DOI: https://doi.org/10.15276/eltecs.42.118.2025.3
УДК 004.932.2:004.89

Оцінка пози людини в інтелектуальних системах відеоспостереження: сучасні підходи та виклики

Р. В. Ковалевич, Д. Д. Курінько, В. І. Кривда
Національний університет «Одеська політехніка»

Анотація. *У статті представлений огляд сучасних методів оцінки пози людини в інтелектуальних системах відеоспостереження. Розглянуто 2D та 3D підходи, включаючи як класичні методи, так і моделі на основі глибокого навчання. Проаналізовано особливості top-down і bottom-up стратегій, їх переваги та обмеження. Визначено основні виклики, пов'язані з реальними умовами спостереження, зокрема оклюзіями, зміною освітлення та обмеженою якістю зображень. Окреслено перспективи розвитку галузі, зокрема використання 6D оцінки пози, мультисенсорних даних і самонавчальних моделей.*

Ключові слова: *інтелектуальні системи, відеоспостереження, аналіз поведінки, оцінка пози, нейронні мережі, теплові карти, багатокамерні системи, оклюзії.*

Вступ

У сучасному світі відеоспостереження стало невід'ємною складовою систем безпеки, управління міською інфраструктурою, моніторингу громадського порядку та контролю доступу.

Традиційні системи відеоспостереження здебільшого виконували пасивну функцію – накопичення відеоданих для подальшого перегляду оператором. Проте зі стрімким розвитком технологій комп'ютерного зору та глибокого навчання виникла нова парадигма – інтелектуальні системи відеоспостереження, здатні до автоматизованої інтерпретації сцени в реальному часі [1].

Одним з ключових напрямів розвитку таких систем є оцінка пози людини (англ. human pose estimation) — процес визначення просторового положення ключових точок людського тіла на основі зображення або відео потоку [2].

Інформація про позу є основою для більш складних завдань, таких як розпізнавання активності, виявлення аномальної поведінки, відстеження осіб, інтерфейси «людина-машина» та навіть передиктивна аналітика у сфері безпеки.

Сучасні методи оцінки пози поділяються на два основні класи — двовимірна (2D) та тривимірна (3D) оцінка. 2D-методи передбачають локалізацію ключових точок тіла на площині зображення й зазвичай реалізуються за допомогою згорткових нейронних мереж, які прогнозують координати основних суглобів у пікселях [3].

Вони є менш ресурс втратними та широко застосовуються в реальному часі у відеоспостереженні.

Водночас 3D-оцінка пози надає глибше розуміння просторової конфігурації тіла, дозволяє враховувати перспективні спотворення та отримувати положення частин тіла у фізичному просторі. Проте такі методи зазвичай вимагають або використання глибинних сенсорів, або складних моделей реконструкції на основі одного або кількох зображень.

Незважаючи на активний прогрес, задача оцінки пози в умовах реальних систем відеоспостереження залишається складною. На результати негативно впливають такі чинники, як змінні умови освітлення, шум, часткове перекриття об'єктів, різні ракурси камер, обмеження роздільної здатності тощо. В умовах публічного середовища додаткову складність створює наявність кількох людей у кадрі та постійна динаміка сцени (рис. 1) [4].

В реальних умовах виявлення людей і прогнозування пози стикаються з різними перешкодами: залізничні колії, кабелі електропередач і трамваї, що рухаються, блокують різні частини зображення; кут нахилу камери створює розбіжності в розмірах людини [4].

З огляду на актуальність проблематики, у науковій спільноті зростає інтерес до систематизації знань у цій галузі, порівняння наявних підходів, аналізу їх ефективності, а також визначення векторів майбутніх досліджень.

Метою даної статті є систематизація сучасних підходів до оцінки пози людини в інтелектуальних системах відеоспостереження, аналіз

Рис.1. Реальні умови виявлення людей і прогнозування пози.

основних методів 2D та 3D оцінки пози, їх переваг, недоліків та сфер застосування, а також окреслення актуальних викликів і перспектив подальшого розвитку цієї області досліджень.

1. Загальна постановка задачі

Оцінка пози (положення та конфігурації тіла) людини є однією з ключових задач у галузі комп'ютерного зору, зокрема в контексті інтелектуальних систем відеоспостереження. Метою такої задачі є автоматичне виявлення та відстеження просторового положення ключових точок тіла людини на основі вхідних відеоданих.

Вхідні дані. Розглянемо відеопотік, який представляється як послідовність кадрів:

$$I = \{i_1, i_2, \dots, i_t\}, \;\; i_t \in \mathbb{R}^{H \times W \times 3}, \qquad (1)$$

де t – кількість кадрів, H та W – висота та ширина кадру відповідно, i_t – кольорове зображення в момент часу t.

