КРИМІНОЛОГІЧНІ ДОСЛІДЖЕННЯ:

Д. Д. Журавльов,
Магістрант з інформаційних систем
та технологій НТУУ «КПІ ім. Ігоря Сікорського»

ВИЯВЛЕННЯ СПУФІНГ-АТАК НА СИСТЕМИ БІОМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ ЗА ОБЛИЧЧЯМ

Анотація: В статті розглянуто методи детектування та розпізнавання спуфінг-атак на системи біометричного захисту за обличчям людини, проаналізовано їх якісні показники та проаналізовано підхід з використанням згорткових нейромереж який б дозволив отримати найкращий показник HTER для розробленої власної системи захисту. Отриманий результат дозволив виділити переваги та недоліки при проектуванні системи виявлення атак у розглянутій області застосування. Сплановано алгоритм роботи системи для виявлення спуфінг– атак на базі згорткової нейромережі.

Ключові слова: анти-спуфінг, ідентифікація обличчя, згорткові нейромережі, ключові точки і ознаки, біометрична автентифікація та ідентифікація.

Вступ

Біометрична ідентифікація людини – доволі давня ідея в сфері інформаційних технологій для розпізнавання людей, яку намагалися технічно реалізувати. Паролі можна вкрасти, скопіювати, забути, ключі

– підробити, а ось унікальні характеристики самої людини підробити та втратити набагато важче. Такими характеристиками можуть бути відбитки пальців, голос, малюнок судин сітківки ока, хода та інше. Сучасні системи розпізнавання обличчя людини демонструють доволі велику точність, з появою великих наборів даних та складних архітектур систем стало мож-

ливим досягти точності розпізнавання обличчя аж до 0,000001 (одна помилка на мільйон), і вони вже зараз придатні для перенесення на мобільні платформи. Але вразливим місцем систем біометричної ідентифікації стала їхня безпека.

Для того, щоб підробити дані та видати себе за іншу людину в нашій технічній реальності, найчастіше використовують маски. Маски бувають абсолютно різної якості, від надрукованого на принтері фото іншої людини, яке тримають перед обличчям, до дуже складних тривимірних масок з підігрівом. Комп'ютерну систему теж намагаються обдурити, представивши перед камерою замість свого обличчя чиєсь ще, на фото або відео, яке зображене на екрані планшета, смартфона.

Неабияку увагу до теми привернула успішна спроба взлому системи «Face ID» на смартфоні “iPhone X”, за допомогою складної маски з кам'яного порошку зі спеціальними вставками навколо очей, що імітують тепло живого обличчя за допомогою інфрачервоного випромінювання [1]. Наявність таких вразливостей дуже небезпечна для банківських або державних систем автентифікації особи користувача по біометричним даним, де проникнення зловмисника спричиняє значні втрати. Відповідно комплекс заходів, щоб протистояти такому обману, називатимемо anti-spoofing, він може бути реалізований у вигляді різних технологій і алгоритмів, що вбудовуються в модуль системи ідентифікації та автентифікації. Саме найбільш ефективну anti-spoofing систему заходів за показником HTER буде визначено у даній роботі.

Аналіз проблеми біометричної ідентифікації за обличчям

Для визначення якості роботи системи часто користуються метрикою HTER (англ. Half-Total Error Rate – половина повної помилки), яку обчислюють у вигляді суми коефіцієнтів помилково дозволених ідентифікацій (англ. FAR – False Acceptance Rate) та помилково заборонених ідентифікацій (англ. FRR – False Rejection Rate), поділеної навпіл:

$$HTER = (FAR + FRR)/2 \qquad (1)$$

Варто зазначити, що в системах біометрії зазвичай найбільшу увагу приділяють FAR показнику, щоб зробити все можливе та не допустити зловмисника в систему і досягають у цьому хороших успіхів. Зворотною стороною виявляється неминуче зростання FRR – кількості простих користувачів, помилково класифікованих як зловмисників, і якщо для державних, оборонних та інших подібних систем, з чутливими даними, цим можна пожертвувати, то мобільні технології, що працюють з їх величезними масштабами, різноманітністю абонентських пристроїв є дуже чутливі до будь-яких факторів, які можуть змусити користувачів відмовитися від послуг, тому на це варто звернути особливу увагу при розробці системи.

Найпопулярнішим засобом спуфінг-атак на системи біометричної ідентифікації за обличчям є маски або як ще називається даний метод Mask attack. Немає нічого більш простого, ніж одягнути маску, яка підроблює зображення обличчя іншої людини, та представити обличчя системі ідентифікації (мал. 1).

Мал. 1. Приклад Mask attack

Ще можна роздрукувати фото себе чи когось ще на аркуші паперу та піднести його до камери, такий тип атаки називають Printed attack(мал. 2). Більш складним є тип атак Replay attack, коли системі пред’являють екран іншого пристрою(планшет, смартфон), на якому відтворюється заздалегідь записане відео з іншою людиною. Складність виконання даного методу ком-

пенсується високою ефективністю такої атаки, оскільки системи захисту часто використовують ознаки, засновані на аналізі часових послідовностей, наприклад, відстеження моргання, мікрорухів голови, наявність міміки, дихання тощо. Все це легко відтворити на відео.

