

РОЗРОБКА СИСТЕМИ РОЗУМНОГО ДОСТУПУ ДО ВИРОБНИЧОГО ПРИМІЩЕННЯ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ КОМП'ЮТЕРНОГО ЗОРУ

М.А. Водяницький

Харківський національний університет радіоелектроніки

Україна, 61166, Харків, пр. Науки 14

E-mail: mykhailo.vodianytskyi@nure.ua

Анотація: у статті розглядаються сучасні системи розумного доступу до приміщення, що базуються на технології комп'ютерного зору. Проаналізовано основні положення теорії розпізнавання образів, на основі теоретичних відомостей описані методи розпізнавання обличчя та способи їх реалізації на базі ESP32-CAM.

Ключові слова: комп'ютерний зір, автоматизація, доступ, ESP32-CAM, розпізнавання лиця.

DEVELOPMENT OF A SYSTEM FOR SMART ACCESS TO PRODUCTION FACILITIES USING COMPUTER VISION TECHNOLOGIES

M. Vodianytskyi

Kharkiv National University of Radio Electronics

Ukraine, 61166, Kharkiv, Nauky av., 14

E-mail: mykhailo.vodianytskyi@nure.ua

Annotation: In the article deals with modern systems of smart access to premises based on computer vision technology. The main provisions of the theory of pattern recognition are analysed, and based on theoretical information, face recognition methods and methods of their implementation on the basis of Arduino ESP32-CAM are described.

Key words: computer vision, automation, access, ESP32-CAM, face recognition.

АКТУАЛЬНІСТЬ РОБОТИ. Безпека виробничих приміщень завжди була актуальною проблемою. З розвитком технологій, коли мега корпорації витрачають мільярди доларів на побудову виробництв, найму працівників, створення технологій – одне виробниче приміщення може зберігати таємниці на мільйони доларів. Тож важливою задачею є охорона таких місць. Одним з методів може бути розробка системи розумного доступу на базі комп'ютерного зору. Метою роботи є підвищення безпеки доступу на підприємстві, мінімізуючи потребу у фізичному контролі шляхом розробки системи розумного доступу на основі технологій комп'ютерного зору. Задачею роботи є розробка такого рішення на базі ESP32.

МАТЕРІАЛ І РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ. Розвиток мікроелектроніки у двадцять першому сторіччі, у сфері відеокарт, дав великий поштовх у розвитку систем комп'ютерного зору. Комп'ютерний зір (computer vision) – теорія та технологія створення машин, які можуть проводити виявлення, відстежування та визначення об'єктів. Як наукова дисципліна комп'ютерний зір належить до теорії та технології створення штучних систем, які отримують інформацію у вигляді зображень. Головним користувачем даної теорії у наш час є нейронні мережі. Збільшення потужності відеокарт дозволило навчати мережі швидше та на більшій кількості вхідних даних. Завдяки цьому нейронні мережі можуть зчитувати інформацію з зображень за доли секунди, оброблювати її та видавати відповідь. Наприклад можна надіслати мережі фото листку з задачею по фізиці і мережа знайде рішення. Але до буму їх розвитку, інформація з зображення зчитувалась іншими технологіями, які підпорядковуються теорії розпізнавання образів.

Теорія розпізнавання образів – розділ кібернетики, що розвиває теоретичні основи й методи класифікації і ідентифікації предметів, явищ, процесів, сигналів, ситуацій і об'єктів, які характеризуються скінченним набором властивостей і ознак. Розпізнавання образів включається в ширшу наукову дисципліну – теорію машинного навчання, метою якої є розробка методів побудови алгоритмів, що здатні навчатися. Прикладом задачі, що потребує такий підхід, може служити система безпілотного керування автомобілем від компанії Tesla. Ця система повинна зчитувати велику кількість зовнішньої інформації та приймати миттєве рішення: колір світла світлофору на перехресті, дорожні знаки, інші машини, стан дорожнього покриття, погодні явища, стан автомобілю. Наш мозок виконує такі задачі автоматично, шляхом спрощення вхідних даних та зведення їх до чогось подібного, тобто наш мозок автоматично проводить класифікацію явищ навколишнього світу на групу схожих явищ. Комплекс апаратних і програмних рішень, здатних симулювати цей процес, називається системою розпізнавання образів.

