УДК 004.93 DOI: 10.30837/0135-1710.2021.177.018

Є.В. ГУБАРЕНКО, М.С. ГУБАРЕНКО, М.В. АНТОНЮК

РОЗПІЗНАВАННЯ ОБ'ЄКТІВ У ВІДЕОПОТОЦІ

Описані проблеми попередньої обробки зображень у відеопотоці для подальшого аналізу і розпізнавання об'єктів. Розглянуті основні способи та підходи до компенсації недоліків чи дефектів зображень, серед яких оптимізація контрастності зображення, різкості, нормалізація освітлення, пошук і відбракування схожих зображень. Запропоновано використання методу еквіваріантного детектора для розпізнавання об'єктів, що швидко рухаються. Синтезована нейрона мережа, яка пройшла навчання для розпізнавання транспортних засобів та обличь людини, ймовірно водія. Запропоновано підхід для вдосконалення алгоритму пошуку локальних ознак при навчанні нейронної мережі. Наведені та проаналізовані результати експерименту з використанням навченої мережі при розпізнаванні окремих об'єктів та композитних сцен.

1. Вступ

Розпізнавання зображень є важливим компонентом систем управління, обробки інформації та прийняття рішень. Завдання, пов'язані з класифікацією і ідентифікацією предметів, явищ і сигналів, що характеризуються кінцевим набором деяких властивостей і ознак, виникають в таких сферах як робототехніка, інформаційний пошук, моніторинг та аналіз візуальних даних, дослідження штучного інтелекту. На даний момент широко використовуються системи розпізнавання рукописного тексту, автомобільних номерів, відбитків пальців або людських обличь, що знаходять застосування в інтерфейсах програмних продуктів, системах безпеки та аутентифікації особи [1].

За останній час з появою методів зниження розмірності, згортальних нейронних мереж, deep learning і констеляціонних моделей у розпізнаванні візуальних образів був досягнутий істотний прогрес. Однак, незважаючи на досягнуті успіхи, сучасні дослідження підтверджують той факт, що алгоритми розпізнавання об'єктів не можуть повноцінно замінити людину.

Актуальним залишається питання розпізнавання зображень тривимірних об'єктів під різними кутами зору, що піддаються перетворенням обертання, масштабування і трансляції. Сучасні підходи до вирішення цього питання, такі як багатошарові згорткові нейронні мережі, а також використання інваріантних детекторів ознак SIFT і ORB [2], в даний момент пропонують часткові рішення, що не забезпечують достатньої точності розпізнавання і втрачають інформацію про структуру об'єкта. Існують проблеми з обробкою потокового відео і виявлення об'єктів, що рухаються. Також проблемою є розпізнавання нечітких розмитих зображень або зображення перекриті іншими об'єктами в отриманих кадрах.

Використання камер як уніфікованого пристрою для визначення множини параметрів рухомих об'єктів (відстані, швидкості, метричні параметри) дозволить знизити собівартість системи і спростити формалізацію одержуваної інформації за рахунок зменшення різновидів застосовуваних технічних пристроїв, а також, без додаткових налаштувань робочого місця і додавання апаратних засобів, підвищити багатофункціональність системи контролю.

Мета дослідження полягає у вдосконаленні методу розпізнавання об'єктів, які рухаються, за рахунок використання еквіваріантного детектора на етапі навчання та використання нейронної мережі, що повинно покращити роботу систем автоматизованого та автоматичного моніторингу.

У рамках дослідження приділялась увага геометричним, кінематичним та динамічним характеристикам стану і поведінки рухомих та нерухомих об'єктів, які виявляються інформаційно-вимірювальними системами моніторингу в системах з рухомими об'єктами.

Теоретичне і практичне значення дослідження полягає у тому, що розроблені і реалізовані алгоритмічне, математичне і програмне забезпечення становлять основу вимірювальної системи, яка може використовуватися для віддаленого контролю і моніторингу транспортних засобів, людей, технологічних процесів в різних сферах діяльності, а також входити

до складу комплексу технічного зору, що забезпечує автономне функціонування транспортно-технологічних комплексів, безпілотних систем.

2. Проблеми на етапах розпізнавання об'єктів у відеопотоці

Процес розпізнавання об'єктів у відеопотоці може буди розділений на наступні етапи (рис. 1).

