

Латипов О.В.

Національний університет «Запорізька політехніка»

Єфименко М.В.

Національний університет «Запорізька політехніка»

Миронова Н.О.

Національний університет «Запорізька політехніка»

Молочков Д.Є.

Національний університет «Запорізька політехніка»

Куляба-Харитоновна Т.І.

Національний університет «Запорізька політехніка»

РОЗРОБКА СИСТЕМИ РОЗПІЗНАВАННЯ ПОЛОЖЕННЯ ОБ'ЄКТІВ У РОБОЧІЙ ЗОНІ ПРОМИСЛОВОГО РОБОТА

У статті представлено дослідження та реалізацію системи розпізнавання просторового положення об'єктів у робочій зоні промислового робота з використанням стереопари USB-камер і глибокої нейронної мережі YOLOv8. Актуальність роботи зумовлена потребою в точному тривимірному позиціонуванні об'єктів у режимі реального часу для задач автоматизованої маніпуляції, складання та взаємодії з людиною в індустріальних і дослідницьких умовах. Об'єктом дослідження є процес визначення просторового положення об'єктів у зоні дії промислового робота. Предметом дослідження є алгоритми та програмно-апаратні засоби для реалізації системи комп'ютерного зору на основі стереозображення і глибокого навчання. У роботі обґрунтовано вибір USB-камер як оптимального апаратного засобу для побудови стереопари: камери забезпечують стабільну передачу зображень, низьку затримку та високу гнучкість налаштувань. Проведено повне стереокалібрування камер із використанням шахового шаблону, що дозволило визначити внутрішні параметри, просторову геометрію та виконати ректифікацію зображень. У програмній частині реалізовано повний цикл обробки: від синхронного захоплення кадрів до виводу координат об'єктів у тривимірному просторі. YOLOv8 використовується для визначення об'єктів на обох зображеннях, після чого за допомогою аналізу диспаритету та методу тріангуляції обчислюється глибина сцени. Проведене експериментальне тестування продемонструвало стабільну роботу системи з точністю до 2–4 см в межах робочого діапазону до 1 метра. Частота обробки становила 15–20 FPS, затримка на кадр – 50–70 мс, навантаження CPU – до 80%. Аналіз впливу чинників виявив критичну роль точності калібрування, стабільності відеопотоку, синхронізації кадрів і базової лінії між камерами. У системі реалізовано візуальний контроль результатів, що забезпечує зручне тестування та подальше налагодження. Отримані результати підтверджують доцільність використання стереопари USB-камер і YOLOv8 для побудови доступної та гнучкої системи тривимірної локалізації об'єктів. Розроблений підхід може бути адаптований до інтеграції з ROS, мобільними платформами або роботизованими маніпуляторами. Подальші дослідження плануються спрямувати на реалізацію трекінгу в динаміці, оптимізацію алгоритмів для edge-пристроїв і розширення функціональності шляхом застосування методів сегментації.

Ключові слова: стереозір, YOLOv8, комп'ютерний зір, тривимірна локалізація об'єктів, глибина сцени, стереопара, калібрування камер, робототехнічні системи, нейронні мережі.

Постановка проблеми. У сучасних робото- з достатнім рівнем точності. Це необхідно для технічних системах важливо забезпечити просторову локалізацію об'єктів у реальному часі виконання операцій маніпуляції, автоматизованого складання, контролю якості продукції,

а також безпечної взаємодії між роботами та людьми. Традиційні методи, засновані на фіксованих координатах і шаблонному програмуванні, не забезпечують достатньої гнучкості в динамічному середовищі. Для вирішення цієї проблеми зростає потреба у використанні інтелектуальних систем комп'ютерного зору, здатних адаптивно розпізнавати положення об'єктів та оцінювати їх координати в просторі. Одним із перспективних підходів є інтеграція методів глибинного навчання з технологією стереозору.

Аналіз останніх досліджень і публікацій. Задача локалізації об'єктів є основною для автономної робототехніки, знаходячи застосування у промисловості, логістиці, медицині та агротехніці. Сучасні підходи до вирішення задачі локалізації об'єктів пройшли шлях від класичних алгоритмів комп'ютерного зору до складних систем, що поєднують глибокі нейронні мережі та гібридних сенсорних даних.

