Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0825U003242

Особливі позначки: відкрита

Дата реєстрації: 31-07-2025

Статус: Наказ про видачу диплома

Реквізити наказу МОН / наказу закладу: Наказ ВНТУ № 315 від 26 вересня 2025р.

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Ся Гуаньсян ..
2. Guanxiang Xia

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 122

Назва наукової спеціальності: Комп’ютерні науки

Галузь / галузі знань: інформаційні технології

Освітньо-наукова програма зі спеціальності: комп'ютерні науки

Дата захисту: 11-09-2025

Спеціальність за освітою: Техніка промислового дизайну

Місце роботи здобувача:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): PhD 10061

Повне найменування юридичної особи: Вінницький національний технічний університет

Код за ЄДРПОУ: 02070693

Місцезнаходження: вул. Хмельницьке шосе, буд. 95, Вінниця, Вінницький р-н., 21021, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Вінницький національний технічний університет

Код за ЄДРПОУ: 02070693

Місцезнаходження: вул. Хмельницьке шосе, буд. 95, Вінниця, Вінницький р-н., 21021, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

V. Відомості про дисертацію

Мова дисертації: Англійська

Коди тематичних рубрик: 28.23.15

Тема дисертації:

1. Моделі та методи підвищення ефективності розпізнавання текстових областей у потоковому відео в режимі реального часу
2. Models and Methods for Improving the Efficiency of Text Area Recognition in Real-Time Video Streams

Реферат:

1. Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 – «Комп'ютерні науки». – Вінницький національний технічний університет, Вінниця, 2025. Стрімка цифровізація сучасного суспільства та бізнес-процесів зумовлює потребу в ефективних рішеннях для автоматизованого вилучення інформації з візуальних джерел. Одним із таких викликів є розпізнавання тексту та структурних елементів документів, що представлені у відеопотоках у реальному часі. Це завдання перебуває на перетині комп'ютерного зору, теорії розпізнавання образів, обробки природної мови (NLP) та технологій глибинного навчання й набуває все більшої актуальності у низці галузей, зокрема: - Оцифрування архівів – обробка історичних або юридичних документів у форматі відеозаписів (включаючи рукописи та плівкові матеріали) з метою структурованого вилучення тексту; - Системи контролю документів – автоматична перевірка паспортів і віз за допомогою відеоспостереження в умовах обмеженого часу; - Освіта та електронне навчання – транскрибування візуального матеріалу під час лекцій чи вебінарів, особливо у форматі дистанційного навчання; -

Промислова автоматизація – розпізнавання серійних номерів, етикеток і пакувальної інформації на відео в логістиці та виробництві. На відміну від традиційного OCR, розпізнавання документів у відео супроводжується додатковими складнощами: змінним освітленням, розмитістю руху, частковими перекриттями об'єктів, а також багатомовністю. Це потребує застосування складних моделей, здатних до оперативного аналізу в неідеальних умовах. У вступі визначено мету, об'єкт, предмет і методи дослідження, обґрунтовано наукову новизну та практичну значущість, наведено структуру дисертаційної роботи. Розділ 1 присвячено аналізу архітектури сучасних систем розпізнавання документів, зосереджуючись на мобільних рішеннях і викликах, пов'язаних із обробкою зображень документів в обмежених умовах. Розділ 2 пропонує математичну модель системи оптичного розпізнавання об'єктів у відео з модулями інтеграції результатів і завершення процесу, формалізує відповідне завдання розпізнавання. Розділ 3 описує алгоритм об'єднання результатів розпізнавання символьних об'єктів із урахуванням альтернативних класифікацій символів. Представлено порівняльний аналіз з алгоритмом ROVER, а також методику синтезу даних для навчання й створення відкритих наборів документів для ідентифікації. Розділ 4 містить новий метод зупинки розпізнавання, що базується на пороговому скороченні очікуваної відстані між поточним та наступним інтегрованим результатом. Задачу формалізовано як монотонну зупинку, і доведено перевагу методу над традиційними підходами. Наукова новизна: - Розроблено динамічну модель розпізнавання об'єктів у відео в реальному часі. На відміну від статичних методів, вона забезпечує акумулювання результатів між кадрами та визначення оптимального моменту зупинки, що дало змогу підвищити точність на 9%. - Запропоновано ймовірнісний метод інтеграції гіпотез на рівні символів, який підтримує поетапне уточнення результатів і перевершує ROVER на 7% за точністю. - Розроблено алгоритм зупинки, що базується на оцінці очікуваного виграшу, формалізований як задача монотонної зупинки. Це дозволило зменшити середню кількість оброблених кадрів на 18% без втрати якості, що критично для мобільних пристроїв. Практичне значення: - Запропоновані методи підвищують надійність оцифрування документів і розпізнавання в умовах динамічного відео. - Продемонстровано успішну адаптацію ROVER для випадків покадрового спостереження за об'єктом. - Покращена точність аналізу паспортних та ідентифікаційних відео підтримує автоматизацію прикордонного контролю. - Метод зупинки оптимізує використання ресурсів без втрати точності, особливо на пристроях з обмеженою обчислювальною потужністю. - Інтеграційні рішення підвищують ефективність OCR-систем у мобільних та відеоспостережних сценаріях. - Перевірка на реальних даних і сумісність із Tesseract підтверджують придатність для мобільних і хмарних застосунків. - Виявлений вплив якості вхідних даних на вибір стратегії об'єднання дозволяє адаптувати системи до конкретних умов. - Формалізація задачі зупинки відкриває нові можливості оптимізації відеоаналітики в реальному часі. - Експерименти на наборі MIDV-500 підтверджують ефективність методів для систем верифікації документів, що створює передумови для комерційного та державного впровадження. - Запропоновані підходи можуть бути адаптовані до інших задач комп'ютерного зору в реальному часі, зокрема розпізнавання об'єктів і динамічного аналізу сцен.

