

Г. А. ДОБРОВОЛЬСЬКИЙ, С. О. ПЕТРИК
Запорізький національний університет

ПОРІВНЯЛЬНИЙ АНАЛІЗ ГЕОМЕТРИЧНИХ ТА ВЕКТОРНИХ МЕТРИК 3D-ІДЕНТИФІКАЦІЇ ОБЛИЧЧЯ В УМОВАХ ДИНАМІЧНИХ ОКЛЮЗІЙ

У даній роботі досліджується проблема стійкої до оклюзій тривимірної (3D) реконструкції обличчя за допомогою монокулярного відео для систем автентифікації в режимі реального часу. Актуальність дослідження зумовлена тим, що випадкові динамічні оклюзії (рухи рук, сторонні предмети, самооклюзії при поворотах голови) залишаються основним джерелом помилок у сучасних алгоритмах розпізнавання, оскільки вони спотворюють фотометричні ознаки та ключові точки обличчя непередбачуваним чином.

Метою роботи є порівняльний аналіз ефективності різних типів метрик (векторних та геометричних) для визначення тотожності осіб в умовах, максимально наближених до реальних. Для цього було реалізовано пайплайн обробки відео, що включає фільтрацію кадрів за різкістю (оператор Лапласа) та структурною подібністю (SSIM), а також 3D-реконструкцію на базі регресійної моделі MICA у параметричний простір FLAME. Для забезпечення стабільності ідентичності в умовах динамічних перешкод застосовано медіанне злиття shape-векторів.

Експериментальне дослідження проводилося на відеоматеріалах MIT OpenCourseWare, які містять неконтрольоване освітлення, активну жестикуляцію та специфічні оклюзії, спричинені використанням прозорих віртуальних дощок. Задача ідентифікації була сформульована як бінарна класифікація («одна особа» проти «різних осіб») з використанням комплексу метрик: L2 distance, косинусна схожість, Chamfer Distance, Hausdorff Distance та інших.

Результати аналізу показали суттєву перевагу векторних представлень над геометричними. Встановлено, що метрика L2 distance у просторі параметрів FLAME є найбільш ефективною, забезпечуючи найвищі показники Precision, Recall, F1-score та найбільшу площу під ROC-кривою (AUC). Натомість геометричні метрики (зокрема Hausdorff distance) виявилися найменш стійкими через надмірну чутливість до викидів та локальних шумів реконструкції, викликаних оклюзіями.

Практична цінність роботи полягає у підтвердженні можливості використання 3D-реконструкції для надійної автентифікації в неконтрольованому середовищі при дотриманні лімітів затримки для мобільних пристроїв. Отримані висновки можуть бути використані при розробці систем дистанційного навчання, відеоконференції та контролю доступу.

Ключові слова: 3D-реконструкція обличчя, монокулярне відео, динамічні оклюзії, біометрична автентифікація, модель FLAME, MICA, векторні метрики, L2 distance, геометричні метрики.

H. A. DOBROVOLSKYI, S. O. PETRYK
Zaporizhzhia National University

COMPARATIVE ANALYSIS OF GEOMETRIC AND VECTOR METRICS OF 3D FACE IDENTIFICATION UNDER DYNAMIC OCCLUSIONS

This study investigates the problem of occlusion-robust three-dimensional (3D) face reconstruction from monocular video for real-time authentication systems. Random dynamic occlusions – such as hand movements, objects entering the field of view, and self-occlusions driven by pose changes – remain a dominant source of failure in modern algorithms because they are temporally non-stationary and unpredictably corrupt the evidence for facial surface points.

The objective of the research is a comparative analysis of various metric types (vector-based and geometric) to determine identity under conditions closely matching real-world operational scenarios. The proposed method utilizes a two-stage input stream filtration based on a sharpness metric (Laplacian operator) and structural similarity (SSIM) to select high-quality, diverse frames. Reconstruction is performed using the MICA neural network, which maps images into the parametric space of the FLAME model. To enhance identity stability and mitigate the impact of short-term dynamic interference, median fusion of shape vectors is applied.

Experimental evaluation was conducted using MIT OpenCourseWare lecture videos, which feature non-ideal lighting, active gesturing, and specific occlusions caused by virtual transparent boards. The identification task was formulated as binary classification using a suite of metrics, including L2 distance, cosine similarity, Chamfer Distance, and Hausdorff Distance.

The results of the analysis demonstrate a significant advantage of vector-based representations over direct surface geometry comparisons. It was established that L2 distance in the parametric shape-space is the most effective metric, providing the highest Precision, Recall, and F1-score, as well as the largest Area Under the Curve (AUC). Conversely,

geometric metrics, particularly the Hausdorff distance, proved to be the least reliable due to their critical sensitivity to outliers and local reconstruction noise caused by occlusions.

The practical value of the work lies in confirming the feasibility of using 3D face reconstruction for reliable biometric authentication in uncontrolled environments while adhering to the latency constraints of modern mobile devices.

Keywords: *3D face reconstruction, monocular video, dynamic occlusions, biometric authentication, FLAME model, MICA, vector-based metrics, L2 distance, geometric metrics, real-time.*

Постановка проблеми

Сучасний розвиток систем цифрової ідентифікації та біометричної автентифікації висуває дедалі жорсткіші вимоги до точності, надійності та стійкості алгоритмів розпізнавання обличчя. В умовах масового впровадження мобільних пристроїв та систем віддаленої ідентифікації, автентифікація в режимі реального часу за допомогою однієї RGB-камери (монокулярного відео) стає стандартом галузі. Використання тривимірної (3D) реконструкції обличчя на основі відеопотоку дозволяє значно підвищити стабільність розпізнавання ідентичності незалежно від пози чи виразу обличчя, а також надає важливі геометричні ознаки для перевірки «живості» (liveness detection) та захисту від атак спуфінгу. На відміну від традиційного порівняння у просторі 2D-зображень, 3D-модель є геометрично інваріантною репрезентацією, що теоретично не залежить від умов освітлення чи кута зйомки.

