

УДК 629.7.05:004.93
УКПП
№ держреєстрації 0117U003934
Інв. №

Міністерство освіти та науки України
Сумський державний університет
(СумДУ)

40007, м. Суми, вул. Римського-Корсакова, 2
тел. (0542) 33-54-79 факс (0542) 33-54-79

ЗАТВЕРДЖУЮ

Проректор з наукової роботи
д-р. фіз.-мат. наук, професор

_____ А.М. Черноус

ЗВІТ

ПРО НАУКОВО-ДОСЛІДНУ РОБОТУ

Інтелектуальна автономна бортова система безпілотного літального апарату
для ідентифікації об'єктів на місцевості

РОЗРОБКА ТЕХНІЧНОГО ПРОЕКТУ БОРТОВОЇ СИСТЕМИ
БЕЗПІЛОТНОГО ЛІТАЛЬНОГО АПАРАТУ
(проміжний)

Начальник НДЧ
канд. фіз.-мат. наук,
старш. наук. співроб.

Д.І. Курбатов

Керівник НДР
канд. техн. наук

В.В. Москаленко

2017

Рукопис завершено 20 грудня 2017 р.

Результати цієї роботи розглянуто на засіданні наукової ради СумДУ, протокол від
29.11.2017 № 2

СПИСОК АВТОРІВ

Керівник НДР,
канд. техн. наук,
провідн. наук. співроб.

(20.12.2017)

Москаленко В. В. (вступ, висновки,
підрозділи 1.1-1.3, 2.2)

Відповідальний виконавець
канд. техн. наук,
старш. наук. співроб.

(20.12.2017)

Москаленко А. С. (підрозділ 1.2)

Виконавці:
канд. техн. наук,
старш. наук. співроб.

(20.12.2017)

Берест О. Б. (підрозділ 2.3)

канд. техн. наук,
старш. наук. співроб.

(20.12.2017)

Мартиненко С. С. (підрозділ 1.3)

канд. фіз.-мат. наук,
старш. наук. співроб.

(20.12.2017)

Журба В. О. (підрозділ 2.1)

канд. техн. наук,
старш. наук. співроб.

(20.12.2017)

Нагорний В. В. (підрозділ 2.3)

мол. наук. співроб.

(20.12.2017)

Окопний Р. П. (підрозділ 2.1)

аспірант,

мол. наук. співроб.

(20.12.2017)

Коробов А. Г. (підрозділ 1.1, 1.3)

мол. наук. співроб.

(20.12.2017)

Бойко О. В. (підрозділ 1.2)

РЕФЕРАТ

Звіт про НДР: 85 с., 1 табл., 31 рис., 21 джерело.

БОРТОВА СИСТЕМА БЕЗПІЛОТНОГО ЛІТАЛЬНОГО АПАРАТУ, ЕКСТРАКТОР ОЗНАК, ІНФОРМАЦІЙНИЙ КРИТЕРІЙ, ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНА ІНТЕЛЕКТУАЛЬНА ТЕХНОЛОГІЯ, МАШИННЕ НАВЧАННЯ, НАВІГАЦІЯ, ОПТИМІЗАЦІЯ, РОЗРІДЖЕНЕ КОДУВАННЯ.

Об'єкт дослідження – процес машинного навчання бортової системи безпілотного апарату та автономної ідентифікації об'єктів на місцевості за даними різнотипних сенсорів.

Мета роботи – підвищення функціональної ефективності бортової системи безпілотного літального апарату, що здійснює у автономному режимі локальну навігацію та класифікаційний аналіз спостережень за даними різнотипних сенсорів за умов ресурсних та інформаційних обмежень.

Методи дослідження – аналіз літературних джерел, інформаційно-екстремальна інтелектуальна технологія оцінки ефективності моделей подання даних.

Виконано аналітичний огляд сучасного стану розроблення бортових систем безпілотних літальних апаратів і сформовано технічне завдання на дослідний зразок інтелектуальної бортової системи для ідентифікації об'єктів на місцевості. Було розроблено структурні та функціональні схеми апаратно-програмного та алгоритмічного забезпечення безпілотного літального апарату. Вперше запропоновано алгоритм самонавчання екстрактора ієрархічного ознакового опису спостережень на основі принципів розрідженого кодування та самоорганізації неструктурованих сіток. Запропоновано критерій оцінки ефективності вирішальних правил автономної бортової системи для ідентифікації об'єктів на місцевості.

Результати виконання роботи впроваджено в навчальний процес при підготовці лекційних курсів з навчальних дисциплін «Інтелектуальні системи керування» та «Мови програмування інтелектуальних систем», а також при написанні трьох магістерських робіт.

ЗМІСТ

| | |
|---|----|
| Скорочення та умовні позначки | 6 |
| Вступ..... | 7 |
| 1 Аналітичний огляд підходів до аналізу та синтезу автономної бортової системи безпілотного літального апарату | 9 |
| 1.1 Методи, моделі та алгоритми екстракції ознакового опису середовища із зображень аеровідеозйомки | 9 |
| 1.2 Методи, моделі та алгоритми ідентифікації об'єктів інтересу на місцевості..... | 23 |
| 1.3 Методи, моделі та алгоритми навігації автономних безпілотних літальних апаратів..... | 47 |
| 2 Аналіз і синтез інтелектуальної бортової системи ідентифікації об'єктів на місцевості | 67 |
| 2.1 Аналіз об'єкту та обґрунтування напрямку дослідження | 67 |
| 2.2 Формалізована постановка задачі інформаційного синтезу автономної системи ідентифікації об'єктів на місцевості, що навчається | 72 |
| 2.3 Оцінка функціональної ефективності навчання системи ідентифікації об'єктів інтересу на місцевості..... | 75 |
| Висновки | 81 |
| Перелік джерел посилання | 83 |

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

БЛА – безпілотний літальний апарат;

ІЕІ – технологія – інформаційно-екстремальна інтелектуальна технологія;

КФЕ – критерій функціональної ефективності.

ВСТУП

Розробка безпілотних літальних апаратів (БЛА) є одним з найважливіших напрямків в сучасній авіації, який переживає справжній бум. Безпілотна авіація знаходить широке застосування у точному землеробстві, розвідувальних та рятувальних операціях, транспорті та відеозйомці. Проте на сьогоднішній день практично відсутні БЛА загального призначення, здатні функціонувати в автономному режимі з можливостями ідентифікації об'єктів інтересу на місцевості. Існуючі рішення або не здатні до переналаштування і адаптації під нову прикладну задачу і умови експлуатації, або використовують дороге програмно-апаратне забезпечення і мають порівняно високу вагу.

За умов обмежених ресурсів бортового комп'ютера безпілотного літального апарату на ньому неможливо розгорнути інтелектуальну систему ідентифікації об'єктів на місцевості, яка буде оптимальним для будь-якого середовища і задач. Зниження вимог до апаратних ресурсів безпілотного літального пристрою та підвищення достовірності рішень, що автономно приймаються бортовою системою ідентифікації об'єктів інтересу на місцевості, за умов нестаціонарності середовища та мінливості об'єктів інтересу є актуальною задачею. Її вирішення дозволяє зменшити вагу, вартість та розширити функціональні можливості бортової системи, що зробить літальний апарат конкурентоздатним на ринку безпілотної авіації.

Розробка, що пропонується, є актуальною для підвищення обороноздатності країни, оскільки дозволяє підвищити ефективність загальновійськової системи збору інформації та прийняття рішень. Також впровадження даної технології в аграрному господарстві України дозволить зменшити накладні витрати на моніторинг рослинних культур, який є важливим етапом у комплексі мір щодо підвищення врожайності у рослинництві.

Проміжний звіт складається із вступу, двох розділів, висновків і переліку посилань.

Перший розділ присвячено аналізу сучасного стану і тенденцій розвитку

автономних бортових систем безпілотних літальних апаратів для ідентифікації об'єктів інтересу на місцевості та аналітичному огляду методів їх аналізу і синтезу.

У другому розділі виконано аналіз об'єкту, запропоновано структурні та функціональні схеми програмно-апаратного забезпечення бортової системи БЛА, обґрунтування напрямку дослідження та формалізована постановка задачі інформаційного синтезу автономної системи ідентифікації об'єктів на місцевості на основі нейромережевої та інформаційно-екстремальної інтелектуальної технологій (ІЕІ - технології) аналізу даних.

Результати наукових досліджень, одержаних виконавцями проекту, опубліковано в працях [1]–[4].

1 АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО АНАЛІЗУ ТА СИНТЕЗУ АВТОНОМНОЇ БОРТОВОЇ СИСТЕМИ БЕЗПІЛОТНОГО ЛІТАЛЬНОГО АПАРАТУ

1.1 Методи, моделі та алгоритми екстракції ознакового опису середовища із зображень аеровідеозйомки

Класифікаційний аналіз зображень є важливою складовою інтелектуальної обробки спостережень безпілотної апарату для реалізації автономної навігації та ідентифікації об'єктів інтересу на місцевості. При цьому ефективність класифікаційного аналізу залежить від оптимальності вибору структури та параметрів екстрактора ознакового опису спостережень та вирішальних правил. Крім того можливість підтримання ефективності функціонування бортової системи апарату на високому рівні залежить від здатності інтелектуальної системи з обмеженими ресурсами адаптуватися до модифікацій об'єкта інтересу та зміни умов чи середовища спостереження.

Зображення аеровідеозйомки характеризуються низькою інформаційною ємністю – тобто, більша частина растрового зображення (наприклад, місця однотонного або рівномірно розподіленого фону) не містять інформації, яка впливає на процес розпізнавання. Це дозволяє застосовувати підхід пошуку компактного подання зображень, яке містить основну інформацію, необхідну для розпізнавання – виділення обмеженого числа деяких інформативних ознак або застосування статистичних методів зниження розмірності, наприклад методу головних компонентів [5].

На основі аналізу праць [3]–[5] можна сформулювати такі загальні положення та принципи, що лежать в основі технік подання даних для задач інтелектуального аналізу інформації та прийняття рішень :

- гладкість – функція f , якій алгоритм буде навчений, у загальному випадку передбачає виконання умови $f(x) \approx f(y)$ при $x \approx y$;
- множинні пояснюючі фактори (першопричини) – вхідні дані є результатом взаємодії багатьох основоположних факторів і навчання моделі

одному новому фактору призводить у загальному випадку до узагальнення його в конфігураціях інших факторів;

- ієрархія в організації пояснюючих факторів (першопричин) – поняття, які використовуються для опису світу довкола нас, можуть бути визначені в термінах інших понять, ієрархічно, з більш абстрактними поняттями вверх по ієрархії;

- часткове навчання з учителем (Semi-supervised learning) – подання даних, яке зручне для обчислення $P(X)$ стає зручним і для навчання $P(Y|X)$, що дозволяє об'єднати статистичний взаємозв'язок між машинним навчанням з учителем і без вчителя;

- спільні фактори від задачі до задачі – в багатозадачних моделях ті задачі, що використовують спільні з іншими задачами пояснючі фактори, дозволяють використовувати статистичний взаємозв'язок навчання між різними задачами;

- різноманіття (Manifolds) – основна щільність ймовірності даних зосереджена біля регіонів, які мають набагато меншу розмірність, ніж оригінальний простір, де знаходяться дані;

- природна кластеризація – локальні зміни в різноманітності мають тенденцію до збереження значень самої категорії, а лінійна інтерполяція між зразками різних класів проходить у загальному випадку через регіони з низькою щільністю ймовірності, (гіпотеза основана на тій ідеї, що люди дають назви (імена) категоріям і класам на базі статистичної схожості об'єктів);

- часова і просторова зв'язаність – спостереження, сформовані у сусідніх областях простору чи отримані послідовно в часі, повинні прагнути асоціюватися з однаковими значеннями відповідної категорії понять, чи приводити до невеликого руху по поверхні різноманіття високої щільності, (іншими словами при спробі здійснити зміни в категоріях, викликане факторами з різним часовим і просторовим масштабом, така зміна повинна відбуватися дуже повільно, тобто повинне бути введено обмеження на зміну «крізь час і простір»);

– розрідженість (Sparsity) – для будь-якого даного спостереження x тільки мала частина з усіх можливих факторів (першопричин події) є значимою, тобто більша частина виділених ознак нечутлива до малих змін спостереження x , тому більша частина детектованих ознак повинна бути нульовою;

– простота залежності факторів – в хороших високорівневих подання даних фактори зв'язані один з одним через прості, як правило лінійні залежності.

Найбільшого успіху в задачі аналізу зображень набула згорткова архітектура нейронної мережі. В даній мережі згідно принципу просторової зв'язаності кожен нейрон дивиться лише на частину зображення (рис. 1.1а), а не зв'язаний з всіма пікселями (рис. 1.1б). Таке архітектурне рішення дозволяє значно знизити кількість параметрів моделі і підвищити оперативність функціонування, оскільки нейрони, що представляють з себе згорткові фільтри, сканують зображення, тобто відбувається повторне використання вагових коефіцієнтів нейронів.

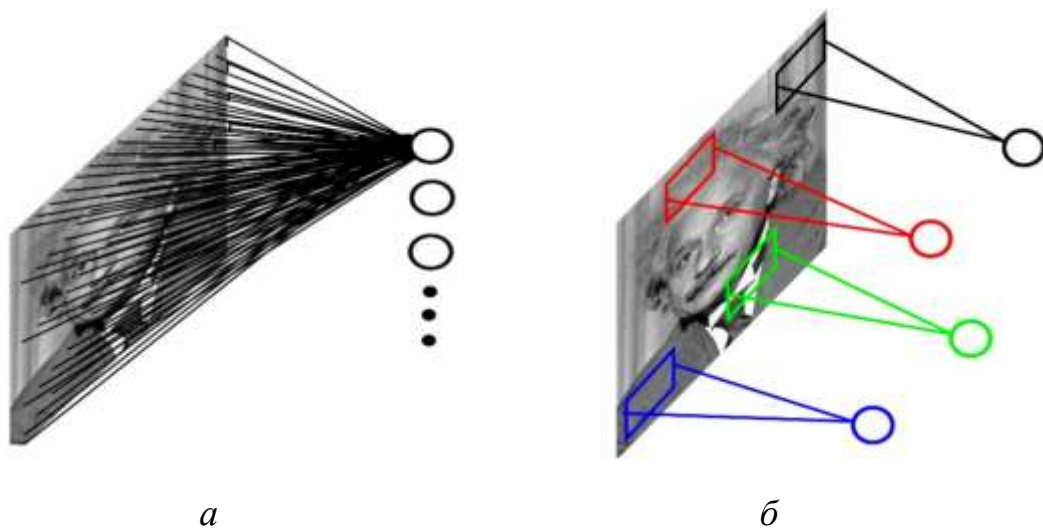


Рисунок 1.1 – Схема нейронного шару : *a* – повнозв'язна мережа;

б – згорткова мережа

У задачах формування ознакового подання можна використовувати як однорівневі так і багаторівневі (глибокі) архітектури моделей. У працях [5]–[6] було доведено, що як однорівневі так і глибокі моделі є універсальними, тобто вони у зв'язці з вирішальними правилами можуть наблизитись до будь-якої неперервної функції n змінних у компактній області значень. Точність

апроксимації функцій моделями як з однорівневою так і глибокою архітектурою є еквівалентною, оскільки неглибокі мережі можна розглядати як частинний випадок глибоких мереж. Проте у випадку високої розмірності вхідного вектора ознак кількість параметрів для глибоких мереж є набагато меншою порівняно з однорівневими, що дозволяє знизити обчислювальну складність алгоритмів функціонування системи. На відміну від однорівневої моделі глибока модель подання інформації представляє в своїй архітектурі структурну інформацію про вхідні дані та може бути адаптована до детального сприйняття такої інформації.

Згідно принципу ієрархії в організації пояснюючих факторів, згорткова мережа є багатошаровою. При цьому у згортковому шарі нейрони, які використовують одні і ті ж ваги, об'єднуються в карти ознак (feature maps), а кожен нейрон карти ознак пов'язаний з частиною нейронів попереднього шару. Приклад архітектури згорткової нейронної мережі представлено на рис. 1.2.

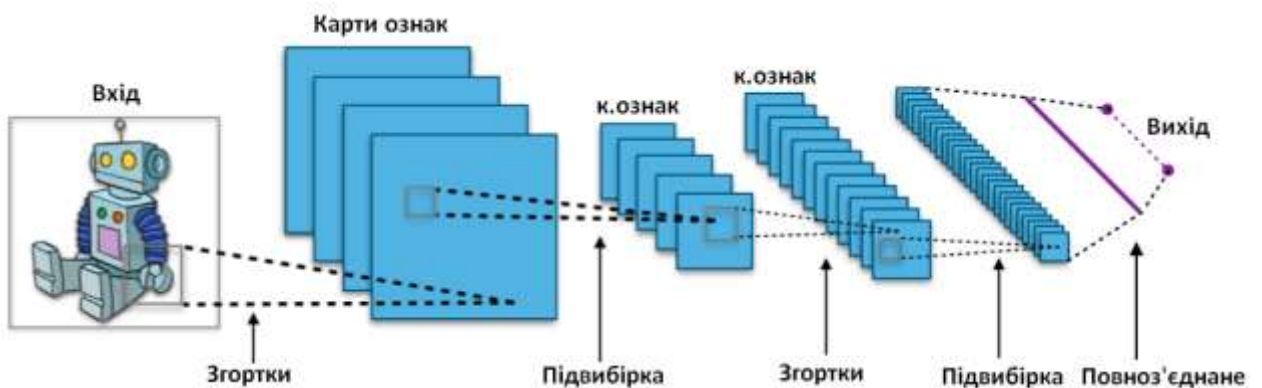


Рисунок 1.2 – Архітектура згорткової нейронної мережі

При обчисленні мережі виходить, що кожен нейрон виконує згортку деякої області попереднього шару (яка визначається множиною нейронів, пов'язаних з даним нейроном).

Особливістю згорткового шару є те, що він дещо зменшує зображення за рахунок крайових ефектів. На рис. 1.3 показано приклад згорткового шару з ядром згортки розміром 3×3 .

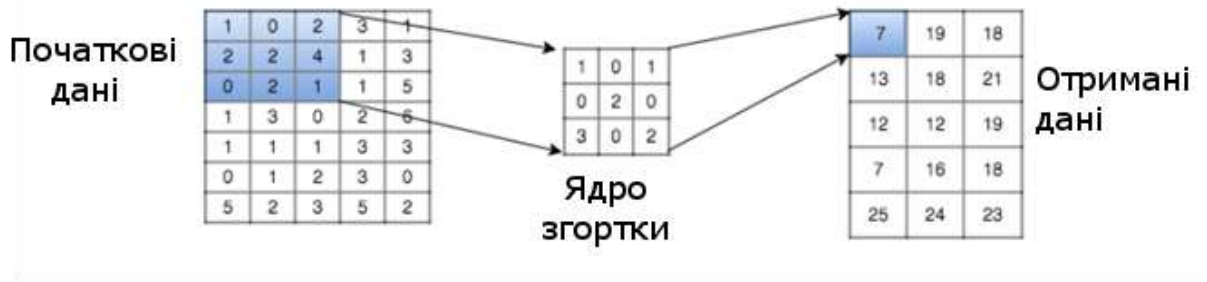


Рисунок 1.3 – Схема роботи згорткового шару

Багато праць присвячено розробці методів навчання з учителем, які тренують всю ієрархію згорткових фільтрів Θ безпосередньо, щоб мінімізувати цільову функцію втрат вищого рівня [7].

У галузі машинного зору набули поширення підходи, основані на застосуванні принципів лінгвістики до аналізу зображення за його фрагментами, які асоціюють з візуальними словами. При цьому визначення координат таких фрагментів відбувається за результатами роботи детектора найбільш помітних фрагментів (ключових точок), які характеризуються стійкістю до геометричних та фотометричних перетворень (рис. 1.4а), або вибираються з незмінним кроком, покриваючи все зображення (рис. 1.4б).



Рисунок 1.4 – Приклад роботи екстрактора візуальних фрагментів :

a – інваріантний до масштабу, повороту і освітленості детектор локальних ознак; *б* – розбиття зображення на патчі однакового розміру з заданим кроком

Як детектори особливих точок часто використовуються алгоритми Harris-Laplace, Hessian-Laplace, Fast Hessian, SUSAN та DoG, серед яких найбільшою оперативністю характеризується Fast Hessian [5]. При цьому фрагменти зображення для більш компактного подання описують числовим вектором, який називають дескриптором локальної ознаки.

Дескриптори локальних ознак поділяють на дійсночислові SIFT (Scale-invariant feature transform) і SURF (Speeded up robust features) та бінарні MSER (Maximally Stable Extremal Regions), BRIEF (Binary Robust Independent Elementary Features), DAISY (A Fast Local Descriptor for Dense Matching), BRISK (Binary Robust Invariant Scalable Keypoints), FREAK (Fast Retina Keypoint) і AKAZE (Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces) [6], [7]. З точки зору відношення точності та оперативності найбільш ефективними є алгоритми SURF та FREAK. Проте досі не існує формалізованих правил щодо вибору значень порогових параметрів, що відповідають за врахування чи відкидання локальних ознак, та параметрів, що впливають на формування глобального дескриптора зображення.

Згідно положення про часткове навчання з учителем, дослідники розглядають різноманітні можливості тренування параметрів Θ , пов'язаних з формуванням ознакового подання зображення, без учителя на вибірках не розмічених зразків [5]. При цьому ознакове подання спостереження x на кожному рівні ієрархії формується відповідним кодером $\Phi(x; \Theta)$ для компактного кодування інформації (рис. 1.5) та оператором пулінгу для агрегації окремих кодів.