Система розпізнавання образів – це комплекс апаратних і програмних засобів, призначених для ідентифікації або класифікації об'єктів (образів) на основі їх характеристик, здатний моделювати розумові процеси, властиві під час прийняття рішень із метою виявлення аналогій серед досліджуваних об'єктів. Система має велику кількість методів роботи із даними, тому для більшої зручності вони класифікуються за спільними ознаками. Самі системи розпізнавання поділяються на складні та прості. Прості характеризуються єдиною фізичною природою ознак. Наприклад тільки геометричні розміри для ідентифікації об'єкту на екрані, або тільки колір певний звук. Складні системи мають неоднорідні фізичні ознаки – автопілот в автомобілі.

Доволі довгий час, системи розпізнавання лише умовно були симуляціями роботи мозку. Головною причиною було складність програмування великих програм та мала обчислювальна потужність систем. У разі роботи із письмовими даними системи частіше за все будувалися за структурними правилами – використання спеціальних граматики, що породжують мови, де кожне речення описує об'єкти конкретного класу. У випадку необхідності класифікувати зображення або реальні об'єкти через комп'ютерний зір, використовували детерміновані методи побудови систем, що базуються на геометричних методах. Цей підхід передбачав створення еталонних класів з якими порівнювались вимірюванні відстані на об'єкті.

Для спрощення програмування детермінованих систем, зазвичай використовували математичні методи фільтрації, що дозволяли виділити на зображенні потрібні ознаки, без обробки зайвого масиву даних. Найбільш простими та ефективними методами, що використовуються і у сучасних системах є бінаризація по порозу, фільтри Гаусса та Габора, оператор Кенні для визначення меж та контурів.

Але, як вже було зазначено, були доволі далекі від симуляції роботи мозку людини, вони не виконували необхідну умову – спрощення вхідних даних та зведення їх до чогось подібного в автоматичному режимі, без використання додаткової перед обробки даних. Довгий час ця проблема не могла бути вирішена через недостатню обчислювальну потужність комп'ютерів. Але справжня революція сталася у 2012 році, коли на міжнародних змагання ILSVRS, присвячених комп'ютерному зору та штучному інтелекту, була представлена система AlexNet. Це була згортована нейронна мережа (CNN), що складалась з восьми шарів, п'яти згорткових шарів (convolutional layers) та трьох повноз'язних шарів (fully connected layers).

Нейронні мережі, зокрема CNN, є революційними інструментами для розпізнавання і обробки зображень, завдяки своїй здатності самостійно виділяти важливі ознаки. CNN мережа складається з трьох типів шарів: перший шар вхідний, другий шар прихований і вихідний шар. Згорткові шари витягують особливості із вхідних даних, вони використовують фільтри, які «ковзають» по зображенню, виконуючи операцію згортання. У результаті формуються карти

ознак (feature maps), які показують, де виявляються певні характеристики. Після згорткового шару данні потрапляють до прихованого шару, або шару об'єднання. Його головна задача зменшити розмірність зображення, для того аби зменшити кількість навчальних параметрів, що знижує обчислювальні витрати. На завершальному етапі CNN обробляє дані через повнозв'язні шари, де кожен нейрон з'єднаний з усіма нейронами попереднього шару. Це об'єднує інформацію, створюючи узагальнену картину, яка може бути використана для класифікації.

CNN можуть виділяти ознаки без втручання людини, що робить їх дуже ефективними. Обробляючи локальні області зображення, зменшують обсяг оброблюваної інформації, дозволяючи мережі працювати швидше. Мають чудові показники точності, оскільки вони налаштовуються на розпізнавання різноманітних ознак об'єкта.

У топі помилок AlexNet досягла 15,3% (проти 26% у найближчого конкурента). Це стало першою демонстрацією того, як глибокі нейронні мережі можуть перевершувати інші алгоритми в завданнях класифікації зображень. Цей результат ставку початком «буму» нейромереж, і в наш час кожна нова система розпізнавання будується на їх базі.

