowiedniego zaprojektowania sposobu interakcji użytkownika z systemem. Na etapie projektowania należy zdecydować czy sprawdzanie autentyczności próbki będzie odbywało się przed ekstrakcją cech służących do uwierzytelniania osoby czy jednocześnie z nią. Oba z tych rozwiązań mają swoje wady i zalety. Strategia zakładająca równoległe wykonanie obu operacji pozwala na szybszą identyfikację użytkownika, oczywiście jeżeli próbka okaże się autentyczna. Jeżeli przed ekstrakcją cech sprawdzana jest autentyczność próbki, to unikamy wykonania niepotrzebnych obliczeń w sytuacji, gdy próbka nie jest autentyczna – jest to próba ataku[3].

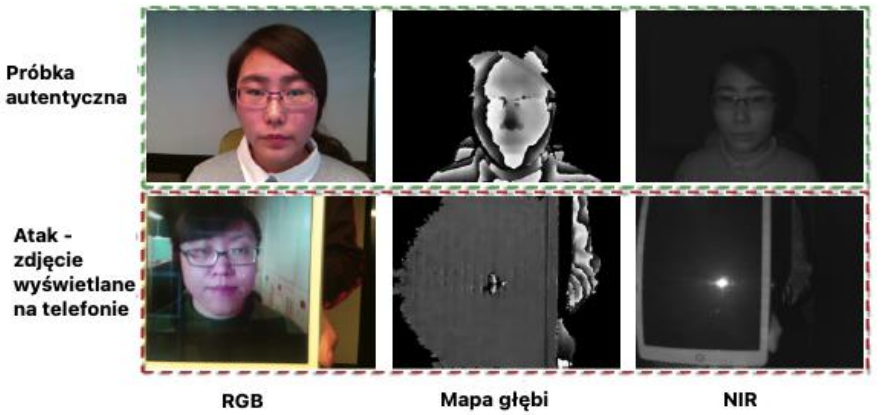
Po wybraniu strategii należy przyjrzeć się źródłom danych. Wykorzystanie wielu różnych jest bardzo pomocne w skutecznym wykrywaniu ataków prezentacyjnych, a identyfikacja tylko na podstawie danych z sensorów RGB wiąże się z ograniczeniami związanymi m.in. z wpływem oświetlenia zewnętrznego na jakość pozyskanych zdjęć.

Żeby pozyskać kompleksowe dane i poprawić skuteczność detekcji ataków prezentacyjnych, można skorzystać z sensorów rejestrujących znacznie więcej niż jedynie obraz w świetlne widzialnym. Przykłady takich sensorów to:

– rejestrujący obraz w bliskiej podczerwieni (ang. *Near Infrared*),
– głębi,
– rejestrujący obraz w podczerwieni krótkofalowej (ang. *Short Wave Infrared*),
– termiczny.

Na rysunku 3 zaprezentowano dwie próbki, tj. fałszywą i autentyczną, zarejestrowane za pomocą różnych sensorów.

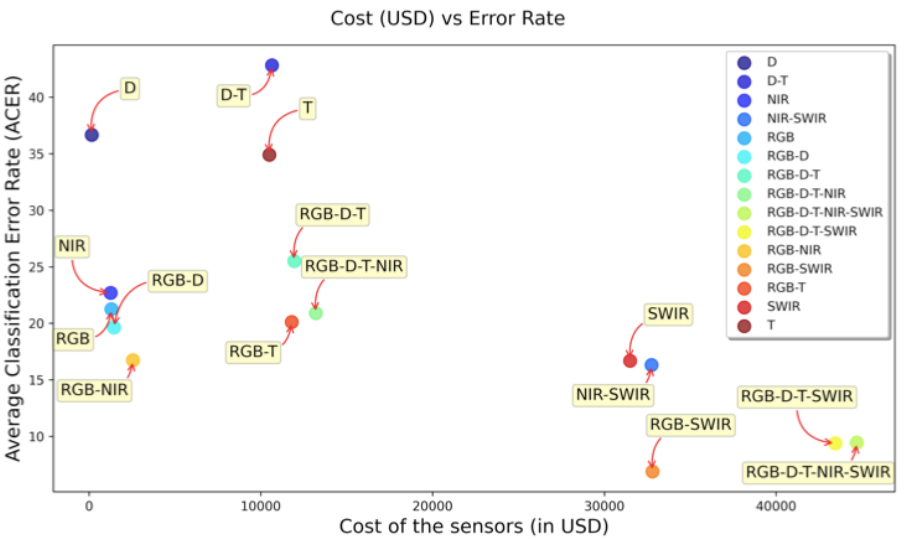
3 Z. Yu, Ch. Zhao, Z. Lei, *Face Presentation Attack Detection*, https://arxiv.org/pdf/2212.03680.pdf [dostęp: 16.01.2024].