Класичні методи локалізації об'єктів формували основу для ранніх роботизованих систем, де точність мала вирішальне значення. Вони ґрунтуються на законах геометрії та аналітичних методах обчислення просторового положення об'єктів на основі зображень з однієї або кількох камер. Наприклад, метод стереозору дозволяє обчислити глибину шляхом аналізу відмінностей між парою зображень однієї сцени, що дає змогу отримати тривимірну інформацію про об'єкти. Методи структури з руху (Structure from Motion) відновлюють просторову сцену за серією двовимірних кадрів, використовуючи зміну положення камери або об'єктів у кадрі.

Іншим поширеним рішенням є використання маркованих систем, таких як ArUco або AprilTag, де положення ідентифікується завдяки наявності спеціальних візуальних міток [1]. Такі системи добре працюють у структурованих лабораторних або промислових середовищах.

З розвитком глибокого навчання з'явилися архітектури, що дозволили суттєво підвищити якість та швидкість локалізації. Модель YOLO та її наступні версії (YOLOv5, YOLOv8) стали стандартом для визначення об'єктів в реальному часі завдяки визначенню об'єктів за один прохід через зображення, що суттєво скорочувало час обчислень [2]. Для завдань, що вимагають вищої точності, застосовується архітектура Faster R-CNN, яка використовує дворівневий підхід для уточнення меж об'єктів [3]. Сучасні рішення, такі як трансформерні моделі DETR, використовують механізми уваги для глобального аналізу сцени,

що покращує врахування контексту [4]. Для мобільних пристроїв з обмеженими ресурсами були розроблені моделі прямої регресії позиції, наприклад PoseNet, що безпосередньо передбачають положення та орієнтацію об'єкта з одного RGB-зображення [5].

Сучасні роботизовані системи все частіше використовують гібридні підходи, що поєднують різні джерела даних для підвищення надійності. Найбільш поширеною є комбінація RGB-зображень з картами глибини (RGB-D), отриманими з сенсорів Intel RealSense [6] або подібних. Такі дані слугують входом для нейронних мереж, що значно підвищує точність тривимірної локалізації. Для навігації в динамічних середовищах використовуються візуально-інерціальні системи та алгоритми одночасної локалізації та картування, прикладом яких є RTAB-Map [7]. Отримані дані про положення об'єктів інтегруються у фреймворки планування руху, такі як MoveIt, для побудови безконфліктних траєкторій маніпуляторів [8].

На сьогодні у сфері застосування комп'ютерного зору в робототехніці все частіше реалізуються системи, що поєднують виявлення об'єктів за допомогою нейронних мереж та локалізацію їх у просторі через глибинну сенсоріку або стереозір.

Наприклад, у дослідженні [9] представлена система для зварювального маніпулятора на базі YOLOv5 та камери Intel RealSense, що досягає середньої похибки позиціонування 2,9 мм. Схожий підхід використано у роботі [10] для точного зрізу плодоніжок томатів за допомогою модифікованої YOLOv8 та RGB-D камери, де точність локалізації склала $\pm 2,5$ мм. Інші дослідження [11] зосереджені на оптимізації YOLO-архітектур для вбудованих платформ, демонструючи високу швидкість обробки навіть без використання глибинних даних.

Проведений аналіз дозволяє зробити висновок про наявність сталих архітектурних рішень у проектуванні робототехнічних систем локалізації об'єктів. Типовими ознаками таких систем є використання згорткових нейронних мереж для визначення об'єктів на зображеннях, інтеграція глибинної інформації, отриманої за допомогою стереозору або RGB-D камер, а також перетворення отриманих даних у метричні координати для подальшого використання в траєкторному плануванні та управлінні маніпуляторами. У ряді реалізацій додатково застосовуються спрощені архітектури, що сприяє зниженню обчислювальної складності та забезпечує можливість розгор-

тання таких систем на ресурсно обмежених апаратних платформах.

Постановка завдання. Метою статті є дослідження та реалізація прототипу системи розпізнавання положення об'єктів у тривимірному просторі на основі стереопари з двох USB-камер і глибокої нейронної мережі YOLOv8.

Виклад основного матеріалу. Вибір та обґрунтування апаратної бази.

