

УДК 004.4 / C88

[0000-0001-5657-9144] **О. Б. Данченко¹**, д.т.н., доцент,

[0000-0002-7435-3533] **О. Є. Іларіонов²**, к.т.н., доцент,

oleg.ilarionov@knu.ua

[0000-0003-1986-6130] **Г. В. Красовська²**, к.т.н., доцент,

Т. С. Короткова², магістрант

¹ Черкаський державний технологічний університет
б-р Шевченка, 460, м. Черкаси, 18006, Україна

² Київський національний університет імені Тараса Шевченка
вул. Б. Гаврилишина, 24, м. Київ, 04118, Україна

РОЗРОБКА СИСТЕМИ РОЗПІЗНАВАННЯ ОБЛИЧЧЯ ЛЮДИНИ У ВІДЕОПОТОЦІ З ДОПОВНЕНОЮ РЕАЛЬНІСТЮ

Розглянуто проблему розпізнавання обличчя людини у відеопотоці з доповненою реальністю, досліджено сучасний стан питання. Вивчено загальний процес розпізнавання обличчя та основні поняття доповненої реальності. Проведено аналіз сучасних підходів до розв'язання задачі розпізнавання обличчя, виявлено сильні та слабкі сторони використовуваних методів. Проведено пошук методу, інваріантного до масштабування, зміни сцени, поворотів голови, зміни освітленості, аксесуарів та зміни емоцій. Розроблено алгоритм, архітектуру та саму програмну систему, що розв'язує задачу розпізнавання обличчя людини у відеопотоці з доповненою реальністю. Методом для виявлення обличчя обрано гістограму нахилених градієнтів (від англ. "Histogram of Oriented Gradients", HOG), розпізнавання обличчя розроблено на основі згорткової нейронної мережі архітектури ResNet34. Проведено експериментальні дослідження, систему протестовано як на одному, так і на кількох обличчях одночасно. Визначено оцінки якості розпізнавання розробленої програмної системи – побудова ROC-кривих, які показують залежність кількості помилкових спрацьовувань алгоритму розпізнавання (false positive) від точності розпізнавання (true positive rate), та обчислення AUC (від англ. "Area Under the Curve"). При розпізнаванні одного обличчя досягнуто значення AUC 0,95, при розпізнаванні кількох обличч (максимум чотирьох) – 0,83.

Ключові слова: розпізнавання обличчя, доповнена реальність, попередня обробка зображення, виявлення обличчя, виявлення ключових ознак.

Вступ. Технологія розпізнавання обличчя використовується у багатьох галузях: безпека, медицина, розважальна індустрія і т. д. Багато з нас щодня користуються цією технологією (наприклад, щоб розблокувати свій смартфон). Проте розпізнавання обличчя, поєднане з технологіями доповненої реальності, може встановити нову планку, наприклад, використовувати Augmented Reality (AR) (з англ. – «доповнена реальність») і Artificial Intelligence (AI) (з англ. – «штучний інтелект»), щоб поєднати зображення людського обличчя з камери та інформацію про реальну людину. Можливості AR-технологій дуже великі – вони можуть знайти застосування чи не в усіх сферах нашого життя. Доповнена реальність збагатить та покращить якість комунікації, отримання інформації та обміну нею, ведення підприємницької діяльності. Віртуальний екскурсовод проведе нас по руїнах замку, та

ще й покаже сценку, як саме цей замок розвалили, і яким він був до того. Мобільний додаток з AR допоможе позбутися паперових інструкцій, провівши інтерактивне заняття з користування щойно придбаним предметом. Віртуальна примірочна одягу, аксесуарів та макіяжу дасть можливість легше створити новий образ. Щоб алгоритм візуального відстеження був корисним у реальних сценаріях, він повинен бути в змозі обробляти та долати випадки, коли зовнішній вигляд об'єкта змінюється від кадру до кадру. Значна і швидка зміна зовнішнього вигляду внаслідок шуму, оклюзії, фонові перешкоди, пози, масштабу та змін освітленості є основними проблемами, які потрібно подолати детектору [1]. Саме тому важливо розуміти, який із методів розпізнавання буде найбільш ефективним для тієї чи іншої задачі

Мета дослідження. Сьогодні існує багато методів розпізнавання облич. При їх розробці бралися до уваги вимоги щодо розміру сформованих біологічних шаблонів облич, швидкості їх формування й порівняння та якості розпізнавання як основного критерію. Є кілька основних підходів. Наразі метрика Хеммінга для порівняння бінарних шаблонів є найшвидшою [2-3]. Проте, у свою чергу, кращі показники якості в розпізнаванні облич демонструють глибокі згорткові нейронні мережі, що використовують небінарні біометричні шаблони [4-5].