2. Dissertation for the degree of Philosophy Doctor in the specialty 122 "Computer Science". – Vinnytsia National Technical University, Vinnytsia, 2025. The rapid digitalisation of modern society and business processes demands effective solutions for the automated extraction of information from visual sources. One such challenge is the recognition of text and structure in documents embedded within real-time video streams. This task lies at the intersection of computer vision, pattern recognition, NLP, and deep learning technologies, and is becoming increasingly relevant in multiple domains, including: - Archival digitisation – processing historical or legal video-recorded documents, including handwritten materials or film archives, for structured text extraction; - Document control systems – automatic passport and visa validation at borders using video surveillance under time constraints; - Education and e-learning – transcription of visual materials during lectures or webinars, especially in remote learning environments; - Industrial automation – recognising serial numbers, labels, and packaging data via video footage in logistics and manufacturing. Compared to traditional OCR, video-based document recognition introduces additional difficulties such as variable lighting, motion blur, occlusions, and multilingual texts. This requires sophisticated models capable of real-time analysis in non-ideal conditions. The introduction defines the research objectives, object, subject, and methods, presents the novelty and practical importance, and outlines the

dissertation's structure. Chapter 1 surveys the architecture of modern document recognition systems, focusing on mobile workflows and the challenges of processing document photos under constrained conditions. Chapter 2 introduces a mathematical model of an optical object recognition system in video, including modules for result integration and termination, and formulates the recognition task accordingly. Chapter 3 proposes an algorithm for combining recognition results of string objects in video, incorporating alternative classifications for each character. It includes a comparative analysis with the ROVER algorithm and discusses a method for generating synthetic training data and building open datasets of identity documents. Chapter 4 presents a novel stopping method based on threshold reduction of expected distance between current and upcoming integrated results. This chapter formalises the stopping task as a monotonic problem and shows experimental superiority over traditional fixed-frame and heuristic methods. Scientific Novelty: - A dynamic model of real-time video-based object recognition is developed. Unlike static methods, it accumulates recognition results across frames and determines optimal stopping points. Experiments show up to 9% improvement in accuracy over frame-by-frame approaches. - A probabilistic method of integrating character-level recognition hypotheses is introduced. It enables iterative result refinement and demonstrates a 7% accuracy gain over the ROVER benchmark. - A stopping algorithm based on expected gain estimation is proposed, framed as a monotonic stopping task. It reduces average frame processing by 18% while maintaining result quality, crucial for mobile applications. Practical Significance: - Developed methods enhance the reliability of document digitisation and real-time recognition tasks in dynamic video environments. - The research demonstrates successful adaptation of ROVER to frame-by-frame object observations, expanding its usability. - Improved accuracy in passport and ID video analysis directly supports automation in border control and identity verification. - The stopping method improves system efficiency by reducing resource use without sacrificing accuracy, especially on devices with limited computing power. - Integration techniques increase performance of OCR systems processing video from surveillance or mobile sources. - Testing on real data and compatibility with the open-source Tesseract OCR engine confirm applicability in mobile and cloud-based applications. - The influence of input data quality on integration strategy selection enables system customisation for different use cases. - The formalisation of the stopping problem offers a new optimisation path for real-time video analytics. - Evaluation on the MIDV-500 dataset confirms method effectiveness for identity document analysis in verification systems, with strong potential for commercial and governmental deployment. - The proposed methods are adaptable to other real-time computer vision tasks, including dynamic object recognition and scene analysis.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Інформаційні та комунікаційні технології

Стратегічний пріоритетний напрям інноваційної діяльності: Освоєння нових технологій транспортування енергії, впровадження енергоефективних, ресурсозберігаючих технологій, освоєння альтернативних джерел енергії

Підсумки дослідження: Теоретичне узагальнення і вирішення важливої наукової проблеми

Публікації:

- 1. X. Guanxiang and V. Kovtun, "The model of the system for objects recognition in the real-time video stream", Computer systems and information technologies, no. 4, pp. 157–162, Dec. 26, 2024. dooi: 10.31891/csit-2024-4-19.
- 2. X. Guanxiang and V. Kovtun, "Methods for implementation of string object recognition results in real-time video streams," Measuring And Computing Devices In Technological Processes, no. 4. Khmelnytskyi National University, pp. 338–347, Nov. 28, 2024. doi: 10.31891/2219-9365-2024-80-41.
- 3. X. Guanxiang and V. Kovtun, "The method for stopping the string object recognition process in a real-time video stream," Measuring And Computing Devices In Technological Processes, no. 3. Khmelnytskyi National University, pp. 243–250, Aug. 29, 2024. doi: 10.31891/2219-9365-2024-79-32.

