

КРИВЕНЧУК Ю. П.

Національний університет "Львівська політехніка"

<https://orcid.org/0000-0002-2504-5833>e-mail: Yurii.P.Kryvenchuk@lpnu.ua

МАРКО О. М.

Національний університет "Львівська політехніка"

e-mail: oleksandramarko@gmail.com

СИСТЕМА РОЗПІЗНАВАННЯ ТА ТРЕКІНГУ ОБ'ЄКТІВ НА ВІДЕО

В роботі проаналізовано та наведено результати дослідження теми розпізнавання об'єктів на відео, а також здійснений опис запропонованих раніше методологій, з метою виявлення параметрів, які можна вдосконалити та в подальшому застосувати оптимізовані методи для створення та поширення аналогічних систем в проекті DRINMIX (автоматизована система обслуговування клієнтів), а також у наближених сферах. Виділено та описано такі основні етапи, як розпізнавання об'єкту, відслідковування об'єкту, класифікація об'єкту. Аналіз результатів підтвердив актуальність тематики та доцільність створення системи розпізнавання та трекінгу об'єктів на відео, на основі проробленої роботи зроблений висновок, що найбільш ефективним інструментом для виконання цієї роботи стане використання методів глибинного навчання.

Ключові слова: об'єкт, відео, розпізнавання об'єктів, трекінг об'єктів, згорткова нейронна мережа.

Yurii KRYVENCHUK, Oleksandra MARKO

Lviv Polytechnic National University

CREATION OF OBJECT RECOGNITION AND TRACKING SYSTEM

In modern society, video recording systems have become widespread, allowing them to recognize objects, their absence, and changes in position. In the vast majority of intelligent systems for video monitoring and determining objects through their image, such objects as human faces, printed publications, and state registration numbers are taken into consideration. At the same time, the circumstances for accepting an image are quite strict, since there is a limit on illumination, background, location relative to the lens, and so on. All this significantly facilitates the joint work of a person with a computer, and creates the prerequisites for using all kinds of systems of artificial origin of the mind. The primary goal in the development of a method and software for the automatic design of a video surveillance system is the purpose of recognize an object whose image is transmitted through the camera. Since the image of any object depends on many moments of its direction about the video camera, illumination, characteristics of the recorder, static and dynamic characteristics of the object, it is rather difficult to arrange and present a picture in the guise of a specific mathematical model. As a result, the methods for implementing a computer representation are significantly dependent on the goals being solved and are occasionally inferior to generalization. As a result, the bulk of these methods is considered non-linear. This affects the need to accumulate the calculation of computer power and the difficulty of algorithms for work acquired through the technical channels of the resulting image. In addition to technical indicators that distort the quality of a digital image, several external moments are considered, and these are: lighting around the scene, moving objects within it, etc. As a result, to obtain the best accuracy of character recognition, it is necessary to take into account all the details. The aim of this work is to study the topic of object recognition in images and video, in order to further use the results of work in creating a system for object recognition in images and video.

Keywords: image, video, object, object detection, object tracking, convolutional neural network.

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

Дивлячись і бачачи, ми пізнаємо світ, у якому живемо. Це середовище оточує нас нескінченними типами предметів і вражень. Інакше кажучи, зір та можливість бачити – це засіб для розуміння навколишнього світу. Як саме працює зорова система, залишається загадкою, яку потрібно розгадати, навіть незважаючи на те, що фізіологи досліджують це явище протягом десятиліть. Говорячи про зір, замінивши живу істоту обчислювальним пристроєм, ми маємо широкі і абстрактні вираження комп'ютерного зору. Його можна узагальнити як процес, коли комп'ютери аналізують цифрові зображення або відео та отримують від цього розуміння світу що їх оточує.