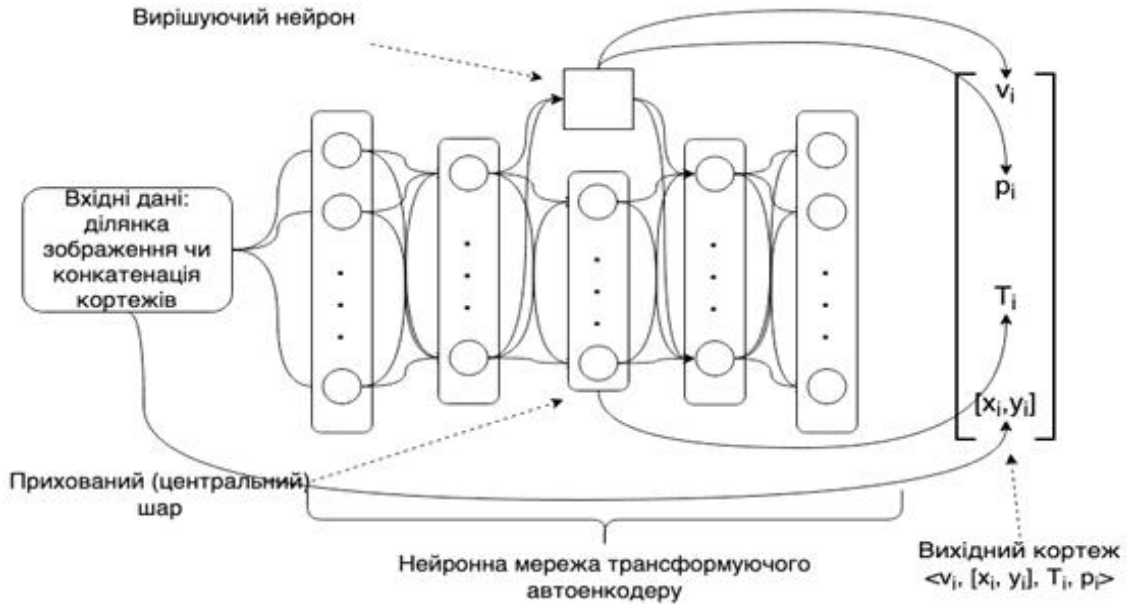

Рис.1. Схема еквіваріантного детектора на базі трансформуючого автоенкодера

Етап 1. Порівняння кадрів у відеопотоці - серед проблем, які виникають на цьому етапі, слід зазначити незмінність кадрів. Частіше за все відеопотік має 16-30 кадрів у секунду, часто виникають ситуації, коли кадри ідентичні, або мають незначні відмінності. Існують три основних підходи до вирішення такої проблеми [3]: порівняння значення хеш-функцій двох кадрів, які змінюються; обчислення коефіцієнту кореляції; побудова та аналіз SURF-дискриптів.

Етап 2. Оцінка якості зображення - серед проблем, які виникають на цьому етапі, слід зазначити розмиття, наявність шумів та засвічення кадрів. З багатьох причин зображення у кадрі може бути пошкоджено, іноді якість кадру може бути дуже низькою і не придатною до розпізнавання, такі кадри слід відкинути. На цьому етапі слід провести оцінку контрастності, різкості та чіткості [4]. Можна також підвищити різкість або компенсувати недоліки якості зображення.

Етап 3. Зменшення розмірності зображення - серед проблем, які виникають на цьому етапі, слід зазначити вхідні кадри високої розмірності. Кадри надходять у вигляді матриці пікселів, чим більша розмірність, тим більше операцій буде виконано і більше часу на розпізнавання буде витрачено. Іноді є можливим зменшити розмірність зображення, при цьому залишивши достатньо даних для обробки [5].

Етап 4. Розпізнавання об'єктів на зображенні - серед проблем, які виникають на цьому етапі, слід зазначити оклюзії та трансформацію. Іноді об'єкти повертаються, віддаляються або перекриваються іншими об'єктами. Такі проблеми вирішуються завдяки використанню відповідних алгоритмів розпізнавання [5].

Етап 5. Отримання результату розпізнавання.

3. Удосконалення методу еквіваріантного детектора розпізнавання об'єктів, що швидко рухаються

Головні особливості методологій та алгоритмів розпізнавання об'єктів, що швидко рухаються, а також програмного забезпечення, створеного на їх основі, визначаються особливостями предметної області та апаратного забезпечення. На даний момент складно створити універсальну систему для розпізнавання різних класів об'єктів та подальшого їх аналізу. Вузькими місцями подібної системи зараз є обчислювальні можливості комп'ютерів та недосконаліть алгоритмів. У дослідженні пропонується розглядати переважно розпізнавання рухомих об'єктів транспортного типу. Програмне забезпечення, яке використовується для вирішення задач подібного типу, зараз є одним з найбільш затребуваних. Таке забезпечення може використовуватися для спостереження за дотриманням правил дорожнього руху на шлагбаумах, пропускних пунктах, залізничних шляхах, магістралях, для моніторингу трафіку, знаходження вузьких місць в трафіку на дорогах, для перепустки в якісь місця тощо.

Вхідними даними є кадри з відеопотоку чи запис трафіку.

Виходом є навчена нейронна мережа, результат розпізнавання у вигляді областей знайдених об'єктів та ідентифікований об'єкт.

Слід виділити наступні етапи роботи алгоритму розпізнавання:

Етап 1. Оптимізація обробки відеопотоку. На даному етапі вибираємо зображення з серії кадрів, які відрізняються, знаходимо кращий кадр за якістю серед схожих для більш детального аналізу зображення. Даний етап потрібен для оптимізації швидкості роботи алгоритму: не має сенсу повністю аналізувати кожен кадр з відеопотоку, адже існує велика ймовірність того, що зображення ідентичні або майже незмінні. Важливо також вибрати для подальшого аналізу чіткіше і контрастніше зображення серед вибірки з декількох схожих задля покращення процесу розпізнавання.

Етап 2. Стиснення, зменшення розмірності зображення з мінімальною втратою інформативності. Даний крок необхідний для оптимізації швидкодії алгоритму, адже, як правило, необроблені зображення мають дуже високі розмірності, що ускладнює обробку і потребує більше ресурсів і часу для аналізу.

Етап 3. Знаходження границь об'єктів на зображенні. На даному етапі знаходимо контури об'єктів, встановлюємо їх кількість і розташування на зображені. Таким чином, надалі ми будемо аналізувати лише частини зображення, які нас цікавлять.

