

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Електронної та біомедичної інженерії  
(повна назва)

Кафедра Мікроелектроніки, електронних приладів та пристроїв  
(повна назва)

**КВАЛІФІКАЦІЙНА РОБОТА**  
**Пояснювальна записка**

Рівень вищої освіти другий(магістерський)

Розробка та дослідження методів комплексування при виявленні  
об'єктів в системах технічного зору  
(тема)

Виконав:

Здобувач 2 року навчання  
групи ЕППМ-24-1

Очеретін В.Д.

(прізвище, ініціали)

Спеціальність 171 Електроніка

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Електронні прилади та  
пристрої

(повна назва освітньої програми)

Керівник д.т.н, професор кафедри МЕЕПП

Стрількова Т. О.

(посада, прізвище, ініціали)

Допускається до захисту

Заф. кафедри

(підпис)

Бондаренко І. М.

(прізвище, ініціали)

2025 р.

## Харківський національний університет радіоелектроніки

Факультет Електронної та біомедичної інженерії  
(повна назва)

Кафедра Мікроелектроніки, електронних приладів та пристроїв  
(повна назва)

Рівень вищої освіти другий(магістрський)

Спеціальність 171 – Електроніка  
(код і назва спеціальності)

Тип програми освітньо-професійна  
(освітньо-професійна або освітньо-наукова)

Освітня програма Електроні прилади та пристрої  
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ:

Заф. кафедри \_\_\_\_\_  
(підпис)

«\_\_» \_\_\_\_\_ 20\_\_ р.

## ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Очеретіну Вячеславу Дмитровичу  
(прізвище, ім'я, по батькові)

1. Тема роботи РОЗРОБКА ТА ДОСЛІДЖЕННЯ МЕТОДІВ КОМПЛЕКСУВАННЯ ПРИ ВИЯВЛЕННІ ОБ'ЄКТІВ В СИСТЕМАХ ТЕХНІЧНОГО ЗОРУ

Затверджена наказом університету від 24 11 2025 р. № 1056Ст

2. Термін подання студентом роботи до екзаменаційної комісії \_\_\_\_\_ 2025 р.

3. Вихідні дані до роботи: \_\_\_\_\_

1. Синтетичні зображення видимого та (VIS) та інфрачервоного (IR) діапазонів

2. Параметри моделювання деградації зображень

4. Перелік питань, що потрібно опрацювати в роботі \_\_\_\_\_

1. Огляд систем технічного зору та принципів багатоспектрального спостереження

2. Методи обробки та комплексування зображень у VIS та IR діапазонах

3. Методи машинного навчання у задачах комплексування

4. Аналіз показників якості зображення та відношення сигнал/шум

5. Розробка алгоритму адаптивного комплексування

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри)\_\_\_\_\_

Презентаційний матеріал – 15 слайдів

### КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналіз технічного завдання	04.08.25	
2	Аналіз літературних джерел та нормативної документації	21.08.25	
3	Систематизація джерел, обробка та постановка завдання	02.09.25	
4	Вибір математичної моделі для алгоритму	07.09.25	
5	Проведення фізико-математичного моделювання процесу комплексування	26.09.25	
6	Аналіз алгоритму SNR-based методу	13.10.25	
7	Оформлення пояснювальної записки	23.11.25	
8	Підготовка презентації	28.11.25	
9	Рецензування, нормоконтроль, антиплагіат		

Дата видачі завдання 4 серпня 2025 р.

здобувач \_\_\_\_\_

(підпис)

Керівник роботи \_\_\_\_\_ проф. каф. МЕЕПІ Стрілкова Т. О.

(підпис)

(посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка містить: 91 сторінку, 25 рисунків, 3 таблиці, 28 джерела, 2 додатки

СИСТЕМИ ТЕХНІЧНОГО ЗОРУ, КОМПЛЕКСУВАННЯ, ВІДНОШЕННЯ СИГНАЛ/ШУМ, SNR, ГЛОБАЛЬНІ ОЗНАКИ, ВИДИМИЙ ТА ІНФРАЧЕРВОНИЙ КАНАЛ, ЕНТРОПІЯ

Об'єкт дослідження – процес формування та обробки багатоканальних зображень у системах технічного зору (видимий та інфрачервоний канали) в умовах деградації якості спостереження.

Мета роботи – розробити та дослідити метод адаптивного комплексування VIS та IR зображень для підвищення інформативності сцени при погіршенні умов спостереження, з автоматичним визначенням порогового діапазону якості за SNR і підтримкою прийняття рішення на основі ознак та дерева рішень.

Метод – аналітичний та математично-імітаційний для моделювання деградації VIS/IR кадрів і оцінювання якості за ознаками з подальшим комплексуванням.

Актуальність – актуальність роботи зумовлена потребою у підвищення надійності виявлення та спостереження об'єктів у системах технічного зору в умовах погіршення видимості.

У роботі сформовано набір синтетичних сцен із нелінійною деградацією у часі, реалізовано механізм автоматичного виділення сигналу й розрахунок набору інформативних ознак для VIS та IR, побудовано SNR-орієнтовану схему комплексування та отримано кількісне підтвердження покращення сигнальної складової на комплексованих кадрах. Розроблено процедуру автоматичного визначення порогового діапазону якості VIS за SNR і навчено дерево рішень, яке в перехідній зоні використовує додаткові статистичні та структурні ознаки.

## ABSTRACT

Explanatory note to the qualification work: 91 pages, 25 figures, 3 tables, 28 references, 2 appendices

MACHINE VISION SYSTEMS, FUSION, SIGNAL-TO-NOISE RATIO, SNR, GLOBAL SIGNS, VISIBLE AND INFRARED CHANNELS, ENTROPY

The object of research is the process of forming and processing multi-channel images in vision systems (visible and infrared channels) in conditions of deterioration of the observed quality.

The purpose of the work is to develop and investigate a method of adaptive fusion of VIS and IR images to increase the informativeness of the scene when the observation conditions are violated, with automatic determination of the SNR quality threshold and support for decision-making based on features and decision trees.

The method is analytical and mathematical-simulation for modeling the degradation of VIS/IR frames and assessing quality by features with subsequent fusion.

Relevance - the relevance of the work is due to the need to increase the reliability of detection and observation of objects in vision systems in conditions of reduced visibility.

In the work, a set of synthetic scenes with nonlinear degradation in time was formed, a mechanism for automatic signal removal and calculation of a set of informative labels for VIS and IR was implemented, an SNR-oriented fusion scheme was constructed and multiple confirmation of the improvement of the signal component on complex frames was obtained. A procedure for automatically determining the VIS threshold quality for SNR was developed and a decision tree was trained, which uses additional statistical and structural features in the transition zone.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	8
ВСТУП.....	9
1 СИСТЕМИ ТЕХНІЧНОГО ЗОРУ. КОМПЛЕКСУВАННЯ .....	10
1.1 Системи технічного зору (СТЗ): Структура та принцип роботи .....	10
1.2 Комплексування інформації в СТЗ .....	15
1.2.1 Комплексування на рівнях. ....	18
1.2.2 Часове та просторове комплексування.....	26
1.2.3 Комплексування VIS–IR .....	28
1.3 Технічні особливості та обмеження сучасних систем технічного зору .....	32
2 ДОСЛІДЖЕННЯ МЕТОДІВ ОБРОБКИ ТА АЛГОРИТМІВ КОМПЛЕКСУВАННЯ ЗОБРАЖЕНЬ У СТЗ.....	36
2.1 Методи машинного навчання в задачах комплексування зображень СТЗ....	36
2.1.1 Поняття машинного навчання та місце ML у задачах комплексування..	36
2.1.2 Типи задач ML у СТЗ .....	37
2.1.3 Методи ML на основі ознак.....	39
2.2 Локальні перетворення та фільтрація зображень у СТЗ.....	42
2.2.1 Операція згортки та фільтри згладжування .....	42
2.2.2 Градієнтні оператори.....	44
2.2.3 Статистичні та інформаційні ознаки для оцінки якості зображення .....	45
2.3 Фільтр Калмана .....	48
2.3.1 Математична модель фільтра Калмана.....	49
2.3.2 Використання КФ для багатосенсорного комплексування.....	52
2.3.3 Оцінка стану при шумових моделях .....	53
2.4 Класичні методи комплексування зображень.....	54
2.4.1 РСА .....	55
2.4.2 Вейвлет-комплексування .....	57
2.4.3 Лапласіан .....	59
2.5 Порівняння методів комплексування та виявлення недоліків .....	61
2.5.1 Недоліки класичних та ML – орієнтованих методів .....	61

2.5.2 Неможливість універсального застосування .....	64
3 РОЗРОБКА АЛГОРИТМУ КОМПЛЕКСУВАННЯ НА ОСНОВІ SNR – ОРІЄНТОВАНОЇ ОЦІНКИ ЯКОСТІ ЗОБРАЖЕННЯ.....	66
3.1 Постановка задачі та логіка побудови алгоритму VIS-IR комплексування ..	66
3.2 Механізм формування локальних та статистичних ознак для VIS та IR каналів .....	70
3.3 Математична модель SNR-орієнтованого злиття зображень .....	78
3.4 SNR-орієнтована модель прийняття рішень: навчання, інтерпретація та оцінювання алгоритму .....	84
ВИСНОВКИ.....	88
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	89
ДОДАТОК А .....	<b>Помилка! Закладку не визначено.</b>
ДОДАТОК Б .....	<b>Помилка! Закладку не визначено.</b>

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

СТЗ – Системи технічного зору

SNR – (Signal-to-Noise Ratio)

VIS – Видимий канал (Visible)

IR – Інфрачервоний канал (Infrared)

ML – Машинне навчання (Machine Learning)

## ВСТУП

Сучасні системи технічного зору та відеоспостереження широко застосовуються для виявлення й супроводження об'єктів у реальних умовах, де освітленість і стан атмосфери можуть швидко змінюватися. У видимому діапазоні (VIS) якість зображення істотно залежить від рівня освітлення: при затемненні зростає вплив шуму, зменшується контраст і погіршується роздільність корисної сигнальної складової від фону. У таких ситуаціях інфрачервоний канал (IR) часто забезпечує кращу спостережуваність об'єкта, однак має власні обмеження та особливості відображення сцени. Тому практично важливою є задача комплексування (злиття) VIS та IR даних для отримання більш інформативного та стійкого результату.

Ключовою проблемою комплексування є автоматичне визначення моменту, коли VIS-канал перестає бути надійним, а також вибір вагового внеску кожного каналу без ручного налаштування під конкретну сцену. У роботі як основний критерій якості використано відношення сигнал/шум (SNR), яке прямо відображає роздільність об'єкта на фоні та є природною основою для прийняття рішення про необхідність комплексування. Для підвищення надійності алгоритму додатково формуються статистичні та структурні ознаки (контраст, характеристики фону, ентропія, градієнтні показники, геометрія виділеної області), які особливо важливі у «перехідній зоні», коли однієї SNR-нормалізації недостатньо для стабільного рішення.

Метою роботи є розробка та дослідження методу автоматичного комплексування VIS та IR кадрів, який на основі ознак якості (насамперед SNR) здатний самостійно визначати порогові режими, класифікувати стан зображення за рівнями якості та обирати коректну стратегію обробки. Для досягнення мети сформовано синтетичну базу сцен із часовою деградацією кадрів, реалізовано автоматичне виділення сигнальної області без апіорних знань про її координати, проведено оцінювання характеристик для різних умов (затемнення та туман).

# 1 СИСТЕМИ ТЕХНІЧНОГО ЗОРУ. КОМПЛЕКСУВАННЯ

## 1.1 Системи технічного зору (СТЗ): Структура та принцип роботи

Системи технічного зору(СТЗ) є ключовим інструментом для автоматизованого отримання та аналізу інформації про об'єкти та середовище, у якому вони знаходяться. На відміну від традиційних оптико-електронних засобів, орієнтованих переважно на візуальне спостереження людиною, системи технічного зору формують та інтерпретують зображення у режимі реального часу з метою подальшої роботи алгоритмів виявлення, класифікації, контролю та прийняття рішень. У сучасних умовах розвитку інтелектуальних датчиків, автономних платформ і роботизованих систем саме технічний зір визначає можливість функціонування пристроїв у різноманітних та складних умовах спостереження. Це обумовлює потребу не лише в якісній апаратній основі, але й у розробці адаптивних алгоритмів, здатних забезпечувати стабільність роботи незалежно від шумових, атмосферних чи освітлювальних факторів.

Системи технічного зорус у більшості застосувань виступають центральною ланкою у процесах автоматизації, оскільки забезпечують можливість отримання даних у формах, придатних для машинної обробки. СТЗ виконують функцію формування “машинно-читаних” представлень сцени, у яких зображення розглядається не як інтерпретаційний об'єкт для оператора, а як множина вимірювальних параметрів, таких як інтенсивності, градієнтів, спектральних характеристик та просторових структур. Саме завдяки цьому СТЗ здатні автоматизувати які є надто складними для візуального спостереження через певні лімітуючі фактори, такі як швидкість події.

На відміну від людини, яка сприймає сцену комплексно та контекстною, технічна система покладається на вимірювані параметри сигналу: інтенсивність, контраст, спектральні характеристики та відношення сигнал/шум. Завдяки цьому СТЗ здатні працювати там, де оператор не може виконати спостереження через

обмеження видимості, перевантаженість інформацією або необхідність високої швидкості реакції[1].

Відеоспостереження є ключовою складовою сучасних систем безпеки, однак традиційні камери, що працюють виключно у видимому спектрі, суттєво обмежені в умовах низької освітленості та несприятливої погоди. Це зумовлено тим, що інтенсивність відбитого світла різко зменшується у нічний час, за туману, опадів чи задимлення, що призводить до втрати контрасту та деградації зображення. У спеціалізованих галузях, таких як охоронні комплекси, технічна діагностика, промислова автоматизація та автономні платформи, системи технічного зору виступають основою для забезпечення безпеки та стабільного функціонування обладнання. Завдяки використанню спеціалізованих сенсорів і алгоритмів обробки зображень такі системи дозволяють здійснювати цілодобовий моніторинг і виявлення об'єктів навіть у складних умовах спостереження, де традиційні засоби візуалізації втрачають ефективність[2].

У таких умовах традиційне відеоспостереження часто втрачає свою ефективність, адже зображення перестає містити достатньо корисної інформації. Натомість системи технічного зору не просто фіксують сцену, а й здатні оцінювати якість сигналу в реальному часі та автоматично змінювати режим роботи. Завдяки цьому вони можуть застосовувати комплексування й адаптивну обробку даних, зберігаючи стабільність та надійність навіть тоді, коли видимий канал вже не забезпечує належної інформативності.

Основними вимогами до технічного зору є високий динамічний діапазон, стійкість до шумів та здатність працювати при низькому контрасті сцени. Системи повинні зберігати стабільність роботи за рахунок внутрішніх механізмів компенсації: адаптивної обробки, корекції оптичних спотворень, фільтрації та інтелектуальних методів оцінки інформативності. У більшості сучасних платформ аналіз сигналу здійснюється у реальному часі, що накладає вимоги на ефективність алгоритмів обробки та оптимізацію всієї архітектури СТЗ[3].

Подальше функціонування системи технічного зору визначається її апаратно-програмною структурою, яка забезпечує формування зображення, його

перетворення, попередню цифрову обробку та подальший аналіз. У загальному випадку робота СТЗ включає послідовність наступних етапів: оптичне формування зображення, реєстрацію випромінювання сенсором, аналого-цифрове перетворення, попереднє покращення якості кадру та інтелектуальне виділення інформативних ознак. Така структурна організація дозволяє системі стабільно функціонувати у змінних умовах середовища та забезпечувати достовірність аналізу сцени незалежно від зовнішніх факторів.

Для узагальнення принципу роботи СТЗ доцільно розглянути її структурну модель, наведено у вигляді блок-схеми (рис. 1.1).

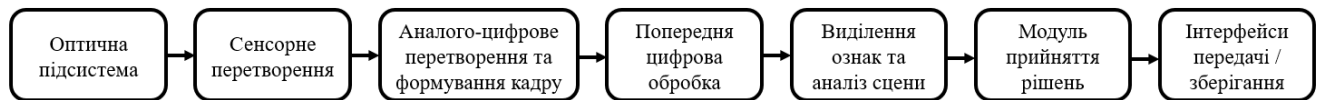


Рисунок 1.1 – Структурна схема типової системи технічного зору

Після формування зображення в оптичній підсистемі світловий потік потрапляє на сенсор, який перетворює його на електричний сигнал. У системах видимого спектра здебільшого застосовуються CMOS або CCD матриці, у ближньому IR — фотодіодні матриці InGaAs, а в довгохвильовому IR діапазоні — мікроболометричні сенсори. Вибір сенсора визначає спектральні можливості СТЗ, рівень шумів, чутливість та здатність працювати при низькому освітленні.

Формування зображення на сенсорі описується моделлю згортки, яка відображає вплив оптики та шумів на кінцевий результат:

$$I(x, y) = O(x, y) \cdot h(x, y) + n(x, y); \quad (1.1)$$

де  $I(x, y)$  — сформоване зображення;

$O(x, y)$  — оптичне відображення об'єкта;

$h(x, y)$  — точкова функція розсіювання (PSF), що характеризує роздільну здатність та оптичні аберації системи;

$n(x, y)$  — шум сенсора.

PSF визначає, наскільки точно система відтворює дрібні деталі та контрасти об'єкта: чим ширша PSF, тим «розмитішим» буде зображення, і тим важче системі виділити контури, текстури чи малі об'єкти.

Подальше аналого-цифрове перетворення включає операції корекції темного струму, компенсації шуму читання, підсилення сигналу та лінеаризації, формуючи цифровий кадр, придатний для подальшої обробки.

На етапі попередньої цифрової обробки система виконує комплекс процедур покращення якості кадру, необхідних для забезпечення достатньої інформативності перед подальшим аналізом. До таких процедур належать:

- фільтрація шумів (медіанна, гаусова, адаптивна);
- компенсація нерівномірності (NUC), що виправляє неоднорідність чутливості матриці;
- нормалізація яскравості, яка забезпечує рівномірність освітлення кадру;
- корекція спектрального відгуку сенсора, що компенсує нерівномірність спектральної чутливості;
- підвищення локального контрасту, що дозволяє виділити об'єкти на складному або темному фоні.

У сучасних системах технічного зору цей етап попередньої обробки реалізується апаратно або за допомогою високопродуктивних обчислювальних модулів. Це дає змогу виконувати всі необхідні операції швидко й забезпечувати стабільну роботу системи навіть у режимі реального часу.

Після покращення зображення система переходить до етапу виділення ознак, на якому формуються структурні характеристики сцени: контури, градієнти, ключові точки, текстурні параметри та сегментовані області. Цей модуль є критичним для подальшої інтерпретації сигналу, оскільки саме ознаки визначають, які інформаційні компоненти будуть використані у наступних алгоритмах. Для їх виділення застосовуються як традиційні методи комп'ютерного зору (каннінга, Sobel, Canny, Watershed), так і сучасні глибокі нейронні мережі, що здатні автоматично формувати ієрархії ознак.

Завершальним етапом роботи структурної схеми є модуль прийняття рішень, у якому СТЗ здійснює інтерпретацію сцени: визначення наявності об'єктів, їх класу, положення та параметрів, аналіз можливих сценаріїв поведінки або оцінку технічного стану. У залежності від конкретної задачі цей блок може виконувати виявлення загроз, ініціювати сигнал тривоги, видавати управляючі команди або передавати результати до зовнішніх систем.

Показовими прикладами реальних систем технічного зору є багатоспектральні комплекси FLIR PT-Series(рис 1.2, а), камери міського спостереження Hikvision DarkFighterX(рис 1.2, б), модуль DJI Zenmuse H20T, що використовується у професійних безпілотних платформах (рис. 1.2, в) і поєднує тепловізійний канал, видимий CMOS-сенсор, гібридний оптичний зум та лазерний далекомір та інші інтегровані рішення.



Рисунок 1.2 – Приклади систем технічного зору у системах відеоспостереження

Такі системи використовують методи високочутливої сенсорики, спектральної адаптації та інтелектуальної обробки для забезпечення стабільної роботи в умовах, де традиційні камери втрачають ефективність. Застосування багатоспектральних сенсорів дозволяє отримувати доповнюючу інформацію з різних діапазонів. Наприклад, тепловізійний канал зберігає контрастність сцени, коли видимий канал повністю втрачає інформативність через темряву або метеорологічні фактори.

У сучасних системах особливе значення має здатність оцінювати якість сцени та адаптувати режими роботи під зовнішні умови. Оскільки інформативність зображення безпосередньо залежить від відношення сигнал/шум (SNR), цей показник виступає базовим критерієм, що визначає можливість коректної роботи алгоритмів детекції та ідентифікації. Для цифрових систем технічного зору, зокрема при аналізі реальних або синтетично згенерованих кадрів, SNR доцільно визначати як відношення різниці середніх інтенсивностей сигналу та фону до шумової складової:

$$SNR = \frac{\mu_{obj} - \mu_{bg}}{\sigma_{noise}}; \quad (1.2)$$

де:  $\mu_{obj}$  — середня яскравість області з об'єктом (ROI);

$\mu_{bg}$  — середня яскравість фону;

$\sigma_{noise}$  — стандартне відхилення шуму, оцінене на фоновій області або з темного кадру.

Падіння SNR призводить до втрати контрасту, появи артефактів і зростання ймовірності помилкового виявлення об'єктів. Саме аналіз SNR дозволяє визначити момент, коли зображення у видимому діапазоні стає недостатньо інформативним, а активування інфрачервоного каналу або режиму комплексування стає доцільним. Це забезпечує об'єктивний критерій перемикання між каналами та дозволяє уникнути ситуацій, коли система продовжує працювати у режимі з недостатньою якістю даних [2].

## 1.2 Комплексування інформації в СТЗ

Комплексування інформації є одним із ключових напрямів розвитку сучасних систем технічного зору, оскільки дозволяє підвищити якість отримуваних даних, розширити діапазон умов роботи та забезпечити стійкість алгоритмів аналізу сцени у складних середовищах спостереження. На відміну від традиційних оптико-електронних систем, які працюють з одним джерелом інформації, СТЗ сучасного покоління дедалі частіше інтегрують дані з різних сенсорів,

спектральних діапазонів або часових кадрів, формуючи більш повне та інформативне представлення оточення. Такий підхід дозволяє отримати характеристики сцени, які недоступні окремому каналу, та компенсувати фізичні обмеження сенсорів, що є критично важливим у задачах моніторингу, безпеки, автономної навігації та промислової діагностики.

Комплексування розглядається як процес поєднання інформації з декількох джерел для формування більш точного, надійного та повного уявлення про середовище, а також для зменшення невизначеності та підвищення достовірності прийняття рішень.

У контексті СТЗ комплексування набуває додаткового сенсу, оскільки візуальна інформація природно чутлива до освітлення, шуму та атмосферних факторів. Зокрема, у видимому діапазоні якість зображення різко залежить від рівня освітленості, тоді як інфрачервоний діапазон краще працює при нестачі світла, але має нижчу просторову роздільну здатність. Саме тому сенсорні канали часто мають комплементарні властивості, що створює основу для їх поєднання[4].