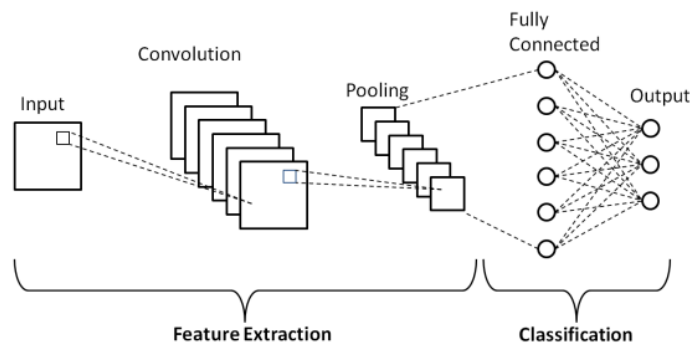


Рисунок 1 – Схема нейронної мережі типу CNN

Швидкий розвиток мікроелектроніки дав поштовх не лише у сфері нейромереж. Ми отримали таку систему як Arduino. Arduino – це відкрита апаратно-програмна платформа, розроблена для швидкого створення електронних пристроїв. Вона популярна серед аматорів, інженерів і розробників завдяки простоті використання та великій спільноті. На базі цих мікрочіпів створюють системи розумного дому, системи відеоспостереження та інше. Одна з можливостей це спільне використання з модулем ESP32-CAM. Цей модуль представляє собою плату із чіпом ESP32S та камерою, який підтримує функції комп'ютерного зору, включно з розпізнаванням облич. Ця функція надається бібліотекою ESP-WHO, яка розроблена Espressif спеціально для комп'ютерного зору. У бібліотеці доступні приклади коду для розпізнавання облич, які можна використовувати з ESP32-CAM.

Камера ESP32-CAM робить знімок у режимі реального часу. Спочатку застосовується алгоритм виявлення облич, наприклад, Haar Cascade або MTCNN. На зображенні виділяються області, які містять обличчя. Після виявлення обличчя мережа витягує ключові особливості (наприклад, відстань між очима, форму носа) і перетворює ці дані в числовий вектор. Вектор обличчя порівнюється зі збереженими в базі даних векторами, через алгоритм k-найближчих сусідів або відстань косинусної близькості.



Рисунок 2 – Розпізнавання обличчя бібліотекою ESP-WHO

Але такий підхід доволі обмежений у своїх можливостях. Потужності чіпу не дають змоги використовувати великі масиви даних для порівняння, не дають змоги використовувати нейронної мережі. Вирішенням цієї проблеми є використання функції ESP32 WebServer, що надає змогу передавати відео або фото на окремий сервер, де вже відбувається обробка. Створити таку систему можливо власноруч, або за допомогою готових програмних рішень типу Home Assistant. Home Assistant запускає сервер на Python і використовує базу даних для зберігання інформації, що працює на локальному пристрої, це може бути Raspberry Pi, окремий сервер на Linux чи віртуальна машина. Сервер зв'язується з компонентами розумного дому (датчики руху, температури, світла), або IP-камерами що передають відео формату RTSP, якою є ESP32. Сам сервер працює як центральний вузол, що зв'язую усі компоненти і застосовує для пристроїв сценарії роботи. Тож у випадку коли потрібно ідентифікувати обличчя використовуються сторонні інтеграції, наприклад Microsoft Azure Face API. Сервер передає зображення для ідентифікації через Azure API на сервери Microsoft, які повертають результат – на зображенні ідентифіковано людину. Якщо обличчя відоме сервер Home Assistant може, наприклад, надати команду відкрити двері, або навпаки – обличчя не відоме, відправити сповіщення.