Етап 4. Аналіз інформативних параметрів об'єкта. На даному етапі знаходимо колір, розміри, віддаленість від камери, швидкість руху, позицію відносно камери, напрям руху. Всі ці характеристики далі використовуються для вирішення бізнес-задач конкретної системи.

Етап 5. Виділення класу об'єкта та його виду. На даному етапі аналізуємо, що за об'єкт був знайдений, його клас та тип, вирішується, чи потрібен він для роботи системи, чи є аномалією, далі він може класифікуватися, а також використовуватися для навчання адаптивних нейронних мереж для самовдосконалення системи.

Було проведено експериментальну оцінку результатів дослідження. Отримані результати роботи алгоритму розпізнавання було порівняно з результатами роботи відомих аналогічних алгоритмів, зокрема, були порівняні долі успішно розпізнаних зображень. Як тестові вибірки було взято зображення, згенеровані спеціальними програмними засобами, а також відомі тестові набори, які знаходяться у відкритому доступі.

4. Синтез моделі нейронної мережі з використанням еквіваріантного детектора на базі трансформуючого автоенкодера

Алгоритм розпізнавання з використанням еквіваріантного детектора, побудованого на базі трансформуючого автоенкодера (рис. 1), де вихідними даними для мережі є ділянка зображення (для першого рівня ієрархії) або конкатенація вихідних кортежів з детекторів нижчих рівнів (для другого та вищих за номером рівнів). Результатом є кортеж розмірністю $n\langle v_i, [x_i, y_i], T_i, p_i \rangle$, де v_i - значення ідентифікатора функції детектора, вираженого як вектор-маска; $[x_i, y_i]$ - відносні координати центру детектора; T_i - значення трансформації, вира-

жене як матриця афінної трансформації; p_i - значення впевненості детектора на відрізку [0;1].

Детектори є частинами ієрархічної структури, кожен рівень ієрархії відповідає локальним ознакам або групі ознак, які описують певну частину об'єкта (рис. 2). Перший рівень ієрархії містить детектори мінімальних локальних ознак об'єкта (наприклад, фара, лобове скло чи колесо автомобіля). Наступний рівень описує групи локальних ознак, які можна об'єднати у більш складну локальну ознаку. Останній рівень ієрархії містить один детектор, який описує весь об'єкт і включає у себе локальні ознаки попередніх рівнів. Між першим та останнім рівнем ієрархії може бути скільки завгодно рівнів.

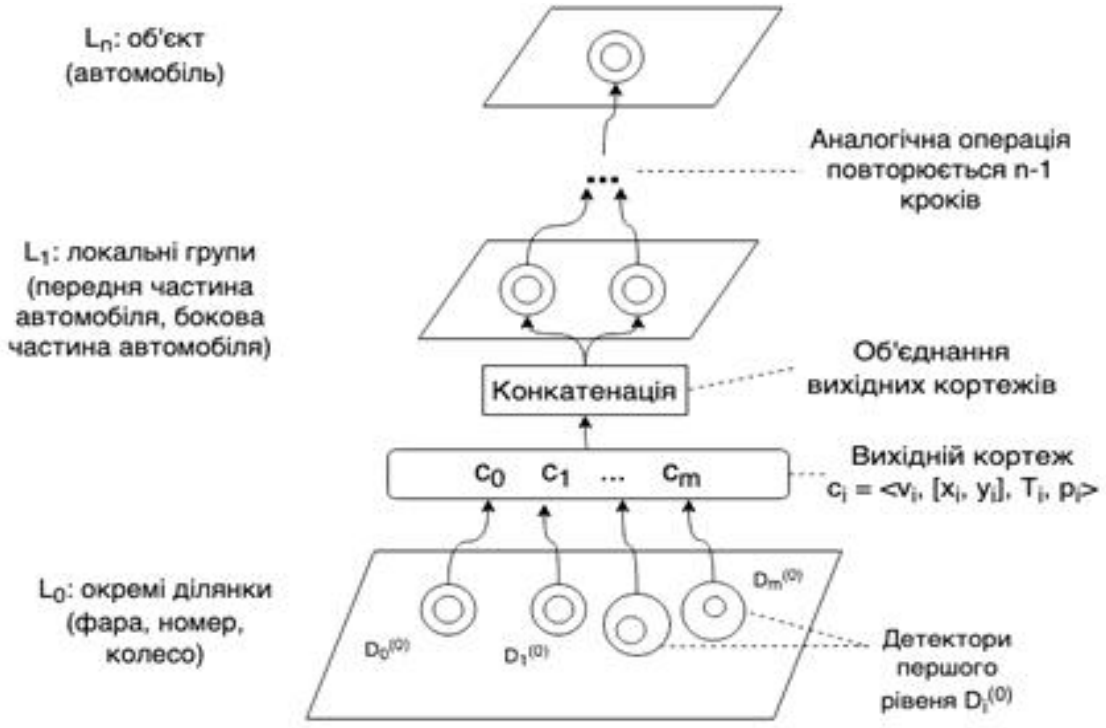

Рис.2. Схема ієрархічної організації локальних еквіваріантних детекторів

5. Вдосконалення алгоритму пошуку локальних ознак

Алгоритм розпізнавання складається з послідовної активації детекторів моделі, починаючи з першого рівня, на заданому зображенні. Нехай є функція розпізнавання $g: X \to Y$, аргументами якої є зображення $x_n \in X$, представлені в вигляді вектору довжини *n*, а значеннями функції - множина класів (категорій) $y \in Y$, варійована в залежності від поставленого завдання. Є також підмножина пар аргументів і значень функції $D = \{(x_0, y_0),...,(x_0, y_0)\}$. Таким чином, ієрархія детекторів, що послідовно активуються, реалізує функцію $h: X \to Y$, яка апроксимує функцію *g* на всій її області визначення, в тому числі в точках, які не включені в *D*. Для розрахунку значення *h(x)* зображення надходить на вхід першого шару навченої системи, потім виконується послідовна активація локальних детекторів на кожному з рівнів. Вихідне значення ієрархії являє собою бінарне число, що визначає приналежність зображення до класу, при цьому вихідний рівень також проводить оцінку параметрів локалізації зображеного об'єкта, якщо значення активації дорівнює 1 (зображення успішно розпізнано і належить класу).