Źródło: Z. Yu, Ch. Zhao, Z. Lei, op. cit.

Rys. 3. Próbka autentyczna i fałszywa rejestrowana za pomocą różnych sensorów

Wykorzystanie wielu, odpowiednio dobranych, źródeł danych zwiększa skuteczność systemu uwierzytelniania, ale skutkuje, niestety, wzrostem kosztów urządzeń wykorzystywanych do identyfikacji osoby.



Źródło: A. George, D. Geissbuhler, S. Marcel, *A Comprehensive Evaluation on Multi-channel Biometric Face Presentation Attack Detection*, „Journal of Latex Class Files" 2015, t. 14, nr 8.

Rys. 4. Zależność skuteczności systemów uwierzytelniania od ceny wykorzystanych sensorów (cena w USD)

Metody detekcji ataków prezentacyjnych wykorzystujące analizę obrazu i jego charakterystykę

Część dostępnych systemów dokonuje identyfikacji osoby na podstawie jej interakcji z zastosowanymi urządzeniami. W celu potwierdzenia tożsamości osoba musi wykonać wskazane czynności. Może to być ruch głową, mrugnięcie oczami lub otwarcie ust. Gdy system, na podstawie wykonanych czynności, rozpozna osobę, wówczas zostaje ona uwierzytelniona. W przeciwnym razie zakłada się, że jest to próba ataku. Takie rozwiązanie może być skuteczne podczas ataków prezentacyjnych z wykorzystaniem wydruków, ale już detekcja ataków z wykorzystaniem nagrania wideo lub zastosowaniem masek 3D może być problematyczna. Dotyczy to przede wszystkim sytuacji, w których sekwencje ruchów oczekiwanych od nas przez system są powtarzalne i odtworzenie ich jest stosunkowo proste. Tego typu metody detekcji ataków prezentacyjnych ponadto mogą być czasochłonne i uciążliwe dla użytkownika.

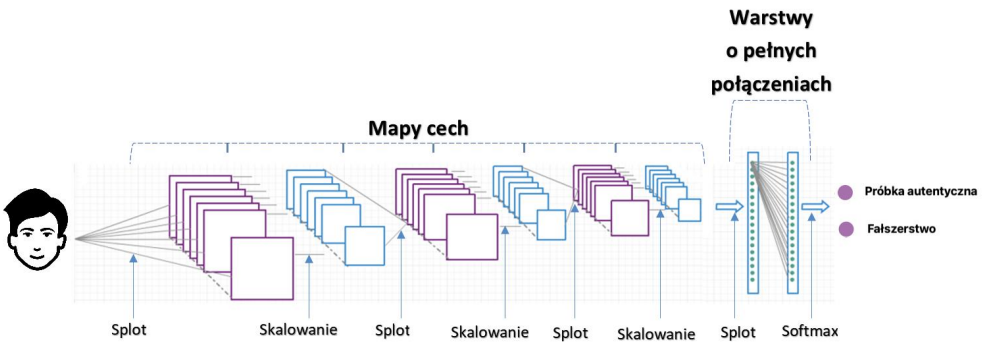
Innym podejściem są algorytmy bazujące na charakterystykach rejestrowanych obrazów. Do wykrywania fałszerstw takich jak wydruki zdjęcia twarzy czy nagrania można wykorzystać algorytmy odtwarzające kształt twarzy w 3D. Jest to, niestety, metoda o stosunkowo dużej złożoności obliczeniowej. Ponadto jest nieskuteczna wówczas, gdy atakujący wykorzystuje maskę lub głowę manekina. Nagrania, słabej jakości wydruki czy maski papierowe charakteryzują się inną reprezentacją kolorów i wyższym zaszumieniem od danych autentycznych. Tego typu właściwości są również wykorzystywane do wykrywania ataków. W tym przypadku algorytmy detekcji są mniej kosztowne obliczeniowo w stosunku do poprzedniego podejścia. Ich wadą jest niestety mała skuteczność detekcji (duży błąd APCER) w przypadku wydruków i masek o wysokiej jakości. Innym rozwiązaniem jest analizowanie tekstury badanego obrazu. Tego typu detektory są skuteczne dla zdjęć dobrej jakości, ale mogą skutkować odrzucaniem autentycznych próbek, które zostały zarejestrowane w słabszej jakości (duży błąd BPCER).