Алгоритм машинного навчання без вчителя приймає на вхід множину нерозмічених зразків $\{x^{(i)} \mid i = \overline{1, n}\}$ і формує на виході натреновану множину параметрів Θ , що кодує знання про розподіл даних в $\{x^{(i)} \mid i = \overline{1, n}\}$. Отриману інформацію можна використати для кодування розмічених зразків, що подаватимуться на вхід алгоритму навчання з учителем, або використати для формування одного шару L-шарової моделі аналізу даних :

$$y \approx f(\Phi^{(L)}(\Phi^{(L-1)}(\dots\Phi^{(1)}(x; \Theta^{(1)}); \dots\Theta^{(L-1)}); \Theta^{(L)}); \theta),$$

де $\Theta^{(1)}, \dots, \Theta^{(L)}$ – параметри кожного рівня (шару) подання даних;

$\Phi^{(l)}(\cdot)$ – нелінійна функція перетворення простору ознак, яка приймає на вхід ознаки нижчого рівня і перетворює їх в ознаки вищого рівня;

θ – ознакове подання спостереження.



Рисунок 1.5 – Стандартний шаблон побудови широкого сімейства алгоритмів навчання ознакового подання без вчителя

Оператор пулінгу використовується для агрегації ознакового представлення окремих просторових регіонів зображення в єдине ціле компактне подання, тобто він приймає на вхід декілька кодів, що відповідають сусіднім фрагментам спостереження з заданої області, які можливо перетинаються, і повертає узагальнений вектор фіксованої довжини. На практиці набули поширення використання операторів Max pooling та Average pooling.

Average pooling обчислює середнє значення кодів по регіону i є методом агрегації у фреймворках типу “мішок візуальних слів” (Bag of Visual Features):

$$h_m = \frac{1}{|N_m|} \sum_{i \in N_m} \phi_i$$

де ϕ_i – код фрагменту в області m ;
 N_m – набір розташувань/індексів кодів в області m .

На рис. 1.6 проілюстровано використання стратегії Average pooling для формування ознакового опису зображення у вигляді гістограми частот появи візуальних слів (Bag of Visual Words). При цьому вектор ознак розширено гістограмами, отриманими при поділі зображення на частини для просторової деталізації опису.

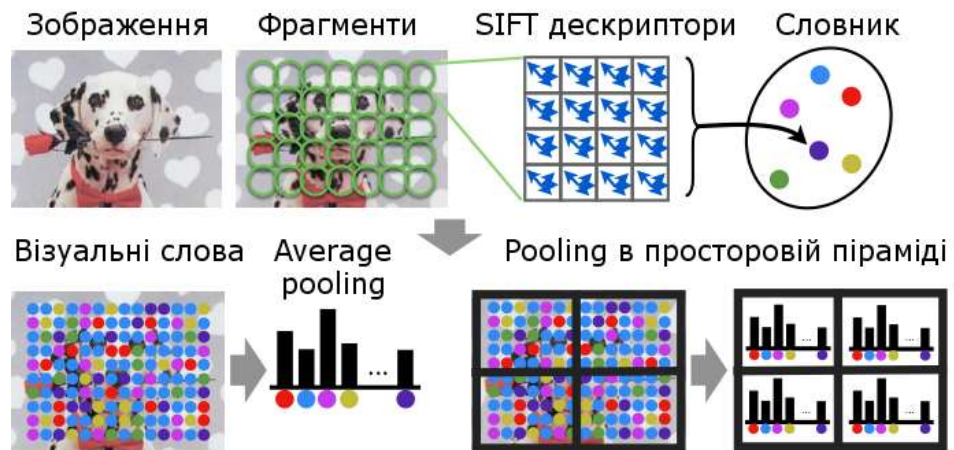


Рисунок 1.6 – Схема застосування Average pooling для побудови результуючого ознакового опису зображення

Оператор Max pooling обчислює максимум кожного компонента замість його середнього значення:

$$h_{m,j} = \max_{i \in N_m} \phi_{i,j}, \quad j = \overline{1, K},$$

де K – кількість компонентів у вихідному векторі ознак.

Оператор max-pooling є важливим компонентом технології згорткових нейронних мереж, реалізуючи субдискретизуючі (підвибіркові) шари, що виконують зменшення розмірності (як правило в кілька разів) – вся карта ознак поділяється на осередки, з яких вибираються максимальні за значенням. На рис. 1.7 показано приклад субдискретизуючого шару, оснований на виборі максимального елемента.



Рисунок 1.7 – Схема роботи субдискретизуючого шару (max-pooling) в згортковій нейронній мережі

Основне призначення субдискретизації — знизити варіативність даних, забезпечуючи стійкість до трансляцій локальної ознаки в межах окремого регіону. Тобто факт наявності ознаки важливіший знань місця його точного положення на зображенні. Таким чином мережа забезпечує інваріантність до просторових викривлень.

У працях [5]–[7] для аналізу даних високої розмірності з невеликим обсягом розмічених зразків пропонується використовувати навчання ознак без учителя на повному обсязі даних, а навчання класифікатора функціональних станів здійснювати на розмічених зразках закодованих навченими ознаками.

Найпростіший метод для навчання згорткового шару при навчанні без учителя це фрагментне (patch-based) навчання: коли класичному алгоритму виділення ознак без учителя подаються на вхід випадкові фрагменти з різних місць вхідних даних. Отриманий детектор ознак може послідовно просканувати екземпляр вхідних даних для отримання на виході згорткової карти ознак. Ця карта може бути використана як новий вхід для наступного шару. Операція може повторитися, в результаті чого можна навчити і з'єднати один за одним декількох шарів. Такий метод використовувався в аналізі незалежних підпросторів Independent Subspace Analysis (Le et al., 2011c) на 3D відео блоках, досягаючи кращих результатів на наборах даних з розпізнавання подій в Hollywood2, UCF, KTH и YouTube [6]. При цьому існує можливість покращення результатів фрагментного навчання згорткових нейронних мереж шляхом використання алгоритму зворотного поширення помилки в процесі мінімізації функції втрат найвищого шару вирішальних правил, хоча це питання малодосліджене. Досі

залишається відкритим питання вибору алгоритму фрагментного навчання без вчителя та архітектури згорткової мережі для забезпечення максимальної достовірності та оперативності рішень в режимі екзамену та навчання.

Важливим кроком аналізу даних є їх попередня нормалізація з метою усунення лінійної кореляції компонент спостереження і уніфікації первинного ознакового подання. Відбілювання даних, зокрема при фрагментному навчанні, за методом ZCA (Zero-phase Component Analysis) є одним з найпоширеніших методів попередньої нормалізації даних. Метод ZCA полягає у виконанні наступних кроків:

- 1) обчислення середнього вибіркового значення ознак $\mu := \text{mean}(X)$;
- 2) обчислення коваріаційної матриці вибірових спостережень $\Sigma := \text{cov}(X)$;
- 3) сингулярний розклад коваріаційної матриці $\Sigma \approx VDV^T$;
- 4) відбілювання кожного спостереження за формулою $x_j := VD^{-1/2}V^T(x_j - \mu)$.

На даний момент існує велика кількість методів навчання без вчителя, які потенційно можуть бути використані для фрагментного навчання, серед яких різноманітні алгоритми кластер-аналізу, розрідженого кодування, випадкового вибору фрагментів (Randomly sampled patches, RP) чи випадкової ініціалізації вагових коефіцієнтів (Random weights, R) [6]–[8].

Значного поширення набув підхід до кластеризації даних, який полягає у визначенні «точок згущення», використовуючи аналогію з «центром ваги». До найбільш відомих представників цієї групи алгоритмів відносяться алгоритми FOREL, K-means, K-medoids, K-modes та IsoData [6]–[8]. Результатом роботи алгоритму є $\Theta = (D)$, де D – матриця, колонки якої $D^{(k)}$, $k = \overline{1, K}$ є центроїдами кластерів. Такі алгоритми характеризуються відносно високою оперативністю виділення компактних сукупностей точок в багатовимірному просторі ознак. При цьому вони придатні для кластеризації, як великого, так і малого набору даних. Однак, алгоритми цієї групи, внаслідок жорсткої конкурентної схеми навчання досить чутливі до початкового наближення стартових параметрів,

характеризуються повільною збіжністю і пристосовані до використання гіперсферичних контейнерів класів розпізнавання, що не адекватно у випадку складних витянутих розподілів реалізацій образів. Крім того, застосування при виборі початкових центроїдів таких традиційних стратегій, як принцип максимальної віддаленості, рівномірного розподілу вздовж діагоналі гіперкубу даних та випадкової ініціалізації, не гарантує стабільності результатів кластеризації [9]. Тому на практиці доводиться здійснювати цілий ряд розбиттів при різних стартових параметрах, включаючи і кількість кластерів, і обирати розбиття, що максимізує критерій якості.

У працях [8]–[10] розвинуто ймовірнісний підхід до кластер-аналізу даних на прикладі алгоритму розщеплення Гаусових розподілів (Gaussian mixture model, GMM), який полягає в розщепленні параметрично заданих одномодальних Гаусових ймовірнісних розподілів векторів-реалізацій. Розщеплення здійснюється на основі припущення про кількість кластерів K і закон ймовірнісного розподілу в процесі визначення найбільш правдоподібних значень параметрів закону розподілу для кожного з кластерів. Остаточне розбиття здійснюється за правилом Байеса. Результатом алгоритму є набір параметрів $\Theta = (\mu, \Sigma, \pi)$, що представляють середні значення, коваріації, та апіорні ймовірності відповідно для кожного компонента суміші. При цьому в алгоритмі, як правило, вводять обмеження, щоб коваріаційна матриця Σ була діагональною. Однак подібні алгоритми непридатні для простору ознак високої розмірності особливо за умов малого обсягу набору даних. Крім цього, для алгоритмів в рамках ймовірнісного підходу характерні висока обчислювальна трудомісткість та проблематичність визначення оптимальної потужності розбиття простору ознак на кластери.

Б. Ольшозен та інші дослідники запропонували алгоритм розрідженого кодування, інспірований властивостями клітин зорової кори головного мозку (так звані “прості клітини” зони V1), який здатний ефективно представляти внутрішню структуру фрагмента зображення [6], [7]. Даний алгоритм представляє дані у вигляді надповного базису векторів, кожен з яких, таким чином, не є лінійно незалежним від інших. Тобто у випадку розрідженого подання, отримане одним із

алгоритмів, базисні компоненти можуть бути нелінійними, інкапсулюючи більш складні функції подання даних. Для вхідного зображення $x \in R^n$ розріджене подання формується зваженою сумою таких векторів $D^{(k)}, k = \overline{1, K}$

$$x = \sum_{k=1}^K D^{(k)} z_k,$$

де $K > n$ (умова надповного базису).

Розрідженість отриманого подання забезпечується тим, що для окремо взятого зображення, більшість компонентів будуть мати коефіцієнт, що дорівнює нулю. Ця умова мотивована тим фактом, що природні зображення, як правило, можуть бути представлені за допомогою комбінації невеликого числа ненульових компонентів-примітивів, таких як краї або межі (в області алгоритмів розрідженого кодування відповідні компоненти зветься «атомами» або «кодовими словами»). Таким чином, розріджене кодування забезпечує великий набір компонентів, які можуть значно відрізнятися один від одного, при цьому гарантуючи, що окремо взяте зображення буде представлено за допомогою суми всього лише деяких з них (рис. 1.8).

Оскільки розріджене подання все ще має апроксимувати вхідні дані досить ефективним для розпізнавання чином, на відповідний базис накладається додаткова умова максимальної відповідності. Таким чином, для вибірки з m зображень компоненти можуть бути отримані за допомогою мінімізації наступної функції [5]

$$J(x, D, z) = \sum_{j=1}^M (x^j - \sum_{k=1}^K D^{(k)} z_k^j)^2 + \lambda \sum_k^K S(z_k^j),$$

де $S(z_k^j)$ – функція умови розрідженості, що обмежує z_k^j таким чином, щоб більша частина значень аргументів була рівною нулю, наприклад, її роль можуть виконувати норми L_1, L_2 , або функція

$$S(z_k^j) = \log(1 + (z_k^j)^2).$$

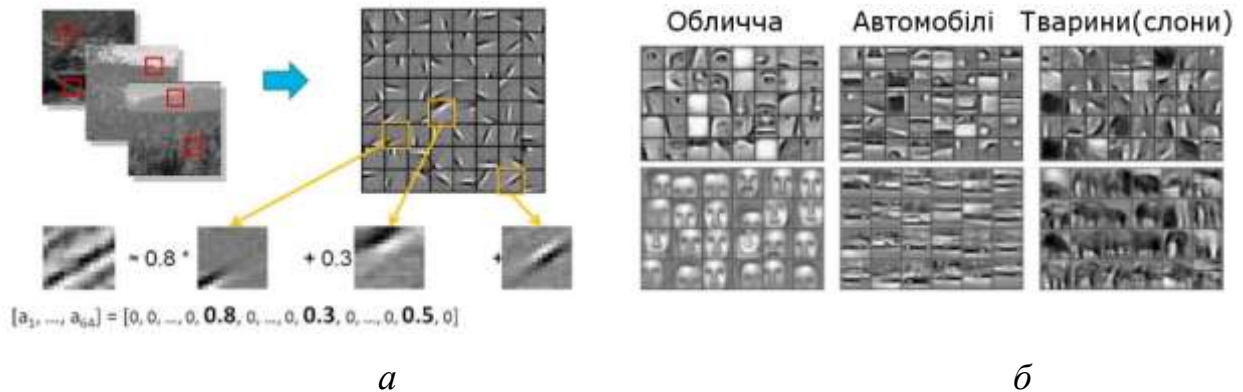


Рисунок 1.8 – Ілюстрація принципів розрідженого кодування: а – схема кодування фрагменту зображення розрідженим кодом; б – приклади компонент для трьох категорій об'єктів (на різних рівнях) [6]

Існують різні алгоритми пошуку розрідженого коду для вибірки зображень, таких як ортогональне узгоджене переслідування, регресія найменшого кута [5] і використання специфічних нейронних мереж – розрідженого автоенкодера (Sparse autoencoder, SAE) (рис. 1.9а), розрідженої обмеженої машини Больцмана (Sparse restricted Boltzmann machine, SRBM) (рис. 1.9б) [7].

Автоенкодер (автоасоціатор) спеціальна архітектура нейронних мереж, що дозволяє застосовувати навчання без учителя при використанні метода зворотного поширення помилки. Найпростіша архітектура автоенкодера приведена на рис. 1.9а — мережа прямого поширення, без зворотних зв'язків, найбільш схожа на перцептрон і містить вхідний шар, проміжний шар з K нейронів та вихідний шар. Головна мета навчання автоенкодера — досягти того, щоб вхідний вектор ознак викликав відгук мережі, рівний вхідному вектору. Тобто задача функціонування автоенкодера зводиться до знаходження апроксимації такої функції, щоб відгук нейронної мережі був рівним значенню вхідних ознак з точністю до заданого значення помилки. При цьому під час навчання слід внести обмеження, що забезпечує розріджену активація нейронів прихованого шару.

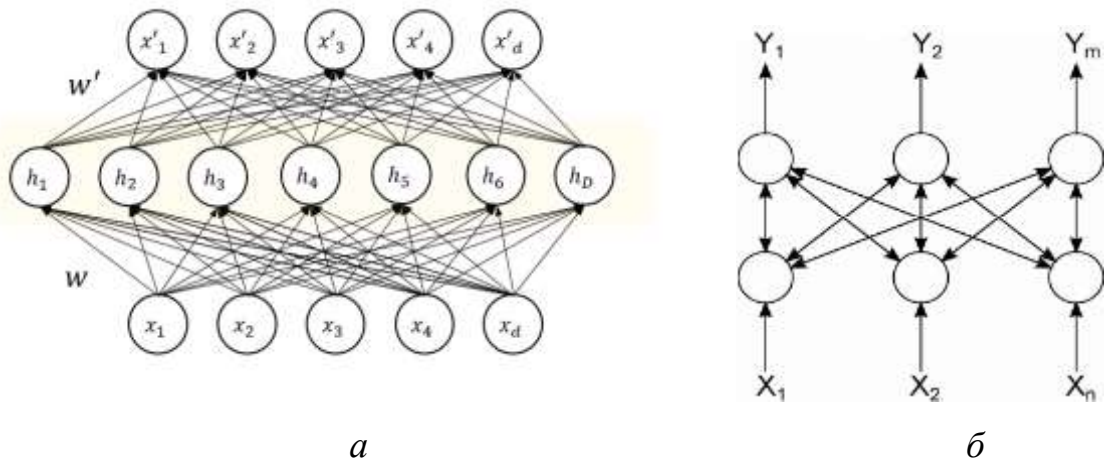


Рисунок 1.9 – Нейромережіві моделі для розрідженого кодування:

a – автоенкодер: *б* – обмежена машина Больцмана

Результатом роботи алгоритму є набір параметрів $\Theta = (W, b^{(1)}, b^{(2)})$, де $W \in R^{K \times n}$ – матриця вагових коефіцієнтів, $b^{(1)} \in R^K$, $b^{(2)} \in R^n$ – зміщення для кодуючого та декодуючого шарів відповідно.

Обмежена машина Больцмана є породжуючою стохастичною нейронною мережею, яка навчається формувати деякий ймовірнісний розподіл своїх входів. Обмежена машина Больцмана представляє модифікацію класичних машин Больцмана, які, в свою чергу, є варіаціями мереж Хопфілда [5]. Обмежена машина Больцмана складається з видимого і прихованого шару. В обмежених машинах Больцмана нейрони прихованого шару є детекторами ознак, які виділяють закономірності вхідних даних. Кожен нейрон має двонаправлений зв'язок з іншими в сусідньому шарі (рис. 1.9б). Обмежена машина Больцмана може генерувати (подавати) будь-який дискретний розподіл, якщо використовується достатня кількість нейронів прихованого шару.

Ключова ідея навчання полягає у відновленні розподілу вхідних даних на основі станів нейронів прихованого шару як можна точніше. Це еквівалентно максимізації функції логарифмічної правдоподібності розподілу вхідних даних $P(x)$ шляхом модифікації синаптичних зв'язків нейронної мережі. Використовуючи цей підхід Хінтон запропонував використовувати метод

контрастної дивергенції (Contrastive divergence, CD) для навчання RBM. Він базується на семплюванні Гіббса [5]. При цьому вводиться штраф за нерозрідженість, аналогічно навчанню автоенкодерів.

Основним недоліком нейромережових підходів є необхідність великого обсягу навчальних зразків, обчислювальних ресурсів. При цьому є необхідність значного нарощування глибини моделі та ускладнення алгоритмів машинного навчання при наявності в одному класі об'єктів зі значними візуальними відмінностями та суттєвими змінами вигляду тривимірного об'єкта при геометричних трансформаціях і кутах зору.

Після навчання без вчителя параметрів моделі подання даних Θ за нерозміченими фрагментами X здійснюється кодування кожного фрагменту x розміченої навчальної вибірки у результуючий вектор ознак. Серед методів кодування, що реалізують функцію $\Phi(x; \Theta)$, можна виділити техніку м'якого призначення (Soft assignment), грубого призначення (Hard assignment), розрідженого кодування (Sparse coding, SC), ортогонального узгодженого переслідування (Orthogonal matching pursuit, OMP-1), сигмоїдна функція (Sigmoid), триангуляція (Triangle) та м'який поріг (Soft threshold) [5].

Таким чином, формування ознакового опису аеровідеоспостережень полягає у виборі алгоритмів навчання параметрів моделі подання даних та виборі алгоритму кодування вхідних даних у вихідний вектор ознак. При цьому оператор просторового пулінгу дозволяє агрегувати коди окремих фрагментів зображення у єдину карту ознак вищого рівня подання. Узагальнюючі властивості відомих моделей подання даних залежать від глибини моделі, обсягу навчальних зразків, однак питання зниження ресурсних потреб для функціонування таких моделей є активною областю досліджень.

1.2 Методи, моделі та алгоритми ідентифікації об'єктів інтересу на місцевості

Основним компонентом інтелектуальної системи аналізу даних є вирішальні

правила, що часто представляють із себе класифікатор ознакового опису спостережень. Задачу класифікації можна визначити як задачу визначення класу об'єкта за набором заданих формальних ознак.

В задачах класифікації схожі об'єкти частіше знаходяться в одному класі, ніж в інших. Тому часто користуються гіпотезою компактності, яка припускає, що класи утворюють компактно локалізовані підмножини в просторі об'єктів. При цьому для формалізації поняття «схожості» вводиться функція відстані або метрика $d(x, y)$ в N -вимірному просторі об'єктів. Алгоритми, основані на аналізі схожості об'єктів, часто називають метричними, навіть у тих випадках, коли функція d не задовольняє всім аксіомам метрики (наприклад, аксіомі трикутника).

Алгоритм найближчого сусіда (nearest neighbor, NN) є найбільш простим метричним алгоритмом, який класифікує об'єкт як реалізацію того класу, якому належить найближчий об'єкт навчальної вибірки. Навчання такого класифікатора зводиться до елементарного запам'ятовування навчальної вибірки $\{y_{m,i}^{(j)} \mid m = \overline{1, M}; j = \overline{1, n}; i = \overline{1, N}\}$, де M – потужність алфавіту класів розпізнавання, n – обсяг спостережень класу X_m^o , N – кількість ознак розпізнавання. Єдиною перевагою цього алгоритму є простота реалізація. Однак недоліків набагато більше. По-перше, наявності викидів у навчальній вибірці призводить до нестійкості і похибок. По-друге відсутні параметри, які можна б було налаштувати за навчальною вибіркою. Алгоритм повністю залежить успішності вибору дистанційної міри $d(x, y)$.

Алгоритм k -найближчих сусідів (k -nearest neighbors, kNN) з метою згладжування шумового впливу викидів, буде класифікувати об'єкти шляхом голосування за k найближчими сусідами. Кожен із сусідів $\{y^{(j)} \mid j = \overline{1, k}\}$ голосує за віднесення об'єктів до свого класу. Алгоритм відносить вхідний об'єкт до того класу, який набере більше число голосів. Оптимальне значення параметра k визначається за критерієм ковзного контролю з виключенням об'єктів по одному (leave-one-out, LOO). Для кожного об'єкта перевіряється чи правильно він класифікується за своїми k найближчими сусідами.

Міри близькості обирають виходячи з властивостей об'єктів. Найбільшого поширення набули такі метрики, як Евклідова, Манхеттенська, Степенева та Журавльова [9].

