

<https://doi.org/10.31891/2307-5732-2026-365-38>

УДК 004.93

ВОЛІВАЧ АНТОНІНА

Київський національний університет технологій та дизайну

<https://orcid.org/0000-0002-7119-7774>

e-mail: volivach.ap@knutd.edu.ua

ЛЕБЕДЕНКО ЮРІЙ

Київський національний університет технологій та дизайну

<https://orcid.org/0000-0002-1352-9240>

e-mail: lebedenko.yo@knutd.edu.ua

СЕЛЯКОВ ЄВГЕН

Київський національний університет технологій та дизайну

<https://orcid.org/0009-0007-6480-8323>

e-mail: seliakov.yo@knutd.edu.ua

УСІКОВ МИКИТА

Київський національний університет технологій та дизайну

<https://orcid.org/0009-0009-8023-8396>

e-mail: usikov.my@knutd.edu.ua

АНАЛІЗ ПІДХОДІВ ДО ВИЗНАЧЕННЯ ПРОСТОРОВОГО ПОЛОЖЕННЯ ОБ'ЄКТІВ ЗАСОБАМИ КОМП'ЮТЕРНОГО ЗОРУ

У роботі проведено аналіз та систематизацію підходів до визначення просторового положення об'єктів засобами комп'ютерного зору, що є важливою складовою сучасних інтелектуальних систем обробки візуальної інформації. Актуальність дослідження зумовлена широким застосуванням комп'ютерного зору в таких сферах як індустрія моди, дизайн інтер'єру, робототехніка, автономні транспортні системи, доповнена та віртуальна реальність, де точне визначення розташування об'єктів є критично необхідним.

Проведено аналіз підходів за двома класифікаційними ознаками: типом вхідних даних та методом обробки інформації. Показано, що монокулярні методи характеризуються простотою реалізації та мінімальними вимогами до обладнання, однак обмежені відсутністю безпосередньої інформації про глибину сцени. Стереоскопічні підходи та методи з використанням глибинних сенсорів забезпечують вищу точність завдяки додатковій просторовій інформації, проте потребують складнішої апаратної реалізації.

Розглянуто підходи обробки даних: геометричні, що базуються на математичних моделях камери та просторових перетвореннях, методи на основі ознак, які використовують характерні елементи зображення, а також нейромережеві методи глибокого навчання, що забезпечують високу ефективність у складних умовах і здатні враховувати нелінійні залежності даних.

Встановлено, що гібридні підходи, які поєднують класичні та нейромережеві методи, дозволяють підвищити точність і стійкість при прийнятній обчислювальній складності.

У результаті дослідження обґрунтовано, що найбільш перспективним напрямом є інтеграція різних типів даних і методів обробки та розвиток гібридних моделей для роботи в умовах невизначеності та в режимі реального часу.

Ключові слова: комп'ютерний зір, просторове положення об'єктів, геометричні методи, методи на основі ознак, нейромережеві методи, гібридні методи.

VOLIVACH ANTONINA, YURI LEBEDENKO, SELIAKOV YEVHEN, USIKOV MYKYTA

Kyiv National University of Technologies and Design

ANALYSIS OF APPROACHES TO DETERMINING THE SPATIAL POSITION OF OBJECTS USING COMPUTER VISION

This paper analyzes and systematizes approaches to determining the spatial position of objects using computer vision, which is an important component of modern intelligent visual information processing systems. The relevance of this research stems from the widespread use of computer vision in fields such as the fashion industry, interior design, robotics, autonomous transportation systems, and augmented and virtual reality, where the precise localization of objects is critically important.

An analysis of approaches was conducted based on two classification criteria: the type of input data and the method of information processing. It is shown that monocular methods are characterized by simplicity of implementation and minimal hardware requirements, but are limited by the lack of direct information about scene depth. Stereoscopic approaches and methods using depth sensors provide higher accuracy due to additional spatial information, but require more complex hardware implementation.

Data processing approaches are considered: geometric methods based on mathematical camera models and spatial transformations; feature-based methods that utilize characteristic image elements; and deep learning neural network methods that ensure high performance under complex conditions and are capable of accounting for nonlinear data dependencies.

It has been established that hybrid approaches, which combine classical and neural network methods, allow for increased accuracy and robustness with acceptable computational complexity.

The study concludes that the most promising direction is the integration of different types of data and processing methods, as well as the development of hybrid models for operation under conditions of uncertainty and in real-time.

Keywords: computer vision, spatial positioning of objects, geometric methods, feature-based methods, neural network methods, hybrid methods

Стаття надійшла до редакції / Received 11.02.2026

Прийнята до друку / Accepted 11.03.2026

Опубліковано / Published 28.05.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© Волівач Антоніна, Лебеденко Юрій, Селяков Євген, Усіков Микита

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

На сьогодні сучасні інформаційні технології активно інтегруються в різні сфери життєдіяльності, що значною мірою зумовлено стрімким розвитком методів комп'ютерного зору. Одним із ключових завдань цього напрямку є визначення просторового положення об'єктів, яке передбачає встановлення їх координат та орієнтації у двовимірному та тривимірному просторі на основі візуальних даних. Такі задачі мають широке практичне застосування, зокрема в індустрії моди, дизайні інтер'єру, робототехніці, автономних транспортних системах, доповненій та віртуальній реальності, а також у системах автоматизації виробничих процесів та інтелектуального проєктування.

