

DOI: <https://doi.org/10.15276/ict.01.2024.32>

УДК 004.89

Задачі візуального аналізу натовпу в системах інтелектуального відеоспостереження

Добришев Руслан Євгенович

Аспірант каф. Штучного інтелекту та аналізу даних

ORCID: <https://orcid.org/0009-0007-8639-3157>; rdobrishev@gmail.com

Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

Сучасне зростання міського населення створює виклики для громадської безпеки, зокрема через скупчення людей. Це стимулює розробку нових методів управління натовпом, які потребують автоматизованого аналізу. Візуальний аналіз натовпу, заснований на технологіях комп'ютерного зору, є ключовим інструментом для вирішення цих завдань. Розвиток глибокого навчання суттєво покращив системи моніторингу, що використовуються для міського спостереження, контролю соціальної дистанції, управління транспортом і заходами. Однак аналіз натовпу залишається складним через оклюзії, варіації масштабу, непередбачувані патерни руху та складну поведінку. Для подолання цих викликів потрібні нові алгоритми, моделі й масштабні набори даних, які дозволяють проводити аналіз у реальному часі. Основні завдання включають підрахунок людей, виявлення об'єктів, аналіз руху, розпізнавання поведінки та виявлення аномалій. Глибокі нейронні мережі та навчання з перенесенням суттєво підвищують точність і адаптивність таких систем, що сприяє покращенню громадської безпеки та управлінню потоками людей.

Ключові слова: інтелектуальне відеоспостереження; машинне навчання; глибоке навчання; розпізнавання поведінки; аналіз руху

Актуальність дослідження. Зростання чисельності населення в містах часто призводить до виникнення ситуацій скупчення людей у густонаселених районах, що створює низку викликів і загроз для громадської безпеки. Щоб забезпечити безпеку людей, стратегії управління натовпом вимагають ефективного аналізу натовпу. У той час як ручний аналіз натовпу є рутинним завданням, автоматичний аналіз натовпу є бажаним у багатьох ситуаціях [1].

Автоматичний моніторинг натовпу за допомогою візуального аналізу є актуальною темою в дослідженнях комп'ютерного зору з багатьма цікавими застосуваннями в міському спостереженні, соціальному дистанціюванні, транспорті, спорті, моніторингу дикої природи і так далі. За останнє десятиліття досягнення в галузі глибокого навчання принесли нові можливості для досягнення найсучасніших результатів у різних завданнях візуального аналізу натовпу, таких як підрахунок натовпу, виявлення об'єктів, розпізнавання активності, виявлення аномалій, аналіз руху і так далі.

Незважаючи на те, що в результаті численних досліджень було досягнуто значних успіхів у візуальному аналізі натовпу, все ще залишається ряд проблем і завдань, які необхідно вирішити. Причина багатьох невирішених проблем криється в основній складності та викликах, пов'язаних зі сценами натовпу, порівняно з іншими завданнями комп'ютерного зору, наприклад, сильні оклюзії, скупчення, варіації масштабу, непередбачувані патерни руху, складна поведінка натовпу, невідомий контекст діяльності натовпу і так далі [2].

Таким чином, аналіз натовпу часто вважається більш складним завданням, ніж інші завдання комп'ютерного зору. Складний характер проблеми впливає на розвиток системи візуального аналізу натовпу і вимагає більш складних алгоритмів і моделей, збору різноманітних і масштабних наборів даних, апаратних ресурсів для роботи в реальному часі, а також системної інтеграції таких алгоритмів з камерами, сенсорами і системами зберігання даних.

Роботи з візуального аналізу натовпу поділяються на шість основних напрямків, а саме: підрахунок натовпу, виявлення та відстеження об'єктів, аналіз руху, розпізнавання поведінки, виявлення аномалій та прогнозування натовпу [3].

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/deed.uk>)

Підрахунок натовпу полягає в оцінці щільності людей або загальної кількості в певній географічній зоні. Виявлення об'єктів має на меті виявити та локалізувати певні об'єкти, що становлять інтерес у натовпі, наприклад, виявлення лише жінок, виявлення людей з транспарантами або палицями тощо. Аналіз руху стосується стану колективної мобільності натовпу, наприклад, чи є натовп нерухомим/рухомим, а також інших статистичних даних про рух, таких як напрямок, швидкість, аналіз потоку тощо. Аналіз поведінки визначає колективну характеристику натовпу, зосереджуючись на діях, які виконують члени натовпу, щоб отримати контекстну інформацію про поведінку, наприклад, чи є натовп спокійним, агресивним і так далі. Виявлення аномалій фокусується на пошуку незвичайних і аномальних подій і дій як на індивідуальному, так і на груповому рівнях. Прогнозування натовпу відноситься до прогнозування активного накопичення (збору) або припливу і відтоку людей в/з певного регіону, що може призвести до утворення натовпу [4].