У такому вигляді розроблена система розпізнавання використовується для вирішення задачі унарної класифікації, коли множина класів $y \in Y$ представлена двома елементами

{0,1}. Функція, таким чином, дорівнює 1 у випадках, коли зображення, що служить її аргументом, містить об'єкт, що належить класу, і 0 в іншому випадку. Для випадків, коли потрібно розпізнати зображення серед кількох можливих класів (завдання мультикласової класифікації), проводиться навчання окремої ієрархії детекторів для кожного конкретного класу, і потім проводиться послідовна перевірка зображення на позитивну відповідність кожного з них. В цьому випадку розглядається функція g', визначена на множині X', значеннями якої є множина класів, така що для обраного *j*-го класу $Y' = \{y_i, U_{j \neq i} y_i\}$.

Нехай дана навчена ієрархія детекторів M_c для деякого класу зображень *c* (наприклад, людських обличь), або кілька ієрархій для завдання мультикласового розпізнавання, і зображення *I*, яке необхідно розпізнати. Алгоритм розпізнавання складається з наступних кроків:

Крок 1. Вибираємо ієрархію детекторів M_c.

Крок 2. Для *l*-того рівня ієрархії, починаючи з першого, і для кожного детектора відповідного рівня $D_j^{(l)}$ складаємо вектор вихідних даних $Z^{(l)}$.

Крок 3. Якщо $l = 0$, то заповнимо вектор вихідних даних наступним чином: розраховуємо значення функції ідентифікації детектора $d_{I_j}^{(l)}$ і значення впевненості $d_{p_j}^{(l)}$ для кожної ділянки зображення *I(x..x+w,y..y+h)*, де *w* та *h* відповідають розмірам локальної ділянки зображення, і проведемо конкатенацію вектора $Z^{(l)}$ з вихідними значеннями детектора для тих ділянок, де $d_{I_j}^{(l)} = 1$, та $d_{p_j}^{(l)} > t$, де *t* - обрана ступінь впевненості.

Крок 4. Якщо $l \neq 0$, вектор вхідних даних отримаємо як $Z^{(l)} = \left(D_0^{(l)}\left(Z^{(l-1)}\right)\right) \| \ldots \| \left(D_j^{(l)}\left(Z^{(l-1)}\right)\right) \| \left(0_0 \| \ldots \| 0_{N_l}\right)$, де N_l - максимальна кількість можливих детекторів на рівні *l*.

Крок 5. Якщо для рівня *l* активації всіх детекторів цього рівня негативні (дорівнюють нулю), то зображення не належить класу *c*.

Крок 6. В іншому разі інкрементуємо значення *l* і повторюємо алгоритм з другого кроку.

Крок 7. Якщо активація останнього рівня позитивна, то зображення належить класу *c*. Інакше повторюємо алгоритм з першого кроку.

Використовуючи той факт, що низькорівневі локальні ділянки зображення (взяті з достатнім масштабом), як правило, представляють собою край, межі і кути, слід зазначити, що має місце тенденція до зростання різноманітності серед еквіваріантних детекторів на більш високих рівнях моделі. Експерименти показують, що кількість детекторів першого рівня при навчанні не перевищує 10, при цьому ієрархії, навчені на об'єктах різних категорій, здатні розділяти між собою частину детекторів першого рівня, демонструючи ефект, що нагадує трансферне навчання або попереднє навчання без вчителя.

6. Аналіз результатів експерименту

Для оцінки ефективності розробленого алгоритму розпізнавання зображень проведемо експерименти для трьох варіантів алгоритму об'єднування в одній категорії зображень, різних з інформатико-теоретичної точки зору (таких, як зображення людського обличчя в профіль і в фас), але відповідних при цьому одній категорії приналежності.

Варіант 1. Навчання локального еквіваріантного детектора в якості окремого класифікатора. Метою експериментів є визначення точності відновлення фрагментів зображення за компактними репрезентаціями, адекватності сформованих репрезентацій, їх відповідність значенням трансформацій. Крім цього, тестується ефективність роботи детектора в якості опції-ідентифікатора. У завершальній частині проводиться зіставлення з іншими алгоритмами унарної класифікації та відновлення трансформації.

Варіант 2. Розпізнавання зображень за допомогою повноцінної ієрархії локальних еквіваріантних ознак. Об'єктами експерименту є показники помилки розпізнавання (в порівнянні з відомими алгоритмами), а також сформовані моделлю просторові структури еквіваріантних детекторів.

Варіант 3. Робота системи на узагальнених відеофрагментах, що містять множину об'єктів різних класів. Досліджується здатність моделі кластеризувати сцену, виявляючи окремі об'єкти і формуючи відповідні їм репрезентації.