Задача. Для кожного кадру i_t необхідно визначити набір ключових точок $K_t = \{k_1, k_2, \dots, k_n\}$, які відповідають основним анатомічним орієнтирам тіла людини: голова, плечі, лікті, кисті, таз, коліна, стопи тощо. Кожна точка k_i задається координатами на площині зображення:

$$k_i = (x_i, y_i), x_i \in [0, W], y_i \in [0, H]. \qquad (2)$$

У випадку тривимірної оцінки пози використовується просторове представлення:

$$k_i = (x_i, y_i, z_i), \;\; (x_i, y_i, z_i) \in \mathbb{R}^3, \qquad (3)$$

де z_i відображає глибину точки в просторі або відстань до камери.

У випадку, коли на зображенні присутні декілька людей, поза кожної людини визначається окремо. Тоді вихідний набір поз для кадру має вигляд:

$$P_t = \left\{K_t^{(1)}, K_t^{(2)}, \dots, K_t^{(M_t)}\right\}, \qquad (4)$$

де M_t – кількість людей у кадрі в момент часу t, $K_t^{(j)}$ – набір ключових точок для j-ої людини.

Мета – знайти відображення:

$$f: I \to \{P_1, P_2, \dots, P_t\}, \qquad (5)$$

яке для кожного кадру i_t забезпечує точне і стійке визначення просторового положення ключових точок усіх людей у сцені.

Вимоги до системи розпізнавання пози.

Точність – висока відповідність оцінених координат ключових точок істинним (референтним) значенням. Може оцінюватися, наприклад, за метрикою PCK (Percentage of Correct Keypoints), як відсоток правильних ключових точок [5]:

$$PCK@\alpha = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\frac{\left\|\hat{k}_i - k_i\right\|_2}{d} < \alpha\right), \qquad (6)$$

де $\hat{k}_i$ – прогнозована позиція точки, k_i – істинна позиція точки, d – нормалізуючий коефіцієнт

(наприклад, відстань між плечима), α – заданий поріг.

Стійкість – система повинна коректно працювати в умовах часткового перекриття, різного освітлення, різних масштабів та ракурсів.

Продуктивність – алгоритм повинен працювати в режимі реального часу або близькому до реального, з мінімальною затримкою обробки (наприклад, не менше 15–30 кадрів за секунду).

Масштабованість – можливість одночасної обробки кількох людей в кадрі.

2. Двовимірна та тривимірна оцінка пози людини в інтелектуальних системах відеоспостереження

У контексті інтелектуальних систем відеоспостереження, які призначені для автоматизованого аналізу поведінки людини, ключовим етапом оцінки пози людини є локалізація та інтерпретація пози у дво- або тривимірному просторі.

Двовимірна оцінка пози. У більшості практичних застосувань відеоспостереження використовується двовимірна оцінка пози (2D pose estimation).

Це пов'язано з відсутністю просторової інформації про глибину сцени у стандартних відеокамерах та вимогою до реального часу обробки. Двовимірна оцінка полягає у знаходженні координат N ключових точок людського тіла на площині зображення:

$$K^{(2D)} = \{k_i \in \mathbb{R}^2 | k_i = (x_i, y_i), i = 1,2, \dots, N\}. \quad (7)$$

Ключові точки можуть включати: верхівку голови, шию, плечі, лікті, кисті, таз, коліна та щиколотки. Визначення таких точок часто реалізується за допомогою згорткових нейронних мереж, які прогнозують теплові карти (heatmaps):

Рис.2. Ілюстрація задачі двовимірної оцінки пози [6]

$$H_i: \mathbb{R}^{H \times W} \to [0,1], \qquad i = 1,2, \dots, N, \quad (8)$$

де значення $H_i(x, y)$ характеризує ймовірність наявності i-ої точки у пікселі (x, y). Найімовірніше положення точки визначається як:

$$k_i = \arg\max_{(x,y)} H_i(x, y). \quad (9)$$

Двовимірна поз є інваріантною до абсолютного масштабу сцени, однак не дозволяє точно визначити глибину об'єкта або просторові кути суглобів.

Тривимірна оцінка пози людини. На відміну від двовимірних підходів, тривимірна оцінка пози (3D pose estimation) передбачає реконструкцію просторового положення ключових точок:

$$K^{(3D)} = \{p_i \in \mathbb{R}^3 | p_i = (x_i, y_i, z_i), i = 1,2, \dots, N\}, \quad (10)$$

де координата z_i відображає глибину (відстань від камери або вертикальну компоненту у світовій системі координат) (рис. 3).