Для вирішення задачі розпізнавання та локалізації об'єктів у тривимірному просторі було розроблено та реалізовано програмно-апаратний комплекс на основі стереозору та глибокого навчання. Дана методологія дозволяє з високою точністю визначати тривимірні координати об'єктів, що є критично важливим для завдань роботизованої маніпуляції. Процес розробки включав наступні етапи: вибір та обґрунтування апаратної бази, підбір програмних інструментів для обробки зображень та визначення об'єктів, а також процедуру калібрування стереосистеми для забезпечення точності вимірювань.

Основою системи є стереопара камер, що дозволяє обчислювати глибину сцени на основі ефекту паралакса. Для вибору оптимального апаратного рішення було проведено порівняльний аналіз наступних типів пристроїв: IP-камер, USB-камер та спеціалізованих глибинних сенсорів. Ключові критерії та результати порівняння наведені в табл. 1.

Аналіз показав, що для створення дослідницького прототипу пара ідентичних USB-камер є оптимальним вибором. На відміну від IP-камер, які вносять значні та непередбачувані затримки через мережеву передачу, USB-камери забезпечують стабільний відеопотік з мінімальною латентністю. Важливою перевагою є можливість програмного відключення автоматичних налаштувань (автоекспозиція, автофокус, баланс білого), що гарантує сталість параметрів зображення, необхідну для точного калібрування. Хоча глибинні сенсори надають готову карту глибини, їх висока вартість та залежність від пропрієтарних SDK роблять їх надлишковими для задач, де потрібен

Таблиця 1

Порівняння типів камер для реалізації стереозору

Критерій	IP-камери	USB-камери	Глибинні сенсори (RealSense, ZED тощо)
Затримка відео	Вища (залежить від мережі Wi-Fi)	Мінімальна	Мінімальна (апаратна обробка)
Стабільність кадрів / синхронізація	Низька, відсутня апаратна синхронізація	Вища, можливість апаратної синхронізації через USB	Вбудована апаратна синхронізація
Якість зображення	Залежить від моделі; часто автообробка на смартфоні	Залежить від моделі, стабільна за умов фіксації параметрів	Висока, спеціалізовані сенсори
Калібрування стереопари	Ускладнене (автокорекція, спотворення)	Просте (при відключенні автоекспозиції та автофокусу)	Не потребує (вбудоване позиціонування)
Гнучкість налаштувань	Середня (через web-інтерфейси), залежить від API	Висока (через прямий доступ до параметрів потоку)	Обмежена власними SDK
Простота підключення	Висока (по Wi-Fi, через IP-адресу)	Висока (Plug-and-Play через USB)	Потребує додаткового живлення, драйверів
Вартість	Дуже низька	Низька	Висока
Інтеграція з OpenCV / Python	Через MJPEG / HTTP потік, можлива затримка	Пряма підтримка, відеопотік стабільний	Потребує власного SDK або драйвера (RealSense SDK, ZED SDK)
Підтримка глибини / 3D-координат	відсутня	Потрібен обчислювальний модуль (disparity map)	Є (вбудована оцінка глибини, PointCloud, 3D frame)
Придатність до прототипування	Висока (швидкий запуск без додаткового обладнання)	Висока (особливо в лабораторних умовах)	Низька через ціну та обмежену доступність
Придатність до промислового застосування	Обмежена (залежність від мережі, нестабільність кадрів)	Висока (при правильному апаратному налаштуванні)	Висока (точність, надійність, промислові інтерфейси)

гнучкий контроль над усім процесом обчислення тривимірних координат.

Програмна реалізація та архітектура системи

Програмний комплекс розроблено на мові Python з використанням бібліотек для комп'ютерного зору. Бібліотека OpenCV виступає як основний інструмент для захоплення та обробки зображень, реалізації алгоритмів стереозору, калібрування камер та візуалізації результатів.

Для визначення об'єктів у реальному часі після розгляду альтернатив (YOLOv5, TensorFlow Object Detection API) було обрано фреймворк Ultralytics YOLOv8. Цей вибір обґрунтований його високою продуктивністю, точністю та повною реалізацією на Python, що спрощує інтеграцію. В системі використовується попередньо навчена модель yolov8n.pt, яка завантажується через API `ultralytics.YOLO()`. Модель послідовно застосовується до лівого та правого кадрів стереопари для виявлення об'єктів. Координати обмежувальних рамок, отримані від YOLOv8, використовуються для обчислення центрів об'єктів, їх зіставлення між кадрами та подальшого розрахунку диспаратності.