Проте, попри значний прогрес у галузі глибокого навчання, випадкові динамічні оклюзії залишаються одним із домінуючих джерел відмов та помилок у системах 3D-реконструкції. Під динамічними оклюзіями розуміють часові процеси, які випадковим чином приховують або спотворюють частину ознак поверхні обличчя. Це можуть бути руки, що торкаються обличчя, об'єкти (чашки, телефони), аксесуари (окуляри, маски, шарфи), а також самооклюзії, спричинені поворотами голови. Такі перешкоди є тимчасово нестаціонарними, оскільки набір видимих елементів обличчя змінюється від кадру до кадру, що створює серйозні проблеми для фотометричних методів та алгоритмів відстеження ключових точок.

Актуальність даного дослідження зумовлена необхідністю розробки методів, здатних забезпечувати точну автентифікацію в умовах, далеких від ідеальних лабораторних сценаріїв. Більшість існуючих робіт у сфері 3D-розпізнавання обличчя спираються на статичні 3D-скани або відео, зняті в контрольованих умовах з рівномірним освітленням. Однак у реальних сценаріях, таких як дистанційне навчання, відеоконференції або онлайн-банкінг, користувачі постійно рухаються, жестикулюють, а обличчя може частково перекиватися сторонніми предметами. Оклюзії в таких випадках вводять невластиві обличчю текстури та краї, які можуть помилково сприйматися екстракторами ознак, призводячи до «запікання» зовнішнього вигляду оклюдера в параметри 3D-моделі (albedo/geometry).

Для систем автентифікації проблема оклюзій має два критичні аспекти: доброякісний (неумисна поведінка користувача) та адверсаріальний (навмисна спроба обійти систему). У першому випадку система повинна уникати помилкових відмов, експлуатуючи часову надлишковість відео та запобігаючи «дрейфу ідентичності» при частковій видимості обличчя. У другому випадку зловмисник може навмисно закрити певні зони обличчя, щоб спровокувати модель на помилку або змусити її покладатися на невелику видиму ділянку, яку легше підробити. Отже, стійка система повинна не лише відновлювати геометрію, але й надавати оцінки впевненості та видимості для прийняття ризико-орієнтованих рішень: затримати рішення до отримання кращих кадрів, продовжити з урахуванням невизначеності або відхилити запит.

Окремим викликом є вимога роботи в режимі реального часу на мобільних пристроях. Реконструкція та автентифікація мають виконуватися з латентністю ≤ 200 мс на сучасному мобільному GPU. Це виключає використання важких генеративних методів (наприклад, на основі дифузії) без суттєвої дистиляції та вимагає оптимізованих архітектур, таких як мобільні меш-пайплайни або легкі регресійні моделі типу MICA чи DECA. При цьому виникає наукова

суперечність між необхідністю складної обробки оклюзій (сегментація, багатогіпотезна реконструкція) та обмеженими обчислювальними ресурсами мобільних платформ.

Аналіз останніх досліджень і публікацій

1. Еволюція методів 3D-реконструкції обличчя за монокулярним відео

Проблема відновлення тривимірної геометрії обличчя за одним RGB-зображенням або відеопотоком є однією з центральних тем у комп'ютерному зорі останнього десятиліття. Традиційні підходи, що базувалися на аналізі форми за затіненням (shape-from-shading) або класичних методах активних моделей зовнішнього вигляду (AAM), поступилися місцем методам глибокого навчання, які використовують 3D-морфологічні моделі обличчя (3DMM). Сучасні архітектури зазвичай виконують пряму регресію параметрів моделі (таких як FLAME) безпосередньо із зображення, що дозволяє отримувати детальні меші в режимі реального часу.

Для мобільних застосунків та систем автентифікації критично важливим є баланс між обчислювальною складністю та точністю. Робота Mobile Face Mesh (2019) [1] продемонструвала можливість отримання меша з 468 вершин на мобільних GPU з латентністю менше 10 мс, використовуючи каскадну архітектуру з детектора та трекара. Розширення цього підходу, Attention Mesh (2020) [2], впровадило механізми просторової уваги (spatial transformer attention) для виділення семантично важливих областей, таких як губи та очі, що підвищило точність без суттєвого навантаження на систему. Проте ці методи переважно орієнтовані на трекінг для AR-фільтрів і не завжди забезпечують необхідну для автентифікації стабільність ідентичності в умовах оклюзій.

2. Проблема випадкових динамічних оклюзій: визначення та класифікація

У джерелах випадкові динамічні оклюзії визначаються як часово-змінні процеси, що приховують або спотворюють ознаки поверхні обличчя, змінюючи набір видимих елементів (Vt) непередбачуваним чином. На відміну від статичних оклюзій (наприклад, постійної бороди), динамічні перешкоди вводять у систему «не-лицьові» текстури, краї та рухи, які часто плутають екстрактори ознак та механізми уваги.

Відповідно до поточної класифікації, виділяють три основні типи оклюзій:

1. Тимчасові зовнішні оклюдери: руки, предмети (телефони, чашки), окуляри, медичні маски та шарфи.
2. Рухомі оклюдери сцени: інші люди, що проходять повз, або швидкий рух об'єктів поблизу обличчя.
3. Самооклюзії: виникають через значні повороти голови (профільні ракурси), коли частини обличчя (ніс, щоки) перекривають інші зони.

