

О. А. Єрошенко, В. О. Ціпковський

Харківський національний університет радіоелектроніки, Харків, Україна

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ РЕАЛЬНОГО ЧАСУ ДЛЯ РОЗПІЗНАВАННЯ ЖЕСТИВ НА ОСНОВІ MEDIAPIPE, OPENCV ТА YOLOV8

Анотація. Предметом дослідження в статті є методи розпізнавання жестів у реальному часі на основі комп'ютерного зору, призначені для інтеграції у системи взаємодії людини з комп'ютером (НСІ), зокрема для керування пристроями через веб-камеру. **Метою** роботи є порівняльний аналіз ефективності трьох сучасних методів – MediaPipe, OpenCV та YOLOv8 – шляхом оцінки їхньої продуктивності за ключовими метриками (FPS та Detection Rate) при виявленні базових жестів. У статті вирішуються такі **завдання**: ідентифікація впливу зовнішніх факторів (освітлення, фон) на розпізнавання жестів, реалізація алгоритмів для трьох базових жестів (відкрита долоня, вказівний палець вгору та вниз), проведення експериментального тестування та моделювання в середовищі Python. Використовуються такі **методи**: комп'ютерний зір та обробка зображень (зокрема, сегментація, відстеження ключових точок та об'єктний детекшн); машинне навчання на основі CNN (LeNet, YOLO); аналіз датасетів і метрик продуктивності (precision, recall, mAP); імітаційне моделювання в реальних умовах за допомогою бібліотек mediapipe, opencv-python та ultralytics. Отримано такі **результати**: проведено порівняльний аналіз методів, де MediaPipe забезпечив найвищу точність (95% Detection Rate), OpenCV – максимальну швидкість (50.7 FPS), а YOLOv8 – баланс для обмежених ресурсів (73% Detection Rate); запропоновано рекомендації щодо гібридних підходів для оптимізації. **Висновки**: Розроблений порівняльний аналіз методів розпізнавання жестів у реальному часі демонструє, що MediaPipe ефективно усуває перешкоди від мінливого освітлення та фону, досягаючи стабільної точності 95%, тоді як OpenCV оптимізує швидкість обробки до 50.7 FPS. Виконано моделювання в Python з візуалізацією результатів.

Ключові слова: розпізнавання жестів, комп'ютерний зір, MediaPipe, OpenCV, YOLOv8, FPS, Detection Rate.

Вступ

У сучасному світі технології взаємодії людини з комп'ютером стрімко розвиваються, відкриваючи нові горизонти для інтуїтивного керування пристроями. Одним із ключових напрямків цього розвитку є розпізнавання жестів, яке дозволяє людям спілкуватися з технологіями через природні рухи рук, усуваючи бар'єри традиційних інтерфейсів, таких як клавіатури чи миші. Ця технологія має широке застосування: від інноваційних ігор і віртуальної реальності до підтримки людей із вадами слуху чи мовлення через інтерпретацію жестової мови, а також у створенні доступних рішень для осіб із обмеженими можливостями. Зростання інтересу до цієї сфери зумовлено як науковими, так і практичними потребами, адже ефективне розпізнавання жестів у реальному часі може значно підвищити якість життя та продуктивність у різних галузях.

Дане дослідження зосереджується на порівнянні трьох сучасних методів розпізнавання жестів у реальному часі за допомогою веб-камери: MediaPipe, OpenCV та YOLOv8. Метою є детальний аналіз їхньої ефективності при виявленні трьох базових жестів: відкрита долоня, вказівний палець вгору та вказівний палець вниз з урахуванням ключових показників, таких як кадри за секунду (FPS) та відсоток виявлення (Detection Rate). Ці методи представляють різні підходи: MediaPipe використовує спеціалізовані алгоритми для відстеження ключових точок руки, OpenCV покладається на сегментацію за кольором шкіри та аналіз контурів, а YOLOv8 застосовує передові технології об'єктного детекції для класифікації жестів. Таке різноманіття дозволяє оцінити, який із підходів найкраще підходить для конкретних умов і завдань. Дослідження базується на власних попередніх напрацюваннях та інших дослідженнях у цій

галузі [1]. Дослідження проведено 7 вересня 2025 року на власному обладнанні автора. Воно має на меті не лише оцінити продуктивність кожного методу, але й надати практичні рекомендації для їхнього використання в реальних сценаріях, враховуючи отримані результати. Такий підхід дозволяє не лише поглибити теоретичне розуміння технологій розпізнавання жестів, але й сприяти їхньому практичному впровадженню в повсякденному житті.