Відстеження об'єктів — це область комп'ютерного зору, яка має багато практичних застосувань, таких як відеоспостереження, взаємодія людини та комп'ютера та навігація роботів. Це добре вивчена проблема і в багатьох випадках складна для вирішення. Проблема відстеження об'єкта у відео можна узагальнити як завдання пошуку положення об'єкта в кожному вилученому кадрі з відеопотоку. Можливість відстежувати об'єкт у відео залежить від багатьох факторів, таких як знання про цільовий об'єкт, тип параметрів, що відстежуються, і тип відео, що показує об'єкт.

Відстеження об'єктів є важливою частиною співпраці людини та комп'ютера в безперервному середовищі, у сенсі, що дозволяє комп'ютеру отримати кращу модель реального світу. Наприклад, у сфері застосування автономних транспортних засобів, де людина не може досить точно та швидко передати стан навколишнього середовища з огляду на вимоги агента. Широка область застосування відображає важливість надійного, точного та ефективного відстеження об'єктів. Існує кілька важливих кроків для ефективного відстеження об'єкта, включаючи вибір моделі для представлення об'єкта та методу відстеження об'єкта, який підходить для виконання завдання.

Системи виявлення об'єктів розпізнають наявність або відсутність об'єктів у певних сценах і точках огляду камер. Завдяки, відслідковуванню об'єктів на відео, в подальшому 'розумні' системи зможуть більш чітко передбачувати ті чи інші потенційні зміни в поведінці об'єктів, що дозволить покращити якість обслуговування людей вищезгаданими системами. Наприклад, за допомогою виявлення об'єктів на відео чи зображенні можна відслідковувати ціль та визначити її точне місце знаходження в просторі, ця інформація дозволить роботизованим системам якісніше, швидше та точніше обслуговувати своїх потенційних клієнтів. Також, варто зазначити, що за допомогою системи визначення та трекінгу об'єктів на відео можна проаналізувати рухи людей, та застосувати такий аналіз в побудові патернів рухів для вище згаданих роботизованих систем, а також можна використовувати аналіз рухів об'єктів у сентиментальному аналізі (визначення рівня сп'яніння та ін.)

Потреба дослідження цієї теми підтверджена малою кількістю публікацій та зростанням попиту на такі розробки у зв'язку з динамічним розвитком індустрії сервісів. Варто зазначити, що такі системи поширені на азіському ринку та в північній Америці, отже найближчим часом схожі потреби виникнуть на українському ринку де, власне і планується застосування результатів цього дослідження.

Аналіз досліджень та публікацій

У статті [1] автор обговорює керування жестами системою за допомогою обробки зображень вилучених з відеопотоку. Робототехнічна промисловість розвиває багато нових тенденцій для підвищення ефективності, доступності та точності систем. Роботи замінюють людей. Але все одно ці роботи потребують уваги та обслуговування людини до того ж для керування роботом потрібна людина. Роботи можуть мати різні типи підключення до комп'ютерної системи: дротові або бездротові. Проте, незважаючи на тип з'єднання усі роботи мають пристрій керування. Для керування робототехнікою за допомогою фізичних пристроїв популярним методом є керування жестами. Основна мета використання жестів полягає в тому, що він забезпечує більш природний спосіб керування та забезпечує насичену та інтуїтивну форму взаємодії з роботизованою системою. Сервісні роботи безпосередньо взаємодіють з людьми, тому знаходження більш природного і простого інтерфейсу користувача має принципове значення. У цьому дослідженні, робота зосереджена на питаннях, пов'язаних з маніпуляцією та навігацією в навколишньому середовищі. Для вирішення цієї проблеми автор реалізував систему, базовану на використанні згорткової нейронної мережі. За допомогою такого сервісу користувач може давати команди бездротовому роботу за допомогою жестів. За допомогою цього методу користувач може керувати роботом за допомогою жестів долонею, взаємодіючи, таким чином, з роботизованою системою.