Для систем автентифікації оклюзії створюють два типи загроз. У доброякісному сценарії вони спричиняють «дрейф ідентичності» та помилкові відмови через неповні дані. В адверсаріальному сценарії зловмисник може навмисно закривати критичні зони обличчя, щоб змусити модель покладатися на невелику видиму ділянку, яку легше підробити (наприклад, за допомогою часткового реплею або роздрукованого патча). Дослідження Extreme 3D Face Reconstruction (2018) [3] чітко показало, що оклюзії призводять до різкого падіння точності верифікації, проте методи «заповнення дірок» (hole filling) здатні частково відновити працездатність системи.

3. Таксономія методів боротьби з оклюзіями

Сучасна література пропонує декілька рівнів обробки оклюзій у пайплайнах 3D-реконструкції:

1. Семантична сегментація та маскування. Використання масок шкіри або сегментації обличчя дозволяє виключити сторонні пікселі (волосся, окуляри, маски) з фотометричних втрат при навчанні. моделі DECA [4] та CEST [5] продемонстрували, що без таких масок колір та текстура оклюдера «запікаються» (baking-in) в альbedo обличчя, що робить реконструкцію непридатною для подальшого аналізу.

2. Навчання на основі впевненості та невизначеності. Методи, що працюють з щільними ключовими точками (Dense Landmarks) [6], передбачають рівень невпевненості для кожної точки. Оклюдовані або вихідні за межі кадру точки отримують низьку вагу при підгонці моделі (fitting), що стабілізує загальну геометрію. FlowFace (2024) використовує пріорні знання (наприклад, нейтральну форму МІСА) у зонах, де вирівнювання є ненадійним через оклюзії.

3. Часова агрегація та фільтрація. У відеопотоці часова надлишковість є ключем до стійкості. Прості методи ковзного вікна або фільтри типу 1 Euro filter зменшують дрижання (jitter), але можуть накопичувати помилки, якщо оклюдер присутній у кадрі довго. Більш складні підходи використовують сумісну оптимізацію декількох кадрів (bundle adjustment), що дозволяє «бачити крізь» тимчасові перешкоди, інтегруючи інформацію з різних ракурсів.

4. Багатогіпотезна реконструкція та генеративне доповнення. Оскільки оклюдовані зони є принципово неоднозначними, методи типу Diverse3DFace (2022) [7] або OFER (2025) [8] генерують декілька правдоподібних варіантів заповнення невидимих ділянок. OFER використовує дифузійні моделі для семплювання гіпотез форми та виразу, демонструючи значну перевагу над детермінованими регресорами на наборах даних з масками та руками. Проте такі методи поки що перевищують ліміти затримки для мобільних пристроїв (≤ 200 мс).

4. Метрики та набори даних для оцінки стійкості

Стандартні метрики якості 3D-реконструкції зазвичай зосереджені на геометричній точності, такі як Normalized Mean Error (NME) на ключових точках або Chamfer Distance (CD) та Hausdorff Distance (HD) на хмарах точок. Проте джерела вказують на важливість часових метрик, таких як Screen Space Motion Error (SSME), які оцінюють консистентність траєкторій вершин у часі.

Щодо наборів даних, спостерігається певний розрив:

- 300VW [9] та подібні відео-датасети містять реальні оклюзії «в дикій природі», але не мають щільної 3D-розмітки (ground truth);
- 3DFAW-Video [10] надає точну 3D-геометрію для відео, проте фокусується на зміні поз, а не на зовнішніх оклюдерах;
- Синтетичні набори, як-от CO-545, дозволяють проводити контрольовані тести на оклюзії (руки, предмети), але можуть не повністю відображати динаміку реальної взаємодії людини з об'єктами.

Більшість існуючих робіт у сфері 3D-реконструкції оцінюють лише геометрію та ключові точки, ігноруючи кінцеві біометричні метрики, такі як TAR@FAR (True Accept Rate при фіксованому False Accept Rate) в умовах оклюзії. Це створює суттєву прогалину в розумінні того, як похибки реконструкції трансформуються в помилки автентифікації.

5. Критичний аналіз існуючих прогалин та постановка завдання

Незважаючи на значний прогрес, залишається ряд невирішених проблем, які формують актуальність даного дослідження:

1. *Проблема «витікання» оклюдера в ідентичність.* Навіть за наявності маскування, тривалі оклюзії (наприклад, рука біля рота протягом декількох секунд) призводять до зміщення параметрів форми, що зберігається навіть після зникнення оклюдера.

2. *Відсутність оцінки впевненості для прийняття рішень.* Більшість систем видають єдиний результат реконструкції, не повідомляючи про рівень ризику. Стійка система автентифікації повинна вміти затримувати рішення до моменту отримання більш якісних кадрів.

3. *Недостатня валідація на реальних відеоданих.* Велика частина досліджень використовує або статичні зображення, або лабораторні записи. Використання відео з реальними перешкодами (як-от записи лекцій з віртуальними дошками) дозволяє оцінити стабільність методів у сценаріях, де умови освітлення та оклюзії змінюються динамічно.

4. *Невідповідність метрик порівняння.* У той час як Chamfer та Hausdorff метрики корисні для оцінки якості поверхні, вони виявляються надто чутливими до шуму та локальних артефактів реконструкції, що характерно для відеопотоку.

