УДК 004.056.5:004.932

В. О. Пальчик
К. В. Безпалий

СТЕК ДЛЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ АВТЕНТИФІКАЦІЇ НА ОСНОВІ РОЗПІЗНАВАННЯ ЖЕСТІВ

Вінницький національний технічний університет

Анотація

Роботу присвячено обґрунтуванню вибору технологічного стека для побудови інтелектуальної системи автентифікації на основі розпізнавання жестів рук. Сформульовано критерії оцінки: робота у реальному часі без графічного прискорювача, кросплатформність, семантично насичене представлення жесту та доведена криптографічна стійкість засобів захисту еталонних шаблонів. Проведено порівняльний аналіз альтернативних рішень за цими критеріями та обґрунтовано вибір комбінації Python + OpenCV + MediaPipe HandLandmarker для модуля розпізнавання, бібліотеки Tkinter для графічного інтерфейсу та схеми Fernet з деривацією ключа через PBKDF2-HMAC-SHA256 для захисту шаблонів. Показано, що обраний стек є компромісним оптимумом між обчислювальною складністю, простотою інтеграції та рівнем безпеки.

Ключові слова: жестова автентифікація, MediaPipe, landmark-представлення, обробка у реальному часі, PBKDF2, Fernet.

Abstract

This paper substantiates the choice of a technological stack for building an intelligent authentication system based on hand gesture recognition. Evaluation criteria are formulated: real-time operation without a graphics accelerator, cross-platform support, semantically rich gesture representation, and proven cryptographic robustness of the reference template protection. A comparative analysis of alternative solutions against these criteria is performed, and the combination Python + OpenCV + MediaPipe HandLandmarker for the recognition module, Tkinter library for the graphical interface, and the Fernet scheme with PBKDF2-HMAC-SHA256 key derivation for template protection is justified. The selected stack is shown to be a compromise optimum among computational complexity, integration simplicity, and security level.

Keywords: gesture authentication, MediaPipe, landmark representation, real-time processing, PBKDF2, Fernet.

Вступ

Ефективність систем жестової автентифікації значною мірою визначається обраним технологічним стеком. Сучасний ринок інструментальних засобів пропонує широкий спектр рішень для комп'ютерного зору — від низькорівневих бібліотек обробки зображень до високорівневих фреймворків глибокого навчання, кожне з яких має власні компроміси між точністю, обчислювальною складністю та вимогами до апаратного забезпечення [1]. У контексті систем автентифікації на типових пристроях користувача, які зазвичай позбавлені графічного прискорювача, вибір технологічного стека стає критичним архітектурним рішенням. Метою роботи є формулювання критеріїв оцінки, порівняльний аналіз альтернативних рішень за цими критеріями та обґрунтування вибору технологічного стека для системи жестової автентифікації.

Результати дослідження

На основі вимог до системи сформульовано чотири ключові критерії оцінки технологічного стека: спроможність роботи у реальному часі (не менше 30 кадр/с на типовому центральному процесорі без графічного прискорювача), кросплатформність (підтримка Windows, Linux, macOS без модифікації коду), семантична насиченість представлення жесту (стійкість до варіацій освітлення та масштабу) та доведена криптографічна стійкість засобів захисту еталонних шаблонів. Додатковими критеріями є відкритість ліцензії, активність спільноти розробників і простота інтеграції компонентів стека між собою.

Як мову реалізації обрано Python. Хоча Python не належить до найшвидших мов за самостійним виконанням коду, його реальною перевагою є те, що він виступає високорівневою оболонкою над оптимізованими бібліотеками на C/C++, забезпечуючи лаконічний інтерфейс до низькорівневих

обчислень [1]. Альтернативні мови поступаються Python за критеріями завдання: C++ ускладнює прототипування та значно подовжує цикл розробки, JavaScript обмежений браузерною екосистемою та має проблеми з доступом до локальних ресурсів камери, а Java/Kotlin потребують додаткових обгорток для бібліотек комп'ютерного зору. Зрілість екосистеми Python, наявність більшості сучасних бібліотек комп'ютерного зору та машинного навчання саме у вигляді Python-біндингів, а також кросплатформний характер інтерпретатора роблять Python природним вибором для задач даного класу.

Для захоплення та попередньої обробки відеоданих обрано бібліотеку OpenCV, яка є фактичним галузевим стандартом у задачах комп'ютерного зору у реальному часі. Її реалізація на C/C++ із використанням SIMD-інструкцій забезпечує обробку кадру за кілька мілісекунд навіть на стандартних процесорах. Модульна архітектура OpenCV дозволяє підключати лише необхідні компоненти, що знижує обчислювальне навантаження. Хоча у складі OpenCV наявний модуль DNN для запуску попередньо навчених нейромереж, його можливості поступаються спеціалізованим фреймворкам, тому у даній роботі OpenCV використовується виключно як засіб захоплення відеопотоку та базової обробки кадрів, а задача детекції ключових точок винесена у спеціалізований фреймворк [1, 2].