Оскільки представлена модель вимагає для навчання обмежену вибірку з повнозв'язних відеофрагментів - тип даних, який не отримав поширення серед сучасних методів розпізнавання зображень, - навчання моделей для експерименту проводилося на даних, представлених автором. Навчальні вибірки включають в себе як відеофрагменти, отримані в результаті обробки згенерованих з використанням комп'ютерної графіки тривимірних моделей, так і фрагменти живої зйомки навколишнього світу.

Оцінка ефективності розпізнавання людських обличь проводилася на відкритих базах даних LFW [6], HPID (Head Pose Image Database) [7]. Деякі моделі, що використовують альтернативні методи розпізнавання, надані бібліотекою Caffe [8]. Також для експериментів використовувалися дані, отримані за допомогою програми тривимірного моделювання FaceGen [9], що представляють собою марковані зображення людських обличь в різних орієнтаціях щодо камери.

Експериментальна перевірка розробленого методу розпізнавання проведена відповідно до рекомендацій з навчання і порівняння моделей розпізнавання образів в рамках апарату теорії розпізнавання образів і машинного навчання [10, 11]. Для оцінки ефективності методів, які використовуються в експерименті, використовувалися показники точності і повноти (що відображають помилки першого і другого роду).

Для оцінки результативних показників застосовувався метод розрахунку довірчих інтервалів. Відповідно до прийнятої практиками постановки експерименту в області розпізнавання і рекомендаціями щодо вибору рівня довіри, значення рівня довіри p вибрано рівним 0.95 [12]. Експериментальні дані продемонстрували, що для досягнення відповідного рівня досить вибірки, яка складається з близько 500 примірників, що для авторського методу відповідає короткому відеофрагменту, який за часом не перевершує 15 секунд. Частота зміни кадрів дорівнює 24 кадри за секунду, за умови використання методу пермутації.

Оскільки специфіка розробленого методу полягає у визначенні просторової локалізації об'єктів під різними кутами зору, то експериментальні вибірки для навчання і тестування містять набори відповідних зображень. Для оцінки здатності методу розпізнавання стійким чином обробляти об'єкти під впливом відповідних інваріантних перетворень, для кожного етапу експерименту результати обчислювалися за допомогою агрегації показників серії експериментальних перевірок. При цьому кожна серія складається з наступних елементів: навчання і тестування на загальній вибірці, навчання і тестування випадковою вибіркою, навчання і тестування на кластеризованій вибірці. Кластеризація включає в себе розбивку вибірки на відеофрагменти, що містять зображення окремих об'єктів (конкретних людських осіб), при цьому дані відеофрагментів навчання не перетинаються з тестовими. Таким чином, результатом експерименту є агрегований показник серії з двох стадій навчання і тестування з урахуванням розрахованої на базі складових показників стандартної помилки.

В рамках експерименту проводиться зіставлення результатів використання розробленого методу і альтернативних методів розпізнавання зображень, що використовуються в даний час. Вибрані альтернативи широко застосовуються як у виробництві (метод Віоли-Джонса [13], згорткові нейронні мережі [14]), так і в академічних дослідженнях. При виборі альтернатив використовувалися дані відкритих змагань з машинного навчання в області розпізнавання зображень, таких як ILSVRC [15]. Реалізації алгоритмів, що використовувалися в ході експерименту, представлені авторами відповідних методів і отримані з відкритих джерел [10,16].

На базі отриманої вибірки, згрупованої попарно, проводиться навчання розрідженого трансформуючого автоенкодера. В експерименті брали участь чотири моделі автоенкодерів, позначені нижче як A, B, C і D, що відрізняються кількістю нейронів першого і останнього прихованих шарів.

- A - 36x36 нейронів;
- B - 64x64 нейрони;
- C - 256x256 нейронів;
- D - 768x768 нейронів.

Вибір кількості нейронів зроблено відповідно до відомих експериментів з навчання автоасоаціативних нейронних мереж [17]. Трансформації представляють собою поворот камери по осях *X* і *Z*.

У табл. 1 наведені результати розпізнавання для різних категорій фрагментів зображень, взятих з відеофрагментів (складання вибірки проводилося в автоматичному режимі, методом трекінгу; категорії промарковані назвами для спрощення і зручності аналізу).

Таблиця 1

	A (36x36), %	B (64x64), %	C (256x256), %	D (768x768), %
Обличчя: очі	79±4	95±3	96±3	96±3
Обличчя: ніс	82±3	92±3	93±3	94±2
Обличчя: рот	84±3	97±4	97±3	97±3
Обличчя: вуха	72±3	91±3	93±4	93±3
Обличчя: деталі контуру	92±4	96±3	97±4	98±3
Автомобіль: колеса	72±4	92±4	94±3	95±4
Автомобіль: фари	74±3	91±3	92±2	92±2
Автомобіль: лобове скло	80±3	94±4	95±2	95±3
Автомобіль: заднє скло	82±4	92±2	93±3	93±3
Автомобіль: бампер	79±4	90±3	92±4	93±4
Автомобіль: деталі контуру	78±2	89±4	90±3	91±3