Як правило, існує логічна послідовність у реалізації цих задач. Наприклад, першою інформацією, яка може зацікавити користувача або службу, що здійснює моніторинг натовпу, є оціночна щільність натовпу, за якою слідують інші аспекти учасників натовпу, такі як вік, стать і так далі, а також будь-які виявлені об'єкти, такі як банери, плакати і палиці, які тримають учасники натовпу. Наступна інформація, яка може бути цікавою – це постійне спостереження за рухом натовпу, щоб знати, чи є він нерухомим, чи рухається в певному напрямку. Вищезгадані аспекти охоплюються першими трьома напрямками аналізу натовпу. Потім складніша система може забезпечити більш детальний аналіз, наприклад, поведінку натовпу (настрій, конкретні дії, що виконуються членами натовпу), а також виявлення конкретних аномальних подій або об'єктів. І, нарешті, дуже корисною інформацією, яку бажано отримати, може бути прогноз щодо формування або розсіювання натовпу в певному регіоні або районі в певний час [5].

Хоча традиційно було запропоновано багато напівавтоматизованих методів комп'ютерного зору, нещодавні досягнення в сучасному глибокому навчанні зробили революцію в розробці повністю автоматизованих додатків для моніторингу натовпу на основі зору. Завдяки використанню можливостей глибоких нейронних мереж, точність, ефективність і загальна продуктивність таких додатків були значно покращені.

Методи глибокого навчання автоматично вивчають і виділяють значущі закономірності та особливості з великомасштабних візуальних даних про натовп, таких як зображення і відео, що дає змогу краще аналізувати натовп. Крім того, використання навчання з перенесенням дозволяє розгортати моделі, навчені на одному наборі даних, в іншому сценарії після тонкої настройки, що прискорює процес навчання [6].

Метою роботи є дослідження перспектив використання моделей та методів візуального аналізу натовпу в системах інтелектуального відеоспостереження.

Результати досліджень. Підрахунок кількості людей та оцінка щільності натовпу. Одним з перших аспектів моніторингу натовпу є оцінка кількості людей (скалярне значення для всього зображення) або щільності в різних частинах сцени. Оскільки люди в натовпі зазвичай об'єднані в групи, оцінка щільності дає більше інформації, ніж просто загальний підрахунок. Оцінка кількості або щільності може забезпечити хорошу обізнаність про ситуацію для суб'єктів моніторингу, таких як правоохоронні органи та організатори заходів.

Фактичним методом підрахунку людей на зображенні або відеокадрі є оцінка щільності. Модель на основі згорткової нейронної мережі навчається оцінювати щільність натовпу. У методі оцінки щільності кожна голова визначається за допомогою гаусової плями навколо центру голови (Рис. 1). Це задача регресії на рівні пікселів, і для навчання моделі використовується загальнозживана евклідова функція втрат (або середньоквадратична помилка MSE) [7].

Нещодавно були запропоновані моделі, що не використовують CNN, але використовують візуальні трансформатори. Моделі на основі трансформаторів застосовують механізми самоуваги для моделювання глобального контексту та фіксації довгострокових

залежностей. Їхнім недоліком є те, що ці моделі потребують більше обчислювальних ресурсів порівняно з CNN через більшу кількість параметрів та операцій самоуваги [8].

Незважаючи на значне покращення точності підрахунку, в даній області все ще залишається кілька проблем. Проблема варіацій масштабу, спричинених ефектом перспективи, було подолано за допомогою багатомасштабних архітектур, але інші проблеми, такі як оклюзії та складне тло, все ще залишаються серйозною проблемою для багатьох складних сцен. Хоча деякі методи пропонують сегментацію фону як крок попередньої обробки, це може збільшити складність завдання [9].

Виявлення та відстеження об'єктів. Існує багато ситуацій, в яких необхідно виявити об'єкт, що цікавить, у відеокадрі, а потім відстежити його в часі в послідовних кадрах (Рис. 2) [10].

