

ПРОЕКТУВАННЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ АВТЕНТИФІКАЦІЇ НА ОСНОВІ РОЗПІЗНАВАННЯ ЖЕСТІВ

Вінницький національний технічний університет

Анотація

Роботу присвячено питанням проектування інтелектуальної системи автентифікації на основі розпізнавання жестів рук із використанням веб-камери. Визначено вимоги до системи та спроектовано алгоритм виділення та нормалізації ознак жесту на базі фреймворку MediaPipe, що формує 73-вимірний вектор ознак (або 146-вимірний у режимі двох рук). Розроблено алгоритм послідовної жестової автентифікації на основі нормованої евклідової метрики подібності, порогового рішення та механізму часової стабілізації. Формалізовано логіку автентифікації у вигляді детермінованого скінченного автомата.

Ключові слова: жестова автентифікація, MediaPipe, landmark-координати, вектор ознак, евклідова відстань, детермінований скінченний автомат.

Abstract

This paper addresses the design of an intelligent authentication system based on hand gesture recognition using a webcam. System requirements were defined, and an algorithm for feature extraction and normalization was designed based on the MediaPipe framework, which generates a 73-dimensional feature vector (or a 146-dimensional one in two-hand mode). An algorithm for sequential gesture authentication has been developed based on a normalized Euclidean similarity metric, a threshold decision, and a time stabilization mechanism. The authentication logic has been formalized as a deterministic finite automaton.

Keywords: gesture authentication, MediaPipe, landmark coordinates, feature vector, Euclidean distance, deterministic finite automaton.

Вступ

Традиційні методи автентифікації, засновані на паролях або PIN-кодах, мають суттєві обмеження: складні комбінації символів важко запам'ятовувати, прості ж є вразливими до атак. Альтернативою слугують біометричні методи, серед яких особливу увагу привертають системи на основі розпізнавання жестів рук. На відміну від статичних біометричних показників (відбитків пальців, сканування сітківки), жестова автентифікація поєднує унікальні поведінкові характеристики з можливістю безконтактної взаємодії, що є надзвичайно актуальним в умовах підвищених вимог до дистанційного обслуговування.

Поява фреймворку MediaPipe від Google та моделі HandLandmarker, яка здатна у реальному часі визначати 21 тривимірну ключову точку кисті зі швидкістю понад 30 кадрів/с навіть на пристроях без GPU [1], відкрила можливість для побудови ефективних систем жестової автентифікації на доступному обладнанні. Landmark-базоване представлення суттєво поступається піксельному за обсягом даних, але перевершує його за семантичною насиченістю та стійкістю до фонових перешкод і варіацій освітлення [2]. Метою цієї роботи є проектування архітектури такої системи, визначення вимог до неї та розробка алгоритмів виділення ознак і прийняття рішень про автентифікацію.

Результати дослідження

Сформулюємо ключові вимоги до спроектованої системи на підставі аналізу існуючих підходів. Основна функціональна вимога – здатність реєструвати та автентифікувати користувача за унікальною послідовністю жестів, зафіксованих через стандартну веб-камеру. «Паролем» у системі виступає послідовність із жестів, яку користувач визначає самостійно під час реєстрації; система підтримує виконання жестів як однією, так і двома руками. Серед нефункціональних вимог центральне місце займає інваріантність розпізнавання до варіацій освітлення, положення руки у кадрі та відстані до камери. Вимога до роботи у режимі реального часу обґрунтовує вибір оптимізованих бібліотек. Важливою вимогою є конфіденційність: у сховищі зберігаються не відеозаписи жестів, а лише компактні числові вектори ознак, що унеможливує відтворення оригінального відео. Архітектурно система складається з чотирьох функціональних алгоритмів: захоплення та попередня обробка

відеоданих (OpenCV); виділення та нормалізація ознак жестів (MediaPipe); порівняння з еталоном та прийняття рішення про автентифікацію; графічний інтерфейс користувача (Tkinter).