- 4. X. Guanxiang, “A Study of the Implementation Techniques and Processes of Live Video Courses in Online Teaching and Learning,” Journal of Online Learning, no. 2, pp. 3224–3225, Jun. 29, 2022. doi: 10.12255/j.issn.1672-6677.2021.11.1603.
- 5. X. Guanxiang, “Post-Processing Software Application Techniques in the Audio-visual Recording of Hunan Chinese Dialect Survey for the Language Protection Project - Taking Adobe Premiere Pro as an Example,” Journal of Audio-visual Studies, no. 5, pp. 162–166, Apr. 15, 2022. [Online]. Available: https://d.wanfangdata.com.cn/periodical/Ch9QZXJpb2RpY2FsQ0hJTmV3UzIwMjUwMTA0MTcwMjI2EhNzbXNqLX
- 6. X. Guanxiang, “Preliminary quality control in audio-visual recording of Hunan Chinese dialect investigation of China language resources protection project,” Journal of Language Protection, no. 3, pp. 241–243, Mar. 15, 2022. [Online]. Available: https://d.wanfangdata.com.cn/periodical/Ch9QZXJpb2RpY2FsQ0hJTmV3UzIwMjUwMTA0MTcwMjI2EhpRS0JKQk
- 7. X. Guanxiang, “Video and audio recording technology in Hunan Chinese dialect investigation of China language resource protection project,” Journal of Media Technology, no. 8, pp. 39–45, Aug. 02, 2019. doi: 10.3969/j.issn.2096-0751.2019.02.009.
- 8. X. Guanxiang, “Analysis of digital film and television special effects production technology,” Journal of Modern Information, no. 3, pp. 48–50, Jun. 16, 2017. doi: 10.3969/j.issn.1674-6708.2017.06.040.

Наукова (науково-технічна) продукція: пристрої; технології; методи, теорії, гіпотези; програмні продукти, програмно-технологічна документація

Соціально-економічна спрямованість: забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Впровадження не планується

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Ковтун В'ячеслав Васильович

2. Vyacheslav V. Kovtun

Кваліфікація: д. т. н., професор, 05.13.06

Ідентифікатор ORCID ID: 0000-0002-7624-7072

Додаткова інформація:

Повне найменування юридичної особи: Вінницький національний технічний університет

Код за ЄДРПОУ: 02070693

Місцезнаходження: вул. Хмельницьке шосе, буд. 95, Вінниця, Вінницький р-н., 21021, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Олійник Юрій Олександрович

2. Yurii Oliinyk

Кваліфікація: к. т. н., 05.13.06

Ідентифікатор ORCID ID: 0000-0002-7408-4927

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Павлова Ольга Олександрівна

2. Olha Pavlova

Кваліфікація: д.філософ, доц., 122

Ідентифікатор ORCID ID: 0000-0001-7019-0354

Додаткова інформація:

Повне найменування юридичної особи: Хмельницький національний університет

Код за ЄДРПОУ: 02071234

Місцезнаходження: вул. Інститутська, буд. 11, Хмельницький, Хмельницький р-н., 29016, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Мокін Віталій Борисович

2. Vitalii Mokin

Кваліфікація: д. т. н., професор, 01.05.02

Ідентифікатор ORCID ID: 0000-0003-1946-0202

Додаткова інформація:

Повне найменування юридичної особи: Вінницький національний технічний університет

Код за ЄДРПОУ: 02070693

Місцезнаходження: вул. Хмельницьке шосе, буд. 95, Вінниця, Вінницький р-н., 21021, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Маслій Роман Васильович
2. Roman V. Masliy

Кваліфікація: к. т. н., доц., 05.13.06

Ідентифікатор ORCID ID: 0000-0003-3021-4328

Додаткова інформація:

Повне найменування юридичної особи: Вінницький національний технічний університет

Код за ЄДРПОУ: 02070693

Місцезнаходження: вул. Хмельницьке шосе, буд. 95, Вінниця, Вінницький р-н., 21021, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради	Квєтний Роман Наумович
Власне Прізвище Ім'я По-батькові головуючого на засіданні	Квєтний Роман Наумович
Відповідальний за підготовку облікових документів	Іванчук Ярослав Володимирович
Реєстратор	УкрІНТЕІ

Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності

Юрченко Тетяна Анатоліївна