Для розв'язання задачі оцінювання просторового положення об'єктів застосовуються різні підходи, зокрема методи обробки стереозображень, аналіз даних сенсорів та алгоритми глибокого навчання, що забезпечують інтеграцію локальної та глобальної інформації. Проте ефективність цих підходів у реальних умовах суттєво знижується під впливом різних факторів, серед яких змінні умови освітлення, наявність шумів у зображеннях, часткове перекриття об'єктів, а також обмеженість або відсутність інформації про глибину сцени при використанні монокулярних камер. Крім того, додаткові труднощі виникають у випадку складної геометрії об'єктів та варіативності ракурсів їх спостереження.

У науковій літературі представлено значну кількість підходів до розв'язання цієї задачі, які відрізняються як за типами вхідних даних, так і за методами їх обробки, що ускладнює обґрунтований вибір ефективних рішень для практичного застосування. Таким чином є актуальною задача систематизації існуючих підходів до визначення просторового положення об'єктів засобами комп'ютерного зору та проведення їх порівняльного аналізу з метою визначення найбільш ефективних рішень для конкретних умов застосування.

Аналіз досліджень та публікацій

Дослідження задачі визначення просторового положення об'єктів пройшло значну еволюцію – від класичних геометричних методів, що базуються на принципах аналітичної геометрії, до сучасних підходів, які використовують методи машинного навчання, зокрема штучні нейронні мережі. При цьому, на ранніх етапах основна увага приділялася розробці точних математичних моделей, методів калібрування камер та визначенню координат об'єктів у просторі на основі відомих параметрів і геометричних залежностей.

Подальший розвиток обчислювальних потужностей і поява великих обсягів даних сприяли активному впровадженню методів комп'ютерного зору. Такий підхід дозволив автоматизувати процес визначення просторового положення об'єктів і підвищити ефективність обробки візуальної інформації.

У сучасних дослідженнях застосування комп'ютерного зору основний акцент зміщено в бік методів глибокого навчання. Зокрема, у роботі [1] представлено комплексний огляд сучасних підходів до визначення просторового положення об'єктів із використанням нейронних мереж, що працюють з різними типами вхідних даних (RGB, RGB-D, point cloud) та дозволяють оцінювати положення об'єктів з різною кількістю ступенів свободи. Систематизовано підходи за рівнем узагальнення (instance-level, category-level, unseen object), що дозволяє класифікувати методи за їх здатністю переносити знання на нові об'єкти. Показано, що методи глибокого навчання забезпечують ефективне моделювання складних нелінійних залежностей і підвищують точність визначення просторового положення об'єктів навіть у складних умовах (зміни освітлення, шуми, часткове перекриття тощо). Водночас такі методи мають низку обмежень, зокрема високу залежність від великих обсягів навчальних даних, значні обчислювальні витрати та потребу в потужних апаратних ресурсах, що ускладнює їх застосування в режимі реального часу та вбудованих системах.

Більшість науковців приділяють значну увагу монокулярним підходам, які дозволяють визначити положення об'єктів за допомогою одного зображення. Так, у роботі [2] запропоновано методи оцінювання положення людини на основі даних з однієї камери як у двовимірному, так і в тривимірному просторі. Наведено класифікацію підходів за типом моделей, способом обробки зображень та формулюванням задачі. Показано, що використання теплових карт забезпечує вищу точність порівняно з прямою регресією координат. Проаналізовано архітектури нейронних мереж і підходи до переходу від 2D до 3D-представлення, що дозволяє частково відновлювати інформацію про глибину сцени навіть за наявності одного зображення. Встановлено, що основними обмеженнями таких методів є складність конфігурацій об'єктів, часткове перекриття та варіативність умов спостереження.

У той же час значна увага приділяється задачам геометрії в комп'ютерному зорі, зокрема оцінюванню просторового положення об'єктів у 2D та 3D за допомогою каліброваних та некаліброваних камер. У роботі [3] авторами розглянуто перехід від класичних каліброваних підходів до сучасних некаліброваних алгоритмів, які дозволяють визначити просторове положення точок і камер у тривимірному просторі за мінімального обсягу вхідних даних. Систематизовано розвиток методів мультивидової геометрії та описано задачі від встановлення відповідностей між зображеннями до відновлення епіпольної, трифокальної та чотирифокусної геометрії, а також проєктної реконструкції на основі множини ракурсів.