Kolejnym podejściem do wykrywania ataku może być badanie mikro ruchów twarzy. Idea jest podobna do analizowania interakcji użytkownika z systemem uwierzytelniania, ale jest to wykonywane automatycznie i nie wymaga intencjonalnego działania użytkownika. Podejście to, niestety, nie sprawdza się w atakach wykorzystujących nagrania wideo lub poruszające się wydruki.

Ze względu na ograniczenia prezentowanych metod często tworzone są rozwiązania hybrydowe polegające na połączeniu kilku podejść. W ten sposób możliwe jest poprawienie skuteczności działania systemu wykrywania ataków prezentacyjnych.

Głębokie sieci neuronowe do wykrywania ataków prezentacyjnych

W ostatnich latach do klasyfikacji, segmentacji i detekcji obiektów w obrazie dość powszechnie wykorzystuje się sztuczne sieci neuronowe (SSN). Stosujące je algorytmy wykazują się bardzo dużą skutecznością. Modele SSN znajdują też zastosowanie w systemach wykrywania aktów prezentacyjnych[4]. Przykład architektury takiej sieci jest przedstawiony na rysunku 5. Identyfikacja użytkownika, w razie możliwości wystąpienia ataku prezentacyjnego, to typowe zadanie binarnej klasyfikacji. Mamy dwie klasy, tj. autentycznych i fałszywych próbek. Analizowane zdjęcie twarzy jest przypisywane do jednej z nich.



Źródło: opracowanie własne

Rys. 5. Przykładowa architektura SSN do detekcji ataków prezentacyjnych

Wytrenowanie modelu SSN wymaga przygotowania odpowiednio dużego zbioru danych uczących. Różnorodność danych i wielkość tego zbioru mają znaczny wpływ na jakość modelu sieci. Ważne jest zagwarantowanie tego, żeby zawierał on różne typy ataków, przed którymi chcemy zabezpieczyć nasz system oraz jego zbalansowanie – podobna liczba reprezentantów obu klas. Należy również pamiętać, że na jakość danych mogą wpływać warunki, w jakich jest pobierana próbka. Należy starać się uwzględnić różne sytuacje. Próbki wykorzystywane do trenowania modeli SSN powinny być zbierane

4 Y. Ma, L. Wu, Z. Li, F. Liu, *A novel face presentation attack detection scheme based on multi-regional convolutional neural networks*, „Pattern Recognition Letters" 2020, t. 131; A. George, S. Marcel, *Robust Face Presentation Attack Detection with Multi-channel Neural Networks* [w:] *Handbook of Biometric Anti-Spoofing. Advances in Computer Vision and Pattern Recognition*, red. S. Marcel, J. Fierrez, N. Evans, Singapore 2023; A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, S. Marcel, *Biometric Face Presentation Attack Detection With Multi-Channel Convolutional Neural Network* [w:] *IEEE Transactions on Information Forensics and Security*, t. 15, IEEE 2020, s. 42–55.

w różnych warunkach oświetleniowych, z różnych odległości oraz zawierać możliwie dużo modyfikacji, które mogą wystąpić w środowisku produkcyjnym.

Uczenie sieci neuronowych na zbyt małym lub niereprezentatywnym zbiorze danych może prowadzić do zjawiska nadmiernego dopasowania do danych uczących (ang. *overfitting*). W takim wypadku system doskonale radzi sobie z detekcją ataków prezentacyjnych, ale tylko na danych, które były wykorzystane w procesie trenowania sieci. W warunkach produkcyjnych, gdy na wejściu systemu pojawią się wcześniej nieznane dane, wówczas jego skuteczność znacząco spada. Podobna sytuacja może wystąpić wtedy, kiedy sieć neuronowa o złożonej architekturze jest uczona na niewystarczająco licznym zbiorze danych. W takiej sytuacji wskazane jest wykorzystanie modeli, które zostały wstępnie wyuczone na bardzo dużych zbiorach danych (ang. *transfer learning*). Tego typu modele są udostępniane w sieci internet. Mogą być one następnie douczane na danych dotyczących rozwiązywanego problemu.

Typowym przykładem pokazującym jak ważne jest, żeby warunki podczas zbierania danych uczących były w miarę możliwości zgodne z przypadkami użycia systemu w środowisku produkcyjnym, są problemy z poprawnym odwzorowaniem rzeczywistości dla modelu SSN, gdy dane są zbierane tylko w laboratorium. Akwizycja przy świetle sztucznym, w sytuacji, gdy odbywa się ona przy założeniu jednego kąta padania promieni świetlnych i tworzenia się refleksów może spowodować, że model SSN nauczy się, że jedna ze stron twarzy jest jaśniejsza od drugiej. Wpłynie to negatywnie na dokładność przyszłej klasyfikacji, a tym samym skuteczność detekcji ataków. Podsumowując, w przypadku modeli bazujących na uczeniu jakość i liczność zbioru trenującego są ważne do uzyskania oczekiwanego rezultatu działania stosujących je detektorów. Pozyskanie i odpowiednie przygotowanie zbiorów danych uczących jest często głównym problemem w budowaniu modeli SSN i wymaga dużego nakładu pracy.

