

**Беілей М.І.**

Національний університет «Львівська політехніка»

**Пастух В.А.**

Національний університет «Львівська політехніка»

**Беілей Г.В.**

Національний університет «Львівська політехніка»

**Андрущак В.С.**

Національний університет «Львівська політехніка»

**Климаш М.М.**

Національний університет «Львівська політехніка»

## РОЗРОБКА ІНТЕЛЕКТУАЛЬНОЇ МОБІЛЬНОЇ СИСТЕМИ КОМП'ЮТЕРНОГО ЗОРУ ТА ГОЛОСОВОГО УПРАВЛІННЯ ДЛЯ ЛЮДЕЙ З ВАДАМИ ЗОРУ

*У статті на основі смартфона розроблено інтелектуальну мобільну систему комп'ютерного зору та голосового управління, спеціально адаптованої для користувачів з вадами зору. Розроблена система працює в режимі реального часу та надає можливість використання як в онлайн, так і в офлайн режимах, розширюючи функціональні можливості сучасних рішень для незрячих та слабозорих осіб. Важливим аспектом є впровадження модуля голосового керування, який використовує комбінацію вбудованих та хмарних API неймережевих моделей, що значно підвищує доступність та ефективність системи. Ми використали передові технології, такі як нейронна мережа MobiNet3 та модель MiDaS 2 Lite для класифікації об'єктів та оцінки глибини зображення. Ми вдосконалили традиційні методи візуального розпізнавання та впровадили ітераційний асинхронний алгоритм обчислення глибини, що забезпечило надзвичайну швидкість системи. Результати тестування підтверджують, що час обробки для нашого ітеративного алгоритму знаходиться в діапазоні від 600 мкс до 2 мс, що є вражаючим покращенням порівняно із традиційними рекурсивними методам. Ми внесли нововведення в систему, включаючи розрахунок коефіцієнтів глибини зображення в обох горизонтальній та вертикальній площинах, що представляє значущий крок удосконалення в порівнянні з попереднім нашим варіантом. Ці зміни сприяють точнішому визначенню просторового оточення системи, поліпшуючи її здатність ефективно уникати перешкод. Крім того, ми впровадили аналіз даних з акселерометра та гіроскопа, що надає важливу інформацію про рух та орієнтацію пристрою. Акселерометр оцінює зміни прискорення, нахилу та обертання, що дозволяє отримати комплексне уявлення про просторову динаміку. Гіроскоп доповнює цей аналіз, надаючи інформацію про кутові зміни в орієнтації пристрою. Ці інтегровані покращення грають вирішальну роль у виконанні функцій, таких як уникнення об'єктів, голосове управління, навігаційна допомога та просторова орієнтація для осіб із вадами зору. Тестування підтвердило, що комбінація голосового управління та розпізнавання зображень дозволяє нашій системі працювати в режимі реального часу та забезпечувати користувачів актуальною інформацією.*

**Ключові слова:** смартфон, навігаційна допомога, комп'ютерний зір, голосове управління, інтелектуальна мобільна система, нейронна мережа, мобільність, ітераційний асинхронний алгоритм, глибина зображення.

**Постановка проблеми.** У світі стрімкого розвитку технологій важливо не лише створювати інноваційні рішення, але й забезпечувати їх доступність та використання в повсякденному житті. Однією з ключових сфер, де технології можуть справжньою мірою змінити життя,

є розробка інтелектуальних мобільних систем для людей з вадами зору [1]. У цьому контексті, покликано вирішити низку проблем, що виникають при використанні сучасних рішень та розширити їх функціонал за допомогою смартфона, який став необхідним елементом повсякденного життя.

На сьогоднішній день багато існуючих технологій, спрямованих на полегшення життя людей з вадами зору, часто обмежені своєю функціональністю, портативністю та високою вартістю. Великі та незручні пристрої можуть ставити під сумнів їхню практичність у повсякденному житті. Багато з пристроїв для незрячих працюють в обмеженому режимі, надаючи базову інформацію про оточуючий світ. Зазвичай, ці рішення не забезпечують детального опису об'єктів навколо та не дозволяють повноцінно взаємодіяти з навколишнім середовищем [2]. Також важливим фактором є не лише точність розпізнавання об'єктів та голосу, а й можливість забезпечення цього процесу в реальному часі. Затримки чи невдалі спроби розпізнавання можуть суттєво впливати на здатність користувача ефективно взаємодіяти з навколишнім світом та отримувати актуальну інформацію. Це стає особливо актуальним, коли користувачі перебувають в русі, переходять дорогу або виконують щоденні обов'язки. Отже, однією з найважливіших характеристик ефективної системи для людей з вадами зору на базі смартфона є її здатність працювати в режимі реального часу [3].

**Аналіз останніх досліджень і публікацій.** Експоненціальне зростання індустрії смартфонів проклало шлях до їх більш глибокого вивчення в дослідженнях навігаційної допомоги. Система навігаційної допомоги, запропонована в роботі [4] автори використовували смартфон для взаємодії з користувачем за допомогою голосового введення. Стереокamera використовувалися для зйомки відео з навколишнього середовища, а потім надсилання його на хмарну обчислювальну платформу. У роботі [5, 6] пропонувані системи вимагали постійного підключення до мережі передачі даних, щоб функціонувати через віртуальні окуляри. Система, представлена в роботі [7], складається з системи розпізнавання зображень, інтегрованої з додатком для смартфонів. Система підтримує два режими роботи залежно від доступності мережі: онлайн і офлайн. Смартфон використовувався для захоплення об'єктів перед користувачем і надсилання їх на внутрішній сервер. Для розпізнавання об'єктів було застосовано два алгоритми: швидший R-CNN та YOLO. Швидший алгоритм R-CNN використовувався в автономному режимі системи для отримання вищої точності. Алгоритм YOLO, навпаки, застосовувався в онлайн-режимі для отримання більшої швидкості обробки. Після визначення перешкод і відстані до них користувач отримував повідомлення про результати в аудіорежимі.