Особлива увага приділяється сучасним методам оптимізації та робастним алгоритмам, зокрема RANSAC, що підвищують точність оцінювання просторового положення об'єктів у складних умовах та забезпечують стійкість до шумів і викидів. Наголошується, що некалібровані методи дозволяють будувати 3D-моделі в реальному часі навіть за відсутності параметрів калібрування камер, зокрема при роботі з архівними відеоданими або рухомими камерами. При цьому підкреслюються практичні труднощі обробки реальних даних,

серед яких наявність шумів, неконсистентні відповідності та складна геометрія сцен, що потребує застосування спеціалізованих алгоритмічних рішень. Такий підхід спрощує обчислення та підвищує гнучкість і ефективність методів відновлення просторової структури сцени.

У роботі [4] представлено огляд сучасних підходів до обробки зображень, включаючи класичні методи на основі дескрипторів ознак, алгоритми нечіткої логіки та моделі глибокого навчання. Встановлено, що класичні методи залишаються ефективними для задач порівняння зображень і навігації за умов обмежених обсягів даних, тоді як нейромережеві моделі забезпечують вищу точність у задачах класифікації та виявлення об'єктів. Підкреслено, що жоден із підходів не є універсальним, а ефективність їх застосування залежить від вимог до точності, швидкодії та доступних обчислювальних ресурсів.

Практичні аспекти застосування методів комп'ютерного зору розглянуто у роботі [5], де досліджується задача визначення стану водія на основі даних з відеокамер. Проаналізовано підходи до класифікації зображень, обробки відеопотоку та визначення ключових точок тіла. Показано, що використання згорткових нейронних мереж забезпечує високу точність розпізнавання, однак не враховує часову динаміку процесів. Для обробки відеоданих запропоновано застосування 3D-конволюційних та рекурентних моделей, що дозволяє враховувати зміни поведінки у часі. Також досліджено підхід на основі оцінювання положення тіла та орієнтації голови, що підвищує інтерпретованість результатів. Встановлено, що комбінування різних підходів дозволяє підвищити ефективність систем аналізу поведінки.

Таким чином, проведений аналіз наукових досліджень та публікацій показує, що існує широкий спектр підходів до визначення просторового положення об'єктів – від класичних геометричних методів до сучасних нейромережевих моделей. Кожен із підходів має свої переваги та обмеження, що зумовлює необхідність їх подальшої систематизації, комбінування та адаптації до конкретних прикладних задач.

Формулювання цілей статті

Метою роботи є аналіз, систематизація та узагальнення сучасних підходів до визначення просторового положення об'єктів засобами комп'ютерного зору. Для досягнення поставленої мети в роботі передбачено вирішення таких завдань: розглянути сутність задачі визначення просторового положення об'єктів у контексті систем комп'ютерного зору; здійснити класифікацію підходів за типом вхідних даних і методами їх обробки; провести порівняльний аналіз.

Виклад основного матеріалу

Просторове положення об'єкта визначається його координатами та орієнтацією відносно обраної системи координат. У задачах комп'ютерного зору воно полягає у визначенні положення об'єкта в 2D- або 3D-просторі на основі аналізу візуальної інформації.

Одним із сучасних підходів до оцінювання просторового положення об'єктів на основі одного зображення є метод PoseNet. Він базується на використанні глибоких згорткових нейронних мереж і дозволяє визначати положення та орієнтацію об'єкта у просторі з шістьма ступенями свободи (6DoF), зокрема положення камери. Метод є ефективним для задач візуальної локалізації в реальному часі, у тому числі за умов змінного освітлення, наявності шумів та часткових перекриттів. У межах підходу вдосконалено функцію втрат для одночасної регресії положення та орієнтації з урахуванням геометрії сцени, а також запропоновано механізм автоматичного узгодження вагових коефіцієнтів, що зменшує залежність від ручного налаштування гіперпараметрів і підвищує точність локалізації порівняно з традиційними методами на основі локальних ознак [6].

Варто зазначити, що перехід від 2D до 3D-представлення є складною задачею, оскільки зображення не містить безпосередньої інформації про глибину сцени. Унаслідок цього одна й та сама двовимірна проекція може відповідати різним конфігураціям просторового розташування об'єктів, що зумовлює неоднозначність відновлення геометрії сцени [7]. При цьому система аналізує лише проекцію сцени без прямого вимірювання відстаней, що потребує використання додаткових припущень і може знижувати точність результатів.

Отже, ключовою проблемою визначення просторового положення об'єктів засобами комп'ютерного зору є неповнота або відсутність інформації про глибину сцени. Це обумовлює доцільність проведення класифікації існуючих підходів за типом вхідних даних, у межах якої виділяють три основні групи: монокулярні, стереоскопічні та методи з використанням глибинних сенсорів.

Монокулярні методи базуються на використанні одного зображення або послідовності зображень як вхідних даних. Їхньою перевагою є простота реалізації та відсутність потреби у спеціалізованому обладнанні, що має широке практичне застосування. Водночас основним обмеженням є відсутність прямої інформації про глибину сцени, що суттєво ускладнює відновлення просторової структури [8].