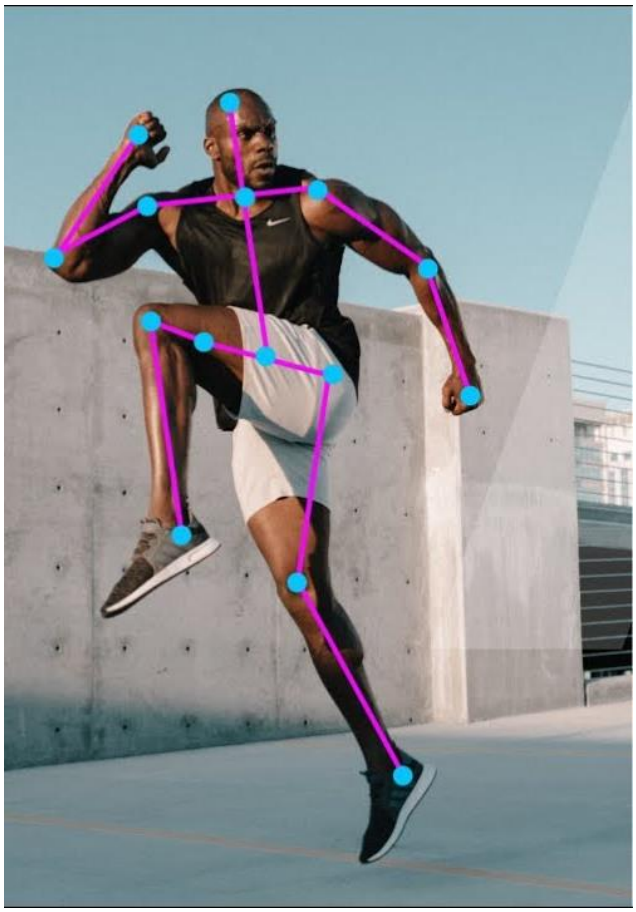


Рис.3. Ілюстрація задачі тривимірної оцінки пози [7]

У системах з однією камерою оцінка z_i є неоднозначною і потребує або навченої моделі для «відновлення» глибини з двовимірних даних, або додаткових сенсорів (наприклад, стерео камер, глибинних камер або мультикамерних систем).

При наявності каліброваної камери (тобто з відомою матрицею внутрішніх параметрів K) можна здійснити проекцію 3D точок у 2D зображення за допомогою проективного перетворення:

$$\lambda \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K \cdot [R|t] \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix}, \quad (11)$$

де (u_i, v_i) – координати точки на зображенні, $R \in \mathbb{R}^{3\times3}$ – матриця обертання, $t \in \mathbb{R}^3$ – вектор трансляції, λ – масштабний коефіцієнт.

Таким чином, задачу можна подати як зворотну до проекції: за відомими 2D точками та параметрами камери – відновити положення в просторі.

В таблиці 1 наведена порівняльна характеристика задач двовимірної та тривимірної оцінки пози людини в інтелектуальних системах відеоспостереження.

Таким чином, у типових інтелектуальних системах відеоспостереження, де головною метою є виявлення, відстеження та базовий аналіз активності людини, використовується 2D оцінка пози через її обчислювальну ефективність та відсутність вимог до спеціалізованого обладнання.

Проте у випадках, коли потрібна більш глибока просторова інтерпретація (наприклад, для виявлення падінь, агресії або складних патернів поведінки), доцільним є використання 3D оцінки пози, за умови наявності відповідних технічних засобів.

Таблиця 1

Порівняльна характеристика задач 2D та 3D оцінки пози людини

Характеристика	2D оцінка	3D оцінка
Вихідні координати	(x, y)	(x, y, z)
Обладнання	Звичайна камера	Глибинна, стерео, багатокамерна система
Обчислювальна складність	Нижча	Вища
Точність положення	В межах зображення	У реальних метричних одиницях
Застосування	Безпека, спостереження, аналітика	Медичні системи, спорт, робототехніка

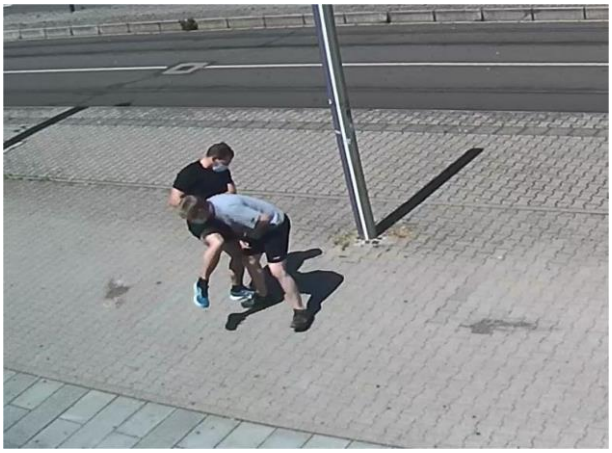
3. Виклики та перешкоди задачі оцінки пози людини

Оцінка пози людини під час відеоспостереження є складним завданням. На відміну від лабораторних умов, камера веде безперервну зйомку вдень і вночі за різних погодних умов протягом усього року. Влітку відбиття від мобільного телефону може частково затуляти камеру, залізничні колії, що заважають людським формам, різки тіні, які можуть призвести до хибно позитивних спрацьовувань, як показано на рисунку 4.