У роботі [8], автори розробили гнучку і портативну систему з використанням мобільного телефону з глибокою камерою для виявлення перешкод без необхідності придбання додаткових апаратних пристроїв. На рис. 1 показано запропоновану систему для носіння. Смартфон вставляється в портативну, складну тканинну сумку-слінг, яка вішається на шию користувачеві. Недоліком даної роботи є обмежена точність виявлення перешкод через використання лише глибокої камери мобільного телефону.



**Рис. 1.** Система допомоги при пересуванні людей з вадами зору за допомогою смартфона на базі Android (праворуч) та слінг-сумки (ліворуч)

Автори роботи [9] представили систему навігаційної допомоги, основними компонентами якої є RGB-D камера, інерційний вимірювальний блок (IMU), встановлений на парі окулярів, і смартфон. На смартфоні був встановлений легкий CNN для виявлення перешкод, визначення їх положення та орієнтації. Хоча автори стверджують, що система була протестована і працює як в приміщенні, так і на вулиці, кабелі, що з'єднують різні компоненти системи, можуть створювати незручності для користувачів під час навігації.

PERCEPT-II - це система навігації в приміщенні на основі смартфона, запропонована в роботі [10]. Це додаток для Android, який дозволяє людям з вадами зору отримувати навігаційні інструкції до місця призначення, коли вони торкаються певних орієнтирів, обладнаних мітками ближнього радіозв'язку. Основне обмеження системи пов'язане з установкою міток NFC, що передбачає високі витрати на обслуговування при широкомасштабному розгортанні.

Підсумовуючи, можна сказати, що значні обмеження більшості розглянутих систем пов'язані з портативністю, що робить їх незручними під час навігації. Хоча існують портативні рішення, багато з них залежать від підключення до мережі та обробки даних на сервері. Користувачі, які потребують підключення до Інтернету під час

навігації, можуть бути стурбовані питаннями конфіденційності та безпеки. Хоча деякі дослідники вивчали моделі виявлення об'єктів загального призначення з кращою точністю, вони не показали, як їх можна інтегрувати в портативну систему з низькими обчислювальними ресурсами. Як правило, при використанні в реальному часі моделі виявлення об'єктів/перешкод повинні мати малий час обробки даних без шкоди для точності. Однак, коли використовуються універсальні моделі об'єктів, вони можуть мати достатню точність, але потребувати більше часу для отримання результату, що може спричинити аварії або зіткнення під час навігації через затримку в часі реагування [1].

Таким чином, метою роботи є розроблення та впровадження інтелектуальної мобільної системи комп'ютерного зору та голосового управління для людей з вадами зору на основі смартфона, що працює в режимі реального часу та надає можливість використання як в онлайн, так і в офлайн режимах. Головний акцент робиться на забезпеченні швидкого та ефективного доступу користувачів до необхідної інформації, навігації та взаємодії з системою в різних умовах використання.

**Виклад основного матеріалу дослідження.** Запропонована система, яка визначає себе як універсальний додаток для смартфонів, призначена для задоволення потреб користувачів із вадами зору. Вона спроектована для роботи на смартфонах з операційною системою Android, і може бути адаптована до інших платформ за умови наявності додаткової програмної та апаратної підтримки. Перелік функцій системи є лише частковим і може бути розширений у майбутньому. Структурно-функціональна схема мобільної системи включає різні модулі, які між собою взаємодіють та знаходяться на різних рівнях: рівень даних, логічний рівень та рівень представлення (рис. 2).

*Рівень даних* у системі включає різні функціональні модулі, кожен з яких має своє власне призначення:

- Модуль обробки зображень, який використовує алгоритми, автономні нейромережеві моделі та комунікаційні інтерфейси для обробки графічних даних.
- Модуль навігації, що забезпечує повну навігаційну функціональність користувача, розраховуючи маршрути, траєкторії та переміщення на основі відкритих карт та даних пристрою.
- Модуль сенсорних даних з прямими інтерфейсами для отримання даних з апаратних компонентів, таких як датчики та комунікаційні інтерфейси.

- Модуль аналітики, що складається з вбудованих і сторонніх компонентів для зберігання та аналізу даних, включаючи базу даних.

- Модуль голосового управління, що використовує комбінацію вбудованих і хмарних систем голосового управління API для підвищення доступності.

- Мережевий модуль, який служить інтерфейсом зв'язку з серверами та різними хмарними сервісами.

*Рівень бізнес логіки* домену складається з динамічних варіантів використання, які розглядаються як атомарні операції і конфігуруються на основі багатопотоковості системи. Приклади включають обробку зображень і зіставлення з даними акселерометрів і гіроскопічних датчиків для прогнозування патернів руху користувача.