Водночас у своїй роботі Khaleghi та співавтори підкреслюють[4], що багатосенсорна інтеграція виконує не лише просте “складання” інформації, а й систематичне зменшення шумів, відмовостійкість та підвищення надійності системи, що особливо важливо для СТЗ, які працюють в реальному часі. Тобто, комплексування виступає механізмом стабілізації інформаційного потоку: навіть при деградації одного із сенсорів система здатна зберігати високу якість даних за рахунок додаткових каналів або часових кадрів.

Важливою особливістю комплексування в СТЗ є те, що кожен спектральний канал має власні фізичні обмеження, які визначають його інформативність у різних умовах. Видимий діапазон забезпечує високу просторову деталізацію та точне відтворення текстур, але його робота залежить від зовнішнього освітлення й атмосферних умов. Натомість інфрачервоний канал зберігає стабільність у ситуаціях низької освітленості, туману або задимленості, формуючи тепловий контраст незалежно від наявності видимого світла. Зображення у видимому діапазоні містять детальну просторову інформацію, тоді як інфрачервоні

зображення забезпечують стабільне відображення об'єктів за умов недостатнього освітлення[5], що наочно демонструє їхню взаємодоповнюваність. Саме тому об'єднання цих каналів дозволяє отримати більш повну модель сцени: структурна й текстурна інформація видимого спектра доповнюється стійкими температурними характеристиками IR-діапазону. Комплементарна природа VIS та IR робить їх особливо придатними для комплексування. У результаті інтеграція спектральних каналів виступає одним із найефективніших способів підвищення надійності, інформативності та стійкості систем технічного зору в широкому спектрі умов експлуатації.

Окремим різновидом комплексування, який відіграє важливу роль у системах технічного зору, є багатокадрове або часове комплексування. Його суть полягає у поєднанні кількох послідовних кадрів одного сенсора для зменшення шумів і підвищення стабільності сигналу в умовах низького освітлення або різких коливань яскравості. На відміну від спектрального поєднання каналів, часовий підхід працює у межах одного діапазону, але дозволяє суттєво збільшити інформативність зображення за рахунок усереднення випадкових шумових компонентів.

Цей принцип лежить в основі так званої «інтеграції по експозиціях» та використовується у нічних камерах, тепловізорах і промислових СТЗ. Теоретично якість зображення при усередненні  $N$  кадрів покращується згідно зі співвідношенням:

$$SNR_N = SNR_1 \sqrt{N} \quad (1.3)$$

Комплексування у системах технічного зору спрямоване на розв'язання широкого спектра ключових проблем, пов'язаних із нестабільністю та неповнотою інформації при роботі окремих сенсорів. Воно дозволяє зменшити залежність системи від зовнішніх факторів — таких як освітлення, погодні умови або рівень шумів — і забезпечити стабільну інформативність навіть тоді, коли один із каналів втрачає якість. Об'єднання даних із різних спектральних діапазонів, сенсорних модальностей або часових кадрів створює багатовимірне, узгоджене представлення

сцени, яке значно повніше, ніж інформація, отримана з одного джерела. Завдяки цьому СТЗ можуть компенсувати недоліки окремих сенсорів та виділяти нові ознаки, що залишаються недоступними в одноканальних системах.

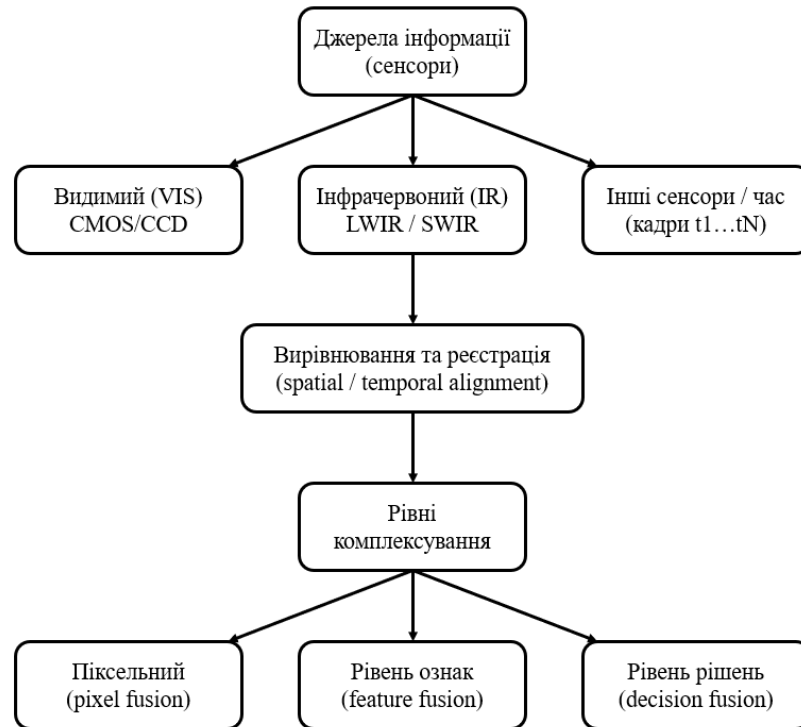


Рисунок 1.3 – Структурна схема процесу комплексування інформації в СТЗ

Процес комплексування включає послідовність етапів — від реєстрації первинних сигналів до їхнього узгодження, нормалізації та інтеграції на різних рівнях: піксельному, ознаковому або рівні прийняття рішень. Ці етапи формують цілісний механізм, який забезпечує узгоджене об'єднання даних та підвищення надійності роботи СТЗ. Узагальнена структура такого процесу подана на рис. 1.3, де послідовно відображено основні стадії формування та поєднання інформації.

### 1.2.1 Комплексування на рівнях.

В контексті багатоканальних систем технічного зору, комплексування має здійснюватися з урахуванням структури даних і фізичних особливостей сигналів, що надходять з різних сенсорів. Кожен канал — видимий, інфрачервоний, тепловізійний або спектрально-специфічний — формує власне, часткове уявлення про сцену, і тому інтеграція цих даних потребує чітко визначених механізмів

об'єднання. ефективність комплексування залежить від того, на якому рівні відбувається інтеграція сигналів — на рівні сирих пікселів, виділених ознак або вже сформованих рішень. Така класифікація визначає, яку інформацію система зберігає, які спотворення компенсує та наскільки стійкою буде до шумів і деградації окремих каналів. Саме тому аналіз рівнів комплексування є базовим етапом при проектуванні СТЗ, що повинні працювати у змінних та потенційно складних умовах спостереження.

Комплексування на рівні пікселів. Комплексування на рівні пікселів передбачає, що інтеграція даних виконується безпосередньо над початковими значеннями зображення, тобто над окремими значеннями яскравості кожного пікселя, ще до будь-яких етапів фільтрації, сегментації чи виділення ознак. Такий підхід працює з максимально «сирими» даними з усіх сенсорів, що дозволяє зберегти повний обсяг інформації про сцену. Саме тому комплексування на цьому рівні є найбільш обчислювально складним, але потенційно й найточнішим, адже будь-які дрібні відмінності між каналами враховуються вже на старті. Коли втрати на етапі формування кадру мінімальні, кінцевий результат найбільше залежить від того, яке правило об'єднання використано та які алгоритми стоять після нього, що робить піксельне комплексування базовим і водночас найбільш вимогливим етапом інтеграції в системах технічного зору[6].

На практиці такий підхід передбачає, що всі канали попередньо приведені до єдиного масштабу, геометрично вирівняні та радіометрично узгоджені, аби відповідні координати містили інформацію про одну й ту саму точку сцени. Після такого вирівнювання інтеграція виконується безпосередньо над значеннями яскравості, що дозволяє уникнути втрат, пов'язаних із проміжними етапами обробки, та зберегти максимально високий рівень деталізації й контрастності.

У загальному випадку результат комплексування двох каналів системи технічного зору наприклад, видимого  $I_{VIS}(x,y)$  та інфрачервоного  $I_{IR}(x,y)$  можна записати у вигляді операторного виразу:

$$F(x, y) = \Phi(I_{VIS}(x, y), I_{IR}(x, y)), \quad (1.4)$$

Базовим способом інтеграції є застосування простих арифметичних операторів. Наприклад, арифметичне середнє дозволяє отримати зображення, у якому інтенсивності інформаційних каналів збалансовано об'єднуються. У цьому випадку значення результативного зображення в точці  $(x, y)$  визначається як[7]:

$$F(x, y) = \frac{I_1(x, y) + I_2(x, y)}{2} \quad (1.5)$$

Такий підхід певною мірою згладжує шум, але водночас зменшує локальний контраст та розмиває деталі, що є небажаним для задач технічного зору з низьким SNR. Більш гнучким є зважене середнє[7,8]:

$$\begin{aligned} F(x, y) &= \omega_{\text{VIS}}(x, y)I_{\text{VIS}}(x, y) + \omega_{\text{IR}}(x, y)I_{\text{IR}}(x, y), \\ \omega_{\text{VIS}}(x, y) + \omega_{\text{IR}}(x, y) &= 1, \end{aligned} \quad (1.6)$$

де вагові коефіцієнти можуть залежати від локальної інформативності каналів. Наприклад, у ділянках з добре вираженою текстурою у видимому діапазоні доцільно збільшувати  $\omega_{\text{VIS}}$ , а в умовах малої освітленості або туману — підвищувати  $\omega_{\text{IR}}$ , використовуючи оцінку локального SNR як критерій.

Якщо для кожного каналу відома оцінка відношення сигнал/шум  $\text{SNR}_{\text{VIS}}$  та  $\text{SNR}_{\text{IR}}$ , ваги можна обрати пропорційними до квадрата SNR:

$$\omega_{\text{VIS}} = \frac{\text{SNR}_{\text{VIS}}^2}{\text{SNR}_{\text{VIS}}^2 + \text{SNR}_{\text{IR}}^2}, \quad \omega_{\text{IR}} = \frac{\text{SNR}_{\text{IR}}^2}{\text{SNR}_{\text{VIS}}^2 + \text{SNR}_{\text{IR}}^2}. \quad (1.7)$$

У такому випадку канал з вищою якістю автоматично робить більший внесок у результат. Наприклад: якщо у сумерках  $\text{SNR}_{\text{VIS}} = 5$ , а тепловізійний канал дає  $\text{SNR}_{\text{IR}} = 15$  то  $\omega_{\text{VIS}} = 0,1$ ,  $\omega_{\text{IR}} = 0,9$ . Інтегроване зображення майже повністю спирається на IR-інформацію, зберігаючи лише залишкові деталі з видимого діапазону.

Окрім алгебраїчних правил, широко застосовуються піксельно-вибіркові оператори, коли в кожній точці сцени обирається яскравість того каналу, який є

більш інформативним за деяким локальним критерієм. Типовий приклад — правило максимуму:

$$F(x, y) = \begin{cases} I_{\text{VIS}}(x, y), & \text{якщо } I_{\text{VIS}}(x, y) \geq I_{\text{IR}}(x, y), \\ I_{\text{IR}}(x, y), & \text{інакше.} \end{cases} \quad (1.8)$$

У практиці СТЗ замість самих значень яскравості часто використовують величини градієнта або локальної дисперсії. Наприклад, можна порівнювати модулі  $|\nabla I_{\text{VIS}}|$  та  $|\nabla I_{\text{IR}}|$  і обирати той піксель, де зміна яскравості більша, що дозволяє віддати перевагу каналу з більш вираженими контурами.

Значну роль у сучасних алгоритмах комплексування на рівні пікселів відіграють багатомасштабні методи, які працюють не безпосередньо з матрицею яскравостей, а з її представленням у декількох просторово-частотних масштабах. При цьому початкові зображення розкладаються на набір компонент, що містять низькочастотну (глобальну) та високочастотну (детальну) інформацію. Формально для кожного каналу виконується перетворення[7]:

$$I_k(x, y) \xrightarrow{\tau} \{C_k^{(0)}(x, y), C_k^{(1)}(x, y), \dots, C_k^{(m)}(x, y)\}, \quad (1.9)$$

де  $C_k^{(0)}$  – наближена (низькочастотна) компонента;

$C_k^{(m)}$ ,  $m \geq 1$  – детальна компонента різних масштабів.

Далі для кожного рівня  $m$  задається власний оператор злиття. Орієнтуючись на формулу 1.4 маємо:

$$F^{(m)}(x, y) = \Psi^{(m)}(C_{\text{VIS}}^{(m)}(x, y), C_{\text{IR}}^{(m)}(x, y)), \quad (1.10)$$

Далі інтегроване зображення відновлюється за допомогою зворотного перетворення:

$$F(x, y) = \mathcal{J}^{-1}(\{F^{(m)}(x, y)\}_{m=0}^M). \quad (1.11)$$

На високочастотних рівнях зазвичай використовують правила, орієнтовані на збереження контрастних деталей (вибір коефіцієнта з більшою абсолютною величиною, з більшою енергією або локальною ентропією), тоді як для низькочастотної компоненти застосовується зважене середнє або оператор, що враховує глобальний SNR каналу.

Математично правило вибору за енергією на рівні  $m$  можна записати як:

$$F(x, y) = \begin{cases} C_{\text{VIS}}^{(m)}(x, y), & \text{якщо } E_{\text{VIS}}^{(m)}(x, y) \geq E_{\text{IR}}^{(m)}(x, y), \\ C_{\text{IR}}^{(m)}(x, y), & \text{інакше,} \end{cases} \quad (1.12)$$

де локальна енергія, наприклад, визначається згорткою:

$$E_k^{(m)}(x, y) = (C_k^{(m)}(x, y))^2 \cdot G_\sigma(x, y), \quad (1.13)$$

При застосуванні згортки з фільтром Гауса, в кожному масштабі відбирається канал де локальна структура більш виражена, що дозволяє поєднати контурну інформацію видимого діапазону з термальними контрастами IR-каналу.

Щоб проілюструвати загальну логіку багатомасштабного піксельного комплексування, доцільно використати узагальнену структурну схему, наведену в роботі Pixel-level image fusion: A survey of the state of the art. На ній показано, як вихідні зображення проходять послідовні етапи декомпозиції, об'єднання та синтезу, утворюючи інтегрований кадр, придатний для подальшого аналізу в СТЗ.

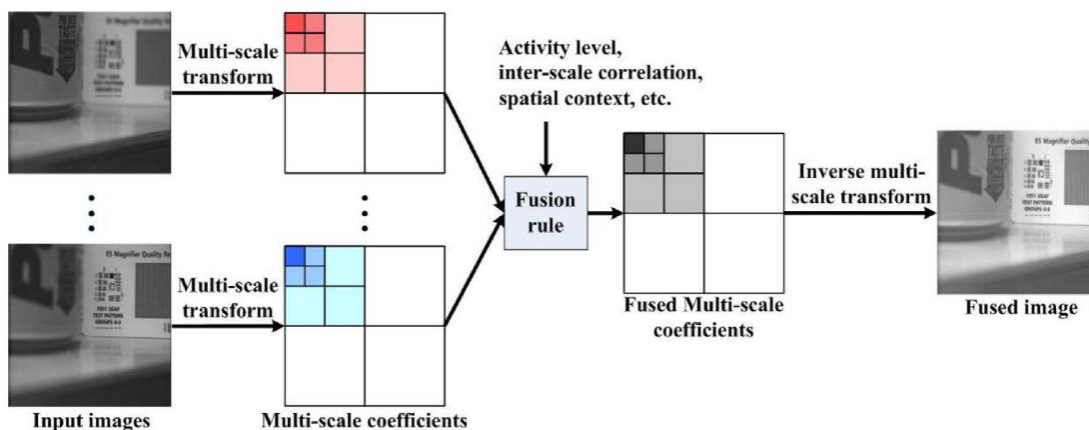


Рисунок 1.4 – Узагальнена схема багатомасштабного піксельного комплексування

Базові етапи, показані на цій схемі, добре узагальнюють типову послідовність операцій для більшості сучасних алгоритмів. Типовий багатомасштабний метод комплексування складається з трьох основних кроків: спочатку кожне вхідне зображення перетворюється у багаторівневе представлення, далі для кожного рівня застосовується відповідне правило об'єднання коефіцієнтів, після чого результуюче зображення відновлюється шляхом зворотного перетворення. Такий підхід дозволяє окремо керувати низькочастотними компонентами, які відповідають за загальну яскравість і контраст сцени, та високочастотними компонентами, що містять деталі й контури, що є особливо корисним при комплексуванні зображень з різних сенсорів[7].

Для систем технічного зору це означає, що на рівні пікселів можна одночасно зберегти глобальну структуру тепловізійного зображення (силуети об'єктів, температурні плями) і детальну текстуру видимого каналу (контури, дрібні об'єкти), що критично для задач виявлення цілей за умов низької освітленості, туману або диму. У подальших розділах саме цей підхід до багатомасштабного піксельного комплексування розглядатиметься як базис для побудови більш складних алгоритмів, керованих показником SNR.

Комплексування на рівні ознак. Комплексування на рівні ознак передбачає інтеграцію не вихідних зображень, а тих характеристик, які попередньо виділяються з кожного каналу: контурів, градієнтів, текстурних дескрипторів, локальних інваріантних точок або статистичних ознак. Такий підхід дає змогу працювати з більш компактним та інформативним представленням сцени, зменшуючи обчислювальні витрати порівняно із піксельним рівнем та підвищуючи стійкість до шумів, освітлення та артефактів.

Комплексування на рівні ознак інтегрує інформативні дескриптори, виділені з кожної модальності, зменшує надмірність та підвищує розрізнявальну здатність, уникаючи при цьому високих обчислювальних витрат, характерних для піксельного рівня [9]. Це твердження повністю узгоджується з природою VIS-IR систем, де теплові ознаки забезпечують стабільність у складних умовах, а ознаки видимого діапазону додають структурну деталізацію.

На відміну від піксельного рівня, де інтеграція виконується над сирими радіометричними значеннями, на рівні ознак кожен канал передає до системи вже інтерпретовану інформацію про сцену. Для видимого спектра це можуть бути контури об'єктів, градієнтні карти, SIFT/ORB-ключові точки або локальні гістограми градієнтів (HOG). Для інфрачервоного каналу — температурні області, сегментовані гарячі об'єкти, теплові контури або статистичні параметри розподілу температури. Об'єднання таких дескрипторів дозволяє формувати комбіновану ознакову множину, що містить як структурні, так і теплові характеристики сцени.

У загальному випадку ознаки кожного каналу формують вектори:

$$f_{\text{VIS/IR}} = (f_1^{\text{VIS/IR}}, f_2^{\text{VIS/IR}}, \dots, f_n^{\text{VIS/IR}}), \quad (1.14)$$

які інтегруються у спільний простір ознак за допомогою операторів конкатенації, вагового об'єднання або статистичного узгодження. Найпростіший варіант полягає у формуванні об'єданого ознакового вектора:

$$F = [f_{\text{VIS}} || f_{\text{IR}}], \quad (1.15)$$

що дозволяє зберегти повний набір характеристик, підвищивши інформативність моделі.

Однак у реальних СТЗ така стратегія може призвести до появи надмірних ознак або корельованих компонентів, що збільшує ризик перенавчання або знижує ефективність алгоритмів. Саме тому на рівні ознак часто застосовуються методи нормалізації, відбору ознак (feature selection) або зменшення розмірності (PCA, LDA), що дозволяє уникнути переобтяження простору ознак.

Ще одним поширеним підходом є вагове комплексування, коли висока інформативність одного каналу компенсує слабкість іншого. У нічних сценах теплові ознаки мають більшу дискримінативність, тому ваги для IR-компонент можуть бути збільшені; вдень — навпаки, сильні текстурні ознаки VIS-каналу домінують. Такий підхід робить систему більш гнучкою і дає їй можливість підлаштовуватися під зміну умов спостереження.

Особливо важливо, що комплексування на рівні ознак менш чутливе до невеликої геометричної невідповідності між каналами. Це є суттєвою перевагою для багатоспектральних камер, де ідеальна піксель-до-піксель відповідність недосяжна через різні оптичні тракти та фокусні відстані. Навіть при невеликих зміщеннях ключові точки, контури та дескриптори зображення залишаються узгодженими, що забезпечує роботу алгоритму у реальних, а не лабораторних умовах.

У сучасних СТЗ цей рівень комплексування є фундаментом для алгоритмів машинного та глибокого навчання, де об'єднані вектори ознак використовуються для класифікації, детекції та відстеження об'єктів. Саме завдяки інформації, отриманій зі структурних, текстурних і теплових характеристик, такі системи демонструють значно вищу стійкість у складних умовах: тумані, дощі, низькій освітленості або частковому перекритті об'єктів.

Комплексування на рівні рішень. Комплексування на рівні рішень є найбільш високорівневим способом інтеграції даних у системах технічного зору. На цьому етапі кожен канал або підсистема виконує повний цикл аналізу самостійно — від отримання зображення до формування висновків щодо наявності об'єктів, їх класів, стану чи інших параметрів. До модуля комплексування надходять уже готові рішення, представлені у вигляді міток, логічних висновків, ймовірнісних оцінок або координат.

Основна ідея полягає в тому, що система поєднує не первинні дані, а результати їх інтерпретації. Це дозволяє уникнути проблем, характерних для нижчих рівнів комплексування: невідповідності роздільної здатності, різної фізичної природи сигналів, шумів, геометричних спотворень або необхідності радіометричного узгодження. Завдяки такому підходу інтеграція стає гнучкішою, оскільки рішення можуть надходити від джерел, що працюють у різних режимах, мають різні моделі обробки або навіть різні типи даних.

Комплексування на рівні рішень зазвичай реалізується через механізми голосування, логічні правила, агрегування ймовірностей або комбінування оцінок достовірності. Перевага цього рівня полягає у високій завадостійкості, навіть якщо

один із модулів помиляється або працює в несприятливих умовах, система може скоригувати підсумок на основі інших джерел. Проте цей підхід має і обмеження: втрачається частина інформації, яка могла би бути корисною на ранніх етапах аналізу, а помилки, що виникають у процесі локальної обробки кожного каналу, вже неможливо компенсувати.

### 1.2.2 Часове та просторове комплексування.

Часове комплексування використовується в тих випадках, коли один сенсор спостерігає сцену протягом певного інтервалу часу, формуючи послідовність кадрів, у яких значення сигналу залишаються відносно стабільними, тоді як шум змінюється випадковим чином. Завдяки цьому інтеграція кількох кадрів дає змогу суттєво підвищити якість зображення без зміни апаратної частини системи. У багатьох сучасних дослідженнях *temporal fusion* розглядається як один із найбільш ефективних способів підвищення SNR у режимах низького освітлення, оскільки випадковий шум зменшується пропорційно  $\sqrt{N}$ , де  $N$  — кількість кадрів, що накопичуються[7,10]. На практиці це означає, що навіть прості методи усереднення або медіанного комбінування серії кадрів здатні витягнути структуру сцени, яка в окремих кадрах «втоплена» у шумі. На основі дослідів SNR-поведінки, проведеного для серій синтетичних кадрів із поступовим зниженням освітленості, було показано, що шум фону наростає значно швидше, ніж корисний сигнал[2].