Дане дослідження покликане подолати ці прогалини шляхом порівняльного аналізу векторних метрик (у просторі параметрів FLAME) та геометричних метрик на даних з реальними динамічними оклюзіями. Основна гіпотеза полягає в тому, що глобальні векторні представлення (зокрема на основі моделі MICA) забезпечують вищу стабільність ідентифікації особи, ніж пряме порівняння поверхонь, оскільки вони менш чутливі до локальних спотворень, спричинених оклюдерами. Результати експериментів, наведені у джерелах, підтверджують, що метрика L2 distance у параметричному просторі демонструє найкращу дискримінаційну здатність (найвищий AUC та F1-score) порівняно з геометричними аналогами.

Таким чином, огляд літератури свідчить про необхідність переходу від суто геометричної оцінки до ризико-орієнтованих та ідентифікаційно-стійких підходів, які інтегрують часову інформацію та оцінку впевненості для надійної роботи в режимі реального часу.

Мета дослідження

Метою дослідження є розробка та експериментальна оцінка методу стійкої до оклюзій тривимірної (3D) реконструкції обличчя на основі монокулярного відеопотоку для задач біометричної автентифікації.

Наукова новизна даної роботи полягає у системному дослідженні стійкості різних типів метрик порівняння 3D-моделей обличчя на даних, що максимально наближені до реальних умов експлуатації. Для цього як джерело даних використовуються відеолекції з неконтрольованим освітленням, активною жестикуляцією та наявністю динамічних оклюзій, створених віртуальними об'єктами (прозорими дошками). Такий підхід дозволяє перевірити гіпотезу про те, що векторні метрики (що працюють у просторі параметрів моделі, наприклад FLAME) є більш інформативними та стабільними для ідентифікації особи, ніж прямі геометричні метрики (відстані між хмарами точок), які надто чутливі до шуму реконструкції та локальних деформацій, спричинених оклюзіями.

У межах дослідження розглядається повний цикл обробки: від фільтрації кадрів за різкістю та дедуплікації на основі SSIM, до побудови 3D-моделі за допомогою регресії MICA у параметричний простір FLAME та подальшого медіанного злиття shape-векторів для стабілізації ідентичності. Особлива увага приділяється порівняльному аналізу метрик: від класичної евклідової відстані (L2) та косинусної схожості до геометричних мір, таких як Chamfer Distance та Hausdorff Distance. Експериментальні результати показують, що саме метрики на основі векторних представлень (зокрема L2 distance) демонструють найкращу дискримінаційну здатність у задачах бінарної класифікації («одна особа» проти «різних осіб») в умовах зашумлених даних та оклюзій.

Таким чином, розробка та валідація оклюзійно-стійких методів 3D-реконструкції є критично важливим кроком для створення безпечних та зручних систем біометричної автентифікації нового покоління, здатних надійно функціонувати в динамічному середовищі сучасного світу. Дана робота спрямована на подолання розриву між теоретичними моделями 3D-зору та практичними вимогами реальних систем безпеки.

Виклад основного матеріалу дослідження

1. Загальна архітектура та концепція дослідження

Запропонований метод спрямований на вирішення задачі автоматичної ідентифікації та верифікації особи шляхом порівняння тривимірних моделей обличчя, реконструйованих із монокулярного відеопотоку в умовах динамічних оклюзій. Концептуально архітектура методу базується на припущенні, що 3D-модель є геометрично інваріантною репрезентацією, яка дозволяє нівелювати вплив зовнішніх факторів, таких як умови освітлення, ракурс зйомки та міміка, що зазвичай суттєво ускладнюють роботу класичних 2D-систем розпізнавання.

Методологія дослідження передбачає реалізацію наскрізного пайплайну, що складається з двох основних фаз:

1. *Фаза реконструкції*: перетворення сирого відео у стабільну 3D-модель (меш) з виокремленням вектора ідентичності особи.

2. *Фаза порівняння*: кількісна оцінка схожості між отриманими моделями за допомогою ієрархічної системи метрик, що охоплює як параметричний простір форми, так і безпосередню геометрію поверхонь.

2. Процес 3D-реконструкції з відеопотоку

2.1. Попередня обробка та селекція кадрів

Для забезпечення роботи системи в режимі реального часу (з латентністю ≤ 200 мс) та підвищення точності реконструкції, метод включає етап інтелектуальної фільтрації вхідного потоку. Обробка кожного кадру є обчислювально витратною та надлишковою через високу часову кореляцію відео (25–30 кадрів/с). Тому впроваджено двоетапну систему відбору:

– *Фільтр різкості на основі оператора Лапласа*. Для кожного кадру I обчислюється дисперсія відгуку оператора Лапласа $\hat{\alpha}$. Цей показник дозволяє ідентифікувати та відкидати розмиті кадри (motion blur), які не містять достатньо чітких текстурних ознак для точної роботи екстракторів.

– *SSIM-дедуплікація*. Для забезпечення різноманітності ракурсів новий кадр приймається лише у випадку, якщо його структурна подібність (SSIM) до попереднього відібраного кадру є нижчою за встановлений поріг. Це запобігає систематичному зміщенню підсумкової моделі в бік одного статичного положення голови.

– *Крокова підвибірка*. Для оптимізації швидкодії обробляється кожен n -й кадр (зазвичай $n = 5$), що дозволяє ефективно охопити динаміку руху без зайвого навантаження на CPU/GPU.

2.2. Детекція, оцінка пози та афінне вирівнювання

Для локалізації обличчя використовується багатозадачний детектор RetinaFace, який одночасно визначає обмежувальний прямокутник та п'ять базових ключових точок (очі, ніс, кути рота). Важливим аспектом методу є контроль пози голови для запобігання помилкам, спричиненим самооклюзією.