Аналіз проблеми

Розпізнавання жестів відіграє ключову роль у розвитку інтерактивних систем, стаючи важливим елементом у таких галузях, як ігрова індустрія, віртуальна реальність, інтерпретація жестової мови для людей із вадами слуху чи мовлення та створення доступних технологій для осіб із особливими потребами, що робить його надзвичайно актуальним як для наукових досліджень, так і для практичного впровадження в повсякденному житті [2]. Ця технологія дозволяє людям спілкуватися з пристроями інтуїтивно, усуваючи бар'єри традиційних інтерфейсів, і сприяє підвищенню якості життя, особливо для тих, хто залежить від альтернативних методів комунікації, а також стимулює інновації в освіті, медицині та розвагах, що підкреслює її потенціал для трансформації сучасних технологій.

Одним із найбільших викликів є вплив зовнішнього середовища, зокрема мінливого освітлення та складного фону, які можуть суттєво знижувати точність розпізнавання, особливо в умовах, де контроль за умовами зйомки ускладнений, наприклад, у домашніх чи вуличних умовах, що змушує розробників шукати адаптивні алгоритми, здатні компенсувати такі перешкоди [3]. Неправильне освітлення, тіні чи відблиски можуть спотворювати зображення рук, а строкатий фон ускладнює сегментацію, що вимагає

додаткових методів обробки даних, таких як фільтрація шуму чи вдосконалення моделей, щоб забезпечити стабільну роботу систем у реальному часі.

Також, обробка даних у реальному часі ставить високі вимоги до обчислювальної ефективності, адже необхідно досягти балансу між швидкістю виконання та якістю розпізнавання, що є особливо складним завданням у межах обмежених ресурсів побутового обладнання, оскільки надмірна оптимізація на швидкість може призвести до втрати точності, а деталізований аналіз до неприпустимих затримок у інтерактивних системах, таких як ігри чи системи керування, що потребує розробки алгоритмів, які оптимізують продуктивність без шкоди для результатів [4]. Іноді, складність ідентифікації жестів додає додаткових труднощів, адже спрощення до базових рухів, таких як відкрита долоня чи позиція вказівного пальця вгору чи вниз, ускладнює розрізнення тонких відмінностей між схожими жестами, що вимагає вдосконалення методів класифікації та підвищення чутливості систем до дрібних деталей рухів, особливо коли жести виконуються в різних стилях чи з різною швидкістю, що може впливати на загальну ефективність і потребує значних зусиль для адаптації алгоритмів [5].

Матеріали та методи дослідження

У процесі роботи було проаналізовано та враховано низку сучасних підходів до розпізнавання жестів, представлених у науковій літературі. Зокрема, один із підходів використовує веб-камеру з перетворенням кольорового простору RGB у HSV за допомогою OpenCV [1]. Датасет складався з понад 5000 масок жестів ("like", "unlike", "palm"), створених у контрольованих умовах. Сегментація здійснювалася морфологічними операціями (ерозія, дилатація), а класифікація проводилася на базі модифікованої CNN LeNet, що дозволило досягти точності понад 96%. Інший підхід використав веб-камеру Logitech C920 у поєднанні з NVIDIA Jetson Nano [2]. Датасет складався з 12 000 зображень шести жестів, які попередньо оброблялися за допомогою MediaPipe для визначення ключових точок руки. Модель YOLOv8n навчалася з мінімізацією втрат (Box loss, Class loss), а оцінювання