У роботі [2] досліджено системи безперервної зйомки в які інтегровані функції розпізнавання об'єктів, оскільки людину можна легко відволікти, і одна помилка може призвести до великої катастрофи. Таким чином, системи відеоспостереження значно полегшують роботу споживачам та забезпечують безпеку та контроль там, де він потрібен весь час. У цій роботі автором запропоновано алгоритм (сформовану архітектуру згорткової нейронної мережі, на основі паралельного вилучення ознак), який допомагає виявляти рухомий об'єкт і класифікувати його як людину та відстежувати рух людини. Це можна зробити без будь-якого використання сенсорних пристроїв. У цій статті запропоновану систему можна розділити на три етапи: виявлення, відстеження та аналіз дій. Виявлення людини здійснюється поєднанням морфологічної процедури та особливого методу екстракції.

Методики, викладені в [3], варіюються від дуже простого алгоритму до опублікованих методів, які класифікуються на основі швидкості, вимог до пам'яті та точності. Вони використовували такі методи, як різниця кадрів, віднімання фону та виявлення тіні в реальному часі, адаптивна модель суміші фону для методу відстеження в реальному часі. Вони використовували алгоритми різного рівня точності та обчислювальної складності. Деякі з них також можуть вирішувати проблеми в реальному часі, як-от сніг, дощ, рухомі гілки, об'єкти, що перекривають досліджуваний об'єкт, інтенсивність світла або повільно рухомі об'єкти. Аналіз зображень відокремлених з відеопотоку відбувався з застосуванням згорткової нейронної мережі.

Проблема досягнення високої швидкості розпізнавання з низькою частотою помилок виявлення і відстеження людини у відеопослідовності полягає в тому, що алгоритмам потрібна максимізація продуктивності та зменшення часу реакції. Зазначеними причинами є вплив складності сцени, зміни масштабу та взаємодії фону та людини. Двоетапне рішення для обробки, яке полягає в виявленні людини, і відстеження людини за допомогою двох шаблонних класифікаторів, представлені у [4].

Існує три основні етапи відеодослідження: виявлення цікавих об'єктів у відеосцені, відстеження таких об'єктів від кадру до кадру та аналіз слідів об'єктів для розпізнавання їх діяльності. Виявлення людей на відео є складною проблемою через рух об'єктів. У [5] дослідники розробили детектор для переміщення людей на відео з можливо рухомими камерами та фоном, перевірили кілька різних схем кодування рухомого об'єкта та показавши, що орієнтовані гістограми диференційного оптичного потоку дають максимальну продуктивність.

Виклад основного матеріалу

Відстеження об'єкта у відео можна виразити як процес пошуку шляху руху об'єкта в часі шляхом визначення його положення в кожному кадрі відео. Відстеження об'єктів стало одним з найпопулярніших

предметів дослідження в області комп'ютерного зору. Хоча це добре вивчене питання, вона залишається проблемою в багатьох аспектах. Створення трекерів для певних класів об'єктів, таких як люди чи обличчя, досягло значного прогресу протягом багатьох років, тоді як трекери для загальних об'єктів продовжують залишатися складною областю. Об'єкти, які різко змінюють зовнішній вигляд через, наприклад, розтягнення або зміни світла, шум у зображеннях, складний рух і складні форми об'єктів, є деякими прикладами того, чому загальне відстеження об'єктів вважається складною проблемою [6, 7]. Уявіть, що ви дивитеся на довільний рухомий об'єкт, який раптово змінює траєкторію руху або особливості зовнішнього вигляду, такі як форма чи колір. Це разом із кількома іншими об'єктами в декорації, можуть вести себе аналогічно. Відстеження об'єкта в більшості випадків було б складним завданням навіть для людського ока. З огляду на це, це означає, що це майже неможливе завдання для комп'ютера. З огляду на це, існують способи спростити відстеження, застосувавши деякі обмеження. Майже всі алгоритми припускають, що рух відстежуваного об'єкта плавний, без різких змін. Постійна швидкість або прискорення, а також попередні знання про зовнішній вигляд об'єкта є іншими способами спрощення завдання [7]. Процес побудови об'єктного трекера зазвичай поділяється на кілька кроків, якими є представлення об'єкта, виявлення об'єкта та відстеження об'єкта [8]. Цей розділ міститиме ретельний огляд різних методів відстеження об'єктів, що є основною темою цієї дипломної роботи. Однак він також коротко охопить область представлення об'єктів та виявлення об'єктів, оскільки це важливі кроки в процесі успішного відстеження об'єктів. Таким чином читач отримує кращу основу та розуміння процесу побудови об'єктного трекера. Загальна методологія опрацювання об'єкта в такій системі описана нижче.