Міра близькості підбирається індивідуально для конкретних типів даних. Інколи адекватної міри близькості підібрати не вдається, і доводиться обирати її евристично.

Метричні алгоритми здійснюють локальну апроксимацію вибірки, при якій обчислення відкладаються до моменту, доки не стане відомий вхідний об'єкт. Тому метричні алгоритми відносяться до методів лінивого навчання (lazy learning).

Суть статистичних методів навчання полягає у відновленні роздільної функції шляхом мінімізації середнього ризику помилкового прийняття рішення [6]. Статистичні методи дозволяють побудувати вирішальні правила у випадках перетину класів розпізнавання, що має місце в практичних задачах контролю та керування слабо формалізованими процесами.

Одним із класичних статистичних методів класифікації є метод Байєса [5], відповідно до якого прийняття класифікаційних рішень здійснюється шляхом знаходження максимальної апостеріорної умовної ймовірності $p(X_m^o/x)$, що обчислена для заданого алфавіту класів розпізнавання $\{X_m^o | m = \overline{1, M}\}$ за формулою

$$p(X_m^o/x) = \frac{P(X_m^o)p(x/X_m^o)}{\sum_{k=1}^M P(X_k^o)p(x/X_k^o)},$$

де $P(X_m^o)$ – безумовна ймовірність появи класу X_m^o .

$p(x/X_m^o)$ – значення функції правдоподібності (щільності розподілу ймовірностей) класу X_m^o для вхідної реалізації x .

Безумовна ймовірність появи класу X_m^o визначається як відношення числа реалізацій, що належать класу X_m^o , до загальної кількості реалізацій

$$P(X_m^o) = \frac{\text{count}(x^{(j)} \in X_m^o)}{n},$$

де $\text{count}(x^{(j)} \in X_m^o)$ – число реалізацій навчальної вибірки, що належать класу X_m^o ;

n – загальна кількість реалізацій образів у навчальній вибірці.

Значення функції правдоподібності класу X_m^o для реалізації x при статистичній незалежності ознак роз-пізнавання обчислюється за формулою

$$p(x / X_m^o) = \prod_{i=1}^N p(x_i / X_m^o),$$

де $p(x_i / X_m^o)$ – значення щільності розподілу ймовірностей i -ї ознаки в класі X_m^o для вхідної реалізації x .

Щільності розподілу ймовірностей $p(x_i / X_m^o)$ можуть бути оцінені в рамках припущення про тип розподілу. Наприклад, може використовуватися гіпотеза про нормальний закон розподілу

$$p(x_i / X_m^o) = \frac{1}{\sigma_{m,i} \sqrt{2\pi}} \exp\left(-\frac{(x_i - \bar{x}_{m,i})^2}{2\sigma_{m,i}^2}\right),$$

де $\sigma_{m,i}^2$ – дисперсія i -ї ознаки в класі X_m^o ;

$\bar{x}_{m,i}$ – математичне очікування i -ї ознаки в класі X_m^o .

Методи розпізнавання, основані на використанні формули Байєса і гіпотези про незалежність ознак, як правило, називають найвними байєсовськими класифікаторами. Перевагою байєсовських класифікаторів є простота реалізації алгоритмів класифікації, а недоліком – низька достовірність класифікації реалізацій образів у випадку обмеженого обсягу вибірок та перетину в просторі

ознак класів розпізнавання. При цьому при побудові байєсовського класифікатора під етапом навчання розуміється набір статистичних даних.

Основними недоліками статистичних методів, які обмежують їх використання на практиці, є необхідність великих обсягів статистики для апроксимації функції щільності розподілу імовірностей, виконання жорстких умов для забезпечення статистичних стійкості та однорідності та висока чутливість до репрезентативності навчальних вибірок.

У рамках найбільш універсального підходу до аналізу і синтезу здатних навчатися систем прийняття рішень, яким є геометричний підхід, набув популярності метод опорних векторів SVM (support vector machine) [7]. В основі SVM лежить ідея розділення згущення векторів гіперплощинами, що знаходяться на максимальній відстані від згущень при мінімізації зміщення реалізацій класу від опорного вектора (рис. 1.10).

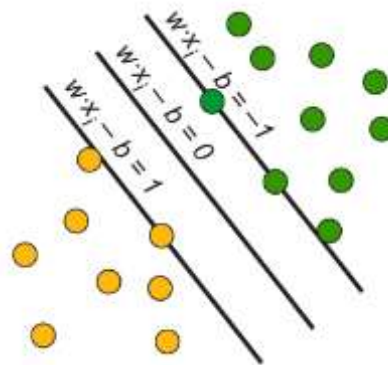


Рисунок 1.10 – Лінійне розділення образів в SVM

Задачу синтезу вирішальних правил SVM зручно формулювати для випадку двохкласової класифікації. При цьому багатокласова класифікація, як правило, зводиться до серії двокласових класифікацій. Прямі багатокласові узагальнення SVM існують, проте навчити один складний багатокласовий класифікатор набагато довше, ніж багато простих однокласових.

Проблема лінійної нероздільності в методі SVM вирішується шляхом переходу від початкового простору ознак до нового розширеного простору за допомогою деякого перетворення. На рис. 1.11 показано приклад переходу від

двовимірного простору до розширеного трьохвимірного шляхом переносу точок на поверхню сфери.

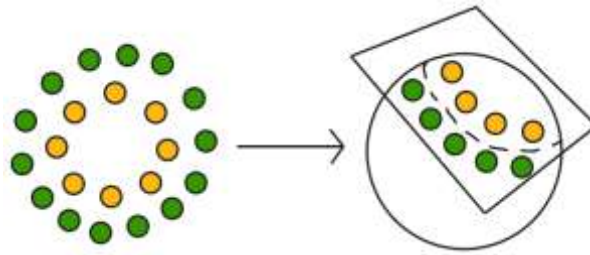


Рисунок 1.11 – Трансформація простору ознак в SVM

Аналіз рис. 1.11 показує в новому просторі більшої розмірності вдалося розділити точки вибірок двох класів площиною.

Вирішальне правило двох класового SVM-класифікатора має вигляд

$$a(x) = \text{sign} \left[\sum_j^n \lambda_j c_j K(x, x^{(j)}) - \rho \right],$$

де ρ – параметр порогу;

λ – вектор коефіцієнтів, ненульові значення компонентів якого відповідають опорним векторам навчальної вибірки;

c_j – мітка класу j -го вектора навчальної вибірки,

яка приймає значення із множини $\{-1, +1\}$;

$K(x, x^{(j)})$ – функція ядра, що відповідає скалярному добутку векторів в деякому розширеному просторі ознак

$$K(x, x^{(j)}) = \phi(x)^T \phi(x^{(j)}),$$

де $\phi(x^{(j)})$ – функція відображення вектора $x^{(j)}$ в розширений простір, де забезпечується лінійна розділимість класів.

Навчання SVM зводиться до пошуку коефіцієнтів λ_j (множники Лагранжа) та порогу ρ , знайти які можна вирішивши таку задачу квадратичної оптимізації

$$L(\lambda) = \sum_{j=1}^n \lambda_j - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j c_i c_j K(x_i, x_j) \rightarrow \max_{\lambda}$$

за умови

$$\begin{cases} \sum_{j=1}^n \lambda_j c_j = 0; \\ 0 \leq \lambda_j \leq C, \end{cases}$$

де C – константа регуляризації, що, як правило, обирається рівною

$$C = \frac{1}{nv}$$

де v – максимальна частка векторів навчальної послідовності, які можуть бути викидами.

Описана задача оптимізації вирішується евристичними алгоритмами шляхом послідовного зменшення цільової функції.

Для вибору функції ядра використовують теорему Мерсера (Мерсер, 1909 [8]): функція $K(x, x')$ є ядром тільки тоді і тільки тоді, коли вона симетрична, $K(x, x') = K(x', x)$, і невід'ємно визначена. Однак вибір ядра для конкретних вхідних даних досі залишається нетривіальною задачею. Повний перебір всіх можливих ядер, ускладнений повільністю навчання SVM.

На усунення недоліків відомих методів машинного навчання спрямовано інформаційно-екстремальну інтелектуальну технологію (ІЕІ-технологію), яка ґрунтується на максимізації інформаційної спроможності системи розпізнавання шляхом введення на етапі навчання додаткових інформаційних обмежень. Основні концептуальні положення ІЕІ-технології такі:

- ІЕІ-технологія ґрунтується на прямій оцінці інформаційної спроможності системи, що навчається;
- ІЕІ-технологія дозволяє оптимізувати просторово-часові параметри функціонування системи, що навчається;
- прийняття рішень у рамках ІЕІ-технології здійснюється в рамках детерміновано-статистичного підходу шляхом побудови відносно простого детермінованого класифікатора, статистична корекція якого здійснюється в процесі навчання з метою побудови безпомилкових, за навчальними матрицями, вирішальних правил;
- методи ІЕІ-технології ґрунтуються на застосуванні гіпотез як чіткої, так і нечіткої компактності реалізацій образу, тобто є працездатним за умови перетину класів розпізнавання, що має місце в практичних задачах контролю та керування;
- методи ІЕІ-технології базуються на вибіркового підході математичної статистики і орієнтовані на застосування прийнятних з практичних міркувань обсягів репрезентативних навчальних вибірок.

Побудова «точного» контейнера класу розпізнавання складної геометричної форми у багатовимірному просторі ознак навіть для сучасних комп'ютерних комплексів має суттєві ускладнення. Тому при обґрунтованості гіпотези нечіткої компактності в працях [7]–[9] згідно з принципом редукції [7] у рамках детерміновано-статистичного підходу пропонується відновлювати контейнери спрощеної форми, наближеної до «точної» деяким оптимальним способом, і які формують вирішальні правила, що забезпечують в режимі екзамену достовірність прийняття рішень, близьку до максимальної асимптотичної.

В основі інформаційно-екстремального машинного навчання лежать методи адаптивного двійкового кодування високорівневого ознакового опису для побудови в бінарному просторі Хеммінга оптимальних в інформаційному і обчислювальному сенсах радіально-базисних вирішальних правил. На відміну від методу опорних векторів в даному методі перетворення простору здійснюється на основі обчислювально ефективних операцій порівняння і виключаючого АБО (рис. 1.12а). При цьому кодування вектор ознак в рамках ІЕІ-технології полягає у

порівнянні значення i -ї ознаки з відповідним нижнім $A_{B,l,i}$ та верхнім $A_{T,l,i}$ межами несиметричного поля контрольних допусків l -го рівня здійснюється за правилом

$$x_{k,(l-1)*N+i}^{(j)} = \begin{cases} 1, & \text{if } A_{B,l,i} \leq y_{k,i}^{(j)} \leq A_{T,l,i}; \\ 0, & \text{else.} \end{cases}$$

де $y_{k,i}^{(j)}$ – значення i -ї ознаки j -ї вектора-реалізації k -го класу розпізнавання;
 N – кількість ознак розпізнавання;
 L – кількість рівнів контрольних допусків.

Застосування популяційного алгоритму оптимізації дозволяє здійснювати швидку регуляризацію моделі даних шляхом пошуку глобального максимуму інформаційного критерію в допустимій області його визначення. Як критерій оптимізації використовується модифікація інформаційного критерію С. Кульбака, подана у вигляді функціоналу від точнісних характеристик рішень – чутливості і специфічності (рис. 1.12б).

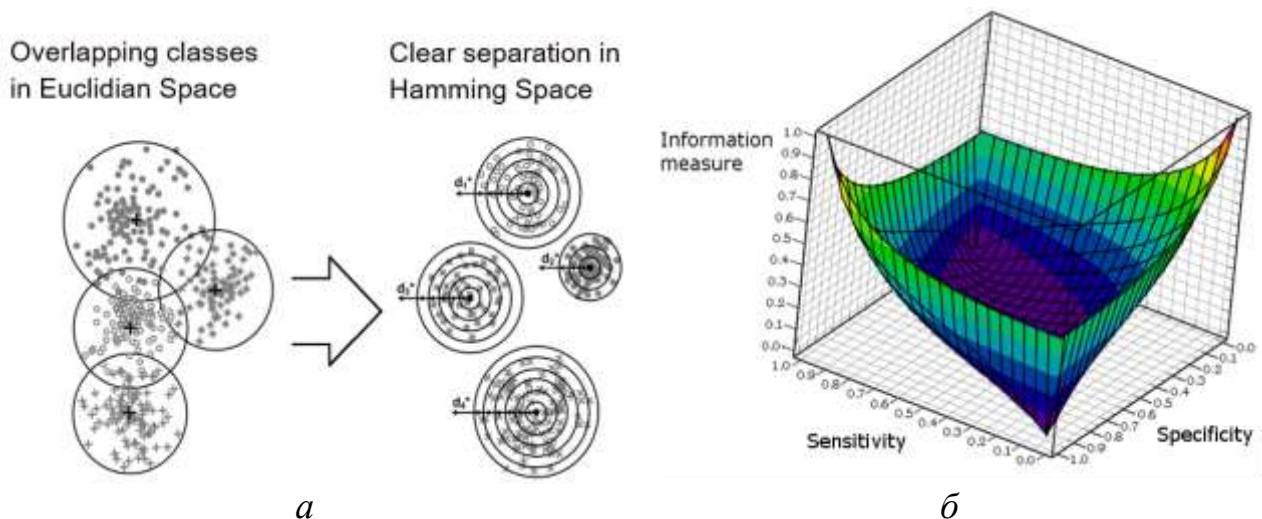


Рисунок 1.12 – Принципи інформаційно-екстремального машинного навчання вирішальних правил: а – перетворення простору ознак; б – залежність критерія ефективності від чутливості і специфічності вирішальних правил

Завдяки згладжуючому ефекту логарифмічної функції критерію вдається отримати вирішальні правила з високою узагальнюючою здатністю для випадку малих навчальних вибірок. Проте питання забезпечення ефективності застосування методу за умов незбалансованості вирішальних правил та питання оптимального вибору кількості рівнів контрольних допусків на значення ознак досі залишається недостатньо дослідженим.

Значної популярності останнім часом набули методи класифікаційного аналізу, основані на застосуванні багат шарових нейронних мереж, де вхідні значення нейрона шару l представляють собою вихідні значення нейронів попереднього шару $l-1$. При цьому нейрони першого (вхідного) шару отримують як вхідні значення безпосередньо дані, що підлягають розпізнаванню.

Вихідний шар мережі може варіюватися в залежності від завдання, але класична архітектура має на увазі формування його числом нейронів, рівній кількості класів розпізнавання, при цьому вихідне значення кожного нейрона нормується по інтервалу $[0,1]$, і являє собою ймовірність належності вхідного зображення до відповідного класу (рис. 1.13).

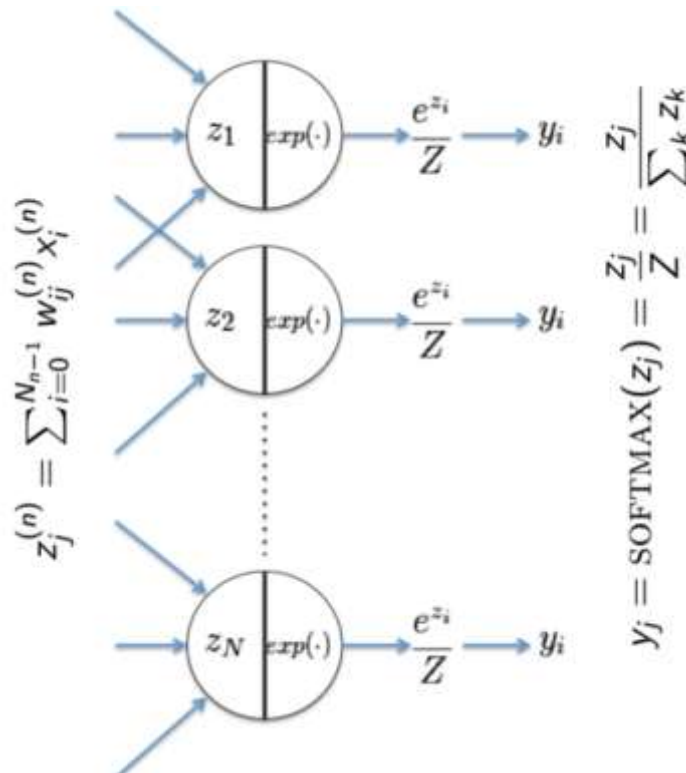


Рисунок 1.13 – Вихідний шар нейронної мережі

Навчання нейронних мереж відбувається за методом градієнтного або стохастичного градієнтного спуску, який ще називають алгоритмом зворотного поширення помилки в нейронні мережі. Однак ці алгоритми часто призводили до проблем затухаючого градієнту, паралічу нейронної мережі, перенавчання та застрягання в локальному мінімумі функції помилки. Тому в кінці 80-х років XX століття стало зрозуміло, що одного алгоритму зворотного поширення помилки недостатньо для ефективного глибокого навчання.

Незважаючи на деякі успішні приклади, використання більш одного прихованого шару рідко давало переваги на практиці [7]. Причина цього була сформульована в 1991 році в роботі [10] – проблема зникає градієнт (vanishing gradient problem). При використанні традиційних функцій активації, сигнали про обернено розповсюджуваних помилки швидко стають дуже маленькими (або навпаки, надмірно великими). У практичних завданнях вони зменшуються експоненціально з кількістю шарів в мережі. Ця проблема також відома як проблема тривалої затримки – (long time lag problem) [11].

У 2010 році Ксав'є Глорія (Xavier Glorot) і Юшуа Бенджі (Yoshua Bengio) в роботі [12] провели дослідження впливу методів початкової ініціалізації ваг і функцій активації на поширення сигналу в мережі як в прямому, так і в зворотному напрямку. Згідно з дослідженням, використання логістичної сигмоїдальної функції спільно з початковою ініціалізацією ваг погано підходить для створення глибоких нейронних мереж, тому що призводить до швидкого насичення. Функція активації гіперболічний тангенс такою проблемою не володіє через симетричності (середнє значення функції 0, область значення – (-1, 1)). Глорія і Бенджі запропонували новий метод ініціалізації ваг нейронної мережі, який вони назвали нормалізована ініціалізація. Початкові значення ваг мережі W визначаються за такою формулою:

$$W \approx U\left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}\right],$$

де U – рівномірний розподіл на відрізку;
 n_j – кількість нейронів на поточному шарі мережі;
 n_{j+1} – кількість нейронів на наступному шарі мережі.

Використання нормалізованої ініціалізації призводить до зниження насичення нейронів і сигнал про помилку поширюється значно краще.

Сергій Йоффе (Sergey Ioffe) і Крістіан Жегеди (Christian Szegedy) в 2015 році [8] запропонували використовувати в нейронних мережах спеціальні шари пакетної нормалізації (batch normalization), які дозволяють підвищити якість навчання глибокої нейронної мережі. У роботі [11] встановлено, що алгоритм зворотного поширення помилки сходиться швидше, якщо вхідні дані нормалізовані (мають нульове маточікування і одиничну дисперсію). Однак при поширенні сигналу по нейронній мережі його маточікування і дисперсія змінюються, причому іноді значно, що негативно позначається на процесі навчання. Сергій Йоффе і Крістіан Жегеди запропонували виконувати нормалізацію не тільки на вході в нейронну мережу, а й перед кожним шаром мережі. Нормалізація виконується окремо для кожного пакета даних (mini-batch) методу стохастичного градієнтного спуску (або його модифікацій). Пакет B містить m елементів вхідних даних $x_i : B = \{x_1, \dots, x_m\}$. Нормалізовані значення \hat{x}_i визначаються за наступною формулою:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}},$$

де μ_B – середнє значення даних в пакеті;
 σ_B^2 – дисперсія;
 ε – константа, введена для стабільності методу.

Потім, щоб зберегти виразність даних, виконується зсув і масштабування:

$$y_i = \gamma \hat{x}_i + \beta,$$

де y_i — результуюче значення;

γ і β — параметри, які визначаються у процесі навчання.

Пакетна нормалізація реалізується у вигляді шарів пакетної нормалізації, які можуть бути вставлені в потрібне місце в нейронну мережу, в тому числі кілька разів. Додатковою перевагою використання пакетної нормалізації є скорочення часу навчання та зниження перенавчання. Методи нормалізованої ініціалізації ваг і шари пакетної нормалізації допомагають на практиці впоратися з проблемою зникаючого градієнта і навчати глибокі нейронні мережі, що складаються з декількох десятків шарів. Це дозволило деяким авторам стверджувати, що проблема зникає градієнт в даний час вирішена [11].

При класифікаційному аналізі зображень треба звернути увагу, що деякі кадри відеопотоку можуть мати низьку різкість за рахунок поривів вітру і сейсмічних впливів на відеокамеру, а у випадку розміщення камери на літальному апараті — із-за складного характеру руху літального апарату в неоднорідному нестабільному повітряному середовищі. Тому для покращення результатів класифікаційного аналізу автори праці [8] попередньо пропонують виділяти найбільш різкі кадри відеопотоку і в подальшому визначати за ними належність об'єктів інтересу до того чи іншого класу розпізнавання.

Ідентифікація об'єктів інтересу на місцевості часто передбачає оцінку місця розташування даного об'єкта. Це можливо лише при реалізації локалізації і визначення меж зони зображення, де знаходиться об'єкт інтересу. Цей процес називається детектуванням.

Якщо камера відеоспостереження є нерухомою, то можна звужити область пошуку до ділянок, що містять рухомі об'єкти у відео послідовності для наступного розпізнавання класу об'єкта в даній зоні інтересу. Під детектуванням рухомих об'єктів розуміється виявлення об'єктів, що рухаються, з одночасним відсіювання помилкових спрацьовувань, викликаних зміною освітленості, фоновим рухом і тінями. Незважаючи на високий інтерес вчених і інженерів до даної проблеми вона не вирішена повною мірою.

Основні проблеми при виявленні руху виникають, якщо задній фон динамічного зображення є складним і неоднорідним, тобто. характеризується різною колірною гамою, наявністю природних або штучних перешкод, присутні рухомі об'єкти, які за кольором і (або) формі схожі з шуканим.