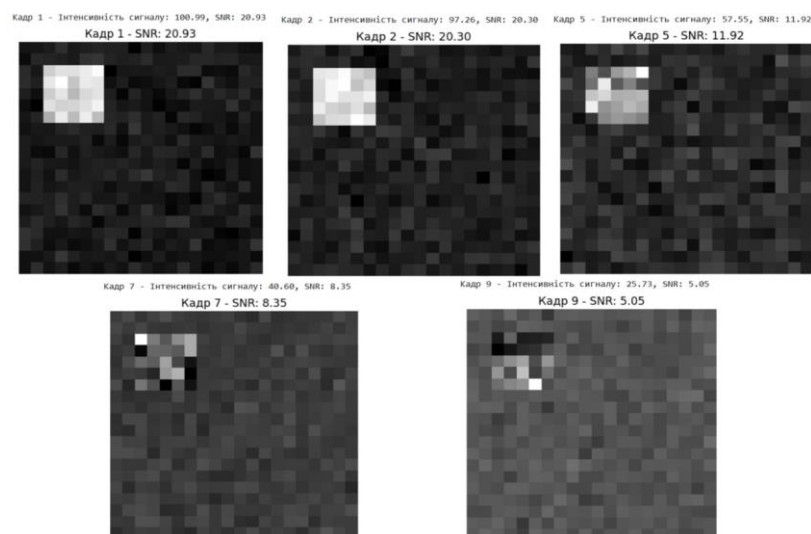


Рисунок 1.5 – Приклад часової деградації SNR у видимому діапазоні

З отриманих кадрів чітко проглядається поступова деградація зображення при зменшенні інтенсивності сигналу, що відповідає умовам зменшення освітленості. Цей процес чітко відображається на поступовому зниженні показника SNR: спочатку чітка та контрастна структура поступово розмивається, а деталі втрачають чіткість через наростаючий рівень шуму. Механізм підсилення сигналу сенсором камери, покликаний компенсувати брак освітлення, парадоксально призводить до того, що зображення замість покращення стає дедалі більш зашумленим та нерозбірливим. Це підтверджує складність забезпечення стабільної якості відеозображення в умовах обмеженої освітленості.

Для часових методів найбільш поширеною є проста модель усереднення:

$$F(x, y) = \frac{1}{N} \sum_{k=1}^N I_k(x, y), \quad (1.16)$$

яка застосовується у задачах шумозниження та накопичення сигналу. Однак сучасні СТЗ дедалі частіше використовують адаптивні алгоритми з ваговими коефіцієнтами, що залежать від локального SNR або стабільності руху[11].

Просторове комплексування передбачає поєднання даних, що належать до однієї сцени, але отримані з різних просторових точок, сенсорів або оптичних каналів. Це можуть бути:

- два об'єкти;
- мультиспектральні VIS–NIR–SWIR камери;
- тепловізійні модулі з різними полями зору;
- панхроматичні та мультиспектральні канали (як у супутникових системах).

Тут ключовим завданням є геометричне вирівнювання, тобто приведення всіх каналів до спільної системи координат. Після цього просторові дані інтегруються на рівні пікселів або ознак. Просторове комплексування дозволяє компенсувати втрату деталізації в одному каналі за рахунок високої просторової роздільності іншого[8].

### 1.2.3 Комплексування VIS–IR.

Комплексування видимого (VIS) та інфрачервоного (IR) діапазонів є одним із найбільш ефективних підходів до підвищення інформативності систем технічного зору, оскільки поєднує дві принципово різні форми візуальної інформації. Зображення у видимому спектрі забезпечують високу просторову роздільну здатність, чітку передачу контурів, текстур та дрібних деталей сцени, проте їхня якість різко падає за умов низької освітленості, перевищення динамічного діапазону або наявності атмосферних перешкод. Інфрачервоний спектр, навпаки, менш чутливий до таких факторів і зберігає стабільний тепловий контраст об'єктів навіть у нічний час, за туману, опадів чи задимленості. IR-зображення добре передають теплові цілі, але природно мають нижчу структурну деталізацію, тоді як VIS-дані містять багаті текстурні й геометричні ознаки, однак швидко деградують за низького рівня освітлення[12]. Саме цей контраст дозволяє виділяти об'єкти тоді, коли VIS-кадр втрачає інформативність через низький SNR.

У реальних умовах такі властивості відіграють взаємодоповнювальну роль: коли видимий канал «провалюється» через шум, недостатнє освітлення чи сильні засвічення, інфрачервоний продовжує формувати коректні сигнатури об'єктів; коли ж IR-канал не передає дрібної текстури, саме VIS відновлює геометрію сцени. Зміст вихідних сигналів є настільки різним, що коректне об'єднання можливе лише за умови попереднього узгодження. Це включає геометричне вирівнювання, підбір масштабу, корекцію паралаксу та приведення каналів до спільного радіометричного діапазону. Успішне комплексування можливе лише тоді, коли інфрачервоні та видимі зображення спочатку геометрично вирівняні та радіометрично нормалізовані. Без такого узгодження навіть найсучасніші алгоритми не зможуть коректно зіставити відповідні точки сцени, що призведе до артефактів та втрати інформативності. [13, 14].

Для того щоб зрозуміти, чому саме поєднання видимого та інфрачервоного каналів є настільки ефективним, доцільно порівняти їх властивості у типових умовах роботи СТЗ. Обидва спектральні діапазони мають принципово різну фізичну природу сигналу, а отже різні сильні й слабкі характеристики. Наведена

нижче таблиця узагальнює ключові параметри VIS та IR каналів, які визначають їхню інформативність та роль у процесі комплексування.

Таблиця 1.1 – Порівняння каналів

Характеристики	Видимий діапазон	Інфрачервоний діапазон
Природа сигналу	Відбите світло у діапазоні 400–700 нм	Власне теплове випромінювання (8–14 мкм)
Залежність від освітлення	Висока. Суттєва деградація вночі та при засвіченні	Мінімальна. Працює вночі, у тумані, при задимленні
Просторова деталізація	Висока (контури, текстури, дрібні структури)	Середня, контури згладжені
Контрастність об'єктів	Висока лише за достатнього освітлення	Висока завдяки температурному контрасту
Чутливість до атмосферних умов	Висока (туман, дощ, задимленість)	Низька (особливо LWIR)
Шуми сенсора	Низькі у денних умовах, різко зростають при темряві	Характерні шумові патерни (NUC), але стабільні при будь-якому освітленні

Порівняльний аналіз підтверджує, що VIS та IR канали мають виражено комплементарний характер: сильні сторони одного компенсують критичні недоліки іншого. Саме тому інтеграція спектральних діапазонів дає можливість отримати зображення, яке одночасно містить структурно детальну інформацію та зберігає високу стійкість до зовнішніх факторів. На практиці це дозволяє підвищити стійкість алгоритмів виявлення, покращити якість сегментації та зменшити кількість хибних спрацювань у складних умовах зйомки.

Комплексування видимого (VIS) та інфрачервоного (IR) діапазонів неможливе без формальної математичної моделі, оскільки саме вона визначає спосіб інтеграції структурної та теплової інформації у єдине представлення сцени. Після виконання попередніх етапів — геометричного вирівнювання, нормалізації та узгодження інтенсивностей — обидва зображення можна розглядати як матриці однакового розміру. Позначимо їх у загальному вигляді як:  $I_{VIS}(i, j)$  та  $I_{IR}(i, j)$ .

Після приведення обох каналів до спільної шкали виконується їхнє інтегрування у відповідності до загального принципу комбінування сигналів, сформульованого раніше (1.4). У найзагальнішому вигляді операцію злиття можна записати як:

$$F = W_{VIS} \odot \hat{I}_{VIS} + W_{IR} \odot \hat{I}_{IR}, \quad (1.17)$$

де  $\hat{I}_{VIS}, \hat{I}_{IR}$  – нормалізовані матриці яскравостей;

$W_{VIS}, W_{IR}$  – матриці вагових коефіцієнтівж;

символ  $\odot$  позначає поелементне множення.

Такий запис відображає базовий принцип VIS–IR злиття: кожна точка вихідного зображення є результатом комбінації відповідних точок обох спектральних каналів з урахуванням їх локальної інформативності.

Найпростішим випадком оператора (1.17) є використання глобальних ваг, однакових для всього кадру:

$$F = \omega \hat{I}_{VIS} + (1 - \omega) \hat{I}_{IR}, \quad (1.18)$$

де  $\omega$  — сталий коефіцієнт, що визначає внесок видимого каналу.

Такий підхід є прийнятним для однорідних сцен, проте в системах технічного зору використовується рідко. VIS та IR мають яскраво виражену просторову неоднорідність: температурний контраст добре проявляється лише на певних ділянках, тоді як структурні деталі важливі в інших. Тому фіксовані ваги не забезпечують достатньої адаптивності.

Ефективніші системи використовують локальні ваги, що змінюються для кожного пікселя. Логічною основою для цього є локальна оцінка співвідношення сигнал/шум, яка у попередніх розділах була визначена через формулу (1.7). Використовуючи цю оцінку, вагові коефіцієнти можна подати як:

$$\omega_{VIS}(i, j) = \frac{\omega_{VIS}(i, j)}{\omega_{VIS}(i, j) + \omega_{IR}(i, j)}, \quad \omega_{IR}(i, j) = 1 - \omega_{VIS}(i, j), \quad (1.19)$$

що означає автоматичне зростання ваги того каналу, який локально демонструє більш сприятливі умови формування сигналу. Якщо у певній області VIS-канал деградує через низьку освітленість, туман або перевищення динамічного діапазону, величина  $\omega_{VIS}$  зменшується, а внесок IR-каналу стає визначальним. І навпаки, при слабкому тепловому контрасті та високій структурній деталізації сцени алгоритм віддає перевагу видимому спектру.

Таким чином, модель (1.19) дозволяє формувати зображення, адаптивно оптимізоване під характеристики сцени, та отримувати збалансований результат, у якому водночас зберігається точна геометрія VIS-каналу та температурна сигнатура IR-каналу. Саме тому SNR-кероване локальне комплексування є найбільш природним і фізично обґрунтованим підходом у системах технічного зору, що працюють у змінних та неконтрольованих умовах.

Для кращого розуміння практичної цінності спектрального комплексування наведемо типовий приклад сцени, де видимий та інфрачервоний канали демонструють суттєво різну інформативність у різних частинах зображення. Такий формат ілюструє, які саме недоліки кожного діапазону компенсуються в процесі інтеграції та чому злитий кадр має вищу якість, ніж будь-який із початкових.

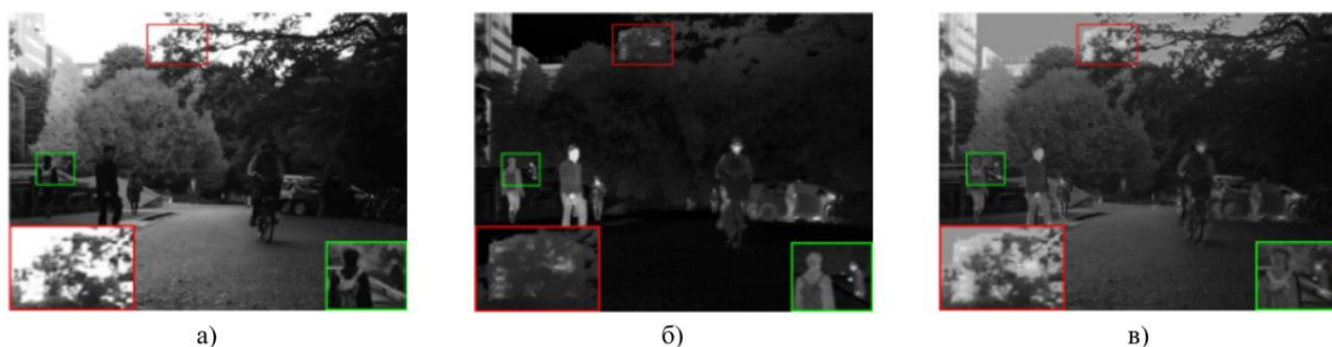


Рисунок 1.6 – Приклад інтегрованого VIS–IR зображення після багаторівневого піксельного комплексування

На рис. 1.5 наведено приклад комплексування міської сцени, що містить пішоходів і велосипедиста. На зображеному кадрі видимого діапазону(а) добре передано текстуру об'єктів і контури дерев, однак у затемнених ділянках

спостерігається суттєве падіння контрасту. Інфрачервоне зображення(б) демонструє протилежну властивість: теплові сигнатури людей залишаються чіткими незалежно від тіней і неоднорідного освітлення, проте дрібні структурні деталі практично відсутні. Третє зображення результат VIS–IR комплексування на якому поєднано деталі видимого спектра зі стійким тепловим контрастом IR-каналу.

Видимі зображення, як правило, втрачають контрастність у затінених областях, тоді як інфрачервоні зображення залишаються стабільними та зберігають теплові сигнатури пішоходів[15]. Це твердження точно відображає ситуацію на фрагментах прикладу: у VIS кадрі об'єкти частково втрачають контури, тоді як IR забезпечує стабільне відображення ключових цілей. Злитий результат демонструє, що саме комбінування структурної інформативності VIS та температурної інваріантності IR дозволяє отримати більш повне та надійне подання сцени.

### 1.3 Технічні особливості та обмеження сучасних систем технічного зору.

Функціональні можливості будь-якої системи технічного зору визначаються її технічними характеристиками, які задають межі, в яких система здатна коректно формувати інформативне зображення. Незалежно від модальності — видимий, ближній IR чи довгохвильовий тепловий діапазон — кожний канал має свою фізичну природу сигналу, тип шумів, динамічний діапазон, часову стабільність та залежність від зовнішніх умов. Саме ці обмеження визначають так звані граничні можливості СТЗ: максимальну дальність виявлення, мінімальну контрастність об'єкта, роботу за низького освітлення та здатність відтворювати структуру сцени в умовах шумів.

У видимому діапазоні ключовим параметром є освітленість. Фотонний потік визначає рівень сигналу, а отже й відношення сигнал/шум (SNR), яке є критичним для виявлення слабо контрастних об'єктів. За зниження освітлення SNR зменшується приблизно пропорційно до  $\sqrt{N}$ , де  $N$  — кількість зібраних фотонів за час експозиції. Таким чином, подовження експозиції (часу накопичення) дозволяє підвищити SNR, але вводить нові обмеження: збільшується розмиття рухомих об'єктів, зростає вплив шуму темного струму та фонове засвічення. Отже,

нарощування експозиції не є універсальним способом покращення якості — існує межа, після якої зображення деградує через рух або тепловий шум сенсора. Зі зменшенням освітлення змінюється і динамічний діапазон системи. Кадри, зняті в сутінках або при локальних джерелах світла, часто мають локальні перенасичення (пересвічені ділянки) та «провали» в тінях. У таких умовах навіть камера з високою світлочутливістю не здатна одночасно відтворити окремі ділянки сцени — виникають граничні стани, за яких одноканальна СТЗ більше не може забезпечити повну інформативність.

Інфрачервоний діапазон має протилежну фізичну специфіку. Тепловізійні сенсори не залежать від видимого освітлення, однак їх межі визначаються температурним контрастом сцени та тепловим шумом. Для нетеплоконтрастних об'єктів (наприклад, людей у теплих костюмах, поверхні будівель у термічній рівновазі) IR канал втрачає виразність: об'єкти стають малопомітними, а деталі — «пласкими». Динамічний діапазон тепловізора залежить від NETD (Noise Equivalent Temperature Difference), який задає мінімальний температурний градієнт, що може бути відділений від шуму. При високій вологості повітря та сильному тумані тепловий контраст падає, що наближає IR-кадр до шумоподібного стану. Таким чином, і в цьому спектральному діапазоні існують власні «граничні» умови, за яких система втрачає інформативність.

Саме зіткнення з граничними характеристиками окремих каналів у реальних умовах і пояснює необхідність розширення СТЗ за рахунок другого або третього каналу. Додавання IR-каналу у систему з видимою камерою змінює не стільки структуру системи, скільки її інформаційну доступність. Наприклад, у нічних умовах дальність виявлення візуальним каналом може знижуватися на порядок, у той час як тепловізійний канал зберігає максимальну дальність майже незмінною. У сумі це означає, що система зі спектральним комплексуванням має ефективну дальність виявлення, яка не визначається мінімальним, а близька до максимального значення серед каналів.

Якщо уявити динамічний діапазон, SNR та дальність виявлення як вектори можливостей, то одноканальна СТЗ має форму вузького «променя», тоді як

двоканальна або трьохканальна система — це значно ширша «область», яка охоплює різні умови середовища. Такий ефект розширення можливостей не є лінійним, а значною мірою визначається фізичною комплементарністю каналів. Зокрема, система з VIS+IR каналами отримує:

- вищу стабільність SNR у темряві;
- розширений діапазон контрастів, оскільки IR не насичується прямим світлом;
- повноту структурної інформації, недоступну для аналогічного тепловізора;
- меншу чутливість до руху, адже коротку експозицію VIS компенсує стійке IR-зображення;
- більшу відмовостійкість — деградація одного каналу не руйнує інформативність системи.

При додаванні третього каналу (наприклад, ближнього IR або MWIR/LWIR) система отримує можливість розрізняти тонкі матеріальні властивості об'єктів або підвищувати точність виявлення холодних/теплих цілей у складних середовищах. Це збільшує інформаційну ємність сцени та дозволяє працювати там, де два канали вже не гарантують стабільності.

Часова інтеграція також відіграє роль у формуванні технічних характеристик СТЗ. Послідовність кадрів може бути використана для накопичення сигналу (temporal denoising), що підвищує SNR, але за рахунок збільшення вимог до стабільності сцени. У нашій попередній тезі було показано, що погіршення освітленості з часом призводить до закономірного зниження SNR у VIS-кадрі, і лише додавання IR-каналу дозволяє компенсувати цей процес, зберігаючи здатність системи виявляти об'єкти на критичних ділянках сцени. Це демонструє фундаментальний взаємозв'язок між тимчасовими характеристиками СТЗ, рівнем шуму та можливістю інтегрувати канали для підвищення інформативності.

У сумі технічні характеристики систем технічного зору не є набором статичних параметрів — це динамічна система взаємопов'язаних обмежень, які в реальних умовах постійно змінюються. Одноканальна система технологічно

обмежена фізикою свого сенсора, тому її можливості завжди залишаються вузькими. Додавання другого або третього спектрального каналу радикально підвищує завадостійкість і дозволяє СТЗ працювати в умовах, де окремий канал втрачає інформативність.

## 2 ДОСЛІДЖЕННЯ МЕТОДІВ ОБРОБКИ ТА АЛГОРИТМІВ КОМПЛЕКСУВАННЯ ЗОБРАЖЕНЬ У СТЗ

### 2.1 Методи машинного навчання в задачах комплексування зображень СТЗ

#### 2.1.1 Поняття машинного навчання та місце МЛ у задачах комплексування.

Принципова відмінність МЛ від класичних “жорстких” алгоритмів полягає в тому, що пороги та межі між станами системи формуються на основі прикладів, а не задаються єдиними фіксованими правилами. У практичних задачах це означає, що система не оперує наперед визначеними граничними значеннями параметрів, а відновлює їх із статистики даних, відображаючи реальні умови спостереження. Для машинного навчання ключовими є поняття навчальних даних та моделі, яка формується за цими даними: алгоритм буде математичне відображення між вимірюваними характеристиками сцени та потрібним рішенням. У найпростішому представленні навчальна вибірка задається як множина пар:

$$D = \{(x_i, y_i)\}_{i=1}^N, \quad (2.1)$$

де  $x_i$  — вектор ознак (кількісних характеристик кадру або сцени);

$y_i$  — цільова мітка (наприклад, клас якості чи режим обробки);

$N$  — кількість прикладів.

Метою навчання є отримання відображення  $f(x)$ , яке відтворює зв’язок між ознаками та рішенням з мінімальною помилкою на прикладах і, водночас, зберігає здатність до узагальнення на нових сценах. У технічному зорі це означає, що межі між станами якості можуть визначатися не лише одним параметром, а комбінацією статистичних та структурних ознак, що описують деградацію сигналу в реальних умовах.

У задачах комплексування зображень у системах технічного зору МЛ доцільно трактувати як формальний механізм вибору режиму за поточною якістю спостереження. При цьому методи машинного навчання не підміняють класичні алгоритми обробки зображень, а використовують результати їх роботи у вигляді

узагальнених ознак для прийняття рішень на більш високому рівні. Тобто модель не замінює фізику формування зображення і не ускладнює сам процес отримання кадру, а впорядковує рішення між кількома станами якості на основі вимірюваних параметрів. За такого підходу вибір режиму (наприклад, доцільність комплексування або зміна внеску каналів) визначається як результат оцінювання якості, а не як застосування наперед заданих правил, що є принципово важливим у перехідних режимах та при значній варіативності сцен.

Окремо слід підкреслити, що в задачах оцінки якості зображень і їх подальшого комплексування використовувані ознаки мають володіти інженерно зрозумілим фізичним змістом. До таких ознак належать статистичні характеристики сигналу та фону, показники контрасту, оцінки рівня шуму, інформаційні міри на кшталт ентропії, а також узагальнені структурні характеристики, зокрема градієнтні показники різкості та деталізації. Застосування подібних ознак дозволяє безпосередньо пов'язати результати роботи моделі з реальними властивостями зображення та умовами спостереження.

У такій постановці машинне навчання доцільно розглядати як механізм адаптивного формування меж між різними зонами якості. Замість використання одного фіксованого порогового значення, модель на основі сукупності прикладів визначає узгоджені області у просторі ознак, що відповідають різним режимам обробки. За наявності перехідних або неоднозначних ситуацій рішення може уточнюватися за рахунок другорядних ознак, тоді як у простих режимах визначальну роль відіграють базові показники якості. Такий підхід забезпечує формалізоване та статистично обґрунтоване прийняття рішень, що є принципово важливим для задач комплексування в системах технічного зору.

### 2.1.2 Типи задач ML у СТЗ.

У системах технічного зору методи машинного навчання застосовуються для формалізованого аналізу візуальної інформації та прийняття рішень на основі кількісних характеристик зображення. Залежно від форми вихідного результату та способу використання навчальних даних методи ML у технічному зорі формулюються у вигляді різних типів задач, кожна з яких відповідає окремому

етапу обробки зорової інформації. До основних типів задач належать класифікація, детекція, регресія параметрів і кластеризація. Зазначені задачі відрізняються формою вихідного рішення, однак у межах єдиної системи вони тісно пов'язані між собою та можуть застосовуватися послідовно або комбіновано, забезпечуючи перехід від вимірювання параметрів сцени до обґрунтованого вибору режимів обробки.

Задачі класифікації в системах технічного зору полягають у поділі зображень, кадрів або їхніх областей на обмежену кількість станів відповідно до наперед заданих критеріїв. На відміну від безперервної оцінки параметрів, класифікація дозволяє перейти до дискретних рішень, що є зручними для керування режимами роботи системи. У технічному зорі такі класи можуть відповідати типам об'єктів, умовам спостереження або рівням якості зображення, що безпосередньо пов'язано з його інформативністю. У задачах комплексування зображень класифікація використовується для поділу кадрів за ступенем інформативності, наприклад на якісні, перехідні та деградовані, що дозволяє формалізувати вибір режиму обробки або визначити доцільність залучення додаткового каналу спостереження на основі сукупності статистичних та структурних ознак, а не за одним фіксованим пороговим значенням [16].