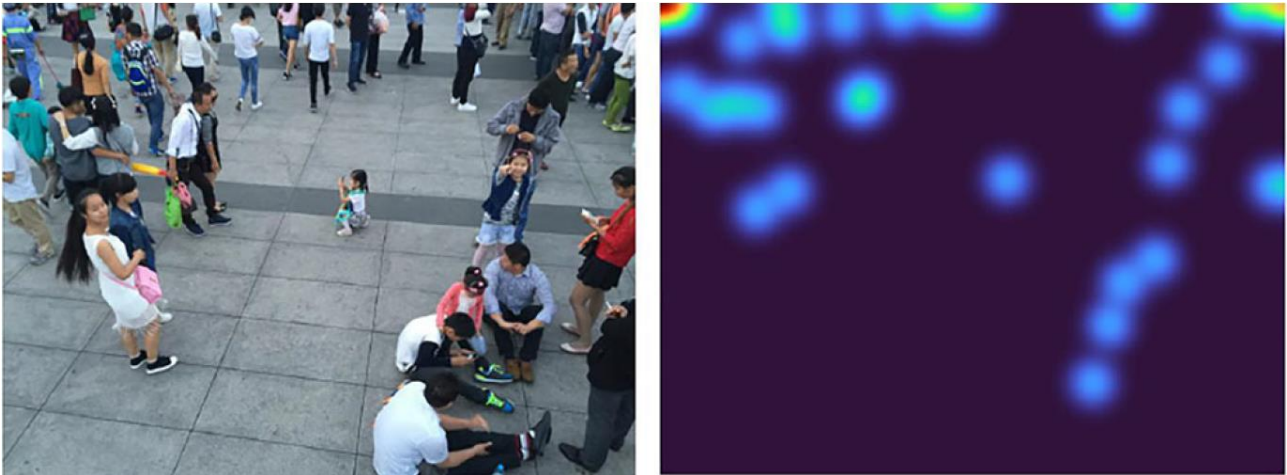


Рис. 1. Зображення натовпу та карта прогнозованої щільності, яка показує щільність натовпу по всій сцені зображення [9]

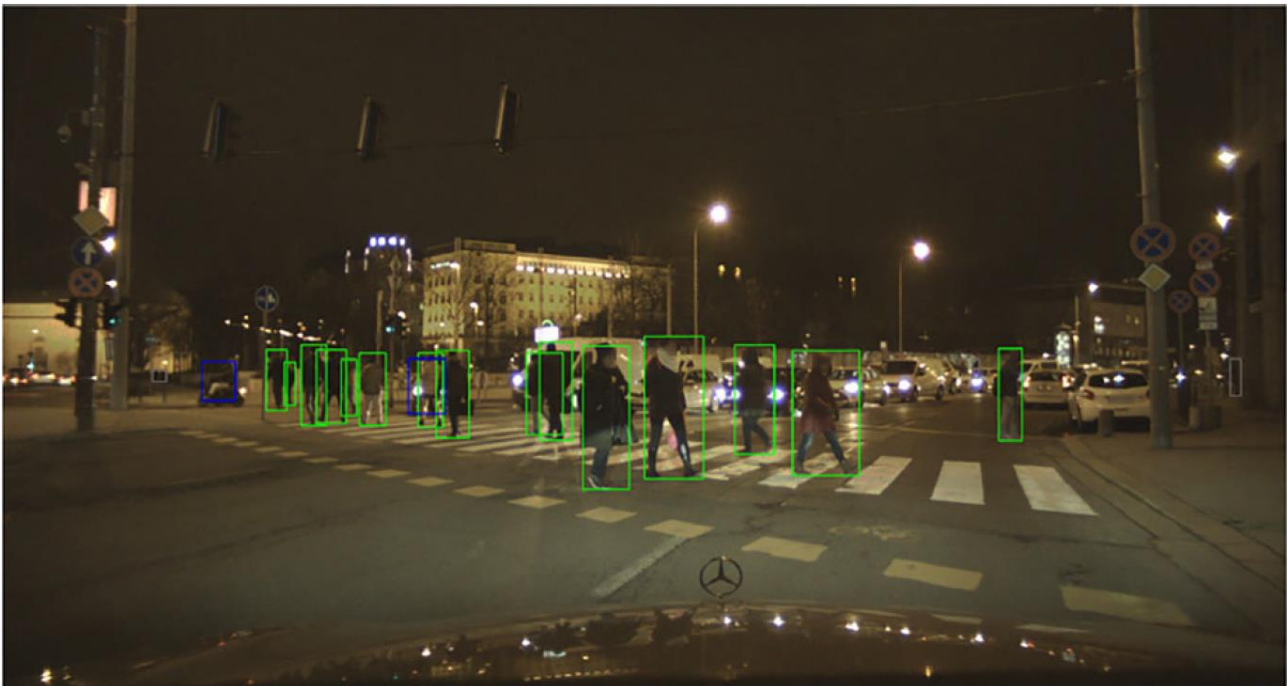


Рис. 2. Виявлення об'єктів з набору даних ECPersons [10]