Оцінювання глибини за одним зображенням є принципово неоднозначною задачею, оскільки одна і та ж сама 2D-проекція може відповідати різним конфігураціям сцени у тривимірному просторі. Для її розв'язання застосовуються глибокі нейронні мережі, які поєднують глобальний аналіз зображення з локальним уточненням деталей. Такий підхід дозволяє формувати карту глибини на основі непрямих візуальних ознак (перспективи, текстури, відносних розмірів), зменшуючи невизначеність масштабу та підвищуючи точність реконструкції сцени [8].

Одним із практичних напрямів застосування монокулярних методів є оцінювання положення людини. У цій задачі підходи реалізуються шляхом побудови нейромережевих моделей, що відновлюють 2D- або 3D-позу з одного зображення або відеопослідовності. Такі моделі поєднують аналіз локальних ознак (контури, текстури) та глобальної структури сцени, що забезпечує високу точність розпізнавання. Водночас ці методи залишаються чутливими до умов зйомки, зокрема перекриттів, зміни ракурсу та варіативності поз [2].

До компенсації відсутності прямої інформації про глибину монокулярні методи активно використовують нейромережеві моделі, які оцінюють просторову структуру сцени на основі непрямих візуальних ознак, таких як перспектива, текстура та відносний розмір об'єктів, формуючи карту глибини. Водночас такі підходи обмежені внутрішньою неоднозначністю задач та значною залежністю від якості й обсягу навчальних даних.

Перехід до стереоскопічних підходів дозволяє отримати додаткову інформацію про глибину сцени за рахунок використання двох або більше камер. Оцінювання просторового положення об'єктів у таких системах здійснюється шляхом аналізу відмінностей між зображеннями, отриманих з різних ракурсів. Основою методу є пошук відповідних точок та обчислення їх зміщення, яке безпосередньо пов'язане з відстанню до об'єкта.

Геометричні принципи двовидової реконструкції описуються в рамках епіпольної геометрії, яка визначає взаємозв'язок між відповідними точками на зображеннях. Цей зв'язок формалізується за допомогою фундаментальної матриці, яка задає обмеження на положення відповідних точок і дозволяє суттєво звужити область їх пошуку. Використання таких обмежень підвищує точність встановлення відповідностей та оцінювання глибини сцени [3].

Для відновлення просторової структури застосовується метод триангуляції, при якому координати точки визначаються як перетин променів, проведених від центрів камер через відповідні проєкції на зображеннях. Це дозволяє не лише відновлювати положення окремих об'єктів, але й оцінювати параметри камер та геометрію сцени.

Перевагою стереоскопічних методів є вища точність визначення просторового положення порівняно з монокулярними підходами завдяки використанню безпосередньої інформації про глибину. Водночас їх застосування вимагає калібрування камер, синхронізації та характеризується більшими обчислювальними витратами.

Такі підходи широко застосовуються в робототехніці, автономних транспортних системах і системах технічного зору, де стереопари камер використовуються для побудови карти глибини та аналізу навколишнього середовища.

Методи з використанням глибинних сенсорів базуються на застосуванні спеціалізованих пристроїв (зокрема RGB-D камер), які безпосередньо вимірюють відстань до об'єктів сцени. У результаті формується глибинна карта, де кожному пікселю відповідає значення відстані, що забезпечує явне представлення тривимірної структури сцени без необхідності її відновлення з непрямих ознак.

Ключовими характеристиками глибинних сенсорів є діапазон і точність вимірювання глибини, просторова роздільна здатність, частота кадрів та стійкість до зовнішніх умов. Також важливим є фізичний принцип роботи (наприклад, структуроване світло або time-of-flight), який впливає на точність та стабільність вимірювань у різних умовах експлуатації.

Використання глибинної інформації (RGB-D, point cloud) дозволяє суттєво зменшити неоднозначність задачі визначення просторового положення об'єктів та підвищити точність алгоритмів за рахунок наявності безпосередніх метричних даних про сцену [1]. Це спрощує побудову тривимірних моделей та підвищує якість оцінювання геометрії простору.

Практична реалізація таких підходів передбачає інтеграцію глибинних даних у єдину тривимірну модель середовища. Зокрема, застосування алгоритмів типу KinectFusion дозволяє об'єднувати послідовність глибинних кадрів у режимі реального часу, одночасно виконуючи оцінювання положення камери та зменшуючи накопичення похибки під час руху [7].

У задачах визначення положення людини використання глибинних сенсорів забезпечує перехід від двовимірного до тривимірного представлення пози. Застосування методів, що працюють із 3D-орієнтирами (ключовими точками тіла), дозволяє враховувати просторову конфігурацію об'єкта та підвищує стійкість до змін ракурсу і часткових перекриттів [9].

Таким чином, методи з використанням глибинних сенсорів забезпечують високу точність, інформативність та стійкість результатів за рахунок прямого вимірювання відстаней до об'єктів. Водночас їх застосування обмежується вартістю обладнання, залежністю від фізичних характеристик сенсорів та меншою універсальністю порівняно з монокулярними підходами.

Результати порівняльного аналізу підходів до визначення просторового положення об'єктів за типом вхідних даних наведено в таблиці 1.