Задачі детекції в системах технічного зору спрямовані на виявлення та локалізацію інформативних об'єктів або областей у зображенні. На відміну від класифікації, де результатом є лише віднесення до певного класу, детекція передбачає визначення просторового положення сигналу та його відокремлення від фону. У технічному зорі детекція застосовується для виділення області інтересу, яка надалі використовується для аналізу властивостей сигналу та обчислення статистичних показників. У задачах оцінки якості та комплексування коректне виділення області сигналу є необхідною умовою для достовірного визначення параметрів якості, зокрема відношення сигнал/шум, а також для узгодження просторової локалізації сигналу з подальшим прийняттям рішень щодо режиму обробки [16].

Задачі регресії параметрів у технічному зорі пов'язані з оцінюванням неперервних величин, що характеризують стан сцени або якість зображення. До таких параметрів належать середні рівні інтенсивності, дисперсія, контраст, рівень шуму, а також інтегральні показники якості, зокрема SNR. На відміну від класифікації, регресійні методи не зводять рішення до дискретних класів, а дозволяють отримати кількісну оцінку, яка може бути використана для подальшого аналізу або керування. У задачах комплексування регресійна оцінка параметрів якості дає можливість не лише описати поточний стан зображення, але й використовувати ці значення як основу для формування вагових коефіцієнтів або вибору режимів обробки в умовах поступової деградації сигналу.

Кластеризація належить до задач некерованого навчання та застосовується для автоматичного групування даних без попереднього задання міток. У системах технічного зору кластеризація дозволяє виявляти природні групи зображень або кадрів, які мають подібні статистичні та структурні властивості. На відміну від класифікації, де класи визначаються наперед, кластеризація формує їх на основі внутрішньої структури даних. У задачах оцінки якості та комплексування зображень кластеризація може використовуватися для попереднього виділення зон якості, що відповідають різним рівням деградації сигналу. Отримані групи можуть надалі застосовуватися як основа для формування класів або для побудови моделей прийняття рішень без ручного задання порогових значень.

Зазначені типи задач машинного навчання в технічному зорі реалізуються на основі кількісного опису кадру у вигляді набору ознак, який відображає статистичні та структурні властивості зображення. Саме вибір інформативних ознак визначає, наскільки коректно модель відокремлює режими якості та забезпечує узгоджене прийняття рішень у простих і перехідних умовах спостереження.

### 2.1.3 Методи ML на основі ознак.

Методи машинного навчання на основі ознак ґрунтуються на попередньому кількісному описі зображення у вигляді набору параметрів, які відображають властивості сигналу, фону та їх взаємодії. На відміну від підходів, що працюють

безпосередньо з піксельними даними, ознакові методи оперують узагальненими характеристиками сцени, що дозволяє зменшити розмірність задачі та забезпечити більш стабільну поведінку моделей у змінних умовах спостереження. Такий підхід є характерним для інженерних систем технічного зору, де важливими є не лише точність, але й керованість та інтерпретованість прийнятих рішень.

У загальному випадку ознаки формуються як числові характеристики, що описують статистичні, контрастні, інформаційні або структурні властивості зображення. Вектор ознак використовується як вхід до моделі машинного навчання, яка встановлює залежність між цими параметрами та шуканим рішенням. При цьому рішення визначається не окремим показником, а їх сукупністю, що дозволяє враховувати складний характер деградації сигналу в реальних сценах. Така постановка задачі робить можливим формування адаптивних меж між різними станами якості без використання жорстко заданих порогових значень.

Методи машинного навчання на основі ознак широко застосовуються у задачах класифікації, регресії та кластеризації, які реалізуються за допомогою відносно простих, але ефективних моделей. До таких моделей належать дерева рішень, ансамблеві методи на їх основі, а також інші алгоритми, орієнтовані на роботу з табличними даними. Спільною рисою цих методів є можливість прийняття рішень на основі послідовного аналізу ознак та формування логічно зрозумілих правил, що є важливою вимогою для технічних систем обробки зображень.

Однією з найбільш поширених моделей у межах ознакового підходу є дерево рішень, яке реалізує процес прийняття рішення у вигляді ієрархічної структури умов. У вузлах дерева здійснюється перевірка окремих ознак або їх комбінацій, а кінцеві листові вершини відповідають певному класу або режиму роботи системи. Така модель дозволяє формалізувати процес вибору на основі кількох параметрів та забезпечує прозорість логіки прийняття рішення. У практичному сенсі це означає, що у простих режимах рішення може визначатися домінуючою ознакою, тоді як у перехідних ситуаціях враховується вплив додаткових характеристик.

У контексті технічного зору методи машинного навчання на основі ознак доцільно розглядати як проміжну ланку між класичними алгоритмами обробки зображень і статистичними моделями прийняття рішень. Вони не замінюють фізичні або математичні методи формування ознак, а використовують їх результати для впорядкованого та адаптивного вибору режимів обробки. Такий підхід дозволяє поєднати аналітичну обґрунтованість обчислюваних параметрів із гнучкістю машинного навчання, що є важливим для задач оцінки якості та комплексування зображень у системах технічного зору.

Для узагальнення підходу до використання машинного навчання у задачах комплексування зображень доцільно розглянути структурну схему (рис 2.1) процесу прийняття рішення. У такій схемі методи машинного навчання не підміняють етапи формування та первинної обробки зображення, а використовують кількісні ознаки, отримані в результаті детекції сигналу та аналізу його властивостей, для формалізованого вибору режиму обробки.

Методи машинного навчання у задачах комплексування не функціонують ізольовано, а спираються на результати попередньої обробки зображень та обчислення кількісних характеристик. Якість і інформативність рішень безпосередньо залежать від того, наскільки коректно сформовані ознаки відображають властивості сигналу, фону та їх взаємодії. У зв'язку з цим важливим етапом є використання локальних перетворень і фільтрації зображень, які забезпечують виділення інформативних областей та стабільне обчислення статистичних і структурних показників.



Рисунок 2.1 – Структурна схема ознакового підходу ML у СТЗ

## 2.2 Локальні перетворення та фільтрація зображень у СТЗ

### 2.2.1 Операція згортки та фільтри згладжування .

У системах технічного зору цифрове зображення можна розглянути як двовимірний дискретний сигнал  $I(x,y)$ , що формується в каналах VIS та IR і містить як корисну сигнальну складову, так і шумові та фонові компоненти. У реальних умовах спостереження (зниження освітленості, туман, нерівномірний фон, артефакти сенсора) випадкові флуктуації інтенсивності призводять до нестабільності параметрів кадру. Тому перед подальшою оцінкою якості та формуванням ознак доцільно застосовувати локальні перетворення, які зменшують вплив шуму й стабілізують характеристики зображення[12,18].

Базовим інструментом локальної обробки є операція згортки, що полягає в обчисленні нового значення пікселя як зваженої суми значень у його околі. У дискретному вигляді згортку з ядром (маскою)  $K(u, v)$  можна записати як:

$$(I \cdot K)_{x,y} = \sum_{u=-a}^a \sum_{v=-b}^b K_{u,v} I(x - u, y - v). \quad (2.2)$$

Сам механізм згортки (переміщення вікна по зображенню, множення пікселів на ваги ядра та підсумовування) наочно відображено на рисунку 2.2:

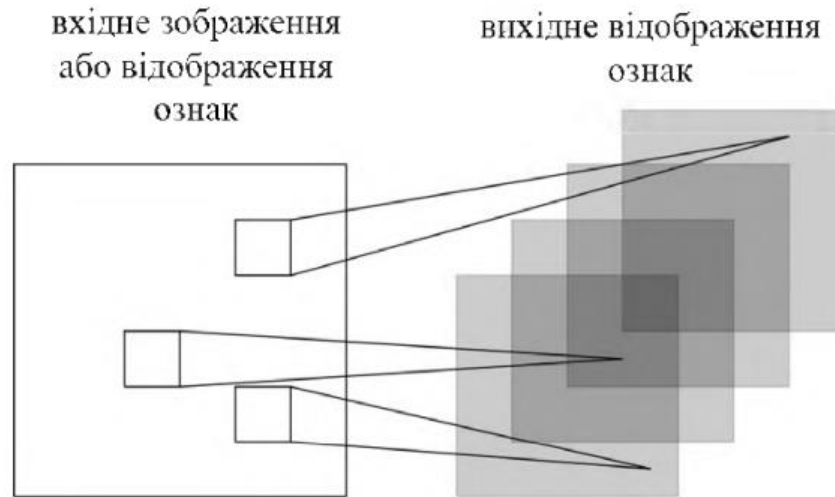


Рисунок 2.2 – Ілюстрація згорткового шару

Рисунок демонструє, що значення в точці  $(x, y)$  визначається не ізольовано, а формується з урахуванням сусідніх пікселів, тобто локальний контекст прямо впливає на кінцеву інтенсивність після фільтрації.

На практиці в задачах попередньої обробки найчастіше використовують фільтри згладжування, мета яких — зменшити вплив високочастотних шумових складових. Найпростішим прикладом є усереднювальний (box) фільтр, ядро якого має рівні ваги:

$$K_{avg}(u, v) = \frac{1}{(2a+1)(2b+1)}, u \in [-a, a], v \in [-b, b]. \quad (2.3)$$

Більш узгодженим для технічних застосувань є гаусівський фільтр, який надає більшу вагу центральним пікселям і плавно зменшує вклад віддалених. Його неперервне подання задається:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right). \quad (2.4)$$

Особливості застосування згладжування залежать від спектрального каналу. У VIS-кадрах інформативність часто пов'язана з дрібними деталями й текстурями, тому надмірне згладжування може погіршувати розрізнюваність сигналу. Водночас при затемненні підсилюється шум, і помірне згладжування дозволяє стабілізувати фон і підвищити надійність оцінок середніх та дисперсійних характеристик. Для IR-кадрів характерна більш "гладка" структура теплового сигналу, і згладжування здебільшого спрямоване на вирівнювання фону та придушення випадкових артефактів сенсора, не руйнуючи при цьому компактну сигнальну складову.

У межах даної роботи згорткові операції та згладжувальні фільтри розглядаються як етап підготовки до оцінки якості та формування ознак. Зокрема, попередня стабілізація кадру може зменшувати вплив одиничних шумових "піків" під час автопошуку області сигналу та робити більш стійкими оцінки параметрів  $i$ , відповідно, показника SNR, що далі використовується для класифікації станів кадру та обґрунтування доцільності комплексування у багатоканальній системі технічного зору.

### 2.2.2 Градієнтні оператори.

Градієнтні оператори застосовуються в системах технічного зору для оцінки локальних змін інтенсивності зображення та аналізу його структурних властивостей. На відміну від згладжувальних перетворень, які зменшують вплив шуму, градієнтні методи підсилюють високочастотні складові сигналу, пов'язані з просторовими переходами між сигналом і фоном. Це робить їх ефективним інструментом для виявлення меж, зон різкої зміни яскравості та оцінки ступеня деталізації зображення. У практичних задачах технічного зору градієнтні оператори використовуються не лише для виділення об'єктів, а й як джерело кількісних ознак, що характеризують якість та інформативність зображення.

У загальному вигляді градієнт зображення визначається як вектор частинних похідних інтенсивності за просторовими координатами:

$$\Delta I(x, y) = \left[ \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \right], |\Delta I| = \sqrt{\left( \frac{\partial I}{\partial x} \right)^2 + \left( \frac{\partial I}{\partial y} \right)^2}. \quad (2.5)$$

У задачах оцінки якості зображень і комплексування в багатоканальних системах технічного зору градієнтні характеристики доцільно використовувати як узагальнені структурні ознаки, а не як самостійний інструмент сегментації. Зокрема, середнє або середньоквадратичне значення модуля градієнта по кадру або по виділеній області сигналу може слугувати індикатором втрати деталізації в умовах затемнення або туману. Градієнтні показники є ефективним та обчислювально простим засобом кількісної оцінки локальних структурних змін і деградації різкості зображення[19].

### 2.2.3 Статистичні та інформаційні ознаки для оцінки якості зображення.

У задачах технічного зору, пов'язаних із комплексуванням VIS/IR-каналів, принципове значення має не лише формування зображення, а й кількісна оцінка його інформативності в поточних умовах спостереження. Реальні сцени характеризуються зміною освітлення, впливом атмосферних факторів та зростанням шумових складових у часі, що призводить до поступової деградації якості сигналу. За таких умов система технічного зору повинна оперувати формалізованими числовими критеріями, які дозволяють відрізнити кадри з достатньою інформативністю від кадрів, у яких видимий канал уже не забезпечує надійного виділення об'єктів і виникає потреба в комплексуванні або зміні внеску спектральних каналів. Для цього використовуються статистичні та інформаційні ознаки, що обчислюються окремо для області сигналу та фону після автоматичного виділення області інтересу.

Нехай  $I(x, y)$  — інтенсивність пікселя з координатами  $(x, y)$ ;  $\Omega_s$  — множина пікселів, що відповідають області сигналу,  $\Omega_b$  — множина пікселів фону. Тоді  $N_s = |\Omega_s|$ ,  $N_b = |\Omega_b|$ .

Базовими характеристиками сцени є середні рівні яскравості в області сигналу та фону, які визначаються як:

$$\mu_s = \frac{1}{N_s} \sum_{(x,y) \in \Omega_s} I(x,y), \mu_b = \frac{1}{N_b} \sum_{(x,y) \in \Omega_b} I(x,y). \quad (2.6)$$

Фізично величина  $\mu_s$  відображає енергетичний рівень корисного сигналу, тоді як  $\mu_b$  характеризує середній рівень фонові складової сцени. У VIS-каналі ці параметри істотно залежать від освітленості та відбивних властивостей об'єктів, тоді як в IR-каналі вони більш стабільні та визначаються тепловим контрастом. Зменшення різниці між  $\mu_s$  та  $\mu_b$  свідчить про ускладнення умов спостереження та зниження візуальної відокремленості сигналу.

Рівень шуму доцільно оцінювати через стандартне відхилення інтенсивності фону:

$$\sigma_b = \sqrt{\frac{1}{N_b - 1} \sum_{(x,y) \in \Omega_b} (I(x,y) - \mu_b)^2}. \quad (2.7)$$

Ця величина характеризує амплітуду випадкових флуктуацій фонові складової, що зумовлені шумом сенсора, електронними завадами або атмосферними впливами. У фізичному сенсі зростання  $\sigma_b$  означає підвищення невизначеності вимірювань, що ускладнює надійне відокремлення сигналу навіть за збереження середнього рівня яскравості.

Для кількісного опису відмінності між сигналом і фоном використовується контраст, визначений як різниця середніх значень:

$$\Delta\mu = \mu_s - \mu_b. \quad (2.8)$$

Контраст безпосередньо пов'язаний із сприйняттям об'єкта на фоні сцени та має чітке фізичне тлумачення як різницю рівнів інтенсивності. У VIS-каналі зменшення контрасту часто пов'язане із затемненням або туманом, тоді як в IR-каналі контраст зберігається довше за рахунок теплової природи сигналу. Саме ця особливість є однією з передумов доцільності комплексування каналів.

Базовим показником якості зображення є відношення сигнал/шум (SNR), яке кількісно описує перевагу корисного сигналу над флуктуаціями фону. У

загальному вигляді ця величина визначається співвідношенням між середнім рівнем сигналу та статистичними характеристиками шуму фону і використовується як інтегральний показник читабельності зображення. Формальний вираз SNR введено у формулі 1.2. Фізично зменшення SNR відповідає погіршенню умов спостереження, зростанню шуму сенсора або ослабленню корисного сигналу, що безпосередньо впливає на можливість виявлення об'єктів.

Для узагальненої оцінки розподілу яскравостей у сцені без прив'язки до локалізації сигналу використовуються дисперсія та стандартне відхилення:

$$\text{Var}(I) = \frac{1}{N-1} \sum_{(x,y)} (I(x,y) - \mu)^2, \quad \text{Std}(I) = \sqrt{\text{Var}(I)}. \quad (2.9)$$

Ці показники відображають глобальний рівень неоднорідності зображення. У фізичному сенсі вони чутливі як до шуму, так і до наявності контрастних структур у сцені, що дозволяє використовувати їх як допоміжні характеристики при оцінці загального стану кадру.

Для кількісної оцінки інформаційної насиченості зображення застосовується ентропія Шеннона, визначена у бітах:

$$H = - \sum_{i=0}^{L-1} p_i \log_2(p_i), \quad (2.10)$$

де  $p_i$  — ймовірність появи  $i$ -го рівня яскравості;

$L$  — кількість рівнів квантування.

Ентропія характеризує ступінь різноманітності яскравісних рівнів у кадрі та дозволяє оцінити зміну структури зображення в умовах деградації. Водночас фізично важливо враховувати, що зростання ентропії може бути пов'язане не лише з появою корисних деталей, а й із підвищенням шумової складової.

У межах задач оцінки якості VIS/IR-кадрів перед комплексуванням наведені статистичні, контрастні та інформаційні характеристики формують узгоджений опис стану сцени. Вони зберігають чітку фізичну інтерпретацію та дозволяють кількісно відобразити різні механізми деградації сигналу, що створює основу для

подальшого використання цих ознак у методах адаптивного прийняття рішень та комплексування зображень.

### 2.3. Фільтр Калмана.

У системах технічного зору, де сенсорні дані містять випадкові шуми, артефакти вимірювання, різні частотні характеристики та неповну інформацію про сцену, ключовою задачею є реконструкція істинного стану середовища на основі множини неточних спостережень. Одним із найефективніших інструментів такого оцінювання є фільтр Калмана (Kalman Filter, KF) — оптимальний рекурсивний алгоритм, що мінімізує середньоквадратичну помилку оцінки для лінійних систем з гаусівськими шумами. Його цінність у СТЗ полягає не лише у здатності згладжувати та прогнозувати сигнали, а й у можливості формально оптимально поєднувати інформацію від різних сенсорів — оптичних, теплових, інерційних, радіолокаційних — у єдиній моделі стану. У класичній роботі Калмана зазначалося, що фільтр “дозволяє поєднати декілька зашумлених вимірювань у найкращу оцінку системи, яка еволюціонує у часі”, що фактично і є центральною задачею багатосенсорного комплексування.

У задачах комплексування VIS та IR сигналів фільтр Калмана відіграє особливу роль, оскільки дозволяє поєднати інформацію двох сенсорів не статично, а у часовій послідовності, формуючи адаптивну оцінку стану сцени. На відміну від звичайних методів комплексування, де комбінування каналів відбувається локально й незалежно для кожного кадру, KF враховує динаміку процесу, накопичуючи корисні дані та пригнічуючи випадкові сплески шуму. Це критично важливо для VIS–IR систем, оскільки видимі зображення зазнають сильних коливань яскравості при зміні освітлення, а теплові матриці можуть демонструвати нестабільність через флуктуації фонові температури. Послідовні методи оцінювання, такі як KF, забезпечують «узгоджену міжкадрову стабільність і значно кращу експлуатаційну надійність, ніж фрейм-орієнтовані підходи», що робить їх природним механізмом для багатоспектрального комплексування[20].

### 2.3.1 Математична модель фільтра Калмана.

У сучасних системах технічного зору об'єднання інформації з видимого та інфрачервоного діапазонів потребує не лише просторової обробки, а й здатності коректно працювати з часовою динамікою сцени. Дані сенсорів надходять у режимі відеопотоку, а отже їхня якість постійно змінюється під впливом освітлення, температурних флуктуацій, шумів детектора та особливостей оптичної системи. Через це моделі статичного комплексування, які обробляють кожен кадр незалежно, не можуть забезпечити стабільність і надійність оцінки. Фільтр Калмана став фундаментальним інструментом у таких умовах, оскільки він дозволяє оптимально інтегрувати інформацію різних сенсорів, використовуючи як нові вимірювання, так і прогноз попереднього стану. Здатність фільтра Калмана адаптивно змінювати вагу каналів робить його особливо ефективним для стабілізації інтегрованих відеопотоків[21].

Модель ґрунтується на припущенні, що стан сцени у момент часу  $k$  можна описати вектором:

$$x_k \in \mathbb{R}^n,$$

який акумулює інтегровану інформацію — від високочастотних структур до теплових профілів. Видимий канал дає детальні контури, інфрачервоний — температурні характеристики, а KF формує узгоджену оцінку, яка враховує обидві складові. Часова еволюція стану моделюється лінійним рівнянням:

$$x_k = F_{k-1}x_{k-1} + w_{k-1}, \quad (2.11)$$

де  $F_{k-1}$  — матриця переходу;

$w_{k-1}$  шум процесу, що моделює непередбачувані зміни середовища, який описується як:

$$w_{k-1} \sim N(0, Q_{k-1}).$$

У СТЗ часто застосовують спрощене припущення  $F_{k-1}=I$ , що відповідає ситуації, коли між сусідніми кадрами сцена не зазнає суттєвих змін. Це дозволяє трактувати оцінку стану як згладжене представлення сцени, яке поступово уточнюється при надходженні нових даних[22].

Вимірювання надходять із двох спектральних каналів — VIS та IR. Вони утворюють загальний вимірювальний вектор який розглядається як лінійна проекція істинного стану:

$$z_k = H_k x_k + v_k, \quad (2.12)$$

де шум вимірювання  $v_k \sim N(0, R_k)$ .

Оскільки природний шум VIS та IR сенсорів некорельований і має різну фізичну природу (фотонні флуктуації, теплові зсуви матриці, електронний шум), матриця covariance  $R_k$  набуває блок-діагонального вигляду:

$$R_k = \begin{bmatrix} R_k^{\text{VIS}} & 0 \\ 0 & R_k^{\text{IR}} \end{bmatrix}. \quad (2.13)$$

Ця структура має принципове значення: вона формує механізм вагового впливу каналів на оцінку. Канал із більшим рівнем шуму автоматично має менший вплив на формування інтегрованого стану. Саме тому фільтр Калмана природним чином реалізує SNR-орієнтоване комплексування, у якому внесок VIS зменшується при падінні освітленості, а внесок IR підсилюється за умов погіршення видимого каналу чи теплових флуктуацій.

Формування оцінки стану відбувається через узгодження інформації прогнозу з новими вимірюваннями. На цьому етапі ключову роль відіграє калманівський коефіцієнт:

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}, \quad (2.14)$$

який визначає міру впливу вимірювань на оновлення стану. Чим надійніші вимірювання певного каналу (менша дисперсія шуму), тим більшу вагу вони

отримують у структурі  $K_k$ . Це робить фільтр Калмана не лише інструментом згладжування, а і повноцінною моделлю оптимального спектрального злиття.

Власне оновлення стану(корекція) виконується як лінійна корекція прогнозу:

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H_k \hat{x}_k^-), \quad (2.15)$$

що інтерпретується як адаптивне додавання до прогнозованого стану інформації нових вимірювань, зважених відповідно до їхньої точності. У контексті VIS–IR інтеграції це означає, що структурні особливості VIS та теплові характеристики IR не просто підсумовуються, а включаються в оцінку з «розумною» інтенсивністю залежно від рівня шумів. Коварація оновлених похибок визначається виразом:

$$P_k = (I - K_k H_k) P_0^-, \quad (2.16)$$

який показує, наскільки зменшилася невизначеність оцінки після об'єднання даних. Саме цей параметр характеризує статистичну надійність інтегрованої карти сцени.