Процедура налаштування та калібрування стереосистеми

Для коректного обчислення глибини необхідним етапом є калібрування стереопари, що дозволяє усунути оптичні спотворення (дисторсію) та визначити точну взаємну орієнтацію камер. В експериментальній установці використовувались дві однакові USB-камери, закріплені на жорсткій основі на відстані близько 6 см між оптичними осями одна одної, спільною площиною кадру і роздільною здатністю кадру кожної камери 640×480 пікселів.

Калібрування здійснювалося за допомогою стандартної процедури з використанням шахового шаблону розміром 9×6 клітинок та функціоналу бібліотеки OpenCV (рис. 1).

Процес включав декілька етапів.

Етап 1. Визначення внутрішніх параметрів (intrinsics). Для кожної камери окремо за допомогою функції `cv2.calibrateCamera` були обчислені матриця внутрішніх параметрів (фокусна відстань, оптичний центр) та коефіцієнти дисторсії.

Етап 2. Стереокалібрування. На основі синхронізованих знімків шаблону з обох камер функція `cv2.stereoCalibrate` розрахувала параметри зовнішньої орієнтації – матрицю обертання R та вектор зсуву T , що описують просторове положення однієї камери відносно іншої.

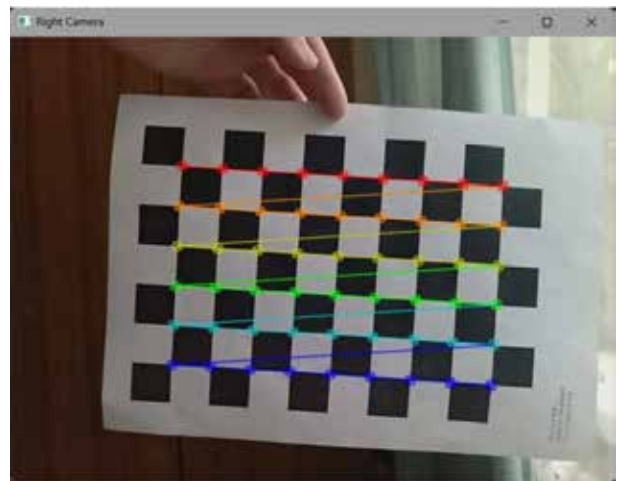


Рис. 1. Процес калібрування камери із використанням шаблону

Етап 3. Ректифікація. За допомогою `cv2.stereoRectify` були обчислені матриці ректифікації, які використовуються для проєктивного перетворення зображень. Цей процес вирівнює зображення таким чином, що епіполярні лінії стають паралельними і горизонтальними. Це гарантує, що відповідні точки об'єкта на лівому та правому зображеннях знаходяться на одній горизонтальній лінії, що значно спрощує пошук відповідностей та обчислення диспаратності.

Отримані калібрувальні параметри зберігаються у файл `stereo_calib.pkl` за допомогою модуля `pickle`, що дозволяє уникнути повторного калібрування при кожному запуску системи.

У реалізованій системі ректифікація зображень виконується в реальному часі перед передачею кадрів до нейронної мережі для розпізнавання. Це дозволяє гарантувати, що знайдені об'єкти будуть знаходитись на однаковій горизонтальній лінії, що є передумовою для точного обчислення диспаратності – різниці координат центрів одного й того самого об'єкта на лівому та правому зображенні.

Також під час калібрування було враховано: вплив освітлення на якість детекції кутів шахівниці; вимоги до стабільного положення камер протягом всієї сесії калібрування; а також проведено перевірку точності за допомогою функцій `cv2.undistort` та `cv2 remap`.

Стереокалібрування є важливим етапом побудови системи тривимірного розпізнавання об'єктів. Завдяки використанню інструментів OpenCV було отримано точні параметри внутрішньої та зовнішньої геометрії камер, що дозволило реалізувати подальше обчислення глибини та положення об'єктів у просторі. Вирівняні зображення, що утворюють епіполярну пару, забезпе-

чують необхідну точність для практичного застосування в робототехніці.

Розробка алгоритму визначення просторових координат об'єктів у стереозображенні

У цій роботі представлено алгоритм, що поєднує технології комп'ютерного зору, стереозору та глибокого навчання для визначення просторового положення об'єктів у робочій зоні промислового робота. Алгоритм реалізовано на основі бібліотек OpenCV, NumPy та нейромережевої моделі об'єктного детектування YOLOv8 від Ultralytics. Його архітектура включає послідовні етапи збору зображень, калібрування камер, ректифікації, виявлення об'єктів, співставлення їхніх проєкцій у стереопарі та обчислення просторових координат за допомогою триангуляції.