Як видно з таблиці 1, кожен із розглянутих підходів до визначення просторового положення об'єктів характеризується власними перевагами та недоліками, що визначають доцільність його застосування в конкретних умовах. При цьому монокулярні методи характеризуються простотою реалізації та універсальністю, однак обмежені відсутністю прямої інформації про глибину сцени, що знижує точність відновлення просторової структури.

Стереоскопічні підходи забезпечують підвищену точність оцінювання за рахунок використання геометричних зв'язків між зображеннями з різних ракурсів, проте потребують складнішої апаратної реалізації, попереднього калібрування камер та більшої обчислювальної складності.

Методи з використанням глибинних сенсорів забезпечують найвищу інформативність вхідних даних завдяки безпосередньому вимірюванню відстаней до об'єктів сцени, проте їх застосування обмежується вартістю обладнання та залежності від умов експлуатації.

Таким чином, вибір підходу за типом вхідних даних визначається компромісом між точністю визначення просторового положення об'єктів, складністю реалізації та вимогами до апаратних ресурсів. Загальна тенденція полягає в тому, що зі збільшенням обсягу та інформативності вхідних даних підвищується точність оцінювання, однак одночасно зростає складність системи.

Таблиця 1

Аналіз підходів до визначення просторового положення об'єктів за типом вхідних даних

Підхід	Тип вхідних даних та принцип роботи	Основні переваги	Основні недоліки	Обчислювальна складність	Типові сфери застосування
Монокулярний	Використовується одне зображення або відеопослідовність; оцінювання просторового положення здійснюється на основі візуальних ознак (перспектива, масштаб, тек-стура) або за допомогою нейромереж, що відно-влюють глибину сцени	Простота реалізації; відсутність потреби у додатковому обладнанні; універсальність	Відсутність прямої інформації про глибину; неоднозначність відновлення просторової структури; знижена точність у складних умовах	Низька – середня (залежно від моделі)	Відеоспос-тереження, мобільні додатки, базові системи комп'ютерного зору
Стереоскопічні	Використовується два або більше зображень з різних камер; глибина визначається шляхом оцінювання диспаратності між відповідними точками	Вища точність оцінювання глибини; безпосереднє використання геометричних обмежень; обґрунтованість моделі	Необхідність калібрування камер; залежність від якості встановлення відповідностей; підвищена складність реалізації	Середня – висока	Робототехніка, автономні транспортні системи, 3D-реконструкція
Використання глибинних сенсорів (RGB-D)	Використовується зображення разом із безпосередньо вимірюваною картою глибини, отриманою сенсором для кожного пікселя	Висока точність; пряме вимірювання глибини; стабільність результатів	Вартість обладнання; обмеження у складних умовах освітлення та дальності	Середня	Доповнена реальність, 3D-сканування, системи взаємодії людина-комп'ютер

Іншим важливим критерієм класифікації підходів до визначення просторового положення об'єктів є спосіб обробки вхідних даних. На відміну від класифікації за типом даних, у даному випадку основна увага приділяється саме алгоритмічним принципам, на основі яких виконується оцінювання положення об'єкта. У межах цього підходу виділяють геометричні методи, методи на основі ознак, нейромережеві (глибокого навчання) підходи, а також гібридні методи, що поєднують різні стратегії обробки інформації.

Геометричні методи базуються на використанні математичних моделей камери та просторових перетворень для опису структури сцени та руху об'єктів [10]. Основою таких методів є встановлення відповідностей між точками зображення та відповідними точками у тривимірному просторі, що дозволяє відновити положення об'єктів відносно камери. Для цього використовуються проєкційні моделі камери, які описують процес формування зображення як проєкцію тривимірної сцени на площину зображення. У найпростішому випадку використовується модель камери типу «pinhole», де кожна точка простору проєктується на зображення відповідно до її геометричного положення відносно камери. Такий підхід дозволяє формалізувати зв'язок між 3D-координатами об'єкта та його 2D-проєкцією у вигляді системи рівнянь.

Визначення просторового положення об'єктів також здійснюється за допомогою методів триангуляції, які базуються на перетині променів, що проходять через відповідні точки на зображеннях, отриманих з різних ракурсів. Окремий напрям становлять методи відновлення структури сцени та руху камери (Structure from Motion), які забезпечують одночасне оцінювання параметрів камери та геометрії сцени на основі набору зображень [10].

Для підвищення точності результатів застосовуються оптимізаційні підходи, зокрема мінімізація похибки перепроєкції (bundle adjustment), що дозволяє узгодити положення точок сцени та параметри камер шляхом мінімізації різниці між спостережуваними та відновленими проєкціями. Перевагою геометричних методів є їхня висока точність, інтерпретованість та теоретична обґрунтованість. За наявності якісних вхідних даних вони забезпечують стабільні результати. Однак їх ефективність може знижуватись у складних умовах, зокрема при наявності шумів, змін освітлення або неповних відповідностей між зображеннями, що ускладнює коректне відновлення сцени.