Виявлення об'єктів є добре дослідженою проблемою, яка привернула значну увагу в комп'ютерному зорі, і з часом було розроблено кілька видатних моделей, які досягли найсучасніших показників. Ці моделі зазвичай поділяються на дві категорії: методи на основі якоря і моделі без якоря. Методи на основі якорів використовують заздалегідь визначений набір опорних точок, розміщених по всьому зображенню, і прогнозують кінцевий набір точок навколо виявлених об'єктів. Методи на основі якорів добре адаптуються до різних масштабів і співвідношень сторін і добре працюють у складних сценах, однак на їхню продуктивність сильно впливають різні фактори, такі як розміри, співвідношення сторін, кількість якірних блоків і варіації форми. Методи без якоря є відносно новими і більш ефективними, ніж методи на основі якорів. Вони використовують підхід виявлення ключових точок [11].

За останні кілька років методи на основі якоря та методи без якоря досягли значного приросту продуктивності, але обидва типи методи все ще стикаються з деякими внутрішніми обмеженнями в контексті сцен з натовпом. Хоча деякі з цих проблем є загальними для будь-яких завдань комп'ютерного зору, деякі з них є більш серйозними в сценах з великим скупченням людей: зміна точки зору – коли об'єкт виглядає по-різному при зйомці з різних кутів; деформація об'єкта – коли об'єкт має різну форму в одному кадрі або в наступних кадрах відео (наприклад, людина нахилиється); сильні оклюзії – коли один або кілька об'єктів частково не видно на зображенні через перекриття іншими об'єктами перед ними; освітлення – коли на зображенні спостерігаються великі варіації значень яскравості пікселів; засмічення – коли зображення містить багато або великі об'єкти, які не є об'єктами інтересу. Варто зазначити, що існуючі моделі для виявлення об'єктів не дуже добре підходять для середовища натовпу, а отже, незважаючи на доопрацювання, дають погані результати. Це серйозна проблема, яка перешкоджає впровадженню серйозних додатків для спостереження (наприклад, тих, що можуть бути використані правоохоронними органами) [12].

Аналіз руху. Розуміння динаміки руху натовпу може надати більше корисної інформації на додаток до підрахунку кількості та оцінки щільності натовпу. Наприклад, може бути цікаво дізнатися, чи є натовп нерухомим або рухомим. Для рухомого натовпу буде цікаво зрозуміти потік натовпу та інші закономірності, пов'язані з рухом натовпу, включаючи траєкторію, напрямок, швидкість і так далі. Сюди також входить виявлення нерухомих груп у натовпі. Аналіз руху має багато цікавих застосувань, наприклад, контроль доступу, ідентифікація людини, аналіз заторів і багатокамерне інтерактивне спостереження. Аналіз руху може стосуватися розпізнавання руху частин тіла людини (наприклад, жестів, дій тощо), але в контексті натовпу його часто називають узгодженим рухом групи людей [13].

Аналіз руху натовпу представляє великий інтерес для розуміння аналізу поведінки натовпу і розуміння сцени, наприклад, класифікації натовпу як нерухомого або рухомого, прогнозування траєкторії руху натовпу, аналізу потоку натовпу і аналізу закономірностей руху натовпу.

Методи аналізу руху натовпу спрямовані на виявлення, відстеження та аналіз патернів руху, щоб зробити важливі висновки про динамічну поведінку натовпу. З технічної точки зору, аналіз руху включає темпо-просторовий аналіз натовпу. Існує кілька ручних і наскрізних автоматизованих методів і математичних моделей для статистики руху натовпу. Традиційно в аналізі руху використовуються такі методи, як сегментація руху (попіксельне відокремлення рухомих об'єктів від фону), часове диференціювання (попіксельне розрізнення послідовних кадрів) та оптичний потік (використання векторів потоку рухомих об'єктів у часі). Наприклад, для вимірювання потоків натовпу використовується підрахунок траєкторій людей, що проходять через визначені користувачем лінії на сцені [14].