На початковому етапі здійснюється синхронне зчитування зображень з двох камер. Для цього використано шаблон шахівниці розміром 9×6 клітин, що дозволяє точно визначати внутрішні кути шаблону на зображеннях. Зображення фіксуються при різних просторових положеннях шахівниці в межах поля зору, але при цьому зберігається фіксована площа шаблону, що забезпечує достатньо варіативну й репрезентативну вибірку для подальшого калібрування. Отримані кадри автоматично розподіляються по окремих каталогах для лівої та правої камери, забезпечуючи узгодженість даних для подальшої обробки.

Після збору даних виконується внутрішнє калібрування кожної камери з використанням відомих метричних розмірів клітин шахівниці, які у даній реалізації становлять 25 міліметрів. Це дає змогу перейти від піксельних координат до фізичних одиниць виміру. За результатами калібрування для кожної камери обчислюються матриця внутрішніх параметрів та вектор дисторсії. Далі проводиться спільне стереокалібрування, в ході якого визначаються просторові трансформації між камерами у вигляді матриці обертання та вектора трансляції. Усі результати серіалізуються у форматі.pkl за допомогою модуля Python pickle, що дозволяє ефективно зберігати складні структури даних. Збережені параметри містять повний набір необхідної інформації для наступних етапів: внутрішні матриці камер, коефіцієнти спотворень, матриця обертання та вектор зміщення правої камери відносно лівої.

Наступним кроком є ректифікація отриманих зображень на основі результатів калібрування. Цей процес усуває лінзові спотворення та вирівнює зображення таким чином, щоб відповідні пари пікселів лежали на одній горизонтальній лінії.

Ректифікація дозволяє перевести вихідні зображення у координатну систему, зручну для обчислення диспаратності. Застосовано стандартні функції бібліотеки OpenCV: cv2.stereoRectify, cv2.initUndistortRectifyMap та cv2.remap, які забезпечують стабільну та відтворювану геометричну трансформацію.

До ректифікованих зображень окремо застосовується нейромережева модель YOLOv8, натренована на відповідному датасеті. Модель виконує виявлення об'єктів на зображеннях та повертає координати обмежувальних прямокутників (bounding boxes), ідентифіковані класи об'єктів та ймовірності розпізнавання. Обробка зображень відбувається незалежно для лівої та правої камери, що дозволяє уникнути впливу можливих локальних шумів або розсинхронізації між потоками. Результати виявлення передаються до наступного модуля зіставлення об'єктів.

Для кожного прямокутника, що відповідає знайденому об'єкту, розраховується центр за формулою:

$$x_c = \frac{x_{min} + x_{max}}{2}, y_c = \frac{y_{min} + y_{max}}{2}. \quad (1)$$

Це значно спрощує процес встановлення відповідностей між об'єктами на лівому та правому кадрах. Для кожного об'єкта виконується пошук пари на іншому зображенні за умови належності до одного класу, мінімальної відстані між центрами та допустимої диспаратності, що не перевищує встановлене порогове значення. Процедура зіставлення реалізована у вигляді окремої функції match_boxes(), яка забезпечує однозначне й стабільне парне зіставлення без перехресних відповідностей. Цей етап є критично важливим для подальших розрахунків, оскільки помилки на цьому рівні можуть суттєво вплинути на точність визначення відстані.

Після успішного встановлення відповідностей між об'єктами розпочинається етап обчислення просторової відстані. Використано класичну формулу триангуляції, що базується на визначенні диспаратності – горизонтального зсуву між проєкціями одного й того самого об'єкта:

$$Z = \frac{B * f}{d}, \quad (2)$$

де Z – відстань до об'єкта, B – база між камерами, f – фокусна відстань у пікселях, а d – різниця між x -координатами центрів об'єкта на лівому та правому зображеннях. Якщо диспаратність дорівнює нулю або надто мала, об'єкт вважається або занадто далеким, або некоректно співставле-

ним і не враховується у фінальних розрахунках. У випадках, коли для об'єкта на лівому кадрі не знайдено пари, система виводить відповідне повідомлення «No match».