З отриманих даних можна зробити висновок про оптимальність моделі - збільшення числа нейронів веде до незначного підвищення точності реконструкції, при цьому збільшуючи обчислювальне навантаження. Також більше число нейронів вимагає більшого числа ітерацій навчання для досягнення збіжності. Найбільш проблемними випадками для роботи еквіваріантного детектора є аномальні ситуації, які не зустрілися у відеофрагменті - наприклад, оклюзія шуканої ділянки зображення сторонніми предметами. Крім ідентифікації ділянок зображення, вторинною функцією еквіваріантного детектора є оцінка параметрів інстанціювання або оцінка позиції об'єкта. Для експериментальної перевірки точності цієї оцінки використовувалися змішані дані вибірки HPID і згенерованої вибірки зображень осіб FaceGen з подальшою пост-обробкою візуальними ефектами розмиття і випадкової оклюзії. Для зіставлення результатів використовувалися такі методи як класичний алгоритм POSIT [18], і навчається з учителем регресор - випадковий ліс [19]. Оскільки прогноз трансформації являє собою регресію, а не класифікацію, як показник ефективності обрано відносну похибку, виражену у відсотках. Результати представлені у табл. 2.

В результаті експерименту виявлено, що еквіваріантний детектор показує більш точну оцінку просторових параметрів об'єкта в порівнянні з альтернативними методами. Істотна перевага над класичними методами комп'ютерного зору пояснюється тим, що алгоритм POSIT вимагає для оцінки позиції наявності маркерів, які можуть бути спроектовані на об'єкт за допомогою методів епіполярної геометрії.

Було проведено дослідження ефективності простої дворівневої моделі на трьох категоріях зображень людських обличь: звичайні зображення осіб, зображення з штучним зашумленням за допомогою оклюзії і розмиття. Ефективність розпізнавання розраховувалася за допомогою показників точності і повноти та порівнювалася з аналогічними показниками альтернативних методів розпізнавання.

Оскільки однією з основних переваг даної моделі є еквіваріантність - здатність до ідентифікації об'єктів на зображенні в різних орієнтаціях - то експериментальні вибірки згруповані таким чином: розглядається деяка вихідна позиція об'єкта з координатами обертання (кутів Ейлера) (0,0,0), при цьому в групу, що характеризується значеннями $(\varphi_{\min}, \varphi_{\max})$, входять зображення об'єктів, які зазнали трансформації обертанням $(\varphi_i, \varphi_j, \varphi_k)$, такої, що для будь-якого $\varphi \in (\varphi_i, \varphi_j, \varphi_k)$ є вірною нерівність $\varphi_{\min} < \varphi < \varphi_{\max}$.

Таблиця 2

	POSIT, %	Випадковий ліс, %	Авторський метод, %
Обличчя: очі	25±3	17±5	11±3
Обличчя: ніс	27±3	14±2	12±3
Обличчя: рот	18±4	16±4	10±4
Обличчя: вуха	24±4	12±4	11±2
Обличчя: деталі контуру	20±2	14±3	14±4
Розмиття: очі	27±3	19±3	12±3
Розмиття: ніс	32±3	18±3	14±2
Розмиття: рот	25±3	15±3	12±4
Розмиття: вуха	21±2	17±2	14±3
Розмиття: деталі контуру	31±4	16±4	12±3
Оклюзія: очі	32±4	20±3	10±5
Оклюзія: ніс	28±2	16±5	14±2
Оклюзія: рот	31±5	16±2	13±3
Оклюзія: вуха	31±3	18±4	11±4
Оклюзія: деталі контуру	30±4	17±4	13±4

Оскільки для цього експерименту потрібна значна кількість зображень об'єктів з різних кутів огляду, для його проведення була використана вибірка, отримана з використанням комп'ютерної графіки і генерації осіб програмою FaceGen.

Тестування розпізнавання осіб проводилося шляхом зіставлення методом Віоли-Джонса, класифікатора SVM в поєднанні з обчисленням гістограми орієнтованих градієнтів і згорткових мереж, навчених на вибірці ImageNet. Навчені моделі були надані бібліотеками Caffe з OpenCV. Результати наведені у табл.3.

Таблиця 3

	Метод Віоли-Джонса, %	SVM+HOG, %	CaffeNet, %	Авторський метод, %
(0°,15°)	85±3	84±5	87±4	92±3
(15°,30°)	72±3	75±2	85±4	91±3
(30°,45°)	67±4	72±4	86±2	89±4
(45°,60°)	66±4	78±4	83±3	86±2
(60°,90°)	86±2	77±3	82±3	90±4
(90°,120°)	67±3	80±3	83±3	86±3

Метод Віоли-Джонса потребує наявності окремої стадії навчання для кожної орієнтації. В рамках експерименту використовувалася попередньо навчена модель, яка продемонструвала зіставні результати для фронтальної і профільної орієнтації осіб, але вкрай низькі результати в проміжних станах. При використанні SVM в поєднанні з методами зниження розмірності і підвищення інваріантності (гістограма орієнтованих градієнтів) для класифікатора характерно зниження точності в міру включення до вибірки зображень різних орієнтацій. Така поведінка пов'язана з тим, що модель в процесі навчання намагається виробити ознаки, які однаково підходять для всіх зображень у вибірці, в результаті отримуємо рівномірно розподілені невисокі значення точності. Серед порівнянних методів авторська модель поступається тільки глибиною згорткової мережі, здатної до навчання різних (таких, що не змішуються) локальних ознак для різних орієнтацій об'єкта.

Слід зазначити, що на відміну від методу Віоли-Джонса, розроблена система розпізнавання дозволяє виділити комплексну структуру голови людини, не обмежуючись ділянкою,

що містить очі і рот. Такий ефект є наслідком використання ознак деталей контуру обличчя.