Раніше було запропоновано ряд підходів до формування біометричних шаблонів зображень облич, які можуть використовуватися для біометричної верифікації або ідентифікації осіб [6-7]. Однак практично всі останні результати в розпізнаванні осіб було отримано за допомогою використання глибоких згорткових нейронних мереж [8]. Існує багато підходів до навчання мереж для розпізнавання людей – мережі можуть бути навчені як за класифікацією, коли кожній особі відповідає свій клас, так і за принципом сіамських мереж на основі попарного порівняння для найкращої верифікації. Також є реалізації, що об'єднують обидва підходи. Найкращі сучасні результати на базі облич Labeled Faces in the Wild (LFW) отримано композицією глибоких мереж, навчених на різних фрагментах (патчах) осіб [9]. Проте використання об'єднання кількох архітектур нейронних мереж вимагає значно більших обчислювальних ресурсів, у той час як деякі однопатчеві мережі можуть бути досить ефективними. Як біометричний шаблон зазвичай використовується вихідний сигнал одного з шарів, найчастіше це останній прихований шар у мережі. Для зниження розмірності отриманих шаблонів зазвичай використовується метод головних компонент, а порівняння шаблонів здійснюється за допомогою косинусної або L2-відстані [9].

У роботі Ю. В. Візілтера, В. С. Горбачевича, А. В. Воротнікова, Н. А. Костромова [8] на основі згорткової мережі з хешуючим шаром було отримано згорткову нейронну мережу з хешуючим лісом (ЗНМХЛ). Ця модифікація дає можливість змінювати розмір шаблону, що використовується, від надкомпактних (десятки байтів) до великих (сотні байтів) шаблонів. Відмінність у них буде в якості розпізнавання. Це дало змогу будувати

компактні біометричні шаблони (250 байт) з одночасно покращеними (порівняно з базовою мережею) показниками верифікації (98,14 %) та ідентифікації (91 % rank-1) на базі осіб LFW, а також ультракомпактні (25 байт) шаблони. З другого боку, описаний підхід дає можливість будувати великі біометричні шаблони, що дають змогу досягати високої якості розпізнавання – ЗНМХЛ з 2000 кодуючими 7-бітними деревами досягає 98,59 % точності верифікації 93 % rank-1 на LFW (надбавка більш ніж у 3 % в rank-1 щодо вихідної базової архітектури мережі).

У роботі J. Betty, I. Bülthoff, B. J. Mohler, I. M. Thorntonb [10] було досліджено, як впливають на якість розпізнавання різні фактори. У результаті досліджень було отримано три основні висновки:

- ефективність розпізнавання є найкращою, коли спостерігач безпосередньо взаємодіє з цільовим об'єктом, досліджуючи його з різних сторін;

- якість розпізнавання покращується, коли спостерігається зміна кута погляду по вертикалі на об'єкт дослідження. Ця маніпуляція не так добре досліджена, як типові повороти по горизонталі, але в цьому випадку спостерігається, що витрати ресурсів збільшуються при взаємодії з людьми у сидячому положенні або стоячи на якомусь підвищенні;

- віртуальний сценарій можна використовувати для вивчення систем розпізнавання обличчя.

Метою статті є аналіз підходів до побудови технології розпізнавання облич у відеопотоці в системах з доповненою реальністю шляхом вивчення моделей, методів та алгоритмів розпізнавання обличчя людини, виявлення сильних та слабких сторін існуючих рішень, вибору найкращого сполучення методів виявлення та розпізнавання облич та доведення ефективності запропонованого підходу на основі гістограми направлених градієнтів та згорткової нейронної мережі архітектури ResNet34 для задачі розпізнавання облич у відеопотоці в системах з доповненою реальністю.

Виклад основного матеріалу

Основні поняття доповненої реальності. Доповнена реальність – це технологія, що розміщує в навколишньому світі графічні або текстові об'єкти у режимі реального часу за допомогою комп'ютерних пристроїв [11].

Необхідно розрізняти доповнену, віртуальну та змішану реальності. У доповненій реальності віртуальні об'єкти переносяться в реальний світ, доповнюють дійсне середовище. У віртуальній реальності світ не є справжнім, хоча й може бути схожим на нього. Він створюється за допомогою комп'ютерних систем, сприйняття якого наразі забезпечує сенсорна система людини (зорові, тактильні й інші відчуття). Змішана, або гібридна реальність поєднує ці поняття. Так, світ, у який занурюється людина, створюється за допомогою VR (з англ. "Virtual Reality" – віртуальна реальність), а додання віртуальних елементів забезпечується AR. Отже, віртуальна реальність взаємодіє лише з користувачами, а доповнена реальність – з усім зовнішнім світом [12].

Створення доповненої реальності у загальному випадку відбувається наступним чином: пристрій, оснащений доповненою реальністю (спеціальні окуляри, смартфон тощо), використовуючи камеру, захоплює об'єкт фізичного світу; за допомогою програмного забезпечення пристрою у режимі реального часу проводиться розпізнавання об'єкта, накладання відповідної цифрової інформації та виведення кінцевого зображення на пристрій візуалізації.

Основні поняття процесу розпізнавання обличчя. Загальний процес розпізнавання обличчя складається з виявлення обличчя, відстеження, вилучення ознак, зменшення розмірів та відповідного розпізнавання (рисунк 1).

Більш детальний опис цього процесу викладено нижче зі специфікою застосування до розробленої системи.