Центральним елементом системи є алгоритм виділення ознак, що перетворює відеопотік у компактне числове представлення жесту. Для детекції ключових точок обрано модель HandLandmarker із фреймворку MediaPipe, яка повертає тривимірні координати 21 анатомічної точки кисті: координати x та y нормовані відносно розміру кадру (діапазон $[0, 1]$), координата z відображає відносну глибину відносно площини зап'ястка [3]. Спроекований вектор ознак для однієї руки формується з трьох компонент. Перша — нормалізовані тривимірні координати всіх 21 точки (63 елементи). Друга — кути нахилу пальців, обчислені за чотириквadrантною функцією $\arctan2$ для кожного з п'яти кінчиків пальців (5 елементів). Третя — евклідові відстані від кінчиків пальців до центру долоні, що відображають ступінь розкриття пальців (5 елементів). Повний вектор однієї руки має розмірність 73, у режимі двох рук — 146 (конкатенація двох векторів). Така комбінація компонент забезпечує практичну інваріантність до трансляційних та часткових масштабних перетворень [4]. Для підвищення стабільності еталонного вектора застосовується часове усереднення по N послідовних кадрах із підтверженою детекцією — підхід, широко застосовуваний у системах розпізнавання жестів для зменшення дисперсії ознак [4].

Ключовою архітектурною ідеєю алгоритму автентифікації є концепція послідовного жестового пароля: користувач реєструє унікальну послідовність із K жестів $\{A, B, C, \dots\}$, і позитивне рішення про доступ приймається лише при успішному підтвердженні повної послідовності. Комбінаторна складність 4-жестового пароля в алфавіті з 4 жестів без повторень становить $A!/(A-K)! = 24$ унікальні комбінації, а при розширенні алфавіту до 8 жестів — 1680 комбінацій, що суттєво ускладнює атаки перебором. Порівняння поточного вектора ознак f з еталонним здійснюється на основі евклідової відстані, яка перетворюється у нормовану оцінку подібності $S = \max(0, 100 \cdot (1 - D/C))$, де $C = 5,0$ — нормувальна константа. При нульовій відстані $S = 100\%$ (ідеальний збіг), при $D \geq C$ маємо $S = 0\%$. Рішення про розпізнавання жесту приймається при перевищенні порогового значення $T = 75\%$, що узгоджується з рекомендаціями для систем жестової автентифікації на основі нормалізованих landmark-ознак [5]. Для запобігання хибним спрацьовуванням від випадкового збігу одного кадру введено механізм часової стабілізації: жест вважається підтвердженим лише після безперервного утримання оцінки $S \geq T$ протягом $T_{sf} = 15$ послідовних кадрів, що при частоті 30 кадр/с відповідає $\approx 0,5$ с витримки [6].

Логіку алгоритму формалізовано у вигляді детермінованого скінченного автомата (ДСА) з множиною станів $Q = \{Idle, Step1, Step2, Step3, Step4, Granted, Denied\}$, вхідним алфавітом $\Sigma = \{match, no_match, cancel\}$, початковим станом $Idle$ та термінальними станами $\{Granted, Denied\}$. Подія $match$ генерується при досягненні лічильником стабільних кадрів значення T_{sf} ; подія no_match — при $S < T$; $cancel$ — при примусовому завершенні сесії. ДСА-представлення підтверджує детермінованість і повноту спроектованої логіки, а також відсутність неоднозначних переходів між станами. З точки зору безпеки, механізм часової стабілізації додатково підвищує захист від автоматизованих атак перебору: для кожного жесту необхідно безперервно утримувати позицію руки протягом $\approx 0,5$ с, що на порядки збільшує часові витрати брутфорс-атаки.

Висновки

У роботі спроектовано архітектуру інтелектуальної системи автентифікації на основі розпізнавання жестів рук, що складається з чотирьох функціональних алгоритмів. Алгоритм виділення ознак формує 73-вимірний вектор на базі landmark-представлення кисті, що поєднує просторові координати суглобів, кутові характеристики пальців та відстані до центру долоні, і є стійким до трансляційних та часткових масштабних перетворень. Алгоритм послідовної автентифікації на основі нормованої евклідової метрики, порогу $T = 75\%$ та часової стабілізації реалізує концепцію «жестового пароля», формальна модель якого у вигляді ДСА підтверджує детермінованість логіки. Спроекована система відповідає сучасним вимогам до безпеки, зручності використання та конфіденційності даних і є достатньою проектною основою для програмної реалізації.

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. 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).

3. Hand landmarks detection guide | Google AI for Developers. URL: https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker (дата звернення: 10.03.2026).
4. Yaseen 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. Real-Time Sign Language Detection using MediaPipe and Deep Learning / T. Chakraborty та ін. International Journal of Computer Applications. 2025. Т. 187, № 57. С. 78–83. URL: <https://doi.org/10.5120/ijca2025925952> (дата звернення: 10.03.2026).
6. Dynamic Hand Gesture Recognition Using MediaPipe and Transformer | MDPI. URL: <https://doi.org/10.3390/engproc2025108022> (дата звернення: 10.03.2026).

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

Науковий керівник: **Безпалій 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: **Bezpalii 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