Для виділення ключових точок кисті проведено порівняння альтернативних рішень. OpenPose, попри високу точність скелетного моделювання, орієнтований на потужні системи з графічним прискорювачем (5–10 кадр/с на CPU) і потребує складного середовища збірки (Caffe, CUDA). Detectron2 від Meta AI забезпечує передові моделі детекції, проте є «важкою» системою з обов'язковою вимогою GPU для роботи у реальному часі. Dlib обмежений переважно задачами обробки облич і не містить готових модулів для трекінгу рук. Фреймворки загального призначення (TensorFlow, PyTorch) потребують самостійного збирання навчальних датасетів і тривалого навчання моделей. На противагу цим рішенням, MediaPipe HandLandmarker від Google забезпечує понад 30 кадр/с на CPU, повертає тривимірні координати 21 анатомічної точки кисті, постачається з попередньо навченою моделлю та має простий Python-API [3]. Це робить його оптимальним вибором за сукупністю обраних критеріїв.

Окремого обґрунтування потребує сам перехід від піксельного представлення жесту до landmark-представлення. Згорткові нейронні мережі, що працюють із сирими пікселями, потребують значних обчислювальних ресурсів та великих навчальних вибірок, а їхня стійкість до варіацій освітлення досягається лише інтенсивною аугментацією [4]. Натомість представлення жесту у вигляді нормалізованих координат суглобів кисті є за побудовою інваріантним до яскравості зображення (на вхід подаються геометричні координати, а не пікселі) та до зміни масштабу (координати нормовані відносно розміру кадру). Це дозволяє відмовитися від навчання власних класифікаторів і реалізувати просте порівняння за евклідовою метрикою, обчислювальна складність якого є лінійною за розмірністю вектора ознак.

Для захисту еталонних шаблонів обрано схему автентифікованого симетричного шифрування Fernet (AES-128 у режимі CBC + HMAC-SHA256), яка одночасно забезпечує конфіденційність та цілісність даних. Альтернативні схеми поступаються Fernet за критеріями завдання: «чистий» AES-CBC не містить вбудованої перевірки цілісності, а власна реалізація комбінованих схем підвищує ризик криптографічних помилок. Для деривації ключа обрано алгоритм PBKDF2-HMAC-SHA256 з кількістю ітерацій 390 000 — значенням, що відповідає рекомендаціям OWASP редакції 2023 р. щодо мінімальної стійкості до офлайн-атак перебору [5]. Сучасніші алгоритми деривації (Argon2, scrypt) забезпечують стійкість також до атак з використанням спеціалізованого обладнання, проте PBKDF2 залишається стандартом FIPS і входить до стандартної бібліотеки Python, що усуває залежність від додаткових пакетів.

Висновки

На основі сформульованих критеріїв оцінки виконано порівняльний аналіз інструментальних засобів та обґрунтовано вибір технологічного стека для інтелектуальної системи жестової автентифікації. Обрано комбінацію Python + OpenCV + MediaPipe HandLandmarker, яка забезпечує обробку понад 30 кадр/с на центральному процесорі без графічного прискорювача та кросплатформну роботу. Обґрунтовано перехід від піксельного до landmark-представлення жесту, що надає за побудовою інваріантність до варіацій освітлення та масштабу. Для захисту еталонних шаблонів обрано схему Fernet з деривацією ключа через PBKDF2-HMAC-SHA256 з 390 000

ітераціями, що відповідає сучасним рекомендаціям OWASP. Обраний стек є компромісним оптимумом між обчислювальною складністю, простотою інтеграції та рівнем безпеки і задовольняє вимоги до систем автентифікації на типовому апаратному забезпеченні масового користувача.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. MediaPipe: A Framework for Building Perception Pipelines / C. Lugaresi та ін. arXiv.org. URL: https://arxiv.org/pdf/1906.08172 (дата звернення: 10.03.2026).
2. Bradski G. The OpenCV Library. Dr. Dobb's Journal of Software Tools. 2000. URL: https://opencv.org (дата звернення: 10.03.2026).
3. MediaPipe Hands: On-device Real-time Hand Tracking / F. Zhang та ін. arXiv.org. URL: https://arxiv.org/pdf/2006.10214 (дата звернення: 10.03.2026).
4. Yasen M., Jusoh S. A systematic review on hand gesture recognition techniques, challenges and applications. PeerJ Computer Science. 2019. Т. 5. С. e218. URL: https://doi.org/10.7717/peerj-cs.218 (дата звернення: 10.03.2026).
5. OWASP Password Storage Cheat Sheet. OWASP Foundation. URL: https://cheatsheetseries.owasp.org/cheatsheets/Password_Storage_Cheat_Sheet.html (дата звернення: 10.03.2026).

Пальчик Владислав Олександрович – студент групи 1КІТС-22б, Факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: vladthefinger@gmail.com

Науковий керівник: ***Безпалий Кирило Валерійович*** – асистент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: kyrylo.bezpalyi@vntu.edu.ua

Palchyk Vladyslav O. – student of group 1KITS-22b, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: vladthefinger@gmail.com

Supervisor: ***Bezpalyi Kyrylo V.*** – assistant of the Department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: kyrylo.bezpalyi@vntu.edu.ua