Більшість досліджень з аналізу руху натовпу ґрунтується на представленні графічної інформації, такої як колір, текстура тощо. Однак ці методи забезпечують лише базовий аналіз патернів руху і не надають достатньої семантичної інформації про рух натовпу. Автоматизований аналіз руху вимагає не тільки великих наборів відеоданих, але й ефективних методів для вилучення мікро- та макростатистики руху натовпу. Існуючим

методам серйозно бракує досконалості, необхідної для реалізації в реальних умовах. Для вилучення більш високорівневої та інтуїтивно зрозумілої інформації про рух моделі повинні автоматично вивчати семантичні ознаки та взаємозв'язки між низькорівневими зображувальними ознаками та високорівневими семантичними ознаками [15].

Розпізнавання поведінки, дій та контексту. Аналіз поведінки – це аналіз натовпу в цілому або його частини на більш тривалому часовому відрізку (від хвилин до годин). Він може включати простіші задачі, такі як визначення стану поведінки натовпу, наприклад, спокійний, активний, буйний натовп, або може включати складне розпізнавання активності.

Розпізнавання дій означає виявлення різних згрупованих дій членів натовпу, таких як протест, танці, бійки тощо, зазвичай у короткому часовому діапазоні (від секунд до хвилин). Це також може стосуватися дій окремого об'єкта, наприклад, оцінки пози. Люди в натовпі можуть взаємодіяти один з одним і займатися різними видами діяльності [16].

У той час як розпізнавання активності виявляє активність натовпу за формою, позою та особливостями руху, контекстний аналіз вивчає соціальну взаємодію між членами групи, використовуючи час, місцезнаходження та іншу контекстну інформацію, а також їхнє відношення до натовпу. Контекстний аналіз активності натовпу є складнішою проблемою, ніж розпізнавання активності, через багато інших факторів, наприклад, навколишнє середовище та інші фактори, пов'язані з психологією та соціологією [17].

Розпізнавання діяльності передбачає створення системи для визначення індивідуальних і групових видів діяльності, а потім присвоєння кожному виду діяльності унікальних дескрипторів.

Наступним кроком є точне вивчення просторових і часових профілів кожної активності за допомогою традиційних методів, таких як математичні моделі або машинне навчання, або застосування наскрізного навчання з використанням моделей глибокого навчання. Аналіз поведінки зазвичай розглядається як макроскопічний аналіз натовпу, тоді як розпізнавання активності можна віднести до мікроскопічного аналізу натовпу.

Розпізнавання активності та аналіз поведінки, як правило, більш складні через семантичні зв'язки між виявленими діями та звичками людини. Одна поведінка зазвичай може бути зіставлена з кількома семантичними концепціями. Таким чином, щоб зробити висновок про значущу семантичну поведінку, необхідно апріорі встановити точний взаємозв'язок між низькорівневими ознаками і семантичною поведінкою. Зазвичай дуже важко виявити і відстежити окремих людей у різних сценах великого скупчення людей. Деякі ключові проблеми в аналізі поведінки включають застосування фонових знань і теорії міркувань для правильного визначення дескрипторів природної мови для семантичної поведінки, а потім вивчення цієї поведінки на основі трансформацій об'єкта в сцені на різних рівнях [18].

Виявлення аномалій. Виявлення аномалій – це пошук аномальних (або ненормальних) подій і має багато важливих застосувань, наприклад, виявлення злочинів, порушень правил дорожнього руху, покинутих об'єктів, виявлення зброї тощо. У контексті моніторингу натовпу аномалії виявляються за допомогою просторово-часового аналізу ознак відеокadrів, а також на одному зображенні. Виявлення аномалій натовпу є складним завданням, головним чином через рідкісне виникнення таких подій і, відповідно, відсутність достатньої кількості даних про аномальні події. Визначення аномалії є суб'єктивним, і одна і та ж подія може бути класифікована як нормальна і аномальна в різний час і в різних місцях. Аномалії можна розділити на (i) точкові (коли окремих об'єктів або особа виглядає або поводиться інакше, ніж інші об'єкти на місці події), (ii) контекстуальні (коли об'єкт або предмет розглядається як аномальний у певній контекстуальній ситуації або середовищі) і (iii) колективні (коли група об'єктів поводиться інакше, ніж решта об'єктів на місці події). Виявлення аномалій також класифікується як глобальна аномалія (чи є аномалія у сцені/кадрі) та локальна аномалія (локалізація аномалії у кадрі або відео) [19].

Найпростішим і найпоширенішим підходом до виявлення аномалій є навчання однокласового класифікатора (ОСС), який навчається на даних, що містять нормальні

прикладі. Модель ОСС, навчена на достатньо великій кількості нормальних навчальних прикладів, може передбачати аномальні події. Однак навіть зібрати дані з усіма нормальними подіями непросто. Виявлення аномалій у переповнених сценах вирішувалося багатьма способами з використанням різноманітних алгоритмів, включаючи класичні схеми машинного навчання, наприклад, k -середні та SVM, GMM тощо [20].