Іншим завданням, яке потребує вирішення при створенні системи автоматичного аналізу відео послідовності, є трекінг рухомих об'єктів, тобто відстеження переміщень об'єкта між кадрами і зіставлення положень цільового об'єкта на послідовності кадрів.

Варто зазначити, що процес детектування і трекінгу рухомих об'єктів оперує лише так званими «блобами» («blob»), тобто зв'язаними областями на кадрі, прийнятими за передній фон. Тому системи автоматичного аналізу потребують вирішення задачі класифікації об'єктів, для того щоб відсіяти об'єкти інтересу від інших об'єктів, які не становлять інтерес для системи інтелектуального відеоаналізу.

Для розпізнавання того, що відбувається на відео-послідовності, а саме, для розпізнавання людських ролей, станів, дій і діяльностей, низькорівнева обробка відео послідовності вимагає також вибору простору ознак, який міг би повною мірою описати характер руху об'єктів і їх властивості, а також отримати ці ознаки для виявлених об'єктів.

Для детектування рухомих об'єктів використовується метод вирахування фону, який здійснює відділення об'єктів переднього плану від статичного фону. Даний підхід показує найбільш оптимальне співвідношення точність / продуктивність.

Основне завдання при використанні даного методу – отримати модель статичного фону. Після отримання моделі фону, необхідно порівняти попіксельно поточний кадр з модельним. Регіони поточного кадру, які значно відрізняються від модельного, вважаються регіонами об'єктів переднього плану:

$$|I_t(X) - I_b(X)| > \tau,$$

де $I_t(x, y)$ і $I_b(x, y)$ означають інтенсивність пікселя X на поточному кадрі в момент часу t і його значення на статичному (змодельованому) фоні I_b відповідно;

τ – деякий поріг бінаризації, що задається як емпірично підібране значення або обчислене за методу Оцу.

Найбільш перспективним є побудова моделі фону за допомогою імовірнісних моделей. Такі алгоритми використовують нормальний розподіл або суміш нормальних розподілів для моделювання пікселя фону. Це дозволяє відокремлювати об'єкти переднього плану більш точно і справлятися зі зміною освітлення, динамічним фоном і т.д.

Для моделювання заднього фону часто використовується метод суміші Гауссіан (Mixture of Gaussians). Даний метод заснований на адаптивному статистичному моделюванні інтенсивності пікселів. Суміш з K Гауссіан (де K , як правило, вибирається в діапазоні від 3 до 5) використовується для побудови моделі заднього фону і детектування рухомих об'єктів. Для пікселя X в момент часу t , ймовірність спостереження поточного значення інтенсивності цього пікселя може бути записана як

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \cdot \eta(X_t, \mu_{i,t}, \Sigma_{i,t}),$$

де K – число розподілів;

$\omega_{i,t}$ – оцінка ваги (частина даних, що враховується цією гауссіаною) i -ї гауссіани в суміші в момент часу;

$\mu_{i,t}$ – середнє значення i -ї гауссіани в суміші в момент часу t ;

$\Sigma_{i,t}$ – матриця коваріації i -ї гауссіани в суміші в момент часу t ;

η – функція щільності ймовірності гауссіани

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)}$$

де n – число вимірів для інтенсивності пікселя X і

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha$$

де α – швидкість навчання.

Передбачається, що червоні, сині та зелені значення пікселів незалежні і мають однакові відхилення, тому вважаємо, що

$$\Sigma_{k,t} = \sigma_k^2 I_t$$

де I_t – послідовність кадрів.

Після впорядкування Гауссіан за значенням ω/α перші B розподілів обираються для моделі фону, де

$$B = \arg \min_b \left(\sum_{k=1}^b \omega_k > T \right),$$

де T — критерій мінімального обсягу даних, що враховується заднім фоном.

Параметри μ і σ для неспівпадаючих розподілів залишаються попередніми. Параметри розподілу, який співпадає з новим спостереженням, оновлюємо за наступним правилом:

$$\begin{aligned} \mu_t &= (1 - \rho)\mu_{t-1} + \rho X_t \\ \sigma_t^2 &= (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t), \end{aligned}$$

де $\rho = \alpha \eta(X_t | \mu_k, \sigma_k)$.

Для кожного кадру розглядається процес зміни пікселів, який представляється у вигляді часових рядів значень пікселів. У будь-який момент часу t для кожного пікселя $\{x_o, y_o\}$ відома його історія:

$$\{X_1, \dots, X_t\} = \{I_t(x_o, y_o, i) : 1 \leq i \leq t\},$$

При зміні освітленості у статичній сцені, необхідно впевнитися, що гауссіана буде відслідковувати ці зміни. Якщо статичний об'єкт був доданий в сцену і не був відображений в тлі до тих пір, поки він не знаходився там довше попереднього об'єкта, то відповідні пікселі можуть прийматися за пікселі переднього плану протягом досить тривалого періоду часу. Це призведе до накопичення помилки в оцінці переднього плану і неправильної сегментації. Ці фактори призводять до того, що недавні спостереження є більш важливими при визначенні оцінок параметрів Гауссіан.

Фонове зображення моделюється з використанням трьох Гауссіан. Якби кожен піксель походив від певної поверхні під певним освітленням, однієї Гауссіани було б достатньо для моделювання значення пікселя, навіть з врахуванням шумів. Якщо тільки освітлення змінюється з часом, однієї адаптивної Гауссіани для кожного пікселя також було б достатньо. На практиці, часто один піксель зображення може відноситися до різних поверхонь, і умови освітленості можуть змінюватися. Таким чином, необхідно використовувати суміш адаптивних Гауссіан для того, щоб апроксимувати цей процес.

З кожним разом параметри Гауссіан оновлюються, після чого Гауссіани оцінюються, використовуючи просту евристику, для того, щоб визначити, які з них найімовірніше є частиною заднього фону. Значення пікселя, яке не збігається з жодною із «фонових» Гауссіан групуються методом зв'язкових компонент. Нарешті, зв'язкові компоненти відслідковуються від кадру до кадру з використанням набору з декількох гіпотез. Перший Гауссовий розподіл показує стійкі пікселі і представляє фонове зображення. Другий Гауссовий розподіл

відображає повторювані відхилення і відносно стаціонарні області. Третій розподіл - пікселі з високою швидкістю зміни.

У випадку рухомої камери, наприклад камери безпілотного літального апарату чи відео реєстратора автомобіля, більшу ефективність матиме метод детектування, що здійснює сканування всього зображення повністю. При цьому найпростішим методом є алгоритм ковзного вікна (sliding window). Під вікном (або вікном обробки) розуміється деяка кінцева область зображення, при цьому її мінімальний розмір вибирається з контексту завдання, тобто визначається кількість пікселів зображення, достатня для детектування об'єкта. Після виділення вікна воно переміщається по зображенню, охоплюючи всі можливі положення, завдяки чому його можна назвати «ковзним». Порядок переміщення можна вибрати будь-яким, проте найчастіше пересування виконується зліва-направо і зверху-вниз. На кожному кроці кожного вікна присвоюється класифікатор, який видає результат: чи присутній об'єкт в цьому вікні чи ні (true або false відповідно). Якщо класифікатор – false, то поточна область вікна відкидається, і алгоритм переходить на наступний крок. Якщо відповідь – true, то це означає, що об'єкт знаходиться в даній області.

Однак, у цього методу є певний недолік. Проходження вікном всіх положень на зображенні на різних масштабах вимагає досить великої кількості часу $O(h * w)$, де h і w – висота і ширина зображення. Однією з модифікацій методу sliding window, створеної для поліпшення ефективності рішення задачі, є алгоритм RASW.

Метод адаптивного кроку ковзного вікна (англ. Run-Time Adaptive Sliding Window, RASW) був запропонований в роботі [8]. RASW – ефективний метод зміни кроку руху скануючого вікна під час обробки без втрати точності. Алгоритм дозволяє знизити витрати часу на обробку областей зображення, які не містять шуканих об'єктів.

Розмір кроку по двох осях може бути різний: dx і dy . Однак в простому випадку: $d = dx = dy$. Точність виявлення зменшується зі збільшенням кроку, але великий крок на «фоновій частині» зображення дозволяє зменшити кількість

помилкових спрацьовувань. Зазвичай кроки дорівнюють одному або двом пікселям, оскільки більш високі значення призводять до погіршення якості виявлення.

Алгоритм RASW має ще одну перевагу в порівнянні з класичним методом ковзного вікна зі статичним кроком. При $d > 1$ ковзне вікно може не потрапити в місце, де присутній шуканий об'єкт, а модифікований алгоритм зменшить крок до 1 пікселя і зупиниться в потрібному місці.

Прискорення можна досягти також шляхом застосування спочатку простих і швидких класифікаторів («сильний класифікатор»), а потім складних («слабкий класифікатор»), які дозволяють отримати задану точність розпізнавання. Сильний класифікатор відкидає області, де відсутні шукані об'єкти за менший час, а в областях, де сильний класифікатор вказав найбільшу ймовірність знаходження об'єкта, застосовується слабкий класифікатор. Цей методи був описаний в роботі [12], де використовується зовнішній класифікатор Selective Search, який визначає області інтересу на зображенні (Region of Interest, ROI) для подальшого детектування об'єктів [13]. В методі R-CNN після відсіву областей для обраних зон застосовується нейронна класифікаційна мережа, яка визначає клас і координати об'єкта.

Поліпшення методу R-CNN, описане в роботі [15], отримало назву Fast R-CNN, де пропонується подавати на вхід мережі повне зображення, але при цьому останній max-pool шар замінити на шар нового типу *RoI pooling*. RoI pooling шар приймає на вхід карту ознак, отриману від останнього згорткового шару нейронної мережі, і RoI претендента (в координатах зображення). RoI перетворюється з координат зображення в координати на карті ознак і на отриманий прямокутник накладається сітка $W \times H$ з наперед заданими розмірами (наприклад, для згорткової мережі VGG16 автори статті пропонують $W = H = 7$). Робиться max pooling по кожному осередку цієї сітки. Таким чином RoI pooling шар перетворює вектор ознак довільного прямокутника з вхідного зображення у вектор ознак фіксованої розмірності. Після RoI pooling шару дані через два повнозв'язних шари подаються паралельно на softmax шар для оцінки

приналежності даного претендента одному з класів об'єкта та на шар, що реалізує регресію, яка уточнює обмежувальний прямокутник об'єкта. Як результат отримуємо наступну схему (рис. 1.14).

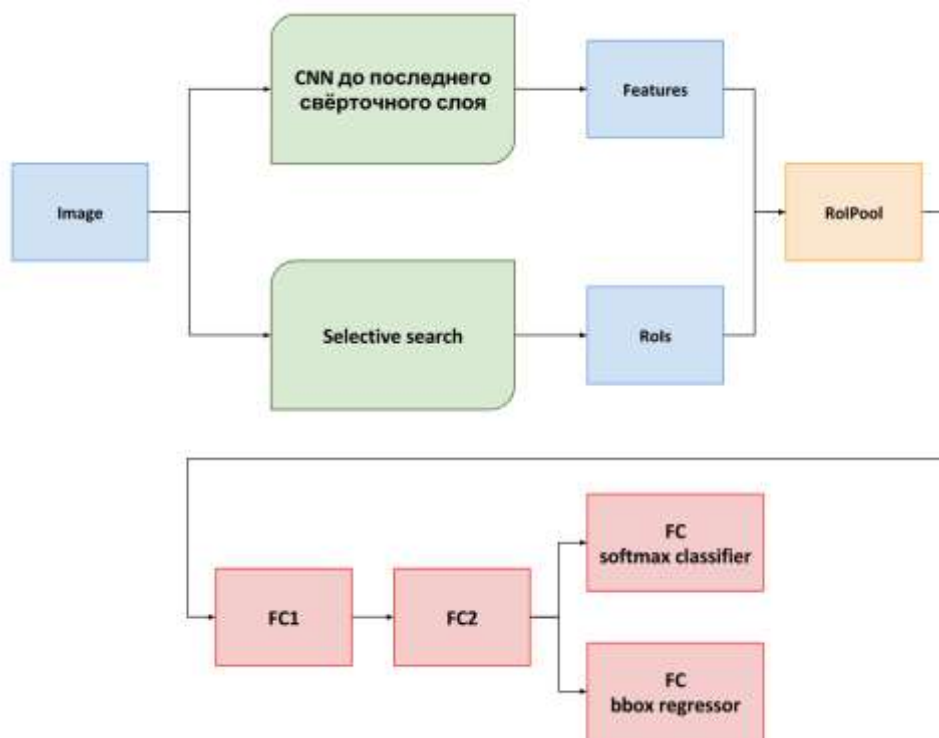


Рисунок 1.14 – Архітектура мережі Fast R-CNN

Щоб натренувати отриману мережу, потрібно вести функцію втрат, що об'єднує помилки класифікації і регресії координат обмежувального прямокутника. Позначимо через $p = (p_0, p_1, \dots, p_N)$ – вихід softmax шару – вектор ймовірностей, що у відповідному RoI міститься об'єкт одного з $N+1$ класів. Другий шар видасть нам матрицю $4 \times N \times N$ відновлених зсувів обмежувальних прямокутників об'єкта для кожного з N класів: $t = (t_x^p, t_y^p, t_w^p, t_h^p, n = \overline{1, N})$. Тут координати обмежувального прямокутника відновлюються по зсувам аналогічно тому як це було у випадку звичайної R-CNN, тільки там використовувався набір лінійних регресій для пошуку t .

Автори пропонують наступну функцію втрат:

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v),$$

де u – клас переданого RoI по версії датасета ($u = 0$ відповідає класу негативів, і в такому разі друга складова в функції втрат зануляється);
 $v = (v_x, v_y, v_w, v_h)$ – bbox об'єкта, що відповідає переданому RoI;
 λ – коефіцієнт балансування впливу помилки класифікації і помилки уточнення координат.

Помилка класифікації обчислюється стандартно: $L_{cls}(p, u) = -\log(p_u)$. У другій складовій як функція втрат береться наступна сума:

$$L_{loc}(t^u, v) = \sum_{i=x,y,w,h} \Psi_1(t_i^u - v_i)$$

де $\Psi_1(x)$ – функція Хубера (з параметром 1), що має такий вигляд

$$\Psi_1(x) = \begin{cases} 0,5x^2, & \text{якщо } |x| < 1; \\ |x| - 0,5, & \text{інакше;} \end{cases}$$

Таким чином, на відміну від R-CNN, де карта ознак генерувалась для кожного претендента на зображенні, в даному випадку генерується карта ознак для всього зображення, а потім за допомогою спеціального шару з неї обчислюють карти для кожного з претендентів, що дозволяє суттєво скоротити час виводу. При цьому автори відмовились від використання окремих лінійних регресорів і натренували, для уточнення локалізації об'єктів, окремий шар нейронної мережі.

У працях [13], [14] було запропоновано замінити процедуру генерації претендентів selective search на невелику нейронну мережу, яка використовує існуючу карту ознак. Як згорткові мережі автори пропонують використовувати або VGG16, або ZF [15]. Для зображення розміру $W_f \times H_f$ на виході останнього згорткового шару (conv5_3) мережа VGG16 видає карту ознак з просторовими розмірами $W_{I/16} \times H_{I/16}$, вектор ознак для кожної точки матиме розмірність 512. При цьому в вектор ознак в точці (x_f, y_f) вносять вклад точки зображення, що

лежать всередині квадрату з центром в $(16x_f, 16y_f)$ і розміру 196×196 . Для кожної точки карти ознак (x_f, y_f) перевіряються k претендентів різних розмірів на зображенні в регіонах з центром в $(16x_f, 16y_f)$. Як претенденти розглядаються 9 прямокутників, варіюючи три масштаби і три відношення сторін (1:1, 1:2, 2:1). Для цієї задачі використовується мережа пропозиції регіонів (Region Proposal Network, RPN) (рис. 1.15).

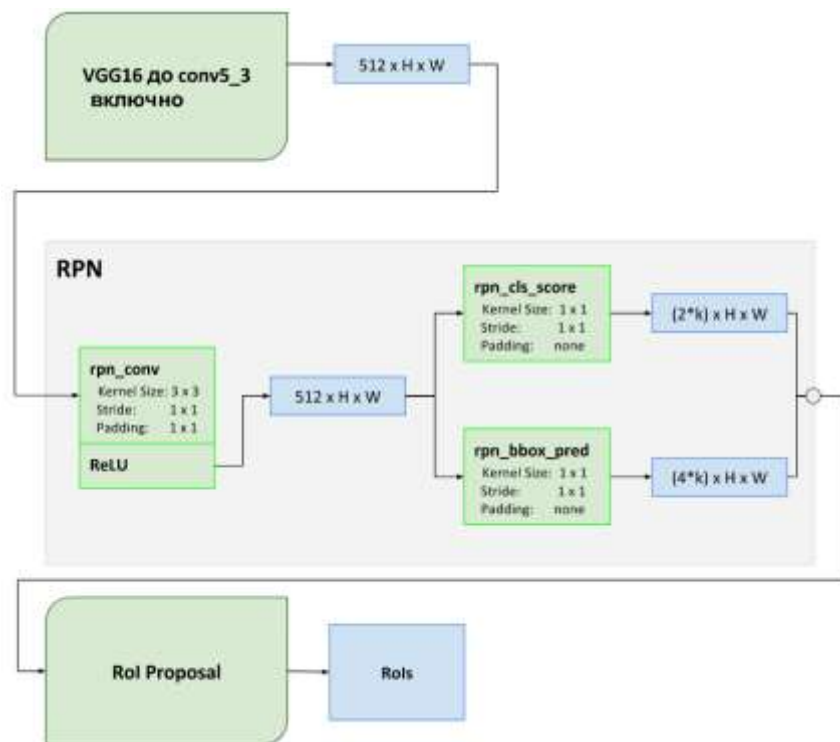


Рисунок 1.15 – Схема роботи мережі Region Proposal Network

Блок RPN використовується для визначення областей, в яких найбільш ймовірно знаходяться об'єкти, що дозволяє скоротити загальний час обчислень. Здійснюється це отриманням k прямокутників, кожен з яких має два значення, що представляють ймовірність присутності і відсутності об'єкта в цьому прямокутнику.

В даному випадку $k=9$: три масштаби по три відношення сторін. RPN використовує вікно розміру 3×3 , ковзаюче по високорівневій карті ознак, що дає розмір 177×177 пікселів вхідного зображення (1.16).

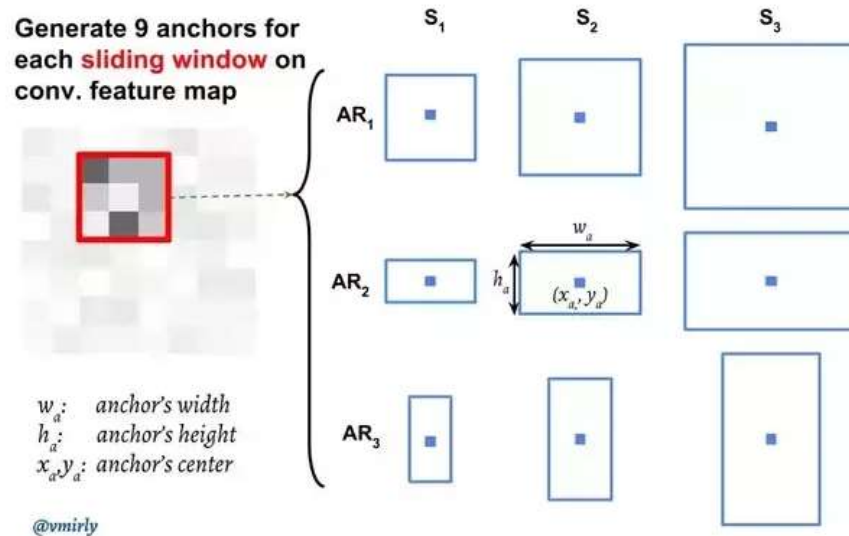


Рисунок 1.16 – Якорі ковзного вікна, що сканує високорівневу карту ознак

Кожне передбачення параметризується відносно центру ковзного вікна і представляється у вигляді дев'яти якорів – прямокутників трьох різних масштабів і трьох відношень сторін. На виході цього блоку для кожного об'єкту буде сформовано $4 \cdot k$ виходів – координат прямокутних областей (шар `grp_bbob_pred`). Даний метод забезпечує інваріантність до розміщення об'єкта на зображенні. Блок RCNN є класифікатором і визначає присутність і відсутність об'єкта для кожної області і містить $2k$ виходів (шар `grp_cls_score`). Отримана з класифікатора інформація перераховується у ймовірність (шар `grp_cls_score_reshape`, `grp_cls_prob`, `grp_cls_prob_reshape`).

Фактично необхідно натренувати дві мережі: Fast R-CNN для класифікації і уточнення координат об'єктів, RPN для генерації претендентів. При цьому зготкові шари повинні бути спільними для цих двох мереж. Таким чином одна нейронна мережа навчається здійснювати детектування і класифікацію об'єктів. Архітектура Faster R-CNN показано на рис. 1.17.

У працях [16],[18] було запропоновано мережу Single Shot Multibox Detector (SSD), яка в багатьох задачах перевищила точність детектування Faster R-CNN, хоч і дещо поступається у швидкодії. Її архітектура зображена на рис. 1.18.

