УДК 004.8:004.032.26:528.3

К.т.н., доц. Андрусенко О.М., магістрант Хавронюк Б.А.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

КОНЦЕПТУАЛЬНА МОДЕЛЬ ТА СИСТЕМА ГЕОПОЗИЦІЮВАННЯ НА ОСНОВІ АНСАМБЛЕВИХ НЕЙРОННИХ МЕРЕЖ З УРАХУВАННЯМ ЧАСОВОЇ ПОСЛІДОВНОСТІ КАДРІВ

Abstract

Olena Andrusenko, PhD, Associate Professor; Bohdan Khavroniuk, student

A conceptual model and system for geolocation based on ensemble neural networks considering the temporal sequence of frames

This paper addresses the problem of achieving reliable and high-precision geopositioning in complex environments where satellite signals are compromised. To address these limitations, we introduce a hybrid, time-aware navigation system that leverages the strengths of multiple deep learning models. Proposed system combines an ensemble of different types of CNNs (for example, ResNet and EfficientNet) to extract spatial features from video streams more robustly. These features are then analyzed by a recurrent neural network, such as an LSTM or GRU, which models the temporal sequence of frames and integrates motion context.

Вступ

Сучасні системи геопозиціювання активно розвиваються у напрямку інтеграції комп'ютерного зору, нейронних мереж та ансамблевих методів машинного навчання. Потреба у точному та надійному визначенні місцеположення є критичною для широкого спектра застосувань, включаючи автономний транспорт, роботизовані системи, доповнену реальність та логістику.

Традиційні підходи, що ґрунтуються лише на супутникових даних (GPS, Galileo), не завжди забезпечують точність і стабільність у складних умовах. На якість позиціювання можуть впливати такі фактори, як погодні умови, рельєф місцевості та щільна міська забудова.

Використання глибоких нейронних мереж для обробки зображень з камер (Візуальна Одометрія та Візуальне Розпізнавання Місця, VPR) дозволяє доповнити класичні навігаційні рішення візуальною інформацією. Згорткові нейронні мережі (CNN) довели свою ефективність у виділенні

стійких до змін умов (освітлення, погода, пора року) дескрипторів зображень [1]. Однак моделі CNN не завжди надійно визначають просторові координати. Вони можуть виявитися чутливими до умов, які не входили в навчальний набір (наприклад, темний час доби, дощ або сніг), а також до змін у сценах та до візуального шуму.

Для подолання цих недоліків буде доцільним застосування ансамблевих архітектур. Ансамбль, що поєднує кілька різнорідних моделей глибинного навчання (наприклад, ResNet, EfficientNet), здатний узагальнювати ознаки набагато краще, ніж окрема модель, компенсуючи слабкі сторони однієї моделі сильними сторонами іншої [2].

Водночас, аналіз окремих, не пов'язаних між собою кадрів, ігнорує важливу часову інформацію. Відео є послідовністю кадрів, що обумовлює часову неперервність (або узгодженість) руху об'єктів. Позиція в момент часу t сильно корелює з позицією в момент t. Рекурентні нейронні мережі (RNN), зокрема їх більш складні варіанти як LSTM (Long Short-Term Memory) [3] або GRU (Gated Recurrent Unit), використовуються для моделювання таких часових залежностей.

Таким чином, виділяється невирішена раніше проблема створення гібридної системи, яка б поєднувала в собі просторову стійкість ансамблів CNN та можливість за допомогою RNN моделювати часовий контекст для досягнення точного та плавного геопозиціювання.

Постановка задачі

Метою дослідження є розробка концептуальної моделі та архітектури системи геопозиціонування, яка, за допомогою поєднання просторової стійкості, ансамблю CNN та часової узгодженості RNN/LSTM, забезпечує підвищену точність визначення місцеположення, стабільність результатів в умовах низької видимості або швидких змін освітлення, та вищу стійкість до зовнішніх перешкод порівняно з традиційними GPS-залежними або простими візуальними моделями.

Концептуальна модель гібридної системи

Запропонована архітектура системи є багатоступеневим конвеєром обробки даних, що складається з п'яти ключових модулів (рис. 1).

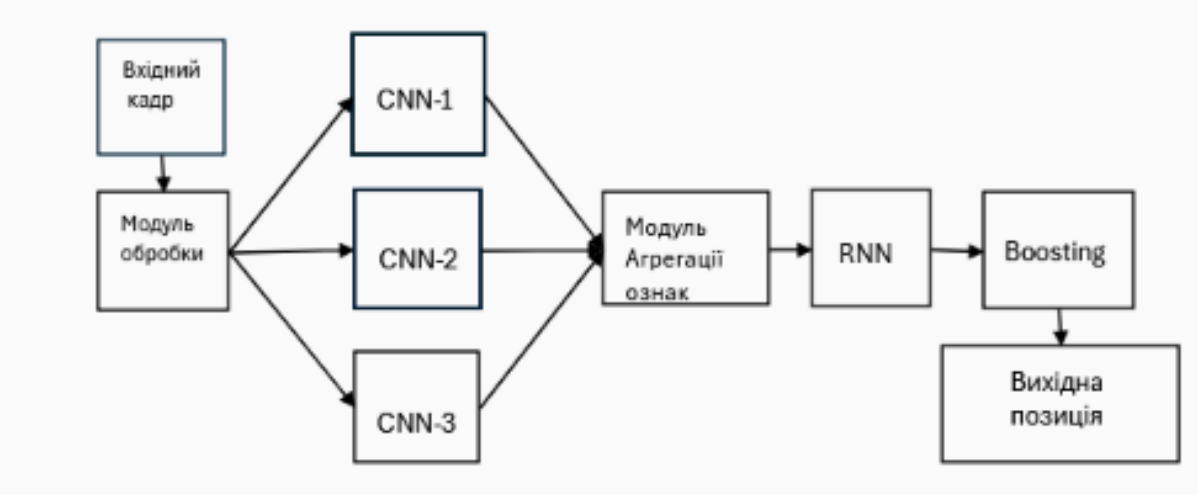

Рис. 1. Діаграма, що візуалізує потік даних

Модуль попередньої обробки Цей модуль відповідає за підготовку вхідних кадрів It до подачі в нейронні мережі. Він виконує стандартизовані операції:

1. Нормалізація: Зміна розміру зображення до єдиного стандарту, що вимагається архітектурами CNN.
2. Фільтрація шуму: застосування фільтрів (наприклад, медіанного або Гаусса) для зменшення візуального шумів на зображенні, хоча сучасні CNN часто стійкі до нього.
3. Аугментація: Для підвищення стійкості моделі, на етапі навчання застосовуються випадкові трансформації: зміна яскравості, контрастності, невеликі повороти та зсуви.

Ансамбль згорткових нейронних мереж є ядром системи, що відповідає за виділення просторових ознак. Замість однієї моделі використовується гетерогенний ансамбль, що включає моделі з різними архітектурними підходами:

1. ResNet (напр., ResNet-50) [1]: Забезпечує глибоке виділення ознак завдяки залишковим з'єднанням (residual connections), що вирішує проблему згасання градієнта.
2. EfficientNet: Масштабує глибину, ширину та роздільну здатність мережі збалансовано, досягаючи високої точності при меншій кількості параметр.
3. MobileNetV3: Оптимізована для мобільних та вбудованих пристроїв, забезпечує надзвичайно швидке виведення (inference), що критично для роботи в реальному часі.

Кожна з цих мереж паралельно обробляє вхідний кадр It і генерує свій вектор ознакю. Ці вектори потім конкатенуються у єдиний, збагачений вектор ознак Ft.

RNN-модуль (LSTM або GRU) конкатенований вектор Ft несе повну просторову інформацію про поточний кадр. Цей вектор подається на вхід

рекурентного модуля. В роботі пропонується використовувати LSTM (Long Short-Term Memory) [2], оскільки ця архітектура здатна ефективно навчатися довгостроковим залежностям у даних завдяки своїм "коміркам пам'яті" та механізмам гейтів (input, forget, output gates). LSTM-модуль обробляє послідовність векторів та оновлює свій внутрішній прихований стан . Цей стан є стисненим представленням не лише поточного кадру, але й усієї релевантної історії руху, що дозволяє системі згладжувати раптові помилки (викиди) в розпізнаванні окремих кадрів та враховувати кінематику об'єкта.

Дерево прийняття рішень/Boosting-модуль використовується замість стандартного повнозв'язного шару. Вихід ht з LSTM є високорозмірним вектором ознак. Просте лінійне відображення може бути недостатньо гнучким. Використання ансамблевих методів, таких як градієнтний бустинг (напр., XGBoost, CatBoost) або випадковий ліс (Random Forest), поверх виходу з LSTM може дати кращі результати.

Для валідації запропонованої концептуальної моделі необхідно провести експериментальне порівняння з базовими підходами на стандартизованих наборах даних, таких як Oxford RobotCar (який містить дані зібрані за різних умов).

Метрики оцінювання:

1. Точність позиціювання: Середня помилка трансляції (в метрах) та орієнтації (в градусах) на 100 метрів пройденого шляху.
2. Стабільність (плавність): у прогнозованій траєкторії порівняно з моделями без RNN.
3. Робастність: Оцінка деградації точності в складних умовах (ніч, дощ, туман), які присутні в датасеті Oxford RobotCar.
4. Обчислювальна ефективність: Кількість кадрів в секунду (FPS) на цільовій платформі.

Базові моделі для порівняння (Baselines):

1. Baseline 1: Окрема CNN + LSTM (для оцінки виграшу від ансамблю).
2. Baseline 2: Ансамбль CNN без RNN (для оцінки виграшу від часової моделі).
3. Baseline 3: Класичний метод (ORB-SLAM2) на тому ж відеопотоці.

Очікується, що запропонована модель (Ансамбль CNN + LSTM + Boosting) продемонструє значне зниження середньої помилки (до 10-15% порівняно з Baseline 1 та Baseline 2 та суттєво вищу стабільність траєкторії, особливо на довгих послідовностях та у складних візуальних умовах.

Висновки

У роботі запропоновано концептуальну модель та архітектуру системи геопозиціювання, що базується на гібридному підході, який поєднує ансамблеві згорткові нейронні мережі та рекурентні нейронні мережі для обробки часових послідовностей. Ключовими перевагами моделі є: робастність, часова узгодженість, гнучкість.

Подальші дослідження будуть зосереджені на трьох основних напрямках:

1. Використання архітектур Трансформерів [5]: Дослідження можливості заміни RNN-модуля на архітектури на основі механізмів уваги (Attention), зокрема Vision Transformers (ViT) або спеціалізованих часових трансформерів.
2. Прототипування та оптимізація: Реалізація розробленої моделі у вигляді реального прототипу на вбудованих обчислювальних пристроях (наприклад, NVIDIA Jetson Xavier NX).

Література

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234-241). Springer.
2. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2961-2969).
3. Huang, Q., & Huang, J. (2023). Comprehensive review of edge and contour detection: from traditional methods to recent advances. *Displays, 79*, 102462.
4. Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8*(6), 679-698.
5. Gonzalez, R. C., & Woods, R. E. (2009). *Digital Image Processing* (3rd ed.). Pearson.