Останні два типи атак(Replay attack та Printed attack) мають ряд характерних ознак, що дозволяють їх виявити, і, таким чином, відрізнити екран планшета або аркуш паперу від реальної особи.

Зведемо характерні ознаки, що дозволяють визначити ці два типи атак у таблицю 1.

Таблиця 1

Printed attack	**Replay attack**
Зниження якості текстури зображення під час друку	Муар
Артефакти передачі напівтонового зображення під час друку на принтері	Відображення (відблиски)
Механічні артефакти друку (горизонтальні лінії)	Плоский малюнок (відсутність глибини)
Відсутність локальних рухів (наприклад, моргання)	Можуть бути видні межі зображення
Можуть бути видні межі зображення	

Аналіз існуючих систем виявлення спуфінг-атак

Перший підхід, який буде доцільно розглянути, заснований на використанні особливості погіршення якості зображення під час друку або відтворення на екрані[2]. На роздрукованих зображеннях або зображеннях на екранах цифрових пристро-

їв будуть виявлені локальні узори(патерни), нехай і невловимі оком, але їх можна виділити, наприклад, порахувавши локальні бінарні патерни (LBP, local binary pattern) для різних зон зображення обличчя людини після виділення з кадру. Вказану систему можна вважати першоджерелом всього напряму алгоритмів face anti-spoofing detection на основі аналізу зображення. Якщо розглядати більш детально даний підхід, то при розрахунку LBP послідовно береться кожен піксель зображення та вісім його сусідів та порівнюється їх яскравість, якщо яскравість більша, ніж на центральному пікселі, то у матрицю, яка відповідає за розміром матриці пікселів зображення, присвоюється одиниця, якщо менше – нуль. Таким чином, для кожного пікселя виходить 8-бітова послідовність. За отриманими послідовностями будується попіксельна гістограма, яка подається на вхід SVM-класифікатора. Показник ефективності HTER для такого підходу становить 15%, і означає, що значна частина зловмисників долає захист без особливих зусиль, хоча слід зазначити, що більша частина відсівається. Система з показником HTER у 15% тестувалася на наборі даних Replay-Attack від IDIAP, що складається з 1200 коротких відео 50 респондентів та трьох видів атак – printed attack, mobile attack, high-definition attack.

Альтернативний підхід для виявлення спуфінг атак був розроблений у 2015 році вченим Букінафітом З. з університету Оулу. Букінафіт розробив алгоритм альтернативного розбиття зображення на канали, крім традиційного RGB, для результатів якого знову підраховувалися локальні бінарні патерни[3], які, як і в попередньому способі, подавалися на вхід SVN класифікатора. Точність HTER, розрахована на датасетах CASIA та Replay-Attack, склала 3%.

Ще одним з існуючих методів виявлення спуфінг-атак є виявлення муару зображень. Пател К. з університету штату Мічіган, США опублікував статтю[4], де запропонував шукати артефакти зображення у вигляді періодичного візерунка, викликані накладанням двох розгорток. Даний підхід виявився ефективним, показавши HTER близько 6% на наборах даних

IDIAP, CASIA та RAFS. Це також була перша спроба порівняти ефективність роботи алгоритму на різних наборах даних.

Щоб виявити спроби атак типу Printed Attack, логічним рішенням було спробувати аналізувати не одне зображення, а їх послідовність, взяту з відео-потоку. Наприклад, Анжос А. із колегами з дослідного інституту «Ідіап» в Швейцарії запропонували виділяти ознаки з оптичного потоку на сусідніх парах кадрів[5], подавати на вхід бінарного класифікатора та усереднювати результати. Підхід виявився досить ефективним, продемонструвавши HTER 1,52% на їхньому власному наборі даних.

Експериментально-дослідне вирішення задачі

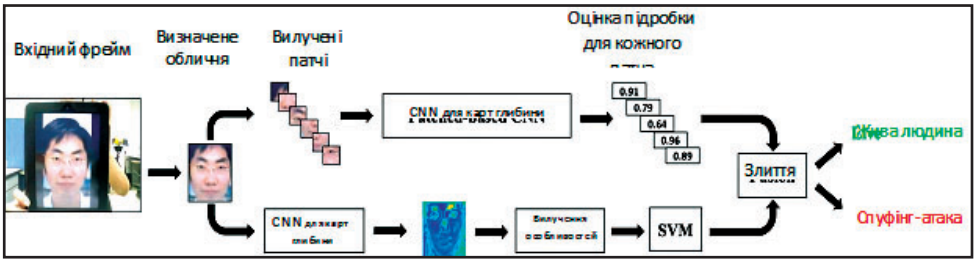
Для того щоб досягти показника HTER, який буде менший, тобто кращий, ніж у розглянутих системах, у рамках даної роботи запропоновано розглянути підхід з використанням згорткових нейромереж або ж CNN-мереж (англ. CNN – convolutional neural network). Даний вибір обумовлений тим, що нейрони всередині шару CNN є з'єднаними лише з невеликою частиною нейронів попереднього шару, що називається рецептивним полем, даний підхід зменшує кількість оперативної пам'яті яку програма використовує для зберігання даних. Поєднання великої кількості таких шарів разом створює нелінійні фільтри, що стають все масштабнішими (тобто, сприймається більша область піксельного простору), так що нейромережа на першому етапі створює представлення дрібних елементів входу, а далі з цих елементів конструює представлення більших областей. В CNN мережі усі фільтри повторюються на всьому зоровому полі зображення. Такі повторні вузли застосовують спільну параметризацію, вектор ваги та упередженості, і утворюють карту ознаки. Тобто всі нейрони у вказаному згортковому шарі реагують на одну й ту ж саму ознаку в межах свого рецептивного поля. При повторюванні вузлів даним способом ознаки зображення мають можливість бути виявленими незалежно від їхнього положення в зоровому полі, і таким чином ми забезпечуємо властивість інваріантності відносно зсуву.