Опис програмного застосування. Для поставленої задачі було розроблено наступну цільову функцію (1):

$$\sum_{i=0}^k BestMatch_i \left(Descr \left(Detec(T_{features}) \right), \right. \\ \left. Descr \left(Detec(E_{features}) \right) \right) \rightarrow \max, \\ BestMatch_i = \{ \overline{Match}_{t,e} \mid t.dist < 0,7 \cdot e.dist, t = \overline{0, n-1}, \\ e = \overline{0, n-1} \} \quad (1)$$

де T – тестове зображення, E – еталонне зображення, $T_{features}$ і $E_{features}$ – особливості T і E відповідно, $Detec()$ – детектор, $Descr()$ – дескриптор, $\overline{Match}_{t,e}$ – вектор відповідностей між дескрипторами зображень T та E , $BestMatch_i$ – кількість відповідностей між дескрипторами, що задовольняють обмеженню, $dist$ – відстань між дескрипторами.

Вхідними даними розробленої програми є база даних обличч людей та інформації про них. У результаті виконання програми знаходиться найкращий збіг обличчя, що відображається у відеопотоці, з обличчям із бази даних та відтворення інформації про цю людину на відео шляхом використання доповненої реальності.

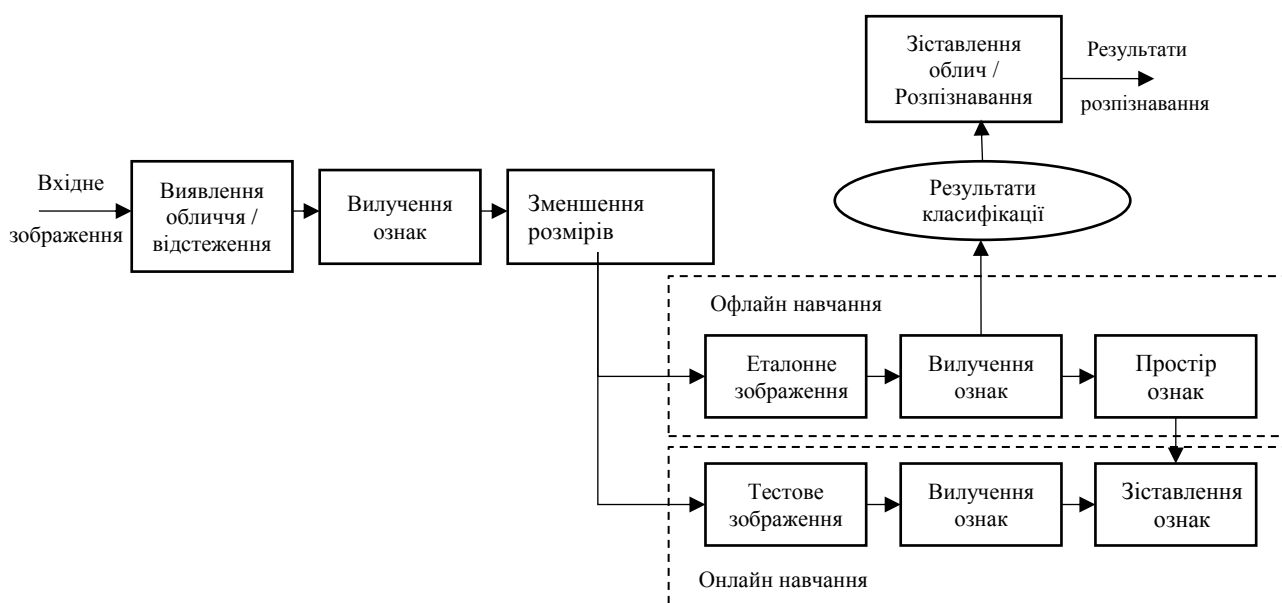


Рисунок 1 – Загальний процес розпізнавання обличчя людини

Для розпізнавання обличчя у програмі використовується наступний алгоритм:

1. Знайти всі обличчя на зображенні з відеопотоку.
2. Розпізнати кожне обличчя, навіть якщо воно знаходиться під кутом або під поганим освітленням.
3. Визначити унікальні риси обличчя, що відрізняють одну людину від іншої, наприклад форма обличчя, розмір очей тощо.
4. Порівняти знайдені особливості цього обличчя з усіма людьми, обличчя яких система вже знає, щоб зрозуміти, хто зараз зображений на відео.

З'єднавши цей ланцюг дій, ми отримаємо розв'язок поставленої задачі (рисунк 2).

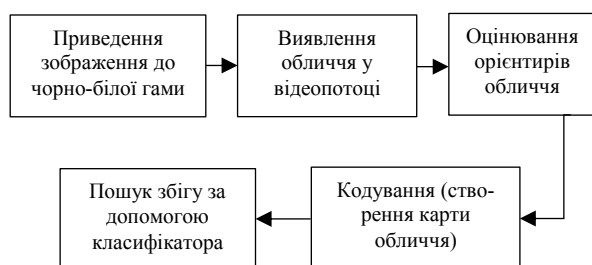


Рисунок 2 – Загальний алгоритм розробленої програми

Розглянемо кожний крок детальніше.