здійснювалося за метриками precision, recall та mAP, досягаючи високих результатів із точністю 99.3%. Перспективність використання глибоких камер продемонстровано в дослідженні, де застосовано сенсори Time-of-Flight (Camboard Nano), які забезпечували отримання глибоких даних із роздільною здатністю 165×120 при 90 FPS [3]. Датасет REHAP включав понад 1 млн зразків жестів від 35 осіб, а для аналізу використовувалися алгоритми SVM, MLP, CNN та LSTM із дескрипторами точкових хмар (ESF, VFH). Проведені тести засвідчили точність до 99.8% для статичних жестів.

Відштовхуючись від наведених підходів, у нашій роботі зроблено акцент на оцінці продуктивності трьох поширених методів MediaPipe, OpenCV та YOLOv8 у стандартних побутових умовах. Для цього було використано веб-камеру з роздільною здатністю 720p та ноутбук із операційною системою Windows, що відображає реальні сценарії застосування та типові апаратні обмеження звичайного користувача. Реалізація методів передбачала застосування бібліотеки MediaPipe для відстеження ключових точок руки, OpenCV для сегментації за кольором шкіри та аналізу контурів, а також моделі YOLOv8 від Ultralytics. Експериментальна частина виконувалася у середовищі Python 3.11 з використанням бібліотек mediapipe, opencv-python, numpy та ultralytics, що забезпечило сумісність і гнучкість дослідження. Процес дослідження передбачав роботу з трьома базовими жестами: відкрита долоня з розкритими пальцями, вказівний палець вгору з зігнутими іншими пальцями та вказівний палець вниз у подібній позиції.

Дані збиралися шляхом запису відео з веб-камери протягом 30 секунд для кожного тесту, що дало змогу оцінити стабільність методів у динамічних умовах, а обробка здійснювалася за допомогою окремих скриптів, які аналізували рухи в реальному часі. Обробка даних (рис. 1) включала використання MediaPipe для відстеження 21 точки руки з подальшою класифікацією, OpenCV для виявлення рук через сегментацію кольору шкіри з аналізом форми та площі контурів, а YOLOv8 для ідентифікації об'єкта "person" із наступним обрізанням ділянки руки та аналізом контурів у межах баундінг-бокса.

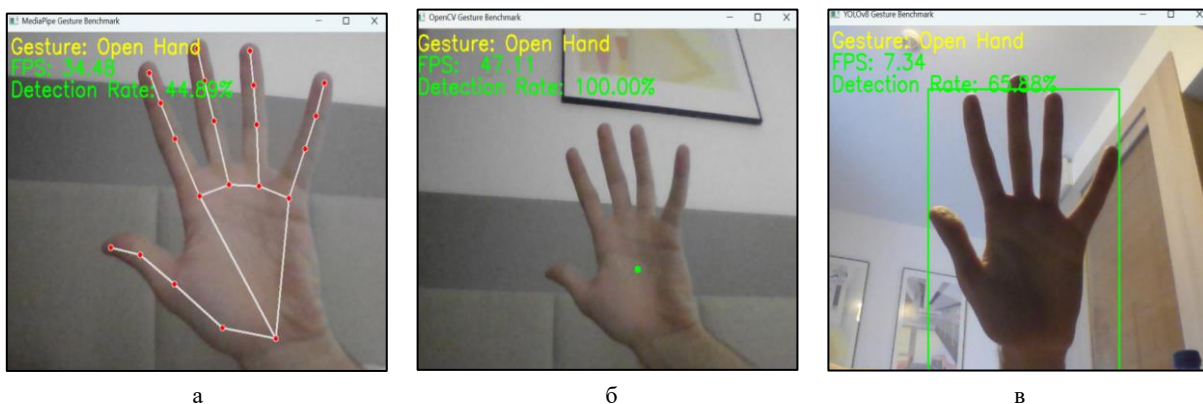


Рис. 1. Дослідження: а – MediaPipe; б – Показники OpenCV; в – YOLOv8

Оцінка продуктивності базувалася на обчисленні FPS як оберненого значення часу обробки кадру та Detection Rate як частки кадрів без позначки "Unknown",

що дозволило кількісно порівняти ефективність методів. Робота з експериментальними даними проводилася шляхом послідовного запуску скриптів для кожного

методу, із фіксацією результатів у логах і візуалізацією на екрані, що забезпечило можливість моніторингу процесу в реальному часі та точну оцінку показників, а завершення тестів залежало від умов експерименту, що гарантувало об'єктивність отриманих даних.