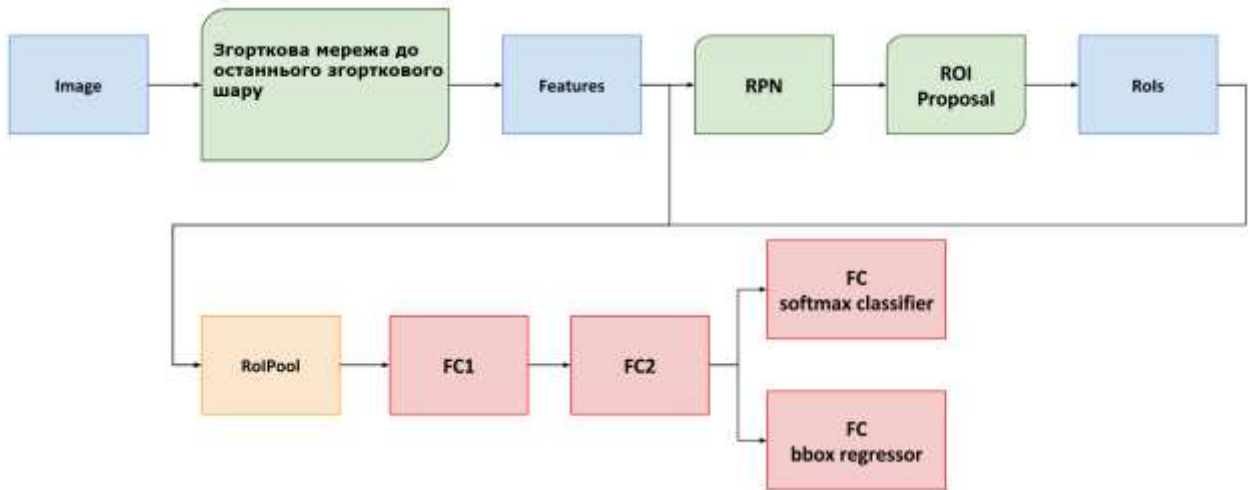


Рисунок 1.17 – Архітектура мережі Faster R-CNN

У мережі SSD вхідне зображення надходить на згорткові шари і після кожного потрапляє в класифікатор-детектор. Завдяки цьому мережа стійка до розміру об'єкта, оскільки після кожного згорткового шару об'єкт стає все меншим. Далі, з усіх знайдених об'єктів обирається найбільш ймовірний методом придушення не максимумів (Non-maximum Suppression, NMS).

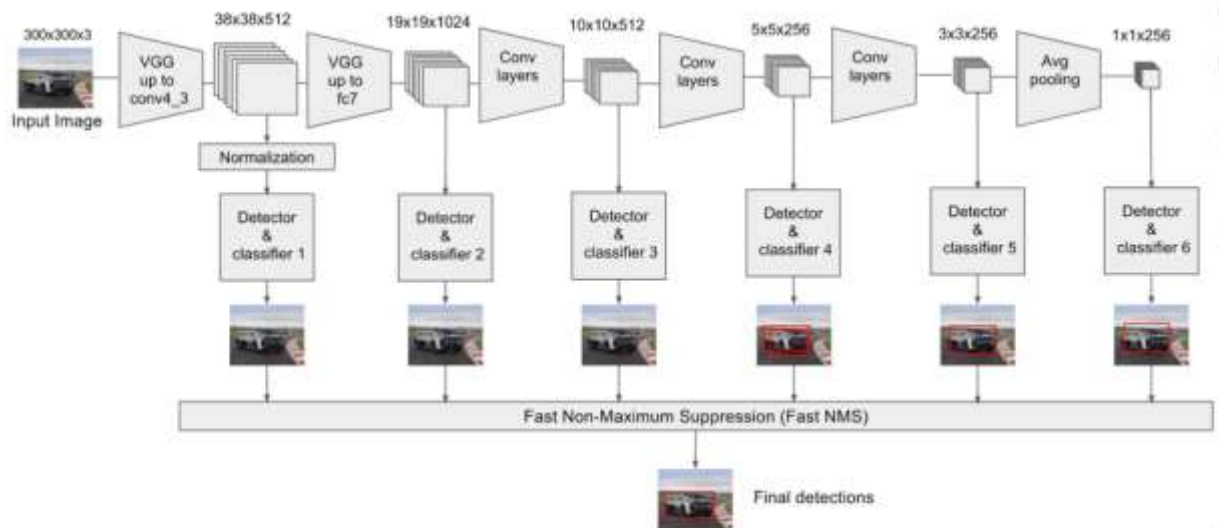


Рисунок 1.18 – Архітектура мережі SSD

Існують і інші мережі для детектування об'єктів, такі як DetectNet [16], Yolo [15] та інші, проте вони поступаються мережам SSD та Faster R-CNN як у швидкодії так і в точності.

Ефективність розглянутих нейронних мереж детектування об'єктів на місцевості залежить, як від ефективності зготкового шару, так і класифікатора об'єктів інтересу. Тому подальше покращення цих архітектур варто здійснювати в напрямках підвищення функціональної ефективності окремих складових.

1.3 Методи, моделі та алгоритми навігації автономних безпілотних літальних апаратів

Пошук та визначення координат об'єктів інтересу на місцевості в процесі аеро- відео- спостереження автономною бортовою системою безпілотного літального апарату потребує вирішення, крім розглянутої задачі детектування об'єктів інтересу на зображенні, також задачі навігації, яка включає в себе підзадачі локалізації літального апарату в просторі, визначення перешкод та корекції траєкторії польоту.

Для навігації та підвищення рівня автономності безпілотний апарат оснащують різноманітними видами сенсорів, які дозволяють відслідковувати стан об'єкта керування в кожен момент часу. До таких сенсорів відносяться інерційні прилади (акселерометри та гіроскопи), за допомогою яких визначається відхилення корпусу літального апарату в земних координатах, оптичні прилади (відеокамери, стереопари, інфрачервоні сенсори, лазерні дальноміри), що можуть бути використані для визначення відстані до об'єктів середовища та оцінки власного переміщення, а також різноманітні радіочастотні пристрої та прилади, серед яких найбільш широкого використання набули приймачі системи глобального позиціонування GPS/ГЛОНАСС.

Вибір сенсорної системи та алгоритмічного забезпечення БЛА визначає його функціональні можливості, ціну та вагу. Кожен вид сенсорів характеризується різною точністю, інформативністю та частотою вимірювань навігаційної інформації. Недоліки окремих сенсорних систем намагаються усунути шляхом використання нових інформаційних технологій обробки та інтерпретації

інформації та на основі комплексного аналізу даних з декількох різномірних сенсорів, для компенсації недоліків один одного.

У табл. 1.1 наведено типи сенсорів та характеристика навігаційної інформації, що ними формується [17].

Таблиця 1.1 – Характеристика навігаційної інформації сенсорів

| | Техніки/сенсори | Навігаційна інформація | Типова точність | Основні характеристики |
|---------------|--|----------------------------|--|---|
| Радіочастотні | GPS/GNSS Координати положення Швидкість | X,Y,Z Vx, Vy, Vz | ~10 м (1-3 м DGPS) ~0.05 м/с ~0.2 м/с | Система для умов прямої видимості. Результат з прив'язкою до глобальної системи позиціонування |
| | Псевдосупутники (псевдоліти) | X, Y, Z, Vx, Vy, Vz | Як з GPS | Система для умов прямої видимості. Функціонування як з частотою GPS, так і з відмінною від GPS частотою |
| | Бездротові мережі(WLAN: 1) метод засно-ваний на силі сигналу (Signal strength-based metho); 2) метод відбитку пальців (Fingerprinting method). | X,Y,Z X,Y,Z | 2-6 м 1-3 м | Система локального позиціонування в приміщеннях. Присутнє ослаблення сигналу залежне від відстані, проходженні через стіни та багатоприменового поширення при відбитті. Перешкоди від інших користувачів в смузі частот 2.4ГГц |

| | Техніки/сенсори | Навігаційна інформація | Типова точність | Основні характеристики |
|-----------------|---------------------------------------|------------------------|--|--|
| Радіочастотні | UWB (надшироко смугові сигнали) | X, Y, Z | Дециметрова точність теоретично досяжна на відстані до 10-20 м | Стійкий до багатопроменевого затування. Висока проникна здатність сигналу. Можливі завади від GPS. Підхід до позиціонування схожий до WLAN |
| Інерційні (IMU) | Акселерометр | Atan, Arad, Az | $<0.03 \text{ м/с}^2$ | Присутній дрейф зміщення нуля. Повинен калібруватися при наявності GPS. |
| | Гіроскоп | Відхилення ϕ | $0.5^\circ - 3^\circ$ | Короткострокова стабільність точності. Не чутливий до зовнішніх впливів. Присутній дрейф зміщення нуля. Повинен калібруватися при наявності GPS/ |
| Оптичні системи | Відеокамера | X, Y, Z | Декілька метрів | Система для умов прямої видимості. |
| | Мережа оптичних сенсорів | X, Y (Z опціонально) | Декілька метрів | Зображення сенсорів повинні перекриватися для реалізації 3D |
| | Лазерний або ультразвуковий дальномір | X, Y, Z | Від сантиметра до дециметра | Прив'язка до глобальної або локальної системи позиціонування |
| Інші | Цифровий | | | Довготривала ста- |

| | Техніки/сенсори | Навігаційна інформація | Типова точність | Основні характеристики |
|--|------------------------|------------------------|-----------------------|---|
| | компас/ магнітометр | Відхилення ϕ | $0.5^\circ - 3^\circ$ | більність точності. Чутливий до магніт-них збурень. Чутливий до нахилу. |
| | Цифровий барометр | Z | 1-3 м | Потребує калібрування на заданій початковій висоті для забезпечення достовірних вимірювань. |

Супутникові радіонавігаційні системи, наприклад, GPS і ГЛОНАСС (глобальні навігаційні супутникові системи) створені спеціально для цілей навігації і дозволяють отримати інформацію про положення в просторі: географічні координати і висота. Принцип роботи оснований на вимірюванні відстані від супутника до антени на пристроях за допомогою радіосигналів. Ефективність і точність подібних пристроїв в значній мірі залежать від характеристик середовища. Оскільки супутникові системи працюють в дециметровому діапазоні радіохвиль, досить важко визначити положення всередині залізобетонної будівлі чи при несприятливих погодних умовах. Похибка даного методу від 1 до 6 метрів, що є неприйнятним для малих і середніх робототехнічних систем. У місцях затухання, блокування чи викривлення сигналу, зокрема в закритому приміщенні, можливо використовувати репітерів, принцип яких оснований на ретрансляції сигналу з супутників. Однак вартість таких системи досить висока.

Позиціонування за допомогою Wi-Fi, основане на вимірюванні відстаней за рівнем сигналу. Радіочастотне позиціонування на основі даної технології є найбільш доступним методом у зв'язку з розвиненою інфраструктурою мереж на базі Wi-Fi точок доступу і не вимагає додаткового дороговартісного обладнання. Практично всі бездротові прийомо-передавачі вимірюють рівень вхідного сигналу

і повертають параметр індикатор сили прийнятого сигналу, на основі якого обчислюються координати. Проблеми даного методу в тому, що при слабкому сигналі в зоні точки доступу Wi-Fi неможна стверджувати, що об'єкт знаходиться далеко. Слабкий сигнал може бути обумовлений зовнішніми факторами : загородження, затухання та іншими [18]. Також точки доступу можуть переміщуватися. Тому на реалізація навігації на базі телекомунікаційної технології Wi-Fi потребує обов'язкового динамічного оновлення бази даних про точки доступу.

Завдяки поширенню стільникових мереж GSM і їх абонентів стали доступними методи позиціювання з використання GSM. Перший, найпростіший метод, оснований на визначенні координат мобільного терміналу за відомим ідентифікатором соти Cell Identifier. Однією з переваг даного методу є те, що немає потреб у оновленні абонентського апарату, все необхідне обладнання встановлюється на базових станціях. Координати обчислюються на основі знань приблизного радіусу дії конкретної базової станції. Однак точність позиціювання при високій щільності базових станцій складає 100-150 м, а на окраїнах містах чи в сільській місцевості знижується до десятку кілометрів. Інший спосіб покладається на оцінку часу прибуття сигналу, тобто на вимірювання часу проходження сигналу від абоненту до трьох найближчих базових станцій. Точність досягає 125 м. Існує і гібридний метод A-GPS (допоміжна GPS), що використовує систему GPS сумісно з GSM для визначення координат. A-GPS швидше знаходить місцезнаходження, проте має дещо меншу точність порівняно з GPS. Для алгоритмів A-GPS необхідний канал зв'язку через інтернет з віддаленим сервером, який надає інформацію для приймача.

У системах локального позиціювання за допомогою інфрачервоного випромінювання координати розраховуються за часом проходження періодичного імпульсу, що випромінюється об'єктом від джерела до приймача. Недоліками даного методу є перешкоди від сонячного світла і невисока відносна точність. Використовуючи інфрачервоний лазер, можна підвищити точність вимірювань до 10 см, однак це дороговартісні рішення. Системи локального позиціювання на

основі ультразвукових сенсорів працюють за аналогічним принципом джерело-приймач. На об'єкт встановлюється джерело ультразвукових хвиль, а приймачі розміщують в приміщенні в певному порядку. Для обчислення координат на площині достатньо отримати дані з трьох приймачів. Для визначення положення в просторі, включаючи висоту, необхідно мати не менше чотирьох приймачів хвиль. Точність вимірювання в ідеальних умовах досягає 3-5 см. Недоліком даного способу є необхідність строгого планування розміщення сенсорів – приймачів ультразвукових хвиль.

У праці [17] було запропоновано здійснювати оцінку положення в 3D-просторі та швидкості руху об'єкта на основі обробки даних про прискорення. За рахунок комбінування інерційних сенсорів та магнітометра передбачалося здійснювати не тільки відслідковування швидких рухів, але й відносно довготривалих змін. При цьому положення об'єкта визначається шляхом оцінки результатів подвійного інтегрування даних від 3-х осьового акселерометра, вбудованого у цифровий компас (рис. 1.19).

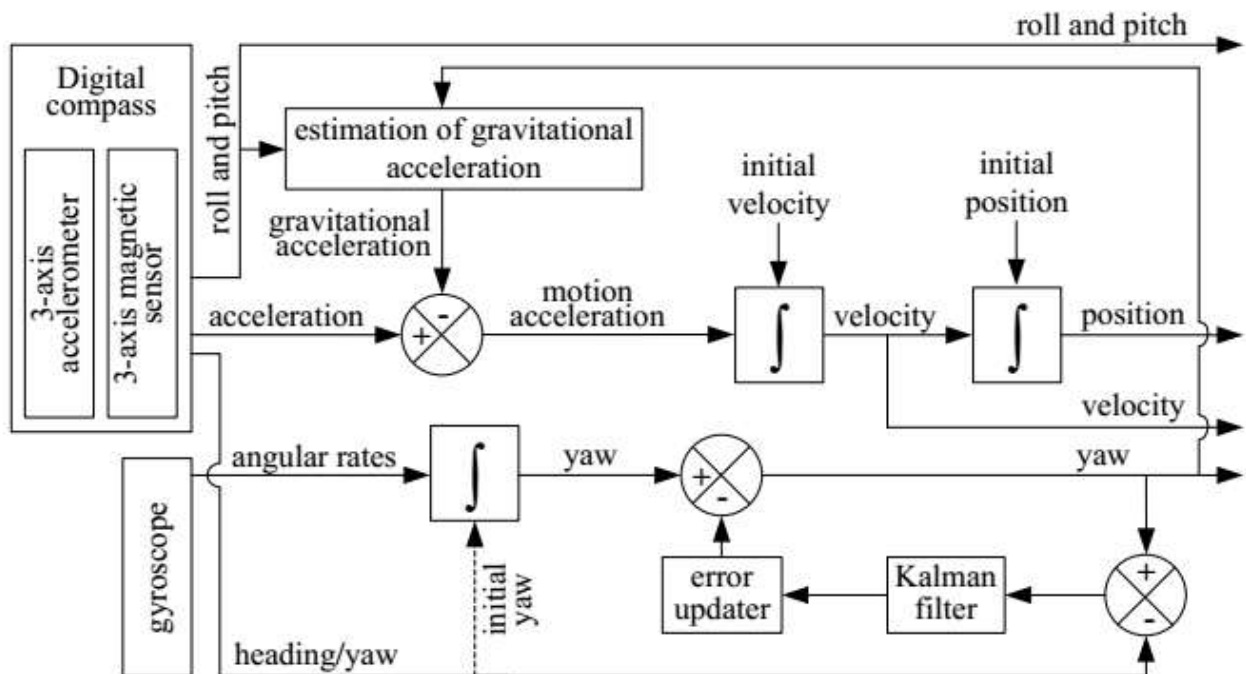


Рисунок 1.19 – Структурна схема системи відслідковування позиції в 3D просторі на основі інерційних сенсорів та магнітометра

Оскільки акселерометр забезпечує як прискорення через рух і силу тяжіння, пропонується метод розділення цих двох компонентів прискорення. Кути крену (Roll) і тангажу (pitch) отримують за допомогою цифрового компасу, а кут ристання (yaw) оцінюють шляхом об'єднання даних від цифрового компасу та гіроскопа з використанням фільтра Калмана. При цьому основним завданням є мінімізація помилки, викликані процесом подвійного інтегрування прискорень. Внаслідок ефекту дрейфу (помилка накопичується з часом) оцінка позиції не може бути виконана з адекватною точністю для періодів довше, ніж кілька секунд. Тому здійснюється виявлення зупинок руху об'єкта з метою інтегрування прискорення лише в ефективні моменти руху. Однак експериментальні дані показали, що накопичена з часом помилка хоч і є меншою, але все одно її значення залишається достатньо високим, що обмежує використання даного методу для розробки навігаційної системи.

Стандартні навігаційні системи безпілотних літальних апаратів залежать від системи глобального позиціонування (GPS) і блоку інерційного вимірювання (Inertial Measurement Unit, IMU) [15]. В літальних апаратах сигнал GPS може бути заблокований або прийнятий з завадами (особливо на цивільних частотах). Однак IMU акумулює помилку і через декілька секунд (для малогабаритних літальних апаратів, де використовуються дешеві IMU) дані вимірювань стають хибними. У зв'язку з цим багато досліджень спрямовано на підвищення ефективності і надійності GPS/IMU навігації з використанням відеокамер БЛА та технологій машинного зору.

Одним із сучасних підходів до вирішення задачі навігації на основі ідей та методів машинного зору є використання алгоритмів на основі схожих ознак (Feature-Based Navigation) або контрольних точок місцевості. Даний підхід оснований на порівнянні поточних зображень, відзнятих з бортової відеокамери, з попередніми еталонними зображеннями, як правило, отриманими, коли GPS сигнал був гарантовано коректним. Система порівнює місце схожих ознак отриманих зображень (для яких в бортовій пам'яті можуть бути записані координати, щоб з їх допомогою визначити місцезнаходження БЛА. Однак такий

підхід потребує наявності бази еталонних зображень і може бути непридатним в ситуаціях необхідності оперативної реакції на зміни зовнішньої обстановки і прийняття рішень в реальному темпі часу. Розробка системи навігація на основі динамічних даних бортової системи літального апарату на даний момент є найбільш актуальною задачею.

Машинний зір використовується вже багато років для виявлення перешкод і навігації транспортних засобів та роботів. Відеокамери мають низькі вартість, масу та енергоспоживання, однак технології обробки візуальних даних характеризуються обчислювальною складністю та варіацією надійності залежно від середовища спостереження. Тому дані системи доповнюють інформацією від інерційних сенсорів та інших сенсорів, а технології обробки візуально-інерційних даних продовжують розвиватися і вдосконалюватися.

Однією з технологій визначення положення літального апарату на основі бортової відеокамери є візуальна одометрія (Visual Odometry), яка покроково оцінює положення апарату залежно від того, як змінюється зображення, отримані з відеокамери, під часу руху літального апарату у навколишньому середовищі. Для того, щоб візуальна одометрія ефективно працювала потрібні достатня освітленість, статичне середовище та насичена текстура [20]. Статичне середовище важливе, оскільки візуальна одометрія працює, відстежуючи зміни місць розташування об'єктів від кадру до кадру. Якщо ознака фактично перемістилася в оточенні навколишнього середовища, трактування її як стаціонарної точки відліку призведе до помилкової оцінки руху літального апарату. Ще однією важливою умовою є наявність перекриття послідовних кадрів з метою відстеження спільних ознак на різних кадрах зображення. Це досягається як за рахунок збільшення частоти кадрів, так і за рахунок збільшення кута зору оптичної системи. Як це показано на рис. 1.20, зображення I_{k-1} сформоване камерою в положенні C_{k-1} в момент часу $k-1$. В наступний момент часу, k , камера матиме положення C_k і сформує зображення I_k . Положення C_{k-1} та C_k можуть бути пов'язані формулою

$$C_k = C_{k-1} T_{k,k-1}$$

де $T_{k,k-1} \in R^{4 \times 4}$ – є матрицею трансформації в наступному вигляді

$$T_{k,k-1} = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ 0 & 1 \end{bmatrix},$$

де $R_{k,k-1} \in R^{3 \times 3}$ – є матрицею повороту;

$t_{k,k-1} \in R^{3 \times 1}$ – вектор переміщення.

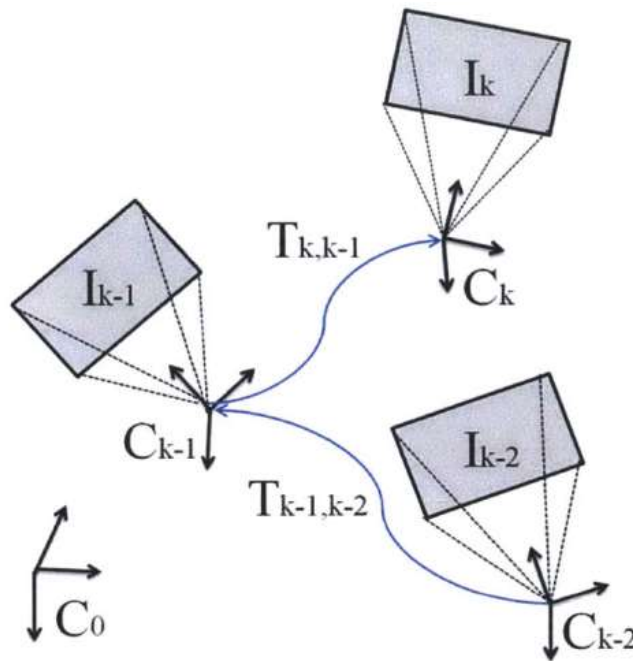


Рисунок 1.20 – Ілюстрація задачі візуальної одометрії

Відносне перетворення $T_{k,k-1}$ між сусідніми положеннями камери C_{k-1} і C_k обчислюється на основі візуальних ознак з зображень I_{k-1} і I_k . Всі трансформації підсумовуються для отримання абсолютного положення C_k відносно координат початкового кадру C_0 . Формалізованою метою візуальної одометрії є відновлення всієї траєкторії переміщення камери шляхом поєднання

множини всіх положень камери, $C_{0:n} = \{C_0, \dots, C_n\}$, починаючи з початкової позиції, C_0 .