Крок 1. Пошук усіх облич

На першому етапі алгоритму потрібно знайти всі обличчя. Для цього було використано гістограму направлених градієнтів (від англ. “Histogram of Oriented Gradients”, HOG).

Спочатку зображення приводиться до чорно-білої гами, тому що для пошуку обличчя дані про колір є надлишковими. Потім кожний піксель розглядається окремо і порівнюється з сусідніми. Мета цих дій – виявити, наскільки темним є поточний піксель порівняно з сусідніми. Потім малюється стрілка, що вказує, в якому напрямку зображення стає темнішим (градієнт). Таким чином, алгоритм отримує градієнтне зображення обличчя. Для кожної людини напрямок зміни кольору на обличчі від темного до світлого буде однаковим при будь-якому освітленні.

Крок 2. Оцінювання орієнтирів обличчя

Після знаходження обличчя необхідно розібратися з проблемою, коли обличчя, повернуті у різні сторони, здаються комп'ютеру різними людьми.

Для розв'язання цієї задачі кожне зображення перетворюється так, щоб очі і губи

завжди знаходилися в якомусь визначеному місці. Це спрощує задачу порівняння облич на наступних кроках. Для цього використовується алгоритм оцінювання орієнтирів обличчя (від англ. – “face landmarks estimation”).

Основна ідея полягає в тому, що ми відмічаємо 68 особливих точок (які називаються орієнтирами), що існують на кожному обличчі, – верхня частина підборіддя, зовнішня точка кожного ока, внутрішня точка кожної брови тощо. Після цього алгоритм машинного навчання знаходить ці 68 точок на кожному обличчі.

Коли відомо, де очі, а де рот, ми можемо повертати, масштабувати і зсувати зображення так, щоб очі та рот були якомога краще відцентровані. Для цього було використано афінні перетворення.

Крок 3. Кодування обличчя

Потрібно знайти спосіб зняти кілька основних вимірів з кожного обличчя, які можна було б порівняти з найближчими відомими вимірами та знайти найбільш схоже обличчя. Наприклад, можна змінити розмір кожного вуха, відстань між очима, довжину носа тощо.

Проблема в тому, що виміри, які здаються людям очевидними (наприклад колір очей), насправді не мають сенсу для комп'ютера, що роздивляється окремі пікселі зображення. Один із підходів – дозволити комп'ютеру самому зробити виміри того, що йому потрібно. Алгоритми глибинного навчання самостійно визначають, які частини обличчя потрібно вимірювати, а які – ні.

Розв'язок полягає у створенні згорткової нейронної мережі, яка може створювати 128 вимірів для кожного обличчя. 128 вимірів кожного обличчя називають картою. Ідея перетворення масиву необроблених даних, наприклад зображення, в список генерованих комп'ютером чисел вкрай важлива для машинного навчання.

Процес навчання згорткової нейронної мережі для отримання мапи облич (кодування обличчя) потребує великої кількості даних і великих обчислювальних потужностей. Саме тому було використано модель, навчену на кількох наборах даних упродовж кількох тижнів. Таким чином, залишається тільки пропустити зображення обличчя через готову навчену нейронну мережу, щоб отримати 128 вимірювань для кожного обличчя.

Крок 4. Пошук інформації людини за кодуванням

На останньому етапі потрібно знайти людину в базі відомих для розробленої програми людей, виміри обличчя якої найближчі до отриманих даних у реальному часі. Це робиться за допомогою класифікатора, що вимірює відстань між значеннями і визначає, наскільки вони схожі.

Для розпізнавання обличчя було обрано згорткову нейронну мережу архітектури ResNet34 [Residual Network (дослівно — «залишкова мережа»)]. ResNet має наступні переваги над базовими нейронними мережами, цим самим усуваючи їх основні недоліки [13]:

1. ResNet – глибинна нейронна мережа.

Згідно з загальною теоремою апроксимації ми знаємо, що мережа прямого поширення з одним шаром може відтворити будь-яку функцію. Проте шар має бути великим, через що нейронна мережа може бути схильною до перенавчання, що дасть погані результати. Для розв'язання цієї проблеми мережева архітектура має бути більш глибинною. ResNet може бути побудована як зі 100, так і з 1000 шарів.

2. ResNet долає проблему зникнення градієнта.

Глибинні мережі важко тренувати через проблему зникнення градієнта – оскільки градієнт повертається назад до більш ранніх шарів, повторне множення може зробити градієнт нескінченно малим. У результаті, оскільки мережа стає глибшою, її продуктивність стає насиченою або швидко починає знижуватись.

Завдяки наявності залишкових блоків (рисунок 3, [13]) ResNet мережа може ігнорувати градієнт деяких підмереж і просто пересувати градієнт від вищих шарів до нижніх шарів без будь-якої модифікації. Це означає, що ResNet може просто передати градієнт з останнього шару, наприклад шару 151, безпосередньо до першого шару.