Рис.4. Приклади спотворення камери [8]

[8], освітлення в межах однієї і тієї ж сцени може сильно відрізнятися, як показано на рисунку 5, а [8]. Крім того, через природу відеоспостереження зображення часто сильно спотворені, а камери здебільшого встановлені на висоті та під нахилом, щоб забезпечити велике поле зору. Така перспектива камери підкреслює затінення іншими об'єктами або самим собою, а іноді навіть власною тінню, як показано на рисунках 5, а і 5, б [8].

а) люди, які частково або повністю затінені тінями навколишнього середовища



б) складні, заплутані і закриті пози

Рис.5. Ілюстрація проблем реальних сценаріїв

Аналіз наукових робіт, присвячених вирішенню задачі оцінки пози людини в інтелектуальних системах відеоспостереження, дозволив визначити такі перешкоди та виклики:

– оклюзії – коли частина об'єкта прихована або заблокована іншим об'єктом, це створює значні труднощі для точної оцінки пози. У багатьох реальних сценаріях частини тіла можуть бути закриті, що ускладнює точне прогнозування положення прихованих суглобів. Ця проблема особливо поширена в сценах великого скупчення людей, де кілька людей можуть закривати один одного [9];

– варіації зовнішнього вигляду – оцінювання пози також ускладнюється зміною точки зору. Зовнішній вигляд об'єкта може кардинально змінюватися, якщо дивитися на нього під різними кутами, що може ускладнити точну оцінку пози. Це особливо складно при 3D-оцінці пози, де метою є оцінка 3D-координат суглобів, на які можуть суттєво впливати зміни точки зору [9];

– відсутність анотованих навчальних даних – ефективність багатьох алгоритмів оцінювання поз, особливо тих, що базуються на глибокому навчанні, залежить від наявності великої кількості анотованих навчальних даних. Однак створення таких наборів даних є тривалим і трудомістким процесом. Відсутність достатньої кількості анотованих навчальних даних може суттєво обмежити продуктивність цих алгоритмів [10];

– обробка в реальному часі – оцінка пози в реальному часі, коли позу потрібно оцінювати в реальному часі під час зйомки відео, створює значні проблеми з точки зору обчислювальних ресурсів і швидкості обробки;

– розмиття руху та якість зображення – при оцінці пози на основі відео, розмиття руху може суттєво вплинути на точність оцінки пози. Швидкі рухи можуть призвести до розмиття зображення, що ускладнює точну ідентифікацію положення суглобів. Аналогічно, низькоякісні зображення або відео, які можуть мати шум або низьку роздільну здатність, також можуть створювати проблеми для точної оцінки пози [10];

– варіації точки зору – оцінювання пози також ускладнюється зміною точки зору. Зовнішній вигляд об'єкта може кардинально змінюватися, якщо дивитися на нього під різними кутами, що може ускладнити точну оцінку пози;

– зміна освітлення – зміна освітлення може кардинально вплинути на зовнішній вигляд об'єкта, що ускладнює точну оцінку пози. Тіні можуть приховувати частини тіла, а сильне освітлення може призвести до пере освітлення ділянок, і обидва ці фактори можуть перешкоджати точному визначенню частин тіла та їхнього положення [11].

4. Огляд наукових робіт в області оцінки пози людини

В сучасних інтелектуальних системах відеоспостереження задача оцінки пози людини (Human Pose Estimation, HPE) відіграє ключову роль у розпізнаванні активності, поведінковому аналізі та безпеці.

Наукові дослідження у цій галузі охоплюють як двовимірну, так і тривимірну оцінку пози, із застосуванням класичних методів, глибокого навчання та гібридних рішень.

4.1 Двовимірна оцінка пози

Розміщення ключових точок у двовимірному просторі відносно кадру зображення або відео можна легко оцінити за допомогою двовимірної оцінки пози. Вона працює шляхом виявлення та аналізу координат X та Y суглобів людського тіла на зображенні. 2D оцінка пози – це процес визначення розташування суглобів тіла на зображенні (у піксельних значеннях).

Традиційні методи оцінки пози людини у 2D включають такі підходи, як HOG, Edgelet та модель пікторіальних структур (PSM) [12], які представляють частини тіла у вигляді геометричних фігур (наприклад, циліндрів) та моделюють просторові взаємозв'язки між ними. PSM, запропонована Фішлером [13], використовує частини, визначені за піксельною площею та напрямком, і може ефективно оброблятись за допомогою динамічного програмування

Методи оцінки пози на базі глибинного навчання. Традиційні методи оцінки пози людини в 2D мають обмежену виразність, не враховують глобальний контекст та базуються на ручних ознаках, що знижує точність та ефективність. Через ці обмеження набули популярності моделі на основі глибокого навчання, які забезпечують кращу точність і дозволяють працювати як з однією, так і з кількома людьми на зображенні.