Окремо варто відзначити здатність KF забезпечувати часову стабілізацію інтегрованого зображення. Під час роботи СТЗ сусідні кадри майже ніколи не є незалежними — і VIS-, і IR-сенсори демонструють міжкадрові коливання інтенсивності, що можуть виглядати як шум, хоча не несуть фізичного змісту. Калманівська модель, використовуючи попередню оцінку стану та її дисперсію, пригнічує такі короточасні аномалії, пропускаючи лише інформацію, яка статистично узгоджується з динамікою сцени. Саме фільтри стану дозволяють уникнути «ефекту мерехтіння» при багатоспектральному комплексуванні — різких перепадів яскравості між кадрами, які істотно ускладнюють подальшу сегментацію або трекінг[22].

Фільтр Калмана фактично створює механізм динамічного VIS–IR злиття: він не просто об'єднує канали, а адаптує їхній внесок на основі статистичних властивостей шуму та поведінки сцени у часі. Адаптивність KF робить його особливо корисним у відеоспостереженні — модель автоматично переходить до IR домінування при зниженні освітленості та до VIS домінування при появі

структурних деталей, які важливі для ідентифікації. Це дозволяє отримати інтегроване зображення, яке стабільне між кадрами, не перенасичене шумами, містить релевантну структурну й теплову інформацію й може використовуватися як вхід для подальших алгоритмів детекції чи класифікації[23].

### 2.3.2 Використання КФ для багатосенсорного комплексування.

Фільтр Калмана є одним із найбільш ефективних підходів до багатосенсорного комплексування, оскільки він забезпечує статистично обґрунтований спосіб об'єднання вимірювань, отриманих з різних джерел. Для систем технічного зору, де VIS та IR канали формують інформацію з принципово різними характеристиками шуму, контрастності та динаміки, КФ виконує роль оптимального інтегратора, який автоматично визначає, наскільки кожному з каналів можна «довіряти» на поточному кроці часу. Завдяки рекурсивній природі алгоритму, комплексування відбувається послідовно у часі, а отже — модель зберігає історію попередніх оцінок і пригнічує міжкадровий шум, що критично важливо для стабільної роботи СТЗ у реальних умовах спостереження.

У випадку багатосенсорного аналізу фільтр Калмана приймає кілька потоків вимірювальних даних, які можуть суттєво відрізнитися за точністю, частотою оновлення та спектральними властивостями. На практиці VIS-канал забезпечує багатшу структуру та деталізацію, але чутливий до низької освітленості й атмосферних перешкод; IR-канал, навпаки, стабільний у темряві, але містить менше високочастотних ознак. Саме тому їх комплексування через КФ дозволяє компенсувати недоліки кожного з сенсорів: модель підсилює той канал, чий рівень шуму є нижчим, і зменшує вплив того, в якому спостерігаються деградації чи флуктуації. Фільтр Калмана забезпечує узгоджене та статистично оптимальне комбінування вимірювань різної природи, що особливо цінно у VIS–IR задачах, де дисбаланс якості сигналів може змінюватися у часі через умови освітлення, температуру та атмосферні явища[4].

Важливою перевагою КФ є його здатність не лише поєднувати вимірювання, але й прогнозувати стан сцени на основі попередньої динаміки. У системах технічного зору це означає стабілізацію інтегрованого зображення, навіть якщо

один із каналів тимчасово дає нестабільні або частково втрачені дані. Наприклад, у VIS-каналі це може бути спричинене раптовою зміною освітлення або падінням експозиції, тоді як IR-канал страждає від термічних шумів і змін емісійних властивостей фону. KF, балансуючи між прогнозом і новими вимірюваннями, зберігає узгодженість сцени та запобігає «стрибкам» фінального результату.

### 2.3.3 Оцінка стану при шумових моделях.

У задачах комплексування VIS та IR каналів ключовою проблемою є те, що рівень шуму в кожному сенсорному модулі змінюється не лише між різними режимами роботи, а й у часі під дією зовнішніх факторів. Це означає, що фіксовані значення шумових параметрів моделі спостереження не здатні описати реальну якість вимірювань, а отже стандартна форма фільтра Калмана стає недостатньо гнучкою. Саме тому в сучасних дослідженнях мультисенсорного комплексування було запропоновано підхід, у якому матриця шумів вимірювань адаптивно модифікується відповідно до оціненого рівня SNR окремо для VIS- та IR-каналів. Такий підхід забезпечує можливість описувати реальні, динамічно змінні умови спостереження, коли оптичний канал деградує через недостатнє освітлення, а інфрачервоний — через температурне насичення або флуктуації мікроболометричної матриці.

У загальному випадку рівень шуму VIS та IR каналів може бути описано через їхній миттєвий SNR. Нехай для моменту часу  $k$  задані оцінки:

$$SNR_k^{(VIS)} = \frac{\mu_{VIS}^2(k)}{\sigma_{VIS}^2(k)}, \quad SNR_k^{(IR)} = \frac{\mu_{IR}^2(k)}{\sigma_{IR}^2(k)}, \quad (2.17)$$

де  $\mu$  та  $\sigma^2$  — оцінені середні значення та дисперсії сигналу на локальних просторових вікнах.

У адаптивного комплексування інформаційна значущість каналу зростає пропорційно його SNR, тоді як еквівалентний шум вимірювання зменшується обернено пропорційно[23]:

$$R_k^{(VIS)} \propto \frac{1}{SNR_k^{(VIS)}}, \quad R_k^{(IR)} \propto \frac{1}{SNR_k^{(IR)}}. \quad (2.18)$$

У цьому контексті адаптивна матриця шумів вимірювання набуває вигляду 2.20. Завдяки такій формі матриця  $R_k$  стає не статичним параметром моделі, а динамічним коефіцієнтом, що відображає фактичну якість даних. Саме це дає фільтру Калмана здатність «адаптуватися» до мінливих умов спостереження — мережа приймає рішення про відносну вагу сенсорів не на основі наперед визначених припущень, а на основі поточного стану сцени.

Адаптивність проявляється безпосередньо у калманівському коефіцієнті 2.14 який задає, наскільки сильно нові вимірювання вплинуть на оновлений стан. Якщо шум певного сенсора зростає, відповідний елемент  $R_k$  збільшується, а внесок цього вимірювання в оцінку зменшується. Якщо ж якість каналу покращується, то елемент  $R_k$  зменшується і нові дані стають більш інформативними. Таким чином, KF автоматично збільшує вагомість надійнішого каналу й приглушує канал зі зниженим SNR. Співвідношення між коваріацією помилки прогнозу та шумовою коваріацією визначає оптимальну структуру ваг, а отже — і якість багатосенсорної інтеграції[20,23]. Внутрішню ефективність такої адаптації можна проілюструвати через очікувану квадратичну помилку:

$$E[\varepsilon_k^T \varepsilon_k] = \text{trace}(P_k), \quad (2.19)$$

яка описує середнє значення накопиченої невизначеності системи. Оскільки матриця  $P_k$  є результатом аналітичної взаємодії матриць  $R_k$ ,  $H_k$  та прогнозу стану, то зниження шуму одного з каналів (тобто підвищення його SNR) прямо впливає на зменшення помилки оцінювання. У експериментальних роботах зазначено, що адаптивні шумові моделі забезпечують помітне зменшення очікуваної помилки порівняно зі схемами з фіксованими параметрами шуму, оскільки динамічна зміна  $R_k$  дозволяє фільтру коректніше «зважувати» інформацію в умовах невизначеності.

## 2.4 Класичні методи комплексування зображень

### 2.4.1. PCA.

Метод головних компонент (Principal Component Analysis, PCA) є одним із найбільш поширених класичних підходів до комплексування зображень, насамперед завдяки своїй математичній прозорості та здатності стисло представляти багатовимірні дані через їх найінформативніші компоненти. На відміну від вейвлет-перетворення, яке виконує локальне частотне розкладання, PCA працює у глобальному просторі спостережень та описує зображення через напрямки максимальної дисперсії. Це робить метод особливо привабливим для задач інтеграції VIS–IR зображень, де кожен канал містить різнорідну статистичну інформацію та має свої унікальні характеристики: високу структурну варіативність у видимому діапазоні та теплову стабільність у інфрачервоному.

У класичному формулюванні PCA вихідні зображення представляються як випадкові величини, об'єднані у векторну форму. Наприклад, для двоканального комплексування можна розглядати вектор:

$$X = \begin{bmatrix} I_{VIS}(x,y) \\ I_{IR}(x,y) \end{bmatrix}, \quad (2.20)$$

де кожна компонента містить значення яскравості у точці  $(x,y)$ . Важливим кроком є оцінка коваріаційної матриці, яка відображає спільну статистику каналів:

$$C = \begin{bmatrix} \sigma_W & \sigma_{VIR} \\ \sigma_{VIR} & \sigma_{II} \end{bmatrix}, \quad (2.21)$$

де  $\sigma_W$  — дисперсія VIS;

$\sigma_{II}$  — дисперсія IR;

$\sigma_{VIR}$  міжканальна кореляція.

Саме ця матриця визначає, якою мірою один канал містить інформацію, спільну або унікальну відносно іншого. PCA може розглядатися як спосіб виявити ті напрямки простору, де найбільше зосереджена енергія сцени, незалежно від природи каналів[24].

Власні вектори матриці  $C$  визначають напрямки головних компонент, а власні значення — відповідну кількість інформації. Перша головна компонента ( $PC_1$ ) є найбільш інформативним лінійним поєднанням каналів, оскільки ефективно зберігає глобальний контраст сцени та основні теплові елементи.  $PC_1$  має вигляд: :

$$PC_1 = w_{VIS} \cdot I_{VIS} + w_{IR} \cdot I_{IR}. \quad (2.22)$$

Такий підхід забезпечує достатньо високу якість інтеграції для задач моніторингу та спостереження.

Однак у VIS–IR комплексуванні PCA має низку важливих обмежень, які впливають із самої природи цього методу. По-перше, PCA є глобальним лінійним перетворенням: воно не враховує локальні зміни освітлення, текстур або теплових контрастів. Це означає, що важливі дрібні структури VIS або слабкі IR-сигнали можуть бути втрачені, якщо їх внесок у дисперсію невеликий. По-друге, PCA не враховує неоднорідність шумів у каналах: VIS у темряві має нижчий SNR, тоді як IR є більш стабільним. Оскільки PCA оптимізує саме дисперсію, а не відношення сигнал/шум, воно може некоректно підсилювати шумові компоненти. Метод PCA схильний обирати компоненти з найбільшою енергією, навіть якщо значна частина цієї енергії є шумовою[25].

Ще однією проблемою є чутливість PCA до кореляції між каналами. Якщо VIS і IR мало корелюють — що часто трапляється в нічних сценах — ваги можуть виявитися некоректними, оскільки напрямки головних компонент слабко відображають реальну семантику зображення. Тому класичні реалізації PCA у технічних системах спостереження нерідко доповнюють нормалізацією каналів, попереднім шумозаглушенням або локальним контрастним підсиленням.

Попри ці обмеження PCA має важливу перевагу: метод є детермінованим, малоресурсним і надзвичайно швидким, що робить його придатним для систем реального часу та малопотужних апаратних платформ. Обчислення коваріації та власних векторів може виконуватися за мілісекунди навіть для великих зображень,

що робить PCA корисним базовим інструментом або початковим етапом перед більш складними алгоритмами.

#### 2.4.2 Вейвлет-комплексування.

Методи вейвлет-перетворення посідають важливе місце серед класичних підходів до комплексування зображень, оскільки вони забезпечують водночас локальність, багатомасштабність та збереження структурних характеристик сцени. На відміну від PCA, яке виконує глобальну лінійну трансформацію у всьому просторі спостережень, вейвлет-аналіз дозволяє розкласти зображення на компоненти різного масштабу та орієнтації, що робить його особливо корисним у задачах інтеграції видимого та інфрачервоного каналів. Для технічних систем спостереження це має фундаментальне значення: VIS-зображення містять багату високочастотну структуру, тоді як IR-канал є більш гладким і характеризується вираженою низькочастотною інформативністю. Вейвлет-простір дає можливість обробляти їх окремо та об'єднувати у спосіб, що мінімізує втрати ключових ознак.

Дискретне вейвлет-перетворення (DWT) виконує декомпозицію зображення на апроксимувальну компоненту низької частоти  $A_1$  та три набори деталей — горизонтальні, вертикальні та діагональні:  $H_1$ ,  $V_1$ ,  $D_1$ . загальному вигляді перетворення можна описати як:

$$W\{I(x, y)\} = \{A_1, H_1, V_1, D_1\}, \quad (2.23)$$

де кожна підсмуга містить певний тип просторової інформації. Подальше комплексування полягає у порівнянні відповідних підсмуг VIS та IR та виборі найбільш інформативних коефіцієнтів відповідно до певного правила злиття. У класичній формі використовують максимум-орієнтовані критерії:

$$C_H(x, y) = \max(|H_{VIS}(x, y)|), (|H_{IR}(x, y)|), \quad (2.24)$$

а для низькочастотної компоненти — вагові або енергетичні схеми:

$$C_A(x, y) = w_{VIS}A_{VIS}(x, y) + w_{IR}A_{IR}(x, y), \quad (2.25)$$

де ваги залежать від локальної енергії або варіаційності фрагмента. Вибір саме локальних частотних характеристик дозволяє уникнути глобальних спотворень структури сцени та зберігати найсуттєвіші контурні елементи VIS-зображення[26].

Для VIS–IR комплексування цей підхід має особливе значення. Вейвлет-деталі високих частот природно зберігають контури, кути та градієнтні структури, що є ключовими для ідентифікації об'єктів у технічних системах спостереження. Водночас IR-канал містить інформацію про теплові контрасти, яка локалізована переважно у низькочастотній апроксимаційній підсмузі. Тому вейвлет-модель дає змогу здійснити семантично коректне розділення інформаційних компонент і виконати їхнє злиття таким чином, щоб яскраво виражені теплові структури IR не нівелювали тонку текстуру VIS.

Попри значні переваги, метод має кілька суттєвих обмежень у контексті VIS–IR комплексування. Він є доволі чутливим до вибору правил злиття та масштабу декомпозиції: невдалий підбір може призвести до втрати дрібних деталей або появи артефактів на межах підсмуг. Оскільки вейвлет-функції мають фіксовану форму та не адаптуються до статистики конкретної сцени, метод може некоректно обробляти слабконтрастні IR-об'єкти або, навпаки, надмірно підсилювати високочастотний шум VIS-каналу. У системах технічного зору це проявляється в тому, що дрібні теплові сигнатури легко губляться, а контури об'єктів можуть спотворюватися залежно від вибору глибини розкладу.

Якість комплексування значною мірою визначається вибором базису вейвлет-перетворення. Простий базис Haar забезпечує чітке виділення переходів і високу швидкодію, проте часто спричиняє блокові артефакти в IR-даних, які зазвичай мають плавні градієнти. Базиси Daubechies (DB2, DB4) або Symlets забезпечують кращу гладкість та дозволяють точніше представляти і високочастотні елементи VIS, і низькочастотні структури IR. Вибір базису є критичним параметром, що визначає баланс між збереженням контурів і передаванням теплових характеристик[27].

Завдяки можливості багаторівневої декомпозиції DWT аналізує зображення одночасно у кількох масштабах:

$$I \xrightarrow{DWT} \{A_L, H_l, V_l, D_l\}, l = 1, \dots, L, \quad (2.26)$$

що забезпечує гнучке виділення дрібних і великих структур. Це особливо важливо для СТЗ, де об'єкти можуть бути як малими (силуети людей, деталі техніки), так і великими (теплові плями, фон). Мультискейлова структура дозволяє зберігати критично важливі геометричні та теплові елементи сцени.

#### 2.4.3 Лапласіан .

Методи, основані на лапласіанних операторів, посідають помітне місце серед класичних алгоритмів об'єднання зображень завдяки здатності ефективно підкреслювати високочастотні компоненти сигналу — контури, різкі переходи яскравості та дрібні деталі сцени. Лапласіан є другим просторовим похідним оператором, чутливим до локальних змін інтенсивності, і тому природно підходить для задач VIS–IR комплексування, де структури видимого діапазону (градієнти, краї, текстури) відіграють ключову роль у формуванні семантичного змісту, а інфрачервоні вимірювання переважно містять низькочастотні теплові сигнатури. Завдяки цьому лапласіан дозволяє окремо обробляти детальну складову зображення та інтегрувати її у фінальну комплексовану карту таким чином, щоб зберегти структурну чіткість сцени за відсутності достатньої освітленості.

У загальному випадку лапласіан визначається як сума других часткових похідних:

$$\nabla^2 I(x, y) = \frac{d^2 I}{dx^2} + \frac{d^2 I}{dy^2}. \quad (2.27)$$

У цифровій формі оператор апроксимується через дискретні ядра, найбільш поширеним з яких є:

$$L = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}, L' = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}. \quad (2.28)$$

Застосування лапласіана до VIS-зображення дозволяє витягнути компоненти, що відповідають різким локальним переходам, тоді як аналогічна операція для IR-каналу дає обмежену кількість структур, оскільки теплові зображення є більш гладкими. Пропонується використовувати лапласіан переважно для видимих зображень, а інфрачервоні сигнали залучати як опорну низькочастотну складову. Таким чином, процедура комплексування набуває форми двокомпонентного поєднання: IR забезпечує глобальну інтенсивність і теплову структуру сцени, а VIS — локальні деталі[28].

Одним із найбільш відомих підходів є використання лапласіанної піраміди, яка забезпечує багаторівневу декомпозицію зображення у спосіб, подібний до вейвлет-методів, але із сильнішим акцентом на високочастотні компоненти. Процес включає побудову послідовності згладжених зображень (гаусової піраміди) та відповідних різницевих карт:

$$L_l = \text{upsampe}(G_l + 1), \quad l = 0, \dots, L - 1, \quad (2.29)$$

де  $G_l$  — усереднені (low-pass) версії початкового зображення.

Лапласіанна піраміда містить на кожному рівні деталі того масштабу, який ще не було усунено розмиванням. Для VIS–IR комплексування це означає, що можна окремо порівняти детальні компоненти  $L_l^{\text{VIS}}$  та  $L_l^{\text{IR}}$  і вибрати найбільш інформативну згідно з певним правилом. Найчастіше застосовують максимум-орієнтовані або енергетичні критерії:

$$C_l(x, y) = \arg \max_{m \in \{\text{VIS}, \text{IR}\}} |L_l^m(x, y)|. \quad (2.30)$$

Таким чином, для кожного рівня декомпозиції зберігаються ті деталі, які виражені сильніше. Це дозволяє досягнути високої чіткості у фінальному зображенні, що є критично важливим у СТЗ під час виявлення об'єктів та локалізації структурних контурів.

Суттєвою перевагою лапласіанних методів є їхня чутливість до контурів незалежно від абсолютної яскравості, що є особливо важливим у ситуаціях, коли VIS-канал деградує через низьке освітлення, а IR — через температурні флуктуації. На відміну від PCA або DWT, лапласіан фокусується саме на другій похідній, що дозволяє уникати втрати дрібних структур під час згладження. Крім того, результати комплексування легко інтерпретуються й можуть бути використані як вхід для подальших алгоритмів сегментації, аналізу контурів або оцінювання руху.

Проте існують і певні обмеження. Лапласіан значно підсилює шум, оскільки диференціювання другого порядку чутливе до будь-яких випадкових флуктуацій. Це особливо актуально в системах технічного бачення, де рівень шуму VIS-каналу в умовах низької освітленості може бути значним. Саме тому лапласіанні методи зазвичай комбінують із попереднім фільтруванням або розмиванням, зокрема через гаусові фільтри або двосторонню фільтрацію. Окрім того, лапласіанна піраміда має більшу обчислювальну складність порівняно з одношаровим лапласіаном, що може бути критичним для реального часу, хоча сучасні оптимізації для GPU значно скорочують затрати часу.

Для технічних систем спостереження лапласіанний підхід є цінним як окремий модуль або як частина гібридних алгоритмів комплексування. Завдяки високій здатності зберігати й підсилювати структурні елементи сцени, він широко застосовується у задачах військового та цивільного спостереження, де якість контурів визначає точність подальшого розпізнавання. У поєднанні з IR-сигналом лапласіанна модель формує збалансовану комплексовану карту, в якій теплові структури поєднуються з деталями видимого спектра, забезпечуючи інформативність навіть у складних умовах зйомки.

## 2.5 Порівняння методів комплексування та виявлення недоліків

### 2.5.1 Недоліки класичних та ML – орієнтованих методів.

Класичні методи комплексування зображень, попри їхню тривалу історію використання та математичну прозорість, демонструють обмеженість у сучасних

умовах функціонування систем технічного зору. Більшість таких методів ґрунтується на фіксованих лінійних або локальних перетвореннях та наперед заданих правилах злиття каналів, що істотно обмежує їхню адаптивність до змінних умов спостереження. Наприклад, метод головних компонент (РСА) виконує проєкцію зображень у глобальний простір ознак, унаслідок чого відбувається усереднення локальних структур сцени та можлива втрата дрібних, але інформативних деталей. Вейвлетні та лапласіанні підходи частково долають цю проблему завдяки багатомасштабному представленню або підсиленню контурів, проте вони залишаються жорстко детермінованими та не враховують змінну інформативність каналів VIS та IR у конкретній сцені.

Суттєвим недоліком класичних алгоритмів є їхня чутливість до деградації сигналу, зокрема в умовах низького відношення сигнал/шум. За таких умов шумові компоненти можуть домінувати над корисною інформацією, а лінійні фільтри та фіксовані правила злиття не мають механізмів для автоматичного зменшення внеску деградованого каналу. У результаті комплексоване зображення може містити артефакти або втрачати інформативні ознаки, що знижує ефективність подальших етапів обробки та аналізу.

Методи машинного навчання на основі ознак значною мірою усувають зазначені обмеження за рахунок переходу від жорстко заданих правил до адаптивного прийняття рішень на основі статистичних характеристик зображення. На відміну від класичних підходів, у таких методах комплексування або вибір режиму обробки визначається не фіксованою формулою, а моделлю, навченою на множині сцен із різними рівнями деградації сигналу. Вхідними даними для моделі є ознаки з чітким фізичним змістом — контраст, рівень шуму, ентропія, відношення сигнал/шум, градієнтні характеристики, — що дозволяє формалізувати зв'язок між параметрами сцени та якістю зображення.

Перевагою ознакових методів машинного навчання є поєднання адаптивності з інтерпретованістю. Моделі, такі як дерева рішень або їх ансамблі, дозволяють не лише враховувати кілька параметрів одночасно, але й відстежувати логіку прийняття рішення у вигляді послідовності умов. Це особливо важливо для

систем технічного зору, де необхідно забезпечити передбачуваність поведінки алгоритму та можливість аналізу його роботи в перехідних режимах. Водночас ефективність таких методів залежить від якості сформованих ознак і повноти навчальної вибірки, що потребує ретельного проектування етапу оцінки якості зображення

Таблиця 2.1 – Порівняння методів

Критерій	Класичні методи (PCA, DWT, Laplacian)	Методи ML на основі ознак
Адаптивність до умов сцени	Низька; фіксовані правила, відсутність навчання	Висока; адаптація на основі навчальних даних
Стійкість до шумів	Слабка; високочастотні методи особливо чутливі	Висока; можливість урахування рівня шуму через ознаки
Урахування спектральних особливостей VIS–IR	Обмежене; немає механізмів для диференційованої ваги каналів	Рішення залежать від інформативності каналів
Обчислювальна вартість	Низька; можливість роботи на вбудованих системах	Помірна; потребує етапу навчання, проста в реалізації
Інтерпретованість	Висока, чітко визначені операції	Висока; рішення формуються у вигляді логічних правил
Масштабованість на складні сцени	Обмежена; локальність або глобальність фіксована	Висока; можливість розширення простору ознак
Стабільність при деградації каналів	Низька; метод не змінює вагу каналів автоматично	Висока; ваги та рішення залежать від оцінки якості

Порівняння методів показує, що класичні підходи до комплексування відзначаються простотою та передбачуваністю, однак є малочутливими до змін якості сигналу та не враховують статистичну різномірність VIS і IR каналів. Методи машинного навчання на основі ознак дозволяють формалізувати прийняття рішень за поточною інформативністю зображення та адаптуватися до умов деградації, зберігаючи інтерпретованість результатів. Це створює основу для побудови

адаптивних схем комплексування, у яких внесок каналів визначається їхньою якістю, а не фіксованими правилами.