Результати обчислення відображаються безпосередньо на зображеннях: біля кожного виявленого об'єкта зазначається відстань у метрах, а з центру прямокутника виводиться лінія до текстового підпису. Завдяки попередній ректифікації всі розрахунки відбуваються вздовж горизонтальної осі, що суттєво спрощує реалізацію алгоритму.

Точність визначення відстані залежить від якості калібрування, ректифікації, стабільності стереопари, а також точності визначення центрів прямокутників. Особливе значення має наявність достатньої диспаратності, що забезпечується за рахунок оптимального розміщення об'єктів у межах робочої зони камер.

Структура розробленої системи передбачає подальшу інтеграцію з модулями керування промисловим роботом. Дані про координати об'єктів, отримані в результаті алгоритмічного оброблення зображень, можуть бути передані до системи керування для розрахунку траєкторії руху маніпулятора з урахуванням кінематичних обмежень. Хоча модулі керування рухом і планування траєкторії не були реалізовані в межах цієї роботи, запропонована система передбачає таку можливість, що робить її перспективною для застосування в реальних роботизованих середовищах.

Експериментальне дослідження точності визначення просторового положення об'єктів

Для перевірки точності та оцінки працездатності розробленої системи було проведено експеримент, у межах якого реалізовано повний цикл: калібрування стереокамери, детекція об'єктів, їх співставлення на зображеннях та обчислення просторової відстані. Як стереопару використано дві USB-камери, закріплені з фіксованим базовим віддаленням.

Попередньо виконано стереокалібрування за шаховим шаблоном розміром 9×6 клітин. З обох камер здійснювався запис узгоджених зображень, після чого визначались внутрішні параметри кожної камери та спільні зовнішні параметри системи. Отримані матриці та вектори збережено у файл `stereo_calib.pkl`, який використовувався на подальших етапах.

Основна програма виконувала обробку потоку з камер у реальному часі. Спочатку зображення ректифікувались, після чого до кожного кадру застосовувалась нейронна модель YOLOv8n для виявлення об'єктів. На основі координат цен-

трів прямокутників об'єктів на лівому та правому зображеннях обчислювалась горизонтальна диспаратність. Враховуючи параметри стереопари, розраховувалась глибина, тобто відстань до об'єкта у метрах. Усі обчислення реалізовано на Python із використанням бібліотек OpenCV, NumPy та PyTorch. У графічному інтерфейсі виводились результати у вигляді класу об'єкта та його відстані.

У ході експерименту, проведеного при денному освітленні, перевірялась точність вимірювання для різних об'єктів: пляшки, руки людини та інших побутових предметів. Наприклад, при реальній відстані до пляшки близько 41,5 см система повідомляла значення в межах 38–42 см. Середнє відхилення складало близько 2 см, що є прийнятним результатом для недорогого обладнання. У випадку руки, що знаходилась на відстані 50–52 см, модель іноді повертала завищене значення (до 60 см), імовірно через особливості детекції YOLO: прямокутник формувався на всю руку, частково розташовану на межі кадру (рис. 2).

Найбільшу точність система демонструвала в діапазоні 0,5–3 м, що відповідає типовому робочому простору промислового маніпулятора. Незначні коливання значень у межах 1–2 см могли бути спричинені низькою якістю камер, неточним калібруванням, відсутністю жорсткої синхронізації та малою довжиною базової лінії (близько 6 см).

Точність системи істотно залежала від низки чинників. По-перше, якість стереокалібрування: недостатня кількість кадрів або неякісне виявлення кутів шаблону призводили до систематичних похибок, особливо у визначенні фокусної відстані чи бази. По-друге, апаратні характеристики камер: обмежена роздільна здатність (640×480), нестабільність автофокусу, розмиття та шуми ускладнювали точне позиціонування об'єктів.

Крім того, на точність впливала довжина базової лінії. При базі 6 см система показувала хорошу точність на малих відстанях, однак на більших – диспаратність зменшувалась до кількох пікселів, що ускладнювало вимірювання. Збільшення бази до 10–12 см дозволило б суттєво покращити чутливість.

Форма та положення об'єктів також були критичними. Центр прямокутника не завжди точно відповідає центру мас об'єкта, особливо у випадках складної геометрії чи часткового перекриття, що спричиняло хибні значення. У таких випадках доцільно використовувати методи сегментації для уточнення контурів.