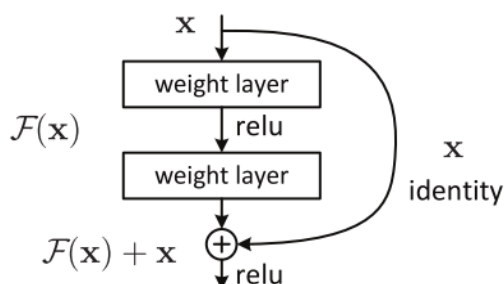


Рисунок 3 – Залишковий блок нейромережі ResNet

Для реалізації алгоритму обрано високорівневу мову програмування загального призначення Python. Для розпізнавання обличчя було використано бібліотеку face_recognition. Вона надає зручний API, який, у свою чергу, використовує бібліотеки OpenFace, dlib, OpenCV. Також ця бібліотека представляє методи роботи з якісно навченими моделями для визначення face landmarks і нейронною мережею ResNet34.

Спроектвану архітектуру системи розпізнавання обличчя у відеопотоці з доповненою реальністю зображено на рисунку 4.

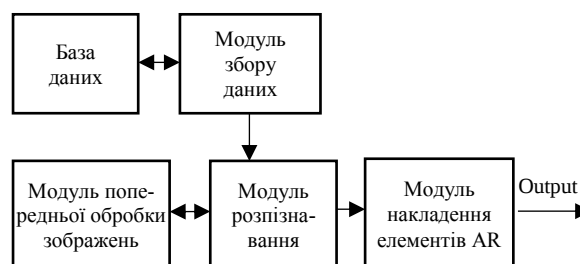


Рисунок 4 – Архітектура системи розпізнавання обличчя у відеопотоці з доповненою реальністю

Розглянемо мережу ResNet34 (нейромережа на архітектурі ResNet з 34 шарами) на рисунку 5 [13].

У першому шарі ResNet використовує згортку 7×7 з кроком 2 для зменшення вхідного сигналу у два рази (від англ. “downsampling”), подібно шару агрегування (від англ. “pooling”). Потім йдуть три блоки ідентичності (або залишкові шари), після чого кількість пікселів вхідного зображення знову зменшується у два рази. Шар семплінгу (зменшення зображення) також є згортковим шаром, але не з’єднується з блоком ідентичності. Ці блоки архітектури повторюються аналогічним чином кілька разів.

Останній шар – це шар усередненого агрегування (від англ. “average pooling”), який створює 1000 карт властивостей (для даних ImageNet) і середнє значення для кожної карти об’єктів. Результатом буде 1000-мірний вектор, який потім подається безпосередньо в шар Softmax, тому він повнозв’язний (від англ. “fully connected”) [13]. Основне призначення функції softmax полягає в перетворенні (ненормалізованого) виводу одиниць K (який, наприклад, представлений у вигляді вектора

К елементів) повнозв'язного шару у розподіл ймовірностей (нормалізований вихід), який часто представлений як вектор К елементів, кожен з яких знаходиться між 0 та 1 (ймовірність), а сума усіх цих елементів дорівнює 1 (розподіл ймовірності).

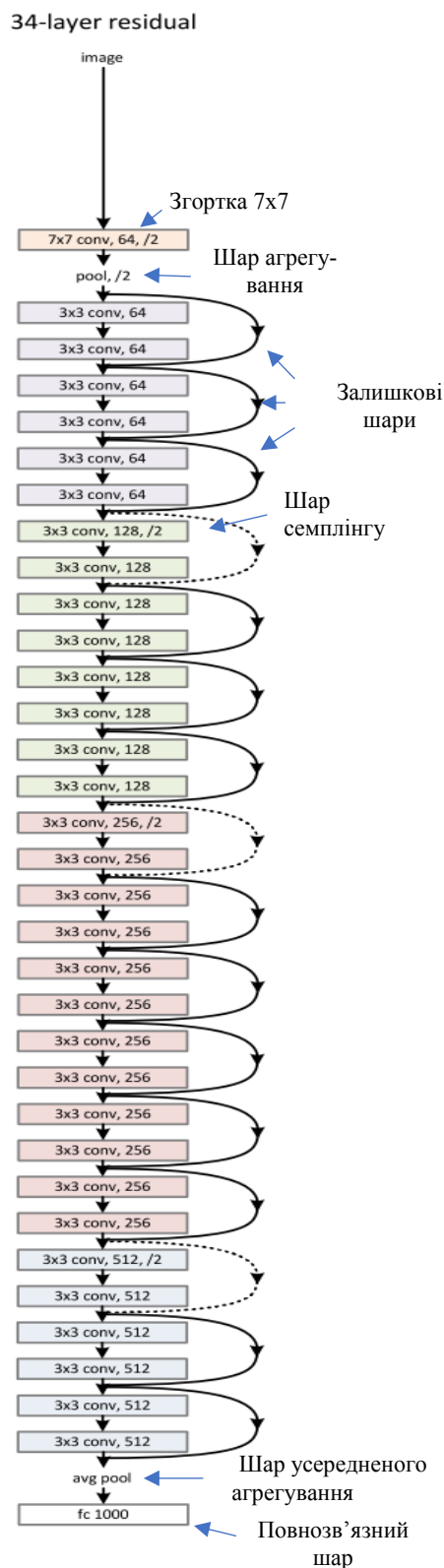


Рисунок 5 – Архітектура ResNet з 34 шарами

У випадку задачі класифікацією i -й елемент вектора, що створюється функцією softmax, відповідає ймовірності входу мережі, що належить до i -го класу.