*Рівень представлення*, залежний від платформи розгортання, має чітко визначені інтерфейси для регулювання взаємодії між користувачем і системою. Ці інтерфейси включають аудіовихід пристрою, вібраційні сигнали, дисплей та зв'язок з периферійними пристроями. Кожен рівень функціонує автономно з логічно відокремленими модулями, що взаємодіють через реалізовані інтерфейси. Типи даних сутностей адаптовані до конкретної робочої області для забезпечення цілісності та безпеки. Такий архітектурний підхід сприяє зручності розширення системи, міграції, уніфікації модулів та крос-платформеній реалізації в майбутньому.

У даній системі реалізована паралельна обробка завдань для оптимізації використання ресурсів. Деякі процеси, такі як початкова ініціалізація та основні функціональності навігації, виконуються в одному потоці виконання для уникнення конфліктів та оптимізації витрат часу. Завдання, які можуть виконуватися паралельно, такі як навігація та обробка зображень, розподіляються на підзадачі, що використовують асинхронний підхід та можуть виділяти ресурси при виході зі стану паузи.

Для запропонованої системи ми успішно розробили та впровадили модуль безперервного голосового управління, використовуючи хмарний сервіс Speech-To-Text API. Цей модуль дозволяє використовувати складну структуру голосових команд, повністю відповідаючи основним вимогам нашої системи. Однак, враховуючи залежність від стабільного інтернет-з'єднання для використання API, ми також впровадили альтернативний модуль обробки голосових команд. Зокрема він включає можливості використання вбудованої

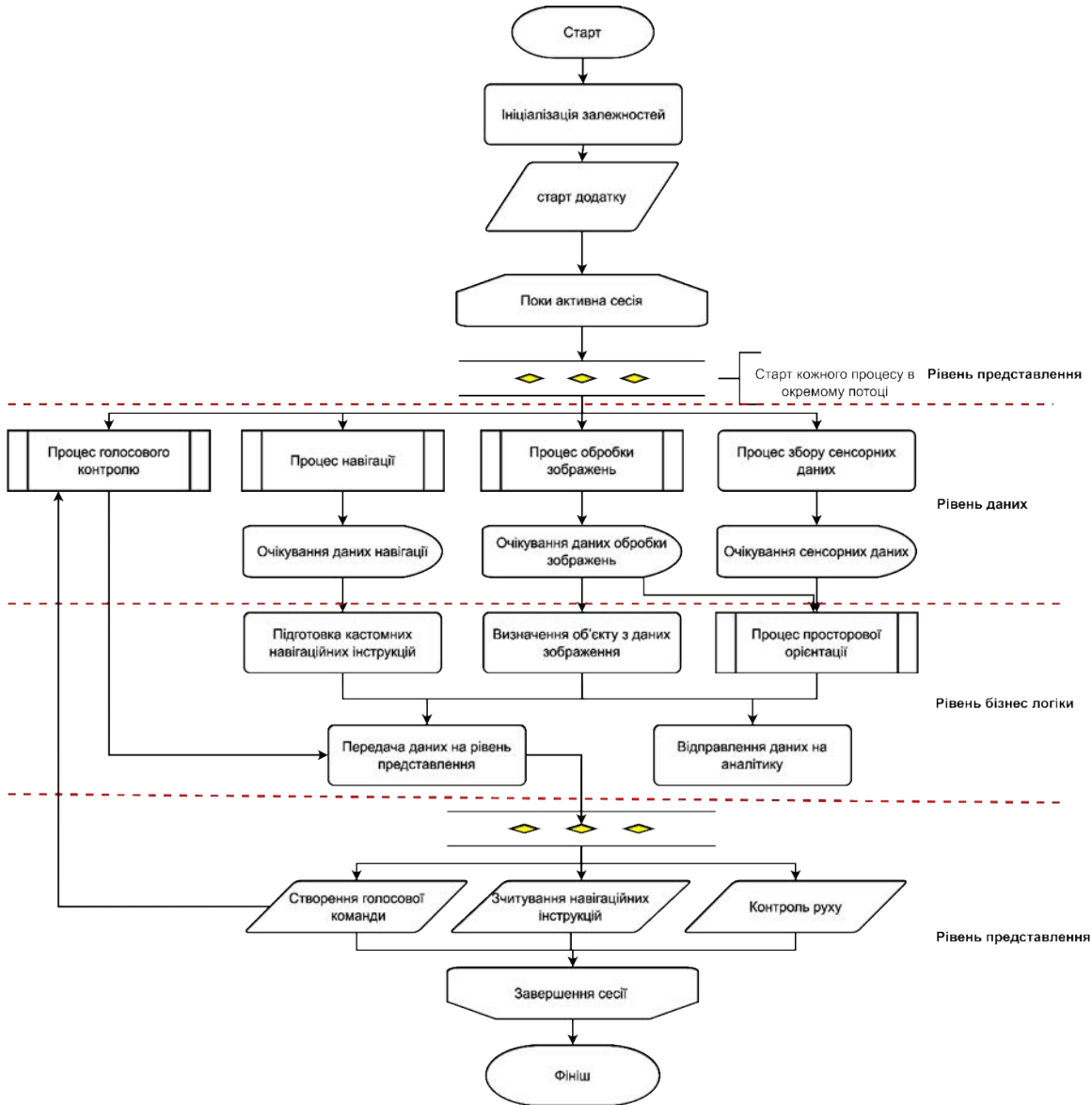


Рис. 2. Структурно-функціональна схема мобільної системи комп'ютерного зору та голосового управління для людей з вадами зору

нейромережевої моделі для класифікації команд та голосового управління, надаючи системі автономність та високу швидкість роботи. Цей модуль може діяти як резервний варіант у випадку недоступності основного Speech-To-Text API, забезпечуючи неперервну функціональність системи в режимі реального часу. З метою оптимізації та інтеграції обох модулів в єдину систему збору даних, ми внесли покращення у загальну архітектуру модуля безперервного голосового управління (рис. 3). Ми розробили модифікований репозиторій VoiceRecognitionRepository, який об'єднує два різних джерела даних: VoiceRecognitionDataSourceApi, що використовує Speech-To-Text API для класифікації команд, та

VoiceRecognitionDataSourceEmbeddedModel, яке використовує вбудовану нейромережеву модель (Embedded AI) для того ж самого завдання.