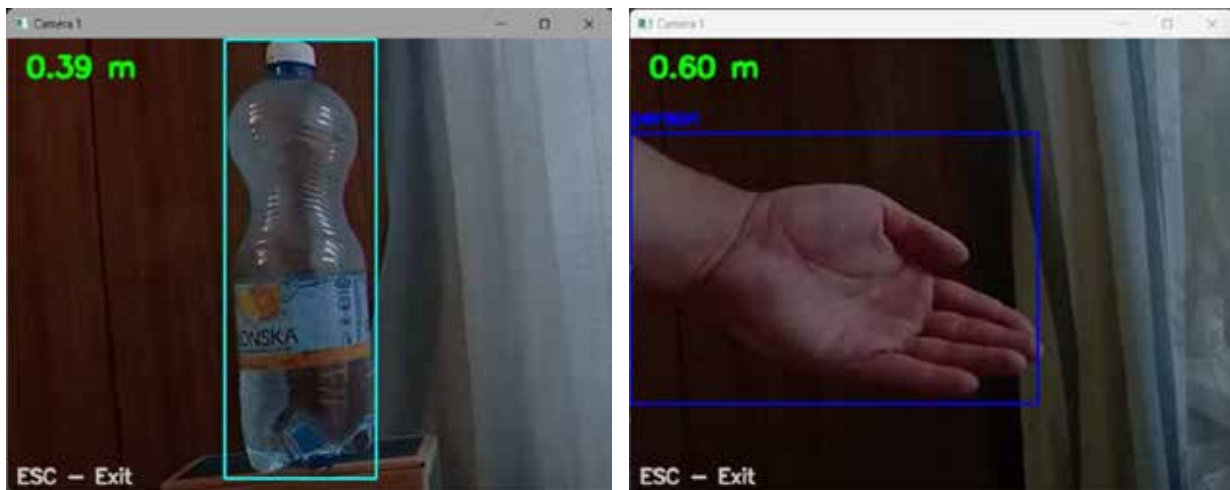


Рис. 2. Тестування роботи системи

На результати впливали також умови освітлення, стабільність відеопотоку та синхронізація кадрів. Варіації яскравості, тіні або відблиски викликали хибну детекцію або зсув bounding box. Для зменшення таких ефектів рекомендується контрольоване освітлення та попередня обробка зображень.

Таблиця 2

Основні параметри продуктивності системи

Метрика	Значення
FPS	15-20
Затримка на кадр	50-70 мс
Навантаження CPU	65-80%
Об'єм ОЗП	~750 МБ

Підсумовуючи, система забезпечила стабільну роботу при обмежених апаратних ресурсах та досягла прийнятної точності для задач середнього рівня складності. Отримані результати підтверджують ефективність запропонованого підходу до визначення просторового положення об'єктів із використанням комп'ютерного зору на основі стереозображень та глибокого навчання.

Висновки. У роботі представлено результати дослідження та реалізації системи визначення просторового положення об'єктів на основі стереопари USB-камер та моделі об'єктного детектування YOLOv8. Запропонований підхід забезпечує повний цикл обробки – від калібрування

камер до розпізнавання об'єктів, співставлення проєкцій та обчислення відстані до цілі – з прийнятною точністю для використання в дослідницьких, лабораторних і промислових умовах.

Ключовими перевагами системи є низька вартість апаратного забезпечення, використання доступних засобів (камер і комп'ютера), гнучкість налаштувань, а також можливість адаптації до різних прикладних сценаріїв. До недоліків слід віднести залежність точності від якості калібрування, роздільної здатності камер, умов освітлення та довжини базової лінії стереопари.

Подальший розвиток системи може бути спрямований на інтеграцію механізмів відстеження об'єктів у динаміці для забезпечення безперервного розпізнавання без повторної ідентифікації у кожному кадрі. Важливим кроком є також поєднання з платформою ROS та інтеграція з роботизованими маніпуляторами або мобільними платформами.

Застосування запропонованої системи можливе в низці практичних задач, включаючи розпізнавання жестів і положень рук у системах людино-машинної взаємодії, сервісну робототехніку, навчальні комплекси для вивчення комп'ютерного зору та роботизованих систем. Запропонований підхід є гнучкою та доступною основою для побудови інтелектуальних систем з візуальним зворотним зв'язком, здатних працювати в умовах реального середовища.