Результати досліджень і їх обговорення. Розроблена програмна система, призначена для розпізнавання обличчя людини у відеопотоці з доповненою реальністю, показала прийнятні результати.

Додаток правильно розпізнає обличчя, наявне у базі даних, в умовах різного освітлення, повороту голови, за наявності аксесуарів, закриття деяких частин обличчя, зміни емоцій тощо, аналогічно – для розпізнавання кількох обличч одночасно. Система тестувалася на 520 прикладах: чотири людини окремо та разом у різних комбінаціях за різних умов освітлення, шумів, перешкод, аксесуарів, емоцій.

Для оцінювання якості детектування за допомогою розробленої програмної реалізації використовується показник, що дорівнює відношенню площі перетину прямокутника, отриманого в результаті детектування, і прямокутника з розмітки до площі їх об'єднання (від англ. "Intersection over Union", IoU). Таким чином, вважається, що обличчя виявлено правильно, якщо цей показник перевищує певний поріг. В іншому випадку приймається, що обличчя не виявлено. Спрацьовування алгоритму детектування на області, де обличчя відсутнє, вважається помилковим спрацьовуванням.

На підставі наведеного показника здійснюється побудова ROC-кривої (рисунки 6, 7), яка показує залежність кількості помилкових спрацьовувань алгоритму детектування (false positive) від точності детектування (true positive rate).

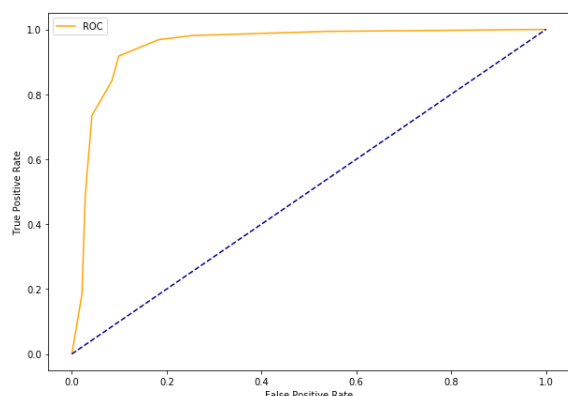


Рисунок 6 – ROC-крива при розпізнаванні одного обличчя, AUC=0,95

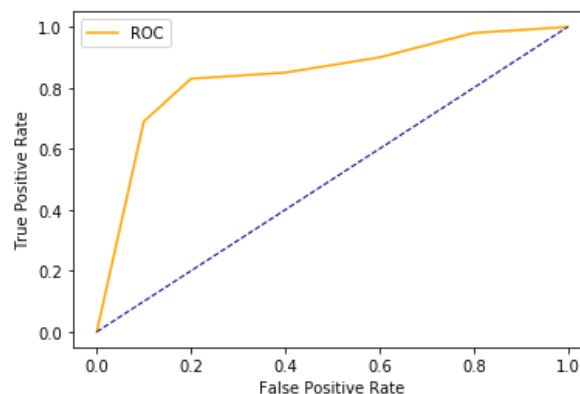


Рисунок 7 – ROC-крива при розпізнаванні кількох обличч (максимум чотирьох), AUC=0,83

Висновки. Застосування нейромережі на архітектурі ResNet з відповідними налаштуваннями для детектування та розпізнавання обличчя людини у відеопотоці з доповненою реальністю є хорошим вибором – цей метод інваріантний до масштабування, зміни сцени, поворотів голови, зміни освітленості, аксесуарів та зміни емоцій.

Експерименти показують стабільні результати. За перешкод, а саме, коли закривається частина обличчя, яка є визначальною для розпізнавання, програма не може визначити, кого зображено, тому що не може порівняти особливі ознаки наявного зображення обличчя та обличчя з бази даних. Система є платформою для подальших розробок. Зокрема, планується проведення експериментальних досліджень з використанням інших методів розпізнавання обличчя у відеопотоці та виконання порівняльного аналізу отриманих результатів, а також створення більш зручного графічного інтерфейсу програми та адаптація для мобільної версії.