Zakończenie

W ostatnich latach można zaobserwować, że biometryczne uwierzytelnianie jest coraz częściej stosowane w wielu dziedzinach życia codziennego. Wygoda i bezpieczeństwo zastosowania odcisku palca czy zdjęcia twarzy do odblokowania telefonu komórkowego czy smartfonu sprawiło, że metody te stały się standardowym wyposażeniem niemal wszystkich produkowanych obecnie urządzeń mobilnych. Jednakże nie należy zapominać, że wraz z rozwojem

i ekspansją systemów wykorzystujących biometrię do rozpoznawania tożsamości rośnie również liczba i zaawansowanie potencjalnych ataków cybernetycznych. W związku z tym w pełni uzasadnione, a wręcz konieczne jest prowadzenie prac badawczych oraz wytwarzanie nowoczesnych, nadążających za pojawiającymi się nowymi wyzwaniami metod i mechanizmów do zabezpieczania tych urządzeń i działającego na nim oprogramowania przed różnego rodzaju atakami, w tym przed atakami prezentacyjnymi. Obecnie jest oferowanych wiele skutecznych metod detekcji ataków prezentacyjnych. Użytkownikom telefonów komórkowych, w których do uwierzytelniania tożsamości wykorzystuje się biometrie twarzy, zaleca się ich rozpoznanie, analizę ich zalet i wad oraz wybranie odpowiedniego narzędzia, jego zainstalowanie w telefonie i dostosowanie do własnych potrzeb.

Bibliografia

Biometrics Market Reports, https://www.biometricupdate.com/biometric-news/biometric-research [dostęp: 12.04.2024.

George A., Geissbuhler D., Marcel S., *A Comprehensive Evaluation on Multi-channel Biometric Face Presentation Attack Detection*, „Journal of Latex Class Files" 2015, t. 14, nr 8.

George A., Marcel S., *Robust Face Presentation Attack Detection with Multi-channel Neural Networks* [w:] *Handbook of Biometric Anti-Spoofing. Advances in Computer Vision and Pattern Recognition*, red. S. Marcel, J. Fierrez, N. Evans, Singapore 2023.

George A., Mostaani Z., Geissenbuhler D., Nikisins O., Anjos A., Marcel S., *Biometric Face Presentation Attack Detection With Multi-Channel Convolutional Neural Network* [w:] *IEEE Transactions on Information Forensics and Security*, t. 15, IEEE, 2020.

Ma Y., Wu L., Li Z., Liu F., *A novel face presentation attack detection scheme based on multi-regional convolutional neural networks*, „Pattern Recognition Letters" 2020, t. 131.

Mostaani Z., George A., Heusch G., Geissbühler D., Marcel S., *The High-Quality Wide Multi-Channel Attack (HQ-WMCA) database*, https://arxiv.org/pdf/2009.09703.pdf [dostęp: 15.03.2024].

The 2022 Duo Trusted Access Report. Logins in a Dangerous Time, https://duo.com/resources/ebooks/the-2022-duo-trusted-access-report [dostęp: 29.02.2024].

Yu Z., Zhao Ch., Lei Z., *Face Presentation Attack Detection*, https://arxiv.org/pdf/2212.03680.pdf [dostęp: 16.01.2024].

Presentation Attacks on Identity Recognition Systems Using Facial Biometrics and Detection Methods

Abstract

Biometrics, as a technique for measuring living beings, focuses on automatically recognizing individuals based on their physical characteristics. One of the most common methods of biometric authentication is facial biometrics. This method is widely used in smartphones, passports, and other devices and systems for identity verification or authentication. With the growing popularity of facial biometrics, there are also security concerns, particularly related to presentational attacks. The goal of these attacks is to

fool biometric systems. Various materials and available data are used for this purpose, including printed photos, videos, and masks. Detecting these types of attacks and protecting them requires various types of security measures and advanced detection algorithms. The article discusses selected methods of detecting presentation attacks, including approaches using user-system interaction, analysis of image properties, and artificial intelligence. Special attention is paid to the quality of the learning datasets used to train the algorithms, including their balance and the diversity of the data they contain. Attention is paid to continuously developing security mechanisms to protect biometric authentication systems from evolving threats.

Key words: face biometrics, presentation attack detection, security of biometrics identity recognition systems