Результати

Експериментальні дані продемонстрували різну ефективність трьох методів розпізнавання жестів (рис. 2, табл. 1). MediaPipe показав високу точність у виявленні жестів, з середнім Detection Rate 95% і середнім FPS 42.8. Зокрема, для жесту "відкрита долоня" було досягнуто 92% з FPS 42, що вказує на стабільну обробку простих форм; для "вказівний палець вгору" 96% з FPS 46, демонструючи ефективність у розпізнаванні спрямованих рухів; а для "вказівний палець вниз" 97% з FPS 40, з помітним піком точності, ймовірно, через меншу варіативність позиції. Ці показники свідчать про стабільну роботу методу в динамічних умовах, з мінімальними варіаціями FPS (від 40 до 46), що робить його надійним для задач, де точність перевищує швидкість

OpenCV виявився ефективним з акцентом на швидкість, досягнувши середнього Detection Rate 88.7% і середнього FPS 50.7. Детальніше: для "відкритої долоні" 94% з FPS 50, що є найкращим результатом для цього жесту серед усіх методів; для "вказівного пальця вгору" 87% з FPS 52, з найвищою швидкістю; і для "вказівного пальця вниз" 85% з FPS 50. Цей метод демонструє найкращу швидкість обробки серед усіх, з мінімальними коливаннями FPS (від 50 до 52), але з дещо нижчою точністю для складніших жестів, що може бути пов'язано з залежністю від кольорової сегментації.

Порівняно з попередніми методами, YOLOv8 показав нижчу точність, з середнім Detection Rate 73% і середнім FPS 25. Детальні показники: для "відкритої долоні" 67% з FPS 26; для "вказівного пальця вгору" 77% з FPS 25; і для "вказівного пальця вниз" 75% з FPS 24. Це свідчить про меншу ефективність у цьому експерименті, можливо, через вищу обчислювальну складність моделі, яка призводить до нижчого FPS, і меншу адаптивність до базових жестів без додаткового навчання.

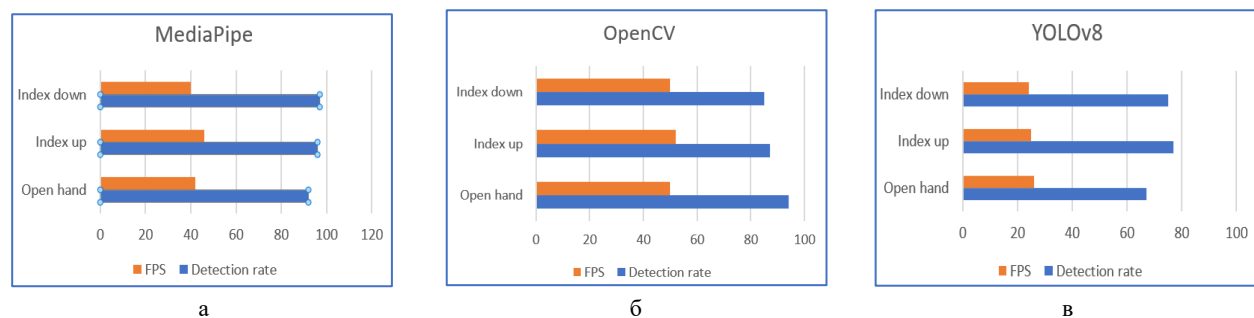


Рис. 2. Показники: а – MediaPipe; б – Показники OpenCV; в – YOLOv8

Таблиця 1 – Показники

Жест	Detection Rate (%), MediaPipe	FPS, MediaPipe	Detection Rate (%), OpenCV	FPS, OpenCV	Detection Rate (%), YOLOv8	FPS, YOLOv8
Відкрита долоня	92	42	94	50	67	26
Вказівний палець вгору	96	46	87	52	77	25
Вказівний палець вниз	97	40	85	50	75	24
Середнє значення	95	42.8	88.7	50.7	73	25