### 2.5.2 Неможливість універсального застосування .

Незважаючи на різноманіття існуючих методів комплексування зображень, жоден з них не може розглядатися як універсальний для всіх умов спостереження. Це пов'язано з тим, що параметри реальних сцен у системах технічного зору змінюються у широких межах і не підкоряються фіксованим статистичним моделям. Видимий канал може втрачати інформативність внаслідок зниження освітленості, атмосферних явищ або зростання шуму, тоді як інфрачервоний канал за певних умов характеризується низьким тепловим контрастом або нестабільним фоном. У таких ситуаціях алгоритм, налаштований на один сценарій, виявляється неефективним в іншому.

Для класичних методів комплексування основним обмеженням є жорсткість правил обробки. Алгоритми, що базуються на фіксованих лінійних перетвореннях або наперед визначених схемах злиття, не враховують зміну інформативності каналів у часі. В умовах низького відношення сигнал/шум такі методи не здатні коректно зменшувати внесок деградованого каналу, що призводить або до підсилення шуму, або до втрати корисних структур зображення. Це особливо критично для багатоканальних систем, де інформативність VIS та IR каналів може суттєво відрізнятись навіть у межах однієї сцени.

Методи машинного навчання на основі ознак забезпечують вищий рівень адаптивності, проте також не є повністю універсальними. Їх ефективність визначається вибором ознак та характером навчальної вибірки. Якщо статистичні властивості нових сцен суттєво відрізняються від тих, що були представлені під час навчання, модель може втрачати здатність коректно розрізняти стани якості або приймати нестабільні рішення. Крім того, використання лише одного узагальненого показника якості не дозволяє повністю охопити різноманіття деградацій, характерних для реальних умов спостереження.

Ключовою проблемою є відсутність універсального механізму, який би дозволяв системі в реальному часі оцінювати надійність кожного каналу та

адаптивно змінювати режим комплексування. Без урахування показників якості сигналу, зокрема відношення сигнал/шум і супутніх статистичних характеристик, поведінка алгоритму залишається статичною. Це обмежує можливість застосування єдиного підходу для широкого спектра сцен і умов.

## 3 РОЗРОБКА АЛГОРИТМУ КОМПЛЕКСУВАННЯ НА ОСНОВІ SNR – ОРІЄНТОВАНОЇ ОЦІНКИ ЯКОСТІ ЗОБРАЖЕННЯ

### 3.1 Постановка задачі та логіка побудови алгоритму VIS-IR комплексування

У системах технічного зору, що працюють у реальних умовах спостереження, якість і стабільність візуальної інформації визначають ефективність подальших етапів аналізу сцени — виявлення, класифікації та відстеження об'єктів. Видимий діапазон забезпечує високу деталізацію, проте значною мірою залежить від освітленості, погодних умов та оптичних перешкод. Інфрачервоний канал, навпаки, формує теплову структуру сцени і майже не реагує на зміну рівня освітлення, але поступається VIS у відтворенні контурів та дрібних текстурних ознак. У багатьох випадках жоден окремих канал не здатний забезпечити повноту й надійність інформації, необхідної для функціонування СТЗ у нічних, перехідних або завадових умовах.

На практиці це призводить до того, що системи, які покладаються лише на VIS або лише на IR, демонструють нерівномірну якість роботи. VIS-кадр може містити багату структурну інформативність, але стає непридатним при падінні освітленості. IR-канал зберігає теплову складову, але не дозволяє виконувати точну локалізацію та розпізнавання малих або слабоконтурних об'єктів. Подібна різниця у властивостях каналів формує природну потребу в комплексуванні, здатному адаптивно використовувати їх сильні та слабкі сторони залежно від умов спостереження.

У більшості традиційних методів об'єднання зображень внесок кожної модальності є фіксованим або визначається евристичними правилами. Такий підхід виявляється недостатнім у динамічних середовищах, де характеристики шуму та контрастності змінюються між кадрами. Занижений рівень освітлення у VIS або зростання шуму датчика призводять до того, що інтегроване зображення може втратити частину ключових ознак, навіть якщо IR-канал залишається інформативним. Постає задача створення методу, який здатний не просто

поєднувати спектральні канали, а оцінювати їхню поточну якість і коригувати ступінь впливу кожного з них.

Ця проблема є особливо актуальною для систем технічного зору, що працюють у змінних або обмежених умовах. Виявлення об'єктів на тлі шуму, робота в сутінках чи при слабкому тепловому контрасті вимагають не лише механічного злиття VIS та IR, а адаптивного аналізу інформативності кожного каналу. Значення SNR стає ключовим індикатором стану даних: воно характеризує, наскільки надійно канал відтворює структуру сцени та чи може бути використаний у поточному циклі обробки. На цій основі виникає потреба у SNR-орієнтованому алгоритмі комплексування, який здатний регулювати вклад каналів у залежності від їх фактичної якості.

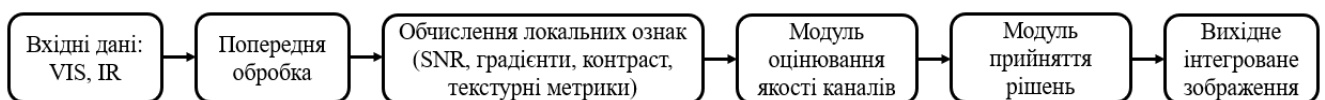


Рисунок 3.1 – Структурна схема SNR-орієнтованого алгоритму комплексування зображень VIS–IR

Другим ключовим етапом роботи алгоритму є попередня підготовка вхідних зображень, яка забезпечує коректність подальшого обчислення SNR та формування локальних ознак. Видимий та інфрачервоний канали мають принципово різну природу: VIS-зображення формуються за рахунок відбитого світла і суттєво залежать від рівня освітлення, тоді як IR-канал відображає теплове випромінювання та зберігає інформативність навіть у повній темряві. Тому перед тим, як переходити до аналізу ознак, необхідно привести обидва канали до єдиного узгодженого формату — як у просторовому, так і в яскравісному сенсі.

На першому кроці виконується геометричне вирівнювання каналів (alignment), що забезпечує точне накладання сцен VIS та IR. Навіть невелике зміщення або невідповідність масштабів призводить до неправильних оцінок градієнтів, некоректного розрахунку локальних метрик і, відповідно, до хибного визначення якості каналів. У практичних системах технічного зору такі

невідповідності виникають через різницю фокусних відстаней, розташування сенсорів або оптичні спотворення. Тому цей крок є критично важливим для будь-якої подальшої моделі — як класичної, так і побудованої на SNR–ML підході.

Наступним етапом є нормалізація яскравості, яка дозволяє усунути неконтрольовані відмінності масштабу інтенсивності між VIS та IR каналами. Без цього порівняння метрик контрасту, варіаційності, градієнтів і тим більше SNR було б некоректним, оскільки абсолютні значення інтенсивності у цих каналах не співставні за фізичною природою. Нормалізація забезпечує можливість оперувати відносними інформативними показниками та дозволяє обчислювати SNR у єдиному узгодженому вигляді, що безпосередньо впливає на рішення моделі у наступних етапах.

Важливою частиною попередньої обробки є також шумозаглушення. У VIS-каналі за низької освітленості шум швидко наростає, часто набуваючи домінуючого характеру на периферії сцени. У IR-каналі, навпаки, основним джерелом завад є теплові флуктуації або шум сенсора. В обох випадках застосування легкого фільтрування (Gaussian smoothing, bilateral filtering) дозволяє стабілізувати локальні оцінки, не знищуючи ключових структур. Це особливо важливо для розрахунку градієнтів та текстурних метрик, оскільки різкі сплески шуму можуть бути помилково інтерпретовані як локальні високочастотні ознаки.

Попередня обробка формує основу для подальшої роботи алгоритму: вирівнює канали, стабілізує локальні характеристики та переводить VIS та IR у спільний простір, де їх можна коректно порівнювати та оцінювати. Без цього етапу побудова достовірної SNR-орієнтованої оцінки була б неможливою, а результат інтеграції — нестабільним і непридатним для подальшого аналізу.

Після виконання попередньої обробки та вирівнювання каналів постає питання, які саме характеристики можуть достовірно відображати їхню фактичну інформативність у кожному окремому фрагменті сцени. У технічних системах спостереження, що працюють у динамічних умовах, зміна освітленості, тепловий фон та сенсорні шуми можуть істотно впливати на якість VIS та IR-даних, причому з різною інтенсивністю у кожному каналі. Саме тому алгоритм потребує

універсального показника, що дозволив би кількісно оцінювати надійність інформації в певній точці зображення. Найбільш придатним для цього є коефіцієнт «сигнал/шум» — SNR, який характеризує співвідношення корисного сигналу до завад і тим самим дає змогу оцінити реальний стан каналу в поточний момент.

Використання SNR у контексті VIS–IR комплексування є не лише технічно доцільним, а й фізично обґрунтованим. У видимому діапазоні падіння SNR означає погіршення яскравісної структури, зникнення контурів і зростання зернистості, що часто спостерігається при низькій освітленості. У інфрачервоному каналі SNR зменшується при зниженні теплового контрасту, нестабільному фоні або підвищенні шуму детектора. У всіх цих випадках низьке значення SNR прямо вказує на те, що канал перестає відображати структуру сцени достовірно. Таким чином, SNR є природним критерієм, який відображає здатність кожного спектрального каналу передавати корисну інформацію у конкретних умовах.

Важливо також підкреслити, що SNR дозволяє побудувати не фіксований, а динамічний механізм управління вкладом каналів. Якщо локально VIS має високу структурність і низький рівень шуму, його внесок повинен бути переважаючим; якщо сцена затемнена, і SNR VIS суттєво падає, система автоматично збільшує роль IR-каналу. Такий підхід відрізняється від класичних методів ф'южну, у яких ваги задаються наперед або визначаються глобальними евристичними правилами. Використання SNR перетворює алгоритм на адаптивну модель, яка реагує на фактичний стан сенсорних даних, забезпечуючи стабільність і високу інформативність інтегрованого зображення. У контексті технічного зору це є принциповою перевагою, оскільки реальні сценарії завжди містять змінні умови, які жоден фіксований метод не здатен належно врахувати.

На основі викладених міркувань задачу комплексування можна сформулювати як процес побудови інтегрованого зображення, у якому кожна локальна область сцени отримує свій внесок від VIS та IR залежно від їхньої миттєвої інформативності, оціненої через SNR та інші локальні характеристики. Вхідною інформацією є два вирівняні й нормалізовані канали, для яких необхідно обчислити показники якості, зокрема локальний рівень сигналу, рівень шуму та

ступінь структурованості даних. Далі ці характеристики оцінюються у спільному просторі, що дозволяє визначити, який із каналів у даній області зберігає більше корисної інформації.

Формально це означає, що для кожного фрагмента зображення алгоритм повинен ухвалити рішення щодо пропорції внеску VIS та IR. Метою є побудова такого інтегрованого подання, яке одночасно передає структурні елементи сцени та забезпечує стійкість до шумів та деградації каналів. Цей підхід вимагає моделі, здатної не просто механічно поєднувати два канали, а аналізувати їхній поточний стан та адаптувати ваги відповідно до локальної якості даних. У результаті злиття повинно мати чітко збережені контури, підсилені теплові області та мінімальну кількість артефактів.

Таким чином, задача комплексування формулюється як проблема побудови адаптивного відображення, яке визначає ступінь впливу кожної модальності у залежності від поточного значення SNR та інших статистичних показників. Очікуваний результат — інтегроване зображення, оптимальне з точки зору структурності, контрастності та стійкості до шумів, придатне для подальших етапів технічного зору, таких як сегментація, детекція або класифікація об'єктів. Саме така постановка створює основу для подальшої розробки SNR-орієнтованої математичної моделі та механізму прийняття рішень у наступних підрозділах.

### 3.2 Механізм формування локальних та статистичних ознак для VIS та IR каналів

У попередньому підрозділі було окреслено логіку роботи SNR-орієнтованого алгоритму та визначено, що ключовим елементом його ефективності є здатність системи адаптивно оцінювати інформативність кожного каналу. Проте така оцінка неможлива без формування набору ознак, які відображають як глобальний стан сцени, так і локальні характеристики її окремих фрагментів. Саме ці ознаки задають основу для подальшої математичної моделі та механізму прийняття рішень,

визначаючи, наскільки кожен канал (VIS або IR) є придатним для внесення інформації в інтегроване зображення.

Оскільки VIS та IR принципово різняться своєю фізичною природою — видимий канал несе високочастотну текстурну структуру, тоді як інфрачервоний відображає тепловий розподіл з переважанням низьких частот — формування ознак повинно відображати ці властивості. Відповідно, у межах цього підрозділу будуть побудовані два класи характеристик: глобальні статистичні ознаки, що описують загальний рівень сигналу, шуму та інформаційної насиченості кадру, і локальні ознаки, що дозволяють оцінити структуру сцени на рівні обмеженого просторового вікна. Саме така комбінація забезпечує можливість детекції як глобальних змін якості (наприклад, деградації VIS у темряві), так і локальних — зокрема, появи або зникнення об'єкта.

Для початкової демонстрації принципів роботи алгоритму були сформовані синтетичні кадри VIS та IR, що відповідають ідеальним умовам спостереження. Такий підхід дозволяє контролювати розташування та контрастність сигналу, а також рівень шуму, що є важливим для коректного пояснення подальших кроків з обчислення ознак. Використання контрольованих даних на цьому етапі забезпечує можливість зосередитися виключно на механізмах аналізу сцени, не враховуючи зовнішні фактори, притаманні реальним зображенням.

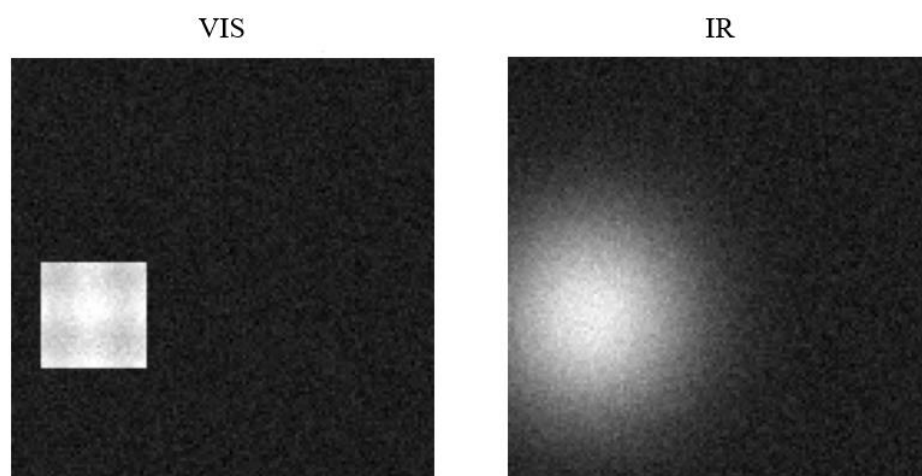


Рисунок 3.2 – Синтетичні VIS та IR кадри для аналізу ознак

На наведених кадрах чітко простежуються ключові особливості обох каналів. VIS-зображення містить структурований квадратний об'єкт із вираженими градієнтами на фоні з відносно рівномірним шумом, що характерно для ідеальних умов зйомки при достатньому освітленні. IR-канал, натомість, демонструє теплову пляму з радіальним спаданням інтенсивності, що є типовим для тепловізійних сенсорів і дозволяє легко вирізнити зону сигналу від фону завдяки природній монотонності теплового профілю. В обох випадках об'єкт чітко відділяється від фонові області, що дає змогу безпомилково простежити методику визначення локальних і глобальних ознак у подальших етапах аналізу.

Автоматичне визначення області сигналу є ключовим елементом побудови SNR-орієнтованої системи, оскільки алгоритм не може спиратися на наперед задану інформацію про положення об'єкта. У реальних умовах зйомки теплові та візуальні структури можуть зміщуватися, змінювати форму, розмиватися або частково перекриватися шумом. Тому було реалізовано процедуру автоматичного пошуку найбільш інформативної ділянки кадру, засновану на локальному аналізі енергії та яскравісних характеристик. Такий підхід дозволяє системі самостійно виокремлювати область сигналу незалежно від її масштабу, інтенсивності чи просторового розташування.

Представлений результат(рис 3.3) демонструє здатність алгоритму адаптивно виявляти сигнальну область у кожному з каналів. У VIS-кадрі квадратна структура легко розпізнається за рахунок різкого переходу від текстури до фону. Для IR-каналу ситуація складніша: тепловий сигнал не має чітких меж і формується як гладка гаусоподібна пляма. Незважаючи на це, методика локального аналізу енергії дозволяє визначити її просторові межі, які автоматично розширюються відповідно до фактичного розміру теплової області. Це наочно демонструє гнучкість алгоритму: розмір і форма виділеного сигналу не фіксовані, а залежать від реальних локальних характеристик кадру.

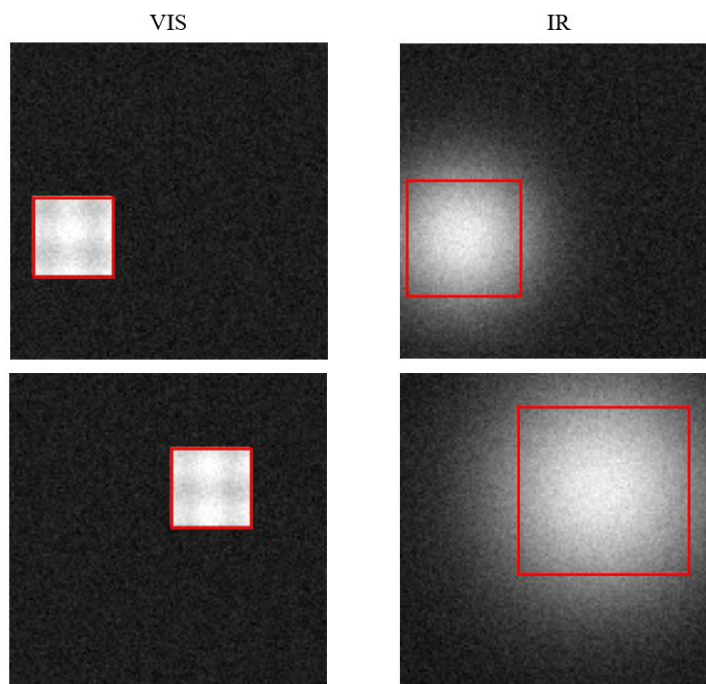


Рисунок 3.3 – Результат автоматичного визначення області сигналу у VIS та IR каналах

Сам принцип роботи полягає в тому, що сцена розбивається на невеликі вікна, для кожного з яких обчислюються локальні метрики — середня інтенсивність, варіаційність, контраст та локальна SNR. Алгоритм обирає ту ділянку, де сумарна інформативність досягає максимуму, та формує сигнал саме на її основі. У реальних кадрах це дає змогу автоматично розпізнавати об'єкти, навіть якщо їх положення та контури неочевидні, а фон містить значний рівень завад.

Щоб продемонструвати динаміку розподілу інтенсивності та проілюструвати, наскільки чітко сигнал відокремлюється від фону, побудовано відповідну тривимірну діаграму. Вона дозволяє інтерпретувати формування сигналу як локальний максимум енергії на поверхні яскравісної карти. Для VIS це пік квадратної структури, тоді як для IR — широкий, плавний максимум теплової плями. Така візуалізація слугує основою для подальшої інтерпретації локальних ознак, оскільки наочно показує, які ділянки сцени є інформативними, а які — шумовими.

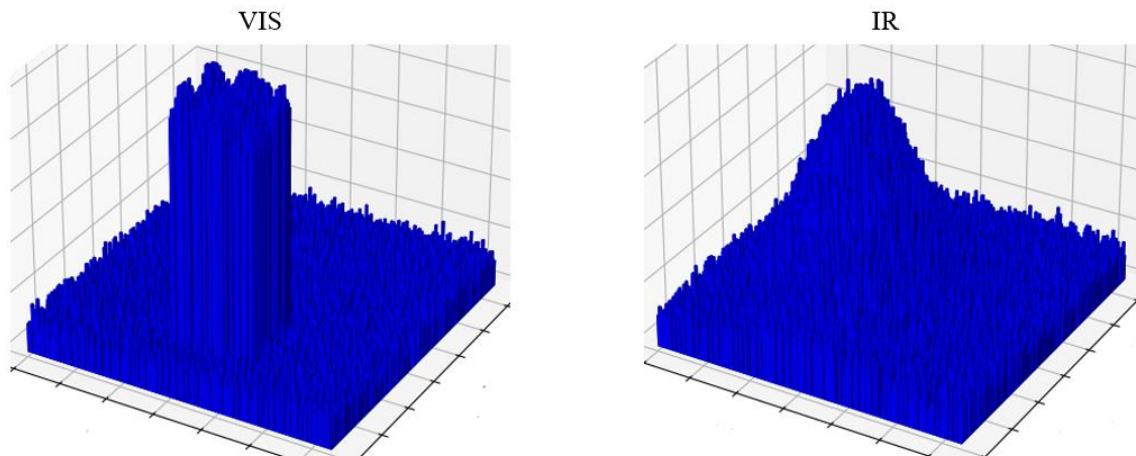


Рисунок 3.4 – Тривимірне представлення інтенсивності(сигнальної складової) VIS та IR зображень

З іншого боку, аналіз гістограм інтенсивності дає змогу оцінити глобальний розподіл яскравостей у кадрі. Для VIS каналу характерне чітке відділення значень, що належать об'єкту, від шумового фону, тоді як в IR-каналі тепловий сигнал формує широкий діапазон інтенсивностей, плавно переходячи у фон. Гістограма дозволяє кількісно оцінити ступінь змішування сигналу з шумом, що є критичним для подальшого визначення глобальних ознак, таких як  $\mu_{\text{signal}}$ ,  $\mu_{\text{bg}}$ ,  $\sigma_{\text{bg}}$  та глобальна SNR. Вона також дає можливість продемонструвати контрастність каналів, що безпосередньо впливає на якість подальшого комплексування.