Окрему групу становлять методи на основі ознак (feature-based methods), які орієнтовані на виявленні характерних елементів зображення, таких як кути, контури або текстурні області. Такі елементи використовуються як опорні точки для встановлення відповідностей між зображеннями або між моделлю та сценою. Процес обробки в межах цих методів включає три основні етапи [10]: виявлення ключових точок (feature detection), формування їх дескрипторів (feature description) та встановлення відповідностей (matching). На практиці для реалізації цих етапів застосовуються алгоритми Harris або Difference of Gaussians (DoG) для детекції ознак, дескриптори SIFT для їх опису, а також алгоритм RANSAC для відфільтрування хибних відповідностей. Отримані відповідності часто використовуються як вхідні дані для геометричних методів відновлення просторової структури сцени та оцінювання положення камери. Таким чином, методи на основі ознак тісно інтегровані з геометричними підходами та часто виступають, як їхній попередній етап у системах комп'ютерного зору.

Перевагами таких методів є стійкість до змін масштабу, повороту та частково освітлення, однак їх ефективність значною мірою залежить від якості виділення ознак і насиченості сцени інформативними деталями.

Сучасний розвиток комп'ютерного зору значною мірою пов'язаний з активним впровадженням нейромережових підходів. Методи глибокого навчання дозволяють автоматично визначати просторове положення об'єктів без явного задання правил або побудови складних математичних моделей. Нейронні мережі стали домінуючим підходом у задачах оцінювання положення об'єктів, демонструючи високу точність та стійкість у складних умовах [1].

Нейромережі здатні виконувати як двовимірну локалізацію об'єктів, так і оцінку їхнього положення у тривимірному просторі. Це досягається за рахунок навчання на великих наборах даних, у процесі якого моделі виявляють складні залежності між вхідними зображеннями та просторовими характеристиками об'єктів. Використання згорткових нейронних мереж суттєво підвищило ефективність задач класифікації, детекції та аналізу сцен [4].

Крім того, сучасні нейромережові підходи застосовуються для оцінювання глибини сцени за одним зображенням. Застосування моделей, що поєднують глобальний аналіз зображення та локальне уточнення, дозволяє формувати карти глибини навіть за відсутності прямої інформації про відстань [8]. Також, вони активно використовуються для визначення положення людини на основі 2D- та 3D-представлень ключових точок, що підтверджує їх універсальність у різних прикладних задачах [9].

Однією з ключових переваг нейромережових підходів є їхня здатність працювати у складних реальних умовах, зокрема при змінному освітленні, часткових перекриттях об'єктів та складних сценах. Водночас їхні обмеження пов'язані з високою обчислювальною складністю, залежністю від обсягу та якості навчальних даних, а також обмеженою здатністю до узагальнення на нові сцени [1].

Окремий напрям становлять гібридні підходи, які поєднують класичні геометричні та нейромережові моделі. У таких системах нейромережі використовуються для попереднього аналізу даних (наприклад, виявлення об'єктів або первинного оцінювання положення), тоді як геометричні методи забезпечують уточнення результатів [4]. Такий підхід дозволяє поєднати переваги обох стратегій і підвищити загальну ефективність систем комп'ютерного зору [1].

У таблиці 2 наведено результати аналізу підходів до визначення просторового положення об'єктів за методом обробки інформації.

Таблиця 2

Порівняльний аналіз підходів за методом обробки інформації

Підхід	Принцип роботи та використовувані методи	Основні переваги	Основні недоліки	Стійкість до шумів та змін	Типові сфери застосування
Геометричні	Використання математичних моделей камери та просторових перетворень; застосування проєкційних моделей епілярної геометрії та триангуляції	Висока точність за наявності якісних даних; теоретична обґрунтованість; інтерпретованість результатів	Чутливість до шумів і похибок відповідностей; Зниження ефективності в реальних умовах	Низька – середня	3D-реконструкція, фотограмметрія, аналіз сцени
На основі ознак	Виявлення ключових точок та їх співставлення; використання алгоритмів SIFT, ORB, SURF та matching-процедур	Стійкість до змін масштабу, повороту та часткового освітлення; ефективність у задачах співставлення	Залежність від якості ознак; можливість помилкових відповідностей	Середня	Відстеження об'єктів, SLAM, доповнена реальність

Продовження таблиці 2

Підхід	Принцип роботи та використовувані методи	Основні переваги	Основні недоліки	Стійкість до шумів та змін	Типові сфери застосування
Нейромережеві	Використання моделей глибокого навчання (зокрема CNN) для оцінювання положення об'єктів; регресійні та детекційні підходи	Висока точність; здатність працювати у складних сценах; адаптивність до різних умов	Висока обчислювальна складність; залежність від обсягу та якості навчальних даних	Висока	Автономні системи, розпізнавання об'єктів, AR/VR
Гібридні	Поєднання нейромережевих моделей із геометричними методами; комбіноване використання глибокого навчання та класних підходів	Баланс між точністю, стійкістю та гнучкістю; підвищена надійність результатів	Складність реалізації та інтеграції різних методів	Висока	Інтелектуальні системи комп'ютерного зору, робототехніка

Як видно з таблиці 2, ефективність визначення просторового положення об'єктів істотно залежить від обраних алгоритмічних принципів. Геометричні методи забезпечують високу точність та теоретичну обґрунтованість, однак характеризуються чутливістю до шумів і складних умов спостереження, що обмежує їх застосування у реальних сценах.