Методи оцінки пози однієї людини (SPPE). Методи оцінки пози однієї людини визначають позу конкретної людини на зображенні. Якщо на зображенні є кілька людей, то зображення обрізається таким чином, щоб на ньому залишилася тільки одна людина. Детектор верхньої частини тіла [14] або детектор всього тіла [15] може виконати це завдання автоматично. Метою методів для однієї людини є визначення місцезнаходження ключових точок у цій області на основі заданої інформації про позицію. Залежно від того, як вони прогнозують ключові точки, методики SPPE поділяються на дві категорії: підходи на основі регресії ключових точок і підходи на основі теплових карт:

– методи на основі регресії ключових точок напряму прогнозують координати кожної точки, але ця задача є складною через необхідність точної відповідності ознак. Для покращення точності використовуються моделі з зворотним зв'язком, наприклад, Iterative Feedback, LCR Network або каскадні регресори [16];

– методи на основі теплових карт прогнозують ймовірність наявності ключової точки в кожному пікселі. Цей підхід дозволяє краще локалізувати частини тіла. Відомі моделі – Convolutional Pose Machines, Stacked Hourglass Network, а також новітні методи SAHR і WAHR, які адаптуються до масштабу і ваг пікселів, значно підвищуючи точність [17]. Також використовуються графові моделі для врахування зв'язків між частинами тіла.

– Пряме регресування координат суглобів є нелінійним і складним для навчання, а також не підходить для випадків з кількома людьми. Водночас теплові карти краще передають просторовий контекст і підходять для складних сценаріїв. Пряме регресування просте, швидке та може застосовуватись у 3D. Теплові карти в поєднанні з великими згортковими ядрами та глибокими моделями покращують точність за рахунок ширшого контексту. У процесі навчання модель поступово пригнічує помилкові відповіді та посилює правильні. Отже, обидва підходи мають свої переваги та недоліки, і універсального рішення не існує.

Методи оцінки пози декількох людей (MPPE). Оскільки положення та об'єм людей на зображенні є невизначеними, оцінка пози кількох людей є складнішою, ніж оцінка пози однієї людини. У більшості випадків цю проблему можна вирішити одним із двох способів:

– найпростіший метод полягає в тому, щоб почати з виявлення людини, потім оцінити її частини і, нарешті, обчислити позу для кожної особи. Він забезпечує високу точність завдяки поетапній обробці. Сучасні моделі, наприклад, Mask R-CNN, GlobalNet і двоетапна модель Papandreou, покращують результати за рахунок багаторівневих ознак і сегментації. Застосування ToF-зображень та виділення ROI підвищує точність для складних областей [18]. Цей підхід називається підходом «зверху-вниз» (top-down);

– інший метод полягає в ідентифікації всіх частин на зображенні (тобто частин кожної людини), а потім групуванні частин, які належать різним людям. Це називається висхідним підходом (bottom-up). Моделі, наприклад, PRN та DeeperCut, покращують точність завдяки глибшим мережам і контекстним зв'язкам [19], а підхід , представлений в роботі [20], поєднує оцінку пози з відстеженням, підвищуючи ефективність.

Підхід «зверху-вниз» забезпечує вищу точність завдяки обробці кожної людини окремо, але він повільніший, оскільки вимагає повторного оцінювання пози для кожного виявленого об'єкта. Висхідний підхід швидший, оскільки обробляє всіх людей одночасно, але часто поступається в точності через нижчу роздільну здатність окремих осіб і складність у виділенні деталей. Глибокі нейронні мережі застосовуються для обох підходів, однак неможливо однозначно сказати, який з них кращий – усе залежить від завдання. Вибір між ними базується на компромісі між точністю та швидкістю.

4.2 Тривимірна оцінка пози

3D оцінка пози (3D HPE) визначає положення суглобів людини у тривимірному просторі (X, Y, Z) за зображенням або відео. Основна мета – обчислити координати ключових точок тіла на

основі RGB-зображення. На відміну від 2D, 3D оцінка є складнішою через неоднозначність, вищі вимоги до обчислювальних ресурсів і вплив зовнішніх факторів (текстура, колір шкіри, тло, оклюзії тощо). Оцінювання також ускладнюється вибором якісного датасету (набору даних). Після визначення суглобів моделі аналізують рух людини у серії кадрів.

Класичні підходи до 3D оцінки пози включають розширення пікторіальних структурних моделей (PSM), які застосовуються як для однієї, так і для кількох осіб [21]. Для покращення точності використовуються структуровані SVM, що навчаються на відповідності між сегментаційними ознаками та положенням суглобів [22]. Інший підхід – використання HOG-ознак і лінійної регресії для оцінки 3D пози, з подальшим зменшенням розмірності за допомогою PCA [23]. Цей метод показав високі результати, здобувши перше місце у COCO 2016 keypoints challenge та перевершивши попередній рекорд на датасеті MPII.