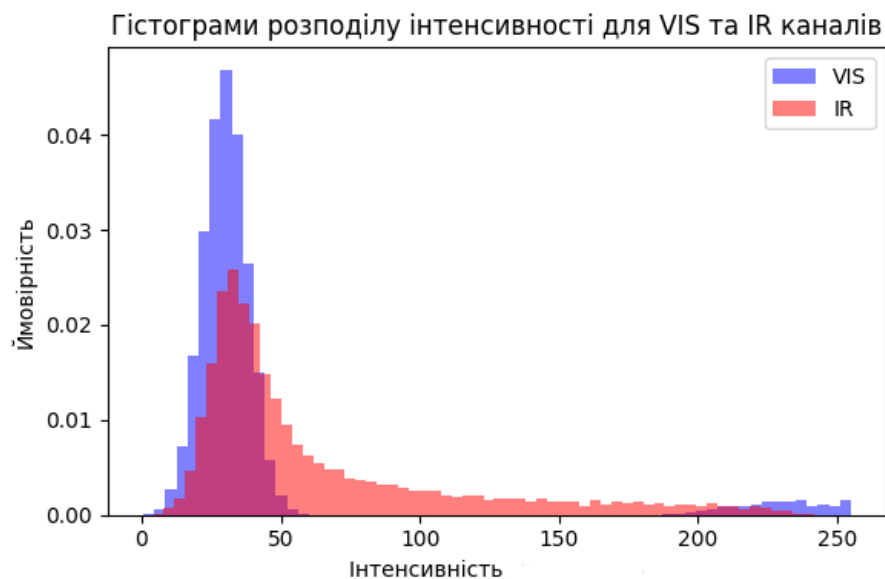


Рисунок 3.5 – Гістограма розподілу інтенсивностей для VIS та IR каналів

Після аналізу просторової структури сигналу необхідно перейти до формування кількісних характеристик зображення, які будуть використані у SNR-орієнтованому алгоритмі. Усі обчислені ознаки можна поділити на два рівні: глобальні, що описують сцену в цілому, та локальні, які характеризують поведінку інтенсивності всередині невеликих областей. SNR обчислювався відповідно до виразу, наведеного у розділі 1.3 (формула (1.х)), де рівень сигналу визначався через середнє значення пікселів у виділеній сигнальній області, а шум — через стандартне відхилення фону. Глобальні метрики дозволяють робити висновки про загальний стан каналу, тоді як локальні забезпечують здатність моделі реагувати на дрібномасштабні зміни структури сигналу і шуму.

Попри те, що в алгоритмі використовується повний набір ознак — контраст, варіаційність, локальні енергетичні показники, ентропія — саме SNR виступає центральною, «ключовою» характеристикою каналу. Це обумовлено тим, що SNR безпосередньо відображає різницю між сигнальною та шумовою складовими й дозволяє кількісно оцінити інформативність фрагмента. Такий показник коректно працює як для VIS, де контури яскраво виражені, так і для IR, де теплові структури згладжені та широкі. У подальших розділах глобальні та локальні SNR значення виступають основою для формування ваг комплексування та для навчання моделі прийняття рішень.

VIS-канал:		
$\mu_{\text{signal}}$	(середнє значення інтенсивності в області сигналу)	= 229.66
$\mu_{\text{bg}}$	(середнє значення інтенсивності фону)	= 29.97
$\sigma_{\text{bg}}$	(стандартне відхилення інтенсивності фону)	= 7.98
SNR	(відношення сигнал/шум для кадру)	= 25.02
Var	(дисперсія інтенсивності всього кадру)	= 2411.37
Std	(середньоквадратичне відхилення інтенсивності)	= 49.11
Entropy	(ентропія сцени, біти Шеннона)	= 5.428
IR-канал:		
$\mu_{\text{signal}}$	(середнє значення інтенсивності в області сигналу)	= 167.70
$\mu_{\text{bg}}$	(середнє значення інтенсивності фону)	= 48.06
$\sigma_{\text{bg}}$	(стандартне відхилення інтенсивності фону)	= 25.17
SNR	(відношення сигнал/шум для кадру)	= 4.75
Var	(дисперсія інтенсивності всього кадру)	= 2376.12
Std	(середньоквадратичне відхилення інтенсивності)	= 48.75
Entropy	(ентропія сцени, біти Шеннона)	= 6.971

Рисунок 3.6 – Глобальні статистичні ознаки

Глобальні ознаки наочно демонструють відмінності між каналами: VIS характеризується високою різницею між середнім рівнем сигналу та фону, тоді як IR має плавний, розтягнутий розподіл інтенсивності, що впливає на підвищене значення ентропії та дисперсії. Разом ці параметри дають повне уявлення про інформативність сцени:  $\mu_{\text{signal}}$  та  $\mu_{\text{bg}}$  показують контрастність об'єкта,  $\sigma_{\text{bg}}$  та  $\text{Var}$  характеризують рівень шуму, а ентропія — ступінь структурованості та невизначеності даних. Саме ці характеристики дозволяють алгоритму формувати кількісну оцінку стану кожного каналу й надалі використовувати їх для адаптивного вибору ваг комплексування.

Формування глобальних ознак дозволяє охарактеризувати канал у цілому, проте для подальшої побудови адаптивного SNR-орієнтованого алгоритму цього недостатньо. У реальних умовах інформативність зображення нерівномірно розподілена: об'єкт може займати лише локальну ділянку кадру, тоді як інші області містять переважно шум або малоінформативний фон. Саме тому виникає потреба у розбитті сцени на локальні вікна, що дозволяє досліджувати просторові зміни статистичних характеристик та визначати ті ділянки, які роблять суттєвий внесок у якість сигналу. Такий підхід є стандартним у методах локального аналізу (sliding-window) й особливо корисним для подальшої роботи дерева рішень, яке оперує ознаками кожної малої області.

Логіка цього підходу полягає у тому, що локальне вікно дає змогу оцінити SNR на обмеженій ділянці сцени, де співвідношення «сигнал/шум» може істотно відрізнятись від глобального. Якщо об'єкт займає невелику частину кадру, то глобальні метрики можуть занижувати справжню інформативність каналу, тоді як локальні ознаки дозволяють точніше виділити структурні фрагменти та теплові контури. Крім того, локальна оцінка дисперсії, ентропії та контрасту робить алгоритм стійким до просторових артефактів, градієнтів фону, неоднорідного освітлення та локальних шумових сплесків — явищ, характерних для VIS та IR модальностей у технічних системах спостереження.

Для демонстрації роботи цього механізму сцена була розбита на рівномірну сітку вікон фіксованого розміру, після чого для одного з них було обчислено

повний набір локальних ознак. На рисунку 3.7 наведено приклад виділення одного вікна у синтетичному VIS-кадрі.

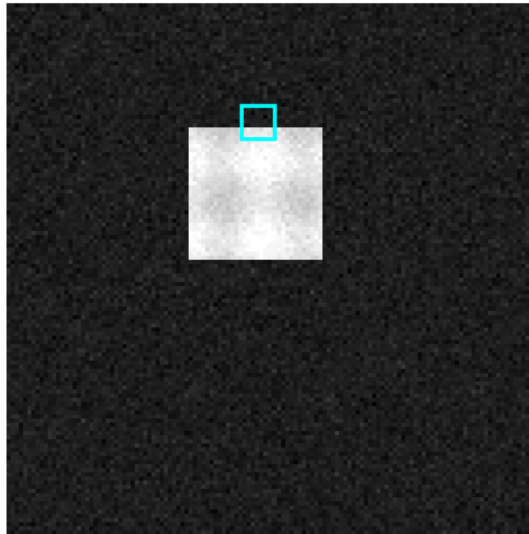


Рисунок 3.7 – Виділене локальне вікно в синтетичному VIS-кадрі

Для обраного вікна алгоритм визначає локальні статистичні характеристики(рис. 3.8):

```

Локальні ознаки для вибраного вікна (VIS):
mu_patch(середнє значення інтенсивності локального вікна): 80.6588
var_patch(дисперсія інтенсивності у вікні): 7494.8223
std_patch(середньоквадратичне відхилення яскравості у вікні): 86.5726
entropy_patch_bits(ентропія локального фрагмента у бітах): 5.0603
contrast(локальний контраст між вікном та фоном): 38.3532
local_SNR(локальне відношення сигнал/шум для вибраного фрагмента): 0.7852

```

Рисунок 3.8 – Локальні статистичні ознаки для одного фрагмента кадру

У реальній роботі алгоритму обчислення таких ознак виконується для кожного вікна сцени, а не лише для одного демонстраційного фрагмента. Усі локальні параметри записуються у внутрішній набір ознак, який у подальшому використовується деревом рішень для визначення ступеня інформативності VIS та IR каналів у різних частинах кадру. Таким чином, модель отримує не лише загальну характеристику сцени, а й її детальну просторову структуру, що є ключовим для адаптивного комплексування.

### 3.3 Математична модель SNR-орієнтованого злиття зображень

Після формування глобальних і локальних ознак основним завданням стає перехід від опису якості каналів до керованого процесу комплексування, у якому внесок VIS та IR визначається не фіксовано, а за поточним станом даних. Для систем технічного зору це принципово, оскільки у реальних умовах параметри сцени змінюються між кадрами: зниження освітленості та наростання шуму у видимому діапазоні призводять до деградації контурів, тоді як інфрачервоний канал частіше зберігає структуру об'єкта за рахунок теплової складової. За цих умов необхідний механізм, який дозволяє кількісно оцінити інформативність кожного каналу та на цій основі регулювати їх вплив у сформованому зображенні.

Як інтегральний показник якості використовується відношення сигнал/шум, обране у вигляді, узгодженому з підходом попередніх розділів: SNR визначається через різницю середнього рівня сигналу та фону, нормовану на стандартне відхилення фону. Такий запис дозволяє інтерпретувати SNR як міру роздільності об'єкта відносно шумового середовища й безпосередньо відповідає задачі виявлення у СТЗ: при малому SNR об'єкт “зливається” з фоном і навіть наявність високої середньої яскравості не гарантує надійної детекції. Отримані значення SNR для VIS та IR надалі використовуються як керуюча інформація, що визначає, який канал має домінувати у конкретному режимі спостереження.

Для перевірки працездатності моделі комплексування застосовано тестовий сценарій сильного погіршення умов, що імітує затемнення та одночасне зростання шуму. У межах цього експерименту формуються дві пари кадрів: базові (умовно “ідеальні”) VIS та IR і відповідні затемнені/зашумлені варіанти, що відтворюють деградацію каналу у складних умовах. Важливою частиною процедури є автоматичне визначення області сигналу на кожному кадрі: алгоритм не використовує апріорних знань про положення чи розмір об'єкта, а виділяє зону з найбільш вираженою енергетичною складовою за результатами порогової сегментації та вибору найбільшої зв'язної компоненти. Саме ця область надалі інтерпретується як сигнал, а решта кадру — як фон, що забезпечує узгоджене

обчислення глобальних параметрів і створює основу для керування злиттям у наступних етапах.

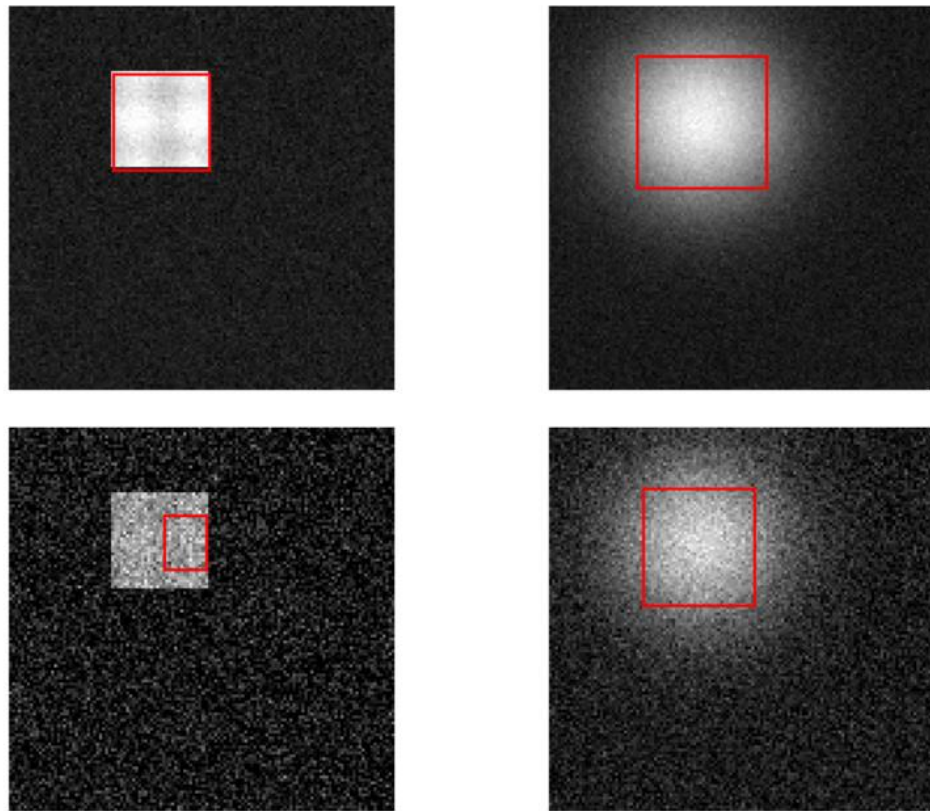


Рисунок 3.9 – VIS/IR кадри з автоматично виділеною областю сигналу

За результатами пошуку сигналу на базових кадрах область об’єкта визначається стабільно, оскільки контраст між сигналом і фоном є значним, а шум не домінує у структурі зображення. У затемненому VIS-кадрі ситуація змінюється: зменшення яскравості та підсилення шуму призводять до втрати частини слабконтрастних фрагментів сигналу, тому алгоритм виділяє не весь об’єкт повністю, а найбільш інформативну його частину: ділянку з максимальними локальними значеннями, що залишились відокремленими від фону після порогоування. Така поведінка є очікуваною для умов низького SNR: у VIS-каналі сигнал “руйнується” шумом нерівномірно, і з точки зору задачі оцінювання якості найбільш суттєвим стає саме збережений фрагмент, який ще можна статистично відрізнити від шумового поля.

Для IR-кадру навіть у затемненому стані зберігається відносно компактна та енергетично виражена теплова пляма, тому область сигналу визначається більш повно і ближче до реальної геометрії об'єкта. Це відображає фізичну специфіку інфрачервоного каналу: тепловий контраст зникає повільніше, ніж візуальна структура у VIS при падінні освітленості. У контексті подальшого комплексування це має прямий наслідок: статистично обґрунтоване визначення сигналу на затемнених кадрах дозволяє отримати коректні значення SNR для кожного каналу та перейти до наступного кроку: побудови керованих ваг і формування комплексованого зображення, у якому домінування IR в умовах деградації VIS є результатом формалізованої оцінки інформативності каналів, а не застосуванням емпіричних правил.

Після етапу автоматичного виділення сигнальної області та обчислення глобальних параметрів якості для кожного каналу виконується безпосереднє комплексування. Його ідея полягає у тому, що інтегроване зображення має формуватися як результат керованого поєднання VIS та IR, де внесок кожного каналу визначається не фіксовано, а на основі оціненої інформативності. У межах запропонованого підходу така інформативність задається значенням SNR, обчисленим за формулою, наведеною у попередніх розділах. Середній рівень сигналу в межах знайденої області порівнюється із середнім фоном, а нормування виконується на стандартне відхилення фону. Саме цей показник використовується як “керуючий” для подальшого злиття.

Формування ваг комплексування реалізується через нормоване співвідношення SNR двох каналів(1.7). Нехай  $SNR_{VIS}$  та  $SNR_{IR}$  — глобальні оцінки якості кадрів VIS та IR відповідно, отримані на основі автоматично визначених областей сигналу. Тоді вага видимого каналу задається як:

$$\omega_{VIS} = \frac{SNR_{VIS}}{SNR_{VIS} + SNR_{IR}},$$

а вага інфрачервоного каналу визначається доповненням:

$$\omega_{\text{IR}} = 1 - \omega_{\text{VIS}}.$$

Такий вибір забезпечує інтерпретованість моделі: якщо SNR одного з каналів зменшується (наприклад, VIS при сильному затемненні), його вага автоматично падає, а внесок іншого каналу зростає. Далі комплексоване зображення формується як піксельне зважене поєднання(1.6):

$$F(x, y) = \omega_{\text{VIS}}(x, y)I_{\text{VIS}}(x, y) + \omega_{\text{IR}}(x, y)I_{\text{IR}}(x, y),$$

де  $I_{\text{VIS}}(x, y)$  та  $I_{\text{IR}}(x, y)$  — значення інтенсивності у відповідних каналах;

$F(x, y)$  — результат злиття у кожній точці кадру.

Отримане зображення потім аналізується аналогічно до окремих каналів: повторно визначається область сигналу та розраховуються глобальні характеристики, що дозволяє оцінити, чи справді комплексування підвищило віддільність об'єкта від фону.

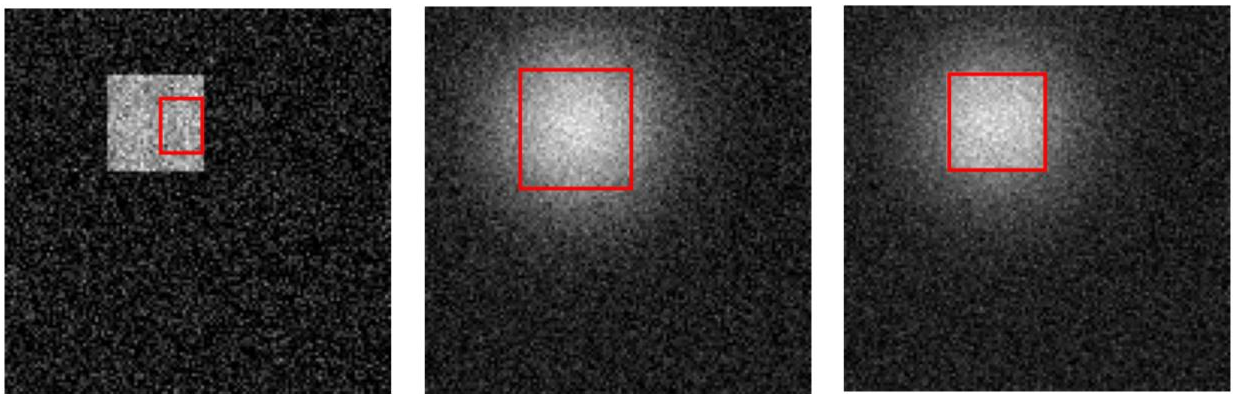


Рисунок 3.10 – Результати SNR-керованого комплексування затемнених кадрів

Представлені результати демонструють узгоджену поведінку моделі з фізичними очікуваннями для умов сильного затемнення. Для затемненого VIS-кадру отримано нижчу оцінку якості, що пов'язано з домінуванням шуму та частковим “розпадом” структурної інформації: алгоритм виявляє лише найбільш енергетично виражений фрагмент у зоні об'єкта, а обчислене значення  $\text{SNR}_{\text{VIS}}$  залишається на помірному рівні. Для IR кадру значення SNR є вищим, оскільки теплова пляма зберігає компактну та контрастну структуру навіть за зростання

шумів, а автоматично знайдена область сигналу краще відповідає реальному положенню найбільш інформативної частини об'єкта.

У результаті нормування SNR модель формує ваги  $\omega_{\text{VIS}}=0.46$  та  $\omega_{\text{IR}}=0.54$ , тобто в умовах деградації видимого каналу інфрачервоний внесок стає домінантним, але VIS не “відключається” повністю. Це важливо з погляду практики: навіть у темних сценах видимий канал може містити частину корисних деталей (локальні текстури, градієнтні переходи), тому часткове збереження його впливу дозволяє уникнути надмірного “розмиття” структури при повному домінуванні IR.

Комплексоване зображення демонструє покращення інтегральної оцінки якості: значення SNR для комплексного зображення є вищим за SNR кожного з каналів окремо, що означає підвищення статистичної віддільності сигналу від фону при збереженні узгодженої структури об'єкта. Додатково зміна дисперсії та ентропії у комплексованому кадрі відображає компроміс між стабілізацією шуму та збереженням інформаційного різноманіття: інтеграція знижує “хаотичність”, але не зводить сцену до надто гладкої структури IR, що є бажаним для подальших етапів аналізу у СТЗ (детекція та локалізація).

Таблиця 3.1 – Глобальні ознаки кадрів

Кадр	$\mu_{\text{signal}}$	$\mu_{\text{bg}}$	$\sigma_{\text{bg}}$	SNR	VAR	Std	Entropy (bits)
VIS ідеальний	228.818	30.008	8.008	24.826	2391.941	48.907	5.439
IR ідеальний	168.859	45.466	24.348	5.068	2209.955	47.010	6.747
VIS зашумлений	56.871	11.666	13.894	3.253	223.468	14.949	5.864
IR зашумлений	90.359	24.318	16.974	3.891	639.746	25.293	7.059
Комплексоване	78.529	18.275	11.055	5.450	333.466	18.261	6.869

Зведення результатів у таблиці 3.1 дозволяє зафіксувати дві принципові закономірності. По-перше, у базовому режимі VIS демонструє дуже високий SNR за рахунок чіткої структурної різниці між об'єктом і фоном, тоді як IR має помірніший SNR через гладкий тепловий профіль і більшу «розмитість» меж. По-друге, після затемнення VIS втрачає запас якості швидше, тоді як IR зберігає

більшу стабільність, що прямо відображається у співвідношенні. Комплексування на основі SNR-ваг переводить систему у режим, де домінуючою стає більш надійна модальність, але внесок другого каналу не обнуляється повністю, що важливо для збереження тих ознак, які залишаються інформативними навіть у деградованому VIS (локальні перепади, залишкові контури або текстурні фрагменти).

Щоб наочно показати зміну сигнальної складової не лише чисельно, а й у просторовому сенсі, додатково розглянуто 3D-представлення інтенсивностей кадрів, де область сигналу виділяється окремим кольором. Це дозволяє інтерпретувати комплексування як перетворення «енергетичного рельєфу» сцени: оцінити форму піків, рівень фону та ступінь відриву сигналу від шумових флуктуацій.

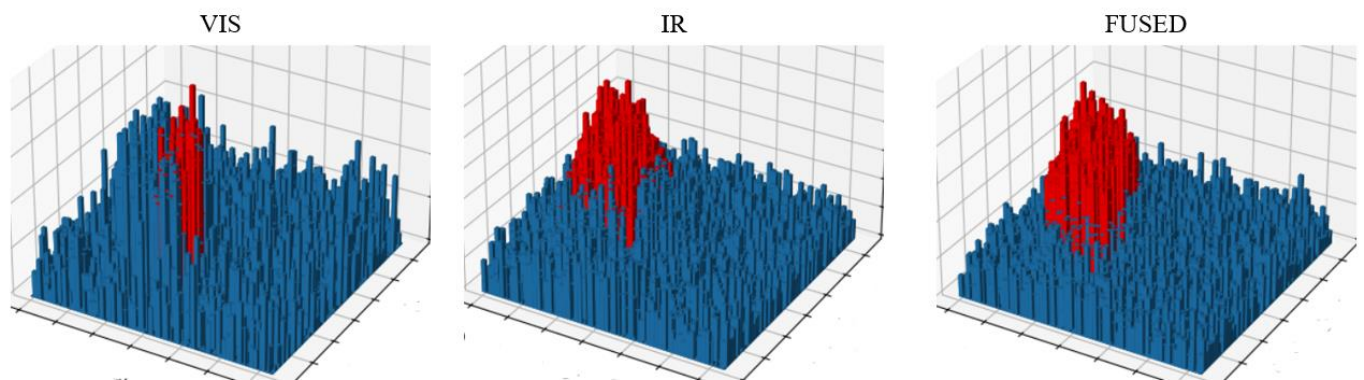


Рисунок 3.11 – Тривимірне представлення інтенсивності(сигнальної складової)  
VIS, IR та FUSED зображень

За результатами можна чітко визначити, що 3D-діаграм комплексоване зображення демонструє більш виражений і компактний «пік» у сигнальній області на фоні менш домінуючих випадкових коливань, ніж у VIS, а також не втрачає локальної структури повністю, як це могло б статись при простому копіюванні IR. У поєднанні з приростом SNR (табл. 3.1) це є практичним підтвердженням того, що SNR-керовані ваги не лише формально перерозподіляють внесок каналів, а реально підсилюють придатність результату для подальших етапів аналізу сцени.