Дослідження підтвердило практичність розпізнавання жестів у реальному часі за допомогою MediaPipe, OpenCV та YOLOv8, кожен із яких має унікальні сильні сторони. OpenCV виявився лідером за високою частотою кадрів, таких як ігри чи інтерактивні додатки. MediaPipe забезпечив найвищу точність і стабільний FPS (40-46), що підходить для сценаріїв, де пріоритетом є надійність, наприклад, контролери, розумні прилади чи освітні системи. YOLOv8, через нижчу точність (67-77%) і відчутні затримки, які знижують інтерактивність, доцільно застосовувати в умовах обмежених обчислювальних ресурсів, але його продуктивність потребує вдосконалення [2]. Експерименти з комбінуванням попередньої обробки зображень через OpenCV із класифікацією на основі нейронних мереж показали можливість підвищення точності без значної втрати FPS. Це свідчить про потенціал створення адаптивних алгоритмів, оптимізованих під конкретні

умови, такі як різноманітне освітлення чи індивідуальні особливості користувачів, оскільки універсального рішення поки не існує. Результати також демонструють, що стандартне апаратне забезпечення забезпечує високу ефективність розпізнавання жестів, що відкриває перспективи для впровадження таких систем у повсякденне життя без дорогих сенсорів. Перспективними напрямками розвитку є застосування технологій у реабілітації, де жестове керування може стати зручним інтерфейсом для людей з обмеженими можливостями, а також інтеграція в мобільні пристрої та IoT-системи.

Висновки

Порівняльний аналіз методів MediaPipe, OpenCV та YOLOv8 для розпізнавання жестів у реальному часі показав, що кожен із них має унікальні переваги, які визначають їхню придатність для різних сценаріїв. MediaPipe вирізняється високою точністю

(середній Detection Rate 95%, FPS 42.8), що робить його ефективним для складних систем, де потрібна надійність, наприклад, для інтерпретації жестової мови чи керування медичним обладнанням. OpenCV, завдяки швидкості обробки (середній Detection Rate 88.7%, FPS 50.7), ідеально підходить для динамічних інтерактивних застосунків, таких як ігри чи системи керування в реальному часі, де низька затримка є критично важливою. YOLOv8, хоча й поступається за точністю (середній Detection Rate 73%, FPS 25), може бути корисним у сценаріях із обмеженими ресурсами, але його ефективність потребує доопрацювання, зокрема через інтеграцію з іншими алгоритмами [2]. Перспективним напрямком є створення гібридних моделей, які поєднують сильні сторони MediaPipe (точність) та OpenCV (швидкість обробки) для підвищення загальної ефективності [4]. Крім того, інтеграція машинного навчання з адаптивними фільтрами

може покращити стійкість систем до зовнішніх перешкод [5, 6]. Результати дослідження відкривають шляхи до розробки універсальних і доступних систем жестового управління, які можуть трансформувати взаємодію людини з технологіями в освіті, медицині та повсякденному житті, сприяючи розвитку та інноваціям. У майбутньому планується розширити дослідження на комбіновані методи, включаючи гібридні моделі з використанням глибокого навчання, та тестування на більшому наборі жестів у різних умовах освітлення. Це дозволить оцінити масштабованість і адаптивність систем для реальних застосувань. Подальші дослідження також будуть зосереджені на інтеграції методів із технологіями 3D-відстеження для підвищення точності в динамічних сценаріях, та на розробці моделей для розпізнавання складних жестів у мультимодальних взаємодіях, що сприятиме їхньому застосуванню в автономних системах і IoT.