CNN нейромережі досягли значного зниження рівня похибки при їх використані для розпізнавання облич, 97.6% рівень розпізнавання на 5600 нерухомих зображеннях понад 10-ти суб'єктів[6]. CNN використовували для оцінки якості відео об'єктивним чином, після тренування вручну; отримана в результаті система мала дуже низьку кореневу середньоквадратичну похибку[7].

В даній роботі в якості експерименту ми розглядаємо систему саме з використанням згорткової нейронної мережі, яка буде виділяти зображення обличчя людини за допомогою карт глибини зображень. Карта глибини є доволі гарною ознакою для визначення площини, де розташоване зображення. Головна з переваг полягає в тому, що зображення на аркуші паперу, «глибини» немає за визначенням. У роботі Атаума 2017 року із зображення витягувалося безліч окремих невеликих ділянок, для них розраховувалися карти глибини, які потім зливалися з картою глибини основного зображення [8].

У нашій розроблювальній системі пропонується разом зливали результати роботи двох згорткових нейромереж(мал. 2), перша з яких буде розраховувати карти глибини для фреймів(частин зображення), а друга – для зображення в цілому. При навчанні на наборах даних із класом Printed Attack буде пов'язуватися карта глибини, що дорівнює нулю, а з тривимірною моделлю особи – серія ділянок, що випадково відбираються. Карта глибини не вирішує проблему в цілому, від неї використовувалася лише деяка індикаторна функція, що характеризує «глибину ділянки». Планується що система покаже значення HTER < 1%. Для навчання нейромережі будуть використані три публічні набори даних – CASIA-MFSD, MSU-USSA та Replay-Attack.

Мал. 2. Архітектура запропонованого підходу проти спуфінгу.



Висновки

В даній роботі були проаналізовані основні види спуфінг-атак на систему ідентифікації за обличчям людини та існуючі системи для виявлення даних атак, їх недоліки та переваги. В якості власного рішення була запропонована система на базі згорткової нейронної мережі з використанням карти глибини для класифікації особливостей зображень та підхід з аналізом зображення як цілого об’єкта, а також аналізом окремих частин цього зображення, як фреймів. На даному етапі розвитку системи зрозуміло, що для покращення показників, треба буде виконувати злиття кількох методів класифікації. Аналіз ураження, розгляду карти глибини повинні використовуватися разом. Перспективним варіантом покращення системи захисту в біометричних системах може слугувати поєднання інших видів ідентифікації та автентифікації, тобто допоможе у системі додатковий потік даних, наприклад, запис голосу людини та якісь комплексні підходи, які дозволяють вмістити кілька технологій в єдиній системі для виявлення спуфінг-атак на систему ідентифікації за обличчям.

Список використаних джерел

1. Face id in businesstransactions [Електронний ресурс]. – 2017. –
Режим доступу до ресурсу: http://www.bkav.com/d/top-news/-
/view_content/content/103968/bkav%92s-new-mask-beats-face– id-in-twin-way-severity-level-raised-do-not-use-face-id-in– business-transactions.
2. Chingovska I. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing / Chingovska Ivana – Martigny, Suisse, 2017
3. Boulkenafet Z. Face anti-spoofing based on color texture analysis
/ Boulkenafet Zinelabidine – University of Oulu, Finland, 2015.
4. Keyurkumar P. Live Face Video vs. Spoof Face Video / Keyurkumar Patel – Michigan State University, USA, 2020.
5. Anjos A. Motion-based counter-measures to photo attacks in face recognition / Anjos André – Paris, France, 2014.
6. Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Networks: 555–559.
7. Callet, Patrick; Christian Viard-Gaudin; Dominique Barba (2006). A Convolutional Neural Network Approach for Objective Video Quality Assessment. IEEE Transactions on Neural Networks: 1316–1327.
8. Atoum Y. Face Anti-Spoofing Using Patch and Depth-Based CNNs [Електронний ресурс] / Yousef Atoum // Michigan State University, East Lansing. – 2020. – Режим доступу до ресурсу: http://cvlab.cse.msu.edu/pdfs/FaceAntiSpoofingUsingPatchandDe pthBasedCNNs.pdf.