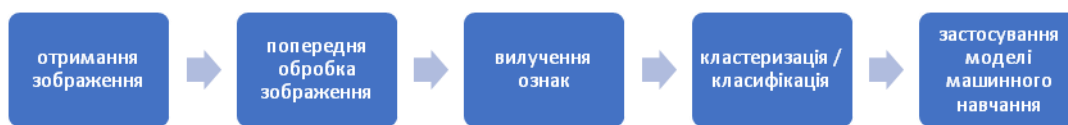


Рис. 1. Методологія опрацювання зображення

- Отримання зображень: обробку зображень можна в широкому розумінні визначити як дію отримання зображення з певного джерела. Зйомка зображення під час обробки зображень завжди є першим кроком у послідовності робочого процесу, оскільки без зображення обробка неможлива. Якщо обладнання не налаштовано належним чином і не вирівняно, то можуть виникнути візуальні артефакти, які можуть ускладнити обробку зображення. Одна з форм отримання зображення в обробці зображень відома як отримання зображення в реальному часі. Зазвичай це передбачає отримання зображень із джерела, яке автоматично знімає зображення. Отримання зображень у реальному часі створює потік файлів, які можна автоматично обробляти, ставити в чергу для подальшої роботи або зшивати в єдиний медіаформат (відео або набір кадрів).

- Попередня обробка зображення: це обробка зображень за допомогою математичних операцій за допомогою будь-якої форми обробки сигналу, для якої вхідним є зображення, серія зображень або відео, наприклад фотографія або відеокадр; Результатом обробки зображення може бути або зображення, або набір характеристик або параметрів, пов'язаних із зображенням.

- Вилучення ознак: це тип зменшення розмірності, який ефективно представляє цікаві частини зображення у вигляді компактного вектора ознак. Цей підхід корисний, коли зображення великі, а для швидкого виконання таких завдань, як зіставлення та пошук зображень, потрібне зменшене представлення функцій.

- Вилучення ознак: це тип зменшення розмірності, який ефективно представляє цікаві досліднику частини зображення у вигляді компактного вектора ознак. Цей підхід корисний, коли зображення великі, а для швидкого виконання таких завдань, як зіставлення та пошук зображень, потрібне зменшене представлення функцій.

- Кластеризація: це завдання поділу сукупності або точок даних (в нашому випадку це пікселів на зображенні чи відеокадрі) на кілька груп таким чином, щоб точки даних в тих самих групах були більш схожими на інші точки даних у тій же групі, ніж в інших групах. Простіше кажучи, мета полягає в тому, щоб відокремити групи зі схожими ознаками та об'єднати їх у кластери.

Щоб мати можливість відслідковувати об'єкт, що цікавить, його спочатку потрібно представити у спосіб, який має сенс для комп'ютера. Властивості, що стосуються зовнішнього вигляду та форми, зазвичай використовуються як основа для представлення. Подання зовнішнього вигляду або подання форми можна використовувати окремо для представлення об'єкта, але також можна використовувати для цієї мети в комбінації. Зовнішні фактори, такі як область застосування, мета та ціль, визначають, як має бути представлений об'єкт або об'єкти. Представлення, у свою чергу, визначає вибір відповідного алгоритму для відстеження.

Уявлення об'єкта = Форма + Зовнішній вигляд

Огляд наборів даних

Зробивши аналіз статей та робіт на схожі теми, зроблено висновок про доцільність використання кількох наборів даних для навчання моделі. Очевидно, що наборів даних, які включають в себе лише зображення одного спрямування варто уникати, адже майбутня система спрямована на трекінг не лише однотипових об'єктів, а навпаки – її завданням є розпізнавання та відслідковування одразу кількох об'єктів в кожному кадрі.