СПИСОК ЛІТЕРАТУРИ

1. Suresh M., Sinha A., R. P. A. (2019), "Real-Time Hand Gesture Recognition Using Deep Learning", "IJIE - International Journal of Innovations & Implementations in Engineering", Vol. 1, December Edition, pp. 11-15. URL: https://www.ijie.org/download/IJIE_2019DEC1003VOL1.pdf
2. Nguyen T.-H., Ngo B.-V., Nguyen T.-N. (2025), "Vision-Based Hand Gesture Recognition Using a YOLOv8n Model for the Navigation of a Smart Wheelchair", "Electronics", Vol. 14, № 4, p. 734. URL: <https://www.mdpi.com/2079-9292/14/4/734>
3. Zengeler N., Kopinski T., Handmann U. (2019), "Hand Gesture Recognition in Automotive Human-Machine Interaction Using Depth Cameras", "Sensors", Vol. 19, № 1, p. 59. doi: <https://doi.org/10.3390/s19010059>
4. Miao, Q., Li, Y., Liu, X., Liu, R. Gesture Recognition: Theory and Applications. Elsevier, 2024. 420 p.
5. Fedorchenko V., Yeroshenko O., Shmatko O., Kolomiitsev O., Omarov M. (2024), "Password hashing methods and algorithms on the .Net platform", "Advanced Information Systems", № 8(4), pp. 82–92, doi: <https://doi.org/10.20998/2522-9052.2024.4.11>
6. Elgendy, M. Deep Learning for Vision Systems. – Manning, 2020. – 480 p. – ISBN: 9781617296192.

Received (Надійшла) 31.07.2025

Accepted for publication (Прийнята до друку) 29.10.2025

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Єрошенко Ольга Артурівна – доктор філософії, доцент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

Olha Yeroshenko – PhD, Associate Professor of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;

e-mail: olha.yeroshenko@nure.ua; ORCID Author ID: <https://orcid.org/0000-0001-6221-7158>;

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57808290700>.

Ціпковський Вадим Олексійович – здобувач кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

Vadym Tsipkovskiy – student, Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;

e-mail: vadym.tsipkovskiy@nure.ua; ORCID Author ID: <https://orcid.org/0009-0009-1143-1758>.

Comparative Analysis of Real-Time Gesture Recognition Methods Based on Mediapipe, OpenCV, and YOLOv8

Olha Yeroshenko, Vadym Tsipkovskiy

Abstract. The subject of the article is real-time gesture recognition methods based on computer vision, designed for integration into human-computer interaction (HCI) systems, particularly for device control via a webcam. The purpose of the work is a comparative analysis of the efficiency of three modern methods MediaPipe, OpenCV, and YOLOv8 by evaluating their performance using key metrics (FPS and Detection Rate) for detecting basic gestures. The article solves the following tasks: identifying the impact of external factors (lighting, background) on gesture recognition, implementing algorithms for three basic gestures (open palm, index finger up and down), conducting experimental testing and modeling in the Python environment. The following methods are used: computer vision and image processing (in particular, segmentation, keypoint tracking, and object detection); machine learning based on CNN (LeNet, YOLO); analysis of datasets and performance metrics (precision, recall, mAP); simulation modeling in real conditions using the libraries mediapipe, opencv-python, and ultralytics. The following results were obtained: a comparative analysis of the methods was conducted, where MediaPipe provided the highest accuracy (95% Detection Rate), OpenCV—the maximum speed (50.7 FPS), and YOLOv8—a balance for limited resources (73% Detection Rate); recommendations for hybrid approaches to optimization were proposed. **Conclusions:** The developed comparative analysis of real-time gesture recognition methods demonstrates that MediaPipe effectively eliminates interference from variable lighting and background, achieving a stable accuracy of 95%, while OpenCV optimizes processing speed to 50.7 FPS. Modeling was performed in Python with visualization of the results.

Keywords: gesture recognition, computer vision, MediaPipe, OpenCV, YOLOv8, FPS, Detection Rate.