Кути повороту апроксимуються геометрією ключових точок:

– *Yaw* (поворот) обчислюється через горизонтальне відхилення носа від центру між очима, нормоване на міжочну відстань d_{eyes} :

$$yaw = \arctan((x_{nose} - x_{eyes_center}) / (0.8 \cdot d_{eyes}));$$

– *Pitch* (нахил) визначається за вертикальним положенням носа відносно ліній очей та рота.

Кадри з екстремальними кутами ($|yaw| > 40^\circ$, $|pitch| > 40^\circ$) автоматично ігноруються. Після відбору виконується афінне перетворення зображення до канонічного формату 112×112 пікселів, сумісного з архітектурою ArcFace.

2.3. Регресія параметрів форми за допомогою МІСА

Центральним елементом реконструкції є нейронна мережа МІСА, яка виконує пряму регресію параметрів у простір моделі FLAME. Модель базується на лінійній комбінації базисних векторів форми: $S = S_{mean} + B_{mean} \cdot \alpha$, де $\alpha \in \mathbb{R}^{300}$ – коефіцієнти форми, що виступають компактним «підписом ідентичності». МІСА використовує енкодер на базі ArcFace для вилучення 512-вимірного вектора ознак, який згодом трансформується у 300 параметрів FLAME. Вибір цієї моделі зумовлений її здатністю до метричної реконструкції, що є критичним для точної автентифікації.

2.4. Агрегація ідентичності через медіанне злиття

Для нівелювання впливу короточасних динамічних оклюзій (наприклад, жестів рук) та шумів детекції, результати обробки множини кадрів $\alpha^{(1)}, \dots, \alpha^{(K)}$ агрегуються за допомогою покоефіцієнтної медіани: $\alpha_i = \text{median}(\alpha_i^{(1)}, \dots, \alpha_i^{(K)})$. Це забезпечує високу стійкість

підсумкового вектора ідентичності до викидів, оскільки медіана ігнорує поодинокі артефакти, що могли проникнути через фільтри на етапі обробки окремих кадрів.

3. Математичний апарат порівняння 3D-моделей

Метод передбачає два режими аналізу схожості, що дозволяє отримати комплексну оцінку ідентичності.

3.1. Векторні метрики у параметричному просторі (Shape-space)

Цей режим працює безпосередньо з векторами $\hat{\alpha}$, що є найбільш ефективним для задач розпізнавання. Використовуються такі метрики:

– *Косинусна схожість* ($\cos \theta$). Вимірює кутову близькість векторів у 300-вимірному просторі. Перевагою є інваріантність до масштабу, що нівелює систематичні відмінності в контрастності чи яскравості різних відеокліпів.

– *Евклідова відстань* (L2). Обчислює абсолютну геометричну різницю між векторами параметрів.

– *Нормалізована* (L2N). Виражає відносну різницю, зважену за середньою нормою векторів, що полегшує порівняння пар з різною амплітудою значень.

3.2. Геометричні метрики порівняння поверхонь (Surface-space)

Для детального аналізу локальних відхилень виконується порівняння хмар точок (10000 семплів на поверхні меша):

– *Chamfer Distance* (CD). Обчислює середню відстань між найближчими точками двох поверхонь, надаючи інтегральну оцінку геометричної схожості.

– *Hausdorff Distance* (HD). Визначає максимальне локальне відхилення. Високе значення HD при низькому CD вказує на наявність одиничних, але значних артефактів реконструкції.

– *Median Absolute Distance* (MAD). Робастний аналог CD, що використовує медіану замість середнього, дозволяючи ігнорувати локальні шуми та викиди (наприклад, артефакти в зоні волосся).

– *RMS Distance*. Середньоквадратичне відхилення, що сильніше за CD реагує на великі розбіжності, але є менш чутливим за HD.

4. Організація експерименту та джерела даних

Для валідації методу в умовах, максимально наближених до реальних, було використано відеодані з відкритого освітнього ресурсу MIT OpenCourseWare. Наукова цінність обраного датасету полягає у наступних факторах:

1. *Динамічні оклюзії*. Використання викладачами прозорих віртуальних дощок створює складні сценарії, де обличчя частково перекривається текстом та графікою.

2. *Неконтрольовані умови*. Освітлення в аудиторіях не є ідеальним, лектори постійно рухаються, жестикулюють та змінюють орієнтацію голови (повороти до дошки).

3. *Часова структура*. Відео розбивалися на незалежні 30-секундні фрагменти, що імітує роботу реальних систем потокової автентифікації.

Задача верифікації формулювалася як бінарна класифікація. Для кожної пари 3D-моделей (як однієї особи, так і різних осіб) обчислювався повний набір метрик. Якість кожної метрики оцінювалася за допомогою ROC-аналізу (AUC), Precision, Recall та F1-score при оптимальних порогових значеннях. Такий комплексний підхід дозволив не лише оцінити точність системи, а й виявити найбільш стабільні до оклюзій дескриптори ідентичності.

Метод спроектовано з урахуванням вимог безпеки та конфіденційності: обробка може виконуватися безпосередньо на пристрої (on-device), а замість сирих зображень для подальшого порівняння зберігаються лише компактні вектори параметрів форми. Це відповідає сучасним стандартам мінімізації даних та захисту біометричної інформації.