На рис. 1.21 показано узагальнену блок-схему алгоритму візуальної одометрії. Спочатку візуальні ознаки детектуються на кожному зображенні I_k і співставляються з ознаками попереднього кадру. Потім обчислюється відносне переміщення T_k між моментами часу $k-1$, k та положення камери $C_k = C_{k-1}T_{k,k-1}$. Останнім кроком є локальна оптимізація за останніми m кадрами для отримання більш точної оцінки локальної траєкторії.



Рисунок 1.21 – Блок-схема основних компонентів алгоритму візуальної одометрії

Для детектування ознак можуть бути використані різноманітні алгоритми виділення країв, такі як Harris, Shi-Tomasi, Moravec та FAST, або алгоритми екстракції точок інтересу, такі як SIFT, SURF та CENSURE [8]. Вибір алгоритму детектування залежить від особливостей середовища та обмежень на обчислювальні ресурси.

Існує багато методів співставлення ознак між зображеннями, найпростішим серед яких є порівняння всіх дескрипторів ознак одного зображення з усіма дескрипторами ознак іншого зображення за допомогою певної міри схожості (наприклад суми квадратів відмінностей або нормалізованої перехресно-

кореляції) [18]. Тип дескриптора впливає на вибір міри схожості. Інший варіант для співставлення ознак – пошук всіх ознак одного зображення, а потім пошук цих ознак в інших зображеннях. Цей метод є кращим, коли рух та зміна зображення між кадрами невеликі [20]. Набір усіх співпадінь, що відповідають одній ознаці, називається треком ознаки.

Метою етапу оцінювання руху є використання всіх треків ознак для обчислення рух камери з попереднього кадру зображення до поточного кадру. На кожному кадрі I_k здійснюється екстракція множини ознак f_k . $3D$ положення i -ї ознаки, f_k^i , представляється як $3D$ координати точки, \tilde{X}_k^i . Так само й i -та ознака на кадрі I_{k-1} представляється як $3D$ координати точок \tilde{X}_{k-1}^i . Переміщення камери обчислюється шляхом пошуку перетворення множини $3D$ точок зображення I_k в множину $3D$ точок зображення I_{k-1} . Явним рішенням є $T_{k,k-1}$, що мінімізує L_2 відстань між двома множинами $3D$ точок

$$\hat{T}_{k,k-1} = \arg \min_{T_{k,k-1}} \sum_i \| \tilde{X}_k^i - T_{k,k-1} \tilde{X}_{k-1}^i \|^2.$$

Ці перетворення мають абсолютну шкалу, оскільки функції представлені у вигляді $3D$ точок [17]. Тоді поточне положення, C_n , обчислюється прямим поєднанням всіх перетворень $T_{k,k-1}$, $k = \overline{1, n}$, за формулою $C_k = C_{k-1} T_{k,k-1}$. При цьому існує два шляхи подання ознак: обидві множини f_{k-1} та f_k задані в $2D$ координатах зображень; f_{k-1} задаються в $3D$ і f_k відповідають $2D$ проєкціям на зображенні I_k .

Перевагою при оцінці переміщення у виконанні вирівнювання лише між останніми парами зображень полягає в тому, що обчислювальна складність алгоритму має постійну тривалість відносно довжини шляху. Недоліком є те, що вводиться дрейф в траєкторію, оскільки невеликі похибки в положенні та поворотах накопичуються з часом [19]. Один із способів зменшення цього ефекту

полягає в застосуванні коригування в межах пакетного вікна (bundle adjustment), що дозволяє здійснювати одночасну оптимізацію параметрів камери та параметрів 3D ознак для набору m з кадрів [17]. Зокрема, коригування в межах пакетного вікна спрямоване на мінімізацію помилки репроекції зображення:

$$[\hat{X}, \hat{C}_{1:k}] = \arg \min_{C, X} \sum_{i=1}^n \sum_{k=1}^m \| p_k^i - g(X^i, C_k) \|^2,$$

де n – кількість ознак;

p_k^i – i -та точка 3D ознаки X^i виміряної на k -му зображенні;

$g(X^i, C_k)$ – репроекція зображення відповідно до поточного положення камери C_k .

Оскільки при коригуванні в межах пакетного вікна використовується вимірювання ознак через m кадрів зображення, то дрейф зменшується в порівнянні з двома кадрами візуальної одометрії. Якщо m збільшується, то дрейф зменшується також, проте обчислювальні витрати збільшуються. Таким чином, вибір розміру вікна, m , відбувається виходячи з обчислювальних обмежень.

Для навігації на невідомій місцевості з перешкодами, крім локалізації, корисною функцією є побудова карта місцевості. Саме для таких цілей використовуються системи одночасної локалізації та картографування (Simultaneous Localization and Mapping, SLAM). Візуальна одометрія може розглядатися як один з кроків, що виконуються SLAM системою. При цьому система візуальної одометрії намагається як найшвидше знайти як можна більше простих орієнтирів для оцінки зміщення літального апарату за час зміни кадрів, а SLAM вимагає опису об'єктів для подальшого розпізнавання сцени.

Необхідність побудови карти навколишнього середовища є двоякою. По-перше, карта часто потрібна для підтримки інших завдань, наприклад, карта може інформувати про планування шляху або забезпечити інтуїтивну візуалізацію для оператора людини. По-друге, карта дозволяє обмежити помилку, здійснену при

оцінці стану робота. За відсутності карти, в навігаційних розрахунках з часом швидко накопичуються помилки; з іншого боку, використовуючи карту, наприклад у вигляді ознакових орієнтирів, літальний апарат може скорегувати свою помилку локалізації, переглянувши відомі ділянки (так зване loop closure). Тому SLAM знаходить застосування у всіх ситуаціях, в яких попередня карта недоступна та потребує побудови.

Архітектура системи SLAM включає дві основні компоненти: front-end та back-end. Front-end описує абстрактні дані сенсорів у моделях, які піддаються оцінці, тоді як back-end виконує виведення з абстрактних даних, що виробляються на front-end. Узагальнена схема SLAM-системи показана на рис. 1.22 [16].



Рисунок 1.22 – Архітектура типової SLAM-системи

Задача SLAM в рамках існуючих підходів формалізується як максимізація апостеріорної оцінки змінних. В SLAM змінна X зазвичай включає траєкторію камери, як дискретний набір її положень, та розміщення орієнтирів в середовищі. Передбачається, що в отриманому наборі вимірів $Z = \{z_k | k = \overline{1, m}\}$ кожне вимірювання можна виразити як функцію від X , тобто $z_k = h_k(X_k) + \varepsilon_k$, де $X_k \subseteq X$ – підмножина змінних X , а $h_k(\cdot)$ – відома функція (модель вимірювання чи спостереження), ε_k – випадковий шум вимірювань. У даній задачі оцінка X^* здійснюється шляхом пошуку максимуму апостеріорної ймовірності $p(X | Z)$ (довіра до X за даними вимірів)

$$X^* = \arg \max_X p(X | Z) = \arg \max_X p(Z | X)p(X)$$

де $p(Z | X)$ – правдоподібність вимірів Z віднесених до X ;

$p(X)$ – апіорна ймовірність X .

Апостеріорна ймовірність включає будь-які попередні знання про X , а у випадку відсутності попереднього знання $p(X)$ стає константою (рівномірний розподіл), що є несуттєвим і може бути виключене з оптимізації. У даному випадку оцінка максимальної апостеріорної оцінки може бути зведена до оцінки максимальної правдоподібності. При цьому варто відмітити, що на відміну від фільтрації Калмана, пошук максимальної апостеріорної оцінки не вимагає чіткого розмежування між моделями руху та спостереження : обидві моделі розглядаються як фактори і незмінно включаються в процес оцінювання. Крім того, варто зазначити, що фільтрація Калмана та визначення максимуму апостеріорної оцінки повертають ту ж саму оцінку в лінійній гауссовій ситуації, тоді як у загальному випадку це не так.

Якщо припустити, що вимірювання Z незалежні (тобто відповідні шуми некорельовані), то попередню формулу пошуку максимальної апостеріорної оцінки можна факторизувати в такому вигляді, де z_k залежить лише від підмножини змінних в X_k

$$X^* = \arg \max_X p(X) \prod_{k=1}^m p(z_k | X) = \arg \max_X p(X) \prod_{k=1}^m p(z_k | X_k)$$

Якщо припустити, що ε_k є Гаусовим шумом з нульовим математичним очікуванням та інформаційною матрицею Ω_k (інвертована коваріаційна матриця). Тоді, правдоподібність $p(z_k | X_k)$ можна виразити так

$$p(z_k | X_k) \sim \exp\left(-\frac{1}{2} \|h_k(X_k) - z_k\|_{\Omega_k}^2\right),$$

де $\|e\|_{\Omega_k}^2 = e^T \Omega e$.

Аналогічним чином можна припустити, що апіорну ймовірність можна записати так

$$p(X) \sim \exp\left(-\frac{1}{2} \|h_0(X) - z_0\|_{\Omega_0}^2\right),$$

де $h_0(\cdot)$, z_0 , Ω_0 – задані функція, апіорні математичне очікування та інформаційна матриця.

В результаті максимізацію апостеріорної оцінки можна подати як мінімізацію негативної логарифмічної правдоподібності з урахуванням згаданих припущень

$$X^* = \arg \min_X -\log\left(p(X) \prod_{k=1}^m p(z_k | X_k)\right) = \arg \min_X \sum_{k=0}^m \left(\|h_k(X_k) - z_k\|_{\Omega_k}^2\right).$$

Отримана формула оцінки X^* подібна до проблеми корегування в межах пакетного вікна (bundle adjustment), але дозволяє включати до розгляду широкий спектр моделей сенсорів, наприклад, інерційні сенсори, кутові енкодери, GPS та ін.. При цьому проблема мінімізації даного виразу зазвичай вирішується шляхом послідовної лінеаризації, наприклад, методами Гауса-Ньютона або Левенберга-Маркварда.

У практичних задачах робототехнічних програм важко безпосередньо писати вимірювання сенсорів як аналітичну функцію стану, як це вимагається в задачі максимізації апостеріорної оцінки. Це справедливо як для відеокамер так і для простих сенсорів у вигляді лазерного дальноміра з одним променем. Для підвищення інформативності і зниження обчислювальної складності, як правило,

здійснюється екстракція ознакового опису. Саме тому перед SLAM back-end міститься модуль front-end, що здійснює екстракцію релевантних ознак з сенсорних даних. Первинна обробка, яка відбувається на front-end, залежить від сенсорів, оскільки поняття ознаки змінюється в залежності від потоку вхідних даних, який розглядається. Крім того front-end виконує асоціацію даних, тобто прив'язку кожного вимірювання з певним орієнтиром середовища (наприклад 3D точкою). В більш абстрактному розумінні можуть асоціації даних асоціює кожне вимірювання z_k з підмножиною невідомих змінних X_k так, що $z_k = h_k(X_k) + \varepsilon_k$.

Модуль асоціації даних на front-end містить короткотерміновий та довгостроковий блоки асоціації даних. Короткотермінові асоціації даних відповідають за взаємодію відповідних ознак у послідовних вимірюваннях сенсорів. Наприклад, короткострокова асоціація даних здійснює відстежування того факту, що 2-піксельні вимірювання в послідовних кадрах відображають одну й ту саму 3D точку. З іншого боку, довгострокова асоціація даних (або замикання петлі) відповідає за зв'язування нових вимірювань із старими орієнтирами. Варто зауважити, що back-end зазвичай подає інформацію на front-end для підтримки виявлення та перевірки замикання петлі.

Сучасні реалізації SLAM та візуальної одометрії добре працюють на потужних комп'ютерах, проте їх запуск на слабких комп'ютерах або вбудованих пристроях може призвести до обмежених можливостей у реальному часі та неадекватних результатів. При цьому побудована в SLAM 3D-структура часто є неефективною в середовищах, де відсутні відстежувані ознаки (наприклад, стіни). SLAM, оснований на стереозорі для оцінки глибини також має низьку продуктивність у регіонах без текстури та може постраждати від дзеркальних відбитків. До того ж більшість загальнодоступних безпілотних апаратів мають лише одну вбудовану камеру, що робить подібні рішення не практичними. Тому розробка навігаційної системи без GPS для малих безпілотних апаратів ускладнена, оскільки вони мають одну камеру, невисокі обчислювальну потужність та вагу корисного навантаження.

Наявність накопичених відеоданих спостереження, реконструйованих карт місцевості та іншої апріорної інформації полегшує задачу розробки навігаційної системи. Отримані дані можуть бути використані для навчання багатошарових нейронних мереж для визначення координат та уникнення перешкод. У праці [20] було запропоновано глибоку нейронну мережу для визначення своїх координат та орієнтації в просторі на основі даних відеокамери (рис. 1.23). При цьому для формування ознакового опису спостережень використовується згорткова нейронна мережа SqueezeNet, а для визначення положення в просторі повнозв'язна мережа та два регресійні шари.

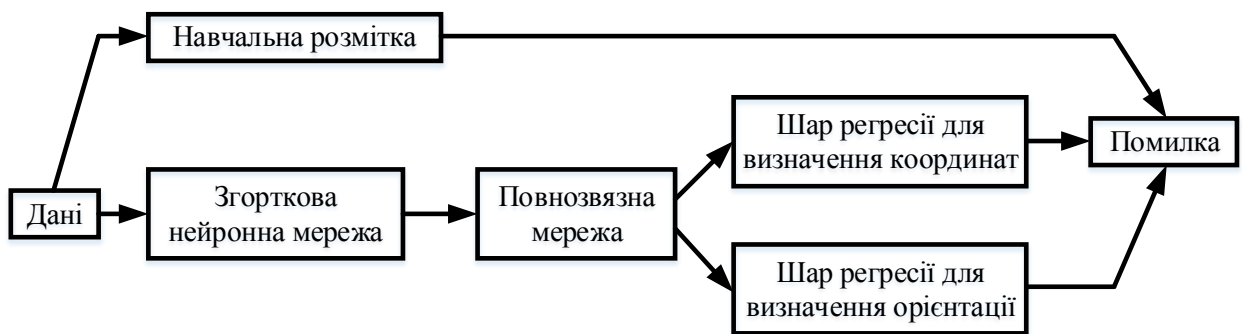


Рисунок 1.23 – Архітектура нейронної мережі для визначення положення в просторі за зображенням відео камери

Для попереднього навчання згорткової нейронної мережі була використана навчання вибірка ImageNet. Повнозв'язна нейронна мережа була попередньо навчена на наборі даних Shop Fac. Регресійні шари були попередньо навчені на наборі Places365. Точна настройка мережі здійснювалась на наборі даних Atrium, що містило дані конкретної місцевості [19], що дозволяє прогнозувати дані в глобальній системі координат.

У праці [20] було запропоновано використовувати згорткову нейронну мережу для уникнення зіткнення з перешкодами при навігації в приміщенні. Для цього було використано 5-ть згорткових шарів, 2-ва повнозв'язних шари та шар класифікації на 6-класів. Архітектуру даної мережі показано на рис. 1.24.

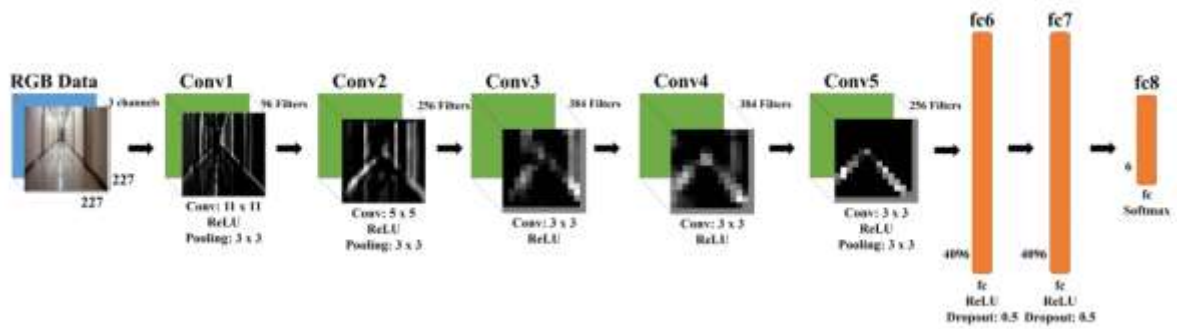


Рисунок 1.24 – Архітектура нейронної мережі для навігації в приміщенні

Навчальна вибірка даної мережі формувалась за різноманітними наборами даних розпізнавання сцен в приміщенні. Приклад ручної розмітки зображень для формування навчальної вибірки показано на рис. 1.25. Кожен клас відповідає командам, що формують поведінку безпілотної апарату.

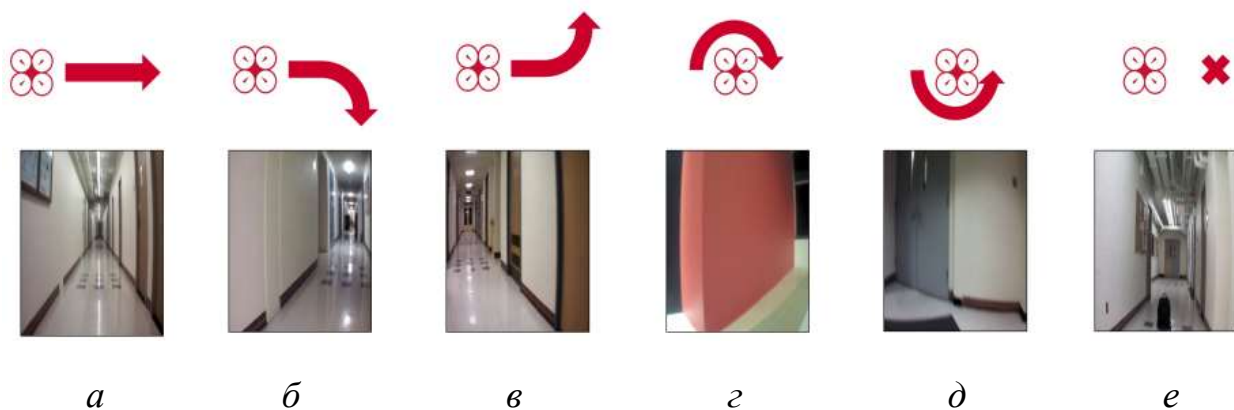


Рисунок 1.25 – Класи розпізнавання, що відповідають поведінці безпілотної апарату залежно від зображення відеокамери: *a* – рух вперед; *v* – рух вправо; *z* – рух вліво; *d* – обертання вправо; *d* – обертання вліво; *e* – зупин

Розглянуті нейромережеві рішення адаптовані до конкретного середовища, наприклад приміщень, лісу, пересічної місцевості тощо. Для усунення даного недоліку у праці [19] було запропоновано нейромережу, що за зображенням камери формує карту глибини, яка не залежить від типу місцевості і може бути використана різними способами для реалізації навігаційних рішень. На рис. 1.26 показано архітектуру даної нейромережі.

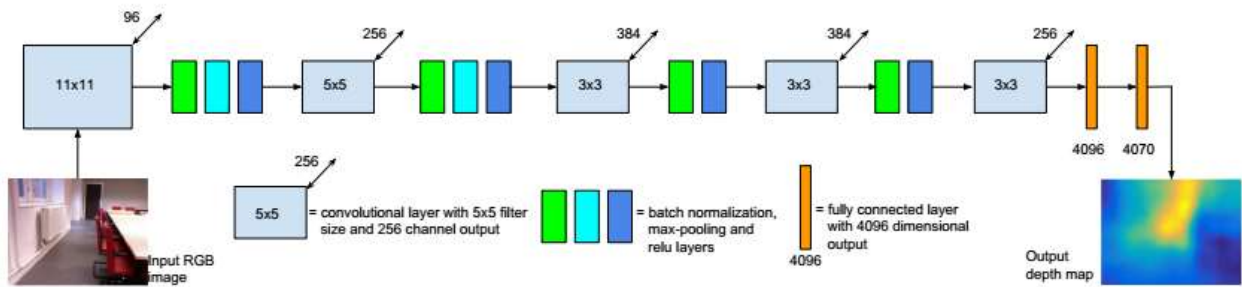


Рисунок 1.26 – Архітектура нейронної мережі для оцінки глибини за одним зображенням

Дана мережа містить 4-ри згорткові та два повнозв'язних шари, які навчаються за даними NYU2 та ESAT, де присутня карта глибини, отримана за допомогою пристрою Microsoft Kinect. Однак отримані результати свідчать, про випадки хибного спрацювання при спостереженні вікон, кутків та при наближенні до стін. Покращення результатів потребує наявності Microsoft Kinect для доповнення навчальної вибірки. Крім того питання формування вирішальних правил для здійснення навігації за даними карти глибини є відкритим.

У праці [20] було запропоновано архітектуру нейронної мережі для визначення положення в невідомому просторі на основі принципів візуальної одометрії та згорткових нейонних мереж. Для цього цього високорівневі ознаки поточного кадру конкатенуються з високорівневими ознаками попереднього кадру. Після цього здійснюється навчання верхніх шарів нейронної мережі для прогнозування зміщення камери в тривимірному просторі. На рис. 1.27 показано архітектуру даної мережі.

Навчання даної мережі здійснюється за даними KITTI Vision, що містять 11 навчальних та тестових послідовностей зображень з заданою реальною траєкторією руху. Ці зображення масштабуються до розміру 256x256. Як згорткова мережа формування ознакового опису використовується модель AlexNet. Як показали результати фізичного моделювання, отриманий ознаковий опис менш адаптований для задач одометрії ніж локальні ознаки, що набули поширення в системах машинного зору.