У класі репозиторію були додані дві окремі функції розпізнавання команд для кожного джерела даних відповідно. Також була впроваджена логіка, яка дозволяє динамічно перемикатися між цими джерелами на основі умов, таких як помилки або відмови в роботі модулів. Результати нашої роботи відображені на блок-схемі, представлений на рис. 4.

Отже, загальний алгоритм безперервного голосового управління був адаптований з урахуванням умов використання різних джерел розпізнавання команд та функціонує наступним чином:

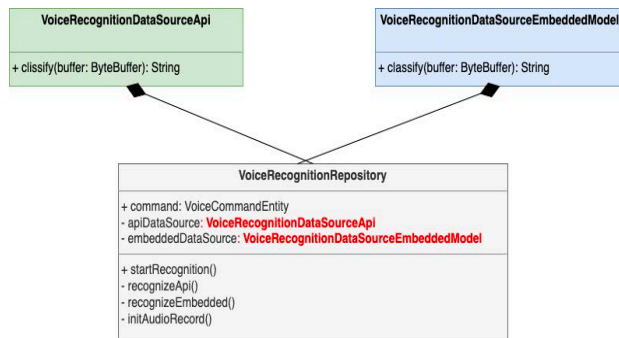


Рис. 3. Діаграма класів модуля безперервного голосового управління

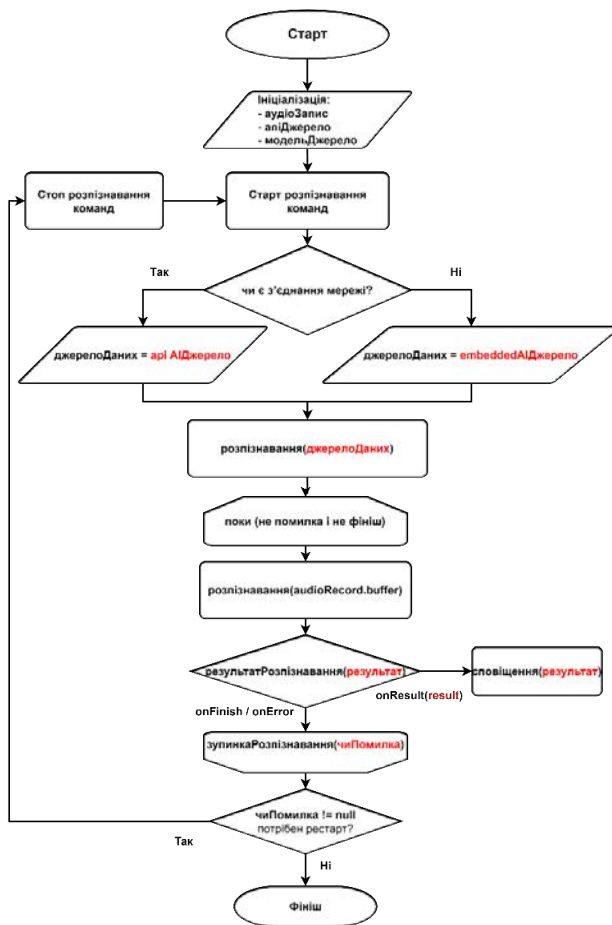


Рис. 4. Блок-схема алгоритму голосового управління для онлайн та офлайн режиму

1. Ініціалізація джерел даних для модулів розпізнавання команд з використанням API та вбудованої нейронної мережі.
2. Початок запису аудіоданих.
3. Перевірка доступності Інтернет-з'єднання (активне/неактивне).
4. Перемикання модуля відповідно до результату перевірки Інтернет-з'єднання.
5. Запуск циклу розпізнавання команд.
6. Обробка результатів розпізнавання (команда / помилка / завершення):

- Команда: активація відповідної команди.
- Помилка: зупинка функції розпізнавання, перезапуск функції розпізнавання.
- Завершення: зупинка функції розпізнавання, завершення роботи модуля голосового керування.

Ми провели порівняльний аналіз двох підходів реалізації голосового контролю. Розглядали параметри швидкості обробки та точності результату для моделі Embedded AI та Speech-to-Text Cloud API (рис. 5).

Модель вбудованого ШІ класифікує аудіокоманди з обмеженим списком на дев'ять елементів, що дозволяє швидко обробляти дані локально. Це, однак, призводить до обмежень у статичному списку команд та можливих помилок у розпізнаванні.

Speech-to-Text Cloud API, реалізований на стороні клієнта, надає більшу гнучкість з можливістю динамічного розширення списку розпізнаваних сутностей. Однак, цей підхід вимагає більше часу на обробку через відправлення аудіоданих на віддалений сервер.