Виявлення аномалій є дуже складним завданням з кількох причин, включаючи доступність даних, вимоги до обчислювальної потужності, справедливості та узагальнення.

Деякі з відкритих проблем і викликів перераховані нижче:

По-перше, не існує універсального визначення аномальної події, тобто подія, яка вважається аномальною, може вважатися аномальною в іншому контексті. Наприклад, людина з пістолетом є ненормальною, але стає нормальною, коли ця людина є поліцейським констеблем. Таким чином, контекст завжди важливий для виявлення аномалій, що робить проблему виявлення аномалій дуже складною. Це серйозний виклик, який необхідно вирішити в першу чергу, щоб покращити результати досліджень з виявлення аномалій. Без достатньої кількості даних і стандартних визначень аномалій натовпу результати досліджень будуть значно обмежені.

По-друге, бракує якісних наборів даних для виявлення аномалій. Існуючі набори даних охоплюють лише невелику кількість аномалій. Методи маркування даних у цих наборах даних також різняться (аномалії на рівні кадру, аномалії на рівні відео, аномалії на рівні сегмента тощо).

По-третє, нелегко створювати великі набори даних з різноманітними аномаліями, оскільки кілька аномальних подій не можуть бути отримані заздалегідь [21].

Прогнозування натовпу. Прогнозування натовпу – це завчасне передбачення скупчення людей у певному регіоні. Завчасне прогнозування натовпу має ключове значення для різних сценаріїв і застосувань, таких як заходи, виявлення інцидентів, туристичні атракції тощо. Прогнозування натовпу, як правило, є складним і часто використовує мультимодальні дані з різних джерел, таких як камери спостереження, соціальні мережі, сигнали мобільних телефонів, датчики тощо. Машинне навчання та методи статистичного моделювання використовуються для вивчення та прогнозування майбутнього натовпу [22]. Цей процес може стосуватися конкретних масових заходів або зібрань, оцінки тривалості заходу, прогнозування пікових годин скупчення людей або передбачення популярності певних районів чи пам'яток під час заходу.

Прогнозування натовпу можна просто змодельовати як задачу прогнозування часових рядів для регресії збільшення щільності натовпу з часом, починаючи зі стану відсутності натовпу, або використовувати просторову статистику натовпу для прогнозування натовпу в просторових регіонах. Традиційно він використовує методи ковзного усереднення, такі як авторегресійне інтегроване ковзне середнє (ARIMA). Однак ці методи не здатні врахувати складні часові та просторові залежності, незважаючи на низку методів інженерії ознак [23].

Щоб впоратися зі складними залежностями, були запропоновані методи глибокого навчання, включаючи ШНМ, LSTM та графові нейронні мережі (GNNs). Хоча мультимодальні дані можуть бути використані для прогнозування натовпів, достатньої кількості таких наборів даних не існує. Використання LSTM та CNN разом є більш перспективним методом для прогнозування натовпів на короткі проміжки часу. Однак для більш тривалих часових проміжків застосовуються методи прогнозування (ARIMA) [24].

Прогнозування натовпу можна або логічно вивести з аналізу моделей руху натовпу, або спрогнозувати безпосередньо за допомогою моделей прогнозування часових рядів. Однак дослідження в цій галузі дуже обмежені, і дуже мало робіт, які часто використовують методи глибокого навчання, такі як CNN і LSTM.

Прогнозування натовпу має як просторову, так і часову залежність, і, як правило, досить складно передбачити скупчення натовпу в регіоні протягом тривалих періодів. Прогнозування у витоках і вузьких місцях є більш складним, ніж у схемі руху по смугах.

Через просторову залежність прогнозування потоку натовпу в регіонах неправильної форми може бути складнішим, ніж у регіонах правильної форми [25].

Висновки. У цій роботі визначено шість основних напрямків візуального аналізу натовпу, які разом утворюють повноцінну автоматизовану систему моніторингу натовпу. Кожна з цих шести сфер передбачає різний рівень складності, а отже, сучасний рівень розвитку в них сильно відрізняється. Наприклад, підрахунок та виявлення натовпу – це дві області, в яких нещодавно було досягнуто значного покращення результатів на великих масивах даних. Як правило, існуючі реальні реалізації моніторингу натовпу охоплюють ці завдання.