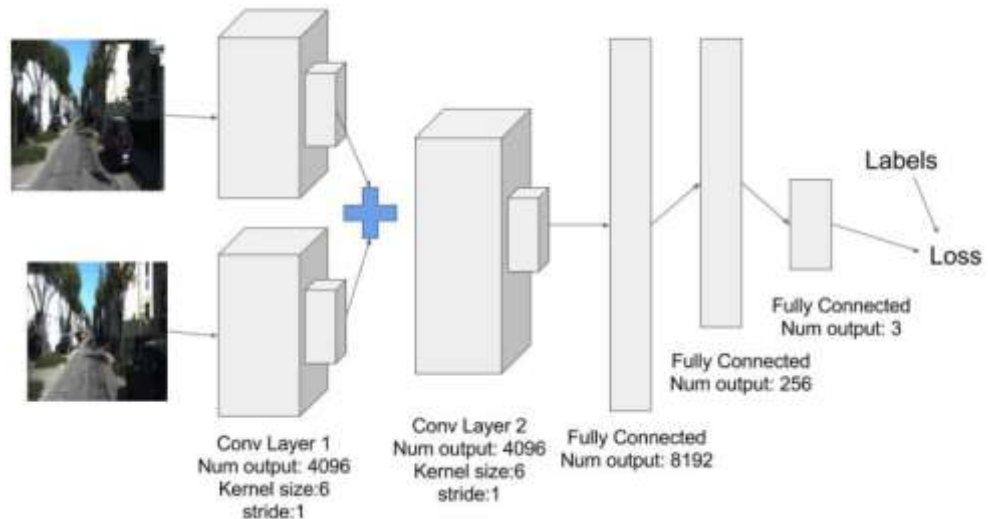


Рисунок 1.27 – Архітектура нейронної мережі візуальної одометрії в 3D

Таким чином, одним з найбільш перспективних напрямків розвитку систем навігації автономних безпілотних апаратів є використання візуальної інформації від бортової системи відеоспостереження. В даному напрямку активно ведуться дослідження, пропонуються нові алгоритми візуальної одометрії та відносної локалізації і картографування. Однак висока обчислювальна складність даних алгоритмів обмежує їх застосування в бортових системах малих безпілотних апаратів. Тому останнім часом більшої уваги приділяють більш вузькоспеціалізованим рішенням на основі нейромереж для локалізації та виявлення перешкод. Ці рішення мають потенціал для удосконалення та є придатними для використання в багатьох практичних застосуваннях.

2 АНАЛІЗ І СИНТЕЗ ІНТЕЛЕКТУАЛЬНОЇ БОРТОВОЇ СИСТЕМИ ІДЕНТИФІКАЦІЇ ОБ'ЄКТІВ НА МІСЦЕВОСТІ

2.1 Аналіз об'єкту та обґрунтування напрямку дослідження

Малі БЛА відіграють важливу роль в задачах ліквідації наслідків різного роду катастроф, стихійних лих, виробничої інспекції і охорони зовнішнього середовища. За аеродинамічною схемою розрізняють літальні апарати літакового та гелікоптерного типу. Гелікоптерні літальні апарати поділяються на однороторні та мультироторні, серед яких найдешевший та найпростіший у конструктивному виконанні є квадрокоптер. Виготовлення квадрокоптера потребує мінімального набору необхідних компонентів, а його конструкцію легко змінювати і розвивати, що робить квадрокоптери оптимальним вибором для експериментальних досліджень систем навігації та класифікаційного аналізу спостережень в БЛА. Структурна схема типового квадрокоптера показана на рис. 2.1.

Правильно зібраний і мінімально налаштований квадрокоптер здатний зависати у повітрі практично без участі пілота. Одним з найважливіших компонентів квадрокоптера є польотний контролер, що здійснює неперервну стабілізацію апарату протидіючи поривам вітру і неоднорідностям повітряних мас.

Функції зовнішньої комунікації, формування команд для польотного контролера, автономної навігації та додаткові функції аналізу даних як правило бере на себе бортовий комп'ютер. Одним з найдешевших варіантів бортового комп'ютера є Raspberry Pi 3, який може комплектуватися відносно недорогим польотним контролером Navio 2, або PXFmini [21].

Компоновочну схему квадрокоптера на базі Raspberry Pi 3 та Navio 2 / PXFmini показано на рис. 2.2.

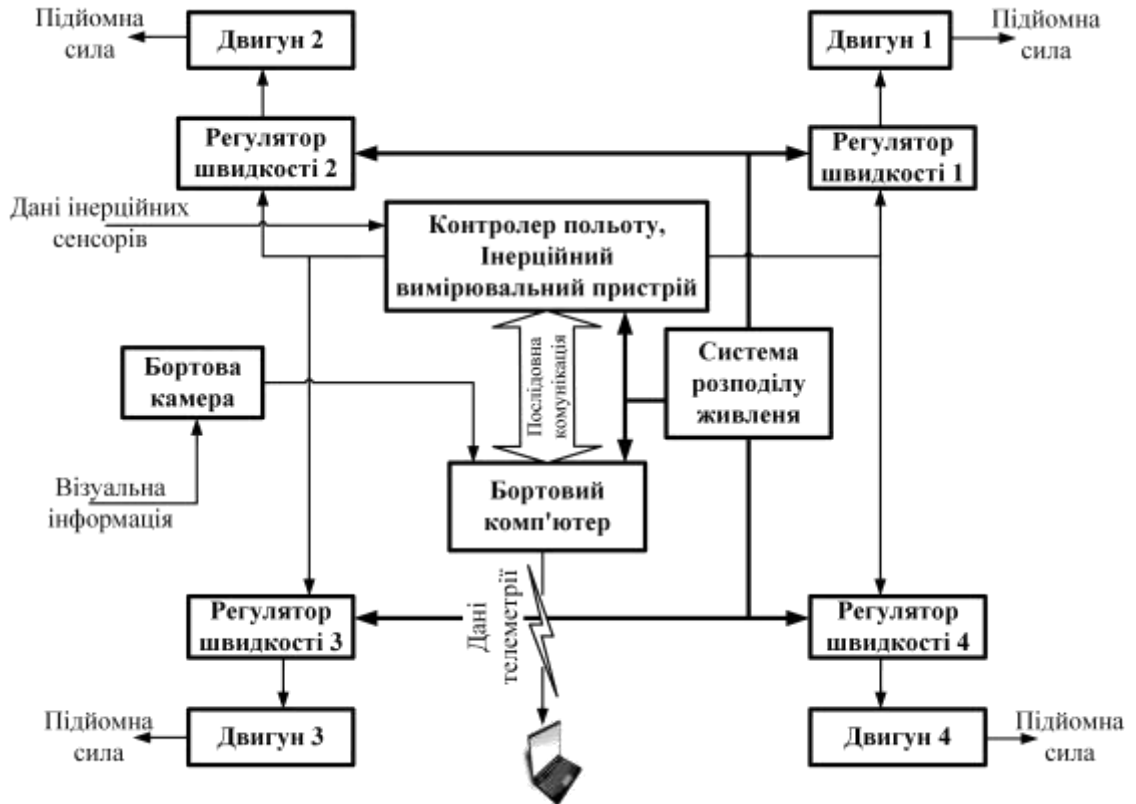


Рисунок 2.1 – Структурна схема апаратного забезпечення квадрокоптера

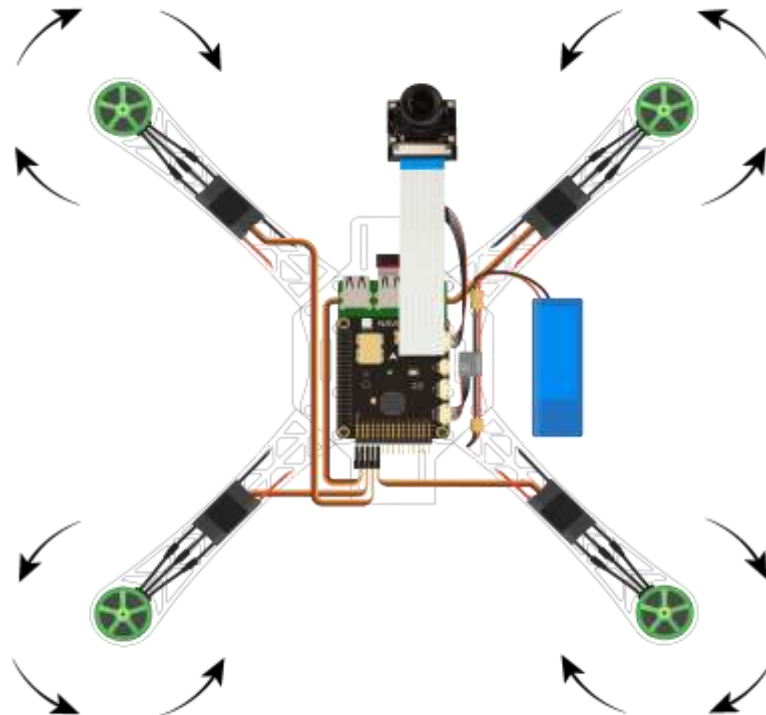


Рисунок 2.2 – Компонувочна схема безпілотного літального апарату

Аналітичний огляд підходів до проектування автономної бортової системи безпілотного апарату для ідентифікації об'єктів інтересу на місцевості, проведений в першому розділі, показав, що до складу програмного забезпечення повинні входити модуль детектування об'єкта інтересу для визначення його координат відносно безпілотного апарату, модуль обчислення позиції апарату в просторі для визначення координат об'єкта інтересу в локальній чи глобальній системі координат та модуль детектування перешкод, для реагування на непередбачувані ситуації. Функціональна схема програмного забезпечення безпілотного апарату для автономної навігації та класифікаційного аналізу спостережень показано на рис. 2.3.

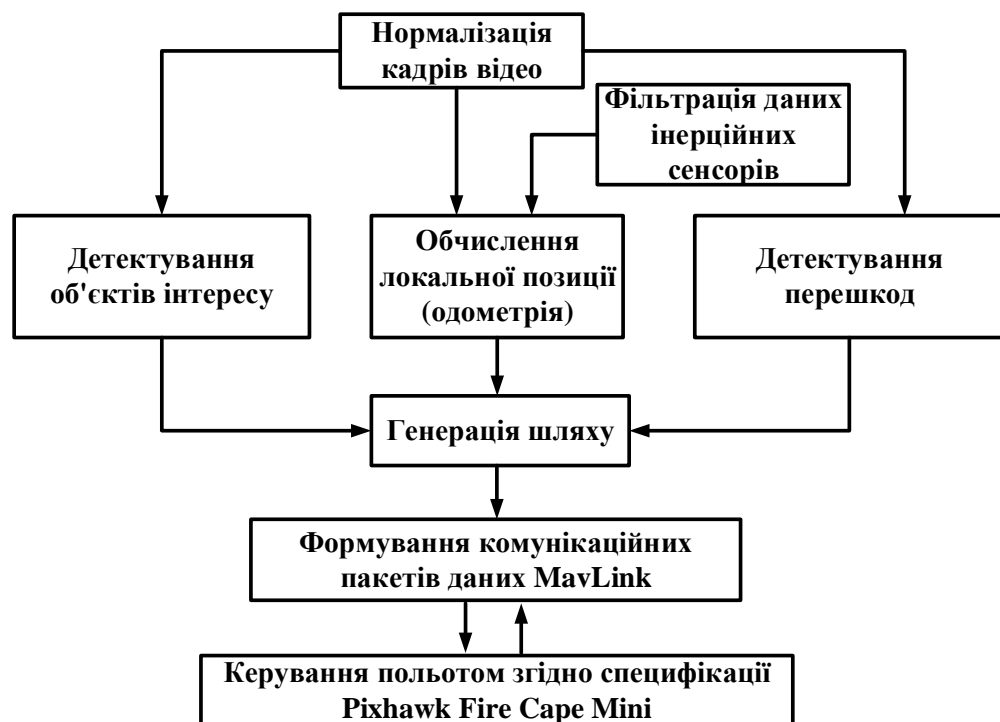


Рисунок 2.3 – Функціональна схема програмного забезпечення безпілотного апарату для автономної навігації та класифікаційного аналізу спостережень

Набір функціональних модулів автономної бортової системи безпілотного апарату, показаний на рис. 2.3, визначає структуру алгоритмічного забезпечення системи навігації та класифікаційного аналізу спостережень. У загальному випадку опис інформаційного та алгоритмічного забезпечення можна виконати у

вигляді структурної схеми моделі аналізу даних та прийняття рішень, яка приведена на рис. 2.4.

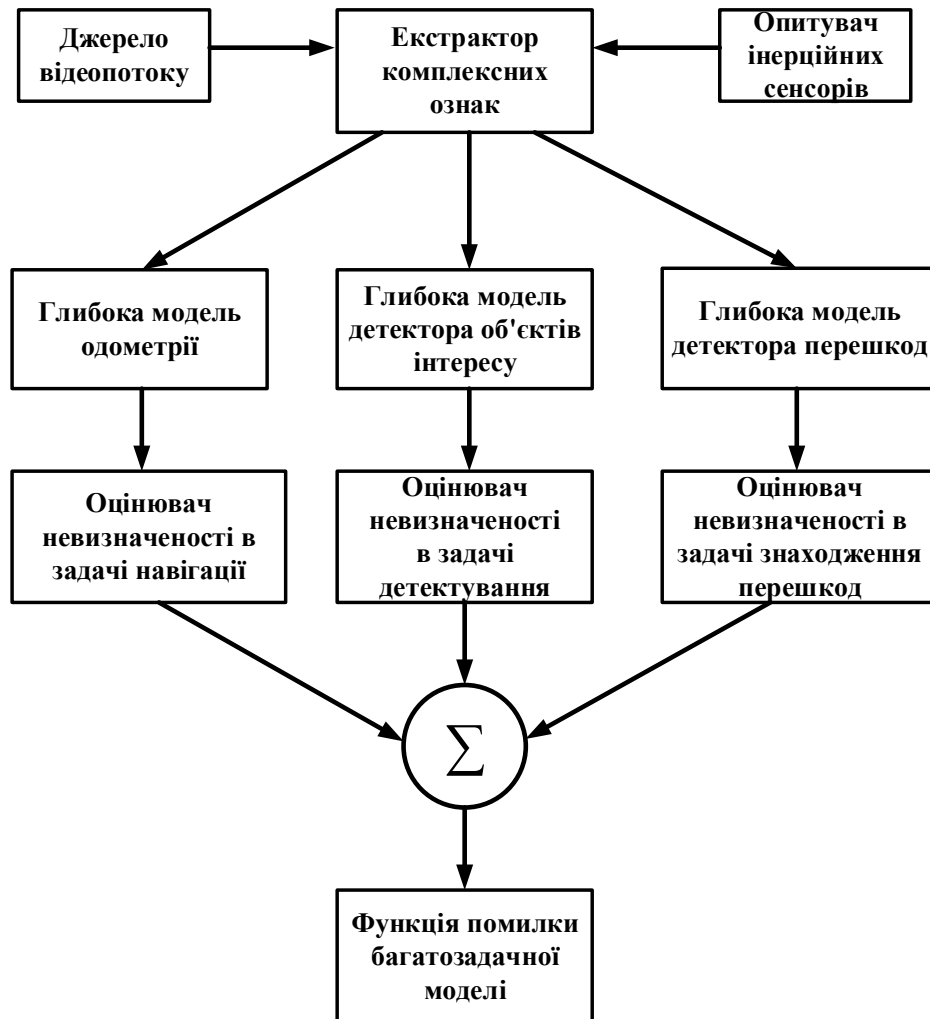


Рисунок 2.4 – Структурна схема моделі аналізу даних для ідентифікації об'єктів на місцевості

Аналіз рис. 2.4 показує, що задача ідентифікації об'єктів на місцевості складається з підзадач, які необхідно вирішувати паралельно. Моделі вирішення кожної з підзадач мають ієрархічну структуру, тобто є глибокими, оскільки БЛА функціонує в реальних умовах, що характеризуються багатофакторністю та мінливістю спостережень. При цьому кожна із задач має спільний ознаковий опис середовища, що у загальному випадку є результатом комплексного аналізу різномірних сенсорних даних.

За умов обмежених ресурсів бортового комп'ютера на ньому неможливо розгорнути універсальний екстрактор ознакового опису спостережень, що буде

оптимальним для будь-якого середовища і задач. Аналогічно й для детектора об'єктів інтересу, він не може бути навчений розпізнавати об'єкти великого алфавіту класів з великою варіацією умов та модифікацій в межах класу. При цьому кожен день можуть з'являтися нові модифікації об'єктів класу, що потребуватиме розширення чи уточнення вирішальних правил. Тому набуває важливого значення здатність системи адаптуватися до зміни умов спостереження, середовища та об'єктів інтересу з урахуванням обмежень на ресурси бортового комп'ютера та обсяг розмічених навчальних даних.

Аналітичний огляд підходів до проектування систем ідентифікації об'єктів на місцевості показав, що в задачі навігації та детектування об'єктів на зображенні одним із найбільш перспективних підходів є використання згорткових нейронних мереж у поєднанні з вирішальними правилами, оптимальними в інформаційному та вартісному сенсі. При цьому питання розробки ефективних методів самонавчання згорткових нейронних мереж для адаптації до середовища функціонування є відкритим, але потенційно може бути вирішене шляхом використання принципів розрідженого кодування та самоорганізації неструктурованих сіток [8].

Як було показано в першому розділі, синтез детектора об'єктів інтересу на місцевості перспективно здійснювати з використанням класифікаторів, отриманих на основі ідей і методів інформаційно-екстремальної інтелектуальної технології (ІЕІ-технології) аналізу та синтезу здатних навчатися систем розпізнавання. Ці методи ґрунтуються на максимізації інформаційної спроможності системи розпізнавання шляхом цілеспрямованої трансформації апріорного, у загальному випадку нечіткого, розбиття простору первинних ознак в еквівалентність класів, контейнери яких відновлюються в радіальному базисі двійкового простору вторинних ознак. У рамках інформаційно-екстремальної технології існує принципова можливість одержання в процесі машинного навчання високостовірних за навчальною та тестовою вибірками обмеженого обсягу обчислювально ефективних вирішальних правил та прогнозувати момент зниження їх функціональної ефективності. Проте питання вибору прогностичних

статистик і розрядності двійкового кодування спостережень є малодослідженим.

Вибір напрямку дисертаційних досліджень потребує розв'язання низки науково-методологічних задач, пов'язаних з аналізом і синтезом здатних навчатися систем ідентифікації об'єктів на місцевості, таких як :

1) перехід від однорівневої до багаторівневої системи контрольних допусків для адаптивного двійкового кодування ознак оптимальною кількістю бітів з метою підвищення функціональної ефективності вирішальних правил системи ідентифікації об'єктів на місцевості;

2) розробка критерію функціональної ефективності машинного навчання системи ідентифікації об'єктів інтересу на місцевості для оптимізації параметрів моделі аналізу даних та прийняття рішень;

3) автоматизація формування вхідного математичного опису здатної самонавчатися системи ідентифікації об'єктів на місцевості, що до теперішнього часу все ще здійснюється людиною – розробником інформаційного забезпечення;

4) дослідження впливу структури та протоколів машинного навчання екстрактора ознакового опису спостережень на функціональну ефективність автономної навігації та класифікаційного аналізу спостережень.

Таким чином, аналіз існуючих рішень для реалізації автономної бортової системи ідентифікації об'єктів інтересу на місцевості показав необхідність підвищення достовірності рішень, що ними приймаються, з одночасним зниженням обчислювального навантаження на бортову систему. При цьому обґрунтовано доцільність використання як напрямку дисертаційних досліджень ідей і методів нейромережевої та інформаційно-екстремальної інтелектуальної технологій аналізу і синтезу здатних навчатися систем розпізнавання образів.

2.2 Формалізована постановка задачі інформаційного синтезу автономної системи ідентифікації об'єктів на місцевості, що навчається

Нехай дано алфавіт класів розпізнавання $\{X_m^o \mid m = \overline{1, M}\}$, де M — кількість класів, що характеризують об'єкт інтересу на зображенні, апріорно класифікована

навчальна матриця $\{y_{m,i}^{(j)}\}, i = \overline{1, N}, j = \overline{1, n}$, де N – розмірність вхідного зображення, n – кількість навчальних зразків кожного класу, множина прямокутників $\{r_m^{(j)}\}$, що задають межі об’єкта інтересу на зображенні, де $r_m^{(j)} = \{x_m^{(j)}, y_m^{(j)}, w_m^{(j)}, h_m^{(j)}\}$ – координати верхнього лівого кутка, ширина та висота обмежувального прямокутника відповідно на j -му зображенні m -го класу, апріорно некласифікована навчальна матриця зображень місцевості $\{\tilde{y}_i^{(j)}\}, i = \overline{1, N}, j = \overline{1, \tilde{n}}$, де \tilde{n} – обсяг апріорно некласифікованих навчальних зразків, для якої задана відповідна множина координат локалізації камери спостереження $p^{(j)} = \{p_x^{(j)}, p_y^{(j)}, p_z^{(j)}\}, j = \overline{1, \tilde{n}}$.

Відомий вектор параметрів функціонування системи розпізнавання, який у загальному випадку має таку структуру:

$$g = \langle g_1, \dots, g_{\xi_1}, \dots, g_{\Xi_1}, f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2} \rangle, \Xi_1 + \Xi_2 = \Xi,$$

де $\langle g_1, \dots, g_{\xi_1}, \dots, g_{\Xi_1} \rangle$ – генотипні параметри функціонування системи інтелектуального аналізу спостережень, які впливають на параметри формування ознакового опису спостережень;
 $\langle f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2} \rangle$ – фенотипні параметри функціонування системи, які прямо впливають на точність ідентифікації об’єкта інтересу на місцевості.

При цьому відомі обмеження на відповідні параметри функціонування:

$$R_{\xi_1}(g_1, \dots, g_{\xi_1}, \dots, g_{\Xi_1}) \leq 0; R_{\xi_2}(f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2}) \leq 0.$$

Необхідно:

1. Треба на етапі навчання системи ідентифікації об'єктів на місцевості отримати оптимальний вектор $\{g_{\xi}^* | \xi = \overline{1, \Xi_1 + \Xi_2}\}$ параметрів функціонування, який забезпечує максимум комплексного критерію ефективності навчання

$$g^* = \arg \max_G \{E_{cls}(g) + E_{rec}(g) + E_{loc}(g)\},$$

де L_{cls} — критерій ефективності класифікації реалізацій класів з $\{X_m^0\}$;

L_{rec} — критерій ефективності локалізації об'єкта інтересу на зображенні;

L_{loc} — критерій ефективності локалізації камери спостереження на місцевості;

G – область допустимих значень параметрів функціонування системи.