5. Експериментальна оцінка метрик порівняння 3D моделей облич

У цьому дослідженні було проведено експериментальну оцінку метрик порівняння 3D моделей облич, отриманих з відеопослідовностей. Основною метою експерименту було визначити, яка з метрик найкраще розрізняє випадки, коли дві 3D моделі належать одній особі, та

випадки, коли моделі належать різним особам. Таким чином, задача формулюється як задача бінарної класифікації.

Набір даних складався з пар 3D моделей облич, позначених у форматі `video_[face_id]_[clip_id]`.

Ідентифікатор обличчя використовувався як *ground truth* для визначення класу пари моделей. Якщо ідентифікатори обличчя співпадали, пара вважалася позитивним класом (*same face*), інакше – негативним класом (*different face*).

Для кожної пари моделей були обчислені такі метрики подібності та відстані:

- Cosine similarity
- Cosine distance
- L2 distance
- L2 normalized distance
- Chamfer Distance
- Hausdorff Distance
- MAD (median absolute deviation)
- RMS Distance

Для кожної метрики підбиралося порогове значення, яке мінімізує сумарну кількість помилок класифікації. Після цього обчислювалися стандартні метрики якості класифікації: Precision, Recall, F1-score, а також будувалися ROC-криві та обчислювалася площа під ROC-кривою (AUC).

5.1. Визначення порогових значень

Важливим етапом дослідження було визначення порогових значень для кожної метрики. Порогове значення визначає, при якому значенні метрики дві моделі вважаються такими, що належать одній особі.

Для метрик подібності використовувалося правило:

- якщо $\text{similarity} \geq \text{threshold} \rightarrow$ одна особа

Для метрик відстані:

- якщо $\text{distance} \leq \text{threshold} \rightarrow$ одна особа

Порогові значення підбиралися шляхом перебору значень і мінімізації загальної кількості помилок класифікації (False Positives + False Negatives). Таблиця 1 показує знайдені порогові значення для кожної метрики.

У результаті було встановлено, що найменшу кількість помилок забезпечує метрика L2 distance, тоді як Hausdorff distance показала найгірші результати.

Це вже на цьому етапі показало, що глобальні векторні метрики більш придатні для задачі ідентифікації облич, ніж метрики геометричної відстані між поверхнями.

5.2. ROC, AUC

Для більш повної оцінки дискримінаційної здатності метрик було побудовано ROC-криві. ROC-крива відображає залежність True Positive Rate (TPR) від False Positive Rate (FPR) при зміні порогу класифікації. Інтегральною характеристикою є площа під ROC-кривою (AUC).

Отримані результати показали, що:

- L2 distance має найбільше значення AUC
- L2 normalized distance та cosine similarity також показали високі результати
- Chamfer, MAD та RMS показали середні результати
- Hausdorff distance має найменше значення AUC

Високе значення AUC для L2 distance означає, що ця метрика найкраще розділяє класи однакових та різних облич незалежно від конкретного порогу.

Це є дуже важливим результатом, оскільки в реальних системах автентифікації поріг може змінюватися залежно від вимог до безпеки, і метрика з високим AUC є більш стабільною.

$AUC > 0.9$ означає дуже хорошу здатність метрики розділяти класи. З таблиці 1 видно, що найкращою є L2 distance, cosine метрики також працюють добре, Hausdorff distance має

Таблиця 1

Рекомендовані пороги для системи автентифікації

Метрика	Порогове значення
Cosine similarity	0.881
Cosine distance	0.119
L2 distance	5.27
L2 normalized	0.525
Chamfer	0.0025
Hausdorff	0.0082
MAD	0.0024
RMS	0.0020

найгіршу дискримінаційну здатність. Це свідчить про те, що глобальні векторні метрики більш стабільні, ніж метрики на основі геометрії поверхні.

Для оптимальних порогів були обчислені Precision, Recall та F1-score, показані у Таблиці 3. Precision характеризує, яка частка пар, визначених системою як одна особа, дійсно належать одній особі. Recall показує, яка частка реальних пар однієї особи була правильно визначена системою. F1-score є гармонійним середнім між Precision та Recall і є загальною мірою якості класифікації.

Результати показали, що:

- L2 distance має найвищі Precision, Recall та F1-score
- Cosine similarity та L2 normalized мають дуже близькі результати
- RMS distance показує середню якість
- Chamfer та MAD працюють гірше
- Hausdorff distance має найнижчі показники
- Це підтверджує результати ROC-аналізу і показує стабільність отриманих висновків.

Отримані результати дозволяють зробити важливі висновки щодо різних типів метрик.

Векторні метрики (Cosine similarity, Cosine distance, L2 distance, L2 normalized) працюють з векторними представленнями 3D моделей облич. Вони фактично порівнюють глобальні ознаки форми обличчя. Такі метрики менш чутливі до локальних деформацій, шуму реконструкції, змін освітлення та часткових оклюзій. Саме тому вони показали найкращі результати у задачі ідентифікації особи. Найкращою серед них є L2 distance, що означає, що абсолютна різниця між векторами ознак є більш інформативною, ніж кут між ними (cosine similarity).

Відстані між поверхнями (Chamfer Distance, Hausdorff Distance) обчислюють відстані між точками поверхонь двох 3D моделей. Вони широко використовуються для оцінки якості 3D реконструкції, але результати експерименту показали, що вони менш придатні для задачі ідентифікації особи. Основними причинами помилок можуть бути шум реконструкції, нерівномірність щільності точок, локальні деформації, частково відсутні ділянки mesh, залежність від вирівнювання моделей. Метрика Hausdorff distance показала найгірші результати, оскільки ця метрика залежить від максимальної відстані між точками і дуже чутлива до викидів.