Методи на основі ознак підвищують стійкість системи за рахунок використання інформативних локальних характеристик зображення, проте їх результативність залежить від якості виділення ознак і кількості ключових точок, що ускладнює застосування в умовах низької текстурної насиченості або значного рівня шумів.

Нейромережеві підходи демонструють найвищу адаптивність і здатність до узагальнення складних просторових залежностей, що забезпечує їх ефективність у неструктурованих та динамічних умовах. Водночас вони потребують значних обчислювальних ресурсів і великих обсягів навчальних даних, що є їх суттєвим обмеженням.

Гібридні методи, що поєднують геометричні та нейромережеві підходи, забезпечують найбільш збалансоване поєднання точності, стійкості та обчислювальної ефективності. Саме вони розглядаються як один із найбільш перспективних напрямів розвитку систем комп'ютерного зору для задач відновлення тривимірної структури сцени та визначення просторового положення об'єктів.

Отже, результати аналізу підтверджують, що вибір методу обробки інформації визначається компромісом між точністю, стійкістю до зовнішніх факторів та обчислювальними витратами, а також залежить від умов задачі та доступних обчислювальних ресурсів.

Перспективи розвитку сучасних методів визначення просторового положення об'єктів пов'язані з інтеграцією різних підходів та ускладненням моделей обробки даних. Подальший розвиток у цій сфері спрямований на поєднання даних з різних сенсорів, таких як камери, глибинні сенсори та LiDAR, що дозволяє зменшити неоднозначність оцінювання просторового положення об'єктів та підвищити точність реконструкції сцени.

Водночас активно розвиваються методи глибокого навчання, які забезпечують ефективну обробку складних сцен та врахування нелінійних залежностей у даних. Значна увага приділяється створенню гібридних моделей, що поєднують геометричні та нейромережеві підходи, забезпечуючи більш збалансовану точність, стійкість та обчислювальну ефективність.

Окремим напрямом є підвищення швидкодії та точності алгоритмів, що є критично важливим для їх використання в системах реального часу. Загальною тенденцією розвитку є перехід до багаторівневих моделей, здатних одночасно враховувати просторові, часові та контекстні залежності [1].

Подальший розвиток цієї області пов'язаний не лише з удосконаленням окремих методів, а й з їх комплексною інтеграцією.

Висновки

У результаті проведеного дослідження розглянуто основні підходи до визначення просторового положення об'єктів засобами комп'ютерного зору. Показано, що ця задача є однією з ключових, оскільки лежить в основі багатьох прикладних систем, включаючи індустрію моди, дизайн інтер'єру, робототехніку, автономний транспорт, доповнену та віртуальну реальність, а також інтелектуальні системи аналізу сцен.

У роботі запропоновано узагальнену класифікацію підходів за двома ознаками: типом вхідних даних та методом їх обробки, що дозволило систематизувати існуючі методи та виявити їхні функціональні особливості, переваги й обмеження.

Проведений аналіз засвідчив, що монокулярні методи відзначаються простотою реалізації та

універсальністю, однак обмежені відсутністю прямої інформації про глибину сцени. Стереоскопічні підходи та методи з використанням глибинних сенсорів забезпечують вищу точність оцінювання просторового положення завдяки наявності додаткової інформації про глибину, але характеризуються підвищеними вимогами до апаратного забезпечення та умов застосування. Нейромережеві методи демонструють високу ефективність і адаптивність у складних та неструктурованих сценах, проте потребують значних обчислювальних ресурсів і великих обсягів навчальних даних. Гібридні підходи, що поєднують геометричні та навчальні моделі, забезпечують найбільш збалансоване співвідношення точності, стійкості та обчислювальної складності.

Узагальнення результатів показало, що не існує універсального методу визначення просторового положення об'єктів, який би однаково ефективно працював у всіх умовах. Найбільш перспективним напрямом розвитку є створення гібридних та багатосенсорних систем, які інтегрують різні типи даних і методи їх обробки для підвищення точності та надійності результатів.

Отримані результати мають практичне значення для задач автоматизованого проектування інтер'єрів, де визначення просторового положення об'єктів є основою для побудови коректних тривимірних моделей приміщень. Застосування сучасних методів комп'ютерного зору дозволяє підвищити точність розміщення елементів інтер'єру, забезпечити автоматичне узгодження їх просторової геометрії та покращити якість візуалізації.

Подальші дослідження доцільно спрямувати на підвищення точності та швидкодії алгоритмів визначення просторового положення об'єктів у режимі реального часу, а також на розробку ефективних моделей, здатних функціонувати в умовах обмежених обчислювальних ресурсів і неповної або зашумленої вхідної інформації, з урахуванням специфіки задач інтелектуального проектування інтер'єрів.