Список використаних джерел

- [1] P. Sharma, R. N. Yadav, and K. V. Arya, "Pose-invariant face recognition using curvelet neural network", *IET Biometrics*, vol. 3, no. 3, pp. 128-138, 2014.
- [2] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, iss. 12, pp. 2916-2929, 2012.
- [3] K. Grauman, and R. Fergus, "Learning binary hash codes for large-scale image

- search", in *Machine Learning for Computer Vision*, R. Cipolla, S. Battiato, and G. M. Farinella, Eds. Berlin, Heidelberg: Springer, 2013, pp. 49-87.
- [4] W. Wang, J. Yang, J. Xiao, S. Li, and D. Zhou, *Face recognition based on deep learning*, Q. Zu, B. Hu, N. Gu, S. Seng, Eds. Human Centered Computing. Springer, 2015.
- [5] X. Wu, "Learning robust deep face representation", arXiv preprint, 2015. [Online]. Available: <https://arxiv.org/pdf/1507.04844.pdf>. Accessed on: April 8, 2020.
- [6] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification", *Proc CVPR*, pp. 3025-3032, 2013, doi: 10.1109/CVPR.2013.389.
- [7] H.-V. Nguyen, and L. Bai, "Cosine similarity metric learning for face verification", *10th Asian Conf. on Computer Vision*, Queenstown, New Zealand, Nov. 8-12, 2010, Revised Selected Papers, Part II. [Online]. Available: https://www.researchgate.net/publication/220745463_Cosine_Similarity_Metric_Learning_for_Face_Verification/link/54dcd4880cf25b09b912d2ed/download. Accessed on: April 8, 2020.
- [8] Ю. В. Визильтер, В. С. Горбачевич, А. В. Воронников, и Н. А. Костромов, "Идентификация лиц в реальном времени с использованием свёрточной нейронной сети и хэширующего леса", *Компьютерная оптика*, № 2, 2017. [Электронный ресурс]. Режим доступа: <https://cyberleninka.ru/article/n/identifikatsiya-lits-v-realnom-vremeni-s-ispolzovaniem-svyortochnyy-neyronnoy-seti-i-heshiruyuschego-lesa>. Дата обращения: Апр. 8, 2020.
- [9] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification", *Proc. 27th Int. Conf. on Neural Information Processing Systems*, 2014, pp. 1988-1996.
- [10] J. Betty, I. Bülthoff, B. J. Mohler, and I. M. Thornton, "Face recognition of full-bodied avatars by active observers in a virtual environment", *Vision Research*, vol. 157, pp. 242-251, 2019.
- [11] P. Milgram, and F. Kishino, "A taxonomy of mixed reality visual displays", *IEICE Transactions on Information and Systems*, vol. 77, no. 12, pp. 1321-1329, 1994. [Online]. Available: https://www.researchgate.net/publication/231514051_A_Taxonomy_of_Mixed_Reality_Visual_Displays. Accessed on: Apr. 8, 2020.
- [12] A. Nayyar, B. Mahapatra, D. Le, and G. Suseendran, "Virtual Reality (VR) & Augmented Reality (AR) technologies", *Int. Journal of Engineering & Technology*, vol. 7, pp. 156-160 2018. [Online]. Available: https://www.researchgate.net/publication/324745910_Virtual_Reality_VR_Augmented_Reality_AR_technologies_for_tourism_and_hospitality_industry. Accessed on: Apr. 8, 2020.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition". [Online]. Available: <https://arxiv.org/pdf/1512.03385.pdf>. Accessed on: Apr. 8, 2020.

References

- [1] P. Sharma, R. N. Yadav, and K. V. Arya, "Pose-invariant face recognition using curvelet neural network", *IET Biometrics*, vol. 3, no. 3, pp. 128-138, Sept. 2014.
- [2] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, iss. 12, pp. 2916-2929, 2012, doi: 10.1109/TPAMI.2012.193.
- [3] K. Grauman, and R. Fergus, "Learning binary hash codes for large-scale image search", in *Machine Learning for Computer Vision*, R. Cipolla, S. Battiato, and G. M. Farinella, Eds. Berlin, Heidelberg: Springer, 2013, pp. 49-87, doi: 10.1007/978-3-642-28661-2_3.
- [4] W. Wang, J. Yang, J. Xiao, S. Li, and D. Zhou, "Face recognition based on deep learning", in *Human Centered Computing*, Q. Zu, B. Hu, N. Gu, S. Seng, Eds. Springer, 2015, doi: 10.1007/978-3319-15554-8_73.
- [5] X. Wu, "Learning robust deep face representation", arXiv preprint, 2015. [Online]. Available: <https://arxiv.org/pdf/1507.04844.pdf>. Accessed on: April 8, 2020.