Список літератури:

1. OpenCV: Detection of ArUco Markers URL: https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html (date of access: 24.06.2025).
2. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. – 2016. – С. 779–788. – DOI: 10.1109/CVPR.2016.91.

3. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks URL: <https://dl.acm.org/doi/10.5555/2969239.2969250> date of access: 28.06.2025).
4. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. End-to-End Object Detection with Transformers. – 2020. – URL: <https://arxiv.org/abs/2005.12872> (date of access: 28.06.2025).
5. Kendall A., Cipolla R. PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization URL: https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Kendall_PoseNet_A_Convolutional_ICCV_2015_paper.pdf (date of access: 28.06.2025).
6. Intel RealSense SDK 2.0 – Depth and Tracking cameras URL: <https://www.intelrealsense.com/sdk-2/> (date of access: 28.06.2025).
7. RTAB-Map – Real-Time Appearance-Based Mapping URL: Режим доступу: <https://introlab.github.io/rtabmap/> (date of access: 28.06.2025).
8. MoveIt 2 Documentation URL: <https://moveit.picknik.ai/main/index.html> (date of access: 04.07.2025).
9. Zhang C., Wang Q. A guidance system for robotic welding based on an improved YOLOv5 algorithm with a RealSense depth camera // Scientific Reports, 2023. URL: <https://www.nature.com/articles/s41598-023-48318-8> (date of access: 04.07.2025).
10. Song H., Zhang Y., Liu J., Li X., Wang Y. Study on the fusion of improved YOLOv8 and depth camera for bunch tomato stem picking point recognition and localization // Frontiers in Plant Science. – 2024. – URL: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1447855/full> (date of access: 04.07.2025).
11. Wang W., Liu Y., Tan M. Transformers for Remote Sensing: A Systematic Review and Analysis // Sensors. – 2024. – V. 24, № 11. – P. 3495. – DOI: <https://doi.org/10.3390/s24113495>.

Latypov O.V., Yefymenko M.V., Myronova N.O., Molochkov D.Ye., Kulyaba-Kharytonova T.I.
DEVELOPMENT OF A SYSTEM FOR RECOGNISING THE POSITION OF OBJECTS
IN THE WORKING AREA OF AN INDUSTRIAL ROBOT

The article presents the study and implementation of a system for recognising the spatial position of objects in the working area of an industrial robot using a stereo pair of USB cameras and the YOLOv8 deep neural network. The relevance of the work is due to the need for accurate three-dimensional positioning of objects in real time for automated manipulation, assembly and interaction with humans in industrial and research environments. The object of research is the process of determining the spatial position of objects in the area of operation of an industrial robot. The subject of the study is algorithms and software and hardware for the implementation of a computer vision system based on stereo imaging and deep learning. The paper substantiates the choice of USB cameras as the optimal hardware for building a stereo pair: the cameras provide stable image transmission, low latency, and high flexibility of settings. The cameras were fully stereo-calibrated using a staggered template, which allowed us to determine the internal parameters, spatial geometry, and perform image rectification. The software part implements a full processing cycle: from synchronous frame capture to the output of object coordinates in three-dimensional space.

YOLOv8 is used to detect objects in both images, after which the scene depth is calculated using disparity analysis and triangulation. The experimental testing demonstrated stable operation of the system with an accuracy of 2-4 cm within the working range of up to 1 metre. The processing frequency was 15-20 FPS, the frame delay was 50-70 ms, and the CPU load was up to 80%. An analysis of the influence of factors revealed the critical role of calibration accuracy, video stream stability, frame synchronisation and baseline between cameras. The system provides visual control of the results, which ensures convenient testing and subsequent debugging. The obtained results confirm the feasibility of using a stereo pair of USB cameras and YOLOv8 to build an affordable and flexible system for 3D object localisation. The developed approach can be adapted for integration with ROS, mobile platforms, or robotic manipulators. Further research is planned to focus on implementing tracking in dynamics, optimising algorithms for edge devices, and extending the functionality by applying segmentation methods.

Key words: stereo vision, YOLOv8, computer vision, three-dimensional object localization, depth estimation, stereo camera pair, camera calibration, robotic systems, neural network.

Дата надходження статті: 24.07.2025

Дата прийняття статті: 29.07.2025

Опубліковано: 27.10.2025