Література

1. Liu J. et al. Deep learning-based object pose estimation: a comprehensive survey. – 2024. [Електронний ресурс]. – Режим доступу: <https://arxiv.org/abs/2405.07801>. – Дата звернення: 29.03.2026.
2. Chen Y., Tian Y., He M. Monocular human pose estimation: a survey of deep learning-based methods // *Computer Vision and Image Understanding*. – 2020. – Vol. 196. – Article 102897. – DOI: 10.1016/j.cviu.2019.102897.
3. Hartley R., Zisserman A. *Multiple view geometry in computer vision*. – 2nd ed. – Cambridge : Cambridge University Press. – 2004. – 672 p.
4. Тимчишин Р. М., Волков О. Є., Господарчук О. Ю., Богачук Ю. П. Сучасні підходи до розв'язання задач комп'ютерного зору // *Control Systems and Computers*. – 2018. – № 6 (278). – С. 46–73. – DOI: 10.15407/usim.2018.06.046.
5. Пашко А. О., Тесленко О. В. Огляд основних підходів до визначення активності водія з камер, встановлених в автомобілі // *Вісник Київського національного університету імені Тараса Шевченка. Серія фізико-математичні науки*. – 2020. – № 1–2. – С. 89–94. – DOI: 10.17721/1812-5409.2020/1-2.15.
6. Kendall A., Cipolla R. Geometric loss functions for camera pose regression with deep learning. – 2017. – arXiv:1704.00390. – DOI: 10.48550/arXiv.1704.00390.
7. Newcombe R. A. та ін. KinectFusion: real-time dense surface mapping and tracking // *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. – 2011. – DOI: 10.1109/ISMAR.2011.6092378.
8. Eigen D., Puhrsch C., Fergus R. Depth map prediction from a single image using a multi-scale deep network. – 2014. – arXiv:1406.2283. – DOI: 10.48550/arXiv.1406.2283.
9. Нескородєва А. Нейромережеві методи автоматичного визначення пози людини, яка виконує вправи з художньої гімнастики // *Ukrainian Journal of Information Systems and Data Science*. – 2023. – Т. 1, № 1. – С. 53–65. – DOI: 10.31558/2786-9482.2023.1.4.
10. Szeliski R. *Computer vision: algorithms and applications*. – London : Springer, 2010. – 812 p. – DOI: 10.1007/978-1-84882-935-0.

References

1. Liu J. et al. Deep learning-based object pose estimation: a comprehensive survey. – 2024. [Elektronnyi resurs]. – Rezhym dostupu: <https://arxiv.org/abs/2405.07801>. – Data zvernennia: 29.03.2026.
2. Chen Y., Tian Y., He M. Monocular human pose estimation: a survey of deep learning-based methods // *Computer Vision and Image Understanding*. – 2020. – Vol. 196. – Article 102897. – DOI: 10.1016/j.cviu.2019.102897.
3. Hartley R., Zisserman A. *Multiple view geometry in computer vision*. – 2nd ed. – Cambridge : Cambridge University Press, 2004. – 672 p.
4. Tymchyshyn R. M., Volkov O. Ye., Hospodarchuk O. Yu., Bohachuk Yu. P. Suchasni pidkhody do rozv'iazannia zadach kompiuternoho zoru // *Control Systems and Computers*. – 2018. – № 6 (278). – С. 46–73. – DOI: 10.15407/usim.2018.06.046.
5. Pashko A. O., Teslenko O. V. Ohliad osnovnykh pidkhodiv do vyznachennia aktyvnosti vodiia z kamer, vstanovlenykh v avtomobili // *Visnyk Kyivskoho natsionalnogo universytetu imeni Tarasa Shevchenka. Serii fizyko-matematychni nauky*. – 2020. – № 1–2. – С. 89–94. – DOI: 10.17721/1812-5409.2020/1-2.15.
6. Kendall A., Cipolla R. Geometric loss functions for camera pose regression with deep learning. – 2017. – arXiv:1704.00390. – DOI: 10.48550/arXiv.1704.00390.
7. Newcombe R. A. та ін. KinectFusion: real-time dense surface mapping and tracking // *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. – 2011. – DOI: 10.1109/ISMAR.2011.6092378.
8. Eigen D., Puhrsch C., Fergus R. Depth map prediction from a single image using a multi-scale deep network. – 2014. – arXiv:1406.2283. – DOI: 10.48550/arXiv.1406.2283.
9. Neskorođieva A. Neiromerezhevi metody avtomatychnoho vyznachennia pozy liudyny, yaka vykonuie vpravy z khudozhnoi himnastyky // *Ukrainian Journal of Information Systems and Data Science*. – 2023. – Т. 1, № 1. – С. 53–65. – DOI: 10.31558/2786-9482.2023.1.4.
10. Szeliski R. *Computer vision: algorithms and applications*. – London : Springer, 2010. – 812 p. – DOI: 10.1007/978-1-84882-935-0.