Монокулярна 3D оцінка пози людини є складним завданням через неоднозначність глибини та оклюзії. Попри це, монокулярна камера є найбільш поширеним засобом для HPE. Методи на основі глибокого навчання поділяються на однокамерні (single-view) та багатокамерні (multi-view) [24]. Single-view підходи працюють із зображеннями з однієї камери, наприклад EpipolarPose використовує лише 2D ключові точки та епіполярну геометрію для побудови 3D без необхідності у 3D анотаціях. Багатокамерні методи покращують оцінку глибини та точність, навіть із несинхронізованими відео потоками. Найпоширеніший підхід – спершу визначати 2D ключові точки, а потім трансформувати їх у 3D. Такі моделі, як PostNet, HRNet, Mask R-CNN та Cascaded Pyramid Network, забезпечують хорошу точність і швидкість роботи в реальному часі [25].

Оцінка 3D пози однієї людини. Більшість робіт для оцінки пози людини використовують одне зображення/відео. Незважаючи на неоднозначність виміру глибини, моделі, навчені на 3D-еталонних даних, показують досить хороші результати для випадку однієї людини без оклюзій.

Оцінка 3D-позиції кількох людей. Основною проблемою при оцінці 3D-позиції декількох осіб є оклюзії. Через обмежену кількість відповідних наборів даних прогрес в оцінці 3D-позиції декількох осіб є обмеженим. Крім того, на жаль, майже не існує анотованих наборів даних 3D-позицій кількох людей, подібних до набору даних Human3.6 [26]. Більшість наборів даних з кількома особами або не мають хороших еталонних даних, або не є реалістичними. Однією з відомих робіт в даній області є робота [27], в якій для оцінки 3D-позицій кількох людей використовується PandaNet (Pose Estimation and Detection Anchor-based Network).

5. Тенденції та перспективи задачі оцінки пози людини

Тенденції задачі оцінки пози людини лежать в площині підходів глибокого навчання, оскільки можна помітити, що ці методи глибокого навчання досягають кращої продуктивності порівняно з іншими сучасними підходами. Успіх підходів глибокого навчання до задачі HPE полягає в доступності величезної кількості даних, що є одним з обмежень застосування глибинного навчання. Незважаючи на те, що для об'єктивної оцінки HPE були створені різні бази даних, додаткові набори даних з адекватними методологіями обстеження все ще бажані. У довгостроковій перспективі можна використовувати додаткові датчики тіла для запису необроблених даних з різних поз [28].

У той час як задача двовимірної оцінки пози людини досягла достатнього рівня точності, тривимірна оцінка потребує багатьох зусиль, якщо не буде розроблено більш збалансованих моделей, особливо для інтерпретації з одного зображення і без деталей глибини.

Перспективи задачі оцінки пози людини є величезними, оскільки задача має велику сферу застосування, яка важлива в нашому повсякденному житті. Існує також можливість досягти хороших результатів на наборах даних вищої розмірності (вище 2D/3D), таких як оцінка 6D-поз, яка оцінює положення і напрямок 6D-поз. Ці пози корисні в робототехнічних додатках. Незважаючи на те, що ідентифікації людських поз з відео або фотографій присвячено багато зусиль, існує значний розрив між теоретичними дослідженнями і реальними застосуваннями [29].

Висновки

В статті представлено систематизований огляд сучасних методів оцінки пози людини в інтелектуальних системах відеоспостереження. Розглянуто двовимірні (2D) та тривимірні (3D) підходи, зокрема як класичні алгоритми на основі ручних ознак, так і сучасні глибинні моделі. 2D оцінка залишається домінуючою у відеоспостереженні завдяки швидкодії та простоті реалізації, тоді як 3D оцінка забезпечує глибшу просторову інтерпретацію, але вимагає складніших моделей та апаратних ресурсів.

Проаналізовано top-down і bottom-up підходи до оцінки пози кількох осіб. Top-down забез-

печує вищу точність, однак є повільнішим і менш масштабованим. Bottom-up дозволяє працювати з великою кількістю об'єктів одночасно, хоча поступається у точності. Методи на основі глибокого навчання, особливо ті, що використовують теплові карти й графові структури, демонструють високу ефективність, однак залежать від наявності якісних і репрезентативних навчальних даних.

Основні виклики у практичному застосуванні HPE пов'язані з перекриттям об'єктів, зміною освітлення, тінями, шумом, низькою якістю відео та вимогами до реального часу. Подолання цих проблем потребує стійких, адаптивних моделей та використання додаткових сенсорів. Перспективи подальших досліджень охоплюють багатомодальні підходи, 6D оцінку пози та розвиток самонавчальних систем, здатних адаптуватися до змінних умов реального середовища. Таким чином, оцінка пози людини залишається ключовим напрямом у розвитку інтелектуального відеоаналізу.