Завдання аналізу руху можуть бути різними, наприклад, виявлення джерела/відтоку натовпу, визначення траєкторії, швидкості тощо. Аналіз поведінки і виявлення аномалій, які іноді перетинаються, є найскладнішими завданнями, і прогрес в цих завданнях все ще дуже обмежений і розрізнений з точки зору методів, підходів, припущень і цілей, незважаючи на їх важливість в ряді випадків використання.

Відсутність визначення аномалій, діяльності та поведінки змушує дослідників використовувати різні цілі та метрики оцінки, що робить бенчмаркінг несправедливим. Майбутні дослідження повинні зосередитися на розробці загальних визначень діяльності та аномалій з урахуванням контексту, а також на розширенні існуючих наборів даних і створенні більших і збалансованих наборів даних.

Підходи, натхненні фізикою (наприклад, енергетичні моделі), є цікавими напрямками для реалізації виявлення аномалій. У найближчому майбутньому прогнозується швидкий прогрес в маловивчених сферах завдяки нещодавнім розробкам у галузі генеративного ШІ, який допоможе впоратися з потребою в більшій кількості навчальних даних.

При спостереженні за допомогою камер відеоспостереження зручніше виконувати всі завдання обробки та висновків на локальному сервері завдяки високошвидкісному дротовому з'єднанню, однак при повітряному спостереженні (за допомогою дронів) в деяких випадках обробка та висновки на пристрої можуть бути більш зручними. Таким чином, перевага надаватиметься легким моделям аналізу натовпу.

СПИСОК ЛІТЕРАТУРИ

1. Alotaibi S. S., et al. “Deep transfer learning enabled intelligent object detection for crowd density analysis on video surveillance systems”. *Applied Sciences*. 2022; 12 (13): 6665. DOI: <https://doi.org/10.3390/app12136665>.
2. Bahmanyar R., Vig E., Reinartz P. “MRCNet: Crowd counting and density map estimation in aerial and ground imagery”. 2019. DOI: <https://doi.org/10.48550/arXiv.1909.12743>.
3. Duong H.-T., et al. “Deep learning-based anomaly detection in video surveillance: a survey”. *Sensors*. 2023; 23 (11): 5024. DOI: <https://doi.org/10.3390/s23115024>.
4. Sharma V., Mir R. N., Singh C. “Scale-Aware CNN for crowd density estimation and crowd behavior analysis”. *Computers and Electrical Engineering*. 2023; 106: 108569.
5. Wang Q., Gao J., Yuan Y. “Learning from synthetic data for crowd counting in the wild”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019. p. 8198–8207. DOI: <https://doi.org/10.48550/arXiv.1903.03303>.
6. Ullah A., et al. “Efficient activity recognition using lightweight CNN and DS-GRU network for surveillance applications”. *Applied Soft Computing*, 2021; 103: 107102. DOI: <https://doi.org/10.1016/j.asoc.2021.107102>.
7. Almagbile A. “Estimation of crowd density from UAVs images based on corner detection procedures and clustering analysis”. *Geo-Spatial Information Science*. 2019; 22 (1): 23–34. DOI: <https://doi.org/10.1080/10095020.2018.1539553>.
8. Xu C., et al. “Learn to scale: Generating multipolar normalized density maps for crowd counting”. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019. p. 8382–8390. DOI: <https://doi.org/10.48550/arXiv.1907.12428>.