Таблиця 1 [1]

Огляд наборів даних, що містять зображення різних класів

База даних	Кількість прикладів	Роздільна здатність	Колір	Кількість категорій	Особливості
Caltech 101	9146	300×200	RGB	101	Орієнтовані об'єкти (наприклад мотоцикли, літаки) завжди розміщені зліва направо, а вертикальні структури (наприклад будинки) перевернуті в нетипові положення відносно своєї стандартної осі.
Caltech 256	30 607	351×351	RGB	256	До кожного зображення додається кілька анотацій – координати рамки, в якій знаходиться об'єкт і детальний опис.
ImageNet	14 197 122	469×387	RGB	21 841	Зображення в базі даних організовані в ієрархію.
Microsoft Common Objects in Context (COCO)	2 500 000	не вказано	RGB	80	Зображення різних побутових сцен та звичайних предметів у буденному використанні.
Open Images	9 178 275	не вказано	RGB	600	в середньому 8 описаних об'єктів на 1 кадр
YFCC100M	100 000 000	не вказано	RGB	не вказано	Великий і різноманітний набір даних зображень і відео

Усі вище наведені набори даних проаналізовано, та виділено переваги та недоліки кожного. Варто зазначити що для побудови повноцінної системи розпізнавання та трекінгу об'єктів варто скористатись набором даних де є не лише статичні зображення, а й динамічні послідовності кадрів з одними і тими ж об'єктами на них. Відповідно як результат подальша робота над реалізацією ідеї проводитиметься з опрацюванням наборів даних Open Images та YFCC100M, адже ці набори даних відповідають вимогам, а також містять велику кількість екземплярів для дослідження та тренування нейромережі.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

В результаті проведеної роботи досліджено наявні методи та підходи до детекції об'єктів на зображеннях та відео. На основі проаналізованих наборів даних, визначено з якими буде проведено подальшу роботу, зокрема тренування згорткової нейронної мережі, а також застосування алгоритму при створенні системи розпізнавання об'єкту на зображеннях (відеокадрах). Загалом, дослідження теми розпізнавання об'єктів довело актуальність обраної теми, адже на теренах України дана тема не є широко досліджуваною, до того ж з кожним днем потреба застосування такої системи зростає. Щодня запускають нові сервіси, що обслуговують людей – розробка такої системи посприє зростанню ефективності та швидкодії вищевказаних сервісів.

Література

1. Zhu H., Wei H., Li B., Yuan X., Kehtarnavaz N. A Review of Video Object Detection: Datasets, Metrics, and Methods. Applied Sciences. 04, 2020;10(21):7834.
2. Arulprakash E., Aruldoss M. A study on generic object detection with emphasis on future research directions. Journal of King Saud University – Computer and Information Sciences. 2021. URL: <https://www.sciencedirect.com/science/article/pii/S1319157821002020>
3. Tamilselvan L., Ravichandran M. A STUDY ON OBJECT DETECTION. 2016.

-
4. Chahal K.S., Dey K. A Survey of Modern Object Detection Literature using Deep Learning. arXiv:180807256. 22, 2018. URL: <http://arxiv.org/abs/1808.07256>
 5. Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:170404861. 2017. URL: <http://arxiv.org/abs/1704.04861>
 6. Wu J., Peng B., Huang Z., Xie J. Research on Computer Vision-Based Object Detection and Classification. B: Li D., Chen Y. Computer and Computing Technologies in Agriculture VI. Berlin, Heidelberg: Springer; 2013. p. 183–8. (IFIP Advances in Information and Communication Technology).
 7. Irram S., Ahmad S.F. Research on Object Detection in Video Streaming Using Deep Learning. 10.
 8. Ahmed M., Hashmi K.A., Pagani A., Liwicki M., Stricker D., Afzal M.Z. Survey and Performance Analysis of Deep Learning Based Object Detection in Challenging Environments. Sensors. 2021; 21(15):5116.