Список використаної літератури

References

1. C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep Learning-based Human Pose Estimation: A Survey," *ACM Computing Surveys*, vol. 56, no. 1, Art. no. 11, pp. 1–37, Jan. 2024. DOI: 10.1145/3603618.

2. S. Dubey and M. Dixit, "A comprehensive survey on human pose estimation approaches," *Multimedia Systems*, vol. 29, no. 1, pp. 167–195, Feb. 2023. DOI: 10.1007/s00530-022-00980-0.

3. Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, and J. Liu, "Human Action Recognition From Various Data Modalities: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3200–3225, Mar. 2023. DOI: 10.1109/TPAMI.2022.3183112.

4. M. Cormier, A. Clepe, A. Specker, and J. Beyerer, "Where are we with Human Pose Estimation in Real-World Surveillance?," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Waikoloa, HI, USA, 2022, pp. 591–601. DOI: 10.1109/WACVW54805.2022.00065

5. M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," in Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 2014, pp. 3686–3693. DOI: 10.1109/CVPR.2014.471.

6. R. Choudhury, K. M. Kitani, and L. A. Jeni, "TEMPO: Efficient Multi-View Pose Estimation, Tracking, and Forecasting," in *Proc. of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, 2023, pp. 14704–14714.

7. J. Rajasegaran, G. Pavlakos, A. Kanazawa, C. Feichtenhofer, and J. Malik, "On the Benefits of 3D Pose and Tracking for Human Action Recognition," *arXiv preprint arXiv:2304.01199*, 2023. [Online]. Available: https://arxiv.org/abs/2304.01199

8. T. L. Munea, Y. Z. Jembre, H. T. Weldegebriel, L. Chen, C. Huang, and C. Yang, "The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation," *IEEE Access*, vol. 8, pp. 133330–133348, 2020. DOI: 10.1109/ACCESS.2020.3010248

9. C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep Learning-Based Human Pose Estimation: A Survey," *arXiv preprint arXiv:2012.13392*, 2023. [Online]. Available: https://arxiv.org/abs/2012.13392

10. R. W. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, Jun. 2010. DOI: 10.1016/j.imavis.2009.11.014

11. M. Wang, J. Tighe, and D. Modolo, "Combining detection and tracking for human pose estimation in videos," *arXiv preprint arXiv:2003.13743*, 2020. [Online]. Available: https://arxiv.org/abs/2003.13743

12. P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial Structures for Object Recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, Jan. 2005. DOI: 10.1023/B:VISI.0000042934.15159.49

13. M. A. Fischler and R. A. Elschlager, "The Representation and Matching of Pictorial Structures," *IEEE Transactions on Computers*, vol. C-22, no. 1, pp. 67–92, Jan. 1973. DOI: 10.1109/T-C.1973.223602

14. A. S. Micilotta, E.-J. Ong, and R. Bowden, "Real-time upper body detection and 3D pose estimation in monoscopic images," in *Proc. of the 9th European Conference on Computer Vision (ECCV)*, Graz, Austria, 2006, vol. Part III, pp. 139–150. DOI: 10.1007/11744078_11

15. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017. DOI: 10.1109/TPAMI.2016.2577031

16. G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy, "Towards Accurate Multi-person Pose Estimation in the Wild," *arXiv preprint arXiv:1701.01779*, 2017. [Online]. Available: https://arxiv.org/abs/1701.01779

17. M. Ben Gamra and M. A. Akhloufi, "A review of deep learning techniques for 2D and 3D human pose estimation," *Image and Vision Computing*, vol. 114, Art. no. 104282, 2021. DOI: 10.1016/j.imavis.2021.104282

18. N. Rodrigues, H. Torres, B. Oliveira, J. Borges, S. Queirós, J. H. Mendes, J. Fonseca, V. Coelho, and J. H. Brito, "Top-Down Human Pose Estimation with Depth Images and Domain Adaptation," in *Proc. of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019), Vol. 5: VISAPP*, Prague, Czech Republic, 2019, pp. 281–288. DOI: 10.5220/0007344602810288

19. Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded Pyramid Network for Multi-person Pose Estimation," in *Proc. of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 7103–7112. DOI: 10.1109/CVPR.2018.00742

20. E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer, 2016, pp. 34–50. DOI: 10.1007/978-3-319-46466-4_3

21. A. Alzughaibi and Z. Chaczko, "Human detection model using feature extraction method in video frames," in *Proc. of the 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Christchurch, New Zealand, 2016, pp. 1–6. DOI: 10.1109/IVCNZ.2016.7804424

22. H. Kim, S. Lee, D. Lee, S. Choi, J. Ju, and H. Myung, "Real-Time Human Pose Estimation and Gesture Recognition from Depth Images Using Superpixels and SVM Classifier," *Sensors*, vol. 15, no. 6, pp. 12410–12427, 2015. DOI: 10.3390/s150612410