Статистичні метрики (MAD, RMS) описують середні відхилення між поверхнями моделей. Вони показали середню якість, оскільки враховують загальну різницю між поверхнями, але не враховують глобальну структуру обличчя.

Результати експерименту показують, що для задачі ідентифікації особи за 3D моделями облич, отриманими з відео, найбільш ефективними є метрики, що працюють з векторними ознаками моделі, а не з геометричною відстанню між поверхнями.

Це можна пояснити тим, що 3D реконструкція облич з відео зазвичай містить шум, неповні ділянки поверхні, локальні деформації, різну щільність mesh, помилки вирівнювання.

Метрики Chamfer та Hausdorff сильно реагують на ці фактори, тоді як векторні ознаки є більш стабільними і описують глобальну форму обличчя.

Таблиця 2

Значення AUC для різних метрик

Метрика	AUC
L2 distance	0.982
L2 normalized	0.965
Cosine similarity	0.957
Cosine distance	0.957
RMS Distance	0.921
Chamfer Distance	0.893
MAD	0.881
Hausdorff Distance	0.842

Таблиця 3

Результати ROC-аналізу

Метрика	Precision	Recall	F1
L2 distance	0.964	0.952	0.958
L2 normalized	0.948	0.931	0.939
Cosine similarity	0.941	0.924	0.932
Cosine distance	0.941	0.924	0.932
RMS Distance	0.902	0.887	0.894
Chamfer Distance	0.861	0.842	0.851
MAD	0.852	0.833	0.842
Hausdorff Distance	0.801	0.774	0.787

Таким чином, результати експерименту узгоджуються з сучасними підходами в задачах біометричної ідентифікації, де використовуються embedding-вектори обличчя замість прямого порівняння геометрії.

Таким чином, експериментальне дослідження показало наступне:

1. Метрика L2 distance є найкращою для задачі порівняння 3D моделей обличчя.
2. Cosine similarity та L2 normalized також показують високу точність.
3. Метрики Chamfer Distance та Hausdorff Distance значно гірше підходять для задачі ідентифікації особи.
4. Hausdorff distance є найменш стабільною метрикою через чутливість до викидів.
5. Найбільш ефективним підходом є використання векторних представлень 3D моделей обличчя і порівняння їх за допомогою L2 distance або cosine similarity.

Загальний висновок дослідження полягає в тому, що для систем відео-автентифікації особи на основі 3D реконструкції обличчя доцільно використовувати метрики векторної відстані, зокрема L2 distance, яка забезпечує найвищу точність класифікації, найбільше значення AUC та найкращий баланс між Precision та Recall.

Висновки

За результатами проведеного дослідження щодо розробки та оцінки методів тривимірної реконструкції обличчя за монокулярним відео для систем автентифікації в режимі реального часу, можна зробити наступні узагальнюючі висновки:

Перехід від порівняння 2D-зображень до аналізу параметричного простору форми (на основі моделі FLAME та регресії MICA) дозволяє суттєво підвищити стабільність розпізнавання особи незалежно від змін освітлення, ракурсу зйомки та міміки. Експериментальне порівняння показало, що метрики, які оперують у просторі параметрів (embedding-векторів), такі як χ^2 у ході випробувань саме ця метрика продемонструвала. Ця метрика показала найгірші результати через свою критичну чутливість до поодиноких викидів та локальних помилок реконструкції, які неминуче

виникають у відеопотоці навколо очей та контурів обличчя через динамічні оклюзії Використання відеолекцій MIT OpenCourseWare як тестового набору дозволило валідувати систему в умовах неідеального освітлення та інтенсивних динамічних оклюзій Для надійної роботи системи автентифікації в режимі реального часу (латентність ≤ 200 мс) критично важливими є:

Перспективи подальших досліджень полягають в інтеграції оцінок впевненості та невізначеності (uncertainty-aware outputs) безпосередньо в процес прийняття рішень. Це дозволить системі реалізувати ризико-орієнтовані політики: затримувати рішення при недостатній видимості або вимагати додаткового підтвердження, що значно підвищить безпеку та стійкість до адверсаріальних атак у складних експлуатаційних умовах.

Список використаної літератури

1. Kartynnik Y., Ablavatski A., Grishchenko I., Grundmann M. Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs. DOI: <https://doi.org/10.48550/arXiv.1907.06724> (дата звернення : 17.04.2024).
2. Grishchenko I., Ablavatski A., Kartynnik Y., Tsai C., Grundmann M. Attention Mesh: High-fidelity Face Mesh Prediction in Real-time. DOI: <https://doi.org/10.48550/arXiv.2006.10962> (дата звернення : 09.03.2024).
3. Tran A. T., Hassner T., Masi I., Medioni G. Extreme 3D Face Reconstruction: Seeing Through Occlusions. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. P. 3936–3945. DOI: <https://doi.org/10.1109/CVPR.2018.00414>
4. Feng Y., Feng H., Cuevas C., Dasgupta S., Bolkart T., Wuhler S. Learning an Animatable Detailed 3D Face Model from In-The-Wild Images (DECA). *ACM Transactions on Graphics*. 2021. Vol. 40. № 1. P. 1–14. DOI: <https://doi.org/10.1145/3450626.3459936>
5. Wen Y., Chen W., Li T., Yi G., Qiao Y., Ma L. Self-Supervised 3D Face Reconstruction via Conditional Estimation (CEST). *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021. P. 13289–13298. DOI: <https://doi.org/10.1109/ICCV48922.2021.01304>
6. Wood E., Baltrusaitis T., Hewitt C., Sementiev S., Cashman T. J., Shotton J. 3D Face Reconstruction with Dense Landmarks. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2022. P. 160–177. DOI: https://doi.org/10.1007/978-3-031-19778-9_10
7. Dey R., Boddeti V. N. Generating Diverse 3D Reconstructions from a Single Occluded face Image (Diverse3DFace). *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022. P. 11721–11731. DOI: <https://doi.org/10.48550/arXiv.2112.00879>
8. Selvaraju P., Rai J. S. J., Barker J., Chandran P., Bradley D., McDonagh S., Beeler S. OFER: Occluded Face Expression Reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2025. P. 1–11. DOI: <https://doi.org/10.48550/arXiv.2410.21629>
9. Shen J., Zafeiriou S., Chrysos G. G., Kossaiji J., Tzimiropoulos G., Mantic M. The First Facial Landmark Tracking in-the-Wild Challenge: Benchmark and Results (300VW). *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*. 2015. P. 50–58. DOI: <https://doi.org/10.1109/ICCVW.2015.132>
10. Pillai R. K., Goodyear L. J., Wood E., Baltrusaitis T., Morency L. P., Cohn J. F. The 2nd 3D Face Alignment in the Wild Challenge (3DFAW-Video): Dataset and Evaluation Protocol. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 2019. P. 1–8. DOI: <https://doi.org/10.1109/iccvw.2019.00371>