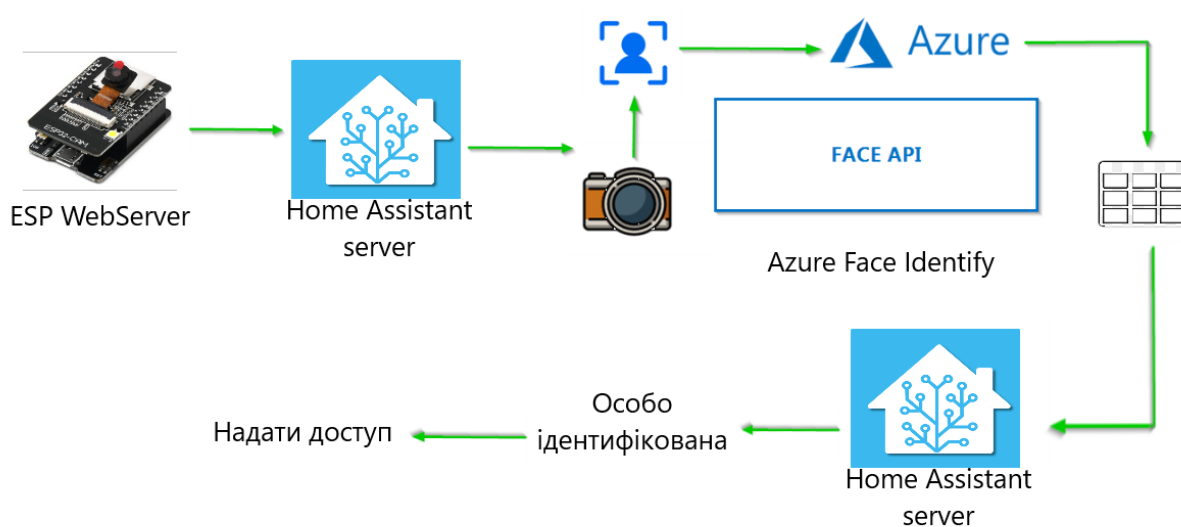


Рисунок 3 – Загальний принцип системи ідентифікації на базі Home Assistant

ВИСНОВКИ. У підсумку: був проведений аналіз положень теорії розпізнавання образів, описані базові типи систем розпізнавання та наданий стислий опис їх роботи із основними особливостями, описано принци роботи мережі CNN, її переваги та вплив на подальший розвиток галузі, описані загальні методи роботи з ESP32-CAM та методи реалізації ідентифікації особи сторонніми сервісами.

ЛІТЕРАТУРА.

1. What is computer vision? // Сайт IBM.com URL: <https://www.ibm.com/topics/computer-vision>.
2. Основи теорії розпізнавання образів, навчальний посібник у двох частинах, А.С. Довбиш, І.В. Шелехов // Сайт НТУ Національний транспортний університет URL: <http://kist.ntu.edu.ua/textPhD/tro2.pdf>.
3. Теорія розпізнавання образів, навчальний посібник, В.В. Гавриленко, Г.Ф. Іванченко, Є.Г. Шевченко // Сайт studfile URL: <https://studfile.net/preview/5607072/>.
4. What is Thresholding in Image Processing? A Guide. // Сайт Roboflow Blog URL: <https://blog.roboflow.com/image-thresholding/>.
5. Face Recognition Techniques – A Review // Сайт Ajay Kumar Garg Engineering College URL: <https://www.akgec.ac.in/wp-content/uploads/2019/06/2-Face-Recognition-Techniques-Bhaskar-Gupta.pdf>.
6. Face Recognition for Beginners // Сайт Medium URL: <https://towardsdatascience.com/face-recognition-for-beginners-a7a9bd5eb5c2>.
7. Convolution Neural Network (CNN) // Сайт Analytics Vidhya URL: <https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-networkcnn/>.
8. Azure Face API: Face Detection & Recognition // Сайт Medium URL: <https://medium.com/codex/azure-face-api-face-detection-recognition-b2b954e3e939>
9. Build ESP32-CAM Projects // Сайт MyArduino URL: <https://myarduino.dk/ESP32-CAM.pdf>
10. Yevsieiv, V., & Demska, N. (2024). *A Model of Using Computer Vision to Monitor the Environment of a Collaborative Manipulator Robot* (Doctoral dissertation).
11. Nevliudov, I., Yevsieiev, V., Maksymova, S., Demska, N., Starodubcev, N., & Klymenko, O. (2023, September). Monitoring System Development for Equipment Upgrade for IIoT. In *2023 IEEE 5th International Conference on Modern Electrical and Energy System (MEES)* (pp. 1-5). IEEE.
12. Nevliudov, I. S., Yevsieiev, V. V., Demska, N. P., & Kostrova, H. Y. (2023). Development and Improvement of the Design of a Lightweight Mobile Robot Manipulator Using Generative Design.
13. Yevsieiev, V., & Demska, N. (2023). *Application of Generative Design Methods for Improving Manipulator Designs for Mobile Robots* (Doctoral dissertation, Черкаського національного університету імені Богдана Хмельницького).
14. Nevliudov, I., Yevsieiev, V., Maksymova, S., Demska, N., Kolesnyk, K., & Miliutina, O. (2022, September). Object Recognition for a Humanoid Robot Based on a Microcontroller. In *2022 IEEE XVIII International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH)* (pp. 61-64). IEEE.
15. Євсєєв, В. В., Демська, Н. П., & Олександров, Ю. М. (2022). *Моделювання виробничої лінії SMT-монтажу в кібер-фізичних виробничих системах* (Doctoral dissertation, Кременчуцький національний університет імені Михайла Остроградського).

Науковий керівник: Гурін Дмитро Валерійович, старший викладач кафедри КІТАР, Харківський національний університет радіоелектроніки