Порівняння результатів розпізнавання зображень обличь, підданих впливу, розмиття проводилося за допомогою методів SVM + HOG, глибинної згорткової мережі і алгоритму випадкового лісу. Вибірка проводилася на базі CVLAB Dataset і включала в себе об'єкти, подані з різних кутів (табл. 4). Шумом є розмиття по Гаусу зі значенням σ = 0.5 ... 2.5.

Таблиця 4

	Випадковий ліс, %	SVM+HOG, %	CaffeNet, %	Авторський метод, %
(0°,15°)	72±4	72±4	85±2	91±4
(15°,30°)	73±3	74±3	81±3	88±3
(30°,45°)	75±2	76±4	80±4	88±4
(45°,60°)	72±3	74±4	82±4	89±4
(60°,90°)	71±3	73±4	83±3	91±3
(90°,120°)	73±3	72±4	83±3	87±4

При впливі ефекту розмиття дисперсія результатів по відношенню до орієнтації об'єкта знижується.

Як завершальні групи експериментальної вибірки використовувалися зображення людських обличь під впливом шуму оклюзії (часткового перекриття). Шум оклюзії згенерований за допомогою випадкового розміщення на зображеннях осіб непрозорих геометричних фігур. При генерації шуму оклюзії параметри розміщення фігур підібрані таким чином, щоб залишати відкритою мінімум дві третини площі зображення. Результати наведені у табл. 5.

Таблиця 5

	Випадковий ліс, %	SVM+HOG, %	CaffeNet, %	Авторський метод, %
(0°,15°)	72±4	72±3	84±3	90±5
(15°,30°)	69±3	66±4	82±3	89±4
(30°,45°)	70±3	68±4	83±2	88±3
(45°,60°)	71±4	70±4	81±4	89±4
(60°,90°)	70±2	67±3	84±5	88±3
(90°,120°)	72±3	67±3	85±4	87±3

Як демонструють результати, оклюзія є суттєвою перешкодою для алгоритмів розпізнавання, що використовують компактні цілісні представлення, таких як випадковий ліс і SVM+HOG - для деяких груп орієнтацій спостерігається падіння точності розпізнавання до 10%. Інструменти розпізнавання, які використовують локальні ознаки (наприклад, розроблена система і мережа CaffeNet), менш чутливі до часткового перекриття локальних областей зображення. У таких ситуаціях продуктивність представленого методу наближається до показників основного конкурента - згорткових мереж.

7. Аналіз композитних сцен

В даному експерименті перевірялася здатність розробленої системи за допомогою багаторівневої моделі аналізувати сцени, що містять безліч об'єктів, і автономно (без наявності вчителя) класифікувати їх між собою. Як вибірки використовувалися відкриті дані, що містять відеозаписи камер спостереження за дорожнім трафіком. Оцінка продуктивності алгоритму проводилася в такий спосіб: підраховувалася кількість об'єктів в кадрі, аналогічним чином вручну робилося розбиття по групах, потім ці значення зіставлялися зі значеннями знайденими алгоритмом (рис. 3).

У табл. 6 значення наведені у відсотках від фактичної кількості об'єктів. Слід зазначити, що на відміну від ситуації з контрольованим трекінгом і рухом камери навколо об'єкта, алгоритм в режимі аналіза композитних сцен зі сторонніх відеофрагментів позбавлений інформації про фактичний рух об'єкта навколо камери. Так система не здатна без додаткової інформації визначити, що рух від камери автомобіля що віддаляється являє собою переміщення по площині, непаралельній площині сенсору камери. У загальному випадку в такій ситуації алгоритм знаходиться в умовах невизначеності і здатний сформувати тільки обмежене представлення об'єкта за допомогою моделі еквіваріантних детекторів. Існує можливість, використовуючи методи проективної геометрії, ввести деякі припущення в розрахунок оптичного трекінгу - так для розглянутих експериментальних випадків стадія оптичного трекінгу була доповнена умовою жорсткості (фіксованою формою) об'єктів в кадрі. При цьому трекер інтерпретував зменшення розмірів об'єкта на зображенні як видалення від камери, і на підставі відносної зміни площі об'єкта і показників оптичного потоку обчислював напрямок його руху.

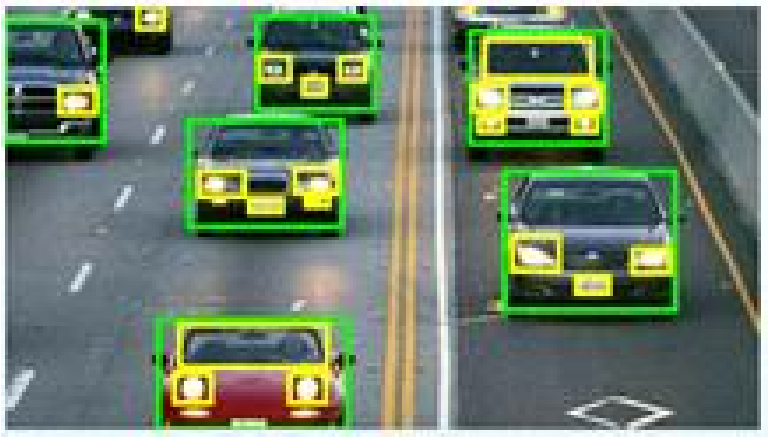

Рис.3. Результат розпізнавання запису «трафік 2»