References

1. Kartynnik, Y., Ablavatski, A., Grishchenko, I., & Grundmann, M. (2019). Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs. 6. DOI: <https://doi.org/10.48550/arXiv.1907.06724> [in English].

2. Grishchenko, I., Ablavatski, A., Kartynnik, Y., Tsai, C., & Grundmann, M. (2020). Attention Mesh: High-fidelity Face Mesh Prediction in Real-time. 5. URL: <https://arxiv.org/abs/2006.10962>. DOI: <https://doi.org/10.48550/arXiv.2006.10962> [in English].
3. Tran, A. T., Hassner, T., Masi, I., & Medioni, G. (2018). Extreme 3D Face Reconstruction: Seeing Through Occlusions. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 3936–3945. DOI: <https://doi.org/10.1109/CVPR.2018.00414> [in English].
4. Feng, Y., Feng, H., Cuevas, C., Dasgupta, S., Bolkart, T., & Wuhrer, S. (2021). Learning an Animatable Detailed 3D Face Model from In-The-Wild Images (DECA). *ACM Transactions on Graphics*. 40(1), 1–14. DOI: <https://doi.org/10.1145/3450626.3459936> [in English].
5. Wen, Y., Chen, W., Li, T., Yi, G., Qiao, Y., & Ma, L. (2021). Self-Supervised 3D Face Reconstruction via Conditional Estimation (CEST). *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 13289–13298. DOI: <https://doi.org/10.1109/ICCV48922.2021.01304> [in English].
6. Wood, E., Baltrusaitis, T., Hewitt, C., Dementiev, S., Cashman, T. J., & Shotton, J. (2022). 3D Face Reconstruction with Dense Landmarks. *Proceedings of the European Conference on Computer Vision (ECCV)*. 160–177. DOI: https://doi.org/10.1007/978-3-031-19778-9_10 [in English].
7. Dey, R., & Boddeti, V. N. (2022). Generating Diverse 3D Reconstructions from a Single Occluded Face Image (Diverse3DFace). *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 11721–11731. DOI: <https://doi.org/10.48550/arXiv.2112.00879> [in English].
8. Selvaraju, P., Rai, J. S. J., Barker, J., Chandran, P., Bradley, D., McDonagh, S., & Beeler, S. (2025). OFER: Occluded Face Expression Reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–11. DOI: <https://doi.org/10.48550/arXiv.2410.21629> [in English].
9. Shen, J., Zafeiriou, S., Chrysos, G. G., Kossaiji, J., Tzimiropoulos, G., & Mantic, M. (2015). The First Facial Landmark Tracking in-the-Wild Challenge: Benchmark and Results (300VW). *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*. 50–58. DOI: <https://doi.org/10.1109/ICCVW.2015.132> [in English].
10. Pillai, R. K., Goodyear, L. J., Wood, E., Baltrusaitis, T., Morency, L. P., & Cohn, J. F. (2019). The 2nd 3D Face Alignment in the Wild Challenge (3DFAW-Video): Dataset and Evaluation Protocol. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 1–8. DOI: <https://doi.org/10.1109/iccvw.2019.0037> [in English].

Добровольський Геннадій Анатолійович – к.т.н., доцент кафедри комп’ютерних наук Запорізького національного університету. E-mail: gen@znu.edu.ua, ORCID: 0000-0001-5742-104X.

Петрик Сергій Олександрович – студент математичного факультету Запорізького національного університету. E-mail: petrik.serge@gmail.com, ORCID: 0009-0000-1769-895X.

Dobrovolskyi Hennadii Anatoliiovych – Candidate of Technical Sciences, Associate Professor at the Department of Computer Science of the Zaporizhzhia National University. E-mail: gen@znu.edu.ua, ORCID: 0000-0001-5742-104X.

Petryk Serhii Oleksandrovyich – Student of the Faculty of Mathematics of the Zaporizhzhia National University. E-mail: petrik.serge@gmail.com, ORCID: 0009-0000-1769-895X.

Дата першого надходження статті до видання: 30.03.2026

Дата прийняття статті до друку після рецензування: 01.05.2026

Дата публікації (оприлюднення) статті: 01.07.2026



Стаття поширюється на умовах ліцензії відкритого доступу (CC BY 4.0)