9. Liu J., Gao C., Hauptmann A. G. “DecideNet: Counting varying density crowds through attention guided detection and density estimation”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. p. 5197–5206. DOI: <https://doi.org/10.1109/CVPR.2018.00545>.
10. Zhang Y., et al. “Single-Image crowd counting via multi-column convolutional neural network”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016. p. 589–597. DOI: <https://doi.org/10.1109/CVPR.2016.70>.
11. Sindagi V. A., Patel V. M. “HA-CCN: Hierarchical attention-based crowd counting network”. *IEEE Transactions on Image Processing*. 2019; 29: 323–335. DOI: <https://doi.org/10.1109/TIP.2019.2928634>.
12. Liu M., et al. “Crowd counting with fully convolutional neural network”. *25th IEEE International Conference on Image Processing (ICIP)*. 2018. p. 953–957. DOI: <https://doi.org/10.1109/ICIP.2018.8451787>.
13. Peng T., Li Q., Zhu P. “RGB-T crowd counting from drone: A Benchmark and MMCCN network”. *Proceedings of the Asian Conference on Computer Vision*. 2020; 12627. https://doi.org/10.1007/978-3-030-69544-6_30.
14. Hsieh M. R., Lin Y. L., and Hsu W. H. “Drone-Based object counting by spatially regularized regional proposal network”. *Proceedings of the IEEE International Conference on Computer Vision*. 2017. p. 4145–4153. DOI: <https://doi.org/10.1109/ICCV.2017.446>.
15. Chen M., Wang Q., and Li X. “Anchor-Based group detection in crowd scenes.” *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017. p. 1378–1382. DOI: <https://doi.org/10.1109/ICASSP.2017.7952382>.
16. Kuehne H., Jhuang H., Garrote E., Poggio T., and Serre T. “HMDB: A Large Video Database for Human Motion Recognition”. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2011. p. 2556–2563. DOI: [10.1109/ICCV.2011.6126543](https://doi.org/10.1109/ICCV.2011.6126543). DOI: <https://doi.org/10.1109/ICCV.2011.6126543>.
17. Gupta A., Johnson J., Fei-Fei L., Savarese S., and Alahi, A. “Social GAN: socially acceptable trajectories with generative adversarial networks”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. p. 2255–2264. DOI: <https://doi.org/10.1109/CVPR.2018.00240>.
18. Wang Z., Cheng C., and Wang X. “A fast crowd segmentation method”. *International Conference on Audio, Language and Image Processing (ICALIP)*. 2018. p. 242–245. DOI: <https://doi.org/10.1109/ICALIP.2018.8455441>.
19. Perez M., Kot A. C., and Rocha A. “Detection of real-world fights in surveillance videos”. *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2019. p. 2662–2666. DOI: <https://doi.org/10.1109/ICASSP.2019.8683676>.
20. Spampinato C., Palazzo S., and Giordano D. “Evaluation of tracking algorithm performance without ground-truth data”. *19th IEEE International Conference on Image Processing*. 2012. p. 1345–1348. DOI: <https://doi.org/10.1109/ICIP.2012.6467117>.
21. Peter T., Richard J. M., Murray M., and Krause, A. Robot “Navigation in dense human crowds: The case for cooperation”. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. 2013. p. 2153–2160. DOI: <https://doi.org/10.1109/ICRA.2013.6630866>.
22. Marsden M., McGuinness K., Little S., and O’Connor N. E. “Holistic features for real-time crowd behaviour anomaly detection”. *IEEE International Conference on Image Processing (ICIP)*. 2016. p. 918–922. DOI: <https://doi.org/10.1109/ICIP.2016.7532491>.
23. Rabiee H., Haddadnia J., Mousavi H., Kalantarzadeh M., Nabi M., and Murino V. “Novel dataset for fine-grained abnormal behavior understanding in crowd”. *13th IEEE International Conference on Advanced Video and Signal Based Surveillance*. 2016. p. 95–101. DOI: <https://doi.org/10.1109/AVSS.2016.7738074>.
24. Dupont C., Tobías L., and Luvison B. “Crowd-11: A Dataset for fine grained crowd behaviour analysis”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017. p. 9–16. DOI: <https://doi.org/10.1109/CVPRW.2017.271>.

25. Azorin-Lopez J., Saval-Calvo M., Fuster-Guillo A., Garcia-Rodriguez J., Cazorla M., and Signes-Pont M. T. “Group activity description and recognition based on trajectory analysis and neural networks”. *International Joint Conference on Neural Networks (IJCNN)*. 2016. p. 1585–1592. DOI: <https://ieeexplore.ieee.org/document/7727387>.

DOI: <https://doi.org/10.15276/ict.01.2024.32>

UDC 004.942

Tasks of visual crowd analysis in intelligent video surveillance systems

Ruslan Ye. Dobryshev

PhD Student, Department of Artificial Intelligence and Data Analysis

ORCID: <https://orcid.org/0009-0007-8639-3157>; rdobryshev@gmail.com

Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

ABSTRACT

Modern urban population growth creates challenges for public safety, in particular due to crowds. This stimulates the development of new crowd management methods that require automated analysis. Visual crowd analysis based on computer vision technologies is a key tool for solving these problems. The development of deep learning has significantly improved the monitoring systems used for urban surveillance, social distancing control, transportation and event management. However, crowd analysis remains challenging due to occlusions, scale variations, unpredictable movement patterns, and complex behavior. To overcome these challenges, new algorithms, models, and large-scale datasets are needed to enable real-time analysis. The main tasks include people counting, object detection, motion analysis, behavior recognition, and anomaly detection. Deep neural networks and transfer learning significantly increase the accuracy and adaptability of such systems, which helps to improve public safety and manage the flow of people.

Keywords: Intelligent video surveillance; Machine learning; Deep learning; Behavioral recognition; Motion analysis