23. K. Chen, S. Gong, and T. Xiang, "Human pose estimation using structural support vector machines," in *Proc. of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, 2011, pp. 846–851. DOI: 10.1109/ICCVW.2011.6130340

24. A. Zakhor and A. Hallquist, "Single view pose estimation of mobile devices in urban environments," in *Proc. of the 2013 IEEE Workshop on Applications of Computer Vision (WACV)*, Clearwater Beach, FL, USA, 2013, pp. 347–354. DOI: 10.1109/WACV.2013.6475039

25. Y. Deng, C. Sun, J. Zhu, and Y. Sun, "SVMAC: Unsupervised 3D Human Pose Estimation from a Single Image with Single-view-multi-angle Consistency," *arXiv preprint arXiv:2106.05616*, 2022. [Online]. Available: https://arxiv.org/abs/2106.05616

26. C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014. DOI: 10.1109/TPAMI.2013.248

27. A. Benzine, F. Chabot, B. Luvison, Q. C. Pham, and C. Achard, "PandaNet: Anchor-Based Single-Shot Multi-Person 3D Pose Estimation," *arXiv preprint arXiv:2101.02471*, 2021. [Online]. Available: https://arxiv.org/abs/2101.02471

28. K. Khan, W. Albattah, R. U. Khan, A. M. Qamar, and D. Nayab, "Advances and Trends in Real Time Visual Crowd Analysis," *Sensors*, vol. 20, no. 18, Art. no. 5073, 2020. DOI: 10.3390/s20185073

29. S. Chang, L. Yuan, X. Nie, Z. Huang, Y. Zhou, Y. Chen, J. Feng, and S. Yan, "Towards Accurate Human Pose Estimation in Videos of Crowded Scenes," in *Proc. of the 28th ACM International Conference on Multimedia (MM '20)*, Seattle, WA, USA, 2020, pp. 4630–4634. DOI: 10.1145/3394171.3416299

Human posture estimation in intelligent video surveillance systems: modern approaches and challenges

R. Kovalevych, D. Kurinko, V. Kryvda
Odesa Polytechnic National University

Abstract. *The article presents an overview of modern methods for estimating human pose in intelligent video surveillance systems. 2D and 3D approaches are considered, including both classical methods and models based on deep learning. The features of top-down and bottom-up strategies, their advantages and*

limitations are analyzed. The main challenges associated with real-world surveillance conditions, such as occlusions, lighting changes, and limited image quality, are identified. Prospects for the development of the industry, including the use of 6D pose estimation, multisensory data, and self-learning models, are outlined.

Keywords: *Intelligent systems, Video surveillance, Behavioral analysis, Pose estimation, Neural networks, Heatmaps, Multi-camera systems, Occlusions*.

Отримано: 20.02.2025

Про авторів

Ковалевич Роман Валерійович, аспірант кафедри штучного інтелекту та аналізу даних, Національний університет «Одеська Політехніка»; проспект Шевченка, 1, Одеса, 65044, Україна. E-mail: 8766639@as.op.edu.ua

Roman V. Kovalevych, postgraduate student of the Artificial Intelligence and Data Analysis Department, Odesa Polytechnic National University; 1, Shevchenko Avenue, Odesa, 65044, Ukraine. E-mail: 8766639@as.op.edu.ua

ORCID: https://orcid.org/0009-0008-9645-4352

Курінько Дмитро Дмитрович, аспірант кафедри штучного інтелекту та аналізу даних, Національний університет «Одеська Політехніка»; проспект Шевченка, 1. Одеса, 65044, Україна. E-mail: dmitrykurinko@gmail.com

Dmytro D. Kurinko, postgraduate student of the Artificial Intelligence and Data Analysis Department, Odesa Polytechnic National University; 1, Shevchenko Avenue, Odesa, 65044, Ukraine. E-mail: dmitrykurinko@gmail.com

ORCID: https://orcid.org/0000-0001-8304-3257

Кривда Вікторія Ігорівна, к. т. н., доцент, доцент кафедри електропостачання та енергетичного менеджменту, завідувач відділу аспірантури і докторантури, Національний університет «Одеська Політехніка»; проспект Шевченка, 1, Одеса, 65044, Україна. E-mail: kryvda@op.edu.ua; тел.: +38 066 930 0875

Viktoriia I. Kryvda, Ph.D., Associate Professor, Associate Professor of the Department of Power Supply and Energy Management, Head of the Department of Postgraduate and Doctoral Studies, Odesa Polytechnic National University; 1, Shevchenko Avenue, Odesa, 65044, Ukraine. E-mail: kryvda@op.edu.ua; ph: +38 066 930 0875

ORCID: https://orcid.org/0000-0002-0930-1163