VoiceControlledApp Embedded model			VoiceControlledApp Speech recognition API		
Result	Accuracy	Time ms	Result	Accuracy	Time ms
go	0.9984401	24	START	0.98762906	230
stop	0.99988866	33	STOP	0.95282274	70
up	0.87787783	37	NEXT	0.9738304	320
right	0.99842393	41	BACK	0.9237496	420
up	0.99994926	39	ALLOW	0.96366084	390
up	0.98488694	36	DISALLOW	0.78784275	450
down	0.99250424	38	EXIT	0.98762906	970
up	0.79631543	30	SETTINGS	0.98762906	510
right	0.40231836	36	AUTH	0.9716112	40
up	0.97831184	32	ABOUT	0.98000145	560
up	0.32232675	51	HELP	0.98472625	940
right	0.9343734	40	UNIQUE	0.98762906	550
up	0.9337167	41			
go	0.99989104	31			
down	0.614984	29			
left	0.7197286	29			
go	0.90625566	40			
left	0.9996886	30			
stop	0.56151736	33			
off	0.84867606	33			
up	0.48891592	34			
up	0.9837015	37			
up	0.80413634	61			
up	0.4909583	31			
go	0.70920146	31			
go	0.92218935	38			
down	0.9959894	29			

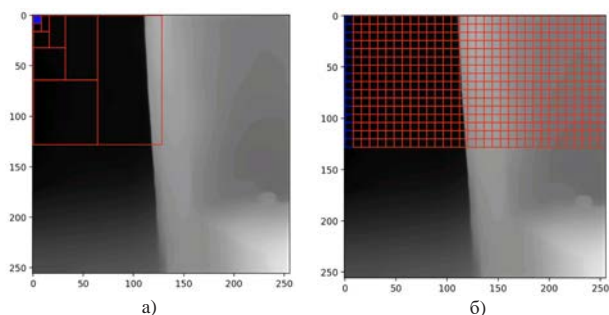
Рис. 5. Порівняння ефективності моделі Embedded AI (а) з моделлю Speech-to-Text Cloud API (б)

Також, у пропонуваній системі використовується нейромережева модель MobiNet третьої версії для класифікації об'єктів на зображенні. Модель приймає оброблений піксельний масив розміром 512x512 пікселів у якості вхідних даних. Результатом роботи моделі є словник із цілими числами та плаваючою комою, що відображає унікальний номер класифікованого об'єкта та

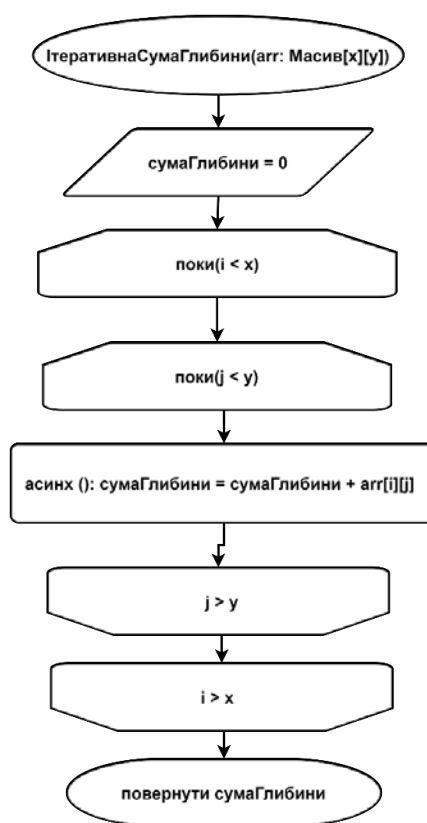
йому відповідне відсоткове значення. Значення з найвищим відсотком визначається як кінцевий результат для обробленого кадру та передається користувачеві. Модель оцінки глибини MiDaS 2 Lite використовує тензор та базується на роботі з простими зображеннями від камери. Вона навчена на 10 різних наборах даних і на вході отримує двовимірний піксельний масив розміром 256x256 пікселів. Результатом моделі є одновимірний масив значень з плаваючою комою, що відповідає квадрату початкового розміру зображення. Отримане зображення використовується для керування переміщенням користувача в просторі за допомогою рекурсивного алгоритму визначення та уникнення перешкод по горизонталі. Глибина зображення визначає відстань до об'єктів, і ці дані використовуються для коригування переміщення користувача в просторі. Під час обробки зображень для оцінки глибини використовується вихідне зображення для розрахунку варіантів руху користувача для уникнення перешкод. У попередній нашій роботі ми використовували рекурсивний метод [11] ітерацій над матрицею карти глибини для обчислення суми значень пікселів зображення (рис. 6). Цей підхід виявився ефективним і давав бажаний результат, але він придатний лише для синхронних обчислень. Використання асинхронного підходу призводило до втрат продуктивності, оскільки кожен рекурсивний крок, виділений в окремий мікропотік, блокував весь контекст асинхронної роботи. Кожен батьківський мікропотік мусив чекати завершення роботи всіх створених ним мікропотоків, що збільшувало час обробки матриці. Для оптимізації часу обробки при використанні асинхронного поділу обчислень матриці карти глибини, алгоритм проходження по елементах матриці було змінено на ітераційний підхід (рис. 7).

Асинхронний ітераційний алгоритм обчислення суми матриці карти глибин працює за звичайним принципом обчислення значень у двовимірному масиві, за винятком того, що обчислення одновимірних масивів відбувається асинхронно (рис. 7).