### 3.4 SNR-орієнтована модель прийняття рішень: навчання, інтерпретація та оцінювання алгоритму

Для реалізації автоматичної реакції системи на погіршення якості зображення було сформовано навчальну базу даних синтетичних сцен, у яких змінювались як просторові характеристики об'єкта (випадкове положення сигналу на кадрі), так і статистичні параметри фону та сигналу. Кожна сцена містила послідовність кадрів, що моделює роботу системи в часі: інтенсивність видимого сигналу деградувала нелінійно, а рівень шуму зростав, імітуючи реальні умови зйомки при зниженні освітленості. Для кожного кадру автоматично визначались найбільш інформативні області(сигнал), як це було виконано у підрозділі 3.2 (рис 3.3) та фон, після чого обчислювались ключові глобальні та структурні ознаки. Таким чином, система отримала узагальнений набір прикладів «якість кадру – рішення», без використання апріорної інформації про координати та розмір сигналу.

Навчання моделі прийняття рішення виконувалось у вигляді дерева рішень, оскільки цей підхід забезпечує інтерпретованість та дозволяє отримати явні правила виду «якщо-то». На першому етапі формувались три класи якості для VIS – каналу (якісний/перехідний/потребує комплексування) на основі автоматичного визначених порогів SNR, визначено на дата сеті зі сцен, та додаткових ознак, після чого дерево навчалось розділяти ці класи. При цьому розбиття даних на навчальну та тестовому вибірці виконувалось по сценах, щоб уникнути витоку інформації між кадрами однієї сцени та забезпечити коректну перевірку здатності моделі узагальнювати на нові умови.

Отримане дерево демонструє, що SNR виступає основною ознакою, однак у граничних/спірних випадках (перехідна зона) модель додатково використовує другорядні статистичні та структурні параметри. Це дозволяє приймати більш стійке рішення, коли одного лише SNR недостатньо через локальні неоднорідності шуму або слабкий контраст між сигналом і фоном. Таким чином, дерево фактично

реалізує логіку «порогового контролю якості», але з адаптацією до різних ситуацій кадру за рахунок додаткових ознак.

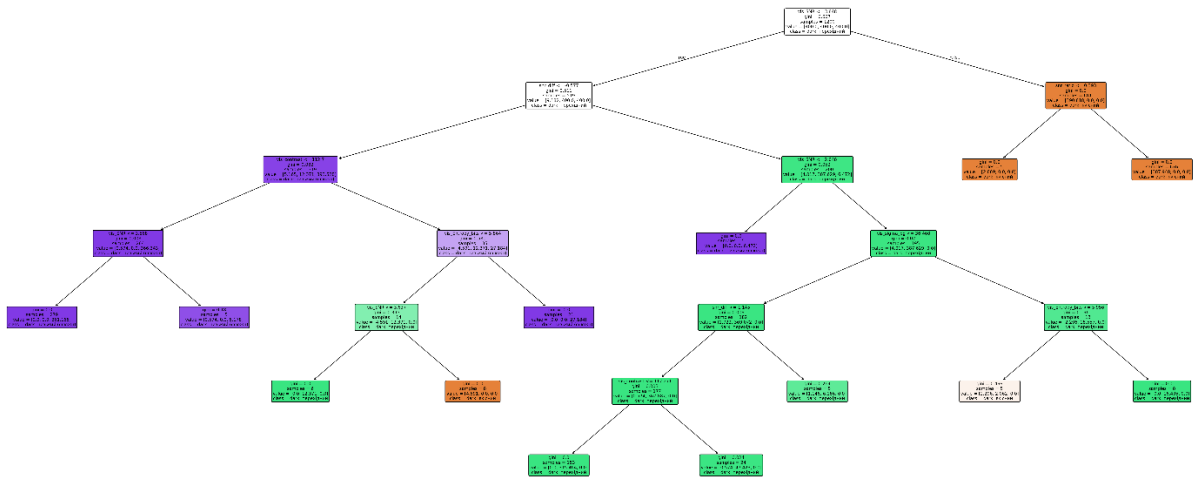


Рисунок 3.12 – Дерево рішень

Наступним кроком виконано перевірку стійкості підходу в іншому типі деградації — моделюванні туману, де погіршення якості зображення зумовлене не тільки загальним зниженням контрасту, а й нерівномірним «перекриттям» частини сигналу та фону. Для цього була сформована окрема база сцен з аналогічною структурою (послідовності кадрів у часі), після чого модель була перенавчена/оновлена для розпізнавання класів якості у нових умовах.

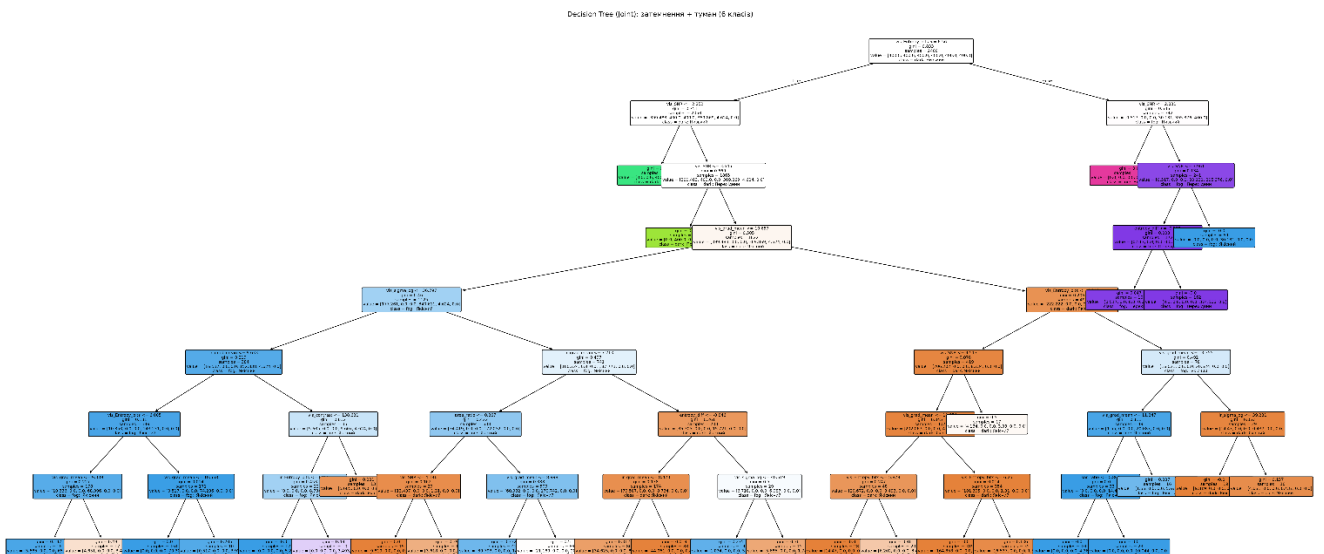


Рисунок 3.13 – Дерево рішень (адапована модель)

На рис. 3.13 наведено узагальнене (спільне) дерево рішень, сформоване на об'єднаній вибірці для двох сценаріїв деградації — затемнення та туману. Структурно дерево працює ієрархічно: на перших розгалуженнях воно використовує комбінацію ознак (SNR, контраст сигнал–фон,  $\sigma_{bg}$ , ентропія та градієнтні характеристики), щоб ідентифікувати домінуючу умову погіршення, оскільки різні фізичні причини деградації по-різному змінюють статистику сцени. Після цього застосовується відповідна “гілка” правил, де рішення щодо класу якості VIS приймається за головним індикатором SNR у видимому каналі, а у прикордонних (“сірих”) значеннях уточнюється додатковими ознаками, що підвищує стабільність класифікації та зменшує хибні переходи між режимами.

Далі для підтвердження цієї логіки наведемо демонстративні графіки, які показують динаміку деградації SNR у часі, розподіл SNR по вибірці та виділення трьох зон якості. Вони потрібні, щоб наочно узгодити порогові області, які “вивчив” класифікатор, із реальною поведінкою ознак під час затемнення та туману, і показати момент, коли VIS-канал переходить у режим, де комплексування стає доцільним.

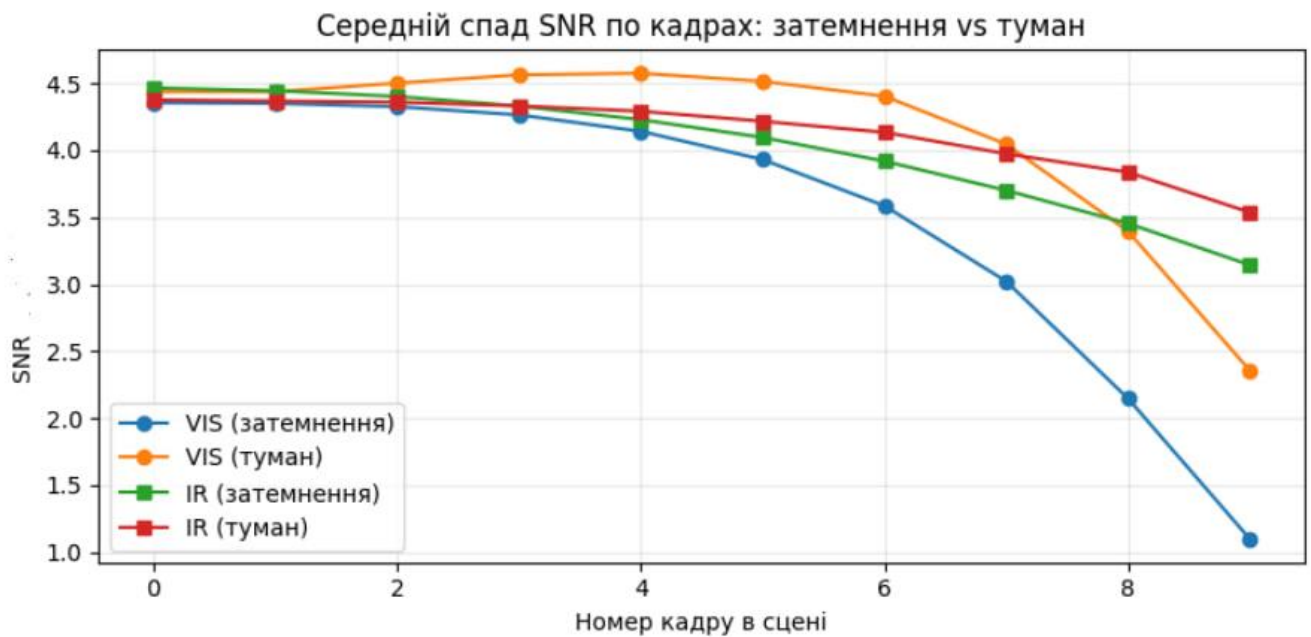


Рисунок 3.13 – Демонстрація деградації SNR за різних умов

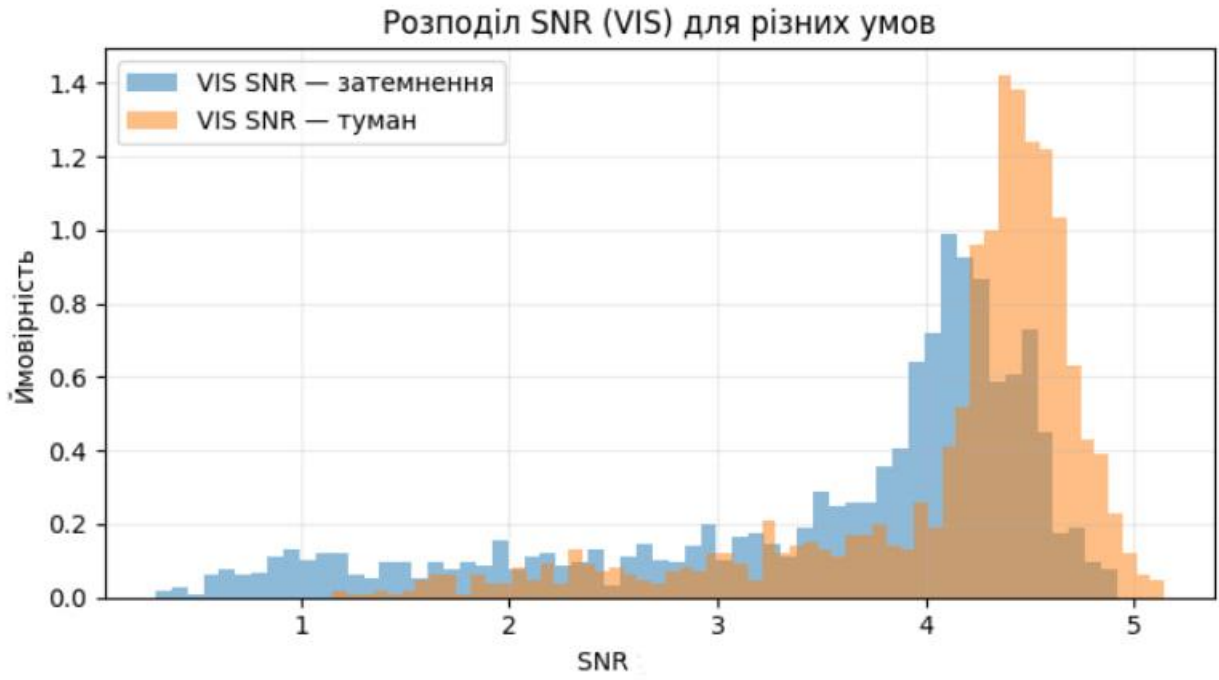


Рисунок 3.14 – Розподіл SNR у видимому спектрі за різних умов

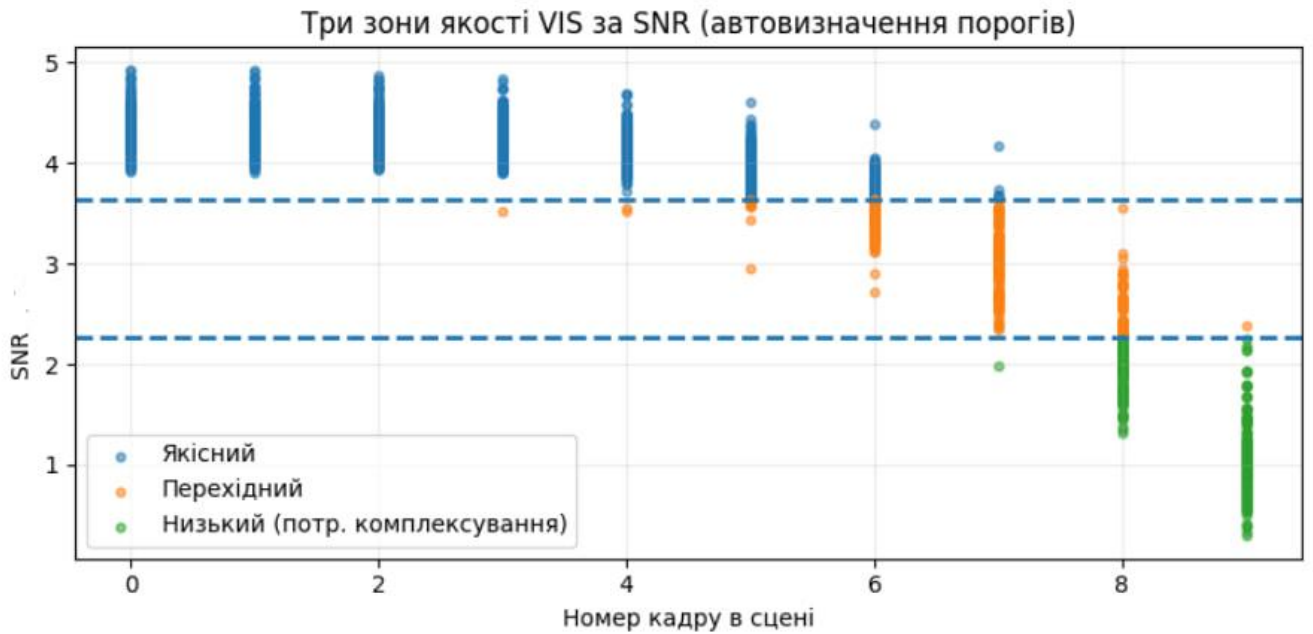


Рисунок 3.15 – Розподіл на 3 порогові рівні SNR для дерева рішень

## ВИСНОВКИ

У кваліфікаційній роботі виконано розробку та дослідження методів комплексування зображень у системах технічного зору (СТЗ), що працюють у змінних умовах спостереження. Основна увага приділялась підвищенню стійкості роботи системи під час деградації зображень у видимому діапазоні та пошуку ефективного критерію для автоматичного перемикання на інфрачервоний канал.

У результаті дослідження проведено аналіз існуючих підходів до комплексування VIS–IR зображень, визначено їхні переваги й обмеження. Розроблено модель оцінки якості кадру за відношенням сигнал/шум (SNR), що дозволяє кількісно оцінювати інформативність сцени та приймати рішення щодо активації IR-каналу.

Було сформовано методику формування локальних і глобальних ознак кадру, на основі яких побудовано дерево рішень для класифікації якості зображення. Навчання здійснювалося на синтетичних сценах із різними умовами деградації — затемненням і туманом. Під час тестування алгоритм показав здатність адаптивно розпізнавати зони низької інформативності та коректно визначати момент перемикання каналів.

Отримані результати підтверджують можливість практичного застосування розробленого методу у системах технічного зору для моніторингу, охоронних комплексів, промислових і робототехнічних платформ. Запропонований підхід дозволяє зменшити вплив погодних і освітлювальних факторів, підвищити надійність виявлення об'єктів і стабільність функціонування СТЗ у реальних умовах.

Перспективи подальших досліджень пов'язані з розширенням бази даних навчальних сцен, впровадженням глибоких нейронних моделей для багатоканального аналізу та інтеграцією розробленого методу у реальні апаратно-програмні комплекси технічного зору.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Richard Szeliski: *Computer Vision: Algorithms and Applications*. 2nd Edition / Springer, 2022.
2. Очеретін В. Д., Стрілкова Т. О. *Методи комплексування в системах відеоспостереження*. – Харків : ХНУРЕ, 2025.
3. Holst G. C. *Electro-Optical Imaging System Performance*. 4th ed. Bellingham: SPIE Press, 2008. 540 p.
4. Khaleghi B., Khamis A., Karray F. O., Razavi S. *Multisensor data fusion: A review of the state-of-the-art* // *Information Fusion*. – 2013. – Vol. 14, No. 1. – P. 28–44.
5. Li S., Kang X., Fang L., Hu J., Yin H. *Multi-Sensor Image Fusion: A Survey of the State of the Art* // *Journal of Sensors*. – 2021.
6. Прудіус І. Н., Лазько Л. В., Семенов С. О. *Багаторівневе комплексування графічної інформації в системах дистанційного зондування* // *Вісник Нац. ун-ту «Львівська політехніка»*. *Радіоелектроніка та телекомунікації*. – 2007.
7. Li S., Kang X., Fang L., Hu J., Yin H. *Pixel-level image fusion: A survey of the state of the art* // *Information Fusion*. – 2017.
8. Pohl C., Van Genderen J. L. *Review article: multisensor image fusion in remote sensing: concepts, methods and applications* // *International Journal of Remote Sensing*. – 1998.
9. Li H., Manjunath B., Mitra S. *Multi-sensor Image Fusion: A Survey of the State of the Art* // *SCIRP Journals*. – 2021.
10. Hasinoff S. W., Geiss R., Krishnan D., Adams A., Ramanan D., Carroll R., Levoy M. *Burst photography for high dynamic range and low-light imaging on mobile cameras* // *ACM Transactions on Graphics (TOG)*. – 2016. – Vol. 35, No. 6. – P. 1–12.
11. O’Byrne M., Dahyot R. *Multi-frame image enhancement for low-light imaging* // *IEEE Transactions on Image Processing*. – 2014.
12. Zhang X., Demiris Y. *Visible and infrared image fusion using deep learning* // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2023.

13. Liu J., Wu G., Liu Z., Wang D., Jiang Z., Ma L., Zhong W., Fan X., Liu R. Infrared and Visible Image Fusion: From Data Compatibility to Task Adaptation // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2025. – Vol. 47, № 15. – DOI / arXiv: 2501.10761.
14. Li H., Wu X.-J. DenseFuse: A fusion approach to infrared and visible images // IEEE Transactions on Image Processing. – 2018. – Vol. 28, No. 5.
15. Liu J., Zhang L., Zeng X., Liu W., Zhang J. MATCNN: Infrared and Visible Image Fusion Method Based on Multi-scale CNN with Attention Transformer // IEEE Transactions on Instrumentation and Measurement. – 2025. – DOI / arXiv: 2502.01959.
16. Gonzalez R.C., Woods R.E. Digital Image Processing. — 4th ed. — Pearson, 2018.
17. Sonka M., Hlavac V., Boyle R. Image Processing, Analysis, and Machine Vision. — Cengage Learning, 2014.
18. Ma J., Liang P., Yu W., Chen C., Guo X., Wu J., Jiang J. Infrared and visible image fusion via detail-preserving adversarial learning // Information Fusion. – 2020. – Vol. 54. – P. 85–98.
19. S. Winkler, Image Quality Assessment: From Error Visibility to Structural Similarity, Springer – 2020.
20. Gao F., Qi J., Zhang X. Kalman-Filter-Based Multi-Spectral Fusion for Robust Target Perception in Adverse Conditions // IEEE Sensors Journal. – 2021. – Vol. 21, No. 4.
21. Zhang Q., Liu J., Sun Y., Zhang J. Multisensor Image Fusion Algorithm Based on Improved Kalman Filter // Sensors. – 2019.
22. Li S., Zhan Y. Multi-sensor Kalman filtering for dynamic scene reconstruction: A review and new insights // Information Fusion. – 2020. – Vol. 56.
23. Simon D. Optimal State Estimation: Kalman,  $H_\infty$ , and Nonlinear Approaches. – 2nd ed. – Hoboken: Wiley-IEEE Press, 2006. – 528 p.
24. Li Y., Wang S. Multimodal Data Fusion: A Review // Information Fusion. – 2021. – Vol. 57. – P. 1–30.

25. Zhang Y., Liu X., Sun X. Low-light Image Fusion: A Survey // Digital Signal Processing. – 2020.
26. Li H., Manjunath B. S., Mitra S. K. Multisensor Image Fusion Using the Wavelet Transform // IEEE Transactions on Instrumentation and Measurement. – 2013. – Vol. 62, No. 6.
27. Zhang J., Zhao X., Liu Y. Wavelet Transform for Image Fusion: A Review // Multimedia Tools and Applications. – 2021. – Vol. 80.
28. Gao S., Ma J. Edge-preserving multispectral fusion using Laplacian operators // Journal of Electronic Imaging. – 2018.