Метод інформаційного синтезу здатної навчатися автономної бортової системи БЛА для ідентифікації об'єктів на місцевості повинен відповідати таким основним вимогам:

- прямо та об'єктивно характеризувати функціональну ефективність системи;

- бути в рамках методології аналізу даних об'єктно-орієнтованим методом проектування;

- у рамках детерміновано-статистичного підходу корегувати математичний опис на вході системи ідентифікації об'єктів інтересу на місцевості з метою побудови вирішальних правил, які забезпечує на екзаміні повну достовірність рішень, наближену до максимальної асимптотичної, що визначається ефективністю навчання системи;

- забезпечувати високу функціональну ефективність навчання системи ідентифікації об'єктів інтересу на місцевості за умов нестационарності середовища та мінливості об'єктів інтересу, що має місце на практиці;

– оптимізувати просторово-часові параметри функціонування здатної навчатися системи ідентифікації об'єктів інтересу на місцевості за комплексним критерієм функціональної ефективності;

– розв'язувати задачі прогнозування зміни функціональної ефективності та надійності системи ідентифікації об'єктів інтересу на місцевості з метою їх корекції;

Таким чином, задача інформаційного синтезу здатної навчатися автономної бортової системи ідентифікації об'єктів інтересу на місцевості полягає в оптимізації за комплексним критерієм ефективності параметрів функціонування системи, які впливають на її ефективність.

2.3 Оцінка функціональної ефективності навчання системи ідентифікації об'єктів інтересу на місцевості

Центральною задачею розробки інформаційно-екстремальних методів аналізу та синтезу здатних навчатися систем класифікаційного аналізу спостережень є оцінка їх функціональної ефективності, що має інформаційну природу і визначає максимальну граничну достовірність рішень, що приймаються в робочому режимі (режимі екзамену). Тому для оцінки функціональної ефективності навчання системи розпізнавання в рамках ІЕІ-технології може бути використаний будь-який статистичний інформаційний критерій, який має властивості інформаційних мір і є функціоналом від точнісних характеристик, що утворюють систему оцінок рішень, що приймаються.

Як відомо критерій функціональної ефективності (КФЕ), щоб відповідати вимогам до інформаційних мір, має бути:

а) знакододатною дійсною функцією від імовірності, яка досягає екстремуму при значенні ймовірності

$$p_i = \frac{1}{M},$$

- де М – кількість альтернативних гіпотез при прийнятті рішення;
- б) дорівнювати нулю для детермінованих змінних ($p_i = 1$ або $p_i = 0$);
- в) відповідати принципу адитивності.

При цьому перевагу слід віддавати статистичним логарифмічним критеріям, оскільки вони, завдяки згладжуючому ефекту логарифмічної функції, забезпечують високу узагальнюючу здатність вирішальних правил, синтезованих за навчальною вибіркою відносно малого розміру. Серед таких критеріїв найбільшого використання знайшли ентропійна міра К. Шеннона та інформаційна міра С. Кульбака [9],[10].

Спочатку розглянемо модифіковану в праці [2] інформаційну міру С. Кульбака для двоальтернативних рішень, яка обчислюється як добуток логарифмічного відношення правдоподібності у вигляді відношення обчисленої на k -му кроці машинного навчання повної ймовірності правильного прийняття рішень $P_{t,m}^{(k)}$ до повної ймовірності помилкового прийняття рішень $P_{f,m}^{(k)}$, на міру відхилень відповідних ймовірностей. Для двоальтернативних гіпотез модифіковану міру Кульбака як КФЕ навчання системи розпізнавати реалізації класу X_m^o представимо у вигляді

$$E_m^{(k)} = \left[P_{t,m}^{(k)} - P_{f,m}^{(k)} \right] \cdot \log_2 \frac{P_{t,m}^{(k)}}{P_{f,m}^{(k)}} =$$

$$= \{ p(\mu_m) [p(\gamma_1^{(k)} / \mu_m) - p(\gamma_2^{(k)} / \mu_m)] + p(\mu_c) [p(\gamma_2^{(k)} / \mu_c) - p(\gamma_1^{(k)} / \mu_c)] \} \times$$

$$\times \log_2 \frac{p(\mu_m) p(\gamma_1^{(k)} / \mu_m) + p(\mu_c) p(\gamma_2^{(k)} / \mu_c)}{p(\mu_m) p(\gamma_2^{(k)} / \mu_m) + p(\mu_c) p(\gamma_1^{(k)} / \mu_c)},$$

де $P_{t,m}^{(k)}$ – повна ймовірність правильного прийняття рішень про належність реалізацій класу X_m^o та сусіднього йому класу X_c^o до контейнера класу X_m^o , побудованому на k -му кроці навчання діагностичної системи розпізнавати

реалізації класу X_m^o ;

$P_{f,m}^{(k)}$ – повна ймовірність помилкового прийняття рішень;

$\gamma_1^{(k)}$ – апіорна основна гіпотеза про належність до контейнера класу

X_m^o своїх реалізацій;

$\gamma_2^{(k)}$ – апіорна альтернативна гіпотези про належність до контейнера

класу X_m^o реалізацій класу X_c^o ;

μ_m – апостеріорна гіпотеза про належність до контейнера класу X_m^o своєї реалізації;

μ_c – апостеріорна гіпотеза про належність до контейнера класу X_m^o реалізації класу X_c^o ;

$p(\mu_m)$ – безумовна ймовірність прийняття гіпотези μ_m ;

$p(\mu_c)$ – безумовна ймовірність прийняття гіпотези μ_c ;

$p(\gamma_1^{(k)} / \mu_m)$ – перша достовірність $D_{1,m}^{(k)}$, яка є апіорною умовною

ймовірністю прийняття гіпотези $\gamma_1^{(k)}$ за умови, що має місце апостеріорна гіпотеза μ_m ;

$p(\gamma_2^{(k)} / \mu_c)$ – друга достовірність $D_{2,m}^{(k)}$;

$p(\gamma_2^{(k)} / \mu_m)$ – помилка першого роду $\alpha_m^{(k)}$;

$p(\gamma_1^{(k)} / \mu_c)$ – помилка другого роду $\beta_m^{(k)}$.

Згідно з принципом Бернуллі-Лапласа, виправдано прийняття рівно-ймовірних гіпотез

$$p(\mu_m) = p(\mu_c) = 0,5,$$

оскільки при оцінці ефективності функціонування системи розпізнавання, що

навчається, апріорна інформація є неповною. Реалізація цього принципу вимагає прийняття рішень діагностичною системою за найгіршими у статистичному розумінні умовами її функціонування. Тому гарантується, що покращення умов функціонування системи не знизить її функціональну ефективність, а навпаки підвищить. З урахуванням прийнятих допущень остаточною формулою модифікації критерію Кульбака для рівноймовірних двоальтернативних гіпотез має вигляд

$$E_m^{(k)} = 0,5 \cdot [(D_{1,m}^{(k)} + D_{2,m}^{(k)}) - (\alpha_m^{(k)} + \beta_m^{(k)})] \cdot \log_2 \left(\frac{D_{1,m}^{(k)} + D_{2,m}^{(k)}}{\alpha_m^{(k)} + \beta_m^{(k)}} \right) = \\ = [D_{1,m}^{(k)} - \beta_m^{(k)}] \cdot \log_2 \left(\frac{1 - [D_{1,m}^{(k)} - \beta_m^{(k)}]}{1 + [D_{1,m}^{(k)} - \beta_m^{(k)}]} \right). \quad (2.1)$$

Навчальна вибірка є обмеженою за обсягом, тому оцінку точнісних характеристик здійснюють за відповідними емпіричними частотами

$$D_{1,m}^{(k)} = \frac{K_{1,m}^{(k)}}{n_m}; \quad \beta_m^{(k)} = \frac{K_{2,m}^{(k)}}{n_m}, \quad (2.2)$$

де $K_{1,m}^{(k)}$ – кількість подій, що характеризують належність реалізацій класу X_m^o

до контейнера класу X_m^o на k -му кроці машинного навчання;

$K_{2,m}^{(k)}$ – кількість подій, що характеризують належність реалізацій класу

X_c^o до контейнера класу X_m^o на k -му кроці машинного навчання;

n_m – обсяг навчальної вибірки класу X_m^o .

При цьому $n_m \geq n_{\min}$, де n_{\min} – мінімальний обсяг репрезентативної навчальної вибірки, який визначається згідно з працею [9].

Робоча формула модифікації КФЕ за Кульбаком після відповідної підстановки оцінок (2.1) у вираз (2.2) набуває вигляду

$$E_m^{(k)} = \frac{1}{n_m} [K_{1,m}^{(k)} - K_{2,m}^{(k)}] \cdot \log_2 \left[\frac{10^{-\omega} + n_m + [K_{1,m}^{(k)} - K_{2,m}^{(k)}]}{10^{-\omega} + n_m - [K_{1,m}^{(k)} - K_{2,m}^{(k)}]} \right], \quad (2.3)$$

де $10^{-\omega}$ – будь-яке мале додатне число, яке вводиться для уникнення невизначеності при діленні на нуль.

На практиці ціла величина ω задається в інтервалі $1 < \omega \leq 3$, тобто дорівнює кількості знаків мантиси в значенні КФЕ.

Для загального випадку багатоальтернативного рішення, коли потужність алфавіту класів більша двох ($M > 2$), задачі оцінки функціональної ефективності та оптимізації параметрів мультикласових вирішальних правил найчастіше зводяться до відповідних задач серії двокласових вирішальних правил, оскільки це дозволяє суттєво підвищити оперативність навчання, не втрачаючи в достовірності результуючого класифікатора. Для реалізації зведення багатокласової оптимізаційної задачі до серії двокласових використовують такі основні методи [22, 23]:

1) «кожний проти всіх» – під час побудови вирішального правила для класу X_m^o кожна j -та реалізація цього класу вважається позитивною, а реалізації решти класів – негативними;

2) «кожний проти найближчого сусіда» – побудова оптимальних роздільних гіперповерхонь здійснюється між найближчими класами розпізнавання, оскільки згідно з гіпотезою компактності (чіткою або нечіткою) вплив реалізацій далеких класів на геометричні параметри вирішальних правил виключається;

3) «кожний проти кожного» – побудова $M(M-1)/2$ класифікаторів, що розрізняють всі пари класів;

4) дихотомія – побудова двійкового дерева, в якому $M-1$ нетермінальним вузлам відповідають двоальтернативні класифікатори, а M термінальним – класи.

Використання принципів методу «кожний проти всіх» призводить до проблеми незбалансованості класів і, відповідно, зменшення середньої точності вирішальних правил. Методи «кожний проти кожного» та дихотомія в рамках ІЕІ-

технології передбачають наявність декількох систем контрольних допусків, що ускладнює структуру вирішальних правил та підвищує обчислювальну складність алгоритмів навчання та екзамену. Найбільш привабливим для застосування в методах інформаційно-екстремального навчання є принцип «кожний проти найближчого сусіда», оскільки відповідає найпростішій структурі вирішальних правил і спрощує процедури оптимізації геометричних параметрів розбиття простору ознак на класи еквівалентності. У базовому алгоритмі ІЕІ-технології близькість класів визначається кодовою відстанню між їх двійковими еталонними (усередненими) векторами, однак це призводить до ігнорування дисперсії векторів-реалізацій (діаметру нечітких множин). Тому для обґрунтування вибору найближчого сусіднього класу слід використовувати оптимізаційну процедуру

$$E_m^* = \min_{c \neq m} \{E_{m,c}^*\},$$

де $E_{m,c}^*$ – значення критерію при оптимальну радіусі гіперсферичного контейнера класу X_m^o за умови, що клас X_c^o обрано як сусідній.

Таким чином, критерій функціональної ефективності діагностичної системи, що навчається, має інформаційну природу і зв'язок з точністними характеристиками.

ВИСНОВКИ

У науково-дослідній роботі розв'язано важливу науково-технічну прикладну задачу аналізу і синтезу здатної навчатися автономної бортової системи БЛА для ідентифікації об'єктів на місцевості. Головні наукові та практичні результати роботи полягають у такому:

1. Незважаючи на широке впровадження інтелектуальних інформаційних технологій для підвищення функціональної ефективності автономних бортових систем ідентифікації об'єктів на місцевості проблема забезпечення прийняття рішень в реальному темпі часу та адаптації до зміни умов функціонування і задач за умов обмежених обчислювальних ресурсів все ще залишається невирішеною.

2. Ускладнення вирішення проблеми забезпечення високої достовірності рішень, що приймаються в реальному темпі часу автономною бортовою системою БЛА, для ідентифікації об'єктів на місцевості за умов нестаціонарності середовища та мінливості об'єктів інтересу пов'язані з відсутністю науково-методологічних основ аналізу і синтезу інтелектуальних систем розпізнавання образів в рамках функціонально-вартісного аналізу.

3. Запропоновано структурну та компоновочну схеми БЛА для проведення досліджень. Розроблено структурні та функціональні схеми програмного та інформаційного забезпечення БЛА, що передбачає інтеграцію різномірної сенсорної інформації для екстракції комплексних ознак, які використовуються алгоритмами навігації, детектування перешкод та об'єктів інтересу, що дозволяє підвищити надійність системи за рахунок взаємної компенсації недоліків окремих сенсорних систем.

4. Самонавчання екстрактора, оснований на композиції згорткових шарів, для ознакового опису середовища є одним із шляхів вирішення проблеми адаптації моделі даних до середовища функціонування. При цьому узагальнюючі властивості моделі подання даних залежать від структури і глибини моделі, обсягу навчальних зразків. Однак питання зниження ресурсних потреб для функціонування таких моделей в режимах навчання та екзамену в реальному

темпі часу є відкритим. Розробка технологій вирішення даної проблеми є активною областю досліджень.

5. Ефективність поширених архітектур детектора об'єктів на зображенні, таких як R-CNN, Fast R-CNN, Faster R-CNN, Detect Net, Yolo та SSD, залежить від ефективності використаного екстрактора ознак та класифікатора ознакового опису об'єктів інтересу.

6. Запропоновано здійснювати синтез класифікатора об'єктів інтересу в рамках інформаційно-екстремальної інтелектуальної технології, яка дозволяє побудувати високодостовірні радіально-базисні вирішальні правила в бінарному просторі Хеммінга шляхом адаптивного двійкового кодування спостережень на основі обчислювально ефективних операцій порівняння та виключаючого АБО. При цьому вперше запропоновано здійснювати самонавчання ієрархічного екстрактора ознакового опису спостережень на основі принципів розрідженого кодування та самоорганізації неструктурованих сіток.

7. Запропоновано критерій функціональної ефективності здатних навчатися систем детектування об'єктів інтересу на місцевості у вигляді модифікацію інформаційної міри Кульбака для двоальтернативної системи оцінок рішень. Також розглянуто обчислювальний аспект критерію при зведенні мультиальтернативного рішення до серії двоальтернативних за принципом «кожний проти найближчого сусіда», що спрощує процедуру оптимізації геометричних параметрів розбиття простору ознак на класи еквівалентності в рамках ІЕІ-технології.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Москаленко В. В. Метод навчання без вчителя ієрархічного екстрактора візуальних ознак на основі модифікації нейронного газу / В. В. Москаленко, А. С. Москаленко, А. Г. Коробов // *Радіоелектронні та комп'ютерні системи*. – 2017. – № 3 (83). – С. 56–61.
2. Moskalenko V. V. Information-extreme algorithm of the system for recognition of objects on the terrain with optimization parameter feature extraction / V. V. Moskalenko, A. G. Korobov // *Radio Electronics, Computer Science, Control*. – 2017. – № 2. – С. 38-45. – DOI 10.15588/1607-3274-2017-2-7.
3. Moskalenko V. V. Development of the method of unsupervised training of convolutional neural networks based on neural gas modification. / V. V. Moskalenko // *Proc. of the International scientific and practical conference «Technology Transfer: fundamental principles and innovative technical solutions: Physical Sciences and Engineering» Tallinn, Estonia, 23.11.2017.* – 2017. – P. 21-27.
4. Коробов А. Г. Автономна система навігації безпілотного літального апарату / Коробов А. Г. // *Матеріал доповіді на IV Міжнародній науково-технічній конференції “Інформатика, управління і штучний інтелект”, м. Харків, 21 – 23 листопада 2017.* – 2017. – С. 10.
5. Bengio Y. Representation Learning: A Review and New Perspectives / Y. Bengio, A. Courville, P. Vincent // *CoRR*. – 2014. – V. abs/1206.5538. – Mode of access: <https://arxiv.org/pdf/1206.5538.pdf>. – Title from the screen.
6. Yongtao Y. Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data / Y. Yongtao, L. Jonathan, W Chenglu, G. Haiyan, L. Huan, W. Cheng // *ISPRS Journal of Photogrammetry and Remote Sensing*. – 2016. – V. 113. – P. 106-123. DOI: 10.1016/j.isprsjprs.2016.01.005.
7. Schmidhuber J. Deep Learning in Neural Networks: an Overview [Electronic resource] / J. Schmidhuber // *Neural Networks*. – 2015. – V. 61. – P. 85–117. – DOI:10.1016/j.neunet.2014.09.003

8. Mhaskar H. N. Deep vs. Shallow Networks: an Approximation Theory Perspective [Electronic resource] / H. N. Mhaskar, T. Poggio // CoRR. – 2016. – V. abs1608.03287. – Mode of access: <https://arxiv.org/pdf/1608.03287.pdf>. – Title from the screen.

9. Comaschi F RASW: A run-time adaptive sliding window to improve Viola-Jones object detection / F. Comaschi, S Stuijk, T Basten, H Corporaal // Seventh International Conference on Distributed Smart Cameras, Palm Springs, California, USA, October 29. – Nov. 1. 2013: proceedings. – IEEE Washington DC, USA, 2013.

10. Ammour N. Deep Learning Approach for Car Detection in UAV Imagery / N. Ammour, H. Alhichri, Y. Bazi, B. Benjdira, et al. // Remote Sens. – 2017. – V. 9. – I. 4. – P. 2–15.

11. Олізаренко С. А. Розробка архітектури нечіткої згорточної нейронної мережі для розпізнавання компактних (точкових) об'єктів на цифровому аерофотознімку [Електронний ресурс] / С. А. Олізаренко, В. А. Капранов, Р. В. Сафронов // Системи озброєння і військова техніка. – 2016. – № 4. – С. 38–41. – Режим доступу: http://nbuv.gov.ua/UJRN/soivt_2016_4_10

12. Ren Sh. Deep vs. Shallow Networks: an Approximation Theory Perspective [Electronic resource] / Sh. Ren, K. He, R. Girshick, Jian Sun // CoRR. – 2016. – V. abs1506.01497. – Mode of access: <https://arxiv.org/pdf/1506.01497.pdf>. – Title from the screen.

13. Grejner-Brzezinska D.A. Positioning in GPS-challenged Environments: Dynamic Sensor Network with Distributed GPS Aperture and Intermodal Ranging Signals / D.A. Grejner-Brzezinska, C.K. Toth, L. Li, J. Park et al. // Proceedings of the 22nd International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2009). – Savannah, GA: Savannah International Convention Center – 2009. – P. 111-123.

14. Ayoul Th., Buckley T., Crevier F. UAV Navigation above Roads Using Convolutional Neural Networks [Electronic resource] / Th. Ayoul, T. Buckley, F. Crevier; Mode of access: World Wide Web: <http://cs231n.stanford.edu/reports/2017/pdfs/553.pdf>. – Title from the screen.

15. Mohanty V. DeepVO: A Deep Learning approach for Monocular Visual Odometry [Electronic resource] / V. Mohanty, S. Agrawal, S. Datta, A. Ghosh et al. // CoRR. – 2016. – V. abs/1611.06069. – Mode of access: <https://arxiv.org/pdf/1611.06069.pdf>. – Title from the screen.

16. Smolyanskiy N. Toward Low-Flying Autonomous MAV Trail Navigation using Deep Neural Networks for Environmental Awareness [Electronic resource] / N. Smolyanskiy, A. Kamenev, J. Smith, S. T. Birchfield // CoRR. – 2017. – V. abs/1705.02550. – Mode of access: <https://arxiv.org/pdf/1705.02550.pdf>. – Title from the screen.

17. Kim D. K. Deep Neural Network for Real-Time Autonomous Indoor Navigation [Electronic resource] / Dong Ki Kim, Tsuhan Chen // CoRR. – 2015. – V. abs/1511.04668v2. – Mode of access: <https://arxiv.org/pdf/1511.04668v2.pdf>. – Title from the screen.

18. Muller M. S. Squeezeposenet: image based pose regression with small convolutional neural networks for real time uas navigation / M. S. Muller, S. Urban, B. Jutzi // ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. – Bonn, Germany. – V. IV-2/W3. – 2017. – P. 49–57. – Mode of access: <http://uavg17.ipb.uni-bonn.de/wp-content/papercite-data/pdf/mueller2017uavg.pdf>.

19. Wu A. D. Autonomous Flight in GPS-Denied Environments Using Monocular Vision and Inertial Sensors / A. D. Wu, E. N. Johnson, M. Kaess, F. Dellaert et al // Journal of Aerospace Information Systems. – 2013. – V. 10. – No. 4. – P. 172–186. – DOI:10.2514/1.I010023.

20. Huang L. Experimental study on the curve evolution method for noise filtering in the inertial state estimation / L. Huang, B. Lotfi // Control Automation and Robotics (ICCAR) 2017 3rd International Conference on, 2017. – P. 95-98.

21. Scott R. Autonomous Navigation and Hazard Evasion Platform for Personal UAV's [Electronic resource] / R. Scott // Mode of access: <http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1411&context=honorsthesis>. – Title from the screen.