- [6] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification", *Proc CVPR*, pp. 3025-3032, 2013, doi: 10.1109/CVPR.2013.389.
- [7] H.-V. Nguyen, and L. Bai, "Cosine similarity metric learning for face verification", *Proc ACCV*, pp. 709-720, 2010, doi: 10.1007/978-3-642-19309-5_55.
- [8] Yu. V. Vizilter, V. S. Gorbatshevich, A. V. Vorotnikov, and N. A. Kostromov, "Real time face identification using convolutional neural network and hash forest", *Kompyuternaya optika*, no. 2, 2017. [Online]. Available: <https://cyberleninka.ru/article/n/identifikatsiya-lits-v-realnom-vremeni-s-ispolzovaniem-svyortochnyy-neyronnoy-seti-i-heshiruyuschego-lesa>. Accessed: April 8, 2020.
- [9] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification", *Proc. 27th Int. Conf. on Neural Information Processing Systems*, 2014, pp. 1988-1996.
- [10] J. Betty, I. Bülthoff, B. J. Mohler, and I. M. Thornton, "Face recognition of full-bodied avatars by active observers in a virtual environment", *Vision Research*, vol. 157, pp. 242-251, 2019.
- [11] P. Milgram, and F. Kishino, "A taxonomy of mixed reality visual displays", *IEICE Transactions on Information and Systems*, vol. 77, no. 12, pp. 1321-1329, 1994. [Online]. Available: https://www.researchgate.net/publication/231514051_A_Taxonomy_of_Mixed_Reality_Visual_Displays. Accessed on: Apr. 8, 2020.
- [12] A. Nayyar, B. Mahapatra, D. Le, and G. Suseendran, "Virtual Reality (VR) & Augmented Reality (AR) technologies", *Int. Journal of Engineering & Technology*, vol. 7, pp. 156-160, 2018. [Online]. Available: https://www.researchgate.net/publication/324745910_Virtual_Reality_VR_Augmented_Reality_AR_technologies_for_tourism_and_hospitality_industry. Accessed on: Apr. 8, 2020.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition". [Online]. Available: <https://arxiv.org/pdf/1512.03385.pdf>. Accessed on: Apr. 8, 2020.

O. B. Danchenko¹, D.Sc., professor,

O. E. Ilarionov², Ph.D., associate professor,
oleg.ilarionov@knu.ua

H. V. Krasovska², Ph.D., associate professor,

T. S. Korotkova², M.Sc.

¹ Cherkasy State Technological University
Shevchenko blvd, 460, Cherkasy, 18006, Ukraine

² Taras Shevchenko National University of Kyiv,
B. Gavrylyshyn st., 24, Kyiv, 04118, Ukraine

DEVELOPMENT OF FACE RECOGNITION SYSTEM IN A VIDEO STREAM WITH AUGMENTED REALITY

In the paper, the problem of face recognition in a video stream with augmented reality is considered. The current state of this problem is investigated. The general process of face recognition and the basic concepts of augmented reality have been studied. The analysis of modern approaches to solving the face recognition problem is carried out, the strengths and weaknesses of the methods used have been found. A search is carried out for a method invariant to scaling, scene changes, head turns, changes in lighting, accessories, and changes in emotions. An algorithm, architecture, and the software system that solves the problem of face recognition in a video stream with the elements of augmented reality have been developed. A histogram of oriented gradients (HOG) is chosen as the method for face detection; face recognition functionality is developed on the basis of the convolutional neural network architecture – ResNet34. Experimental studies are carried out, the system has been tested on both one and several faces simultaneously. Estimate methods of the recognition quality for the developed software system are determined – the plotting of ROC-curves that show the dependence of the number of false positives on the detection accuracy (true positive rate) and measuring AUC. AUC =

0.95 has been achieved during recognition of one face, and $AUC = 0.83$ – during recognition of several faces (maximum 4).

Statement of the problem. Investigation and analysis of existing approaches to building face-to-face recognition technology in augmented reality systems by analyzing models, methods and algorithms for human face recognition, identifying strengths and weaknesses of existing solutions, choosing the best combination of detection and recognition methods.

Analysis of recent research and publications. Approaches have been proposed for the formation of biometric face image templates that can be used for biometric verification or face identification. However, all recent facial recognition results have been obtained through the use of deep convolutional neural networks. In the work of Yu. V. Visilter et al., a convolutional neural network with a hash forest (ZNMHL), based on a convolutional network with a hash layer, has been obtained. J. Betty et al. have investigated how different factors influence recognition quality.

The purpose of the article is to prove the effectiveness of the proposed approach based on the histogram of directed gradients and convolutional neural network architecture ResNet34 for the problem of face recognition in a video stream in augmented reality systems.

Presenting main material. The basic concepts of augmented reality are analyzed. The process of face recognition is described. The description of the software for the solution of the problem is given and mathematical model is developed. The algorithm of work of the program of face recognition program developed by authors is detailed. The architecture of face recognition system in augmented reality video stream is designed.

Results. A software system designed to recognize human faces in augmented reality video streams has shown satisfactory results. The application correctly recognizes the face, available in the database, in different conditions of lighting, head rotation, with the presence of accessory ditches, the closure of some parts of the face, changes in emotions, etc., similarly for recognizing multiple faces at the same time. The system has been tested on 520 examples: 4 people separately and together in different combinations under different conditions of lighting, noise, interference, accessories, emotions.

Conclusion. Applying a neural network to the ResNet architecture with appropriate settings for detecting and recognizing human faces in augmented reality video streams is a good choice – this method is invariant to scaling, scene changes, head turns, light changes, accessories, and emotion changes. The system is a platform for further development. In particular, it is planned to conduct experimental studies using other methods of face recognition in a video stream and to perform a comparative analysis of the results, as well as to create a more convenient graphical interface of the program and adaptation for the mobile version.

Keywords: face recognition, augmented reality, image pre-processing, face detection, face landmarks estimation.