Таблиця 6

	Об’єктів знайдено, %	Груп об’єктів знайдено, %
Запис трафіку 1	70	96
Запис трафіку 2	72	100
Запис трафіку 3	75	75
Запис трафіку 4	74	75

8. Висновки

Результатом дослідження стала модернізація алгоритму розпізнавання. Запропоновано використання еквіваріантного детектора на базі трансформуючого автоенкодера алгоритму розпізнавання. Створена та навчена нейронна мережа для розпізнавання об'єктів у відеопотоці. Проведено експеримент, у процесі якого ефективність розробленої системи порівнювалася з показниками альтернативних відомих методів розпізнавання. Точність розпізнавання у разі використання запропонованого методу зростає на 3-5%. Розроблена система розпізнавання більш стійка до локального шуму: для зображень, що піддаються розмиттю і оклюзії, падіння точності розпізнавання розробленої системи становить 3-6% проти 5-10% у альтернативних відомих методах.

Результати досліджень становлять практичний інтерес при проектуванні систем управління та обробки інформації в області комп'ютерного зору і розпізнавання зображень, для тих завдань, де існує необхідність визначення просторових параметрів зображених об'єктів. Запропонований підхід навчання системи базується на використанні відеозаписів, тобто система може навчатися на встановленому пункті моніторингу для адаптації до локальних образів. Для використання системи для моніторингу в режимі реального часу слід проаналізувати і спроєктувати систему на базі розподілених обчислень для паралельного аналізу кадрів, адже результати локальних експериментів показують труднощі при багатьох кадрах в секунду, активація детекторів багатьох рівнів потребує значних обчислювальних ресурсів.

Список літератури: 1. *Nechiporenko A.S., Gubarenko E.V., Gubarenko M.S.* Authentication of users of mobile devices by their motor reactions. Telecommunications and Radio Engineering. 2019. V. 78 (11). P. 987-1003. doi: 10.1615 / TelecomRadEng.v78.i11.60. 2. *Ebrahim Karami, Siva Prasad, and Mohamed Shehata.* Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images. 2017. arXiv preprint arXiv:1710.0272. 3. *Локтєв Д.А., Кочнев В.А., Локтєв А.А.* Вивчення функцій розмиття зображення у вигляді інформативного параметра стану і поведінки аналізованого об'єкта. Динаміка складних систем - XXI століття. 2020. № 2. С. 16-27. 4. *Loktev D., Loktev A.* Image blurring function as an informative criterion. Advances in Intelligent Systems and Computing. 2021. V. 1258. P. 173-183. 5. *Фісенко В.Т., Фісенко Т.Ю.* Комп'ютерна обробка і розпізнавання зображень: навчальний посібник. СПб: СПбГУ ІТМО, 2008. 192 с. 6. *Huang G.B., Ramesh M., Berg T., Learned-Miller E.* Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts. 2007. No. 1 (2). pp. 3-37. 7. *Gourier N., Hall D., Crowley J.L.* Estimating face orientation from robust detection of salient facial structures. FG Net Workshop on Visual Observation of Deictic Gestures, 2004. V. 6(4). 8. *Gourier N., Hall D., Crowley J.L.* Caffe: Convolutional architecture for fast feature embedding. FG Net Workshop on Visual Observation of Deictic Gestures. 2004. P. 1-9. 9. *Singular* Inversions. FaceGen modeller (Version 3.3). Singular Inversions, 2008. 10. *Bishop C.M.* Neural networks for pattern recognition. Oxford: Oxford university press, 1995. P. 482. 11. *Bishop C.M.* Pattern recognition and machine learning. New York: Springer, 2006. P. 758. 12. *Tan X., Triggs B.* Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. IEEE Transactions on image processing. 2010. Vol. 19, No 6. P. 1635-1650. 13. *Viola P., Jones M.* Rapid object detection using a boosted cascade of simple features. 2001. No. 1. P. 502-511. 14. *Szegedy* et al., Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. P. 1-9, doi: 10.1109/CVPR.2015.7298594. 15. Hinton G.E. A practical guide to training restricted Boltzmann machines. Momentum. 2010. No. 9 (1). P. 926. 16. Hubel D.H., Wiesel T.N. Brain and visual perception. ISBN13, 2005. 17. *Hinton G.E., Salakhutdinov R.R.* Reducing the dimensionality of data with neural networks. Science. 2006. No.313 (5786). P. 504-507. 18. *Duin R.P., Pekalska E.* Open issues in pattern recognition. Computer Recognition Systems. 2005. P. 27-42. 19. *Leo B.* Random forests. Machine learning. 2001. No. 45 (1). P. 5-32.

Надійшла до редколегії 24.05.2021

Губаренко Є.В., кандидат технічних наук, доцент, доцент кафедри системотехніки ХНУРЕ. Наукові інтереси: теорія прийняття рішень, управління соціально-економічними системами, системи штучного зору. Адреса: Україна, 61166, м. Харків, пр. Науки, 14, тел. +38 (050) 741 01 74.

Губаренко М.С., асистент кафедри системотехніки ХНУРЕ. Наукові інтереси: згорткові нейронні мережі, глибоке навчання, розпізнавання зображень. Адреса: Україна, 61166, м. Харків, пр. Науки, 14, тел. +38 (050) 532 61 23.

Антонюк М.В., магістрант кафедри системотехніки ХНУРЕ. Наукові інтереси: проблеми розпізнавання зображень. Адреса: Україна, 61166, м. Харків, пр. Науки, 14, тел. +38 (068) 342 43 74.