Ключові етапи алгоритму включають вибір змінної для суми, використання зовнішнього циклу для ітерації значень  $x$  у масиві, де кожен  $x$  представляє собою одновимірний масив  $y$ , та асинхронне обчислення значень одновимірних масивів та збереження суми  $y$  у відповідній змінній. Алгоритм завершується виходом з циклів та поверненням готової суми значень масивів. Для проведення експериментів і тестування різних



**Рис. 6.** Рекурсивний (а) та ітераційний (б) підхід обрахунку матриці карти глибини зображення



**Рис. 7.** Асинхронний ітераційний алгоритм обчислення суми матриці карти глибини зображення

підходів було використано тестове середовище з картами глибин різних розмірів. Було розраховано середні часи обробки для рекурсивного та асинхронного ітераційного підходів, провівши по 500 експериментів для кожного варіанту розмірності матриці. Результати тестування алгоритму з рекурсивним підходом до обчислення значень матриці карти глибин представлені в таблиці 1. Середні значення часу обробки коливаються в діапазоні від 600 мкс до 70 мс.

Таблиця 1

## Дані тестування алгоритму з рекурсивним підходом

	index	count	mean	std	min	25%	50%	75%	max
0	64	500.0	689.812	2342.05144	3.0	9.00	152.5	439.75	33609.0
1	128	500.0	2214.28	4013.85656	22.0	75.00	702.5	2950.50	34894.0
2	256	500.0	4088.71	7716.00429	104.0	191.75	1893.0	3214.50	65868.0
3	512	500.0	8604.23	14591.0224	344.0	547.00	5465.0	11803.25	102867.0
4	1024	500.0	23498.7	31941.7372	1257.0	1636.75	17788.	40079.75	334632.0
5	2048	500.0	70002.6	78368.7497	4309.0	4717.75	66035.	118967.5	999279.0

Таблиця 2

## Дані тестування алгоритму з асинхронним ітеративним підходом

	index	count	mean	std	min	25%	50%	75%	max
0	64	500.0	639.746	4396.02188	119.0	278.75	330.5	436.25	98358.0
1	128	500.0	842.992	4454.18267	206.0	447.00	535.0	662.50	99803.0
2	256	500.0	1237.098	4528.91724	329.0	794.75	950.5	1075.25	100069.0
3	512	500.0	1194.190	438.767702	691.0	875.50	1022.0	1520.25	3591.0
4	1024	500.0	1268.908	378.224885	611.0	899.00	1326.5	1559.75	2324.0
5	2048	500.0	2191.168	613.270983	1530.0	1863.00	1999.5	2278.00	9053.0

Дані тестування алгоритму з асинхронним ітеративним підходом до обчислення значень матриці карти глибини наведені в таблиці 2. Середні значення часу обробки для цього підходу коливаються від 600 мкс до 2 мс.

Отже, за результатами дослідження двох підходів до обчислення карти глибин виявилось, що ітеративний асинхронний підхід у середньому в 35 разів швидший, ніж рекурсивний (рис. 8). Тому він був обраний як основа для алгоритму визначення коефіцієнтів руху користувача, оскільки він дозволяє ефективніше передавати результати користувачеві у вигляді команд.

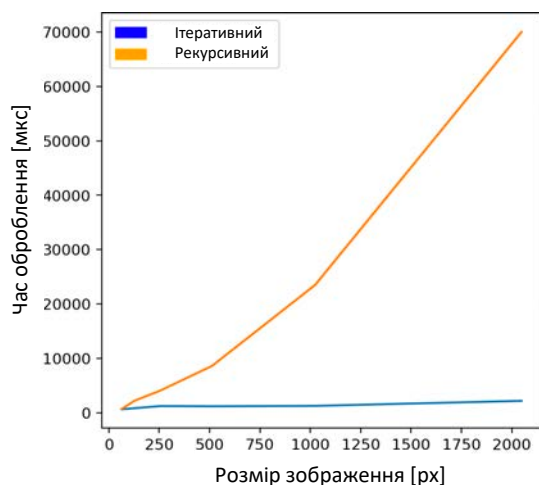


Рис. 8. Порівняння часу оброблення карти глибини зображення різних підходів

Ми також вдосконалили алгоритм, додаючи розрахунок перешкод як по горизонтальній, так і по вертикальній осі на основі визначення коефі-

цієнтів глибини зображення. Для цього площина зображення ділиться на чотири частини для кожної з них визначається центр, з якого буде виконуватися розрахунок. Отримані половинки вихідного масиву передаються як параметр на вхід інтерактивного алгоритму для обчислення суми значень глибини та їх усереднення. Це поліпшення дало можливість отримати більш точні та повні дані про обстановку, що в свою чергу дозволяє ефективніше визначати команди руху для користувача. Чим більше перешкод з одного боку по відношенню до іншого, тим вищий коефіцієнт, тим і ближчим є перешкода.

Процес перевірки працездатності розробленого прототипу інтелектуальної мобільної системи був проведений на високотехнологічному мобільному пристрої Google Pixel 6. Цей пристрій обладнаний всіма необхідними характеристиками для успішного виконання експериментів та валідації роботи системи, зокрема, операційною системою Android, першим поколінням тензорного процесора та вбудованою 50-мегапіксельною камерою. Такий вибір конкретного пристрою для тестування був обумовлений його високим технічним рівнем та можливостями, що відповідають вимогам системи. Це забезпечило надійність та точність отриманих результатів, а також підтвердило робочу здатність системи в реальних умовах експлуатації.

У ході тестування було здійснено візуалізацію результатів роботи функції просторової орієнтації. Для цього були налаштовані компоненти, які автоматично регулюють яскравість на основі коефіцієнтів, отриманих під час обробки глибини

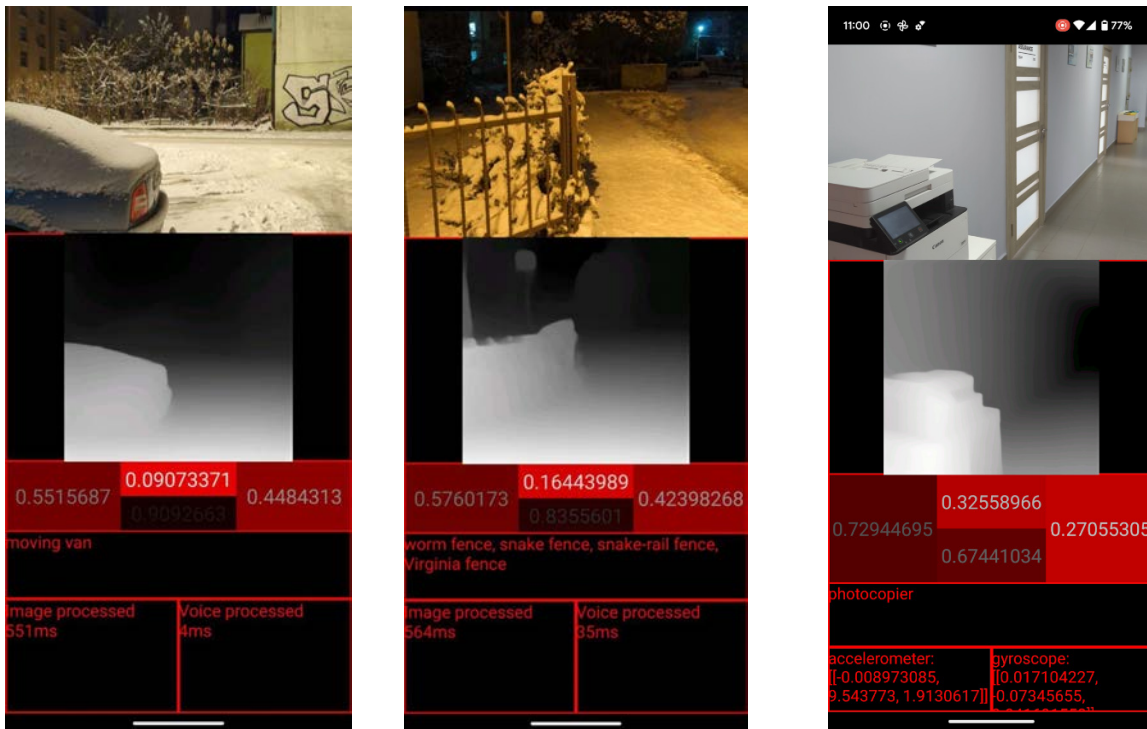


Рис. 9. Тестування розробленої інтелектуальної мобільної системи комп'ютерного зору та голосового управління

зображення. Цей підхід сприяє ефективній візуалізації та виведенню користувачеві інформації щодо просторового середовища. У процесі тестування також проводилася оцінка часу, витраченого на обчислення при обробці зображення та виконанні голосових команд. Це важливий аспект, який дозволяє визначити ефективність та продуктивність системи в реальному часі.

Під час тестування в різноманітних сценах (рис. 9) помітно, що об'єкти, що знаходяться ближче до камери, відображаються більше світла, що призводить до яскравіших зображень (коефіцієнт глибини зображення наближається до 1). З іншого боку, об'єкти, розташовані далеко, відбивають менше світла, що робить їх зображення темнішим (коефіцієнт глибини зображення наближається до 0). Шляхом прив'язки значень яскравості до відстані, система створює карту глибини, яка передає тривимірну структуру оточення та полегшує виявлення об'єктів.

У порівнянні з попередньою розробкою, удосконалення включають обчислення коефіцієнтів не лише в горизонтальній, але й в вертикальній площинах [11]. Це покращення дозволяє системі точніше оцінювати просторове оточення для ефективного уникнення перешкод. Аналіз даних з акселерометра та гіроскопа забезпечує важливу інформацію про рух та орієнтацію при-

строю. Акселерометр оцінює зміни прискорення, нахилу та обертання, надаючи комплексне розуміння просторової динаміки. Гіроскоп доповнює цей аналіз, надаючи інформацію про кутові зміни в орієнтації пристрою. Цей синергетичний підхід виконує ключову роль у виконанні функцій, таких як уникнення об'єктів, навігаційна допомога та просторова орієнтація в реальному часі для осіб із вадами зору. Додатково, система вдосконалено інтегрує функції голосового управління, які працюють у гібридному режимі, залежно від доступності Інтернет-з'єднання.

Серед ключових переваг системи слід відзначити її можливість працювати в режимі реального часу. Це означає, що система забезпечує миттєву реакцію та швидкі результати обробки без помітних затримок. Наприклад, у контексті системи просторової орієнтації на мобільному пристрої для осіб із вадами зору, можливість отримувати дані та повідомлення про оточуюче середовище в реальному часі допомагає користувачеві ефективно реагувати на перешкоди та небезпеки. Режим реального часу розширює функціональні можливості системи та гарантує її актуальність в різних сценаріях використання.

**Висновки.** Розроблено та протестовано прототип інтелектуальної системи комп'ютерного зору та голосового управління на базі смартфона для



осіб із вадами зору. Ми спроектували нашу систему так, щоб вона була легкою у використанні, портативною та доступною за ціною, забезпечуючи доступність для різних користувачів. Важливим кроком в удосконаленні є впровадження голосового управління, що ґрунтується на комбінації вбудованих та хмарних нейромережових моделей. Це значно покращує доступність та продуктивність нашої системи, забезпечуючи ще більш ефективну взаємодію з користувачем. Із застосуванням передових технологій, таких як нейронна

мережа MobiNet3 та модель MiDaS 2 Lite, ми досягли високої точності у класифікації об'єктів та визначенні глибини зображення. Реалізовано алгоритм просторового орієнтування на основі ітеративного обчислення суми значень масиву глибини зображень, продемонструвавши його ефективність у реальному часі по відношенню до рекурсивного підходу. Одержані результати тестування підтверджують високий рівень точності та швидкодії, зробивши систему практичною та вигідною для широкого кола користувачів.

#### Список літератури:

1. Kuriakose B., Shrestha R., and Sandnes F. E. DeepNAVI: A deep learning based smartphone navigation assistant for people with visual impairments. *Expert Syst. Appl.* 2023. vol. 212 no. 118720. p. 118720.
2. Li Q., Li L., Li X., Li C., Zhang L. A new navigation system for the visually impaired based on visual and audio feedback. *Journal of Healthcare Engineering*, 2019. pp. 1–9. DOI: 10.1155/2019/1670586.
3. Aadi F. Z. Proposed real-time obstacle detection system for visually impaired assistance based on deep learning. *Int. J. Adv. Trends Comput. Sci. Eng.* 2020. Vol. 9. No. 4. pp. 6649–6652.
4. Bai J., Liu D., Su G. and Fu Z. A cloud and vision-based navigation system used for blind people. *Proceedings of the 2017 International Conference on Artificial Intelligence Automation and Control Technologies*. pp. 1–6. 2017.
5. Ali H. A., Rao S. U., Ranganath S., Ashwin T. S. and Reddy G. R. M. A Google Glass Based Real-Time Scene Analysis for the Visually Impaired. in *IEEE Access*. 2021. vol. 9. pp. 166351–166369. doi: 10.1109/ACCESS.2021.3135024.
6. Mukhiddinov M. and Cho J. Smart glass system using deep learning for the blind and visually impaired. *Electronics (Basel)*. 2021. vol. 10. no. 22. p. 2756.
7. Lin B.-S., Lee C.-C., and Chiang P.-Y. Simple smartphone-based guiding system for visually impaired people. *Sensors (Basel)*. 2017. vol. 17. no. 6, p. 1371.
8. See A. R., Sasing B. G., and Advincula W. D. A smartphone-based mobility assistant using depth imaging for visually impaired and blind. *Appl. Sci. (Basel)*. 2022. vol. 12. no. 6. p. 2802.
9. Bai J., Liu Z., Lin Y., Li Y., Lian S., and Liu D., Wearable travel aid for environment perception and navigation of visually impaired people. *Electronics (Basel)*. 2019. vol. 8. no. 6, p. 697.
10. Ganz A., Schafer J. M., Tao Y., Wilson C. and Robertson M. PERCEPT-II: Smartphone based indoor navigation system for the blind. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Chicago. IL. USA. 2014. pp. 3662–3665. doi: 10.1109/EMBC.2014.6944417.
11. Beshley M., Volodymyr P., Beshley H., and Gregus Jr M. A smartphone-based computer vision assistance system with neural network depth estimation for the visually impaired, in *Artificial Intelligence and Soft Computing, Cham: Springer Nature Switzerland*, 2023, pp. 26–36.

#### **Beshley M.I., Pastukh V.A., Beshley H.V., Andrushchak V.S., Klymash M.M. DEVELOPMENT OF AN INTELLIGENT MOBILE SYSTEM OF COMPUTER VISION AND VOICE CONTROL FOR VISUAL IMPAIRED PEOPLE**

*In this article, a smartphone-based intelligent mobile system of computer vision and voice control is developed, specially adapted for visually impaired users. The developed system operates in real time and provides the possibility of using both online and offline modes, expanding the functionality of modern solutions for blind and visually impaired persons. An important aspect is the implementation of a voice control module that uses a combination of embedded and cloud-based APIs of neural network models, which significantly increases the system's accessibility and efficiency. We used advanced technologies such as the MobiNet3 neural network and the MiDaS 2 Lite model for object classification and image depth estimation. We improved traditional visual recognition methods and implemented an iterative asynchronous depth calculation algorithm, which ensured the system's outstanding performance. The test results confirm that the processing time for our interactive algorithm is in the range of 600  $\mu$ s to 2 ms, which is an impressive improvement over traditional recursive methods. We have introduced innovations to the system, including the calculation of image depth coefficients in both horizontal and vertical planes, which represents a significant step of improvement over our previous version. These changes contribute to a more accurate determination of the system's spatial environment, improving its ability to effectively avoid obstacles. In addition, we have implemented the analysis of accelerometer and gyroscope data, which provides important information about the movement and orientation of the device. The accelerometer assesses changes in acceleration, tilt, and rotation to provide a comprehensive view of spatial dynamics. The gyroscope complements this analysis by providing information about angular changes in the device's orientation. These integrated enhancements play a crucial role in functions such as object avoidance, voice control, navigation assistance, and spatial orientation for the visually impaired. The testing proved that the combination of voice control and image recognition allows our system to work in real time and provide users with up-to-date information.*

**Key words:** Smartphone, navigation assistance, computer vision, voice control, intelligent mobile system, neural network, mobility, iterative asynchronous algorithm, image depth.